



Review

# Plasticity and Constraints of tRNA Aminoacylation Define Directed Evolution of Aminoacyl-tRNA Synthetases

Ana Crnković <sup>1,\*</sup> , Oscar Vargas-Rodriguez <sup>1</sup> and Dieter Söll <sup>1,2</sup>

<sup>1</sup> Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA; oscar.vargas@yale.edu (O.V.-R.); dieter.soll@yale.edu (D.S.)

<sup>2</sup> Department of Chemistry, Yale University, New Haven, CT 06520, USA

\* Correspondence: ana.crnkovic@yale.edu

Received: 15 April 2019; Accepted: 7 May 2019; Published: 9 May 2019



**Abstract:** Genetic incorporation of noncanonical amino acids (ncAAs) has become a powerful tool to enhance existing functions or introduce new ones into proteins through expanded chemistry. This technology relies on the process of nonsense suppression, which is made possible by directing aminoacyl-tRNA synthetases (aaRSs) to attach an ncAA onto a cognate suppressor tRNA. However, different mechanisms govern aaRS specificity toward its natural amino acid (AA) substrate and hinder the engineering of aaRSs for applications beyond the incorporation of a single L- $\alpha$ -AA. Directed evolution of aaRSs therefore faces two interlinked challenges: the removal of the affinity for cognate AA and improvement of ncAA acylation. Here we review aspects of AA recognition that directly influence the feasibility and success of aaRS engineering toward D- and  $\beta$ -AAs incorporation into proteins *in vivo*. Emerging directed evolution methods are described and evaluated on the basis of aaRS active site plasticity and its inherent constraints.

**Keywords:** aminoacyl-tRNA synthetases; noncanonical amino acids; directed evolution; genetic code expansion; synthetic biology

## 1. Introduction

Synthesis of proteins with exotic building blocks, noncanonical amino acids (ncAAs), can be achieved *in vivo* and *in vitro*. However, the repertoire of ncAAs that can be introduced into peptides and proteins *in vitro* is larger than *in vivo*, due to the fact that the aminoacyl (AA)-tRNA substrates can be generated by Flexizyme, an artificial ribozyme with no ncAA specificity [1]. *In vitro* studies have also revealed a fairly accommodating nature of the ribosome, which can allow introduction of highly diverged ncAAs, with little [2,3] or no modification [4,5]. *In vivo*, however, ncAA-tRNAs have to be synthesized by aminoacyl-tRNA synthetases (aaRSs) which have been naturally optimized during evolution to stringently recognize only one, canonical amino acid [6]. On the other hand, the majority of ncAAs have not been present in the course of natural evolution and tRNA synthetases have not yet evolved to reject them; this fact is exemplified by the apparent flexibility of wild-type synthetases which can, *in vitro*, aminoacylate even *N*-methylated and  $\beta$ -amino acids, as well as  $\alpha$ ,  $\alpha$ -disubstituted amino acids [7]. Supplementation based methods exploit this flexibility of natural aaRSs [8] but 100% incorporation of ncAA may not be feasible using this technology. An ncAA is aminoacylated by a wild-type aaRS and introduced in a residue-specific manner; the matching natural amino acid concentration is kept low by the use of a cAA-auxotrophic strain, while the ncAA is added to the growth medium. However, the target protein must be further mutated in order to achieve site-specific incorporation (e.g., for selenocysteine (Sec) incorporation by cysteinyl-tRNA synthetase

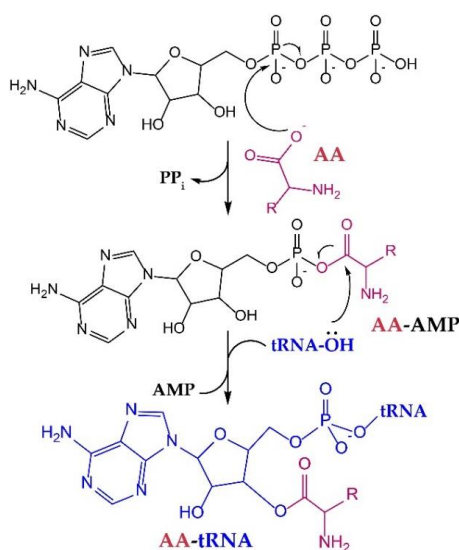
(CysRS) and in response to cysteine (Cys) codons, non-target Cys residues need to be replaced with serine (Ser) [9]; similarly, for incorporation of fluorinated tryptophan (Trp) analogs, non-target Trp residues need to be mutated to tyrosine (Tyr) [10].

Yet, even with the use of an auxotrophic strain, some natural amino acid is still present in the cellular pool and the aaRS will preferentially activate it. This effectively lowers the extent of ncAA insertion. Thus, to generate a target protein decorated with an ncAA at a chosen position, and without additional mutations, one needs to employ the nonsense-suppression method [8]. To that end, an orthogonal aaRS(o-aaRS)-tRNA<sub>CUA</sub> pair needs to be introduced to the cell, and the targeted position in the protein of choice mutated to the amber (TAG) stop codon. An aaRS is orthogonal if it does not react with any of the tRNAs in the host; conversely, an orthogonal tRNA will not be a substrate for the host's aaRSs. The determinants governing aaRS-tRNA<sub>CUA</sub> orthogonality have been reviewed elsewhere in detail [11] and will not be discussed here. With intention to present the current state and advances in directed evolution of aaRSs we will focus on aspects of AA/ncAA recognition and ncAA availability in cellulo.

## 2. Flexibility of Wild-Type Synthetases

### 2.1. Implications of Substrate Recognition and Catalysis for Engineering of aaRSs

All aaRSs catalyze aminoacylation in a two-step reaction, wherein the AA is first “activated” by condensation with ATP and then transferred to the 3′-terminal adenosine of tRNAs in the second step (Figure 1). Mechanistically, aaRSs use binding energy between active-site residues and the substrates to gain specificity and enhance the rate of catalysis [12]. Moreover, although only the last three nucleotides of the tRNA (the conserved 3′-CCA terminus) enter the active site, recognition of the tRNA anticodon, which occurs ~40 Å away from the active site, is essential for proper organization of the active site. Thus, mutations in the anticodon of a tRNA [13] to create a suppressor tRNA do not only reduce its affinity with the aaRS [14], but can also impact AA binding and aminoacylation efficiency [15]. In some cases, aaRS-tRNA complex formation is required for AA or ATP binding [16]. Interestingly, some of these aaRSs retain the capacity to aminoacylate tRNA anticodon variants *in vivo* with no additional mutations to the anticodon-binding domain [17,18]. Therefore, the effect of tRNA mutations on the efficiency of the aminoacylation reaction may vary in severity between different species [18], or can simply be compensated by aaRS overexpression and, possibly, by other unknown mechanisms [17].

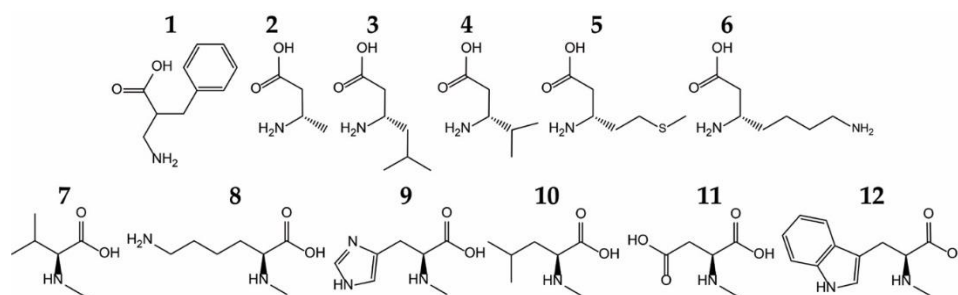


**Figure 1.** Aminoacyl-tRNA synthetases (aaRSs) catalyze aminoacyl (AA)-tRNA formation in two steps. First, the aminoacyl adenylate (AA-AMP) is formed and the pyrophosphate (PP<sub>i</sub>) is released. Next, the oxygen atom from the ribose hydroxyl group of the tRNA attacks the carbonyl carbon of the AA-AMP, forming AA-tRNA with the release of AMP.

