



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Engineering pyrrolysyl-trna synthetase for the incorporation of non-canonical amino acids with smaller side chains

Citation for published version:

Koch, NG, Goettig, P, Rappsilber, J & Budisa, N 2021, 'Engineering pyrrolysyl-trna synthetase for the incorporation of non-canonical amino acids with smaller side chains', *International Journal of Molecular Sciences*, vol. 22, no. 20, 11194. <https://doi.org/10.3390/ijms222011194>

Digital Object Identifier (DOI):

[10.3390/ijms222011194](https://doi.org/10.3390/ijms222011194)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

International Journal of Molecular Sciences

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Article

Engineering Pyrrolysyl-tRNA Synthetase for the Incorporation of Non-Canonical Amino Acids with Smaller Side Chains

Nikolaj G. Koch ^{1,2} , Peter Goettig ³ , Juri Rappsilber ^{2,4} and Nediljko Budisa ^{1,5,*}

¹ Institut für Chemie, Technische Universität Berlin, 10623 Berlin, Germany; nikolaj.koch@mail.tu-berlin.de

² Institut für Biotechnologie-Bioanalytik, Technische Universität Berlin, 10623 Berlin, Germany; juri.rappsilber@tu-berlin.de

³ Structural Biology Group, Department of Biosciences, University of Salzburg, 5020 Salzburg, Austria; peter.goettig@plus.ac.at

⁴ Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3BF, UK

⁵ Department of Chemistry, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

* Correspondence: nediljko.budisa@tu-berlin.de; Tel.: +49-30-314-28821

Abstract: Site-specific incorporation of non-canonical amino acids (ncAAs) into proteins has emerged as a universal tool for systems bioengineering at the interface of chemistry, biology, and technology. The diversification of the repertoire of the genetic code has been achieved for amino acids with long and/or bulky side chains equipped with various bioorthogonal tags and useful spectral probes. Although ncAAs with relatively small side chains and similar properties are of great interest to biophysics, cell biology, and biomaterial science, they can rarely be incorporated into proteins. To address this gap, we report the engineering of PylRS variants capable of incorporating an entire library of aliphatic “small-tag” ncAAs. In particular, we performed mutational studies of a specific PylRS, designed to incorporate the shortest non-bulky ncAA (S-allyl-L-cysteine) possible to date and based on this knowledge incorporated aliphatic ncAA derivatives. In this way, we have not only increased the number of translationally active “small-tag” ncAAs, but also determined key residues responsible for maintaining orthogonality, while engineering the PylRS for these interesting substrates. Based on the known plasticity of PylRS toward different substrates, our approach further expands the reassignment capacities of this enzyme toward aliphatic amino acids with smaller side chains endowed with valuable functionalities.

Keywords: genetic code expansion; pyrrolysyl-tRNA synthetases; non-canonical amino acids; aliphatic amino acids; bioorthogonal reactive handles; azidohomoalanine; photo-methionine; protein engineering; stop codon suppression; S-allyl-L-cysteine



Citation: Koch, N.G.; Goettig, P.; Rappsilber, J.; Budisa, N. Engineering Pyrrolysyl-tRNA Synthetase for the Incorporation of Non-Canonical Amino Acids with Smaller Side Chains. *Int. J. Mol. Sci.* **2021**, *22*, 11194. <https://doi.org/10.3390/ijms222011194>

Academic Editor: Kensaku Sakamoto

Received: 19 September 2021

Accepted: 14 October 2021

Published: 17 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Research in the field of reprogrammed protein translation has now reached experimental and intellectual maturity: More than 200 non-canonical amino acids (ncAAs, i.e., a diversity that is an order of magnitude higher than that of the canonical amino acid repertoire) were introduced into proteins via various genetic code expansion routes: Selective pressure incorporation, stop codon suppression (SCS), fragment condensation, protein semisynthesis, and peptidomimetics [1]. It has been shown that AAs with non-proteinogenic functional groups can be used to manipulate, design, and elucidate protein structure, dynamics, function, allostery, interactions, catalysis, folding, synthesis, trafficking, degradation, and aggregation [2–8].

