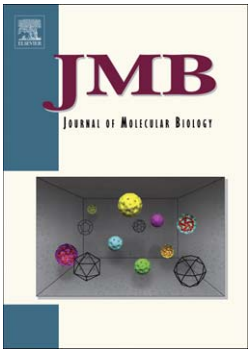


Overcoming challenges in engineering the genetic code

M.J. Lajoie, D. Söll, G.M. Church

PII: S0022-2836(15)00492-1
DOI: doi: 10.1016/j.jmb.2015.09.003
Reference: YJMBI 64846

To appear in: *Journal of Molecular Biology*

Received date: 2 June 2015
Revised date: 19 August 2015
Accepted date: 1 September 2015

Please cite this article as: Lajoie, M.J., Söll, D. & Church, G.M., Overcoming challenges in engineering the genetic code, *Journal of Molecular Biology* (2015), doi: 10.1016/j.jmb.2015.09.003

Overcoming challenges in engineering the genetic code

Lajoie MJ[1,2,‡], Söll D[3], Church GM[1,4]

1. Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

2. Program in Chemical Biology, Harvard University, Cambridge, MA 02138, USA

3. Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520-8114, USA

4. Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA

To whom correspondence should be addressed:
E-mail: mlajoie@genetics.med.harvard.edu
[‡]Present address: Department of Biochemistry, University of Washington, Seattle, WA 98195, USA.

Abstract: Withstanding 3.5 billion years of genetic drift, the canonical genetic code remains such a fundamental foundation for the complexity of life that it is highly conserved across all three phylogenetic domains. Genome engineering technologies are now making it possible to rationally change the genetic code, offering resistance to viruses, genetic isolation from horizontal gene transfer, and prevention of environmental escape by genetically modified organisms. We discuss the biochemical, genetic, and technological challenges that must be overcome in order to engineer the genetic code.

Keywords: genomically recoded organism (GRO); expanded genetic code; genome engineering

Abbreviations: genomically recoded organism (GRO), non-standard amino acids (nsAAs), aminoacyl-tRNA synthetases (aaRSs), multiplex automated genome engineering (MAGE), conjugative assembly genome engineering (CAGE)

Introduction

Nearly all organisms share a common genetic code—the language that specifies how genetic information is interpreted to produce proteins. Although some organisms and mitochondria use non-canonical genetic codes, they typically involve a small number of codon reassignments in response to strong selective pressures [1]. For example, mitochondria are under selection to minimize genome size [1,2] and G+C content [3], while bacteria may use codon reassignment to evade viruses or to restrict horizontal gene transfer [4,5].

Why has the genetic code been so refractory to change? While it is possible that more divergent genetic codes have yet to be discovered [5], several factors help to conserve the genetic code. Evolution tends to increase biological complexity [6], leading to the large genomes of modern free-living organisms (the smallest known genome with a full complement of essential genes has 580,070 base pairs encoding 470 predicted open reading frames [7]). With a few exceptions [4,8–10], these organisms use all 64 codons to encode their proteins while simultaneously accommodating overlapping sequence features such as protein binding sites, promoters, splicing signals, and RNA secondary structure [11]. In this context, the genetic code provides fundamental biochemical constraints that guide how a genome is put together. Therefore, the genetic code shapes how mutations affect an evolving genome, while the genome relies on a stable genetic code to faithfully produce proteins essential for life. Any change in codon function must be tolerated at all instances genome-wide. Therefore small changes in the genetic code must be accompanied by many compensatory changes in the rest of the genome. This scenario is unlikely to occur by random mutagenesis, but current genome engineering technologies are now making it possible to rationally change the genetic code.

What does it mean to "change the genetic code?"

The term "genetic code" has been used with different meanings in different contexts. Since this can lead to confusion, we propose less ambiguous alternatives for "changing the genetic code." (1) Changing the genome sequence (*e.g.*, synonymous codon swaps in one or more genes) will be referred to as genome editing. (2) Introducing new amino acid assignments of one or more codons without removing the original function (*e.g.*, UAG decoded as both a stop and an amino acid [12]) will be referred to as codon suppression. (3) Changing the amino acid assignments of one or more codons genome wide (*e.g.*, genomically recoded organisms [13]) will be referred to as codon reassignment. (4) Adding a new codon to the translation code table (*e.g.*, using quadruplet codons [14–16] or codons composed of unnatural bases [17]) will be referred to as codon creation. Broadly speaking, the term "genetic code" will be used throughout this review to describe the codon assignments in the translation code table. Codon suppression, reassignment, and creation are all ways of changing a genetic code.

Applications of changing the genetic code

Since the 1970s, biotechnology has relied on a ubiquitous code to permit the transgenic production of drugs [18], materials [19], and food [20]. Current DNA synthesis technologies [21–23] have liberated us from reliance on the canonical genetic code, yet we are still subjected to its challenges—viral infection and undesired horizontal gene transfer (*e.g.*, antibiotic resistance [24], recombinant DNA dissemination [25–27]). Furthermore, while the 20 canonical amino acids support an astonishing diversity of biochemistry, they nevertheless limit the potential for new and useful protein functions. This fact is exemplified by the natural existence

of selenocysteine and pyrrolysine as specialized catalytic residues for redox reactions and methanogenesis [1].

Organisms possessing alternate genetic codes have the potential to overcome these challenges (Figure 1). This hypothesis motivated the development of genomically recoded organisms (GROs), whose codons have been unambiguously reassigned to create alternate genetic codes [13]. GROs are able to efficiently incorporate non-standard amino acids (nsAAs) [28] that have been developed to enhance enzyme activity [29,30], to improve the performance of drugs [31,32], and to function as molecular probes [33]. Redesigning essential proteins to depend on these nsAAs for proper translation and function provides a robust strategy for restricting undesired survival outside of controlled environments [34,35]. Additionally, by interpreting genetic information differently, GROs would mistranslate foreign genes from natural organisms. This would prevent viruses from hijacking their translation machinery (potentially saving hundreds of millions of dollars from lost productivity [36,37]) and would thwart transfer of functional genetic information with natural organisms [13]. Thus, GROs have the potential to be safe and powerful chassis for biofermentation, bioremediation, and agriculture.

We recently reported the first GRO, which has an unambiguously reassigned UAG codon. However, while this GRO has demonstrated promising properties such as increased virus resistance [13] and metabolic dependence on nsAAs [34,35], evolution can readily overcome genetic isolation from one reassigned stop codon. Indeed, UAG suppression is well-documented in natural organisms [5], and bacteriophage T7 readily evolves tolerance to a host with a reassigned UAG codon [38]. Therefore, radically altered genetic codes will be required in order to realize the full potential of these applications [39]. While only a subset of sense codons likely need to be reassigned in order to achieve robust genetic isolation, this review will consider the

fundamental limits to changing the genetic code based on biochemical principles, genetic knowledge, and genome engineering technologies.

Engineering expanded genetic codes

In vitro translation systems offer the ultimate flexibility to implement alternate genetic codes [40–42]. Since translation components can be prepared separately, non-specific aminoacylation methods such as CA ligation [43] and flexizyme [44] can be used to incorporate a broad selection of nsAAs. As a result, *in vitro* translation has been the best way to produce unnatural backbones [45,46] for synthetic nonribosomal peptide mimetics [47,48] and polymer materials [46,49]. Furthermore, the production of ribosomes *in vitro* may accommodate extensive modifications that could otherwise compromise fitness *in vivo* [50]. Recent reports of orthogonal 16S rRNA [51], orthogonal tRNAs [52], tethering 16S to 23S rRNA [53], provide the infrastructure to evolve ribosomes with radically modified functions.

Codon suppression

In vivo systems are well-suited for inexpensive, simple, and scalable translation using nsAAs, but must be compatible with essential cellular processes. While sense codons have been transiently diverted to incorporate diverse nsAAs by metabolic labeling [54], persistent metabolic labeling is likely to be highly deleterious. Even evolving tolerance for structurally similar Trp analogs has met varying success in different systems [55–58]. In contrast, ambiguous decoding of stop codons is well-tolerated in *E. coli* [12,59], making it possible to introduce orthogonal translation machinery capable of producing high yields of nsAA-containing proteins *in vivo* [60–62]. The implementation of orthogonal translation machinery [33] has led to an

explosion in the number of nsAAs (currently more than 167 nsAAs [28]) that can be site-specifically incorporated into proteins for applications in medicine [31] and bioremediation [30].

Codon reassignment

While ambiguous decoding has long made it possible to produce nsAA-containing proteins, only recently has the translation function of a codon been unambiguously reassigned, enabling the sustained expression of proteins containing one or more nsAAs [13,63,64]. While a surprisingly small number of changes permit the disruption of UAG termination [63,65], the remaining natural UAG codons provide a selective pressure for efficient UAG translation. This destabilizes the genetic code by selecting for spontaneous suppressor mutations that incorporate canonical amino acids at UAG codons [13]. This strategy could prove even more problematic for sense codon reassignment, since stop codons only occur once at the end of genes, limiting the impact of codon reassignment on the proteome [1]. Therefore, the most general strategy to expand the genetic code using reassigned codons involves (1) identifying all genomic instances of a target codon, (2) replacing them with synonymous codons, (3) abolishing the target codon's natural function by inactivating its translation factors, and (4) introducing new translation function by integrating orthogonal translation systems, and (5) introducing new instances of the target codon to specifically and efficiently incorporate nsAAs into desired proteins (Figure 2A-D) [13]. Using this strategy, expanded genetic codes can be stabilized by redesigning essential proteins to functionally depend on a specific nsAA for survival [34,35]. However, it remains a major biochemical, genetic, and technical challenge to reassign codons that are commonly utilized throughout a genome.

Codon creation

Beyond repurposing one or more of the existing 64 codons, it may be possible to add a new base pair [66–70] or to engineer quadruplet [14–16,71,72] or quintuplet [73] genetic codes, which could give $6^3 = 216$, $4^4 = 256$, or $5^4 = 625$ codons, respectively. Indeed, exciting progress has hinted at the promise of creating new codons, which would need to be replicated, transcribed, and translated. Several unnatural base pairs exhibiting high fidelity replication by PCR and compatibility with proofreading mechanisms of Exo+ polymerases have been developed [67–70]. Additionally, transcription (using T7 RNA polymerase) [74–77] and reverse transcription [77] have been demonstrated. Finally, codons containing unnatural base pairs have been implemented to translate peptides containing unnatural amino acids using an *E. coli*-derived *in vitro* translation system [17,78]. This means that codons containing unnatural base pairs can be immediately implemented for *in vitro* translation of proteins containing nsAAs.

Recently, Malyshev et al. [79] have taken a crucial step toward *in vivo* implementation of unnatural base pairs with the demonstration that the d5SICS-dNaM base pair can be replicated *in vivo* [79]. Still, several major challenges must be overcome to fully implement an unnatural base pair for *in vivo* translation. Malyshev et al. [79] demonstrated that the bioavailability of nucleoside triphosphates is crucial and that heterologous transporters provide one solution to this problem. Additionally, stability is crucial to prevent loss of information. Replication error rates below 10^{-3} per bp per replication have been recommended for PCR [80], but replication fidelity better than 10^{-8} per bp per replication may be necessary *in vivo* in the absence of a strong selective pressure to maintain the unnatural base pair (as has been demonstrated for nsAAs [34,35]). The fidelities of transcription and translation are more flexible as long as they do not interfere with normal cell function, but net translational fidelity should be comparable with that of current suppression systems [81,82]. Finally, unnatural base pairs must be compatible with

essential components of the host replication, transcription, and translation machinery. Fortunately, previous studies have demonstrated that codons containing unnatural base pairs are compatible with the *E. coli* translation system reconstituted *in vitro* [17,78]. However, while thermostable PCR enzymes have been used for replication *in vitro* [67–70] and *E. coli* PolI has been implemented *in vivo* [79], compatibility with *E. coli* PolIII replication has yet to be demonstrated. Additionally, T7 RNA polymerase is routinely used for transcription *in vitro* [74–77], and *E. coli* RNA polymerase transcription has yet to be demonstrated.