It is considered that aaRSs stabilize the transition state of both steps of aminoacylation [19], which, for the purposes of aaRS engineering, may confound the visual inspection of the active site and the roles of individual residues within it. Thus, although structures of aaRS·AA and aaRS·ncAA complexes reveal significant differences in the hydrogen bonding patterns and steric interaction between wild-type aaRS·AA and aaRS·ncAA complexes [20], separating the specific role of active-site residues in initial substrate binding and stabilization of the transition state remains challenging. Sophisticated analysis and redesign of the active site may therefore require >15 residues to be mutagenized around the active site alone, in order to create a superior, ncAA-specific aaRS [21]. Concomitantly, more structural data may be needed to aid the design and discovery of improved, second-generation aaRS variants.

## 2.2. The Inherent AA Polyspecificity of aaRSs

AaRSs recognize the unique side chain of their cognate AA with high specificity, both sterically and through hydrogen bonds and salt bridges. In addition, at least half of the aaRSs also possess a conserved acidic (Asp/Glu) residue that binds the  $\alpha$ -amino group. However, this specificity was developed with regard to other natural AAs. Actually, when tested against different unnatural ncAAs, aaRSs show remarkable flexibility in ncAA recognition, and can aminoacylate these ncAAs *in vitro* [7] and *in vivo* [22] (Figure 2). The most surprising ncAA substrates include some bulky ncAAs, *N*-methylated and  $\beta$ -amino acids, as well as  $\alpha$ ,  $\alpha$ -disubstituted amino acids [7,22,23].



**Figure 2.** Examples of noncanonical amino acids (ncAAs) acylated by wild-type aaRSs *in vitro*. Phenylalanyl-tRNA synthetase (PheRS) acylates a  $\beta_2$ -amino acid, 2-aminomethyl-3-phenyl-propionic acid (1). Five  $\beta_3$ -amino acids (Ala (2), Leu (3), Val (4), Met (5), and Lys (6)), are substrates for their cognate aaRSs. *N*-methylated amino acids that are also substrates for wild-type aaRSs include *N*-methyl-valine (7), *N*-methyl-lysine (8), *N*-methyl-histidine (9), *N*-methyl-leucine (10), *N*-methyl-aspartate (11), and *N*-methyl-tryptophan (12) [7].

## 2.3. Recognition of Fluorinated AA Analogs

Fluorinated AAs have been widely used for protein labeling purposes in *Escherichia coli* auxotrophic strains, through missense suppression [23]. Interestingly, these isosteric ncAAs may be rejected. For instance 3'-trifluorinated isoleucine (Ile) cannot be aminoacylated by *E. coli* isoleucyl-tRNA synthetase (IleRS) [24]. Similarly, hexafluoroisoleucine incorporation *in vivo* requires leucyl-tRNA synthetase (LeuRS) overexpression to overcome impaired aminoacylation [25]. On the other hand, fluorinated histidine (His) is normally aminoacylated by histidyl-tRNA synthetase (HisRS) [26]. Fluorinated Trp analogs are readily accepted by bacterial tryptophanyl-tRNA synthetases (TrpRSs) [27,28], as well as Tyr derivatives [22,29]. Both 4-fluorinated and 4,4-difluorinated glutamate (Glu) can be aminoacylated *in vitro* and introduced into proteins *in vivo* [7,30].

## 2.4. Improving ncAA-tRNA Synthesis Using AA-tRNA Deacylases

Almost half of the aaRSs mis-aminoacylate other natural AAs. They also possess a mechanism to clear (edit) erroneously formed aminoacyl adenylate or mis-aminoacylated tRNA [16]. The latter reaction is catalyzed by a separate “editing” domain. The hydrolytic site of the editing domain can catalyze deacylation of noncognate natural AAs, as well as some ncAAs. For instance, while

trifluoroethylglycine can be acylated by *E. coli* IleRS without any modification to the synthetic active site; it does, however, require a mutation in the editing domain that prevents hydrolysis of the trifluoroethylglycyl-tRNA [31]. Similarly, experiments with an editing-deficient variant of valyl-tRNA synthetase revealed that the enzyme is capable of acylating >10 different ncAAs including  $\alpha,\alpha$ -disubstituted and cyclic  $\beta$ -amino acids [32]. Therefore, accurately identifying the activities of these editing domains in aaRS reaction with an ncAA may be critical for the ncAA incorporation into proteins.

Editing domains have been exploited to increase the purity of ncAA-tRNA synthesis in vivo and in vitro. For example, the editing domain of phenylalanyl-tRNA synthetase (PheRS), which naturally deacylates Tyr-tRNA<sup>Phe</sup> [33], has been fused to an *E. coli* TyrRS variant designed to acylate 3-iodo-L-tyrosine. The appended domain allows the mutant enzyme to clear the residual affinity for Tyr by specific hydrolysis of Tyr-tRNA<sup>Tyr</sup>. Thus, without additional mutations to the synthetic site, the accurate synthesis of 3-iodo-Tyr-tRNA was increased [34]. Interestingly, this editing domain of archaeal PheRS also shows stereospecificity, and hydrolyzes L-Tyr-tRNA, but not D-Tyr-tRNA. A TyrRS-PheRS chimera has been shown to possess increased enantioselectivity toward D-Tyr-tRNA<sup>Tyr</sup> [35]. A similar approach was used to design a human mitochondrial PheRS variant that is “resistant” to misacylation of the reactive oxygen species (ROS) derived from Phe (such as Tyr and *m*-Tyr). The mitochondrial version inherently lacks the editing domain but discriminates between Tyr and Phe with a factor of 10<sup>3</sup>. Human cytosolic PheRS possesses an editing domain, but it cannot clear *m*-tyrosyl-tRNA<sup>Phe</sup>. By-fusing the editing domain of *E. coli* PheRS a variant in which Tyr-tRNA<sup>Phe</sup> and *m*-Tyr-tRNA<sup>Phe</sup> can be enzymatically deacylated was created [36].

### 2.5. Evolvability of 20 aaRSs

Although more than 150 ncAAs have been adapted for in vivo incorporation to date [23], the vast majority are activated by only two orthogonal aaRSs: archaeal TyrRS and pyrrolysyl-tRNA synthetase (PylRS). Both synthetases encompass some of the critical features that a synthetic-biology ready aaRS needs: good orthogonality [37,38] and a flexible active site [39]. Another useful feature of PylRS is that it does not recognize the tRNA anticodon [37,40]. Both enzymes possess a spacious active site that can be mutagenized with relative ease in order to facilitate incorporation of bulky ncAAs. PylRS even has a flexible mode of  $\alpha$ -amino group recognition, thereby allowing activation of  $\beta$ -substituted amino acids [41]. TyrRS active site has been proven extremely flexible, possibly due to similarities with TrpRS [42].

Are aaRSs in general amenable to laboratory design and directed evolution? If we consider that “A good parent protein for directed evolution . . . exhibits enough of the desired function that small improvements (expected from a single mutation) can be reliably discerned in a high-throughput screen” [43], then the answer is yes, especially since many wild-type aaRSs can acylate divergent amino acid substrates in vitro. Also, the majority of evolved aaRS variants contain only a few mutations that improve their activity for a desired ncAA. The problem may lie more in the nature of assays/reporters employed to screen for superior variants. So far, aaRSs have been evolved using reporters/assays that tie the success of ncAA incorporation into proteins to the success of ncAA aminoacylation. Due to such setup we may have failed to identify aaRS variants with improved affinity toward such exotic AAs that are rejected by other components of the translation apparatus (e.g., elongation factor EF-Tu/EF1a or the ribosome), rather than the aaRS itself. Furthermore, because directed evolution of aaRSs has been almost ubiquitously employed in vivo, we may have overlooked prospective aaRS candidates with starting ncAA affinities that cannot be achieved intracellularly, or those that cannot be tested in vivo, due to their lack of orthogonality.

