aBIOTECH

REVIEW

Genome recoding strategies to improve cellular properties: mechanisms and advances

Tanya Singh[1] , Sudesh Kumar Yadav[2] , Alexander Vainstein[3] ,
Vinay Kumar[1✉]

[1] Department of Botany, School of Basic Sciences, Central University of Punjab, Bathinda 151001, India
[2] Center of Innovative and Applied Biotechnology, Mohali 140306, India
[3] Institute of Plant Sciences and Genetics in Agriculture, The Hebrew University of Jerusalem, Rehovot, Israel

Abstract The genetic code, once believed to be universal and immutable, is now known to contain many variations and is not quite universal. The basis for genome recoding strategy is genetic code variation that can be harnessed to improve cellular properties. Thus, genome recoding is a promising strategy for the enhancement of genome flexibility, allowing for novel functions that are not commonly documented in the organism in its natural environment. Here, the basic concept of genetic code and associated mechanisms for the generation of genetic codon variants, including biased codon usage, codon reassignment, and ambiguous decoding, are extensively discussed. Knowledge of the concept of natural genetic code expansion is also detailed. The generation of recoded organisms and associated mechanisms with basic targeting components, including aminoacyl-tRNA synthetase–tRNA pairs, elongation factor EF-Tu and ribosomes, are highlighted for a comprehensive understanding of this concept. The research associated with the generation of diverse recoded organisms is also discussed. The success of genome recoding in diverse multicellular organisms offers a platform for expanding protein chemistry at the biochemical level with non-canonical amino acids, genetically isolating the synthetic organisms from the natural ones, and fighting viruses, including SARS-*CoV*2, through the creation of attenuated viruses. In conclusion, genome recoding can offer diverse applications for improving cellular properties in the genome-recoded organisms.

Keywords Genetic code, Codon usage, Protein engineering, Attenuated virus

INTRODUCTION

The initial concept of a universal genetic code now has many exceptions. The investigation of translation machinery, as well as sequencing, has provided many examples of genetic code variations; in fact, these are routinely observed in numerous organisms, including fungi, bacteria, archaea, and viruses (Ambrogelly et al. 2007). Although genetic code variation has been observed under natural conditions, it is often restricted to the stop codon. The first report on the redefinition of

a stop codon, UGA, was in RNA *phage Qβ* (Weiner and Weber 1971). Selenocysteine (Sec), a noncoding amino acid, was also documented to be encoded by the stop codon UGA in 1986. Collectively, these variations from the standard genetic code were initially documented as deleterious. However later, this mechanism was determined to be widespread in nature and to impart microbial fitness under specific stress conditions (Pan 2013). In another example, the CUG codon is used to code for both Leu and Ser amino acid residues in *Candida albicans* in its natural state, increasing the likelihood of proteome destabilization by a single mismatch in charging tRNA[CAG] by either leucyl-tRNA synthetase or

✉ Correspondence: vinayihbt@gmail.com (V. Kumar)

seryl-tRNA synthetase (Suzuki et al. 1997). However, under some environmental constraints, the same mechanism, as well as the destabilized proteome, have been found to be more beneficial for this species (Suzuki et al. 1997). *Escherichia coli* was also found to be more successful in dealing with some mismade proteins through an upregulated heat-shock response under survival stress conditions by correcting missense mutations in many crucial enzymes (Min et al. 2003). All of these investigations concluded that genetic code variations should not always be considered lethal, and may even be considered advantageous under certain conditions. Thus, genetic code flexibility offers the possibility of codon reassignment without hampering the fitness of the target organism.

Genetic code conservation among organisms allows them to share beneficial traits by horizontal gene transfer (Vetsigian et al. 2006). However, this mechanism also allows viruses to easily use the host translation machinery for the translation of its genetic material (Krakauer and Jansen 2002). Biosafety and virus resistance are among the major problems in biotechnology. Moreover, biotechnology is limited by the use of only 20 canonical amino acids. These limitations can be overcome by changing the genetic code. Genome recoding is a strategy that involves the reassignment of codons to create an alternate genetic code. The term "recoding" was coined by Gesteland et al. (1992). In this approach, a specific codon is replaced with its synonymous codon and the corresponding tRNA is subsequently removed; this renders the complementary codon blank, so that it can be reassigned for another use (Plotkin and Kudla 2011; Kuo et al. 2018). Genetically recoded organisms (GRO) can be generated with diverse valuable features, including genetic isolation, virus resistance, reduced genome size, and improved cellular functions, by in vivo incorporation of specific non-canonical amino acids (ncAA).

In fact, genome recoding allows the synthesis of a nonfunctional protein by mistranslating the transferred gene in the non-recoded organism, resulting in a unique organism that exhibits genetic isolation (Ma and Isaacs 2016). This approach also renders the organism resistant to virus attack, because recoded organisms read the viral genome differently. Moreover, genetic recoding of an organism also allows the site-specific incorporation of desired ncAA (Lajoie et al. 2016). Most organisms targeted for genome recoding are bacteria (especially *E. coli* and *Salmonella*), yeast, and viruses (Coleman et al. 2008). In the bacterial system, TAG recoding has been documented in *E. coli* (Isaacs et al. 2011). Similarly, the recoding of Leu in *Salmonella* has been reported by Lau et al. (2017). The viral (*Arbovirus*) genome has also

been recoded to rebalance its preference for mammals versus insects (Shen et al. 2015). Thus, this technique is emerging as a good alternative for virus attenuation, as well as for protein engineering. To cover all significant aspects, the compiled information about the basic concept of genetic code and variation, successful examples of GROs, and their associated potential applications are discussed. This article provides an overview in the field of recoding and its diverse applications for improving cellular properties, including developing resistance against virus-mediated responses.

BASIC CONCEPTS OF GENETIC CODE

A genetic code is a unique sequence of nucleotides in DNA and RNA that determines the specific complementary sequence of amino acid residues in translated protein product (Jukes 1965). The concept of genetic codon has many unique characteristics: triplet, degenerate, nonoverlapping, commaless, nonambiguous, and universal. Among these unique features, the degeneracy feature permits decoding of each specific amino acid by more than one type of codon; different types of codons that encode the same amino acid are defined as synonymous and are recognized by different available isoacceptor tRNAs (Lagerkvist 1978). Out of a total 64 genetic codons, 61 code for 20 amino acids and the other 3 are stop codons. The corresponding codon in the mRNA is matched with its complementary anti-codon, which is associated with a specific tRNA that carries a specific amino acid during the translation process. After establishing the correct pairing between the specific mRNA codon and tRNA anticodon region, the corresponding amino acid is incorporated into the growing polypeptide chain as shown in Fig. 1. Although the origin of the genetic code is not well documented, two different theories have been proposed to gain in-depth insight into the concept of the origin of codes. The first theory is "Frozen Accident Theory" which proposes that the linkage between the amino acid and its triplet codon came about purely by chance; whereas, the "Stereochemical Theory" proposes that this linkage is based on stereochemistry (Crick 1968). Being universal, the genetic code was assumed to have become frozen in its existing form once a certain level of cellular complexity had been achieved. However, the discovery of deviation from the universal code indicates codon flexibility, as compared to the established concept of nonevolving and frozen genetic code.

Several deviations from the standard code in prokaryotes as well as eukaryotes have been documented using genome analysis. In the nuclear code,
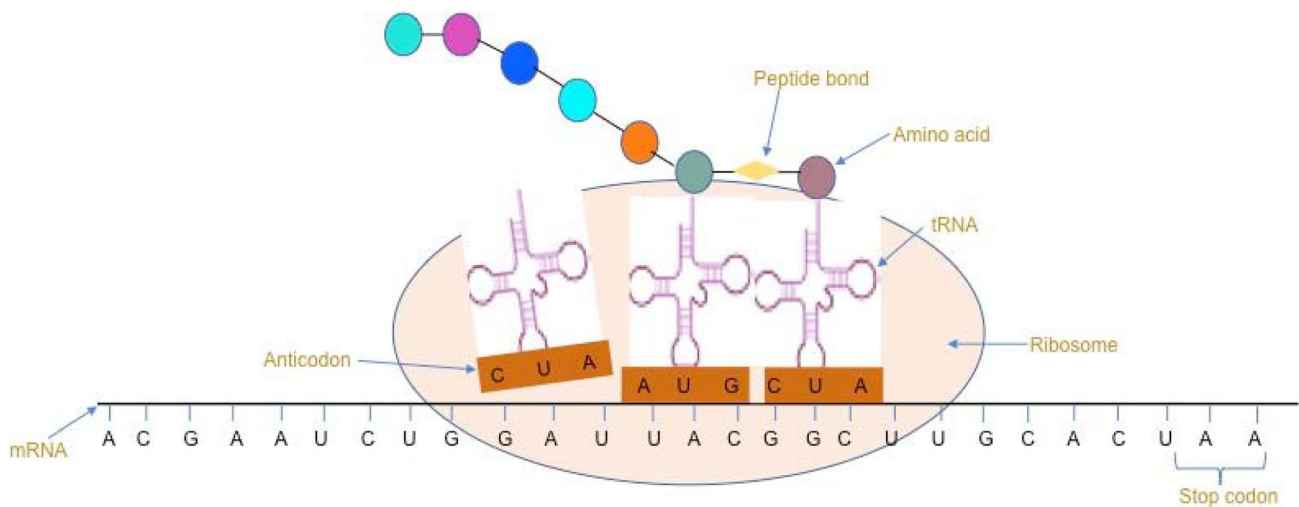
Fig. 1 Mechanism of translation

more than 10 codon reassignments have been reported; in mitochondrial codes, more than 16 changes have been reported (Osawa et al. 1992; Knight et al. 2001; Miranda et al. 2006). The reassignment of the CUG codon to Ser instead of Leu in *Candida* yeast was reported by Kawaguchi et al. (1989). Keeling (2016) has reported the reassignment of UAA and UAG stop codons to encode for a canonical amino acid in many ciliates. Similarly, reassignment of the UGA stop codon to code for either Trp or Gly residue has also been documented in bacteria. Riley et al. (2016) documented the reassignment of CUG to Ala instead of Leu in *Pachysolen tannophilus*. The reassignment of stop codons to amino acids has also been reported in *Condylostoma magnum*, *Parduczia* sp. and *Blastocrithidia* spp. trypanosomatids. However, the occurrences of stop codons are still being documented as a termination signal at the end of the reading frame (Heaphy et al. 2016; Swart et al. 2016). All of these reassignment patterns of the genetic codon prove that the genetic code is not frozen, but rather continuously evolving. The evolution of these genetic variations has been documented through four distinct biochemical mechanisms: biased codon usage, codon reassignment, ambiguous decoding, and natural genetic code expansion (Ling et al. 2015). A simple illustration (Fig. 2) is also provided for in-depth insight of the different mechanisms involved in codon reassignment.

