Efforts and Challenges in Engineering the Genetic Code

Xiao Lin, Allen Chi Shing Yu and Ting Fung Chan *

School of Life Sciences, The Chinese University of Hong Kong, Sha Tin, NT, Hong Kong, China;
xlin@link.cuhk.edu.hk (X.L.); allenyu@cuhk.edu.hk (A.C.S.Y.)

* Correspondence: tf.chan@cuhk.edu.hk; Tel.: +852-3943-6876

Abstract: This year marks the 48th anniversary of Francis Crick's seminal work on the origin of the genetic code, in which he first proposed the "frozen accident" hypothesis to describe evolutionary selection against changes to the genetic code that cause devastating global proteome modification. However, numerous efforts have demonstrated the viability of both natural and artificial genetic code variations. Recent advances in genetic engineering allow the creation of synthetic organisms that incorporate noncanonical, or even unnatural, amino acids into the proteome. Currently, successful genetic code engineering is mainly achieved by creating orthogonal aminoacyl-tRNA/synthetase pairs to repurpose stop and rare codons or to induce quadruplet codons. In this review, we summarize the current progress in genetic code engineering and discuss the challenges, current understanding, and future perspectives regarding genetic code modification.

1. Introduction

In 1968, Francis Crick first proposed the frozen accident theory of the genetic code [1]. The 20 canonical amino acids were once believed to be immutable elements of the code. The genetic code appears to be universal, from simple unicellular organisms to complex vertebrates. Yet in contrast to studies of the natural selection of lifeforms wherein the gradual evolution of species can be observed in a myriad of taxa, relatively few examples of natural genetic code variations (e.g., selenocysteine [2], pyrrolysine [3,4], and stop codon read through [5,6]) have been observed. Different explanations have been proposed to address these variations, such as the codon capture hypothesis [7], the ambiguous intermediate hypothesis [8], and the genome streamlining hypothesis [9]. These hypotheses have been reviewed elsewhere [10,11]. Although some existing noncanonical amino acids (NCAAs) are known to be compatible with enzymatic aminoacylation [12–20], the 20 canonical amino acids in the standard genetic code have been stringently selected over the course of biological evolution. Organisms that require peptides with modified side chains will often resort to pre-translational or post-translational modifications to incorporate NCAAs [21–26]. Some organisms require alternative genetic codes to survive in harsh living conditions [27].

Genetic code engineering refers to the modification, or the directed evolution of cellular machineries, in order to incorporate NCAAs into the proteome of an organism. In general, NCAAs can be artificially incorporated in a site-specific or proteome-wide manner. In the former, scientists have attempted to artificially engineer organisms for compatibility with various NCAAs by employing orthogonal tRNA/aminoacyl-tRNA synthetase pairs [28–32]. In the latter, an organism is forced to take up specific NCAAs, followed by isolating mutants in media containing NCAAs [33–35]. Currently, researchers are recording cellular responses and genetic changes in engineered organisms to understand the mechanisms behind the use of alternative genetic codes. Such efforts enable

an understanding of the evolutionary course of the genetic code and provide a foundation for the derivation of additional alternative codes, a particularly important feature in the era of synthetic biology given the increased focus on engineering synthetic organisms with modified genetic codes [36,37]. Engineered genetic code holds tremendous potential in the field of protein engineering and xenobiology, which was extensively reviewed by Budisa et al. in 2017 [38].

In this review, we first give a brief introduction of current studies on both site-specific and proteome-wide incorporation of NCAAs. Next, we will focus on the challenges of engineering organisms to use modified genetic codes and their implications, such as inhibitory effects caused by NCAAs. Finally, we will discuss current trends in this research area.