Similarly, improved systems for quadruplet decoding have improved translation yields to a level rivaling reassigned triplet systems [16,83]. However, wobble rules for quadruplet and quintuplet systems are not yet well-understood [84,85], requiring an increased reliance on empirical validation of anticodon specificities. Furthermore, triplet codons must still be removed to prevent ambiguous decoding in a genetic code that utilizes both triplet and quadruplet codons [16]. Taken together, significant challenges must still be overcome to implement sustained *in vivo* translation of novel codons. We expect there to be many exciting advances in this field over the next several years.

Biochemical barriers

Ribosomes translate proteins by adding an amino acid to the nascent polypeptide in response to three-nucleotide codons on messenger RNAs (mRNAs). The identity of the amino acid is controlled at multiple discrete steps. First, aminoacyl-tRNA synthetases (aaRSs) charge transfer RNAs (tRNAs) with their correct amino acids. The aminoacyl-tRNA is then shuttled into the active site of the ribosome by elongation factor Tu (EF-Tu), where base pairing between the mRNA codon and tRNA anticodon allows transfer of the amino acid onto the nascent peptide

chain regardless of the amino acid identity. Translation involves more than 100 proteins and RNAs, a subset of which have been engineered to expand the genetic code (*e.g.*, [33,51–53,62,86,87]) (Figure 3). Together with insights from natural non-canonical genetic codes [1,88–91], this work suggests that the biochemistry of the genetic code is remarkably flexible.

Studying natural codon reassignment can provide insights into how to change the genetic code. Interestingly, many of the same codon reassignments appear to have independently evolved several times, suggesting that certain codons have a predisposition for reassignment [1,92]. Stop codons may be favored because they are only used once at the end of genes, so their reassignment is expected to cause minimal damage to the proteome [1]. Indeed, suppression of the stop codons is well-tolerated in *E. coli* [12] and broadly observed in metagenomic data [5]. Alternatively, mitochondrial genomes are under strong selective pressure to reduce genome size and G+C content, which can lead to the genomic depletion of certain codons [2,3]. When this occurs, small tweaks to post-transcriptional anticodon modifications can change codon assignments without affecting aaRS recognition [1,92–94]. For example, a 7-methylguanosine modification on $tRNA^{Ser}_{GCU}$ allows it to decode all four AGN codons in squid mitochondria [1]. In this way, small mutations in tRNA anticodons or changes to their post-transcriptional modifications can change the genetic code [92]. Our analysis will focus on the minimal and maximal variants of the canonical genetic code using *E. coli* as an example (Figure 4). Although the proposed genetic codes may be far from optimal [11,92,95] and challenging to implement, it is instructive to consider how to change the number of unique anticodons that can be achieved based on existing translation machinery.

Minimal genetic code

E. coli strain MG1655 has 87 tRNA genes (with 42 unique tRNA anticodons; including fMet and selenocysteine) and two release factors that unambiguously decode all 64 codons and incorporate all 20 canonical amino acids (Figure 4A) [94]. In contrast, a minimal genetic code requires one tRNA for each of 20 amino acids, a formylmethionine $tRNA^{fMet}$ for translation initiation [96], and a release factor for translation termination (Figure 4B). First, Cys, Trp, Met, fMet, Asp, Glu, Lys, Asn, Gln, His, and Tyr could use one of their natural tRNAs without any anticodon modifications. This is particularly helpful for *E. coli* $tRNA^{Glu}$, $tRNA^{Lys}$, and $tRNA^{Gln}$, which utilize mnm^5S^2U to recognize their full complement of codons [93]. Therefore, even though GluRS requires mnm^5S^2U for efficient glutamylation [97], no aaRS/tRNA engineering is necessary for a single tRNA (naturally modified with mnm^5S^2U) to decode all necessary codons for these amino acids. Ile requires two tRNAs with non-redundant anticodons, and selenocysteine is not essential in *E. coli*. Additionally, as demonstrated by mitochondrial genetic codes, an unmodified uracil in the anticodon wobble position can recognize all four codons in a family group (codons that are identical at the first two positions and differ at the third position) [89]. Therefore, 10 tRNAs with a uracil in the anticodon wobble position could unambiguously assign 40 codons to decode 9 amino acids (Arg and Ser each have six codons, so the AGN family group is redundant; see Figure 4B). This leaves all three stop codons, which can be recognized by a single E167K release factor 2 variant [98]. Therefore, simply by mutating the anticodon wobble position to uracil in 10 tRNAs and deleting all redundant tRNAs, a minimal genetic code would require 23 tRNAs and one release factor in order to decode all 64 codons. More radically, it may be possible to achieve adequate protein function using a code composed of fewer than 20 amino acids [99–102]. Preliminary studies propose that Ile [103] and Trp [104]

could be replaced by natural amino acids that have similar side chains. If Ile and Trp are removed and blank codons are tolerated, the minimal genetic code would require only 19 tRNAs and one release factor (Figure 4B).

Maximal genetic code

To expand the genetic code, codons that are naturally used for canonical amino acids must be reassigned to have new functions. This requires the inactivation of natural translation machinery and the introduction of orthogonal aaRS/tRNA pairs that can decode the reprogrammed codons [62]. Despite the complexity of protein translation, simply modifying tRNA anticodons could free up enough codons to more than double the number of amino acids in the genetic code. While many different strategies could expand the genetic code, we propose three tiers of tRNA manipulations to systematically maximize the number of unique codon assignments (Figure 4C-D).

Tier 1: Simply by leveraging the degeneracy of genetic code, up to ten anticodons could be reassigned in order to provide seven unambiguous and three ambiguous anticodons for nsAA incorporation (Figure 4C). GUN, GCN, and CCN were not included in this list, since the cmo^{5}U-containing anticodon has been empirically shown to recognize all four codons in their respective family groups [94]. Although reassigning the codons chosen in Figure 4C require considerably more genome modifications in *E. coli* MG1655 compared to the most efficient strategy, we chose these codons because they can be further fractionated by changing their anticodon modifications (Figure 4D). A conceptually simpler strategy is outlined in Figure S1. These changes are likely adequate for complete genetic isolation from multiple viruses and horizontal gene transfer.

Tier 2: Inconveniently, six family groups (CUN, GUN, UCN, CCN, ACN, and GCN) use anticodons with overlapping codon specificity, making it difficult to unambiguously reassign their functions. The overlapping tRNA specificities are caused by cmo^5U wobble bases, which are able to base pair with A, G, U, and sometimes C [105]. Unlike these six family groups, UUN and GGN use alternative wobble bases that allow two unambiguous anticodons in each family group. By inactivating tRNA-modifying enzyme CmoB (modifies U34 of tRNAs with cmo^5U) [105] and reengineering MnmE and MnmG to modify U34 of additional tRNAs (naturally modifies $tRNA^{Gln}$, $tRNA^{Lys}$, $tRNA^{Glu}$, and $tRNA^{Arg}$ with mnm^5) [106,107], six ambiguous codons could be disambiguated, accommodating up to 33 total amino acids (Figure 4D, blue features). Although no attempts to reengineer the specificity of tRNA-modifying enzymes have been reported, genetic code expansion might provide an incentive to try.

Tier 3: In wild type *E. coli*, the AUA (Ile) and AUG (Met) codons are unambiguously decoded. Similarly, the other NNR codons could be split into unique singlet codons by exploiting anticodons modified with lysidine (specifically base pairs with A) and cytosine (specifically base pairs with G) to decode NNA and NNG codons, respectively (Figure 4D, magenta features). In order to accomplish this, TilS would need to be engineered to lysidinylate more anticodons in addition to its natural target, $tRNA^{Ile}$ [108]. Although wobble codons do not usually coincide with tRNA identity determinants, lysidine is a crucial identity determinant for IleRS and a crucial antideterminant for MetRS [109]. Therefore, it could potentially impact the orthogonality of heterologous tRNAs introduced for genetic code expansion. Additionally, mnm^5-modified wobble bases are minor tRNA identity determinants for GluRS, GlnRS, and LysRS [110]. Reduced aminoacylation could be addressed by co-evolving the aaRS/tRNA pair to better recognize the modified anticodon loop [111]. Finally, by engineering one of the release factors to

terminate translation exclusively at UAA codons, both UAG and UGA could be reassigned to incorporate nsAAs. Combined with the changes proposed for tier 1 and tier 2, these proposed changes could produce 27 new anticodons (up to 47 total amino acids). While the NNY codons cannot be split into singlet codons based on known anticodon modifications, such modifications have not been ruled out.

aaRS/tRNA reassignment to expand the amino acid repertoire

The aaRSs are evolutionarily ancient enzymes crucial for transmission of the genetic message [112]. They display superb specificity against all metabolites in the cell, and possess editing and proofreading mechanism to correct mistakes in aminoacylation [113]. However, they show little specificity against the large number of nsAAs that are used in genetic code expansion studies [114,115]. This 'polyspecificity' is observed in all the orthogonal aaRS/tRNA systems that have been developed [116–119]. This fact must be kept in mind when the design of better aaRSs is planned, and will be an issue in the production of proteins containing multiple different nsAAs. In such cases, it may be necessary to further evolve aaRS/tRNA pairs not simply to be orthogonal from the organism's natural aaRS/tRNA systems, but also to be mutually orthogonal from each other [111].

Once codons have been liberated for reassignment, new translation machinery must be introduced in order to expand the amino acid repertoire. Orthogonal aaRS/tRNA pairs have been evolved to specifically and efficiently incorporate more than 167 nsAAs into proteins, while minimizing interactions with endogenous aaRS/tRNA pairs (extensively reviewed [4,28,33,120]). In some respects, the ability to incorporate more than 167 nsAAs with diverse functional groups suggests that creating custom orthogonal aaRS/tRNA pairs is a solved

problem. However, nearly all nsAAs are analogs based on a small number of natural amino acids, and they have been developed for a small number of tRNA anticodons that lack strong aaRS specificity determinants. Furthermore, activity and specificity remain low compared to natural aaRS/tRNA pairs [121,122], mainly because current over-expression systems [81,82] do such a great job of outcompeting natural codon function that they compensate for low aminacylation activity. This results in orthogonal translation systems that are adequate for batch protein production, but are suboptimal for sustained translation with nsAAs [34,121]. Especially in strains with unambiguously reassigned codons [13], reducing expression levels to more closely match natural aaRS/tRNA pairs could provide an incentive to evolve orthogonal systems exhibiting superior aminoacylation efficiency and specificity without having to change the standard selection systems (see ref. [33]) used for orthogonal aaRS optimization. Taken together, aaRS/tRNA engineering is certainly not a solved problem.

In addition to reducing aaRS/tRNA expression levels, engineering improved aminoacylation efficiencies and specificities will require more sophisticated design methods. To date, aaRS directed evolution libraries have been constructed based on manual inspection of crystal structures, and the combinatorics have limited the number of residues that can be randomized to those that directly contact the ligand (six fully randomized residues is $20^6 = 6.4 \times 10^7$ unique sequences). Given that second and third shell interactions crucially stabilize binding competent conformations [121,123], expanding the search space could lead to better-optimized enzymes. Although enzyme design remains extremely difficult, notable successes [124–126] provide encouragement that protein design software [127] could narrow the search space enough to probe additional residues in the second and third coordination shells of nsAA binding pockets, while maintaining realistic population sizes for directed evolution. More complex mechanisms of

amino acid and tRNA determination, such as the extensive hydrogen bond network in *E. coli* GlnRS [128], have been addressed by transplanting key elements from other aaRSs and tRNAs to alter specificity [121]. Taken together with the robustness of aaRSs (extensive redesign of aaRS cores is tolerated [34,121]), structural and mechanistic insights can provide actionable information for rational aaRS/tRNA engineering [129].