Biased codon usage

Biased codon usage is defined as the usage of synonymous codons over others with reasonable frequency. This was first observed in a comparison of degenerate codons of sequenced mRNAs which revealed that each gene in the existing genome exhibits similar preference patterns for some specific synonymous codons over others (Grantham et al. 1980). This phenomenon has been observed in most organisms, including *E. coli*, where Arg residue-coding triplets CGU and CGC are observed 10 times more frequently in the genome than AGA and AGG (Ling et al. 2015). Codon preference was also seen to be correlated with the abundance of isoaccepting tRNAs. The degree of this correlation, as explained by the biased codon theory, might be related to protein production levels. Two hypotheses—selection theory and mutation theory—have been proposed to explain the correlation of codon preference with an abundance of isoaccepting tRNAs (Shabalina et al. 2013).

According to the selection theory, the level, as well as accuracy of protein expression can be optimized using biased codon usage, which offers a selective advantage to the organism. In contrast, mutation theory states that mutational pressure is a major driving force for codon biasness. According to the mutation–selection–drift model, which is based on both selection and mutation theories, the abundant codons that are frequently used are preferred during selection, and mutational pressure maintains the rare codons in the genome of the documented organisms (Bulmer 1991). The involvement of rare codons has been established in slowing down the translation processes. Under normal conditions, when all amino acids are available in appropriate amounts in the nutrient media, the supply of amino acid–tRNA matches the requirements of the protein translation process and all of the synonymous codons are translated at similar speed; however, when specific amino acids are scarce in the nutrient media, the tRNA isoacceptors are not charged at similar levels, resulting in a significant difference in translation rate of the
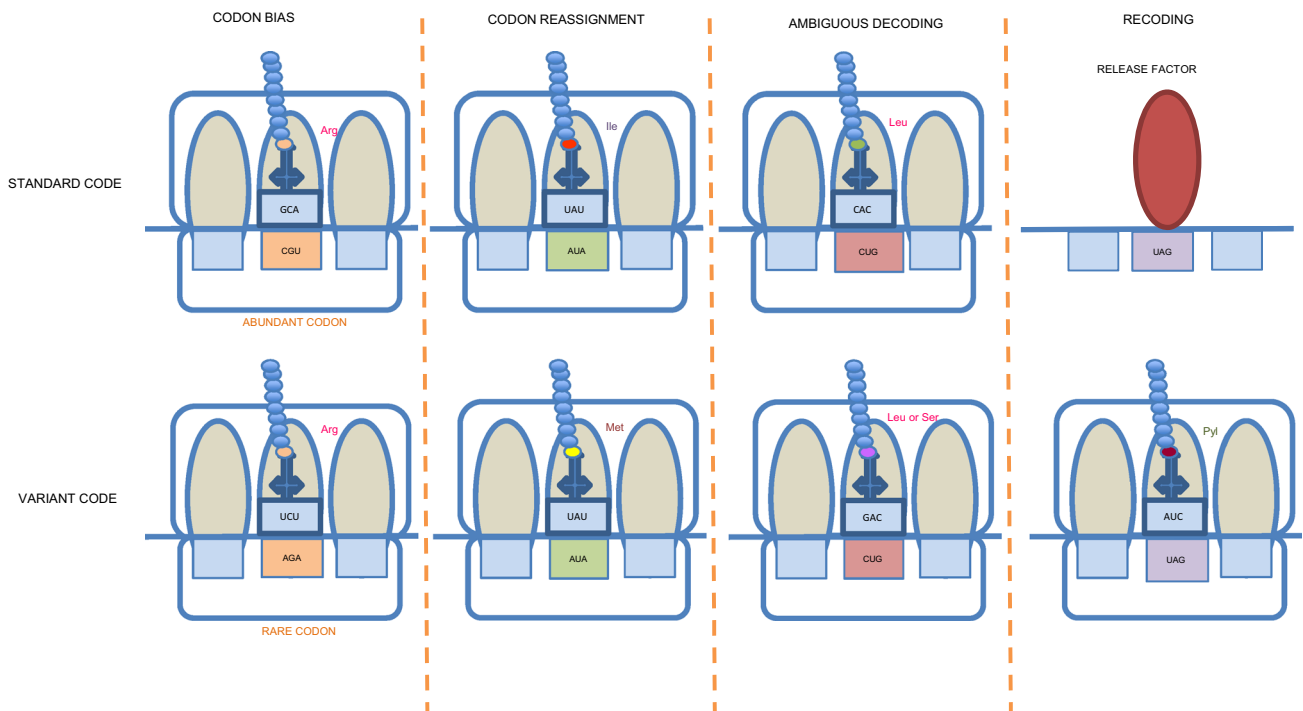
Fig. 2 Mechanisms of genetic code flexibility

synonymous codons. In nature, when microorganisms do not get a proper supply of nutrients, biased codon usage could play an important role in maximizing the overall translation process and subsequently improving the growth and development of the organism (Ling et al. 2015).

Similar to the observation that optimized usage of synonymous codons might be responsible for increasing the rate of protein synthesis, the use of nonoptimal codons at a specific location could lead to improved bacterial fitness under certain conditions. Subramaniam et al. (2013) documented the usage of nonoptimal Ser codons in the encoding of *Sin R* which is involved in biofilm formation. This adaptation strategy can work as a molecular sensor (for Ser residues) for cellular level and enables *Bacillus subtilis* to manipulate the expression of genes related to biofilm in response to environmental conditions (Ling et al. 2015). In fact, this nonoptimal strategy provides a selective advantage for adaptation to variations in the organism's surrounding nutrient environment.

Codon reassignment

The codon reassignment phenomenon includes reassignment of a stop codon to a sense codon, a sense codon to a sense codon, and a sense codon to a stop codon. The reassignment of a sense codon to a stop codon only requires the loss of cognate tRNA, whereas the other two reassignments (stop to sense and sense to sense) require the formation of new tRNA variants. The new tRNA variants may evolve from either potential modifications or duplications of existing tRNA, as described by Lang et al. (2012). In addition, modifications in aminoacyl-tRNA synthetases (aaRS) have been documented as potential contributors in codon reassignment.

The codon capture theory has been proposed to explain codon reassignment (Osawa and Jukes 1989). According to this theory, some codon might have disappeared from the coding sequence, and similarly, some tRNA with the anticodon might have also vanished during evolution. However, during a later stage of evolution, the same codon might have reappeared, but this time the anticodon might be assigned to either the same amino acid or a different one, so that the codon gets reassigned or captured. All of these sequential events can occur due to directional mutation pressure which mediates variations in the GC content of genomic DNA (Sueoka 1962; Jukes and Bhushan 1986; Muto and Osawa 1987). The first example of codon reassignment was observed in *Mycoplasma capricolum*, where the UGA codon codes for Trp instead of working as a stop codon (Yamao et al. 1985). In ciliated protozoans, both UAA and UAG stop codons code for Gln (Hanyu et al. 1986). Codon reassignment has not only been observed in genomic DNA; it occurs in mitochondrial DNA as well. In yeast mitochondria, UGA encodes for Trp instead of

working as a stop codon (Macreadie et al. 1983). Similarly, in *Neurospora* mitochondria, CUN has been reported to code for Thr instead of Leu. The Ile residue is also coded by AUA in green plants, and the CGG codon encodes Trp in mitochondrial DNA (Fox 1987).

An understanding of the consequences of codon reassignment can be gained by investigating specific microorganisms in which the genetic code has already been reengineered or reassigned. Release Factor 1 (RF1) was deleted in an attempt to reassign UAG to a sense codon in *E. coli*. When native UAG codons are detected, deletion of RF1 and the assignment of UAG as a sense codon has been observed with many growth defects (Heinemann et al. 2012; Johnson et al. 2012); at the same time, a rescue growth pattern has been observed when all of the endogenous UAG codons are converted to UAA (Lajoie et al. 2013a). Collectively, these reports suggest that the observed toxicity is due to the extension of proteins past the intended stop. In an RF1 deletion mutant, UAG is also translated by natural amino acids (mainly Gln, Lys, and Tyr) that compete with the insertion of ncAA, resulting in the formation of immature proteins (Aerni et al. 2014). Thus, the concept of radical codon reassignment can decrease fitness, including proteome stability and growth. During the emergence of the codon reassignment, the benefits gained by the organism need to outweigh the associated negative effects. However, a comprehensive understanding of the underlying molecular mechanisms governing a cell's adaptation to expansion of the genetic code is still lacking in the scientific literature and need to be explored.

Ambiguous decoding

Ambiguous decoding refers to the simultaneous decoding of the same codon by two or more amino acids in one cellular compartment; it is also termed mistranslation. Possible reasons for misassignment of amino acids are either codon misreading during decoding of the mRNA in the ribosome, or tRNA mischarging by aaRS. This mischarging is generally caused by either activation of the incorrectly bonded amino acid or failure of an aaRS to recognize the cognate tRNAs. Some aaRS have an editing mechanism that discards the chemically similar amino acid from the active site, thereby minimizing mischarging frequency.