2. Genetic Code Engineering
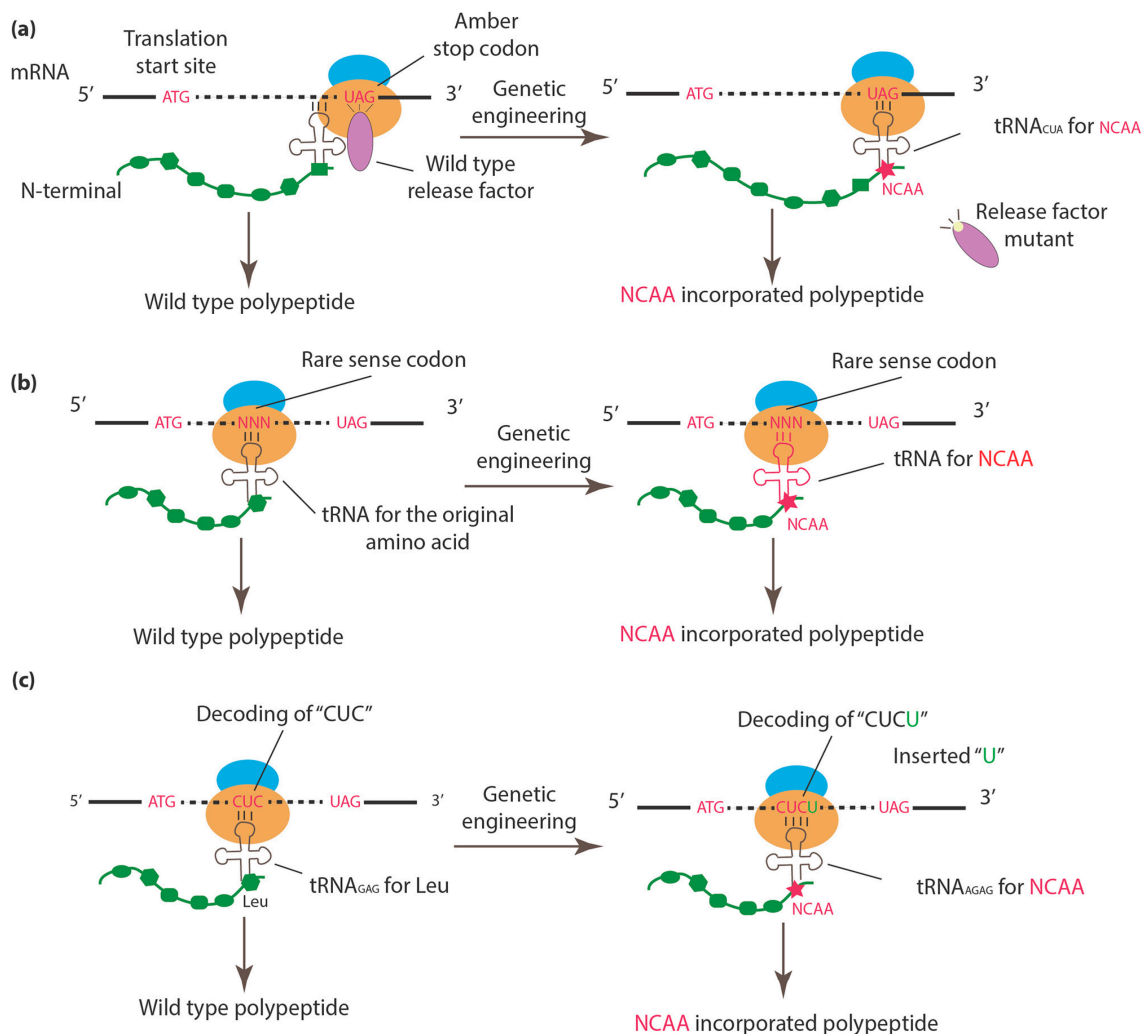
2.1. Incorporation of NCAAs into Specific Sites

Currently, three major approaches are used to engineer the genetic code in a site-specific manner: (1) amber codon suppression; (2) rare sense codon reassignment; and (3) quadruplet codon. Figure 1 provides a schematic illustration of each method. Because organisms such as *E. coli* BL21 rarely use the amber stop codon (UAG) (only 275 of 4160 stop codons in BL21 are amber codons), which minimizes disturbances to existing protein termination signals, this codon has been preferably selected for NCAA encoding [30,39,40]. To enhance the efficiency of amber codon recognition by the orthogonal tRNA$_{CUA}$, Release Factor 1 [41,42] is usually mutated or knocked out [43], thus enabling orthogonal tRNA$_{CUA}$ to recognize and increase its competitive binding to the amber codon [30] (Figure 1a). The role of different artificial tRNA/tRNA synthetase pairs, as well as their structural relationship with different NCAAs, were extensively reviewed by Anaëlle et al. [44].

Rare sense codon assignment [45–47], which is based on a similar principle, repurposes rare sense codons, particularly rare codons including AGG [45,46] and AUA [47], using newly designed tRNA/aminoacyl-tRNA synthetase pairs. In this method, the introduction of a NCAA during protein synthesis requires either competition between a genetically modified tRNA and the corresponding wild-type tRNA [45] or the inhibition of wild-type tRNA via the deletion of its tRNA synthetase [47] (Figure 1b).

To circumvent the limitations of reprogramming existing codons, some researchers have explored NCAA encoding via expansion of the genetic code using quadruplet codons [48–53]. In brief, a single-base (e.g., "U") is inserted after a canonical triplet codon (e.g., a "CUC" triplet codon) to form a frameshift mutation at the specific position (Figure 1c). The additional base also creates a new quadruplet codon (e.g., "CUCU") at this position, which can be recognized by an engineered quadruplet tRNA (e.g., tRNA$_{AGAG}$). Early versions of the quadruplet in vivo coding system were initially tested in *E. coli* [48,51,52], followed by *Xenopus* oocytes [49] and mammalian cells [50,53]. It is also worth mentioning that noncanonical RNA translations, such as the use of tetra- and penta-codon, were observed in mitochondria; however, the 4th and 5th nucleotides were found to be silent during translation [54]. More mechanistic studies would be required to establish their roles in genetic code engineering.

Although the site-specific incorporation approach is arguably the most widely used to produce artificial proteins with NCAAs, some challenges can limit the stability of the engineered code. The efficiency of an engineered tRNA/aminoacyl-tRNA synthetase pair must be high enough to minimize the generation of truncated proteins [55]. Methods such as orthogonal ribosome use can lead to a threefold improvement in the efficiency of unnatural amino acid incorporation [55]. Endogenous tRNA/aminoacyl-tRNA synthetase pairs can also be engineered to incorporate unnatural amino acids. For example, by changing the phenylalanyl-tRNA synthetase amino acid recognition site, phenylalanine analogs such as p-Cl-phenylalanine or p-Br-phenylalanine can be successfully charged to tRNA^{Phe} [14,56]. Advances in genome editing techniques, such as multiplex automated genomic

engineering [31] and CRISPR/Cas [51], may further increase the efficiency and accuracy of NCAA incorporation in specific sites of the proteome.

Figure 1. An overview of approaches to incorporate NCAAs into specific sites. (**a**) The wild-type release factor is mutated or knocked out, allowing the newly introduced tRNA$_{CUA}$ to read through the stop codon, followed by NCAA incorporation with assistance from the compatible aminoacyl-tRNA synthetase. (**b**) The tRNA and corresponding tRNA synthetase for a rare sense codon are genetically engineered to confer the ability to encode NCAA. (**c**) A single-base is inserted after the canonical codon (e.g. "CUC" for Leu). The newly introduced quadruplet tRNA (e.g., tRNA$_{AGAG}$) can encode NCAA by targeting the quadruplet codon "CUCU."

2.2. Proteome-Wide Incorporation of NCAAs

Proteome-wide incorporation offers an alternative approach toward unnatural amino acid incorporation. In the most common approach, amino acid uptake is artificially controlled by feeding auxotrophs with NCAAs [33–35] (Figure 2a). Attempts to control NCAA synthesis have involved supplying organisms with NCAA precursors [57–62] (Figure 2b), which is also known as metabolic engineering [63,64]. In one example, the precursor L-β-thieno [3,2-b]pyrrolyl ([3,2]Trp) was fed to a tryptophan (Trp)-auxotrophic *E. coli* capable of synthesizing [3,2]Tpa (a Trp analog) to generate mutants that could propagate on L-β-(thieno [3,2-b]pyrrolyl)alanine ([3,2]Tpa) [57] (Figure 2b). Although directly feeding auxotrophs with NCAAs is a simpler approach, metabolic engineering could reduce the unwanted effects of impure commercial NCAAs [57].
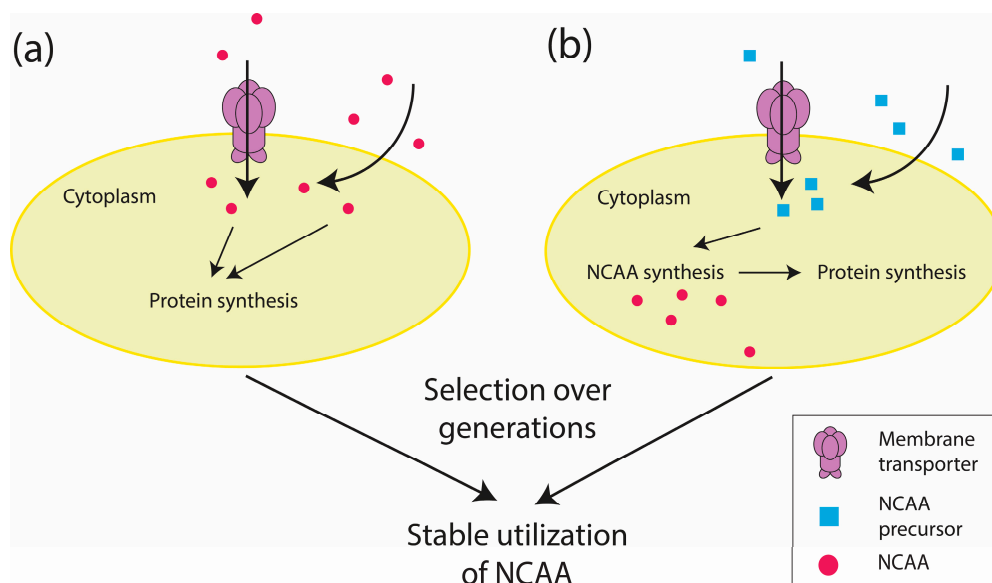
Figure 2. An overview of proteome-wide approaches to incorporate NCAAs. (**a**) The NCAA enters a cell via membrane transporters or diffusion across the membrane. (**b**) The NCAA precursor similarly enters a cell in which it will be used to synthesize NCAAs. Following several generations of propagation with either the NCAA or its precursor, cells that can stably utilize the NCAA are selected.