## 2.6. Availability of ncAA In Vivo May Dictate ncAA Incorporation

Affinity of aaRSs for their cognate amino acid is typically in a micromolar range, but can vary between different aaRSs up to 100-fold. In organisms where the absolute concentration of an amino acid has been measured, and the corresponding aaRS biochemically characterized, we can observe that the amounts of free AA are several fold higher than the  $K_M$  (AA) for this synthetase. Affinity toward cognate AA is frequently investigated as a part of general kinetic studies and Michaelis constants for AA substrates are readily available for all 21 aaRSs. Absolute concentrations of amino acids in a glucose-fed, exponentially growing *E. coli* show that aaRSs affinity matches the intracellular concentration of its cognate AA and that individual AA concentrations exceed the  $K_M$  (AA) of a given aaRS [44]. For instance, *E. coli* TyrRS catalyzes Tyr-AMP formation with a  $K_M$  (Tyr) of  $7.13 \pm 0.81 \mu\text{M}$  [45]. On the other hand, *E. coli* AlaRS catalyzes Ala-AMP formation with almost 100 times lower affinity ( $K_M$  (Ala)  $300 \pm 120 \mu\text{M}$ ). Although these values show a wide range of affinities toward cognate AA, they are consistent with intracellular AA concentration, where Ala is 100 times more abundant than Tyr (2.6 mM vs.  $29 \mu\text{M}$ , respectively [44]).

Similarly, for the purpose of directed evolution in vivo, the intracellular concentration of an ncAA needs to match the aaRS initial affinity. The initial prerequisites are therefore intertwined: the intracellular supply of an ncAA must compensate for the aaRS affinity towards it. Wild-type *E. coli* TyrRS has a  $K_M$  (azatyrosine) of  $17.7 \pm 2.9 \mu\text{M}$ , which is only around five times lower in affinity compared to cognate Tyr; by adding  $50 \mu\text{M}$  azatyrosine to the growth medium, a significant fraction of the cellular proteins will be labeled with this ncAA [46]. On the other hand, the archaeal *O*-phosphoserine-tRNA synthetase (SepRS) accepts its cognate AA phosphoserine (Sep) with much lower affinity ( $K_M$  (Sep) of  $270 \pm 40 \mu\text{M}$  for Sep-AMP formation) [47]. At the same time, intracellular levels of Sep in wild-type *E. coli* grown in rich medium are below the detection limit; miniscule amounts of intracellular Sep are observed only when 5 mM Sep is added to the culture [48].

Sufficient intracellular levels of ncAA can be achieved by supplementation of such ncAAs to the growth media. In that case, the ncAA is imported via canonical transport systems [49]. In cases where the physicochemical properties of the ncAA prevent its recognition by the carrier proteins or its diffusion across the membrane, the ncAA may be supplied covalently linked to a natural AA and introduced to the cell through portage transport [50]. Binding proteins of *E. coli* oligo- and dipeptide permease systems show broad substrate specificity and are able to accept a wide range of peptide substrates [51]. This promiscuity of peptide import has been exploited to transport *O*-phosphotyrosine and its derivatives into *E. coli*, thereby allowing identification of an aaRS capable of acylating *O*-phosphotyrosine [52]. Interestingly, the binding protein of the OppABCDF permease system has been shown to recognize the backbone of the oligopeptide, therefore limiting the spectrum of ncAA that could be introduced to the cytosol in this manner [53]. Recently, an interesting approach to circumvent this issue was reported [54]. This strategy combines the promiscuity of peptide import with the activity of a cytosolic enzyme,  $\gamma$ -glutamyl transferase. The ncAA is attached to the  $\gamma$ -carboxyl group of glutamate residue of an oligopeptide. Upon import, the ncAA is liberated through the action of cytosolic  $\gamma$ -glutamyl transferase. This approach may be useful for the import of ncAAs with alternative backbones [54].

Alternatively, some ncAAs can be produced intracellularly by manipulation of the host genome. For instance, increased intracellular levels of Sep can be achieved by the deletion of phosphoserine phosphatase gene, both in *E. coli* [48,55] and HEK293 cells [56]. A more complex strategy needs to be employed for ncAAs that are not normally part of the cellular metabolome. This may require introduction of complete biosynthetic operons (as with pyrrolysine (Pyl) [57]) or unique genes from another organism. For example, to generate sufficient amounts of *O*-phosphothreonine (pThr) in *E. coli* it was necessary to introduce the threonine kinase gene from *Salmonella enterica*, which catalyzes pThr formation using threonine (Thr) and ATP as substrates [58]. High expression levels of this enzyme in *E. coli* is necessary to achieve intracellular concentrations of up to 1.7 mM [59]. Although SepRS shows



lower affinity for pThr than 1.7 mM ( $K_M$  (pThr) in aminoacylation  $8.4 \pm 4.0$  mM) [47]; these intracellular levels were sufficient to select for a variant capable of introducing pThr into proteins in *E. coli* [59].

How well can we predict intracellular concentration of a desired ncAA? Initial tests to explore the substrate range of an aaRS include selection/screening outcomes in the presence and absence of an ncAA. However, improved synthesis of a reporter in the presence of the ncAA may not necessarily signal ncAA acylation unless the reporter is sensitive/specific to ncAA insertion. Some ncAAs may be metabolically active leading to a decrease of their intracellular level and/or decreased survival due to the toxicity of their byproducts [60]. The function of ncAAs as allosteric inhibitors in cellular biosynthetic pathways for the similar natural AAs may be conceivable, although it is fairly unexplored [61]. The choice of growth media may also influence correct identification of ncAA substrates. For instance, phosphorylated ncAAs like Sep and pThr enter *E. coli* via a complex phosphonate transporter whose expression is induced by phosphate limitation [48,62]. In cases where a satisfactory intracellular concentration of ncAA cannot be achieved, ancestral functions may become prominent and lead to aaRS variants capable of noncognate, natural AA acylation [63].

### 3. Directed Evolution of Orthogonal tRNA Synthetases

Directed evolution of aaRSs is typically conducted in vivo and for that reason the principal requirement is the aaRS and tRNA orthogonality, i.e., the absence of cross-talk with endogenous aaRSs and tRNAs. Generally, the orthogonal tRNA is set to decode the amber (UAG) stop codon by mutating its anticodon to CUA. Successful acylation of the orthogonal tRNA<sub>CUA</sub> is monitored by UAG read-through, where the reporter gene contains an in-frame UAG codon.

As with other enzymes, the success of directed evolution rests on the quality/diversity of aaRS libraries and an unambiguous phenotype that allows rapid and accurate detection of proficient aaRS variants. Commonly used reporters for identification of aaRS candidates, chloramphenicol resistance [64] and GFP fluorescence [65], have been adapted for the directed evolution of aaRSs.

#### 3.1. Sequence Randomization Methods

A good, rationally designed aaRS library requires accurate identification of active site residues that participate in substrate binding and/or catalysis (see above). These residues are the most frequent targets for saturation mutagenesis where every possible AA substitution is tested simultaneously. Advances in primer design, most notably in removing codon redundancy that exists with NNK randomization, facilitate screening and allow a more balanced distribution of individual amino acid replacements in the library [66]. However, even with the utilization of non-redundant primers it is not feasible to target more than seven residues ( $\sim 1.3 \times 10^9$  library members) with conventional saturation-mutagenesis approaches. Advanced computational design using Rosetta3 [67] enabled systematic investigation of variants in which 17 residues were targeted but not all substitutions were tested [15]; this approach led to the discovery of TyrRS variant with high specificity for photocaged 3,4-dihydroxyphenyl-alanine (DOPA).

Significant advances have been made in creating aaRS libraries in situ [68–70]. These approaches circumvent the challenge of creating plasmid-borne libraries: the demand for high transformation efficiency required for introducing large populations of mutant clones. Multiplex automated genome engineering, MAGE [71], uses a previously established technique that exploits the fact that externally introduced  $\lambda$ -phage protein (protein  $\beta$ ) binds and anneals single stranded DNA oligonucleotides to a complementary DNA strand in the vicinity of a DNA replication fork [72]. By transforming oligonucleotides in which the target residue has been randomized, a particular part of the bacterial chromosome can be mutated in a focused fashion. This technique was employed to evolve a genome-borne TyrRS variant able to facilitate introduction of an ncAA in up to 30 individual positions of a reporter protein [68]. One potential disadvantage of the technique are off-target incorporations of mutagenic oligonucleotides. Because several cycles of mutagenesis are needed to create a fully randomized library [68], the resulting strain may have decreased reproductive strength. Nonetheless,

the strains are sufficiently healthy to survive the screening process [68,71], crucial to identification of the aaRS variant. One important advantage of the method is that genomic placement of the aaRS variant unmasks subtle differences between gain-of-function phenotypes which are often present with plasmid-borne libraries.