However, tRNA mischarging is a mandatory requirement in many organisms, including most archaea and bacteria, for the synthesis of Asn, Cys, Sec, and Gln residues. Most decoding errors are caused by misplacing of the nonsense codon and sense codon by either non-cognate or near-cognate tRNA (Moura et al. 2009).

In fact, ambiguous decoding leads to the formation of a statistically relevant pool of different protein products, having amino acid substitutions at different positions (Ling et al. 2015). The higher level of mistranslation may result in a collection of misfolded proteins, leading to toxicity to the native cell. Thus, the reduced fidelity during translation, due to either mutation or stress, leads to ambiguous decoding as documented by Netzer et al. (2009) and Zaher and Green (2009). Mutation in ribosomal genes has also been documented in the mismatching of tRNA anticodons with mRNA codons, and oxidative stress was found to be mainly responsible for decreasing fidelity during aminoacylation. Microorganisms have been documented to use ambiguous decoding as an adaptive mechanism to survive under harsh conditions (Ling et al. 2015). In *Mycobacterium smegmatis*, the frequency of ambiguous decoding for codon Asn increases under low pH conditions as well as during the stationary phase (Javid et al. 2014). This ambiguous translation has been reported to increase resistance against the antibiotic rifampicin via the production of the variant protein RPoB, which is not recognized by rifampicin. Thus, epigenetically enhanced fitness of *M. smegmatis* has been reported using ambiguous decoding by Javid et al. (2014). Fan et al. (2015) also reported that mistranslation due to a mutation in the ribosome activates the general stress response in bacteria, and it was also found to enhance tolerance against hydrogen peroxide. In fact, codon ambiguity increases proteome diversity by creating statistically relevant populations of proteins, and it can lead to genetic and phenotypic diversity that can be exploited by natural selection for regulatory, developmental, and metabolic innovation (Moura et al. 2009).

Natural genetic code expansion

Natural genetic code expansion includes changes in the genetic code pattern that have evolved in the organism in the natural state and enable protein synthesis with more than the usual 20 canonical amino acids. Organisms' utilization of two ncAA residues (Sec and pyrrolysine [Pyl]) is an major example of natural genetic code expansion. However, the mechanisms of these two amino acids' translation in the same organism are distinct and it occurs only rarely. One codon, UGA, has ambiguous meaning and functions as a stop codon, and while present in the frame, it codes for Sec, which is the 21st amino acid to be inserted cotranslationally. This specific function requires a tRNA with the anticodon UCA (tRNA^Sec), mRNA with a stem-loop structure (Sec insertion sequence [SECIS]), and a specialized elongation factor (EF), SelB, as mentioned by Böck et al.

(2005). The SelB protein has two important functions for proper recoding of UGA as Sec. The first function is mediated by its N-terminal domain which is homologous to EF-Tu and binds to Sec–tRNA^{Sec} with high specificity. Interestingly, in contrast to EF-Tu, SelB does not bind to any other aa-tRNA (Forchhammer et al. 1989). Second, binding of SelB to the SECIS element through the C-terminal domain is required for the delivery of Sec–tRNA^{Sec} to the A site of the ribosome, by outweighing RF binding to the UGA codon (Ambrogelly et al. 2007). This dual property of SelB protects the fidelity of translation by confirming that only UGA codon present in the selenoprotein mRNA gets recoded. The amino acid residue Sec and selenoprotein have been reported in the algae *Emiliania huxleyi* and *Chlamydomonas* spp., (Fu et al. 2002; Obata and Shiraiwa 2005). The tRNA^{Sec} has also been reported in lower plants (Hatfield et al. 1992). However, this Sec has not been documented in higher plants or fungi.

In Methanosarcinaceae, an archaebacterial family, the codon UAG is ambiguous; apart from being used as a stop codon, it also codes for the 22nd cotranslationally added amino acid Pyl (Hao et al. 2002; Srinivasan et al. 2002), which also requires pyrrolysyl-tRNA synthetase for the recoding. The mechanism for the genetic encoding of Pyl is distinct and orthogonal to the system that has evolved for the insertion of Sec into protein. Free Pyl is synthesized inside the cell from two free Lys residues with the help of three enzymes encoded by *pylB*, *pylC*, and *pylD* (Gaston et al. 2011). The Pyl residue is then ligated to the tRNA^{Pyl} by the PylRS (Blight et al. 2004). The UAG codon is read by the CUA anticodon of tRNA^{Pyl} to reassign its meaning from a stop codon to Pyl. Pyl has been reported in the bacterium *Desulfitobacterium hafniense* (Ambrogelly et al. 2007).

Of more than 140 amino acids that are found in natural proteins, only 2 (Pyl and Sec) have been added to the list of standard amino acids. Phosphoseryl-tRNA (in form of tRNA^{Cys} and tRNA^{Sec}) may act as a signal for the discovery of other cotranslationally modified amino acids (Ambrogelly et al. 2007).

GENERATION OF RECODED ORGANISMS

Genome recoding involves many aspects, including engineering of the endogenous translational component and orthogonal translational component, and genome engineering to modulate global codon usage. There have been many milestones in the genome recoding of organisms (Fig. 3). For whole-genome recoding, all instances of the targeted codon are identified and replaced by a synonymous codon, the translation factor

associated with the targeted codon is inactivated leaving the targeted codon blank, and all of the enzymes that degrade the ncAA are inactivated. For the incorporation of ncAA, an orthogonal translation system is introduced into the organism. The orthogonal aaRS charges the desired tRNA with a ncAA which is either produced by the cell or supplemented in the media, and it is taken up from there by the cell through the transporter. The charged tRNA binds to its complementary codon on the mRNA and the ncAA is added to the growing polypeptide chain. A general outline for the generation of a recoded organism is depicted in Fig. 4.

Orthogonal translation system

The ribosome translates a protein by reading the codon sequence on the mRNA and adding the desired amino acid residues to the growing polypeptide chain. The identity of the amino acid residue to be incorporated is checked at multiple steps. First, the aaRS charges the tRNA with the correct amino acid. Then, the aminoacyl-tRNA is transferred to the active site of the ribosome by EF-Tu, where base pairing of the mRNA codon with the tRNA anticodon permits transfer of the amino acid to the growing chain. To expand the genetic code, it is important to introduce new translation components such that the function and fidelity of the endogenous translation system remain uncompromised. To achieve this, engineering of different components—including aaRS, tRNA, EF-Tu, and ribosomes—has been considered.

Engineering of aaRS–tRNA pair

The desired orthogonal aaRS–tRNA pair should incorporate the desired amino acid without crossreacting with all of the other incorporated ncAA, tRNA, cellular amino acids, or aaRS. When the newly introduced aaRS–tRNA pair is from a phylogenetically distant organism, it is quite distinct, with diverse tRNA identity elements, reducing crossreactivity between it and the native aaRS–tRNA pairs. The aaRS selects tRNA species by recognizing a small number of bases in the tRNA and also specifies an amino acid with an amino acid-binding pocket as explained by Giegé et al. (1998). This pairing enables discrimination among any component of the cell's metabolome. Orthogonality can be achieved by diversifying the amino acid-binding pockets, but the orthogonal aaRS are lacking any mechanism by which they can reject the many ncAA that are being used by the synthetic biologist, and seem to be poly-specific for ncAAs (Guo et al. 2014). The aaRS variants show very low catalytic activity with ncAA and this can lead to low
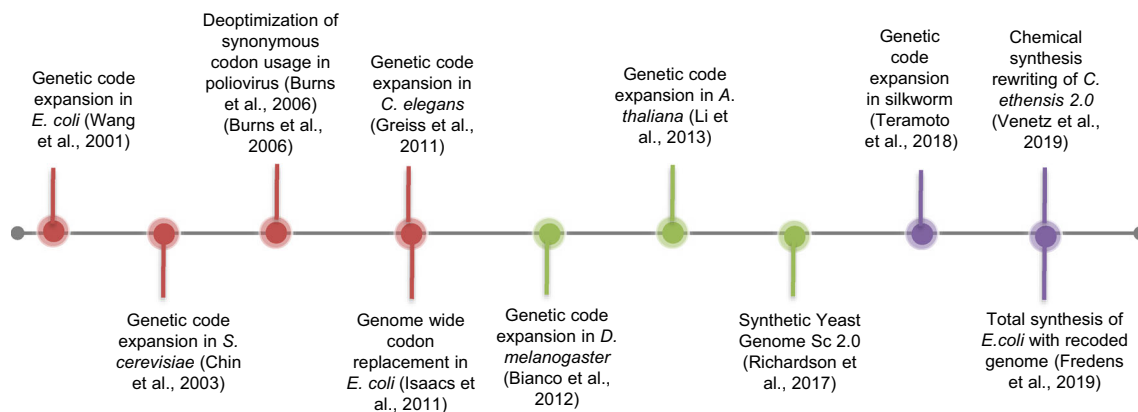
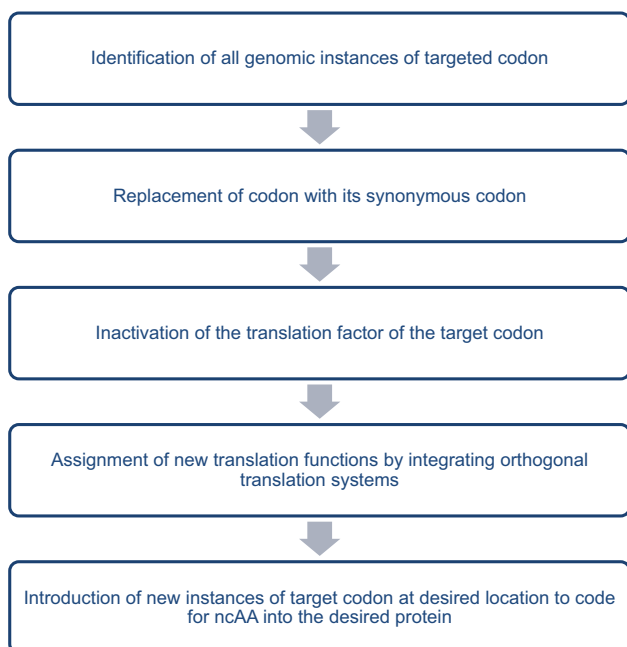Fig. 3 Major milestones in genome recoding of various organisms

Fig. 4 The general process used for the reassignment of a codon

ncAA–tRNA levels in the cell, which can be easily out-competed by endogenous tRNA which are engaged in codon–anticodon interactions. These interactions may lead to the addition of a canonical amino acid (which is not desired) at the codon of interest. This modification can be compensated for by overexpression of an orthogonal tRNA–aaRS pair along with elevated levels of ncAA. Orthogonal aaRS equipped with a heterologous editing domain can also improve the amino acid specificity. The specificity of the aaRS–tRNA pair can be fine-tuned by generating large libraries of mutations in the amino acid-binding pockets or at residues that are involved in the aaRS–tRNA interaction. Among them, the functional variants can be identified by performing alternate cycles of positive and negative selections in

the presence or absence of ncAAs in a particular experiment.