Fidelity during protein translation requires all aaRSs to be selective both for their amino acids and their tRNAs [129]. In order to meet these requirements, orthogonal aaRS/tRNA pairs must have tRNA identity determinants that differ from those in the target organism [62]. Unfortunately, the majority of *E. coli* tRNAs have identity determinants in their anticodons [130]. Therefore, isolating orthogonal tRNAs that are not charged by endogenous *E. coli* aaRSs is a considerable challenge. For instance, a heterologous $tRNA^{Pyl}_{CCG}$ is mischarged by ArgRS, presumably due to anticodon recognition [131]. In contrast, suppressors of the UAG stop codon have been more successful, since *E. coli* lacks an aaRS that recognizes the CUA anticodon. For this reason, the *M. jannaschii* TyrRS/$tRNA_{CUA}$ [28,33,62] and the *M. barkeri* PylRS/$tRNA^{Pyl}_{CUA}$ [132,133] pairs have become the most popular systems for nsAA incorporation in *E. coli*, although orthogonal aaRS/$tRNA_{CUA}$ pairs have also been isolated for Lys [14], Glu [134], Leu [135], Pyl [136], Trp [137,138], Pro [139], and His [140]. Encouragingly, the *E. coli* tRNA identity determinants have been extensively characterized [110,130], and their antideterminants have been predicted [141], providing a starting point for the directed evolution of additional orthogonal aaRS/tRNA pairs. Additionally, preliminary work with the orthogonal selenocysteine translation machinery provides encouraging results for 60 of the 64 codons [142,143].

Although allosteric amino acid discrimination mechanisms [144], tRNA anticodon recognition [130], and editing domains [129] may still prove difficult to overcome while

reengineering aaRSs, metagenomics offers a rich source of aaRS/tRNA pairs that can differ considerably among organisms separated by large evolutionary distances [129]. From a synthetic point of view, domesticating and reengineering only a small subset of the available aaRS/tRNA pairs available in nature is enough to radically expand the genetic code.

Optimal genetic codes minimize the impact of mutational and translational errors

Even if codons can be liberated and reassigned by orthogonal aaRS/tRNA pairs, the error minimization theory suggests that extensively changing the genetic code could be deleterious [11]. Similar amino acids are grouped with similar codons in the canonical genetic code to minimize the effects of spontaneous mutations and translation errors, smoothening the fitness landscape to facilitate evolution [145,146]. The canonical code does this remarkably well [95,147], utilizing all 64 codons for translation throughout the proteome and providing a disincentive for genetic code expansion [116]. While the natural persistence of less-optimal genetic codes [92,95] demonstrates that error minimization is not strictly essential, it nevertheless suggests that GROs should be made explicitly dependent on the expanded genetic code [34,35] in order to overcome the countervailing evolutionary pressure for error minimization. It remains to be seen to what extent expanded genetic codes will compromise fitness and whether the resulting GROs will evolve more accurate ribosomes.

Genetic barriers

Despite the impressive biochemical flexibility of the genetic code, our inadequate understanding of how to design genomes remains a major barrier for creating organisms with radically new genetic codes. Even in the age of chemically synthesized chromosomes [148–150],

genomes must be designed based on incomplete information, and even the best-annotated genomes remain incompletely understood at all levels of complexity from single nucleotide variants to genome architecture (Figure 5).

Single nucleotide variants

To a large extent, the effect of single nucleotide mutations can be predicted based on how they change the coding DNA sequence—frameshifts, premature stop codons, and non-conservative amino acid changes are more likely to disrupt function than are synonymous changes [151]. From this perspective, synonymous codon swaps should be unlikely to be deleterious [13,39]. However, while most synonymous mutations are well-tolerated [13,39], disrupting overlapping sequence features such as ribosome binding motifs [152] and small RNAs [153] can impact crucial cellular functions. Additionally, synonymous mutations are sometimes unpredictably rejected even when non-synonymous mutations are not [39], perhaps due to mRNA structure [154,155] or codon usage preferences [155,156]. Metagenomic conservation data provides a valuable opportunity to de-risk specific mutations [157], but only when the new sequence closely resembles a natural sequence.

Isolated genetic components

Complex systems can be better understood by breaking them into component parts and establishing models to predict the function of each part. Automated annotation remains a major challenge, requiring empirical testing to characterize the function of genetic components. Furthermore, while recent advances have improved our understanding of promoters [158,159], ribosome binding sites (RBSs) [159,160], and terminators [161], sequence context can strongly impact function [155], requiring empirical testing for each genetic circuit.

Isolated genetic pathways

Improved understanding of genetic parts has made it possible to refactor pathways in order to replace cryptic regulation mechanisms with more modular and predictable systems [162–164]. The premise of refactoring is that the functions of non-coding DNA sequences are difficult to predict, so it is easier simply to replace them. This becomes especially important when the goal is to transfer such pathways into heterologous hosts. However, since it remains difficult to predict how gene expression translates to pathway behavior, the iterative process of testing and redesigning biological pathways remains an important strategy for building systems from defined genetic parts [163,164].

Genome construction

Engineering at the genome scale encompasses all of the challenges of engineering its parts plus the added the challenge of preserving essential cellular functions, replication, transcription, DNA structure, and DNA repair. Some notable studies have taught important lessons about engineering genomes. For example, the separation of overlapping coding DNA sequences in T7 bacteriophage [162] provided a proof of concept for modular genomes that are easier to manipulate. The removal of mobile DNA elements and cryptic virulence genes from *E. coli* produced a more genetically stable genome [165]. Integration of the *Synechocystis* PCC6803 genome into *B. subtilis* demonstrated the importance of balancing replichore size, while highlighting the challenges of preserving the function of each species as part of a chimeric genome [166]. Finally, the introduction of large genome rearrangements was used to explore the effects of splitting genomes into multiple chromosomes [167] and to demonstrate the importance of co-orientating transcription with replication in circular chromosomes [168]. Future work will characterize additional crucial genome-scale features that are essential for viability (*e.g.*, although each individual instance of DNA gyrase [169] and chi [170] sites may not be essential,

their presence throughout the genome is crucial for genome maintenance). While it has been difficult to produce a minimal genome based on information learned from single gene knockout studies [171], continued work in this area [7,148,171–175] is likely to uncover additional genome design rules involving synthetic lethality (two mutations are tolerated individually, but are lethal when combined [176]), operon structure, and genome structure.

Genome engineering barriers

Despite much progress, it remains difficult to predict the correct changes to make at every level of genome complexity from single nucleotide changes to megabase/gigabase genome construction. Accepting the fact that existing information is inadequate, draft genomes must be a best guess based on as much information as possible. To put this in perspective, a logical next step in genome recoding is to reassign all instances of the rarest sense codons in *E. coli* [39]. However, with 4228 AGR codons, exhaustive sampling of all synonymous CGN codon alternatives would require testing of 4^{4228} genomes.

Given our tenuous understanding of how to design genomes, effective genome engineering technologies must integrate the information that is known, and overcome the inevitable design flaws that will arise based on our incomplete knowledge [39,177]. We know that all natural instances of a codon must be removed from the genome in order to abolish its natural translation function [13]. We know that orthogonal aaRS/tRNA pairs [62] can introduce new translation functions [28]. We know that we can stabilize an expanded genetic code by establishing functional dependence on an unnatural amino acid [34,35]. The challenge is to produce such genomes by making hundreds or thousands of changes without introducing any lethal design flaws.

The past decade has seen many impressive achievements in genome engineering (reviewed in [178]), although few attempts have been made to produce new sequences that cannot be found in nature. The *de novo* synthesis and transplantation of an intact *Mycoplasma mycoides* JCVI-syn1.0 genome demonstrated that a small, natural prokaryotic genome can be built from simple chemical components [148]. Such an approach could allow the synthesis of any user-defined genome sequence. However, genome design remains the major barrier because even a single design flaw could prevent the function of the entire genome [148]. Given the high stakes for design flaws, *de novo* genome synthesis is most effectively used in combination with exhaustive empirical tests [7,171,172,175] and complete computational models [179].

Genome engineering technologies

Engineering the genetic code requires extensive genome manipulation that can affect fitness in unpredictable ways [39]. With this in mind, we have developed multiplex automated genome engineering (MAGE) [180] and conjugative assembly genome engineering (CAGE) [181] for rapidly prototyping and manufacturing genotypes *in vivo*. MAGE uses the λ bacteriophage β recombinase and ssDNA oligonucleotides [182] to simultaneously introduce multiple defined mutations at multiple locations throughout a replicating bacterial genome [180]. Meanwhile, CAGE uses bacterial conjugation to precisely transfer up to several million base pairs of contiguous DNA [181], allowing the production of extensively modified genomes from small segments that are easily prototyped in parallel using MAGE. Together, MAGE and CAGE exploit evolution to combinatorially explore a broad pool of synthetically defined genotypes *in vivo*, allowing natural selection to remove deleterious design flaws from the population.

MAGE and CAGE were used to remove all 321 known instances of the UAG codon from *E. coli* MG1655 at a fraction of the predicted cost for genome synthesis [13]. Still, DNA

synthesis can be invaluable for extensively modifying genome sequences, provided that the synthetic genome fragments are small enough for efficient troubleshooting. For instance, we tested 6496 total mutations across 42 essential genes [39] using inexpensive, chip-based DNA synthesis [22]. Because we tested each essential gene individually, design flaws could be rapidly mapped and overcome using MAGE [39]. A similar strategy has been successful for the synthetic yeast 2.0 project [150] and could be extended to diverse organisms using an ever-growing arsenal of powerful genome engineering tools [183].

As genome designs increase in complexity, integrated CAD/analysis strategies will become essential for monitoring design clashes, managing genome builds, and analyzing genotypes. There are many useful genome engineering design and analysis tools available: searchable genome annotation databases [184], automated MAGE oligo [181] or CRISPR [185] design tools, sequence manipulation and synthetic circuit design tools [186], sequencing analysis tools for single nucleotide variants [151,187] and structural variants [188–190], and computational models of whole organisms [179]. Integrating these design and analysis tools into a cohesive and efficient software platform will greatly benefit efforts to produce GROs with radically altered genetic codes.

Outlook and conclusions

While more than 167 nsAAs have already vastly expanded protein function [28], radically different genetic codes will be required to achieve virus resistance, genetic isolation, and stable expansion of the genetic code. Thirteen out of thirteen codons tested have already shown promise for reassignment [39]. To implement radically expanded genetic codes, a mechanistic understanding of biochemical principles will be crucial to engineer orthogonal

translation machinery that is capable of reassigning such sense codons. Additionally, genome engineering methods capable of interrogating genetic landscapes containing thousands of potentially deleterious changes will be crucial for producing organisms with reassigned sense codons [13,39]. Advances in understanding codon usage [155,191], gene function [171,172,175], operon structure [159,163,164], and genome architecure [162,165–168] will help establish better guidelines for genome design, but diversity will remain a crucial aspect in prototyping genomes with new and useful biological functions.

Figure captions:

Figure 1. Expanded biological functions of GROs. GROs provide dedicated codons for efficient translation of nsAAs at multiple sites in a protein [13,64], increase resistance to bacteriophages [13], and enable biocontainment/niche restriction [34,35].

Figure 2. Engineering a GRO and its properties. GROs are produced by **(A)** identifying all instances of a target codon (*e.g.*, UAG = stop) in the genome of the starting organism, **(B)** replacing all instances of the target codon with a synonymous codon (*e.g.*, UAA = stop), **(C)** deleting translation factors responsible for decoding the target codon (*e.g.*, release factor 1 terminates translation at UAG and UAA codons), **(D)** introducing orthogonal translation machinery capable of decoding the target codon with a nsAA and then reinserting the reassigned codon into the genome [13].

Figure 3. Components of the translation system have been engineered *in vivo*. The amino acid repertoire has been expanded by more than 167 nsAAs through directed evolution of orthogonal aaRS/tRNA pairs [28,33]. Amino acid specificity has been tuned by directed evolution of the aaRS amino acid binding pocket [62]. Meanwhile, target codon specificities have been controlled by altering the acceptor stem of the tRNA [14], or by mutually evolving the tRNA anticodon and the aaRS anticodon loop binding domain [111]. Additionally, orthogonal ribosomes have been engineered to exhibit useful properties that would also be deleterious for translating the rest of the proteome. An orthogonal 16S rRNA with a modified anti-Shine-Dalgarno sequence has been evolved to promote better UAG [51] and AGGA [15] suppression, an orthogonal 23S rRNA has been modified to accept orthogonal tRNAs with modified 3' ends [52], and a 16S-23S tethered rRNA has been developed to facilitate engineering of peptidyl transferase activity [53]. Finally, release factors have been engineered to recognize additional codons [98], and EF-Tu has been modified to accept highly charged amino acids [87]. Although, tRNA-modifying enzymes such as CmoB, MnmE, and MnmG have not yet been reengineered, they offer the potential to radically alter the genetic code (see Figure 4).