As with any focused-mutagenesis method, the quality of the emerging variants is only as good as the initial design of the library. In other words, if a critical residue is not targeted for mutagenesis, superior variants will not be present in the library. Residues that are essential for stabilizing the transition state, or that participate in long-range signaling between the anticodon-binding domain and the active site may not be apparent from the crystal structure. In addition, rational design does not include improvements in aaRS expression and folding in the host.

Although conventional methods to create randomly mutated aaRS sequences via error-prone PCR have been employed with success [73,74], controlled random mutagenesis *in vivo* allows a less labor-intensive library creation and direct coupling between mutational and selection cycles [75]. Convenient constructs for achieving tunable mutation rates *in vivo* are the MP-plasmids that contain various mutator genes placed under the control of an arabinose promoter [75]. These constructs have been employed in combination with phage-assisted continuous evolution technique (PACE [76]) to rapidly identify PylRS variants with improved kinetic parameters.

### 3.2. Selection Methods

Selection methods employed for aaRS evolution traditionally link ncAA insertion into a protein with cell survival. Typically, the chloramphenicol acetyltransferase (CAT) gene is used and a stop codon is placed at position 112; aaRS variants that facilitate stop-codon read-through are identified by growth on chloramphenicol, and in the presence of the ncAA. Because any AA is tolerated at position 112, this positive selection step must inevitably be combined with a negative selection. Toxic gene products such as barnase [77], CcdB toxin [78] or TolC [68] are used to eliminate aaRS variants that facilitate read-through using a natural AA (i.e., false positives). A recent negative selection protocol was developed where a toxic variant of *E. coli* PheRS is expressed by nonspecific stop-codon read-through; this variant acylates tRNA<sup>Phe</sup> with *p*-Cl-phenylalanine and the incorporation of this amino acid mimic leads to proteome-wide incorporation and ultimately cell death [79]. One significant advantage of this mechanism is the ability to fine-tune the stringency of the negative selection step through changing the concentration of *p*-Cl-phenylalanine in the medium.

Bacteriophage protein pIII needed for phage infectivity was used as a reporter to evolve PylRS using PACE, and the related, non-continuous version of the method [40,69]. Gene III variants with up to three TAG codons were used to link ncAA acylating activity of PylRS with successful phage propagation. Importantly, gene III, normally present in the phage genome, is replaced with the aaRS gene; gene III (with one, two or three TAG codons) is supplied on a plasmid. Alternatively, a less stringent selection can be used where gene III expression is driven from the T7-promoter and the corresponding T7 RNA polymerase is added to the system containing a minimum of 2 TAG in-frame codons. For the purposes of improving the ncAA specificity of a polyspecific *Methanocaldococcus jannaschii* TyrRS variant initially designed to acylate *p*-nitro-*L*-phenylalanine (*p*-NF), a dominant-negative variant of pIII (pIII-neg) was used to switch its affinity to *p*-iodo-*L*-phenylalanine and eliminate existing affinity to *p*-NF [69]. The principle of pIII-neg action is to reduce the infectivity of the developing phage, resulting in reduced propagation of pIII-neg synthesizing phages and subsequent removal of the misacylating clone from the lagoon.

### 3.3. Screening Methods

Colorimetric [80], fluorescent [17,68,73,81] and luminescent [69,82] reporters have been adapted for medium/high-throughput screens of ncAA-acylating aaRS libraries. In all cases the reporter gene is translated with in-frame UAG codons and the cells with ncAA-acylating aaRS variants are collected based on the observable color, fluorescence or luminescence. High-throughput screening by

fluorescence-activated cell sorting (FACS) allows cross-examination of large populations of cells in a reasonable amount of time [83]. Through alternating rounds of sorting (where highly fluorescent cells are collected when growing in the ncAA-supplemented media, and weakly fluorescent ones are collected when ncAA is absent) aaRS library members with the best activity and/or specificity can be identified [65]. To decrease the risk of collecting of false-positives and improve plasmid maintenance GFP-mediated screening benefits from using one-plasmid systems (GFP, tRNA<sub>CUA</sub> and aaRS variant all encoded from the same vector) [73,84]. Similarly, genomic integration of the aaRS-tRNA<sub>CUA</sub> pair leads to a decrease in GFP-reporter yields but unmasks differences in GFP expression that are arising from fluctuating expression levels of aaRS variant [68].

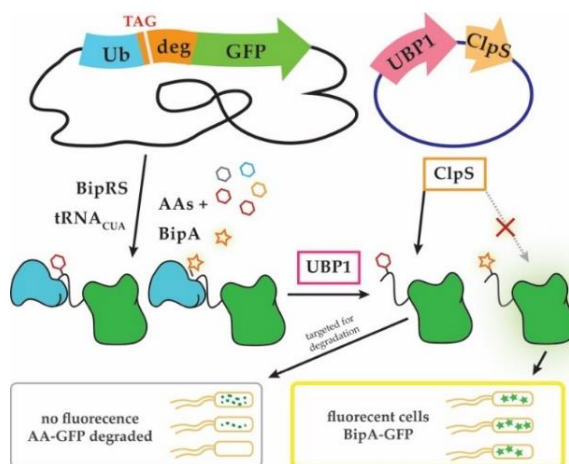
### 3.4. ncAA-Specific Reporters

Most engineered aaRSs are polyspecific [85–87]. This means that the variant retains the capacity to acylate the original, natural AA [88] or that it can discriminate among similar ncAAs [68,74,85,89]. While negative-selection markers can be optimized against particular natural amino acids [90], ncAA-dependent positive-selection reporters remain largely unexplored. In addition to facilitating the discovery of aaRS variants with increased specificity, ncAA-dependent screens and selection assays allow evolution of aaRS-tRNA pairs with partial orthogonality by eliminating the necessity of the negative selection step.

**Selenocysteine (Sec)-dependent reporters:** In contrast to other reporters of ncAA insertion, screen/selection methods to improve EF-Tu-dependent Sec insertion [91] have always been specific for this ncAA. This is due to the fact that Sec is synthesized in a two-step manner, where tRNA<sup>Sec</sup> variant is first misacylated with serine (by seryl-tRNA synthetase, SerRS) and then converted to Sec-tRNA<sup>Sec</sup> (by selenocysteine synthase SelA) [92]. Because both Ser-tRNA<sub>CUA</sub> and Sec-tRNA<sub>CUA</sub> can be accepted by EF-Tu and the ribosome, the reporter protein must be sensitive to Ser vs. Sec insertion. A number of Cys-dependent enzymes have been repurposed for Sec-specific selections due to similarities in chemical properties of Cys and Sec [93–95]. However, these reporters are yet to be employed to enhance SerRS activity with tRNA<sup>Sec</sup> [96].

**Unnatural ncAA-dependent reporters:** A recently reported method that allows selective discovery of ncAA-acylating TyrRS variants showed that the activity of protein degradation machinery can be harnessed to specifically identify ncAA insertion in reporter proteins (Figure 3, [89]). The method exploits the N-end rule of protein degradation, which defines the half-life of a protein by the identity of its N-terminal amino acid [97]. The method was used to improve the specificity of the TyrRS variant (BipRS) that acylates biphenylalanine (BipA), and to lesser extent Tyr and several Tyr analogs. The reporter fusion protein contains a cleavable ubiquitin domain on the N-terminus, a UAG codon and a conditionally strong N-degron in the middle, followed by a GFP reporter on the C-terminus. Fluorescence is initially observed nonspecifically with UAG-read-through. However, upon expression of a yeast ubiquitin cleavase protein (UBP1) the fusion protein is cleaved exposing the AA inserted in response to UAG. Because the AA/ncAA now occupies the first position on the N-terminus, it can be recognized by the adaptor protein ClpS, which preferentially binds destabilizing Tyr residues and directs them for degradation. Thus, cells containing aaRS variants that enable BipA insertion will retain fluorescence in UBP1 presence, while those that preferentially facilitate Tyr incorporation will not. Importantly, the most destabilizing residues Tyr, Phe and Leu are also the ones most frequently misincorporated by engineered TyrRS-, PylRS- and LeuRS variants. Furthermore, wild-type ClpS and its variants were shown to have different affinities for chemically similar ncAAs which could be useful to rapidly engineer aaRS variants with unique ncAA specificity [74,89].