Engineering of elongation factor EF-Tu and ribosome

EF-Tu plays a role in the recognition of amino acids and it tends to reject those aminoacyl-tRNA that are bulky or have negatively charged amino acids (Doi et al. 2007). Modification over time in the amino acid-binding region of an EF has been found to make it compatible for the incorporation of specific amino acids (Sep(O-Phospho-l-serine), phosphotyrosine, and Sec) into proteins in E. coli. The orthogonal translation system also includes another important component, the engineered ribosome. The engineered ribosome should have reduced affinity for the native mRNA and increased affinity for the engineered mRNA and subsequently, the native and the engineered subunits of the ribosome should not crossreact. To increase the specificity between the engineered ribosome and mRNA, a specific sequence (anti-Shine–Dalgarno sequence) at the 3′ end of 16S rRNA can act as a substitute for replacement with the production of an orthogonal 16S molecule that only translates orthogonal mRNAs having a complementary synthetic ribosome binding site (Rackham and Chin, 2005). The large and small subunits of the orthogonal ribosome can be physically linked through a liner sequence to reduce the crossreaction between engineered and native subunits of the ribosomes (Schmied et al. 2018). These strategies can be combined to enable extensive engineering of ribosomal functions.

Recoding in viruses

Synonymous genome recoding has been reported for the generation of viruses with altered phenotypes, in which synonymous mutations were introduced in the protein-coding region of the virus genome without

altering the encoded protein sequence (Martínez et al. 2016). Burns et al. (2006) also documented the replacement of codons of nine amino acids with corresponding synonymous codons in the vaccine Sabin type 2 oral poliovirus. In that work, of the 2555 total nucleotides of the virus capsid region, about 542 substitutions were reported to be introduced and correlated with reduced fitness of the virus in HeLa cells. De-optimization of synonymous codon usage was also carried out by replacing 680 nucleotides of the capsid-encoding region of poliovirus type 1, resulting into a nonviable virus (Mueller et al. 2006). Rare codons have also been used to synthesize the same amino acid sequence as the wild type in poliovirus. Using this strategy, two new polioviruses were designed, named as PV-Max and PV-Min. The capsid precursor PI region contains overrepresented and underrepresented codon pairs in PV-Max and PV-Min, respectively; 566 synonymous mutations were also documented in recoded PV-Max, and 631 in PV-Min. The recoded PV-Min was found to be nonviable due to poor translation with underrepresented codon pairs, leading to virus attenuation. Similarly, a random codon-recoding strategy was used to attenuate tick-borne encephalitis virus (TBEV) via the introduction of 273 synonymous mutations into a 1400-nucleotide region present in the NS5 coding region. The recoded TBEV showed an attenuated phenotype in a mouse model of nonlethal encephalitis (De Fabritus et al. 2015). Dengue virus (DENV) has also been recoded using a codon pair de-optimization strategy targeting three DENV protein-coding sequences. In all three selected regions, strongly underrepresented human codon pairs were incorporated and the recoded viruses showed reduced replication capacity in mammalian LLC-MK2 cells (Shen et al. 2015). Recoding of the virus genome can also offer a platform for studying virus interactions with innate immune response, producing attenuated forms, and improving our knowledge of virus biology (Martínez et al. 2016).

Recoding in bacterial system

Escherichia coli

Genome recoding can be achieved through (1) editing in the existing genome by a site-specific mutation in the target codon, (2) rebuilding the native genome with a new recoded fragment, and (3) completely de novo synthesis from a synthesized or recoded fragment in the existing genome, as reported by Isaacs et al. (2011). Isaacs et al. (2011) also reported use of the same strategy for the application of conjugative assembly genome engineering (CAGE) and multiplex automated genome engineering (MAGE) for changing all 321 UAG codons to UAA in *E. coli*.

Many different strategies have been tested and applied to recode the *E. coli* genome. The first expansion of the genetic code in *E. coli* was reported by Wang et al. (2001), in which *O*-methyl-L-tyrosine was incorporated using a tRNA–aaRS pair generated from tyr tRNA–aaRS of *Methanococcus jannaschii*. Lajoie et al. (2013b) selected the UAG codon as the prime target for genome-wide reassignment because UAG is the rarest codon in *E. coli* strain MG1655. Wild-type *E. coli* MG1655 has a total of 321 known UAG codons that are decoded as a translation stop by RF1. In one strategy, a variant of RF2 was used that exhibited enhanced UAA termination (Johnson et al. 2011) and weak UAG termination (Ohtake et al. 2012). In another strategy, the UAA stop codon was substituted into all seven essential genes that naturally terminate with a UAG stop codon; later, ribosome toxicity was found to be reduced by incorporating amino acids at the remaining 314 UAGs (Mukai et al. 2010). The MAGE approach was utilized to develop different strains that were labeled as C0.B*.ΔA::S (enhanced expression of RF2 variant), C7.ΔA::S (UAG is changed to UAA in seven essential genes), and C13.ΔA::S (UAG is changed to UAA in seven essential genes plus six nonessential ones). In the labeling of these strains, C refers to the number of codons changed, and A and B refer to prf A (RF1) and prf B (RF2) manipulations. RF1 was also deleted from these strains.

To check the fitness effect of UAG reassignment and RF1 removal, the doubling time and maximum cell density of the bacterial population were measured. All three developed strains (C0.B*.ΔA::S, C7.ΔA::S, and C13.ΔA::S) had impaired fitness. RF1 removal significantly reduced fitness for C0.B*.ΔA::S and codon reassignment made the situation worse, probably because the ncAA incorporation outcompeted the weak UAG termination activity exerted by the RF2 variant (Ohtake et al., 2012). Similarly, C7.ΔA::S and C13.ΔA::S also showed reduced fitness, probably due to the more than 300 nonessential UAG codons stalled the translation process in the absence of RF1-mediated translation at the UAG codons. All of these strategies showed reduced fitness of the modified bacterial system. Another strain was then developed, labeled C321.ΔA, in which all 321 UAG codons were substituted by UAA codons and prf A (encoding RF1) was also deleted. This GRO displayed normal prototrophy and morphology. Overall, C321.ΔA exhibited better performance than the other strains developed by the previously applied strategies. This observation allowed the complete reassignment of the UAG stop codon to a sense codon that is capable of incorporating ncAA into proteins in a bacterial system.

The MAGE approach was used to remove all of the known UAG codons in groups of 10 across 32 *E. coli* strains and another approach, CAGE was used to consolidate these codon changes in a group of $\sim$ 80 across four different strains. The MAGE approach uses ssDNA oligonucleotides and λ bacteriophage β recombinase (Ellis et al. 2001) for the simultaneous introduction of multiple defined mutations at multiple locations of a replicating bacterial genome. The CAGE approach utilizes the process of bacterial conjugation to precisely transfer up to several million base pairs of contiguous DNA. This strategy has been utilized for the production of an extensively modified genome from small segments (Isaacs et al. 2011). To examine whether these GROs can obstruct viral infection, different strains were challenged with bacteriophages T4 and T7. The absence of RF1 in strain C321.ΔA was not found to impact infection with T4 but interestingly, it was found to improve resistance against bacteriophage T7 (Lajoie et al. 2013b). The basic steps in the formation of strain C321.ΔA of *E. coli* are illustrated in Fig. 4.

Recently created *E. coli* Syn61 carries the largest recoded genome, in which more than > 18 k codons have been recoded to develop a genome with 61 codons. *E. coli* Syn61 used 59 codons to encode 20 amino acids and enabled the deletion of a previously essential tRNA. Two sense codons, TCA and TCG, encoding Ser, and a stop codon TAG in the reading frame of *E. coli* MDS42 were replaced by their synonymous AGT, AGC, and TAA, respectively, as documented by Fredens et al. (2019).

Caulobacter crescentus

Caulobacter ethensis-2.0 (*C. eth*-2.0), a minimized bacterial genome with fundamental functions, was created by rebuilding the genome of *C. crescentus* using a chemical synthesis rewriting approach. The overall number of encoded genetic features was reduced from 6290 to 799, and a total 133,313 base substitutions were documented for the rewriting of 123,562 codons. A list of essential DNA parts was generated computationally. The DNA parts were extracted from the genomic sequence of *Caulobacter* and were oriented using digital tools and renamed *C. ethensis*-1.0. The *Caulobacter* origin of replication was replaced by the pMR10Y shutter vector sequence to ensure low copy replication in *E. coli*, *Caulobacter*, and *Saccharomyces cerevisiae*. Furthermore, computational sequence design algorithms were used to add 123,141 base substitutions in protein-coding sequences for development of the rewritten *C. eth*-2.0 design. A total of 56.1% of all codons were replaced by their synonymous versions. The identification of toxic genes was also carried out by

forming 37 chromosome segments of *C. eth*-2.0. The conjugation transfer from *E. coli* to *Caulobacter* was also observed and 25 chromosome segments (out of a total 37 segments) without any toxic genes were confirmed by sequencing. A base substitution in the noncoding region was tolerated in this specific case. For confirmation of *C. eth*-2.0's stable maintenance in *S. cerevisiae*, whole-genome sequencing was performed over 60 generations and no mutation or chromosomal rearrangement was detected. In conclusion, these findings showed that decoding of fundamental functions of the genome can be successfully carried out using chemical synthesis rewriting (Venetz et al. 2019).