Figure 4. Minimal and maximal genetic codes using triplet codons composed of four nucleotide types (U, C, A, G). The proposed genetic codes are one possible permutation representing several possible ways to reassign redundant codons. Dashed brackets represent codon recognition ranges for a given anticodon: black is codon recognition agreeing with wobble rules [192,193]; gray is empirical data that supersedes wobble rules [94]; blue and magenta are proposed new tRNAs that can be assigned to nsAAs. Labels correspond to the wobble nucleotide at tRNA position 34 (cmo^5U = uridine 5-oxyacetic acid, mnm^5U = 5-methylaminomethyluridine, $cmnm^5U$ = 5-carboxymethylaminomethyluridine, $cmnm^5Um$ = 5-carboxymethylamino-methyl-2′-*O*-methyluridine, mnm^5s^2U = 5-methylaminomethyl-2-thiouridine, $cmnm^5s^2U$ = 5-carboxymethylamino-methyl-2-thiouridine, I = inosine, k^2C = lysidine, Q = queuosine, GluQ = glutamylqueuosine) [173]. Green letters indicate natural tRNA identity determinants that may be difficult to change without disrupting aminoacylation. Red letters indicate natural anticodon modifications that increase anticodon promiscuity. Blue and magenta letters represent proposed changes in the tRNA wobble position that would alter codon recognition. Amino acid

assignments are indicated in the yellow sidebars. M refers to both Met and fMet (translation initiation). Codons available for new amino acids are indicated by blue and magenta boxes with white numbers. Selenocysteine is not shown. **(A)** The *E. coli* genetic code is presented based on Björk *et al.* [94] and tRNA identity determinants are from Giegé *et al.* [110]. All 64 codons are used to encode 20 amino acids. **(B)** A minimal genetic code utilizing all 64 codons would require initiation at AUG, one release factor (RF2 E167K mutants can terminate all 3 stop codons [98]), and one tRNA for each of the 20 amino acids. Unmodified uracils in the wobble positions would allow tRNAs to recognize all codons in a family group, allowing redundant tRNAs to be deleted. Gray shaded boxes represent additional anticodons that could be potentially deleted ($tRNA^{Arg}_{UCG}$ would encode the same amino acid as $tRNA^{Arg}_{UCU}$ and it may be possible to remove Ile [103] and Trp [104] from the genetic code). Conveniently, the wobble nucleotide is rarely a tRNA identity determinant [110]. The two relevant exceptions, $tRNA^{Phe}_{GAA}$ (G34) and $tRNA^{Glu}_{UUG}$ ($cmnm^5s^2U34$), are weak identity determinants [110], so the proposed changes may be tolerated by their respective aaRSs. **(C)** The genetic code can be expanded to provide 7 unambiguous and 3 ambiguous anticodons by simply deleting tRNAs and introducing orthogonal aaRS/tRNA pairs encoding new amino acids. This analysis assumes that the original aaRS/tRNA identity determinants/antideterminants can be overcome by a metagenomic search for an orthogonal aaRS/tRNA pair and subsequent directed evolution to optimize their orthogonality. Red shaded boxes represent the three codons that would gain ambiguous translation function upon introduction of an orthogonal aaRS/tRNA pair. GUN, GCN, and CCN were not included, since the cmo^5U-containing anticodon has been empirically shown to recognize all four codons in their respective families [94]. The UAG codon has already been reassigned [13]. **(D)** Replacing the cmo^5U and inosine wobble nucleotides with mnm^5U nucleotides could liberate 13 unambiguous anticodons for reassignment (33 total amino acids; changes indicated in blue). Doing so would require the inactivation cmoB [105] and the engineering of mnmE and mnmG to recognize additional tRNAs [106,107]. Taken a step further, the maximal genetic code would have unique amino acid assignments for all NNA and NNG codons (NNA: engineer *tilS* [108] to lysidinylate additional tRNAs so that they only base pair with A; NNG: change the tRNA wobble bases to cytosine so that they only base pair with G). Anticodon modifications capable of splitting NNY codons into unambiguous NNU and NNC codons have not been reported, but such modifications have not been ruled out. Additionally, it may be possible to engineer a release factor to terminate translation only at UAA codons, thereby liberating both UAG and UGA codons. The proposed changes would liberate 27 unambiguous anticodons (47 total amino acids; changes indicated in magenta). This strategy may require directed evolution to overcome the tRNA identity determinants for Glu, Gln, and Lys [110]. Another potential complicating factor is that G + C anticodon content may affect cognate and near-cognate decoding efficiencies, just as the G + C rich anticodons for Val, Ala, and Pro break the wobble rules [94].

Figure 5. Design flaws at each level of genome complexity. Genome design flaws can impair fitness or alter the desired functions of an engineered organism. Genome engineering can introduce such design flaws by several mechanisms, including intentional genome changes [39,166], spontaneous mutations [13,148], transposition [165], and genome rearrangements [13]. Point mutations produced during genome engineering can introduce frame shifts, cause amino acid substitutions, de-optimize codon usage, or disrupt the function of overlapping non-coding sequences. Genetic parts such as genes and expression signals can impact crucial cellular functions or the desired function of the engineered organism (*e.g.*, nsAA incorporation, virus

resistance, biocontainment). Refactoring genetic pathways provides an opportunity to increase modularity [163], but cryptic regulation mechanisms and polar effects make it difficult to design a *de novo* pathway architecture with optimal activity [164,180]. Finally, while genome-scale design rules will continue to be discovered, we already know that it is important to balance the size of replichores in circular chromosomes [166], to co-orient transcription of essential operons with translation [168], to preserve sequences involved in DNA structure [169] and repair [170], and to consider how chromosome size impacts its structural integrity [167].

Acknowledgements:

We thank J. Ling, F. Isaacs, D. Mandell, A. Chatterjee, M. Jewett, A. Forster, M. Sismour, M. Napolitano, and K. Lajoie for helpful comments. Funding was from the U.S. Department of Energy (DE-FG02-02ER63445 to G.M.C. and DE-FG02-98ER20311 to D.S.), NSF (SA5283-11210 to G.M.C.), NIH (GM22854 to D.S.), and Defense Advanced Research Projects Agency (N66001-12-C-4040 to G.M.C., N66001-12-C-4020 to G.M.C. and D.S., N66001-12-C-4211 to G.M.C. and D.S.).

Conflict of interest:

GMC has potentially relevant patent applications and connections to companies: arep.med.harvard.edu/gmc/tech.html

References:

[1] Ambrogelly A, Palioura S, Söll D. Natural expansion of the genetic code. Nat Chem Biol 2007;3:29–35. doi:10.1038/nchembio847.

[2] Andersson G, Kurland C. An extreme codon preference strategy: codon reassignment. Mol Biol Evol 1991;8:530–44.

[3] Osawa S, Jukes TH. Evolution of the genetic code as affected by anticodon content. Trends Genet 1988;4:191–8. doi:10.1016/0168-9525(88)90075-3.

[4] Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, et al. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. Proc Natl Acad Sci U S A 2013;110:5540–5. doi:10.1073/pnas.1303090110.

[5] Ivanova NN, Schwientek P, Tripp HJ, Rinke C, Pati A, Huntemann M, et al. Stop codon reassignments in the wild. Science 2014;344:909–13. doi:10.1126/science.1250691.

[6] Edelman GM, Gally JA. Degeneracy and complexity in biological systems. Proc Natl Acad Sci U S A 2001;98:13763–8. doi:10.1073/pnas.231499798.

[7] Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, et al. The minimal gene complement of Mycoplasma genitalium. Science 1995;270:397–403.

[8] Oba T, Andachi Y, Muto A, Osawa S. CGG: an unassigned or nonsense codon in Mycoplasma capricolum. Proc Natl Acad Sci U S A 1991;88:921–5.

[9] Inagaki Y, Bessho Y, Osawa S. Lack of peptide-release activity responding to codon UGA in Mycoplasma capricolum. Nucleic Acids Res 1993;21:1335–8.

[10] Kano A, Ohama T, Abe R, Osawa S. Unassigned or nonsense codons in Micrococcus luteus. J Mol Biol 1993;230:51–6. doi:10.1006/jmbi.1993.1125.

[11] Itzkovitz S, Alon U. The genetic code is nearly optimal for allowing additional information within protein-coding sequences. Genome Res 2007;17:405–12. doi:10.1101/gr.5987307.

[12] Eggertsson G, Söll D. Transfer ribonucleic acid-mediated suppression of termination codons in Escherichia coli. Microbiol Rev 1988;52:354–74.

[13] Lajoie MJ, Rovner AJ, Goodman DB, Aerni H-R, Haimovich AD, Kuznetsov G, et al. Genomically recoded organisms expand biological functions. Science 2013;342:357–60. doi:10.1126/science.1241459.

[14] Anderson JC, Wu N, Santoro SW, Lakshman V, King DS, Schultz PG. An expanded genetic code with a functional quadruplet codon. Proc Natl Acad Sci U S A 2004;101:7566–71. doi:10.1073/pnas.0401517101.

[15] Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature 2010;464:441–4. doi:10.1038/nature08817.

[16] Chatterjee A, Lajoie MJ, Xiao H, Church GM, Schultz PG. A bacterial strain with a unique quadruplet codon specifying non-native amino acids. Chembiochem 2014;15:1782–6. doi:10.1002/cbic.201402104.

[17] Bain JD, Switzer C, Chamberlin AR, Benner SA. Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. Nature 1992;356:537–9. doi:10.1038/356537a0.

[18] Goeddel D V, Kleid DG, Bolivar F, Heyneker HL, Yansura DG, Crea R, et al. Expression in Escherichia coli of chemically synthesized genes for human insulin. Proc Natl Acad Sci U S A 1979;76:106–10.

[19] Nakamura CE, Whited GM. Metabolic engineering for the microbial production of 1,3-propanediol. Curr Opin Biotechnol 2003;14:454–9.

[20] Padgette SR, Kolacz KH, Delannay X, Re DB, LaVallee BJ, Tinius CN, et al. Development, Identification, and Characterization of a Glyphosate-Tolerant Soybean Line. Crop Sci 1995;35:1451. doi:10.2135/cropsci1995.0011183X003500050032x.

[21] Tian J, Gong H, Sheng N, Zhou X, Gulari E, Gao X, et al. Accurate multiplex gene synthesis from programmable DNA microchips. Nature 2004;432:1050–4. doi:10.1038/nature03151.

[22] Kosuri S, Eroshenko N, Leproust EM, Super M, Way J, Li JB, et al. Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. Nat Biotechnol 2010;28:1295–9. doi:10.1038/nbt.1716.

[23] Quan J, Saaem I, Tang N, Ma S, Negre N, Gong H, et al. Parallel on-chip gene synthesis and application to optimization of protein expression. Nat Biotechnol 2011;29:449–52. doi:10.1038/nbt.1847.

[24] Neu HC. The Crisis in Antibiotic Resistance. Science (80-) 1992;257:1064–73. doi:10.1126/science.257.5073.1064.

[25] LU B-R, SNOW AA. Gene Flow from Genetically Modified Rice and Its Environmental Consequences. Bioscience 2005;55:669. doi:10.1641/0006-3568(2005)055[0669:GFFGMR]2.0.CO;2.

[26] Harris A, Beasley D. Bayer Will Pay $750 Million to Settle Gene-Modified Rice Suits - Bloomberg Business. Bloomberg 2011. http://www.bloomberg.com/news/articles/2011-07-01/bayer-to-pay-750-million-to-end-lawsuits-over-genetically-modified-rice (accessed May 18, 2015).