**Figure 3.** “Posttranslational proofreading” allows monitoring site-specific ncAA incorporation [89]. A ubiquitin-GFP fusion reporter is encoded from the genome, while the necessary enzymes UBP1 and ClpS are overexpressed from a plasmid. The fusion reporter is synthesized via nonsense suppression where different BipRS variants acylate either biphenylalanine (BipA) (star) or a natural AA (hexagons). Ubiquitin-GFP fusions are substrates for UBP1, which removes ubiquitin (Ub) and exposes N-terminally incorporated BipA or natural AA. Truncated fusion proteins that contain destabilizing tyrosine (Tyr), Phe or Leu residues in place of the amber (TAG) stop codon are directed for degradation by ClpS. In contrast, GFP fusions containing BipA are stable and not degraded (see text for details).

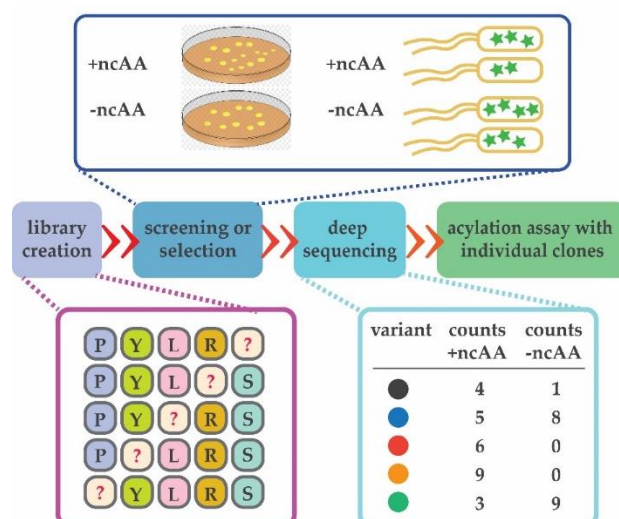
### 3.5. Adaptation of Next-Generation Sequencing (NGS) to Sidestep a Negative Selection

Next-generation sequencing (NGS) became invaluable for directed evolution of enzymes as it allows rapid inspection of tens of thousands of protein variants [98]. In other words, the entire population of a given library can be observed before, during or after the process of directed evolution (Figure 4, [99]).

One remarkable feature of NGS-based screening in directed evolution is the ability to remove the necessity of negative selection. As mentioned earlier, reporters used in positive selection of aaRS variant mostly tolerate various AAs/ncAAs in stop codon read-through; the engineered specificity of an aaRS variant therefore depends on multiple rounds of positive and negative selections/screens in order to sift through a population containing false positives or promiscuous variants.

The possibility to remove the negative selection step may be critical for aaRS-tRNA pairs where the tRNA orthogonality is not complete, or where the starting affinity for the ncAA is very low. Orthogonality is not an absolute value; the orthogonal tRNA may be weakly recognized by some aaRS in the host, a tendency that may not be important in the presence of cognate aaRS. However, due to variation in copy numbers of plasmids from which the tRNA and aaRS are encoded, or different strength of promoters, the levels of orthogonal tRNA and aaRS may not be matched. Furthermore, if the ncAA acylation is very inefficient, the minor fraction of the orthogonal tRNA that is acylated by the host’s aaRS may become prominent. In these cases negative selection may be detrimental for orthogonal aaRS variants that have gained some improvement in ncAA acylation, but this activity is still less effective than misacylation by the host’s aaRS.

A report of special interest where NGS was employed to identify aaRS variants with improved acylation activities demonstrated that a single round of stop-codon read-through with CAT, done in parallel (with and without ncAA) allows identification of library members with improved activity towards the ncAA. Comparison between the populations growing at high chloramphenicol levels with and without the ncAA allows precise identification of the ncAA-specific library member [59].



**Figure 4.** Deep mutational scanning is invaluable for the entire process of directed evolution. It can be used to inspect the quality of the initial aaRS library and quickly identify variants selective for the ncAA. Deep sequencing may be used to ensure even distribution of generated variants after mutagenesis (in purple). After screening or selection is executed in the presence and absence of an ncAA (dark blue rectangle), entire populations of cells are collected and DNA isolated. Individual variants are counted in the populations collected from both selections (light blue rectangle). Variants observed in the populations selected in the presence of ncAA can be then directly tested in an aminoacylation assay in vitro (green). Randomized positions are depicted with “?”.

#### 4. Perspectives

Translation systems to introduce ncAAs into proteins in vivo may cause different levels of cellular toxicity. While it is possible to develop customized strains that have evolved mechanisms to help them reduce this toxicity [81], for the purposes of comprehensive aaRS engineering a more forgiving platform would be desired. In addition to the inherent lack of toxic effects, in vitro evolution of proteins allows much larger focused libraries to be tested, because the library is not limited by the transformation efficiency. Furthermore, the orthogonality requirement can be manipulated, because the canonical aaRS-tRNA pairs can be added in desired quantities and from desired sources. Finally, the ncAA concentration can be controlled far better than with in vivo systems. To date, in vitro evolution of aaRSs remains rare [100]. It is, however, foreseeable that this method might be employed more in the future. Combination of the directed evolution in vitro with ncAA-specific reporter assays in vivo may facilitate de novo protein design and the discovery of novel aaRS variants that do not resemble aaRSs found in nature.

**Author Contributions:** Conceptualization, A.C., O.V.-R., and D.S.; Writing—Original Draft Preparation, A.C., O.V.-R., and D.S.; Writing—Review and Editing, A.C., O.V.-R., and D.S.; Visualization, A.C., and O.V.-R.; Supervision, D.S.; Funding Acquisition, D.S.

**Funding:** Research in the Söll laboratory is supported by grants from the US National Institutes of Health (NIH) (R35GM122560) and from the Division of Chemical Sciences, Geosciences and Biosciences, Office of Basic Energy Sciences of the Department of Energy (DE-FG02-98ER20311) to D.S.

**Acknowledgments:** We thank members of the Söll laboratory for inspiring discussions during the preparation of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

aaRS	Aminoacyl-tRNA synthetase
AA	Amino acid
ncAA	Noncanonical amino acid
PACE	Phage-assisted continuous evolution technique
PANCE	Phage-assisted noncontinuous evolution technique
MAGE	Multiplex automated genome engineering
NGS	Next-generation sequencing
GFP	Green fluorescent protein