Regardless of approach, the incorporation of NCAAs in the proteome may negatively affect the growth of an organism. The inherent toxicities of many unnatural amino acids could suppress propagation of the wild-type strain and select mutants that respond favorably to the NCAA, ultimately causing rejection of the expanded genetic code [34,35]. In the following section, we will focus on the challenges in genetic code modification.

3. Challenges of Genetic Code Engineering

3.1. Inhibitory Effects of Engineered Genetic Codes

The growth inhibitory effects caused by NCAAs, which have been demonstrated in different species including bacteria [65–67], yeasts [68], insects [69,70], and mammals [71], comprise one major challenge encountered during genetic code modification. The inhibitory effects of NCAAs are mainly attributable to two aspects. First, minor structural and chemical differences between NCAAs and their canonical counterparts can drastically affect enzymatic activities [72–75]. Second, these structural and chemical differences may also negatively affect protein synthesis, as some NCAAs cannot be efficiently charged to tRNAs by aminoacyl-tRNA synthetases [15,76]. A better understanding of the key genes and cellular responses associated with these modified genetic codes is of paramount importance to alleviating these inhibitory effects.

3.2. Discovering the Key Genes Controlling the Genetic Code

The growth inhibitory potentials of NCAAs create negative selective pressure, while the organism adapts to the modified genetic code. One effective strategy for overcoming this evolutionary barrier comprises an increase in the mutation rate via mutagenesis with the expectation of generating beneficial mutations that would favor the NCAA. Wong and colleagues isolated mutants from a Trp auxotroph (*Bacillus subtilis* str. QB928) via sequential mutagenesis in an early attempt to modify the genetic code. The resultant HR23 strain could propagate indefinitely on 4-fluoro-tryptophan (4FTrp) but became inviable on canonical Trp [34,35]. As Trp is encoded by a single codon (UGG), the research by Wong and colleagues provided the first evidence of codon membership malleability under external

selection pressure. Subsequently, Yu et al. traced mutations in intermediate mutants, as well as the HR23 strain [77]. A nonsense mutation in the Trp operon RNA-binding attenuation protein (TRAP), which controls transcriptional attenuation of the Trp operon [78,79] and translational repression of Trp transporters [80–83], was shared by all mutants. This lack of TRAP would increase 4-FTrp uptake to compensate for the relatively low charge rate of 4-FTrp to tRNA^{Trp} [15].

In a separate attempt, Bacher et al. isolated *E. coli* mutants that could propagate in medium wherein 4-FTrp comprised ~99% of available Trp. However, the mutant strains could not grow indefinitely under these conditions and required minimal canonical Trp [33]. *E. coli* mutants were found to harbor several mutations affecting genes such as *aroP*, which encodes an aromatic amino acid transporter [84], and *tyrR*, which encodes the associated regulator [85]. Mutated *aroP* and *tyrR* might cooperatively increase 4-FTrp uptake, similar to the effect of TRAP knockout in *B. subtilis.* Taken together, these findings suggest that an efficient NCAA uptake system is essential to accommodation of the modified genetic codes.

RNA polymerase might also play a key role in controlling the genetic code. The above-mentioned *B. subtilis* mutant HR23 was found to harbor a nonsynonymous mutation in the RNA polymerase subunit gene (*rpoB*) that was absent from all other intermediate strains that could still propagate on Trp, suggesting a potential role for this mutation in switching membership of the UGG codon from Trp to 4-FTrp [77]. In an independent study of amber codon-directed 3-iodotyrosine (3-iodoTyr) incorporation in *E. coli*, a *rpoB* mutation was found to confer rifampicin resistance via amber suppression at Gln513 [86], and the same research group also engineered a bacteriophage, T7, that could incorporate 3-iodoTyr at amber codons [29]. In that study, Hammerling et al. observed high mutation frequencies in genes encoding RNA polymerase and the lysis timing regulator type II holin. The authors suggested that these two genes played important roles in the evolution of the expanded genetic code [29]. These studies have shed light on the previously unexplored roles of key genes in genetic code identity.

3.3. Lack of Transcriptomic and Proteomic Studies Related to Engineered Genetic Codes

In addition to mutations, gene and protein expression profiles might also reveal key factors needed to fine-tune the use of modified genetic codes. Technologies such as RNA-seq and mass spectrometry can be used to investigate the cellular responses of organisms in high resolution. RNA-seq was used to compare the cellular responses between mutant (grown on 4-FTrp) and wild-type strains of the above-mentioned *B. subtilis* HR23 mutant (unpublished data). Here, a gene ontology analysis of the gene expression profiles of these strains demonstrated enrichment of genes related to reactive oxygen species responses and branched-chain amino acid biosynthetic processes among upregulated genes, and enrichment of genes related to siderophore biosynthetic processes among downregulated genes (unpublished data). Unsurprisingly, stress response genes were modulated in response to the new genetic code, and the downregulation of siderophore biosynthetic process related genes was consistent with a previous observation of the reduced growth rate of HR23 cells grown on 4-FTrp [77] because iron homeostasis is closely related to bacterial growth [87]. This unique set of data was the first to demonstrate the adaptation of an organism to a new genetic code at the transcriptomic level.