[27] Xia H-B, Wang W, Xia H, Zhao W, Lu B-R. Conspecific crop-weed introgression influences evolution of weedy rice (Oryza sativa f. spontanea) across a geographical range. PLoS One 2011;6:e16189. doi:10.1371/journal.pone.0016189.

[28] Dumas A, Lercher L, Spicer CD, Davis BG. Designing logical codon reassignment – Expanding the chemistry in biology. Chem Sci 2015;6:50–69. doi:10.1039/C4SC01534G.

[29] Jackson JC, Duffy SP, Hess KR, Mehl RA. Improving nature's enzyme active site with genetically encoded unnatural amino acids. J Am Chem Soc 2006;128:11124–7. doi:10.1021/ja061099y.

[30] Ugwumba IN, Ozawa K, Xu Z-Q, Ely F, Foo J-L, Herlt AJ, et al. Improving a natural enzyme activity through incorporation of unnatural amino acids. J Am Chem Soc 2011;133:326–33. doi:10.1021/ja106416g.

[31] Cho H, Daniel T, Buechler YJ, Litzinger DC, Maio Z, Putnam A-MH, et al. Optimized clinical performance of growth hormone with an expanded genetic code. Proc Natl Acad Sci U S A 2011;108:9060–5. doi:10.1073/pnas.1100387108.

[32] Axup JY, Bajjuri KM, Ritland M, Hutchins BM, Kim CH, Kazane SA, et al. Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. Proc Natl Acad Sci U S A 2012;109:16101–6. doi:10.1073/pnas.1211023109.

[33] Liu CC, Schultz PG. Adding new chemistries to the genetic code. Annu Rev Biochem 2010;79:413–44. doi:10.1146/annurev.biochem.052308.105824.

[34] Mandell DJ, Lajoie MJ, Mee MT, Takeuchi R, Kuznetsov G, Norville JE, et al. Biocontainment of genetically modified organisms by synthetic protein design. Nature 2015;518:55–60. doi:10.1038/nature14121.

[35] Rovner AJ, Haimovich AD, Katz SR, Li Z, Grome MW, Gassaway BM, et al. Recoded organisms engineered to depend on synthetic amino acids. Nature 2015;518:89–93. doi:10.1038/nature14095.

[36] Sturino JM, Klaenhammer TR. Engineered bacteriophage-defence systems in bioprocessing. Nat Rev Microbiol 2006;4:395–404. doi:10.1038/nrmicro1393.

[37] Bethencourt V. Virus stalls Genzyme plant. Nat Biotechnol 2009;27:681–681. doi:10.1038/nbt0809-681a.

[38] Hammerling MJ, Ellefson JW, Boutz DR, Marcotte EM, Ellington AD, Barrick JE. Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness. Nat Chem Biol 2014;10:178–80. doi:10.1038/nchembio.1450.

[39] Lajoie MJ, Kosuri S, Mosberg JA, Gregg CJ, Zhang D, Church GM. Probing the limits of genetic recoding in essential genes. Science 2013;342:361–3. doi:10.1126/science.1241460.

[40] Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, et al. Cell-free translation reconstituted with purified components. Nat Biotechnol 2001;19:751–5. doi:10.1038/90802.

[41] Forster AC, Tan Z, Nalam MNL, Lin H, Qu H, Cornish VW, et al. Programming peptidomimetic syntheses by translating genetic codes designed de novo. Proc Natl Acad Sci U S A 2003;100:6353–7. doi:10.1073/pnas.1132122100.

[42] Passioura T, Suga H. Reprogramming the genetic code in vitro. Trends Biochem Sci 2014;39:400–8. doi:10.1016/j.tibs.2014.07.005.

[43] Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG. A general method for site-specific incorporation of unnatural amino acids into proteins. Science 1989;244:182–8.

[44] Murakami H, Ohta A, Ashigai H, Suga H. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. Nat Methods 2006;3:357–9. doi:10.1038/nmeth877.

[45] Hartman MCT, Josephson K, Lin C-W, Szostak JW. An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. PLoS One 2007;2:e972. doi:10.1371/journal.pone.0000972.

[46] Ohta A, Yamagishi Y, Suga H. Synthesis of biopolymers using genetic code reprogramming. Curr Opin Chem Biol 2008;12:159–67. doi:10.1016/j.cbpa.2007.12.009.

[47] Josephson K, Hartman MCT, Szostak JW. Ribosomal synthesis of unnatural peptides. J Am Chem Soc 2005;127:11727–35. doi:10.1021/ja0515809.

[48] Kang TJ, Suga H. Ribosomal synthesis of nonstandard peptides. Biochem Cell Biol 2008;86:92–9. doi:10.1139/O08-009.

[49] Fahnestock S, Rich A. Ribosome-catalyzed polyester formation. Science 1971;173:340–3.

[50] Jewett MC, Fritz BR, Timmerman LE, Church GM. In vitro integration of ribosomal RNA synthesis, ribosome assembly, and translation. Mol Syst Biol 2013;9:678. doi:10.1038/msb.2013.31.

[51] Wang K, Neumann H, Peak-Chew SY, Chin JW. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. Nat Biotechnol 2007;25:770–7. doi:10.1038/nbt1314.

[52] Terasaka N, Hayashi G, Katoh T, Suga H. An orthogonal ribosome-tRNA pair via engineering of the peptidyl transferase center. Nat Chem Biol 2014;10:555–7. doi:10.1038/nchembio.1549.

[53] Orelle C, Carlson ED, Szal T, Florin T, Jewett MC, Mankin AS. Protein synthesis by ribosomes with tethered subunits. Nature 2015;524:119–24. doi:10.1038/nature14862.

[54] Van Hest JCM, Kiick KL, Tirrell DA. Efficient Incorporation of Unsaturated Methionine Analogues into Proteins in Vivo. J Am Chem Soc 2000;122:1282–8. doi:10.1021/ja992749j.

[55] Wong JT. Membership mutation of the genetic code: loss of fitness by tryptophan. Proc Natl Acad Sci U S A 1983;80:6303–6.

[56] Bacher JM, Ellington AD. Selection and characterization of Escherichia coli variants capable of growth on an otherwise toxic tryptophan analogue. J Bacteriol 2001;183:5414–25.

[57] Bacher JM, Bull JJ, Ellington AD. Evolution of phage with chemically ambiguous proteomes. BMC Evol Biol 2003;3:24. doi:10.1186/1471-2148-3-24.

[58] Hoesl MG, Oehm S, Durkin P, Darmon E, Peil L, Aerni H-R, et al. Chemical Evolution of a Bacterial Proteome. Angew Chemie Int Ed 2015;54:n/a – n/a. doi:10.1002/anie.201502868.

[59] Hendrickson WA, Horton JR, LeMaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J 1990;9:1665–72.

[60] Furter R. Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in Escherichia coli. Protein Sci 1998;7:419–26. doi:10.1002/pro.5560070223.

[61] Sharma N, Furter R, Kast P, Tirrell DA. Efficient introduction of aryl bromide functionality into proteins in vivo. FEBS Lett 2000;467:37–40. doi:10.1016/S0014-5793(00)01120-0.

[62] Wang L, Brock A, Herberich B, Schultz PG. Expanding the genetic code of Escherichia coli. Science 2001;292:498–500. doi:10.1126/science.1060077.

[63] Mukai T, Hayashi A, Iraha F, Sato A, Ohtake K, Yokoyama S, et al. Codon reassignment in the Escherichia coli genetic code. Nucleic Acids Res 2010;38:8188–95. doi:10.1093/nar/gkq707.

[64] Mukai T, Hoshi H, Ohtake K, Takahashi M, Yamaguchi A, Hayashi A, et al. Highly reproductive Escherichia coli cells with no specific assignment to the UAG codon. Sci Rep 2015;5:9699. doi:10.1038/srep09699.

[65] Johnson DBF, Xu J, Shen Z, Takimoto JK, Schultz MD, Schmitz RJ, et al. RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. Nat Chem Biol 2011;7:779–86. doi:10.1038/nchembio.657.

[66] Piccirilli JA, Krauch T, Moroney SE, Benner SA. Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. Nature 1990;343:33–7. doi:10.1038/343033a0.

[67] Yang Z, Chen F, Alvarado JB, Benner SA. Amplification, mutation, and sequencing of a six-letter synthetic genetic system. J Am Chem Soc 2011;133:15105–12. doi:10.1021/ja204910n.

[68] Yamashige R, Kimoto M, Takezawa Y, Sato A, Mitsui T, Yokoyama S, et al. Highly specific unnatural base pair systems as a third base pair for PCR amplification. Nucleic Acids Res 2012;40:2793–806. doi:10.1093/nar/gkr1068.

[69] Malyshev DA, Dhami K, Quach HT, Lavergne T, Ordoukhanian P, Torkamani A, et al. Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. Proc Natl Acad Sci U S A 2012;109:12005–10. doi:10.1073/pnas.1205176109.

[70] Dhami K, Malyshev DA, Ordoukhanian P, Kubelka T, Hocek M, Romesberg FE. Systematic exploration of a class of hydrophobic unnatural base pairs yields multiple new candidates for the expansion of the genetic alphabet. Nucleic Acids Res 2014;42 :10235–44. doi:10.1093/nar/gku715.

[71] Roth JR. Frameshift suppression. Cell 1981;24:601–2. doi:10.1016/0092-8674(81)90086-6.

[72] Wang K, Schmied WH, Chin JW. Reprogramming the genetic code: from triplet to quadruplet codes. Angew Chem Int Ed Engl 2012;51:2288–97. doi:10.1002/anie.201105016.

[73] Hohsaka T, Ashizuka Y, Murakami H, Sisido M. Five-base codons for incorporation of nonnatural amino acids into proteins. Nucleic Acids Res 2001;29:3646–51.

[74] Ohtsuki T, Kimoto M, Ishikawa M, Mitsui T, Hirao I, Yokoyama S. Unnatural base pairs for specific transcription. Proc Natl Acad Sci U S A 2001;98:4922–5. doi:10.1073/pnas.091532698.

[75] Morohashi N, Kimoto M, Sato A, Kawai R, Hirao I. Site-specific incorporation of functional components into RNA by an unnatural base pair transcription system. Molecules 2012;17:2855–76. doi:10.3390/molecules17032855.

[76] Kim H-J, Leal NA, Hoshika S, Benner SA. Ribonucleosides for an artificially expanded genetic information system. J Org Chem 2014;79:3194–9. doi:10.1021/jo402665d.

[77] Leal NA, Kim H-J, Hoshika S, Kim M-J, Carrigan MA, Benner SA. Transcription, reverse transcription, and analysis of RNA containing artificial genetic components. ACS Synth Biol 2015;4:407–13. doi:10.1021/sb500268n.

[78] Hirao I, Harada Y, Kimoto M, Mitsui T, Fujiwara T, Yokoyama S. A two-unnatural-base-pair system toward the expansion of the genetic code. J Am Chem Soc 2004;126:13298–305. doi:10.1021/ja047201d.

[79] Malyshev DA, Dhami K, Lavergne T, Chen T, Dai N, Foster JM, et al. A semi-synthetic organism with an expanded genetic alphabet. Nature 2014;509:385–8. doi:10.1038/nature13314.

[80] Hirao I, Kimoto M. Unnatural base pair systems toward the expansion of the genetic alphabet in the central dogma. Proc Jpn Acad Ser B Phys Biol Sci 2012;88:345–67.

[81] Young TS, Ahmad I, Yin JA, Schultz PG. An enhanced system for unnatural amino acid mutagenesis in E. coli. J Mol Biol 2010;395:361–74. doi:10.1016/j.jmb.2009.10.030.

[82] Chatterjee A, Sun SB, Furman JL, Xiao H, Schultz PG. A versatile platform for single- and multiple-unnatural amino acid mutagenesis in Escherichia coli. Biochemistry 2013;52:1828–37. doi:10.1021/bi4000244.

[83] Wang K, Sachdeva A, Cox DJ, Wilf NW, Lang K, Wallace S, et al. Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. Nat Chem 2014;6:393–403. doi:10.1038/nchem.1919.