Recoding in yeast

Expansion of the genetic code in eukaryotes was first demonstrated in *S. cerevisiae* because it is a useful eukaryote model, and because its translation machinery is homologous to that of higher eukaryotes and genetic manipulations are feasible. Chin et al. (2003) demonstrated the incorporation of five ncAA into a protein in response to an amber stop codon using an *E. coli* tRNA–aaRS pair after modifications (Chin et al. 2003). Following the same concept, the synthetic yeast genome project was started with the aim of building complete yeast chromosomes from scratch. Richardson et al. (2017) demonstrated the design of synthetic genome Sc2.0, a modified *S. cerevisiae* genome with an 8% reduction in the natural genome's size. The chromosomes of the Sc2.0 yeast version have a modified genetic code with all TAG stop codons changed to TAA. Interestingly, eukaryotes can easily survive with a single stop codon, as observed in several ciliates in which stop codons have been found to encode amino acids (Salim et al. 2008). Therefore, elimination of the stop codon TAG seemed unlikely to compromise yeast fitness. The introduction of slight alterations in the nucleotide sequence of a specific synthetic or wild-type gene has also been achieved through synonymous recoding within the open reading frame (ORF) (Richardson et al. 2017). The yeast's natural homologous recombination mechanism was exploited for the integration of 30- to 60-kb recoded DNA segments in the Sc2.0 version. This strategy has also been documented as an option for boosting genetic studies and subsequently, to be useful for industrial purposes (Cubillos 2016).

Recoding in multicellular organisms

Extension of the genetic code and site-specific incorporation of different ncAA in multicellular organisms will facilitate our understanding of diverse cellular

processes. This strategy will help us gain in-depth knowledge on complex cellular processes. However, the incorporation of ncAA into multicellular organisms offers new challenges. The prime condition for genome recoding in multicellular organisms is the functioning of the whole translation machinery, including the orthogonal translation system, in the same cell. Maintaining the availability of ncAA in the cytoplasm of the targeted cell for continuous translation is another challenge. Moreover, nonsense-mediated mRNA decay (NMD) further complicates the recoding strategy in eukaryotes, because this process selectively destroys the mRNA with a premature amber codon (UAG) and eventually limits the expression of the targeted protein; this process is also termed "mRNA surveillance". Nevertheless, these challenges have been addressed, and some successful recoding experiments in multicellular organisms have been documented. The type-3 Pol III bacterial promoter has been used for the expression of orthogonal tRNA in multicellular organisms. The first report of genome recoding was documented in *Caenorhabditis elegans* (Greiss and Chin 2011). This organism was selected for genome recoding because its whole genome has already been sequenced and it has a transparent body which makes it easy to visualize expression in cells using fluorescent proteins. To execute this strategy, *C. elegans* strain smg-2, which is deficient in NMD, was used, along with the orthogonal tRNA–aaRS system from *Methanosarcina mazei* (already optimized to code for Lys analogs). Parrish et al. (2012) also reported utilization of the recoding strategy in *C. elegans* by incorporation of a UAG-containing reporter gene in the genome and documented less accuracy of ncAA incorporation when expressed on the extrachromosomal array, as had been done by Greiss and Chin (2011). Due to the occurrence of dipeptide transporters in most *C. elegans* cells, the non-natural amino acid (Dan Ala) was supplied as a food ingredient and incorporated into the desired protein. Site-specific incorporation of ncAA has also been demonstrated in *Drosophila melanogaster* (Bianco et al. 2012). Pyrrolysyl-tRNA synthetase was adopted from *M. mazei*, and shown to be orthogonal to the translation machinery in flies. Incorporation of ncAA was detected in specific tissues at different developmental stages, and in a subset of cells within the same tissue (Bianco et al. 2012). Li et al. (2013) reported recoding in the model plant *Arabidopsis thaliana*. They utilized the upf1-5 mutant with nonfunctional NMD, and expression of *Phy tRNA* was carried out using the type 3 Pol III *Arabidopsis* promoter. They achieved incorporation of N-ε-acryllysine in the protein of interest.

Similarly, Han et al. (2017) documented the integration of a transgene encoding an N-ε-acetyl-lysyl-tRNA synthetase pair in the *Mus musculus* genome, which enabled incorporating ncAA into the targeted protein. However, genome recoding had only been achieved in model multicellular organisms. Recently, genetic code expansion was achieved in an industrial animal, the silkworm *Bombyx mori* (Teramoto et al., 2018). The bacterial system was screened for phenylalanyl-tRNA synthetase with altered amino acid specificity and subsequently, four transgenic *B. mori* lines were created expressing variants of selected synthetases in the silk gland. Among these, two lines were found to have incorporated azidophenylalanine into silk fiber. These findings have paved the way toward an extension of molecular tools for controlling processes inside the cell; further, they have expanded our knowledge of processes that need to be studied at the organismal level, including neural processing and development in animals, and photosynthesis, and responses of plants to different environmental stresses.

Stabilization of recoded genetic codes

Under natural conditions, reassignment of a codon needs to be followed by stabilization of the manipulations. The tolerance for a close analog of natural amino acids can be evolved through the metabolic supplementation approach. In this approach, an auxotrophic bacterial strain that cannot synthesize a common amino acid is placed on media in which the common amino acid has been replaced with its analog. The bacterial strain accumulates mutant protein containing ncAA during translation and the bacteria become dependent on the ncAA in the absence of the natural amino acid (Lemeignan et al. 1993). However, the codons can only be reassigned to diverse ncAA after replacing all of the essential instances of the codon with synonymous codons (Lajoie et al. 2013b). Even then, the new genetic code remains fragile and only becomes established when the cell's fitness becomes dependent on the new translation function. This may happen over time due to natural genetic drift, but it can also be accomplished by re-engineering the essential proteins to be dependent on the specific ncAA for their proper translation, folding, and function.

Utilization of selective ncAA

Despite the success achieved in codon recoding with *E. coli*, not many instances have reported the utility of codon recoding for improved protein function. This might be due to the structural diversity of all available ncAAs (mostly analogs of Pyl and Tyr). Most of the ncAAs are known to have large, hydrophobic side chains

and are subsequently not able to fit in the functional position of many natural proteins (Dumas et al. 2015). However, the use of small-size ncAA could solve this problem. The small-size ncAAs are more versatile for packing at active sites, and their hydrophilic side chain might be more beneficial for the expansion of catalytic functions. However, use of these ncAA is more challenging for engineering the aaRS design: first, it is more difficult to distinguish smaller amino acids from the canonical amino acids in orthogonal aaRS active sites; second, it is more challenging to produce aaRS variants that can properly fulfil the requirement of hydrogen bonds to accommodate the hydrophilic side chain (Mukai et al. 2015).

POTENTIAL DIVERSE APPLICATIONS OF GENOME RECODING

The concept of genetic codon repurposing provides a very powerful tool for obtaining a genome that is enriched in new functions that are not commonly found under natural conditions, and it could provide a highly beneficial platform for genomic manipulations in the future. The formation of a GRO by replacing all 321 UAG codons with UAA has been achieved in *E. coli* (Lajoie et al. 2013b). The recoded organism has improved properties, with the incorporation of ncAA that expand the chemical diversity of the proteins in vivo, and improved resistance against T7 bacteriophage. This finding offers promise for new genetic codes enabling an organism to exhibit diverse applications, including improved viral resistance.

Protein engineering using ncAA as building blocks

As already noted, biotechnology has been limited by the 20 amino acids of the canonical genetic code. The use of ncAA expands the repertoire and offers new functions that were previously not found in nature (Agostini et al. 2017). Several approaches have been used for the incorporation of ncAA in proteins, including selective pressure incorporation, stop or sense codon suppression, and frameshift suppression. Among these, the stop or sense codon suppression and frameshift suppression approaches permit site-specific ncAA incorporation into the growing protein with the help of orthogonal tRNA–aaRS pairs that are specific for the codon and the ncAA (Chin 2014). Nevertheless, decoding methods are ambiguous, resulting in the production of statistical population of proteins, to encode ncAA unambiguously. The codons can be reassigned by eliminating either

native tRNAs or release factors that originally decoded specific codons (Mukai et al. 2017). The useful functional groups that are commonly attached to ncAA are alkynes, azides, ketones, alkenes, and aryl halides. The known cases of natural genetic code expansion are of Sec and Pyl residues. The Sec residue is similar to the Cys residue at the chemical level, but due to the lower redox potential and higher nucleophilicity of the Sec residue, some selenoenzymes are more reactive (100 times) than their Cys-containing counterparts. Moreover, the Sec residue is more resistant to irreversible oxidation than the Cys residue, proving that protein properties can be improved by the incorporation of ncAA as building blocks in the growing polypeptide chain. Incorporation of the multiple bromo/chlorotyrosine residues into the polypeptide chains of redox enzymes has been found to increase the enzyme's thermostability due to better side-chain packing (Ohtake et al. 2015). Protein functionalization can also be improved by the incorporation of ncAA, and specifically for the production of antibody–drug conjugates or bispecific antibodies, and to attach polyethylene glycol to the chain to improve protein stability (Huang and Liu 2018). A single-point mutation of the Phe264 residue to form p-isothiocyanate phenylalanine in the enzyme MetA has been found to increase the enzyme's thermal stability compared to the wild-type protein (Li et al. 2019). Incorporation of ncAA as building blocks in growing polypeptides offers many new and diverse applications, including trapping transient protein–protein interactions, helping in structural determination, monitoring protein functions in vivo using probe labeling, protein purification, drug screening, diagnostics and green energy (Wang et al. 2012).