Methanosarcina acetivorans is a methanogenic archaea strain that uses the alternative genetic codon UAG to encode pyrrolysine (Pyl) [88]. O'Donoghue et al. attempted to reduce the genetic code of this strain by deleting tRNA^{Pyl}, thus blocking the incorporation of Pyl in the proteome. A comparison of the proteomes of mutant and wild-type *M. acetivorans* strains revealed that most upregulated peptides were related to methanogenesis, protein synthesis, and the stress response [89], suggesting that, in this organism, various stress response genes must be fine-tuned before a reduced genetic code can be used.

Very few transcriptomic and proteomic studies of organisms with modified genetic codes have been conducted, and we have only glimpsed the potential factors involved in adaptation to modified genetic codes. Additional genes that contribute to this adaptation might remain to be discovered. In the future, studies of gene and protein expression in organisms with modified genetic codes will be necessary.

3.4. Environmental Factors Affecting Adaptation to Engineered Genetic Codes

Environmental factors, such as the growth medium and selection method, are important when optimizing the use of a modified genetic code. The amino acid source is the first and most obvious factor, as an organism can either take up NCAAs directly from the environment or synthesize them using environmentally available molecules. If the source of NCAAs is from the environment, mutations in amino acid transporters are often needed to facilitate NCAA uptake [33,77].

Positive selection pressure is also needed to maintain stability of the modified genetic code. In a previous study, incorporation of the methionine analog azidohomoalanine (Aha) into the coat protein of a human adenovirus and the subsequent addition of a folate group to Aha facilitated adenoviral infection in mouse hosts [90]. In other words, adenovirus strains that can use modified Aha have a selective survival advantage over other strains. In a more recent study of different *E. coli* strains, the site-specific incorporation of two tyrosine analogs in β-lactamase was selected, and enzymatic function was found to depend on the presence of these analogs [91]. As described above regarding adenovirus, *E. coli* mutants that could utilize NCAAs enjoyed a selective advantage under growth mediums containing certain classes of antibiotics [91]. In one interesting example, even the carbon source may affect the selection of genetic codes by the Pyl-utilizing bacteria *Acetohalobium arabaticum* [92]. *A. arabaticum* used the standard genetic code when grown on pyruvate, but gained the ability to use an expanded genetic code that included Pyl in the presence of the alternative carbon source trimethylamine [92].

4. Future Directions

Current efforts in genetic code engineering have reshaped our ideas regarding genetic code evolution and have paved the way for expanding the genetic alphabet. Based on these studies, we have outlined the key steps by which an organism accommodates a modified genetic code (Figure 3). During adaptation, mutations in amino acid transporters and/or their key regulators allow more efficient NCAA uptake, possibly by increasing the number of amino acid transporters [33,77]. Mutations in the key genes might also favor the use of a modified genetic code [29,33,77,86]. Additionally, environmental positive selection forces contribute to stability of the modified genetic code [77,91,92]. Currently, the genomic changes in organisms with modified genetic codes have been well explored [29,33,77,86] relative to transcriptomic (unpublished data) and proteomic [89] changes. Future trends in elucidation of the biological mechanisms underlying genetic code modifications include the integration of genomic, transcriptomic, and proteomic data and the refining of functional study targets.

From the viewpoint of synthetic biology and xenobiology, genetic code engineering increases the repertoire of building blocks available for protein engineering, thus enabling the development of novel proteins that would be impossible with canonical amino acids [38,93]. Xenobiology is an emerging field that involves synthesizing xenonucleic acids other than the canonical nucleic acids with adenine (A), thymine (T), cytosine (C), and guanine (G) as bases, with alternative pairing rules for protein engineering [94]. It has been demonstrated experimentally that two such xenonucleic acids can be integrated into the current DNA backbone [95–97], and more have been tested for their potentials as novel building blocks of DNA [98]. With the addition of xenonucleic acids, the number of encoded amino acids is likely to be increased to far beyond 20 [94].

High-throughput genome editing technologies, such as MAGE [36] and the emerging CRISPR/Cas technology [99], allow an organism's genetic code to be directly rewritten [37,100] and facilitate the creation of synthetic life [101]. Although the first synthetic minimal bacterial genome still uses the standard genetic code [101], it is now possible to synthesize genomes based on alternative genetic codes. A full exploration of the possibilities enabled by genetic code engineering requires an understanding of the key molecular biological and biochemical mechanisms underlying the modifications. Gradual efforts to address this main question may improve our understanding of the process of genetic code evolution and lay a better foundation for future synthetic biology research.
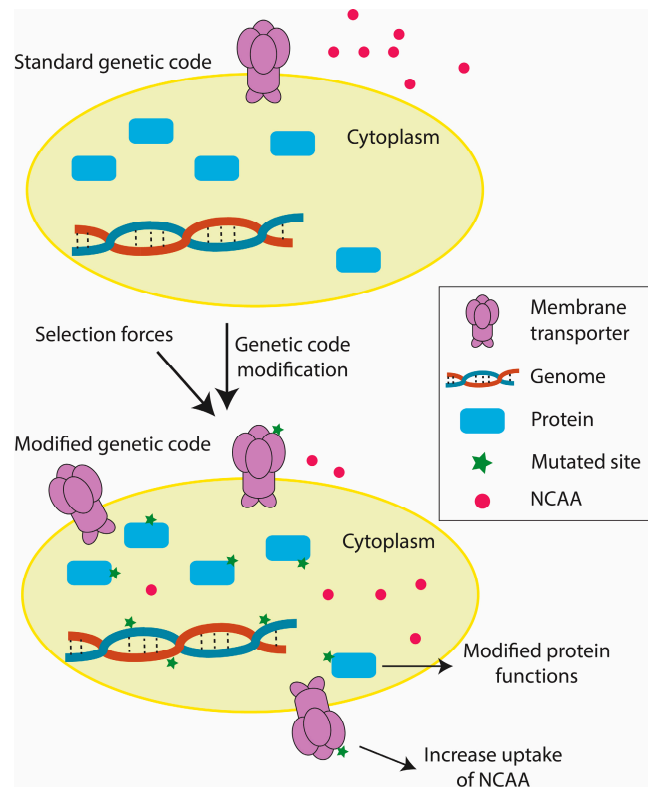
Figure 3. Key steps in the accommodation of a modified genetic code.