[84] Moore B, Persson BC, Nelson CC, Gesteland RF, Atkins JF. Quadruplet codons: implications for code expansion and the specification of translation step size. J Mol Biol 2000;298:195–209. doi:10.1006/jmbi.2000.3658.

[85] Magliery TJ, Anderson JC, Schultz PG. Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in Escherichia coli. J Mol Biol 2001;307:755–69. doi:10.1006/jmbi.2001.4518.

[86] Ohuchi M, Murakami H, Suga H. The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. Curr Opin Chem Biol 2007;11:537–42. doi:10.1016/j.cbpa.2007.08.011.

[87] Park H-S, Hohn MJ, Umehara T, Guo L-T, Osborne EM, Benner J, et al. Expanding the genetic code of Escherichia coli with phosphoserine. Science 2011;333:1151–4. doi:10.1126/science.1207203.

[88] Moura GR, Paredes JA, Santos MAS. Development of the genetic code: insights from a fungal codon reassignment. FEBS Lett 2010;584:334–41. doi:10.1016/j.febslet.2009.11.066.

[89] Knight RD, Freeland SJ, Landweber LF. Rewiring the keyboard: evolvability of the genetic code. Nat Rev Genet 2001;2:49–58. doi:10.1038/35047500.

[90] Santos MAS, Moura G, Massey SE, Tuite MF. Driving change: the evolution of alternative genetic codes. Trends Genet 2004;20:95–102. doi:10.1016/j.tig.2003.12.009.

[91] Watanabe K, Yokobori S-I. tRNA Modification and Genetic Code Variations in Animal Mitochondria. J Nucleic Acids 2011;2011:623095. doi:10.4061/2011/623095.

[92] Sengupta S, Higgs PG. Pathways of Genetic Code Evolution in Ancient and Modern Organisms. J Mol Evol 2015;80:229–43. doi:10.1007/s00239-015-9686-8.

[93] Agris PF, Vendeix FAP, Graham WD. tRNA's wobble decoding of the genome: 40 years of modification. J Mol Biol 2007;366:1–13. doi:10.1016/j.jmb.2006.11.046.

[94] Björk GR, Hagervall TG. Transfer RNA Modification: Presence, Synthesis, and Function. EcoSal Plus 2014;1. doi:10.1128/ecosalplus.ESP-0007-2013.

[95] Freeland SJ, Knight RD, Landweber LF, Hurst LD. Early Fixation of an Optimal Genetic Code. Mol Biol Evol 2000;17:511–8. doi:10.1093/oxfordjournals.molbev.a026331.

[96] Sherman F, Stewart JW, Tsunasawa S. Methionine or not methionine at the beginning of a protein. Bioessays 1985;3:27–31. doi:10.1002/bies.950030108.

[97] Sylvers LA, Rogers KC, Shimizu M, Ohtsuka E, Söll D. A 2-thiouridine derivative in tRNAGlu is a positive determinant for aminoacylation by Escherichia coli glutamyl-tRNA synthetase. Biochemistry 1993;32:3836–41.

[98] Ito K, Uno M, Nakamura Y. Single amino acid substitution in prokaryote polypeptide release factor 2 permits it to terminate translation at all three stop codons. Proc Natl Acad Sci U S A 1998;95:8165–9.

[99] Akanuma S, Kigawa T, Yokoyama S. Combinatorial mutagenesis to restrict amino acid usage in an enzyme to a reduced set. Proc Natl Acad Sci U S A 2002;99:13549–53. doi:10.1073/pnas.222243999.

[100] Walter KU, Vamvaca K, Hilvert D. An active enzyme constructed from a 9-amino acid alphabet. J Biol Chem 2005;280:37742–6. doi:10.1074/jbc.M507210200.

[101] Kawahara-Kobayashi A, Masuda A, Araiso Y, Sakai Y, Kohda A, Uchiyama M, et al. Simplification of the genetic code: restricted diversity of genetically encoded amino acids. Nucleic Acids Res 2012;40:10576–84. doi:10.1093/nar/gks786.

[102] Lu M-F, Ji H-F, Li T-X, Kang S-K, Zhang Y-J, Zheng J-F, et al. Reconstructing a flavodoxin oxidoreductase with early amino acids. Int J Mol Sci 2013;14:12843–52. doi:10.3390/ijms140612843.

[103] Pezo V, Metzgar D, Hendrickson TL, Waas WF, Hazebrouck S, Döring V, et al. Artificially ambiguous genetic code confers growth yield advantage. Proc Natl Acad Sci U S A 2004;101:8593–7. doi:10.1073/pnas.0402893101.

[104] Pezo V, Louis D, Guérineau V, Le Caer J-P, Gaillon L, Mutzel R, et al. A metabolic prototype for eliminating tryptophan from the genetic code. Sci Rep 2013;3:1359. doi:10.1038/srep01359.

[105] Nasvall SJ, Chen P, Bjork GR. The modified wobble nucleoside uridine-5-oxyacetic acid in tRNAPro(cmo5UGG) promotes reading of all four proline codons in vivo. RNA 2004;10:1662–73. doi:10.1261/rna.7106404.

[106] Shi R, Villarroya M, Ruiz-Partida R, Li Y, Proteau A, Prado S, et al. Structure-function analysis of Escherichia coli MnmG (GidA), a highly conserved tRNA-modifying enzyme. J Bacteriol 2009;191:7614–9. doi:10.1128/JB.00650-09.

[107] Pearson D, Carell T. Assay of both activities of the bifunctional tRNA-modifying enzyme MnmC reveals a kinetic basis for selective full modification of cmnm5s2U to mnm5s2U. Nucleic Acids Res 2011;39:4818–26. doi:10.1093/nar/gkr071.

[108] Nakanishi K, Bonnefond L, Kimura S, Suzuki T, Ishitani R, Nureki O. Structural basis for translational fidelity ensured by transfer RNA lysidine synthetase. Nature 2009;461:1144–8. doi:10.1038/nature08474.

[109] Salowe SP, Wiltsie J, Hawkins JC, Sonatore LM. The catalytic flexibility of tRNAIle-lysidine synthetase can generate alternative tRNA substrates for isoleucyl-tRNA synthetase. J Biol Chem 2009;284:9656–62. doi:10.1074/jbc.M809013200.

[110] Giegé R, Sissler M, Florentz C. Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res 1998;26:5017–35.

[111] Neumann H, Slusarczyk AL, Chin JW. De novo generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs. J Am Chem Soc 2010;132:2142–4. doi:10.1021/ja9068722.

[112] Woese CR, Olsen GJ, Ibba M, Soll D. Aminoacyl-tRNA Synthetases, the Genetic Code, and the Evolutionary Process. Microbiol Mol Biol Rev 2000;64:202–36. doi:10.1128/MMBR.64.1.202-236.2000.

[113] Ibba M, Soll D. Aminoacyl-tRNA synthesis. Annu Rev Biochem 2000;69:617–50. doi:10.1146/annurev.biochem.69.1.617.

[114] Richmond MH. The effect of amino acid analogues on growth and protein synthesis in microorganisms. Bacteriol Rev 1962;26:398–420.

[115] Fan C, Ho JML, Chirathivat N, Söll D, Wang Y-S. Exploring the substrate range of wild-type aminoacyl-tRNA synthetases. Chembiochem 2014;15:1805–9. doi:10.1002/cbic.201402083.

[116] O'Donoghue P, Ling J, Wang Y-S, Söll D. Upgrading protein synthesis for synthetic biology. Nat Chem Biol 2013;9:594–8. doi:10.1038/nchembio.1339.

[117] Stokes AL, Miyake-Stoner SJ, Peeler JC, Nguyen DP, Hammer RP, Mehl RA. Enhancing the utility of unnatural amino acid synthetases by manipulating broad substrate specificity. Mol Biosyst 2009;5:1032–8. doi:10.1039/b904032c.

[118] Young DD, Young TS, Jahnz M, Ahmad I, Spraggon G, Schultz PG. An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. Biochemistry 2011;50:1894–900. doi:10.1021/bi101929e.

[119] Guo L-T, Wang Y-S, Nakamura A, Eiler D, Kavran JM, Wong M, et al. Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. Proc Natl Acad Sci U S A 2014;111:16724–9. doi:10.1073/pnas.1419737111.

[120] Neumann H. Rewiring translation - Genetic code expansion and its applications. FEBS Lett 2012;586:2057–64. doi:10.1016/j.febslet.2012.02.002.

[121] Hadd A, Perona JJ. Recoding aminoacyl-tRNA synthetases for synthetic biology by rational protein-RNA engineering. ACS Chem Biol 2014;9:2761–6. doi:10.1021/cb5006596.

[122] 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. doi:10.1093/nar/gku1087.

[123] Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, et al. Computational design of ligand-binding proteins with high affinity and selectivity. Nature 2013;501:212–6. doi:10.1038/nature12443.

[124] Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, et al. De Novo Computational Design of Retro-Aldol Enzymes. Science (80-) 2008;319:1387–91. doi:10.1126/science.1152692.

[125] Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, et al. Kemp elimination catalysts by computational enzyme design. Nature 2008;453:190–5. doi:10.1038/nature06879.

[126] Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St Clair JL, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science 2010;329:309–13. doi:10.1126/science.1190239.

[127] Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol 2011;487:545–74. doi:10.1016/B978-0-12-381270-4.00019-6.

[128] Corigliano EM, Perona JJ. Architectural underpinnings of the genetic code for glutamine. Biochemistry 2009;48:676–87. doi:10.1021/bi801552y.

[129] Perona JJ, Hadd A. Structural diversity and protein engineering of the aminoacyl-tRNA synthetases. Biochemistry 2012;51:8705–29. doi:10.1021/bi301180x.

[130] Giegé R, Eriani G. Transfer RNA Recognition and Aminoacylation by Synthetases. eLS, Chichester: John Wiley & Sons, Ltd; 2014. doi:10.1002/9780470015902.a0000531.pub3.

[131] Krishnakumar R, Prat L, Aerni H-R, Ling J, Merryman C, Glass JI, et al. Transfer RNA Misidentification Scrambles Sense Codon Recoding. ChemBioChem 2013;14:1967–72. doi:10.1002/cbic.201300444.

[132] Nguyen DP, Lusic H, Neumann H, Kapadnis PB, Deiters A, Chin JW. Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA(CUA) pair and click chemistry. J Am Chem Soc 2009;131:8720–1. doi:10.1021/ja900553w.

[133] Wan W, Tharp JM, Liu WR. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. Biochim Biophys Acta 2014;1844:1059–70. doi:10.1016/j.bbapap.2014.03.002.

[134] Santoro SW, Anderson JC, Lakshman V, Schultz PG. An archaebacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in Escherichia coli. Nucleic Acids Res 2003;31:6700–9.

[135] Anderson JC, Schultz PG. Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. Biochemistry 2003;42:9598–608. doi:10.1021/bi034550w.

[136] Neumann H, Peak-Chew SY, Chin JW. Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. Nat Chem Biol 2008;4:232–4. doi:10.1038/nchembio.73.

[137] Hughes RA, Ellington AD. Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA. Nucleic Acids Res 2010;38:6813–30. doi:10.1093/nar/gkq521.

[138] Chatterjee A, Xiao H, Yang P-Y, Soundararajan G, Schultz PG. A tryptophanyl-tRNA synthetase/tRNA pair for unnatural amino acid mutagenesis in E. coli. Angew Chem Int Ed Engl 2013;52:5106–9. doi:10.1002/anie.201301094.

[139] Chatterjee A, Xiao H, Schultz PG. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli. Proc Natl Acad Sci U S A 2012;109:14841–6. doi:10.1073/pnas.1212454109.

[140] Ko J, Llopis PM, Heinritz J, Jacobs-Wagner C, Söll D. Suppression of amber codons in Caulobacter crescentus by the orthogonal Escherichia coli histidyl-tRNA synthetase/tRNAHis pair. PLoS One 2013;8:e83630. doi:10.1371/journal.pone.0083630.

[141] Ardell DH. Computational analysis of tRNA identity. FEBS Lett 2010;584:325–33. doi:10.1016/j.febslet.2009.11.084.