Genetic isolation from horizontal gene transfer

Genetically modified organisms (GMO) carry the risk of releasing functional recombinant DNA into the environment, raising concerns of biosafety and ethics associated with transgenic technology. Horizontal gene transfer can occur through various modes, including transformation, conjugation, and transduction. Horizontal gene transfer causes phenotypic changes that destabilize the engineered biological system through manipulations of cellular fitness or gene–gene interactions (Baltrus 2013). The whole-genome synonymous codon replacement concept offers a mechanism for constructing a unique organism that exhibits genetic isolation and enhanced biological functions (Ostrov et al. 2016). GRO are virtually recalcitrant to horizontal gene transfer (both outcoming and incoming), as the incoming DNA cannot be read in the engineered strains

and the synthetic genes are not translated properly in the wild strain (Santos Moreno and Schaerli 2019). Passive incompatibility is also improved by the incorporation of a non-recoded toxin gene in the recoded organism, which will further prevent the transfer of artificial genetic material to wild strains (Kuo et al. 2018). The survival of GRO depends on the presence of ncAA through the insertion of ncAA in the essential genes (Mandell et al. 2015).

Improved biocontainment potential of GRO can have major scientific and economic implications. Maximum biocontainment is desirable from a scientific perspective, unless the strain has been engineered specifically for genetic exchange. The public might be more accepting of improved biocontainment, allowing revision of the restrictive regulations; this could have huge economic implications for biotechnology and agricultural industries (Santos Moreno and Schaerli 2019).

Virus resistance

Viruses are one of the deadliest pathogens to all organisms, including humans and crop plants. The virus relies completely on its host for the expression of proteins which are necessary for its propagation. Virus resistance is a desirable property for cell cultures which are extensively used in the production of therapeutic proteins and chemicals or food in industries. In industrial fermentation, bacteriophage contamination is a potential issue and phage infection of starter culture is the main bottleneck resulting in fermentation failure in the dairy industry. In a virus-infected cell with repurposed tRNA, the viral genes are not translated properly (Lajoie et al. 2013b). Thus, genome recoding could be an option for providing resistance to natural viruses (Fig. 5) (Fernández et al. 2017). Apart from this promising feature, viruses with a recoded genome have been reported to trigger a wild-type immune response and are a good option for vaccine development. While triggering the wild-type immune response, the protein sequence remains unaltered but has reduced viability because of suboptimal replication or translation processes, creating a safer vaccine (Wang et al. 2018). Attenuated virus strains have already proven useful for protecting against diseases (including COVID-19) caused by the same or related virus species. Development of in vivo-attenuated viruses mediated by genome recoding is reported to be insensitive to gene transfer, thereby reducing the risk of recombination, and it also involves thousands of nucleotide substitutions, thus reducing the risk of phenotypic reversion by a point mutation (Martínez et al. 2016). Redesigning the virus genome based on alterations in codon usage and
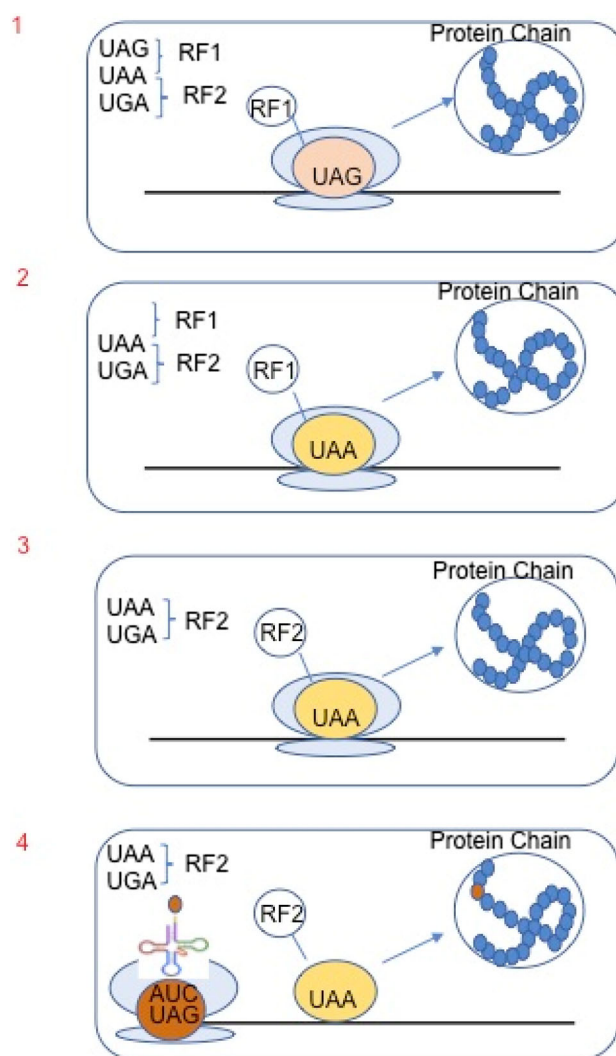

Fig. 5 Outline of steps involved in the formation of strain C321.ΔA of E. coli. (1) In the wild, UAG functions as a stop codon. (2) All of the UAG codons in the genetic code are replaced by UAA. (3) RF1, which binds to UAG, is deleted. The deletion of RF1 renders UAG a blank codon that can be assigned a new ncAA. (4) UAG is added to the desired positions in the genome

dinucleotide content could also lead to the formation of attenuated viruses by the introduction of a large amount of underrepresented synonymous substitutions in the viral genome. Attenuated forms of many RNA viruses have also been recently designed and reported, including human immunodeficiency virus type 1 (Martrus et al. 2013), chikungunya virus (Nougairede et al. 2013), DENV (Shen et al. 2015), vesicular stomatitis virus (Wang et al. 2015), arbovirus (Wimmer et al. 2019) and New Castle disease virus (Wang et al. 2019). This strategy could also be effective against coronaviruses, including the deadly SARS-CoV2.
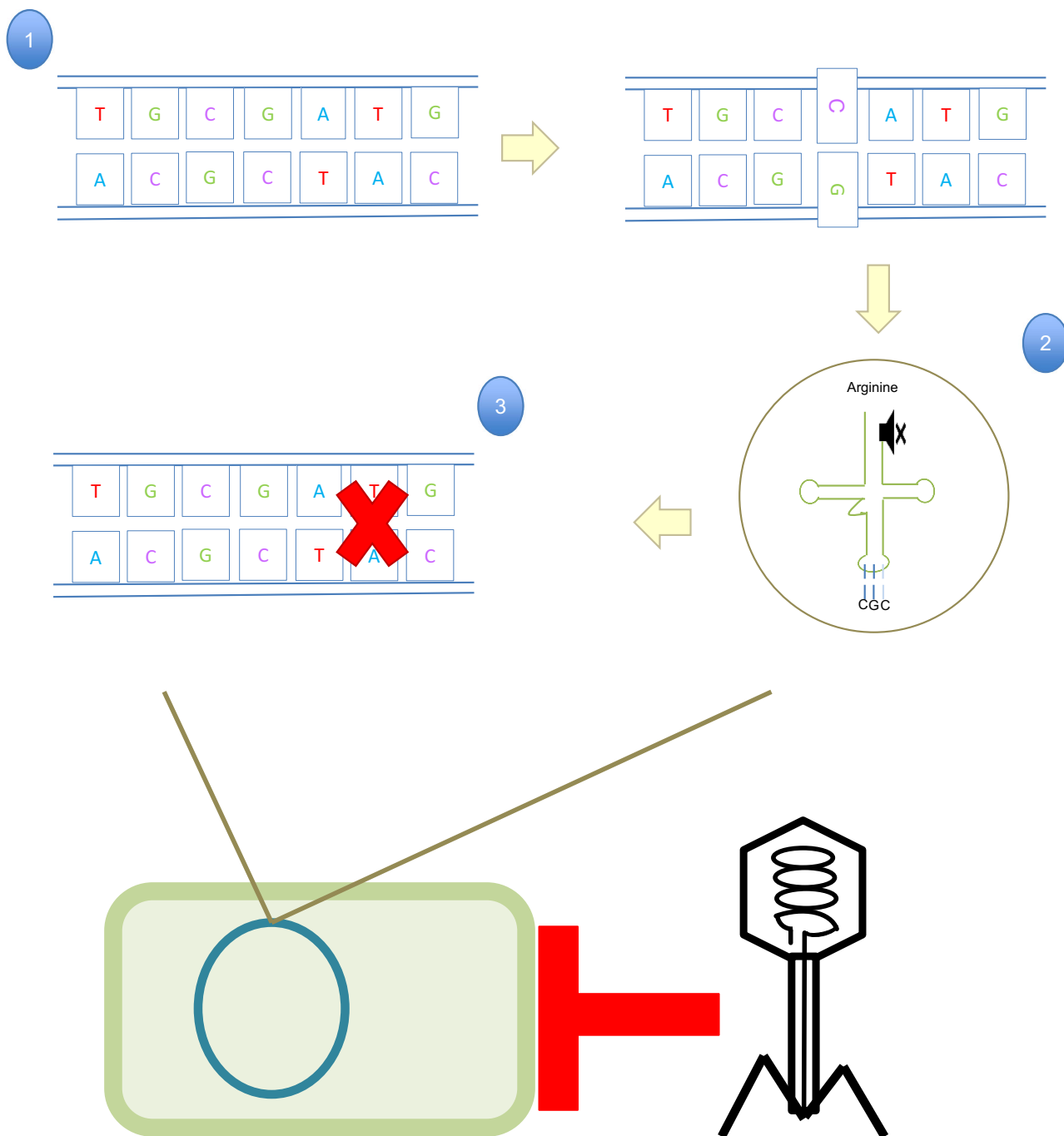
Fig. 6 Virus resistance through recoding. In the recoded organism, (1) the GCG codon is replaced with GCC. The cell will function normally because the replaced codon and deleted codon are synonymous. (2) The gene that encodes the tRNA which is involved in the translation of the deleted codon from mRNA is also deleted. The cell will no longer be able to translate DNA sequences with the GCG codon. (3) When a virus infects this cell, the cell will not be able to translate proteins from the viral DNA because of the missing tRNA in the host cell. Thus, the viral DNA becomes incompatible with the cell's translation toolkit and thus, the host becomes resistant to the virus

FUTURE PERSPECTIVES

Reassignment of a sense codon is a challenging task as codon usage can greatly affect gene regulation (Frenkel-Morgenstern et al. 2012), translation efficiency (Tuller et al. 2010), protein folding (Angov 2011), and ribosome spacing (Ingolia et al. 2009). The *E. coli* Syn61 organism has the largest reported recoded genome to date (Fredens et al. 2019), and research is progressing toward recoding of eukaryotic genomes, including that of humans (Santos Moreno and Schaerli 2019). Recoding of virus genomes can serve to form attenuated vaccines

against SARS-CoV2 or similar deadly viruses, as well as for the development of virus-resistant plants (Kumar and Singh 2020), helping reduce losses in agriculture due to viral attack. The addition of ncAA to the polypeptide chain could also have applications in protein therapeutics, including immunotoxins, vaccines, and bispecific antibodies, thus having a positive impact on medicine (Kuo et al. 2018), and may be useful against COVID-19 and future viral threats (Fig. 6).