Author Contributions: X.L., A.C.S.Y., and T.F.C. wrote the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Crick, F.H.C. The origin of the genetic code. *J. Mol. Biol.* **1968**, *38*, 367–379. [CrossRef]
2. Cone, J.E.; Del Río, R.M.; Davis, J.N.; Stadtman, T.C. Chemical characterization of the selenoprotein component of clostridial glycine reductase: Identification of selenocysteine as the organoselenium moiety. *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 2659–2663. [CrossRef] [PubMed]
3. Hao, B.; Gong, W.; Ferguson, T.K.; James, C.M.; Krzycki, J.A.; Chan, M.K. A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **2002**, *296*, 1462–1466. [CrossRef] [PubMed]
4. Srinivasan, G.; James, C.M.; Krzycki, J.A. Pyrrolysine encoded by UAG in archaea: Charging of a UAG-decoding specialized tRNA. *Science* **2002**, *296*, 1459–1462. [CrossRef] [PubMed]
5. Adachi, M.; Cavalcanti, A.R.O. Tandem stop codons in ciliates that reassign stop codons. *J. Mol. Evol.* **2009**, *68*, 424–431. [CrossRef] [PubMed]
6. Beznosková, P.; Gunišová, S.; Valášek, L.S. Rules of UGA-N decoding by near-cognate tRNAs and analysis of readthrough on short uORFs in yeast. *RNA* **2016**, *22*, 456–466. [CrossRef] [PubMed]
7. Osawa, S.; Jukes, T.H. Codon reassignment (codon capture) in evolution. *J. Mol. Evol.* **1989**, *28*, 271–278. [CrossRef] [PubMed]
8. Schultz, D.W.; Yarus, M. Transfer RNA mutation and the malleability of the genetic code. *J. Mol. Biol.* **1994**, *235*, 1377–1380. [CrossRef] [PubMed]
9. Andersson, S.G.; Kurland, C.G. Genomic evolution drives the evolution of the translation system. *Biochem. Cell Biol.* **1995**, *73*, 775–787. [CrossRef] [PubMed]

10. Söll, D.; RajBhandary, U.L. The genetic code—Thawing the "frozen accident". *J. Biosci.* **2006**, *31*, 459–463. [CrossRef] [PubMed]

11. Koonin, E.V.; Novozhilov, A.S. Origin and evolution of the genetic code: The universal enigma. *IUBMB Life* **2009**, *61*, 99–111. [CrossRef] [PubMed]

12. Hartman, M.C.T.; Josephson, K.; Lin, C.-W.; Szostak, J.W. An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. *PLoS ONE* **2007**, *2*, e972. [CrossRef] [PubMed]

13. Wang, L.; Xie, J.; Schultz, P.G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249. [CrossRef] [PubMed]

14. Ibba, M.; Hennecke, H. Relaxing the substrate specificity of an aminoacyl-tRNA synthetase allows in vitro and in vivo synthesis of proteins containing unnatural amino acids. *FEBS Lett.* **1995**, *364*, 272–275. [CrossRef]

15. Xu, Z.J.; Love, M.L.; Ma, L.Y.; Blum, M.; Bronskill, P.M.; Bernstein, J.; Grey, A.A.; Hofmann, T.; Camerman, N.; Wong, J.T. Tryptophanyl-tRNA synthetase from Bacillus subtilis. Characterization and role of hydrophobicity in substrate recognition. *J. Biol. Chem.* **1989**, *264*, 4304–4311. [PubMed]

16. Lovett, P.S.; Ambulos, N.P.; Mulbry, W.; Noguchi, N.; Rogers, E.J.; Rogers, E.J. UGA can be decoded as tryptophan at low efficiency in *Bacillus subtilis*. *J. Bacteriol.* **1991**, *173*, 1810–1812. [CrossRef] [PubMed]

17. Cataldo, F.; Iglesias-Groth, S.; Angelini, G.; Hafez, Y. Stability toward high energy radiation of non-proteinogenic amino acids: Implications for the origins of life. *Life* **2013**, *3*, 449–473. [CrossRef] [PubMed]

18. Richmond, M.H. The effect of amino acid analogues on growth and protein synthesis in microorganisms. *Bacteriol. Rev.* **1962**, *26*, 398–420. [PubMed]

19. Lea, P.J.; Norris, R.D. The use of amino acid analogues in studies on plant metabolism. *Phytochemistry* **1976**, *15*, 585–595. [CrossRef]

20. Rodgers, K.J.; Hume, P.M.; Dunlop, R.A.; Dean, R.T. Biosynthesis and turnover of DOPA-containing proteins by human cells. *Free Radic. Biol. Med.* **2004**, *37*, 1756–1764. [CrossRef] [PubMed]

21. Commans, S.; Böck, A. Selenocysteine inserting tRNAs: An overview. *FEMS Microbiol. Rev.* **1999**, *23*, 335–351. [CrossRef] [PubMed]

22. Berry, M.J.; Banu, L.; Chen, Y.; Mandel, S.J.; Kieffer, J.D.; Harney, J.W.; Larsen, P.R. Recognition of UGA as a selenocysteine codon in Type I deiodinase requires sequences in the 3′ untranslated region. *Nature* **1991**, *353*, 273–276. [CrossRef] [PubMed]

23. Silva, R.M.; Paredes, J.A.; Moura, G.R.; Manadas, B.; Lima-Costa, T.; Rocha, R.; Miranda, I.; Gomes, A.C.; Koerkamp, M.J.G.; Perrot, M.; et al. Critical roles for a genetic code alteration in the evolution of the genus Candida. *EMBO J.* **2007**, *26*, 4555–4565. [CrossRef] [PubMed]

24. Krzycki, J.A. Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. *Curr. Opin. Chem. Biol.* **2004**, *8*, 484–491. [CrossRef] [PubMed]

25. Polycarpo, C.; Ambrogelly, A.; Bérubé, A.; Winbush, S.M.; McCloskey, J.A.; Crain, P.F.; Wood, J.L.; Söll, D. An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12450–12454. [CrossRef] [PubMed]

26. Mann, M.; Jensen, O.N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **2003**, *21*, 255–261. [CrossRef] [PubMed]

27. Rother, M.; Krzycki, J.A. Selenocysteine, pyrrolysine, and the unique energy metabolism of methanogenic archaea. *Archaea* **2010**, *2010*, 453642. [CrossRef] [PubMed]

28. Bacher, J.M.; Hughes, R.A.; Tze-Fei Wong, J.; Ellington, A.D. Evolving new genetic codes. *Trends Ecol. Evol.* **2004**, *19*, 69–75. [CrossRef] [PubMed]

29. Hammerling, M.J.; Ellefson, J.W.; Boutz, D.R.; Marcotte, E.M.; Ellington, A.D.; Barrick, J.E. Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness. *Nat. Chem. Biol.* **2014**, *10*, 178–180. [CrossRef] [PubMed]

30. Wang, L.; Brock, A.; Herberich, B.; Schultz, P.G. Expanding the genetic code of *Escherichia coli*. *Science* **2001**, *292*, 498–500. [CrossRef] [PubMed]