## References

1. Ramaswamy, K.; Saito, H.; Murakami, H.; Shiba, K.; Suga, H. Designer ribozymes: programming the tRNA specificity into flexizyme. *J. Am. Chem. Soc.* **2004**, *126*, 11454–11455. [[CrossRef](#)] [[PubMed](#)]
2. Chen, S.; Ji, X.; Gao, M.; Dedkova, L.M.; Hecht, S.M. *In cellulo* synthesis of proteins containing a fluorescent oxazole amino acid. *J. Am. Chem. Soc.* **2019**. [[CrossRef](#)]
3. Melo Czekster, C.; Robertson, W.E.; Walker, A.S.; Söll, D.; Schepartz, A. In vivo biosynthesis of a  $\beta$ -amino acid-containing protein. *J. Am. Chem. Soc.* **2016**, *138*, 5194–5197. [[CrossRef](#)]
4. Katoh, T.; Iwane, Y.; Suga, H. Logical engineering of D-arm and T-stem of tRNA that enhances D-amino acid incorporation. *Nucleic Acids Res.* **2017**, *45*, 12601–12610. [[CrossRef](#)] [[PubMed](#)]
5. Katoh, T.; Suga, H. Ribosomal incorporation of consecutive  $\beta$ -amino acids. *J. Am. Chem. Soc.* **2018**, *140*, 12159–12167. [[CrossRef](#)]
6. Woese, C.R.; Olsen, G.J.; Ibba, M.; Söll, D. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 202–236. [[CrossRef](#)] [[PubMed](#)]
7. Hartman, M.C.; Josephson, K.; Szostak, J.W. Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4356–4361. [[CrossRef](#)]
8. Hoesl, M.G.; Budisa, N. Recent advances in genetic code engineering in *Escherichia coli*. *Curr. Opin. Biotech.* **2012**, *23*, 751–757. [[CrossRef](#)] [[PubMed](#)]
9. Mueller, S.; Senn, H.; Gsell, B.; Vetter, W.; Baron, C.; Böck, A. The formation of diselenide bridges in proteins by incorporation of selenocysteine residues: Biosynthesis and characterization of (Se)<sub>2</sub>-thioredoxin. *Biochemistry* **2002**, *33*, 3404–3412. [[CrossRef](#)]
10. Campos-Olivas, R.; Aziz, R.; Helms, G.L.; Evans, J.N.; Gronenborn, A.M. Placement of <sup>19</sup>F into the center of GB1: effects on structure and stability. *FEBS Lett.* **2002**, *517*, 55–60. [[CrossRef](#)]
11. Reynolds, N.M.; Vargas-Rodriguez, O.; Söll, D.; Crnkovic, A. The central role of tRNA in genetic code expansion. *BBA-Gen. Subjects* **2017**, *1861*, 3001–3008. [[CrossRef](#)]
12. Winter, G.; Fersht, A.R.; Wilkinson, A.J.; Zoller, M.; Smith, M. Redesigning enzyme structure by site-directed mutagenesis: Tyrosyl tRNA synthetase and ATP binding. *Nature* **1982**, *299*, 756–758. [[CrossRef](#)]
13. Giegé, R.; Sissler, M.; Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* **1998**, *26*, 5017–5035. [[CrossRef](#)]
14. Rogerson, D.T.; Sachdeva, A.; Wang, K.; Haq, T.; Kazlauskaitė, A.; Hancock, S.M.; Huguenin-Dezot, N.; Muqit, M.M.; Fry, A.M.; Bayliss, R.; et al. Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat. Chem. Biol.* **2015**, *11*, 496–503. [[CrossRef](#)]
15. Rauch, B.J.; Porter, J.J.; Mehl, R.A.; Perona, J.J. Improved incorporation of noncanonical amino acids by an engineered tRNA<sup>Tyr</sup> suppressor. *Biochemistry* **2016**, *55*, 618–628. [[CrossRef](#)]
16. Perona, J.J.; Gruic-Sovulj, I. Synthetic and editing mechanisms of aminoacyl-tRNA synthetases. *Top. Curr. Chem.* **2014**, *344*, 1–41. [[CrossRef](#)]
17. Santoro, S.W.; Anderson, J.C.; Lakshman, V.; Schultz, P.G. An archaeobacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in *Escherichia coli*. *Nucleic Acids Res.* **2003**, *31*, 6700–6709. [[CrossRef](#)]
18. Liu, D.R.; Schultz, P.G. Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4780–4785. [[CrossRef](#)]

19. Avis, J.M.; Day, A.G.; Garcia, G.A.; Fersht, A.R. Reaction of modified and unmodified tRNA<sup>Tyr</sup> substrates with tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). *Biochemistry* **1993**, *32*, 5312–5320. [[CrossRef](#)] [[PubMed](#)]
20. Zhang, Y.; Wang, L.; Schultz, P.G.; Wilson, I.A. Crystal structures of apo wild-type *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for *O*-methyl-L-tyrosine. *Protein Sci.* **2005**, *14*, 1340–1349. [[CrossRef](#)] [[PubMed](#)]
21. Hauf, M.; Richter, F.; Schneider, T.; Faidt, T.; Martins, B.M.; Baumann, T.; Durkin, P.; Dobbek, H.; Jacobs, K.; Möglich, A.; et al. Photoactivatable mussel-based underwater adhesive proteins by an expanded genetic code. *ChemBioChem* **2017**, *18*, 1819–1823. [[CrossRef](#)] [[PubMed](#)]
22. Fan, C.; Ho, J.M.L.; Chirathivat, N.; Söll, D.; Wang, Y.S. Exploring the substrate range of wild-type aminoacyl-tRNA synthetases. *ChemBioChem* **2014**, *15*, 1805–1809. [[CrossRef](#)]
23. 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](#)]
24. Wang, P.; Tang, Y.; Tirrell, D.A. Incorporation of trifluoroisoleucine into proteins in vivo. *J. Am. Chem. Soc.* **2003**, *125*, 6900–6906. [[CrossRef](#)] [[PubMed](#)]
25. Tang, Y.; Tirrell, D.A. Biosynthesis of a highly stable coiled-coil protein containing hexafluoroisoleucine in an engineered bacterial host. *J. Am. Chem. Soc.* **2001**, *123*, 11089–11090. [[CrossRef](#)]
26. Eichler, J.F.; Cramer, J.C.; Kirk, K.L.; Bann, J.G. Biosynthetic incorporation of fluorohistidine into proteins in *E. coli*: a new probe of macromolecular structure. *ChemBioChem* **2005**, *6*, 2170–2173. [[CrossRef](#)]
27. 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](#)]
28. Zhang, Q.S.; Shen, L.; Wang, E.D.; Wang, Y.L. Biosynthesis and characterization of 4-fluorotryptophan-labeled *Escherichia coli* arginyl-tRNA synthetase. *J. Protein Chem.* **1999**, *18*, 187–192. [[CrossRef](#)]
29. Gao, R.; Zhang, Y.; Dedkova, L.; Choudhury, A.K.; Rahier, N.J.; Hecht, S.M. Effects of modification of the active site tyrosine of human DNA topoisomerase I. *Biochemistry* **2006**, *45*, 8402–8410. [[CrossRef](#)]
30. Hartman, M.C.; 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](#)]
31. Völler, J.S.; Dulic, M.; Gerling-Driessen, U.I.; Biava, H.; Baumann, T.; Budisa, N.; Gruic-Sovulj, I.; Koks, B. Discovery and investigation of natural editing function against artificial amino acids in protein translation. *ACS Cent. Sci.* **2017**, *3*, 73–80. [[CrossRef](#)] [[PubMed](#)]
32. Iqbal, E.S.; Dods, K.K.; Hartman, M.C.T. Ribosomal incorporation of backbone modified amino acids via an editing-deficient aminoacyl-tRNA synthetase. *Org. Biomol. Chem.* **2018**, *16*, 1073–1078. [[CrossRef](#)]
33. Roy, H.; Ling, J.; Irnov, M.; Ibba, M. Post-transfer editing in vitro and in vivo by the  $\beta$  subunit of phenylalanyl-tRNA synthetase. *EMBO J.* **2004**, *23*, 4639–4648. [[CrossRef](#)]
34. Oki, K.; Sakamoto, K.; Kobayashi, T.; Sasaki, H.M.; Yokoyama, S. Transplantation of a tyrosine editing domain into a tyrosyl-tRNA synthetase variant enhances its specificity for a tyrosine analog. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13298–13303. [[CrossRef](#)]
35. Richardson, C.J.; First, E.A. Altering the enantioselectivity of tyrosyl-tRNA synthetase by insertion of a stereospecific editing domain. *Biochemistry* **2016**, *55*, 1541–1553. [[CrossRef](#)]
36. Kartvelishvili, E.; Peretz, M.; Tworowski, D.; Moor, N.; Safro, M. Chimeric human mitochondrial PheRS exhibits editing activity to discriminate nonprotein amino acids. *Protein Sci.* **2016**, *25*, 618–626. [[CrossRef](#)] [[PubMed](#)]
37. Nozawa, K.; O'Donoghue, P.; Gundllapalli, S.; Araiso, Y.; Ishitani, R.; Umehara, T.; Söll, D.; Nureki, O. Pyrrolysyl-tRNA synthetase-tRNA<sup>Py1</sup> structure reveals the molecular basis of orthogonality. *Nature* **2008**, *457*, 1163–1167. [[CrossRef](#)] [[PubMed](#)]
38. Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Mol. Biol.* **2003**, *10*, 425–432. [[CrossRef](#)] [[PubMed](#)]
39. Yanagisawa, T.; Umehara, T.; Sakamoto, K.; Yokoyama, S. Expanded genetic code technologies for incorporating modified lysine at multiple sites. *ChemBioChem* **2014**, *15*, 2181–2187. [[CrossRef](#)]