Engineering aminoacyl-tRNA synthetase (aaRS)/tRNA pairs capable of recognizing, activating, and loading ncAAs onto their cognate tRNAs is now a well-established strategy. It enables the site-specific ribosomal incorporation of ncAAs in response to a reprogrammed codon. The most commonly used approach for this purpose is in-frame stop codon suppression, targeting the amber stop codon [9–11]. Hereby, the ncAA is

incorporated in response to an in-frame stop codon placed at a predefined position in the protein coding sequence, ribosomally expressed either *in vivo* or *in vitro* [9–11]. Most aaRS variants used for SCS so far are derived from *Methanosarcina mazei/barkeri* pyrrolysyl-tRNA synthetases (*MmPylRS*/*MbPylRS*) or *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase (*MjTyrRS*) [9–12]. The archaeal origin and therefore distant phylogeny is responsible for their orthogonality in bacterial and eukaryotic cells [12].

The native substrate of the PylRS is the rare proteinogenic amino acid pyrrolysine (Pyl, 1a), a lysine analog with a 4-ethyl-pyrroline-5-carboxylate ring attached to the terminal amino function of the side chain (Figure 1). The wild type enzyme can activate several Pyl variants resembling ncAAs [12]. Moreover, catalytic promiscuity is widely exploited in both native and genetically engineered classes of PylRS enzymes to enable recognition, activation, and tRNA loading of the majority of all translationally active ncAAs. [13–16]. It should be noted, however, that the majority of incorporable ncAAs with the PylRS system are characterized with flexible, long-chained, and bulky pyrrolysine analogs [13,14] or shorter but still bulky aromatic substrates, especially phenylalanine [16], tryptophan [17], and histidine [18] analogs. Therefore, a new class of PylRS enzymes capable of recognizing, activating, and tRNA loading with shorter chain ncAAs endowed with useful functional groups is of great interest. Small ncAAs with shorter side chains containing azido, thioene, fluoro, cyano, and nitroso groups can be particularly useful, e.g., for FTIR, NMR, crosslinking, and spin labeling, because longer side chains are too flexible which usually results in a loss of spectral information or the necessary proximity for specific bioorthogonal reactions [19]. Moreover, still no efficient non-canonical counterparts are available for Glu and Asp, which often form structurally important salt bridges or hydrogen-bond networks. It would be useful to modify these acidic residues, e.g., by removing their hydrogen bond donors or acceptors. Mimicking post-translational modifications of canonical amino acids (cAAs) with their genetically encoded ncAA counterparts is also an attractive application to elucidate their functions.

The main reason for the substrate promiscuity of PylRS is most likely the unique substrate binding mode with relatively nonspecific hydrophobic interactions in the large binding pocket of this enzyme. Therefore, it is not surprising that a multitude of ncAAs can be recognized and activated with very few mutations (the majority has just 2–4) mainly in the binding pocket [12]. For this reason, the PylRS is predestined for the implementation of new functions [20]. In general, an enzyme should possess two key features to ensure successful recognition of new substrates. First, the target enzyme should have low levels of the desired new activity, which in case of PylRS means enzymatic activity toward ncAAs that are highly divergent from the native substrate [21,22]. Second, sufficient stability is required to buffer destabilizing mutations necessary for active site remodeling [23–25]. Unfortunately, PylRS is marginally stable under standard cultivation conditions in *Escherichia coli* (*E. coli*) [26], which is also reflected by the low *in vitro* solubility of the enzyme [20,27]. We demonstrated that this drawback can be partly remedied with a solubility tag fused to the N-terminus of *MbPylRS* (manuscript in preparation) which made this our enzyme of choice.

In our study, we performed mutational analyses to elucidate the structure activity relationship of a PylRS designed to incorporate *S*-allyl-L-cysteine (Sac, 1). Based on this knowledge, we engineered several new *MbPylRS* variants on a rational and semi-rational basis, in order to incorporate a variety of small side chain ncAAs. We used an entire library of small side-chain-containing ncAAs that can be structurally and functionally categorized into five classes (Figure 1, detailed discussion in Appendix A). Briefly, they include (i) aliphatic analogs of Sac (1); (ii) bio-orthogonal tags; (iii) small ncAAs with useful spectroscopic probes; (iv) methionine analogs; and (v) substrates with a terminal alkene as site-specific chemical cleavage site (being also bio-orthogonal tags).