31. Park, H.-S.; Hohn, M.J.; Umehara, T.; Guo, L.-T.; Osborne, E.M.; Benner, J.; Noren, C.J.; Rinehart, J.; Söll, D. Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science* **2011**, *333*, 1151–1154. [CrossRef] [PubMed]

32. Mukai, T.; Yanagisawa, T.; Ohtake, K.; Wakamori, M.; Adachi, J.; Hino, N.; Sato, A.; Kobayashi, T.; Hayashi, A.; Shirouzu, M.; et al. Genetic-code evolution for protein synthesis with non-natural amino acids. *Biochem. Biophys. Res. Commun.* **2011**, *411*, 757–761. [CrossRef] [PubMed]

33. Bacher, J.M.; Ellington, A.D. Selection and characterization of *Escherichia coli* variants capable of growth on an otherwise toxic tryptophan analogue. *J. Bacteriol.* **2001**, *183*, 5414–5425. [CrossRef] [PubMed]

34. Wong, J.T. Membership mutation of the genetic code: Loss of fitness by tryptophan. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 6303–6306. [CrossRef] [PubMed]

35. Mat, W.K.; Xue, H.; Wong, J.T.F. Genetic code mutations: The breaking of a three billion year invariance. *PLoS ONE* **2010**, *5*, e12206. [CrossRef] [PubMed]

36. Gallagher, R.R.; Li, Z.; Lewis, A.O.; Isaacs, F.J. Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA. *Nat. Protoc.* **2014**, *9*, 2301–2316. [CrossRef] [PubMed]

37. Ostrov, N.; Landon, M.; Guell, M.; Kuznetsov, G.; Teramoto, J.; Cervantes, N.; Zhou, M.; Singh, K.; Napolitano, M.G.; Moosburner, M.; et al. Design, synthesis, and testing toward a 57-codon genome. *Science* **2016**, *353*, 819–822. [CrossRef] [PubMed]

38. Budisa, N.; Völler, J.-S.; Koksch, B.; Acevedo-Rocha, C.G.; Kubyshkin, V.; Agostini, F. Xenobiology meets enzymology: Exploring the potential of unnatural building blocks in biocatalysis. *Angew. Chem. Int. Ed.* **2017**. [CrossRef]

39. Wang, L.; Brock, A.; Schultz, P.G. Adding L-3-(2-Naphthyl)alanine to the genetic code of *E. coli*. *J. Am. Chem. Soc.* **2002**, *124*, 1836–1837. [CrossRef] [PubMed]

40. Chin, J.W.; Martin, A.B.; King, D.S.; Wang, L.; Schultz, P.G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11020–11024. [CrossRef] [PubMed]

41. Scolnick, E.; Tompkins, R.; Caskey, T.; Nirenberg, M. Release factors differing in specificity for terminator codons. *Proc. Natl. Acad. Sci. USA* **1968**, *61*, 768–774. [CrossRef] [PubMed]

42. Petry, S.; Brodersen, D.E.; Murphy, F.V.; Dunham, C.M.; Selmer, M.; Tarry, M.J.; Kelley, A.C.; Ramakrishnan, V. Crystal structures of the ribosome in complex with release factors RF1 and RF2 bound to a cognate stop codon. *Cell* **2005**, *123*, 1255–1266. [CrossRef] [PubMed]

43. Johnson, D.B.F.; Xu, J.; Shen, Z.; Takimoto, J.K.; Schultz, M.D.; Schmitz, R.J.; Xiang, Z.; Ecker, J.R.; Briggs, S.P.; Wang, L. RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* **2011**, *7*, 779–786. [CrossRef] [PubMed]

44. Dumas, A.; Lercher, L.; Spicer, C.D.; Davis, B.G. Designing logical codon reassignment—Expanding the chemistry in biology. *Chem. Sci.* **2015**, *6*, 50–69. [CrossRef]

45. Zeng, Y.; Wang, W.; Liu, W.R. Towards reassigning the rare AGG codon in *Escherichia coli*. *ChemBioChem* **2014**, *15*, 1750–1754. [CrossRef] [PubMed]

46. Mukai, T.; Yamaguchi, A.; Ohtake, K.; Takahashi, M.; Hayashi, A.; Iraha, F.; Kira, S.; Yanagisawa, T.; Yokoyama, S.; Hoshi, H.; et al. Reassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli*. *Nucleic Acids Res.* **2015**, *43*, 8111–8122. [CrossRef] [PubMed]

47. Bohlke, N.; Budisa, N. Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: Rare isoleucine codon AUA as a target for genetic code expansion. *FEMS Microbiol. Lett.* **2014**, *351*, 133–144. [CrossRef] [PubMed]

48. Anderson, J.C.; Wu, N.; Santoro, S.W.; Lakshman, V.; King, D.S.; Schultz, P.G. An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7566–7571. [CrossRef] [PubMed]

49. Rodriguez, E.A.; Lester, H.A.; Dougherty, D.A. In vivo incorporation of multiple unnatural amino acids through nonsense and frameshift suppression. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8650–8655. [CrossRef] [PubMed]

50. Niu, W.; Schultz, P.G.; Guo, J. An expanded genetic code in mammalian cells with a functional quadruplet codon. *ACS Chem. Biol.* **2013**, *8*, 1640–1645. [CrossRef] [PubMed]

51. Magliery, T.J.; Anderson, J.C.; Schultz, P.G. 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*, 755–769. [CrossRef] [PubMed]

52. Anderson, J.C.; Magliery, T.J.; Schultz, P.G. Exploring the limits of codon and anticodon size. *Chem. Biol.* **2002**, *9*, 237–244. [CrossRef]

53. Taki, M.; Matsushita, J.; Sisido, M. Expanding the genetic code in a mammalian cell line by the introduction of four-base codon/anticodon pairs. *ChemBioChem* **2006**, *7*, 425–428. [CrossRef] [PubMed]

54. Seligmann, H. Natural chymotrypsin-like-cleaved human mitochondrial peptides confirm tetra-, pentacodon, non-canonical RNA translations. *Biosystems* **2016**, *147*, 78–93. [CrossRef] [PubMed]

55. Wang, K.; Neumann, H.; Peak-Chew, S.Y.; Chin, J.W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.* **2007**, *25*, 770–777. [CrossRef] [PubMed]

56. Kast, P.; Hennecke, H. Amino acid substrate specificity of *Escherichia coli* phenylalanyl-tRNA synthetase altered by distinct mutations. *J. Mol. Biol.* **1991**, *222*, 99–124. [CrossRef]