40. Suzuki, T.; Miller, C.; Guo, L.T.; Ho, J.M.L.; Bryson, D.I.; Wang, Y.S.; Liu, D.R.; Söll, D. Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.* **2017**, *13*, 1261–1266. [[CrossRef](#)]
41. Kobayashi, T.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. Recognition of non- $\alpha$ -amino substrates by pyrrolysyl-tRNA synthetase. *J. Mol. Biol.* **2009**, *385*, 1352–1360. [[CrossRef](#)]
42. Doublé, S.; Bricogne, G.; Gilmore, C.; Carter, C.W., Jr. Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase. *Structure* **1995**, *3*, 17–31. [[CrossRef](#)]
43. Romero, P.A.; Arnold, F.H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 866–876. [[CrossRef](#)]
44. Bennett, B.D.; Kimball, E.H.; Gao, M.; Osterhout, R.; van Dien, S.J.; Rabinowitz, J.D. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **2009**, *5*, 593–599. [[CrossRef](#)]
45. Venkat, S.; Gregory, C.; Gan, Q.; Fan, C. Biochemical characterization of the lysine acetylation of tyrosyl-tRNA synthetase in *Escherichia coli*. *ChemBioChem* **2017**, *18*, 1928–1934. [[CrossRef](#)]
46. Hamano-Takaku, F.; Iwama, T.; Saito-Yano, S.; Takaku, K.; Monden, Y.; Kitabatake, M.; Söll, D.; Nishimura, S. A mutant *Escherichia coli* tyrosyl-tRNA synthetase utilizes the unnatural amino acid azatyrosine more efficiently than tyrosine. *J. Biol. Chem.* **2000**, *275*, 40324–40328. [[CrossRef](#)]
47. Hauenstein, S.I.; Hou, Y.M.; Perona, J.J. The homotetrameric phosphoserine-tRNA synthetase from *Methanosarcina mazei* exhibits half-of-the-sites activity. *J. Biol. Chem.* **2008**, *283*, 21997–22006. [[CrossRef](#)]
48. Steinfeld, J.B.; Aerni, H.R.; Rogulina, S.; Liu, Y.; Rinehart, J. Expanded cellular amino acid pools containing phosphoserine, phosphothreonine, and phosphotyrosine. *ACS Chem. Biol.* **2014**, *9*, 1104–1112. [[CrossRef](#)]
49. Burkovski, A.; Kramer, R. Bacterial amino acid transport proteins: occurrence, functions, and significance for biotechnological applications. *Appl. Microbiol. Biot.* **2002**, *58*, 265–274.
50. Boehm, J.C.; Kingsbury, W.D.; Perry, D.; Gilvarg, C. The use of cysteinyl peptides to effect portage transport of sulfhydryl-containing compounds in *Escherichia coli*. *J. Biol. Chem.* **1983**, *258*, 14850–14855.
51. Smith, M.W.; Tyreman, D.R.; Payne, G.M.; Marshall, N.J.; Payne, J.W. Substrate specificity of the periplasmic dipeptide-binding protein from *Escherichia coli*: Experimental basis for the design of peptide prodrugs. *Microbiology* **1999**, *145*, 2891–2901. [[CrossRef](#)] [[PubMed](#)]
52. Luo, X.; Fu, G.; Wang, R.E.; Zhu, X.; Zambaldo, C.; Liu, R.; Liu, T.; Lyu, X.; Du, J.; Xuan, W.; et al. Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. *Nat. Chem. Biol.* **2017**, *13*, 845–849. [[CrossRef](#)] [[PubMed](#)]
53. Tame, J.R.; Dodson, E.J.; Murshudov, G.; Higgins, C.F.; Wilkinson, A.J. The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands. *Structure* **1995**, *3*, 1395–1406. [[CrossRef](#)]
54. Kuenzl, T.; Sroka, M.; Srivastava, P.; Herdewijn, P.; Marliere, P.; Panke, S. Overcoming the membrane barrier: Recruitment of  $\gamma$ -glutamyl transferase for intracellular release of metabolic cargo from peptide vectors. *Metab. Eng.* **2017**, *39*, 60–70. [[CrossRef](#)] [[PubMed](#)]
55. 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](#)]
56. Beránek, V.; Reinkemeier, C.D.; Zhang, M.S.; Liang, A.D.; Kym, G.; Chin, J.W. Genetically encoded protein phosphorylation in mammalian cells. *Cell Chem. Biol.* **2018**, *25*, 1067–1074. [[CrossRef](#)]
57. Longstaff, D.G.; Larue, R.C.; Faust, J.E.; Mahapatra, A.; Zhang, L.; Green-Church, K.B.; Krzycki, J.A. A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1021–1026. [[CrossRef](#)]
58. Fan, C.; Fromm, H.J.; Bobik, T.A. Kinetic and functional analysis of L-threonine kinase, the PduX enzyme of *Salmonella enterica*. *J. Biol. Chem.* **2009**, *284*, 20240–20248. [[CrossRef](#)]
59. Zhang, M.S.; Brunner, S.F.; Huguenin-Dezot, N.; Liang, A.D.; Schmied, W.H.; Rogerson, D.T.; Chin, J.W. Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nat. Methods* **2017**, *14*, 729–736. [[CrossRef](#)]
60. Niemann, C.; Rapport, M.M. The toxicity of 3-fluoro-*d*(+)- and *l*(-)-tyrosine. *J. Am. Chem. Soc.* **1946**, *68*, 1671. [[CrossRef](#)]