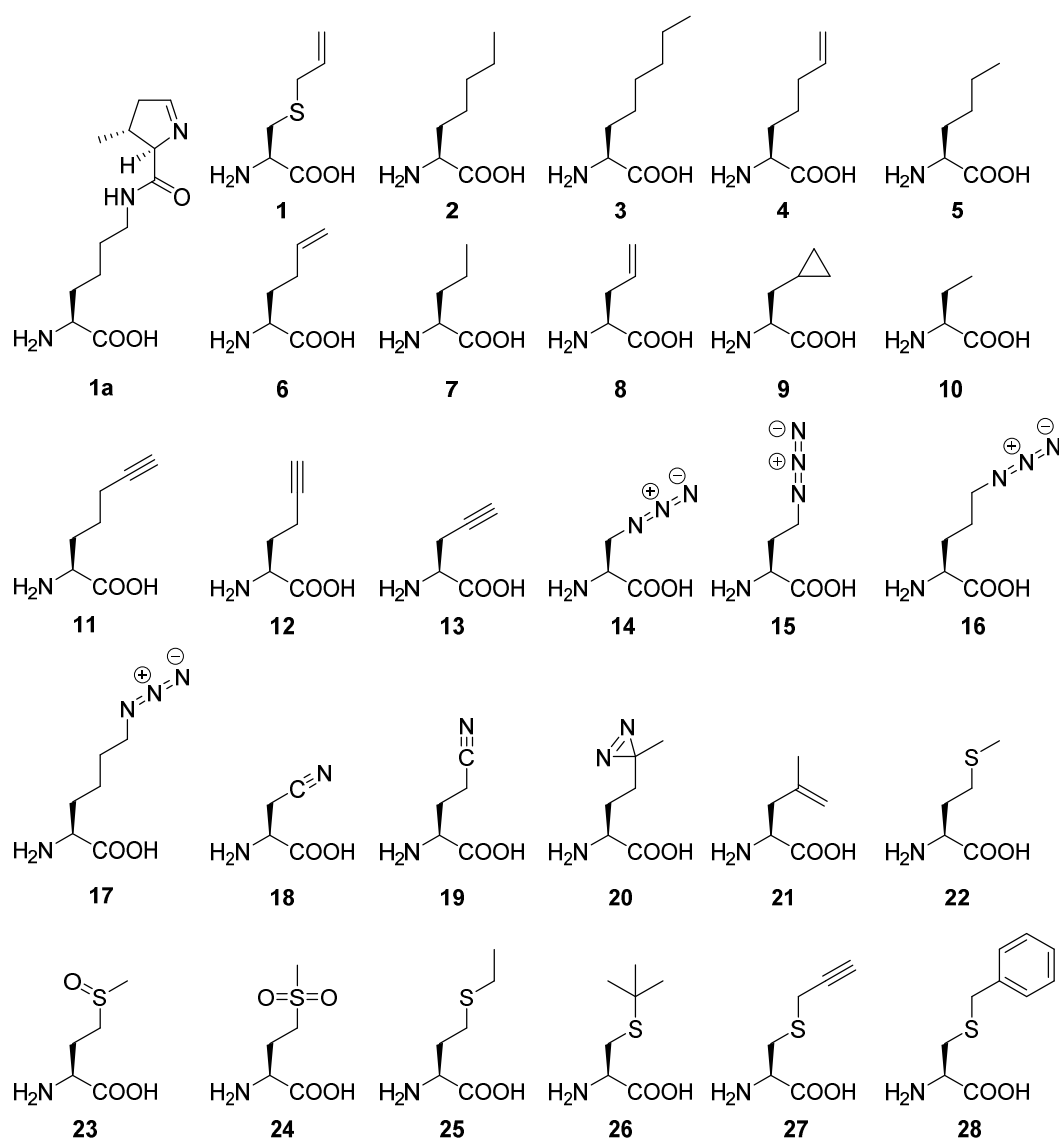


Figure 1. Survey of amino acids used in this study. Chemical structure of pyrrolysine (1a), *S*-allyl-L-cysteine (1), (*S*)-2-aminoheptanoic acid (2), (*S*)-2-aminooctanoic acid (3), (*S*)-2-aminohept-6-enoic acid (4), (*S*)-2-aminohexanoic acid (5), (*S*)-2-aminohex-5-enoic acid (6), (*S*)-2-aminopentanoic acid (7), (*S*)-2-aminopent-4-enoic acid (8), (*S*)-2-amino-3-cyclopropylpropanoic acid (9), (*S*)-2-aminobutyric acid (10), (*S*)-2-aminohept-6-ynoic acid (11), (*S*)-2-aminohex-5-ynoic acid (12), (*S*)-2-aminopent-4-ynoic acid (13), (*S*)-2-amino-3-azidopropanoic acid (14), (*S*)-2-amino-4-azidobutanoic acid (15), (*S*)-2-amino-5-azidopentanoic acid (16), (*S*)-2-amino-6-azidohexanoic acid (17), (*S*)-2-amino-3-cyanopropanoic acid (18), (*S*)-2-amino-4-cyanobutanoic acid (19), (*S*)-2-amino-5,5'-azi-hexanoic acid (20), (*S*)-2-amino-4-methylpent-4-enoic acid (21), L-methionine (22), L-methionine sulfoxide (23), L-methionine sulfone (24), L-ethionine (25), *S*-tert-butyl-L-cysteine (26), *S*-propargyl-L-cysteine (27), *S*-benzyl-L-cysteine (28).

2. Results and Discussion

2.1. General MbPylRS and ncAA Incorporation Readout Setup

All PylRS variants used in this study are equipped with an N-terminal SmbP solubility tag [28]. As mentioned earlier, the tag restores activity by dramatically increasing the abundance of soluble and active enzyme compared to the untagged aaRS. Most likely, this phenomenon is due to an increase in kinetic stability and builds an improved and solid starting point for our enzyme engineering efforts. This fusion enzyme is used throughout this work (for sequence information see supplement part 2.4). The Y349F mutation was included by default, as it is known to generally enhance aminoacylation and orthogonal translation systems (OTS) efficiency [13,29].