CONCLUSIONS

The discovery of genetic code flexibility paved the way for the development of genome recoding as a technique for improving the genome plasticity of an organism. This could prove beneficial in the near future by enabling the manipulation of genomes according to our requirements. It could also provide new opportunities for protein engineering by increasing the number of building blocks, including ncAA, that can be incorporated in vivo. This technique is also promising because it solves the major concern associated with the negative impacts of GMO—the release of functional DNA into the environment. Thus, it offers a technique that solves the biosafety issues of transgenic technology. Genome recoding could also be used to provide virus resistance. Despite all of these advantages, a great deal of research is still required before we can fully exploit this technique. Moreover, the codon usage rules have not yet been fully explored, and the recoded genome designs are likely to contain some unknown lethal elements. It is therefore important to work on a vast genetic landscape, assess phenotypes that are arising from individual changes and combinations, and rapidly make designs to change the genetic code at the genome level.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Aerni HR, Shifman MA, Rogulina S, O'Donoghue P, Rinehart J (2014) Revealing the amino acid composition of proteins within an expanded genetic code. Nucleic Acids Res 43:e8–e8

Agostini F, Völler JS, Koksch B, Acevedo-Rocha CG, Kubyshkin V, Budisa N (2017) Biocatalysis with unnatural amino acids: enzymology meets xenobiology. Angew Chem 56:9680–9703

Ambrogelly A, Palioura S, Söll D (2007) Natural expansion of the genetic code. Nat Chem Biol 3:29

Angov E (2011) Codon usage: nature's roadmap to expression and folding of proteins. Biotechnol J 6:650–659

Baltrus DA (2013) Exploring the costs of horizontal gene transfer. Trends Ecol Evol 28:489–495

Bianco A, Townsley FM, Greiss S, Lang K, Chin JW (2012) Expanding the genetic code of *Drosophila melanogaster*. Nat Chem Biol 8:748

Blight SK, Larue RC, Mahapatra A, Longstaff DG, Chang E, Zhao G et al (2004) Direct charging of tRNA CUA with pyrrolysine in vitro and in vivo. Nature 431:333

Böck A, Thanbichler M, Rother M, Resch A (2005) Selenocysteine. In: Ibba M, Francklyn CS, Cauack S (eds) Aminoacyl-tRNA synthetases. Landes Bioscience, Georgetown, pp 320–327

Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage. Genet 129:897–907

Burns CC, Shaw J, Campagnoli R, Jorba J, Vincent A, Quay J et al (2006) Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. J Virol 80:3259–3272

Chin JW (2014) Expanding and reprogramming the genetic code of cells and animals. Annu Rev Biochem 83:379–408

Chin JW, Cropp TA, Anderson JC, Mukherji M, Zhang Z, Schultz PG (2003) An expanded eukaryotic genetic code. Science 301:964–967

Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S (2008) Virus attenuation by genome-scale changes in codon pair bias. Science 320:1784–1787

Crick FH (1968) The origin of the genetic code. J Mol Biol 38:367–379

Cubillos FA (2016) Exploiting budding yeast natural variation for industrial processes. Curr Genet 62:745–751

De Fabritus L, Nougairède A, Aubry F, Gould EA, De Lamballerie X (2015) Attenuation of tick-borne encephalitis virus using large-scale random codon re-encoding. PLoS Pathog 11:e1004738

Dsssoi Y, Ohtsuki T, Shimizu Y, Ueda T, Sisido M (2007) Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system. JACS 129:14458–14462

Dumas A, Lercher L, Spicer CD, Davis BG (2015) Designing logical codon reassignment-expanding the chemistry in biology. Chem Sci 6:50–69

Ellis HM, Yu D, DiTizio T (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci 98:6742–6746

Fan Y, Wu J, Ung MH, De Lay N, Cheng C, Ling J (2015) Protein mistranslation protects bacteria against oxidative stress. Nucleic Acids Res 43:1740–1748

Fernández L, Escobedo S, Gutiérrez D, Portilla S, Martínez B, García P, Rodríguez A (2017) Bacteriophages in the dairy environment: from enemies to allies. Antibiotics 6:27

Forchhammer K, Leinfelder W, Böck A (1989) Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. Nature 342:453

Fox TD (1987) Natural variation in the genetic code. Annu Rev Genet 21:67–91

Fredens J, Wang K, de la Torre D, Funke LF, Robertson WE, Christova Y et al (2019) Total synthesis of *Escherichia coli* with a recoded genome. Nature 569:514

Frenkel-Morgenstern M, Danon T, Christian T, Igarashi T, Cohen L, Hou YM, Jensen LJ (2012) Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. Mol Syst Biol 8:572

Fu LH, Wang XF, Eyal Y, She YM, Donald LJ, Standing KG, Ben-Hayyim G (2002) A selenoprotein in the plant kingdom mass spectrometry confirms that an opal codon (UGA) encodes selenocysteine in *Chlamydomonas reinhardtii* glutathione peroxidase. J Biol Chem 277:25983–25991

Gaston MA, Zhang L, Green-Church KB, Krzycki JA (2011) The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. Nature 471:647

Gesteland RF, Weiss RB, Atkins JF (1992) Recoding: reprogrammed genetic decoding. Science 257:1640–1642

Giegé R, Sissler M, Florentz C (1998) Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res 26:5017–5035

Grantham R, Gautier C, Gouy M, Mercier R, Pave A (1980) Codon catalog usage and the genome hypothesis. Nucleic Acids Res 8:197

Greiss S, Chin JW (2011) Expanding the genetic code of an animal. J Am Chem Soc 133:14196–14199

Guo LT, Wang YS, Nakamura A, Eiler D, Kavran JM, Wong M, Kiessling LL, Steitz TA, Donoghue P, Söll D (2014) Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. Proc Natl Acad Sci 111:16724–16729

Han S, Yang A, Lee S, Lee HW, Park CB, Park HS (2017) Expanding the genetic code of *Mus musculus*. Nat Commun 8:1–7

Hanyu N, Kuchino Y, Nishimura S, Beier H (1986) Dramatic events in ciliate evolution: alteration of UAA and UAG termination codons to glutamine codons due to anticodon mutations in two Tetrahymena tRNAsGln. EMBO J 5:1307–1311

Hao B, Gong W, Ferguson TK, James CM, Krzycki JA, Chan MK (2002) A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science 296:1462–1466

Hatfield D, Choi IS, Mischke S, Owens LD (1992) Selenocysteyl-tRNAs recognize UGA in *Betavulgaris*, a higher plant, and in *Gliocladium virens*, a filamentous fungus. Biochem Biophys Res Commun 184:254–259

Heaphy SM, Mariotti M, Gladyshev VN, Atkins JF, Baranov PV (2016) Novel ciliate genetic code variants including the reassignment of all three stop codons to sense codons in *Condylostoma magnum*. Mol Biol Evol 33:2885–2889

Heinemann IU, Rovner AJ, Aerni HR, Rogulina S, Cheng L, Olds W et al (2012) Enhanced phosphoserine insertion during *Escherichia coli* protein synthesis via partial UAG codon reassignment and release factor 1 deletion. FEBS Lett 586:3716–3722

Huang Y, Liu T (2018) Therapeutic applications of genetic code expansion. Synth Syst Biotechnol 3:150–158

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223

Isaacs FJ, Carr PA, Wang HH, Lajoie MJ et al (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science 333:348–353

Javid B, Sorrentino F, Toosky M, Zheng W, Pinkham JT, Jain N et al (2014) Mycobacterial mistranslation is necessary and sufficient for rifampicin phenotypic resistance. Proc Natl Acad Sci 111:1132–1137

Johnson DB, Xu J, Shen Z, Takimoto JK et al (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. Nat Chem Biol 7:779

Johnson DB, Wang C, Xu J, Schultz MD, Schmitz RJ, Ecker JR, Wang L (2012) Release factor one is nonessential in *Escherichia coli*. ACS Chem Biol 7:1337–1344

Jukes TH (1965) The genetic code, II. Am Sci 53:477–487

Jukes TH, Bhushan V (1986) Silent nucleotide substitutions and G + C content of some mitochondrial and bacterial genes. J Mol Evol 24:39–44

Kawaguchi Y, Honda H, Taniguchi-Morimura J, Iwasaki S (1989) The codon CUG is read as serine in an asporogenic yeast *Candida cylindracea*. Nature 341:164

Keeling PJ (2016) Genomics: evolution of the genetic code. Curr Biol 26:R851–R853

Knight RD, Freeland SJ, Landweber LF (2001) Rewiring the keyboard: evolvability of the genetic code. Nat Rev Genet 2:49

Krakauer DC, Jansen VA (2002) Red queen dynamics of protein translation. J Theor Biol 218:97–109

Kumar V, Singh T (2020) Genome recoding: a review of basic concepts, current research and future prospects of virus attenuation for controlling plant viral diseases. J Plant Biochem Biotechnol. https://doi.org/10.1007/s13562-020-00583-8