61. Reichau, S.; Blackmore, N.J.; Jiao, W.; Parker, E.J. Probing the sophisticated synergistic allosteric regulation of aromatic amino acid biosynthesis in *Mycobacterium tuberculosis* using D-amino acids. *PLoS ONE* **2016**, *11*, e0152723. [[CrossRef](#)]
62. Wanner, B.L.; Metcalf, W.W. Molecular genetic studies of a 10.9-kb operon in *Escherichia coli* for phosphonate uptake and biodegradation. *FEMS Microbiol. Lett.* **1992**, *100*, 133–139. [[CrossRef](#)]
63. Ko, J.H.; Wang, Y.S.; Nakamura, A.; Guo, L.T.; Söll, D.; Umehara, T. Pyrrolysyl-tRNA synthetase variants reveal ancestral aminoacylation function. *FEBS Lett.* **2013**, *587*, 3243–3248. [[CrossRef](#)]
64. Liu, D.R.; Magliery, T.J.; Pasternak, M.; Schultz, P.G. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10092–10097. [[CrossRef](#)]
65. Santoro, S.W.; Schultz, P.G. Directed evolution of the substrate specificities of a site-specific recombinase and an aminoacyl-tRNA synthetase using fluorescence-activated cell sorting (FACS). *Methods Mol. Biol.* **2003**, *230*, 291–312.
66. Acevedo-Rocha, C.G.; Reetz, M.T.; Nov, Y. Economical analysis of saturation mutagenesis experiments. *Sci. Rep.* **2015**, *5*, 10654. [[CrossRef](#)]
67. Richter, F.; Leaver-Fay, A.; Khare, S.D.; Bjelic, S.; Baker, D. De novo enzyme design using Rosetta3. *PLoS ONE* **2011**, *6*, e19230. [[CrossRef](#)]
68. Amiram, M.; Haimovich, A.D.; Fan, C.; Wang, Y.S.; Aerni, H.R.; Ntai, I.; Moonan, D.W.; Ma, N.J.; Rovner, A.J.; Hong, S.H.; et al. Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. *Nat. Biotechnol.* **2015**, *33*, 1272–1279. [[CrossRef](#)]
69. Bryson, D.I.; Fan, C.; Guo, L.T.; Miller, C.; Söll, D.; Liu, D.R. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **2017**, *13*, 1253–1260. [[CrossRef](#)]
70. 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. Edit.* **2015**, *54*, 10030–10034. [[CrossRef](#)]
71. Wang, H.H.; Isaacs, F.J.; Carr, P.A.; Sun, Z.Z.; Xu, G.; Forest, C.R.; Church, G.M. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **2009**, *460*, 894–898. [[CrossRef](#)] [[PubMed](#)]
72. Ellis, H.M.; Yu, D.; DiTizio, T.; Court, D.L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6742–6746. [[CrossRef](#)] [[PubMed](#)]
73. Kuhn, S.M.; Rubini, M.; Fuhrmann, M.; Theobald, I.; Skerra, A. Engineering of an orthogonal aminoacyl-tRNA synthetase for efficient incorporation of the non-natural amino acid O-methyl-L-tyrosine using fluorescence-based bacterial cell sorting. *J. Mol. Biol.* **2010**, *404*, 70–87. [[CrossRef](#)] [[PubMed](#)]
74. Kwok, H.S.; Vargas-Rodriguez, O.; Melnikov, S.V.; Söll, D. Engineered aminoacyl-tRNA synthetases with improved selectivity towards non-canonical amino acids. *ACS Chem. Biol.* **2019**, *14*, 603–612. [[CrossRef](#)] [[PubMed](#)]
75. Badran, A.H.; Liu, D.R. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* **2015**, *6*, 8425. [[CrossRef](#)] [[PubMed](#)]
76. Esvelt, K.M.; Carlson, J.C.; Liu, D.R. A system for the continuous directed evolution of biomolecules. *Nature* **2011**, *472*, 499–503. [[CrossRef](#)] [[PubMed](#)]
77. Wang, L.; Schultz, P.G. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* **2001**, *8*, 883–890. [[CrossRef](#)]
78. Umehara, T.; Kim, J.; Lee, S.; Guo, L.T.; Söll, D.; Park, H.S. N-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo. *FEBS Lett.* **2012**, *586*, 729–733. [[CrossRef](#)]
79. Maranhao, A.C.; Ellington, A.D. Evolving orthogonal suppressor tRNAs to incorporate modified amino acids. *ACS Synth. Biol.* **2017**, *6*, 108–119. [[CrossRef](#)]
80. Owens, A.E.; Grasso, K.T.; Ziegler, C.A.; Fasan, R. Two-tier screening platform for directed evolution of aminoacyl-tRNA synthetases with enhanced stop codon suppression efficiency. *ChemBioChem* **2017**, *18*, 1109–1116. [[CrossRef](#)]
81. Thyer, R.; Shroff, R.; Klein, D.R.; d’Oelsnitz, S.; Cotham, V.C.; Byrom, M.; Brodbelt, J.S.; Ellington, A.D. Custom selenoprotein production enabled by laboratory evolution of recoded bacterial strains. *Nat. Biotechnol.* **2018**, *36*, 624–631. [[CrossRef](#)]

82. Volkwein, W.; Maier, C.; Krafczyk, R.; Jung, K.; Lassak, J. A versatile toolbox for the control of protein levels using N<sup>e</sup>-acetyl-L-lysine dependent amber suppression. *ACS Synth. Biol.* **2017**, *6*, 1892–1902. [[CrossRef](#)]
83. McKinnon, K.M. Flow cytometry: An overview. *Curr. Protoc. Immunol.* **2018**, *120*, 5 1 1–5 1 11.
84. Young, T.S.; Ahmad, I.; Yin, J.A.; Schultz, P.G. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J. Mol. Biol.* **2010**, *395*, 361–374. [[CrossRef](#)]
85. Guo, L.T.; Wang, Y.S.; Nakamura, A.; Eiler, D.; Kavran, J.M.; Wong, M.; Kiessling, L.L.; Steitz, T.A.; O'Donoghue, P.; Söll, D. Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 16724–16729. [[CrossRef](#)]
86. Young, D.D.; Jockush, S.; Turro, N.J.; Schultz, P.G. Synthetase polyspecificity as a tool to modulate protein function. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 7502–7504. [[CrossRef](#)]
87. Stokes, A.L.; Miyake-Stoner, S.J.; Peeler, J.C.; Nguyen, D.P.; Hammer, R.P.; Mehl, R.A. Enhancing the utility of unnatural amino acid synthetases by manipulating broad substrate specificity. *Mol. Biosyst.* **2009**, *5*, 1032–1038. [[CrossRef](#)]
88. Aharoni, A.; Gaidukov, L.; Khersonsky, O.; Gould, S.M.; Roodveldt, C.; Tawfik, D.S. The “evolvability” of promiscuous protein functions. *Nat. Genet.* **2005**, *37*, 73–76. [[CrossRef](#)]
89. Kunjapur, A.M.; Stork, D.A.; Kuru, E.; Vargas-Rodriguez, O.; Landon, M.; Söll, D.; Church, G.M. Engineering posttranslational proofreading to discriminate nonstandard amino acids. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 619–624. [[CrossRef](#)]
90. Bajaj, K.; Chakrabarti, P.; Varadarajan, R. Mutagenesis-based definitions and probes of residue burial in proteins. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16221–16226. [[CrossRef](#)]
91. Aldag, C.; Bröcker, M.J.; Hohn, M.J.; Prat, L.; Hammond, G.; Plummer, A.; Söll, D. Rewiring translation for elongation factor Tu-dependent selenocysteine incorporation. *Angew. Chem. Int. Edit.* **2013**, *52*, 1441–1445. [[CrossRef](#)]
92. Yoshizawa, S.; Böck, A. The many levels of control on bacterial selenoprotein synthesis. *Biochim. Biophys. Acta* **2009**, *1790*, 1404–1414. [[CrossRef](#)]
93. Haruna, K.; Alkazemi, M.H.; Liu, Y.; Söll, D.; Englert, M. Engineering the elongation factor Tu for efficient selenoprotein synthesis. *Nucleic Acids Res.* **2014**, *42*, 9976–9983. [[CrossRef](#)]
94. Thyer, R.; Robotham, S.A.; Brodbelt, J.S.; Ellington, A.D. Evolving tRNA<sup>Sec</sup> for efficient canonical incorporation of selenocysteine. *J. Am. Chem. Soc.* **2015**, *137*, 46–49. [[CrossRef](#)]
95. Mukai, T.; Sevostyanova, A.; Suzuki, T.; Fu, X.; Söll, D. A facile method for producing selenocysteine-containing proteins. *Angew. Chem. Int. Edit.* **2018**, *57*, 7215–7219. [[CrossRef](#)]
96. Fu, X.; Crnkovic, A.; Sevostyanova, A.; Söll, D. Designing seryl-tRNA synthetase for improved serylation of selenocysteine tRNAs. *FEBS Lett.* **2018**, *592*, 3759–3768. [[CrossRef](#)]
97. Humbard, M.A.; Surkov, S.; de Donatis, G.M.; Jenkins, L.M.; Maurizi, M.R. The N-degradome of *Escherichia coli*: Limited proteolysis in vivo generates a large pool of proteins bearing N-degrons. *J. Biol. Chem.* **2013**, *288*, 28913–28924. [[CrossRef](#)]
98. Fowler, D.M.; Araya, C.L.; Fleishman, S.J.; Kellogg, E.H.; Stephany, J.J.; Baker, D.; Fields, S. High-resolution mapping of protein sequence–function relationships. *Nat. Methods* **2010**, *7*, 741–746. [[CrossRef](#)]
99. Wrenbeck, E.E.; Faber, M.S.; Whitehead, T.A. Deep sequencing methods for protein engineering and design. *Curr. Opin. Struct. Biol.* **2017**, *45*, 36–44. [[CrossRef](#)]
100. Uyeda, A.; Watanabe, T.; Kato, Y.; Watanabe, H.; Yomo, T.; Hoshaka, T.; Matsuura, T. Liposome-based in vitro evolution of aminoacyl-tRNA synthetase for enhanced pyrrolysine derivative incorporation. *ChemBioChem* **2015**, *16*, 1797–1802. [[CrossRef](#)]