To test the efficacy of ncAAs incorporation, we used superfolder-GFP (sfGFP) as a model protein. The sfGFP-based fluorescence readout is the simplest approach for this purpose as the fluorescence intensity of intact cells is directly correlated to the amount of protein produced. The sfGFP reporter construct comprises an in-frame stop codon at position 2 (instead of an arginine triplet; sfGFP(R2 amber)). This construct has been routinely used as a readout vehicle for the in vivo suppression efficiency of the in-frame amber stop codon [30]. This reporter construct is an integral part of the OTS established in *E. coli* BL21(DE3) expression host strain. To avoid deficiency of ncAA in the cell and to detect even very low incorporation activity, we used high concentrations of ncAAs (10 mM, unless otherwise stated) in this study.

2.2. Testing *MbSacRS* for Aliphatic Substrates

We hypothesized that the previously reported SacRS could accommodate close structural aliphatic analogs of Sac (1). Therefore, this variant would be a good starting point for the evolution of SacRS toward similar small-tag substrates. To scrutinize this hypothesis, a SacRS variant (hereafter referred to as *MbSacRS*) was created based on a codon optimized *MbPylRS* sequence by introducing the two crucial active site mutations C313W:W382S. In addition, two previously identified advantageous N-terminal mutations T13L:I36V ([31], cf. Figure S1) were introduced as well and maintained them for all other constructs. Surprisingly, we found very low incorporation of the aliphatic substrates for the *MbSacRS* (Figure S2). Since this specificity among very close structural analogs is relatively uncommon for both native and mutant PylRS enzymes, we set out to investigate the structure–activity relationships of the *MbSacRS*.

2.3. Elucidating the Structure–Activity Relationships of *MbSacRS* via Rational Mutation Studies

To date, no high-resolution crystal structure of SacRS is available. Since there are only two PylRS mutations responsible for altering the substrate specificity to Sac (1), we decided to perform rational mutation studies to elucidate the role of each residue in Sac-incorporation activity. An overview of all relevant residues, whose mutations were guided by crystal structures, can be found in Figure 2. Figure 2B was especially helpful because it contains the same C313W mutation as the *MbSacRS*.