[142] Bröcker MJ, Ho JML, Church GM, Söll D, O'Donoghue P. Recoding the genetic code with selenocysteine. Angew Chem Int Ed Engl 2014;53:319–23.

[143] Miller C, Bröcker MJ, Prat L, Ip K, Chirathivat N, Feiock A, et al. A synthetic tRNA for EF-Tu mediated selenocysteine incorporation in vivo and in vitro. FEBS Lett 2015;589:2194–9. doi:10.1016/j.febslet.2015.06.039.

[144] Li L, Carter CW. Full implementation of the genetic code by tryptophanyl-tRNA synthetase requires intermodular coupling. J Biol Chem 2013;288:34736–45. doi:10.1074/jbc.M113.510958.

[145] Woese CR. On the evolution of the genetic code. Proc Natl Acad Sci U S A 1965;54:1546–52.

[146] Epstein CJ. Role of the Amino-Acid "Code" and of Selection for Conformation in the Evolution of Proteins. Nature 1966;210:25–8. doi:10.1038/210025a0.

[147] Buhrman H, van der Gulik PTS, Klau GW, Schaffner C, Speijer D, Stougie L. A realistic model under which the genetic code is optimal. J Mol Evol 2013;77:170–84. doi:10.1007/s00239-013-9571-2.

[148] Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R-Y, Algire MA, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. Science 2010;329:52–6. doi:10.1126/science.1190719.

[149] Gibson DG, Smith HO, Hutchison CA, Venter JC, Merryman C. Chemical synthesis of the mouse mitochondrial genome. Nat Methods 2010;7:901–3. doi:10.1038/nmeth.1515.

[150] Dymond JS, Richardson SM, Coombes CE, Babatz T, Muller H, Annaluru N, et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature 2011;477:471–6. doi:10.1038/nature10403.

[151] Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 2012;6:80–92. doi:10.4161/fly.19695.

[152] Li G-W, Oh E, Weissman JS. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature 2012;484:538–41. doi:10.1038/nature10965.

[153] Shinhara A, Matsui M, Hiraoka K, Nomura W, Hirano R, Nakahigashi K, et al. Deep sequencing reveals as-yet-undiscovered small RNAs in Escherichia coli. BMC Genomics 2011;12:428. doi:10.1186/1471-2164-12-428.

[154] Bulmer M. The selection-mutation-drift theory of synonymous codon usage. Genetics 1991;129:897–907.

[155] Goodman DB, Church GM, Kosuri S. Causes and effects of N-terminal codon bias in bacterial genes. Science 2013;342:475–9. doi:10.1126/science.1241934.

[156] Gingold H, Dahan O, Pilpel Y. Dynamic changes in translational efficiency are deduced from codon usage of the transcriptome. Nucleic Acids Res 2012;40:10053–63. doi:10.1093/nar/gks772.

[157] Capra JA, Singh M. Predicting functionally important residues from sequence conservation. Bioinformatics 2007;23:1875–82. doi:10.1093/bioinformatics/btm270.

[158] Brewster RC, Jones DL, Phillips R. Tuning promoter strength through RNA polymerase binding site design in Escherichia coli. PLoS Comput Biol 2012;8:e1002811. doi:10.1371/journal.pcbi.1002811.

[159] Kosuri S, Goodman DB, Cambray G, Mutalik VK, Gao Y, Arkin AP, et al. Composability of regulatory sequences controlling transcription and translation in Escherichia coli. Proc Natl Acad Sci U S A 2013;110:14024–9. doi:10.1073/pnas.1301301110.

[160] Salis HM. The ribosome binding site calculator. Methods Enzymol 2011;498:19–42. doi:10.1016/B978-0-12-385120-8.00002-4.

[161] Chen Y-J, Liu P, Nielsen AAK, Brophy JAN, Clancy K, Peterson T, et al. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. Nat Methods 2013;10:659–64. doi:10.1038/nmeth.2515.

[162] Chan LY, Kosuri S, Endy D. Refactoring bacteriophage T7. Mol Syst Biol 2005;1:2005.0018. doi:10.1038/msb4100025.

[163] Temme K, Zhao D, Voigt CA. Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. Proc Natl Acad Sci U S A 2012;109:7085–90. doi:10.1073/pnas.1120788109.

[164] Smanski MJ, Bhatia S, Zhao D, Park Y, B A Woodruff L, Giannoukos G, et al. Functional optimization of gene clusters by combinatorial design and assembly. Nat Biotechnol 2014;32:1241–9. doi:10.1038/nbt.3063.

[165] Pósfai G, Plunkett G, Fehér T, Frisch D, Keil GM, Umenhoffer K, et al. Emergent properties of reduced-genome Escherichia coli. Science 2006;312:1044–6. doi:10.1126/science.1126439.

[166] Itaya M, Tsuge K, Koizumi M, Fujita K. Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. Proc Natl Acad Sci U S A 2005;102:15971–6. doi:10.1073/pnas.0503868102.

[167] Val M-E, Skovgaard O, Ducos-Galand M, Bland MJ, Mazel D. Genome engineering in Vibrio cholerae: a feasible approach to address biological issues. PLoS Genet 2012;8:e1002472. doi:10.1371/journal.pgen.1002472.

[168] Srivatsan A, Tehranchi A, MacAlpine DM, Wang JD. Co-orientation of replication and transcription preserves genome integrity. PLoS Genet 2010;6:e1000810. doi:10.1371/journal.pgen.1000810.

[169] Reece RJ, Maxwell A. DNA gyrase: structure and function. Crit Rev Biochem Mol Biol 1991;26:335–75. doi:10.3109/10409239109114072.

[170] Friedman-Ohana R, Karunker I, Cohen A. Chi-dependent intramolecular recombination in Escherichia coli. Genetics 1998;148:545–57.

[171] Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, et al. Essential genes of a minimal bacterium. Proc Natl Acad Sci U S A 2006;103:425–30. doi:10.1073/pnas.0510013103.

[172] Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, et al. Global transposon mutagenesis and a minimal Mycoplasma genome. Science 1999;286:2165–9.

[173] Forster AC, Church GM. Towards synthesis of a minimal cell. Mol Syst Biol 2006;2:45. doi:10.1038/msb4100090.

[174] Jewett MC, Forster AC. Update on designing and building minimal cells. Curr Opin Biotechnol 2010;21:697–703. doi:10.1016/j.copbio.2010.06.008.

[175] Suzuki Y, Assad-Garcia N, Kostylev M, Noskov VN, Wise KS, Karas BJ, et al. Bacterial genome reduction using the progressive clustering of deletions via yeast sexual cycling. Genome Res 2015;25:435–44. doi:10.1101/gr.182477.114.

[176] Hartman JL, Garvik B, Hartwell L. Principles for the buffering of genetic variation. Science 2001;291:1001–4.

[177] Pál C, Papp B, Pósfai G. The dawn of evolutionary genome engineering. Nat Rev Genet 2014;15:504–12. doi:10.1038/nrg3746.

[178] Carr PA, Church GM. Genome engineering. Nat Biotechnol 2009;27:1151–62. doi:10.1038/nbt.1590.

[179] Karr JR, Sanghvi JC, Macklin DN, Gutschow M V, Jacobs JM, Bolival B, et al. A whole-cell computational model predicts phenotype from genotype. Cell 2012;150:389–401. doi:10.1016/j.cell.2012.05.044.

[180] Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, et al. Programming cells by multiplex genome engineering and accelerated evolution. Nature 2009;460:894–8. doi:10.1038/nature08187.

[181] Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, et al. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science 2011;333:348–53. doi:10.1126/science.1205822.

[182] Ellis HM, Yu D, DiTizio T, Court DL. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci U S A 2001;98:6742–6. doi:10.1073/pnas.121164898.

[183] Esvelt KM, Wang HH. Genome-scale engineering for systems and synthetic biology. Mol Syst Biol 2013;9:641. doi:10.1038/msb.2012.66.

[184] Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Muñiz-Rascado L, et al. EcoCyc: a comprehensive database of Escherichia coli biology. Nucleic Acids Res 2011;39:D583–90. doi:10.1093/nar/gkq1143.

[185] Naito Y, Hino K, Bono H, Ui-Tei K. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. Bioinformatics 2014. doi:10.1093/bioinformatics/btu743.

[186] Xia B, Bhatia S, Bubenheim B, Dadgar M, Densmore D, Anderson JC. Developer's and user's guide to Clotho v2.0 A software platform for the creation of synthetic biological systems. Methods Enzymol 2011;498:97–135. doi:10.1016/B978-0-12-385120-8.00005-X.

[187] Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 2011;27:1691–2. doi:10.1093/bioinformatics/btr174.

[188] Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods 2009;6:677–81. doi:10.1038/nmeth.1363.

[189] Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 2009;25:2865–71. doi:10.1093/bioinformatics/btp394.

[190] Deatherage DE, Traverse CC, Wolf LN, Barrick JE. Detecting rare structural variation in evolving microbial populations from new sequence junctions using breseq. Front Genet 2014;5:468. doi:10.3389/fgene.2014.00468.

[191] Quax TEF, Claassens NJ, Söll D, van der Oost J. Codon Bias as a Means to Fine-Tune Gene Expression. Mol Cell 2015;59:149–61. doi:10.1016/j.molcel.2015.05.035.

[192] Crick FH. Codon--anticodon pairing: the wobble hypothesis. J Mol Biol 1966;19:548–55.

[193] Yokoyama S, Nishimura S. Modified nucleosides and codon recognition. In: Soll D, RajBhandary UL, editors. tRNA Struct. biosynthesis, Funct., Washington, DC: American Society for Microbiology (ASM), Books Division; 1995, p. 207–23.

- Barriers to changing the genetic code
  - biochemical barriers
  - biological knowledge (ability to design functional genomes)
  - biotechnology (ability to create the genomes)
- Approach to engineer genomes with new biological functions