57. Hoesl, M.G.; Oehm, S.; Durkin, P.; Darmon, E.; Peil, L.; Aerni, H.-R.; Rappsilber, J.; Rinehart, J.; Leach, D.; Söll, D.; et al. Chemical evolution of a bacterial proteome. *Angew. Chem. Int. Ed.* **2015**, *54*, 10030–10034. [CrossRef] [PubMed]

58. Ehrlich, M.; Gattner, M.J.; Viverge, B.; Bretzler, J.; Eisen, D.; Stadlmeier, M.; Vrabel, M.; Carell, T. Orchestrating the biosynthesis of an unnatural pyrrolysine amino acid for its direct incorporation into proteins inside living cells. *Chem. Eur. J.* **2015**, *21*, 7701–7704. [CrossRef] [PubMed]

59. Ma, Y.; Biava, H.; Contestabile, R.; Budisa, N.; di Salvo, M. Coupling bioorthogonal chemistries with artificial metabolism: Intracellular biosynthesis of azidohomoalanine and its incorporation into recombinant proteins. *Molecules* **2014**, *19*, 1004–1022. [CrossRef] [PubMed]

60. Exner, M.P.; Kuenzl, T.; To, T.M.T.; Ouyang, Z.; Schwagerus, S.; Hoesl, M.G.; Hackenberger, C.P.R.; Lensen, M.C.; Panke, S.; Budisa, N. Design of S-allylcysteine in situ production and incorporation based on a novel pyrrolysyl-tRNA synthetase variant. *Chembiochem* **2017**, *18*, 85–90. [CrossRef] [PubMed]

61. Ou, W.; Uno, T.; Chiu, H.-P.; Grunewald, J.; Cellitti, S.E.; Crossgrove, T.; Hao, X.; Fan, Q.; Quinn, L.L.; Patterson, P.; et al. Site-specific protein modifications through pyrroline-carboxy-lysine residues. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10437–10442. [CrossRef] [PubMed]

62. Mehl, R.A.; Anderson, J.C.; Santoro, S.W.; Wang, L.; Martin, A.B.; King, D.S.; Horn, D.M.; Schultz, P.G. Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* **2003**, *125*, 935–939. [CrossRef] [PubMed]

63. Acevedo-Rocha, C.G.; Budisa, N. Xenomicrobiology: A roadmap for genetic code engineering. *Microb. Biotechnol.* **2016**, *9*, 666–676. [CrossRef] [PubMed]

64. Völler, J.-S.; Budisa, N. Coupling genetic code expansion and metabolic engineering for synthetic cells. *Curr. Opin. Biotechnol.* **2017**, *48*, 1–7. [CrossRef] [PubMed]

65. Grant, M.M.; Brown, A.S.; Corwin, L.M.; Troxler, R.F.; Franzblau, C. Effect of L-azetidine 2-carboxylic acid on growth and proline metabolism in *Escherichia coli*. *Biochim. Biophys. Acta* **1975**, *404*, 180–187. [CrossRef]

66. Unger, L.; DeMoss, R.D. Action of a proline analogue, L-thiazolidine-4-carboxylic acid, in *Escherichia coli*. *J. Bacteriol.* **1966**, *91*, 1556–1563. [PubMed]

67. Moran, S.; Rai, D.K.; Clark, B.R.; Murphy, C.D. Precursor-directed biosynthesis of fluorinated iturin A in *Bacillus* spp. *Org. Biomol. Chem.* **2009**, *7*, 644. [CrossRef] [PubMed]

68. Téllez, R.; Jacob, G.; Basilio, C.; George-Nascimento, C. Effect of ethionine on the in vitro synthesis and degradation of mitochondrial translation products in yeast. *FEBS Lett.* **1985**, *192*, 88–94. [CrossRef]

69. Rosenthal, G.; Lambert, J.; Hoffmann, D. Canavanine incorporation into the antibacterial proteins of the fly, *Phormia terranovae* (Diptera), and its effect on biological activity. *J. Biol. Chem.* **1989**, *26417*, 9768–9771.

70. Teramoto, H.; Kojima, K. Incorporation of methionine analogues into bombyx mori silk fibroin for click modifications. *Macromol. Biosci.* **2015**, *15*, 719–727. [CrossRef] [PubMed]

71. Poirson-Bichat, F.; Lopez, R.; Bras Gonçalves, R.A.; Miccoli, L.; Bourgeois, Y.; Demerseman, P.; Poisson, M.; Dutrillaux, B.; Poupon, M.F. Methionine deprivation and methionine analogs inhibit cell proliferation and growth of human xenografted gliomas. *Life Sci.* **1997**, *60*, 919–931. [CrossRef]

72. Merkel, L.; Budisa, N. Organic fluorine as a polypeptide building element: In vivo expression of fluorinated peptides, proteins and proteomes. *Org. Biomol. Chem.* **2012**, *10*, 7241. [CrossRef] [PubMed]

73. Schlesinger, S. The effect of amino acid analogues on alkaline phosphatase. Formation in *Escherichia coli* K-II. Replacement of tryptophan by azatryptophan and by tryptazan. *J. Biol. Chem.* **1968**, *243*, 3877–3883. [PubMed]

74. Pine, M.J. Comparative physiological effects of incorporated amino acid analogs in *Escherichia coli*. *Antimicrob. Agents Chemother.* **1978**, *13*, 676–685. [CrossRef] [PubMed]

75. Wong, H.; Kwon, I. Effects of non-natural amino acid incorporation into the enzyme core region on enzyme structure and function. *Int. J. Mol. Sci.* **2015**, *16*, 22735–22753. [CrossRef] [PubMed]

76. Kwon, I.; Tirrell, D.A. Site-specific incorporation of tryptophan analogues into recombinant proteins in bacterial cells. **2007**, *129*, 10431–10437. [CrossRef] [PubMed]

77. Yu, A.C.-S.; Yim, A.K.-Y.; Mat, W.-K.; Tong, A.H.-Y.; Lok, S.; Xue, H.; Tsui, S.K.-W.; Wong, J.T.-F.; Chan, T.-F. Mutations enabling displacement of tryptophan by 4-fluorotryptophan as a canonical amino acid of the genetic code. *Genome Biol. Evol.* **2014**, *6*, 629–641. [CrossRef] [PubMed]

78. Gollnick, P.; Ishino, S.; Kuroda, M.I.; Henner, D.J.; Yanofsky, C. The mtr locus is a two-gene operon required for transcription attenuation in the trp operon of *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8726–8730. [CrossRef] [PubMed]

79. Yanofsky, C. RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria. *RNA* **2007**, *13*, 1141–1154. [CrossRef] [PubMed]