Kuo J, Stirling F, Lau YH, Shulgina Y, Way JC, Silver PA (2018) Synthetic genome recoding: new genetic codes for new features. Curr Genet 64:327–333

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

Lajoie MJ, Kosuri S, Mosberg JA, Gregg CJ, Zhang D, Church GM (2013a) Probing the limits of genetic recoding in essential genes. Science 342:361–363

Lajoie MJ, Rovner AJ, Goodman DB et al (2013b) Genomically recoded organisms expand biological functions. Science 342:357–360

Lajoie MJ, Söll D, Church GM (2016) Overcoming challenges in engineering the genetic code. J Mol Biol 428:1004–1021

Lang BF, Lavrov D, Beck N, Steinberg SV (2012) Mitochondrial tRNA structure, identity, and evolution of the genetic code. Organelle genetics. Springer, Berlin, pp 431–474

Lau YH, Stirling F, Kuo J, Karrenbelt MA, Chan YA, Riesselman A et al (2017) Large-scale recoding of a bacterial genome by iterative recombineering of synthetic DNA. Nucleic Acids Res 45:6971–6980

Lemeignan B, Sonigo P, Marlière P (1993) Phenotypic suppression by incorporation of an alien amino acid. J Mol Biol 231:161–166

Li F, Zhang H, Sun Y, Pan Y, Zhou J, Wang J (2013) Expanding the genetic code for photoclick chemistry in *E. coli*, mammalian cells, and *A. thaliana*. Angew 52:9700–9704

Li JC, Nastertorabi F, Xuan W, Han GW, Stevens RC, Schultz PG (2019) A Single reactive noncanonical amino acid is able to dramatically stabilize protein structure. ACS Chem Biol 14(6):1150–1153

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

Ma NJ, Isaacs FJ (2016) Genomic recoding broadly obstructs the propagation of horizontally transferred genetic elements. Cell Syst 3:199–207

Macreadie IG, Novitski CE, Maxwell RJ, John U, Ooi BG, McMullen GL, Lukins HB, Linnane AW, Nagley P (1983) Biogenesis of mitochondria: the mitochondrial gene (aap1) coding for mitochondrial ATPase subunit 8 in *Saccharomyces cerevisiae*. Nucleic Acids Res 11:4435–4451

Mandell DJ, Lajoie MJ, Mee MT, Takeuchi R, Kuznetsov G, Norville JE et al (2015) Biocontainment of genetically modified organisms by synthetic protein design. Nature 518:55

Martínez MA, Jordan-Paiz A, Franco S, Nevot M (2016) Synonymous virus genome recoding as a tool to impact viral fitness. Trends Microbiol 24:134–147

Martrus G, Nevot M, Andres C, Clotet B, Martinez MA (2013) Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture. Retrovirology 10:78

Min B, Kitabatake M, Polycarpo C et al (2003) Protein synthesis in *Escherichia coli* with mischarged tRNA. J Bacteriol 185:3524–3526

Miranda I, Silva R, Santos MA (2006) Evolution of the genetic code in yeasts. Yeast 23:203–213

Moura GR, Carreto LC, Santos MA (2009) Genetic code ambiguity: an unexpected source of proteome innovation and phenotypic diversity. Curr Opin Microbiol 12:631–637

Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E (2006) Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. J Virol 80:9687–9696

Mukai T, Hayashi A, Iraha F, Sato A, Ohtake K, Yokoyama S, Sakamoto K (2010) Codon reassignment in the *Escherichia coli* genetic code. Nucleic Acids Res 38:8188–8195

Mukai T, Yamaguchi A, Ohtake K et al (2015) Reassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli*. Nucleic Acids Res 43:8111–8122

Mukai T, Lajoie MJ, Englert M, Söll D (2017) Rewriting the genetic code. Annu Rev Microbiol 71:557–577

Muto A, Osawa S (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. Proc Natl Acad Sci 84:166–169

Netzer N, Goodenbour JM, David A et al (2009) Innate immune and chemically triggered oxidative stress modifies translational fidelity. Nature 462:522

Nougairede A, De Fabritus L, Aubry F, Gould EA, Holmes EC, De Lamballerie X (2013) Random codon re-encoding induces stable reduction of replicative fitness of Chikungunya virus in primate and mosquito cells. PLoS Pathog 9:e1003172

Obata T, Shiraiwa Y (2005) A novel eukaryotic selenoprotein in the haptophyte alga *Emiliania huxleyi*. J Biol Chem 280:18462–18468

Ohtake K, Sato A, Mukai T, Hino N, Yokoyama S, Sakamoto K (2012) Efficient decoding of the UAG triplet as a full-fledged sense codon enhances the growth of a prfA-deficient strain of *Escherichia coli*. J Bacteriol 194:2606–2613

Ohtake K, Yamaguchi A, Mukai T, Kashimura H et al (2015) Protein stabilization utilizing a redefined codon. Sci Rep 5:9762

Osawa S, Jukes TH (1989) Codon reassignment (codon capture) in evolution. J Mol Evol 28:271–278

Osawa S, Jukes TH, Watanabe K, Muto A (1992) Recent evidence for evolution of the genetic code. Microbiol Mol Biol Rev 56:229–264

Ostrov N, Landon M, Guell M, Kuznetsov G et al (2016) Design, synthesis, and testing toward a 57-codon genome. Science 353:819–822

Pan T (2013) Adaptive translation as a mechanism of stress response and adaptation. Annu Rev Genet 47:121–137

Parrish AR, She X, Xiang Z, Coin I, Shen Z, Briggs SP, Wang L (2012) Expanding the genetic code of *Caenorhabditis elegans* using bacterial aminoacyl-tRNA synthetase/tRNA pairs. ACS Chem Boil 7:1292–1302

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

Rackham O, Chin JW (2005) A network of orthogonal ribosome·mRNA pairs. Nat Chem Biol 1:159

Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE et al (2017) Design of a synthetic yeast genome. Science 355:1040–1044

Riley R, Haridas S, Wolfe KH, Lopes MR, Hittinger CT, Göker M et al (2016) Comparative genomics of biotechnologically important yeasts. Proc Natl Acad Sci 113:9882–9887

Salim HM, Ring KL, Cavalcanti AR (2008) Patterns of codon usage in two ciliates that reassign the genetic code: *Tetrahymena thermophila* and *Paramecium tetraurelia*. Protist 159:283–298

Santos Moreno J, Schaerli Y (2019) Changing the biological Rosetta stone: the (commercial) potential of recoded microbes. Microb Biotechnol. https://doi.org/10.1111/1751-7915.13466

Schmied WH, Tnimov Z, Uttamapinant C, Rae CD, Fried SD, Chin JW (2018) Controlling orthogonal ribosome subunit interactions enables evolution of new function. Nature 564:444

Shabalina SA, Spiridonov NA, Kashina A (2013) Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. Nucleic Acids Res 41:2073–2094

Shen SH, Stauft CB, Gorbatsevych O, Song Y, Ward CB, Yurovsky A et al (2015) Large-scale recoding of an arbovirus genome to rebalance its insect versus mammalian preference. Proc Natl Acad Sci 112:4749–4754

Srinivasan G, James CM, Krzycki JA (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science 296:1459–1462

Subramaniam AR, DeLoughery A, Bradshaw N, Chen Y, O'Shea E, Losick R, Chai Y (2013) A serine sensor for multicellularity in a bacterium. Elife 2:e01501

Sueoka N (1962) On the genetic basis of variation and heterogeneity of DNA base composition. Proc Natl Acad Sci USA 48:582

Suzuki T, Ueda T, Watanabe K (1997) The 'polysemous' codon-a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. EMBO J 16:1122–1134

Swart EC, Serra V, Petroni G, Nowacki M (2016) Genetic codes with no dedicated stop codon: context-dependent translation termination. Cell 166:691–702

Teramoto H, Amano Y, Iraha F, Kojima K, Ito T, Sakamoto K (2018) Genetic code expansion of the silkworm *Bombyx mori* to functionalize silk fiber. ACS synth Biol 7:801–806

Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell 141:344–354

Venetz JE, Del Medico L, Wölfle A, Schächle P, Bucher Y, Appert D, Deutsch S (2019) Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. Proc Natl Acad Sci 116:8070–8079

Vetsigian K, Woese C, Goldenfeld N (2006) Collective evolution and the genetic code. Proc Natl Acad Sci 103:10696–10701

Wang L, Brock A, Herberich B, Schultz PG (2001) Expanding the genetic code of *Escherichia coli*. Science 292:498–500

Wang A, Nairn NW, Marelli M, Grabstein K (2012) Protein engineering with non-natural amino acids. Protein Eng. IntechOpen 5:6. https://doi.org/10.5772/28719

Wang B, Yang C, Tekes G, Mueller S, Paul A, Whelan SP, Wimmer E (2015) Recoding of the vesicular stomatitis virus L gene by computer-aided design provides a live, attenuated vaccine candidate. MBio 6:e00237-15

Wang L, Jiang S, Chen C, He W, Wu X, Wang F et al (2018) Synthetic genomics: from DNA synthesis to genome design. Angew Chem 57:1748–1756

Wang W, Cheng X, Buske PJ, Suzich JA, Jin H (2019) Attenuate Newcastle disease virus by codon modification of the

glycoproteins and phosphoprotein genes. Virology 528:144–151

Weiner AM, Weber K (1971) Natural read-through at the UGA termination signal of Qβ coat protein cistron. Nat New Biol 234:206

Wimmer E, Mueller S, Futcher B, Shen S, Stauft C, Ward C (2019) Recoded arbovirus and vaccines. Vaccine 38:2943–2948

Yamao F, Muto A, Kawauchi Y, Iwami M, Iwagami S, Azumi Y, Osawa S (1985) UGA is read as tryptophan in *Mycoplasma capricolum*. Proc Natl Acad Sci 82:2306–2309

Zaher HS, Green R (2009) Fidelity at the molecular level: lessons from protein synthesis. Cell 136:746–762