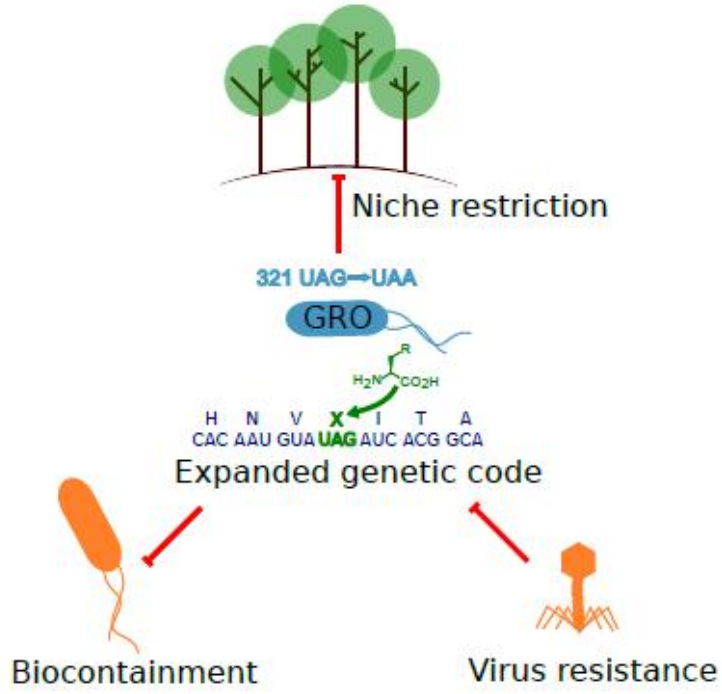


Fig. 1

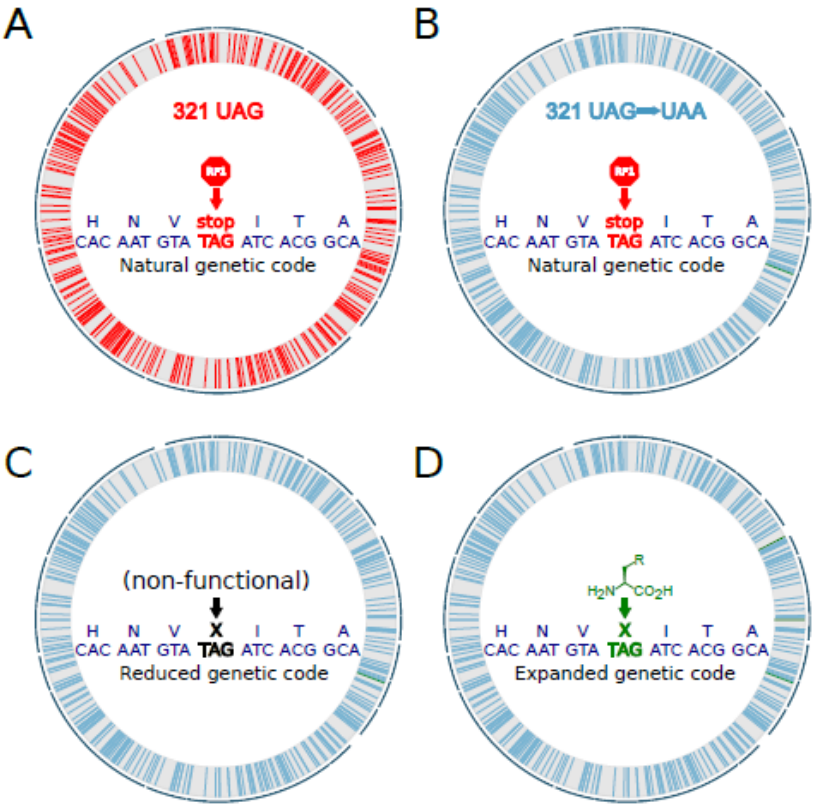


Fig. 2

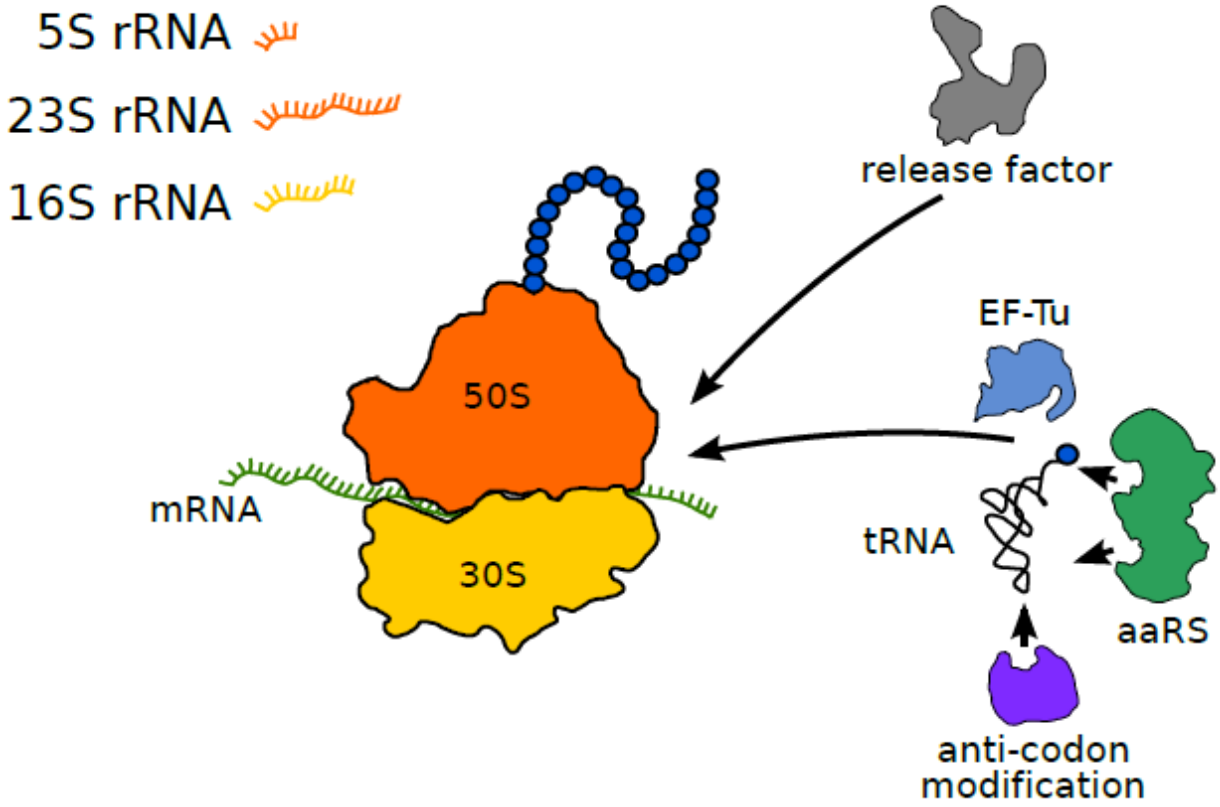


Fig. 3

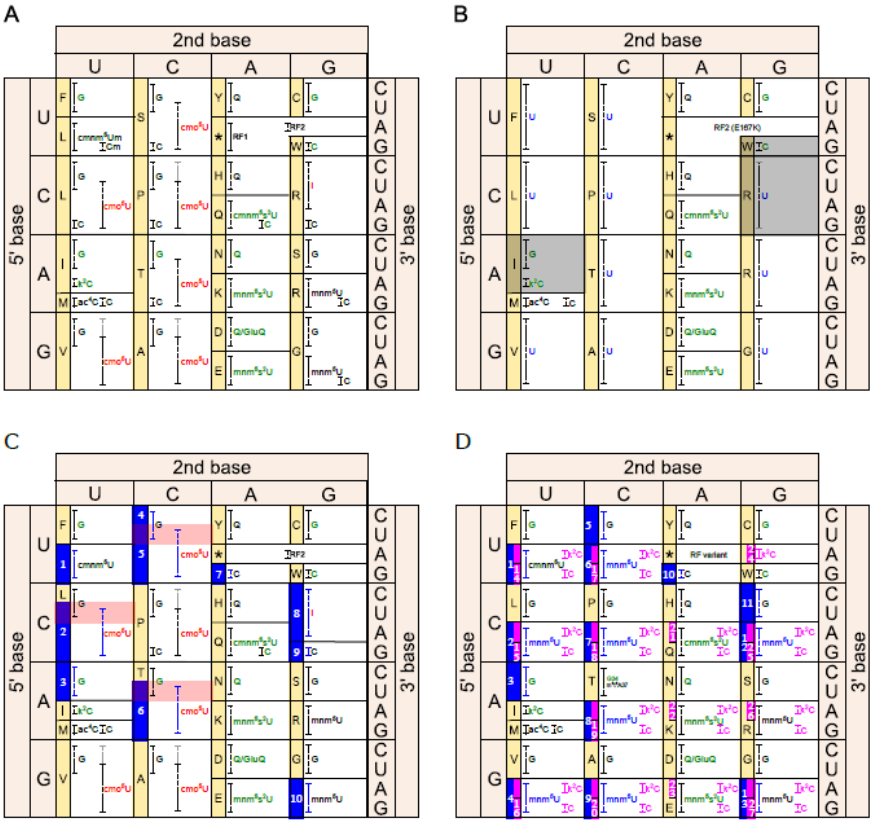


Fig 4

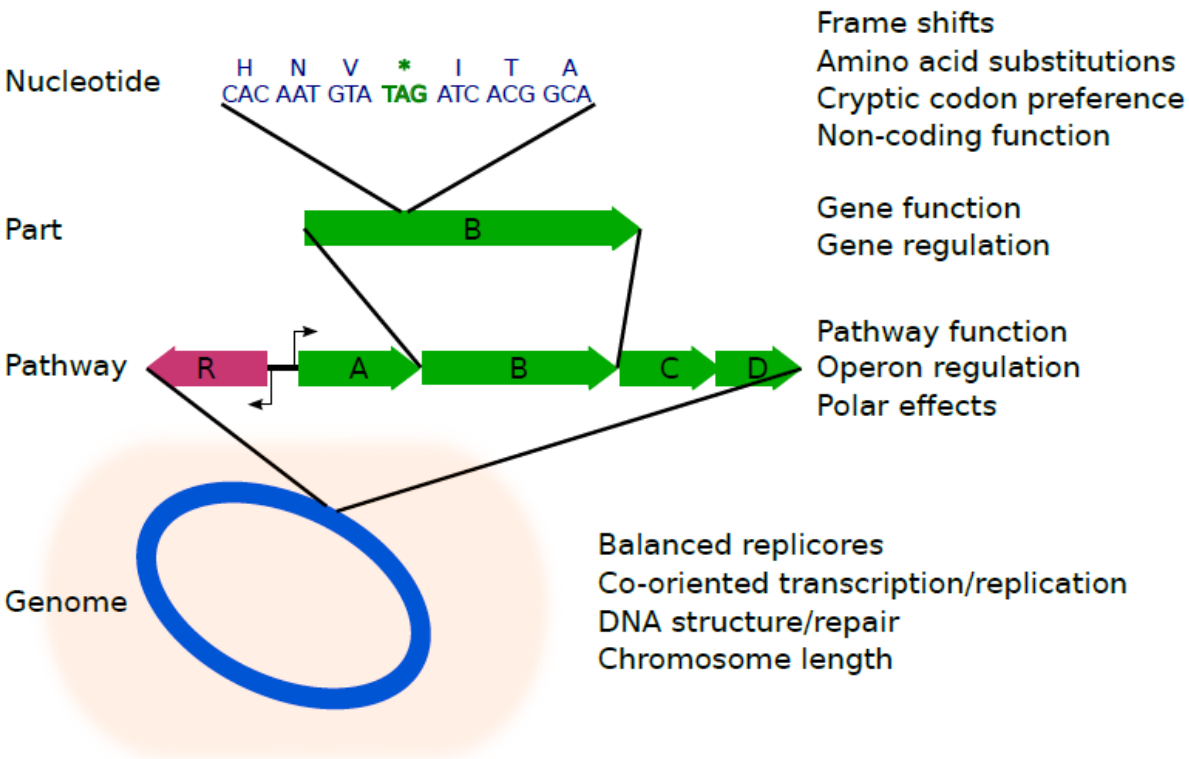
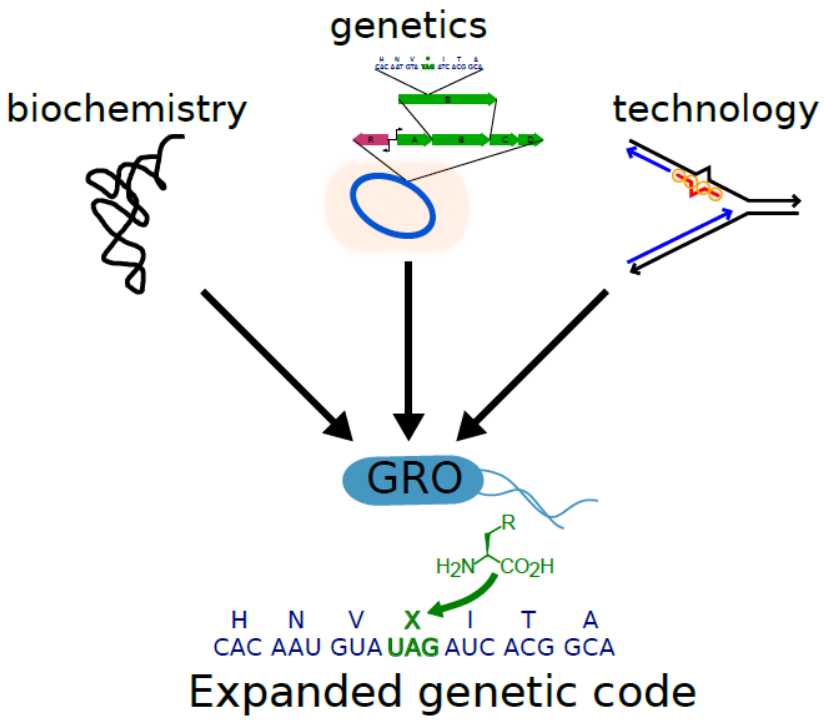


Fig. 5



Graphical abstract