80. Sarsero, J.P.; Merino, E.; Yanofsky, C. A Bacillus subtilis gene of previously unknown function, *yhaG*, is translationally regulated by tryptophan-activated TRAP and appears to be involved in tryptophan transport. *J. Bacteriol.* **2000**, *182*, 2329–2331. [CrossRef] [PubMed]

81. Yakhnin, H.; Yakhnin, A.V.; Babitzke, P. The trp RNA-binding attenuation protein (TRAP) of Bacillus subtilis regulates translation initiation of *ycbK*, a gene encoding a putative efflux protein, by blocking ribosome binding. *Mol. Microbiol.* **2006**, *61*, 1252–1266. [CrossRef] [PubMed]

82. Yakhnin, H.; Zhang, H.; Yakhnin, A.V.; Babitzke, P. The trp RNA-binding attenuation protein of Bacillus subtilis regulates translation of the tryptophan transport gene trpP (*yhaG*) by blocking ribosome binding. *J. Bacteriol.* **2004**, *186*, 278–286. [CrossRef] [PubMed]

83. Du, H.; Tarpey, R.; Babitzke, P. The trp RNA-binding attenuation protein regulates TrpG synthesis by binding to the trpG ribosome binding site of *Bacillus subtilis*. *J. Bacteriol.* **1997**, *179*, 2582–2586. [CrossRef] [PubMed]

84. Honoré, N.; Cole, S.T. Nucleotide sequence of the *aroP* gene encoding the general aromatic amino acid transport protein of *Escherichia coli* K-12: Homology with yeast transport proteins. *Nucleic Acids Res.* **1990**, *18*, 653. [CrossRef] [PubMed]

85. Wang, J.G.; Fan, C.S.; Wu, Y.Q.; Jin, R.L.; Liu, D.X.; Shang, L.; Jiang, P.H. Regulation of *aroP* expression by *tyrR* gene in *Escherichia coli*. *Acta Biochim. Biophys. Sin.* **2003**, *35*, 993–997. [PubMed]

86. Hammerling, M.J.; Gollihar, J.; Mortensen, C.; Alnahhas, R.N.; Ellington, A.D.; Barrick, J.E. Expanded genetic codes create new mutational routes to rifampicin resistance in *Escherichia coli*. *Mol. Biol. Evol.* **2016**, *33*, 2054–2063. [CrossRef] [PubMed]

87. Andrews, S.C.; Robinson, A.K.; Rodríguez-Quiñones, F. Bacterial iron homeostasis. *FEMS Microbiol. Rev.* **2003**, *27*, 215–237. [CrossRef]

88. Borrel, G.; Gaci, N.; Peyret, P.; O'Toole, P.W.; Gribaldo, S.; Brugère, J.F. Unique characteristics of the pyrrolysine system in the 7th order of methanogens: Implications for the evolution of a genetic code expansion cassette. *Archaea* **2014**, *2014*, 374146. [CrossRef] [PubMed]

89. O'Donoghue, P.; Prat, L.; Kucklick, M.; Schäfer, J.G.; Riedel, K.; Rinehart, J.; Söll, D.; Heinemann, I.U. Reducing the genetic code induces massive rearrangement of the proteome. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 17206–17211. [CrossRef] [PubMed]

90. Banerjee, P.S.; Ostapchuk, P.; Hearing, P.; Carrico, I.S. Unnatural amino acid incorporation onto adenoviral (Ad) coat proteins facilitates chemoselective modification and retargeting of Ad type 5 vectors. *J. Virol.* **2011**, *85*, 7546–7554. [CrossRef] [PubMed]

91. Tack, D.S.; Ellefson, J.W.; Thyer, R.; Wang, B.; Gollihar, J.; Forster, M.T.; Ellington, A.D. Addicting diverse bacteria to a noncanonical amino acid. *Nat. Chem. Biol.* **2016**, *12*, 138–140. [CrossRef] [PubMed]

92. Prat, L.; Heinemann, I.U.; Aerni, H.R.; Rinehart, J.; O'Donoghue, P.; Söll, D. Carbon source-dependent expansion of the genetic code in bacteria. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 21070–21075. [CrossRef] [PubMed]

93. Zhang, W.H.; Otting, G.; Jackson, C.J. Protein engineering with unnatural amino acids. *Curr. Opin. Struct. Biol.* **2013**, *23*, 581–587. [CrossRef] [PubMed]

94. Schmidt, M. Xenobiology: A new form of life as the ultimate biosafety tool. *Bioessays* **2010**, *32*, 322–331. [CrossRef] [PubMed]

95. Yang, Z.; Sismour, A.M.; Sheng, P.; Puskar, N.L.; Benner, S.A. Enzymatic incorporation of a third nucleobase pair. *Nucleic Acids Res.* **2007**, *35*, 4238–4249. [CrossRef] [PubMed]

96. Sismour, A.M.; Lutz, S.; Park, J.-H.; Lutz, M.J.; Boyer, P.L.; Hughes, S.H.; Benner, S.A. PCR amplification of DNA containing non-standard base pairs by variants of reverse transcriptase from Human Immunodeficiency Virus-1. *Nucleic Acids Res.* **2004**, *32*, 728–735. [CrossRef] [PubMed]

97. Yang, Z.; Hutter, D.; Sheng, P.; Sismour, A.M.; Benner, S.A. Artificially expanded genetic information system: A new base pair with an alternative hydrogen bonding pattern. *Nucleic Acids Res.* **2006**, *34*, 6095–6101. [CrossRef] [PubMed]

98. Leconte, A.M.; Hwang, G.T.; Matsuda, S.; Capek, P.; Hari, Y.; Romesberg, F.E. Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet. *J. Am. Chem. Soc.* **2008**, *130*, 2336–2343. [CrossRef] [PubMed]

99. Cong, L.; Ran, F.A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P.D.; Wu, X.; Jiang, W.; Marraffini, L.A.; et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **2013**, *339*, 819–823. [CrossRef] [PubMed]

100. Lajoie, M.J.; Rovner, A.J.; Goodman, D.B.; Aerni, H.-R.; Haimovich, A.D.; Kuznetsov, G.; Mercer, J.A.; Wang, H.H.; Carr, P.A.; Mosberg, J.A.; et al. Genomically recoded organisms expand biological functions. *Science* **2013**, *342*, 357–360. [CrossRef] [PubMed]

101. Hutchison, C.A.; Chuang, R.-Y.; Noskov, V.N.; Assad-Garcia, N.; Deerinck, T.J.; Ellisman, M.H.; Gill, J.; Kannan, K.; Karas, B.J.; Ma, L.; et al. Design and synthesis of a minimal bacterial genome. *Science* **2016**, *351*, aad6253. [CrossRef] [PubMed]