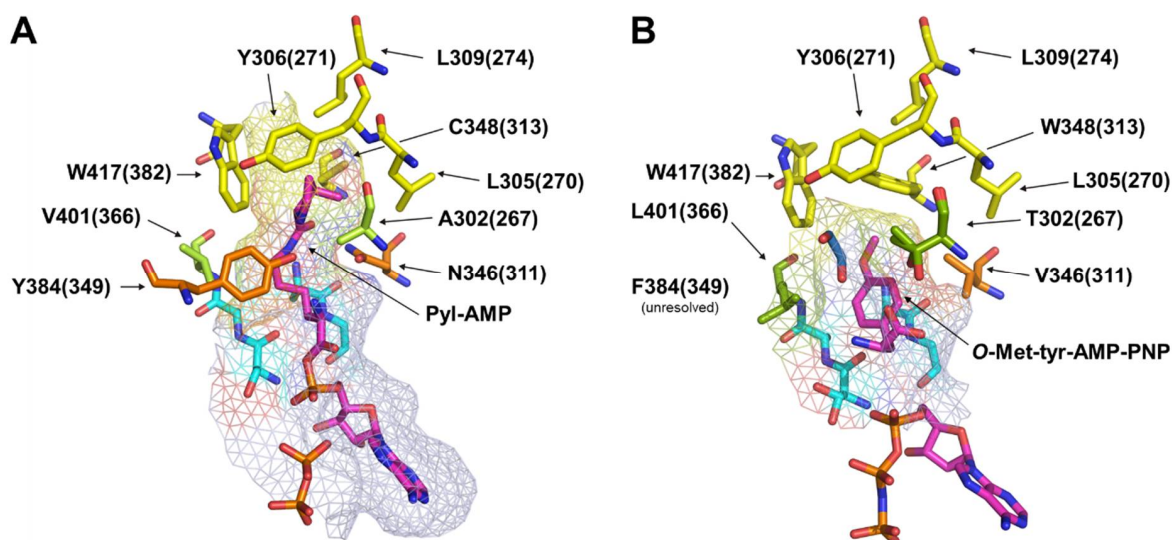


Figure 2. Microenvironment of the active sites derived from the crystal structures of *MmPylRS* and the *MmOmeRS* mutant. These structures guided the rational mutation approach. Shown are critical residues forming the active site. Since only structures of *M. mazei* are available, these were used in a homology model for *M. barkeri*. Residue numbers in brackets reflect the numbering of *M. barkeri*, while numbers not in brackets refer to *M. mazei*. (A) Wild-type *MmPylRS* (PDB ID: 2Q7H) [20] with bound Pyl-AMP. (B) Mutant *MmOmeRS* (PDB ID: 3QTC) [32] with bound O-Methyl-tyrosine-AMP-PNP (O-Met-tyr-AMP-PNP).

Starting with residue S382, we reverted this position to wild-type Trp and tested less bulky amino acids. As serine was located at position 382 in the original SacRS, we investigated whether polar functional groups are necessary for Sac (1) incorporation. As shown in Figure 3A, all constructed mutants resulted in comparable Sac incorporation, with the exception of the C313W:W382F and C313W constructs. This was quite surprising since the authors in the original SacRS report found only three variants for Sac (1) incorporation and just one variant possessed the C313W mutation [33]. This finding highlights the enormous importance of quality control when constructing libraries and the need for sufficient analytics when analyzing newly found variants after screening. Figure 3A does not provide a clear picture of whether a hydroxyl group at position 382 is an advantage, since the C313W:W382A variant performs at comparable levels. In contrast, small size clearly plays an important role, with the Phe and Trp mutations being the two most inefficient variants in the group so far. Interestingly, the Trp mutant is found to have a strong cAA background incorporation. This phenomenon is perfectly in line with the literature reports. It is known that PylRS enzymes with C313W mutations and a small residue at position N311 incorporate Phe [34]. The original finding of SacRS was based on positive and negative selection rounds. The inclusion of the negative step in the selection against variants with high background incorporations clearly shows why the variant W382 was overlooked. Considering the gathered data, the W382 mutation, to smaller residues, is most likely important for restoring orthogonality of *Mb*SacRS.

We tested the four best variants as depicted in Figure 3A in a concentration-dependent manner to gain detailed information about the OTS performance *in vivo*. The best constructs were (C313W:W382T/Y) and showing similar activity, but twice the OTS efficiency (at 0.6 mM Sac (1)) compared to the original SacRS (Figure S3). To determine whether this was a specific property of the *Mb*PylRS, we transferred the mutations to the *Mm*PylRS, the variant resulting in the SacRS. For this variant, the equivalent of the (C313W:W382T) mutations was also observed as the best outcome, indicating a robust result (Figure S4). Interestingly the second-best variant of *Mb* and *Mm*SacRS were different. This result highlights the fact that the best mutation found in one species is not necessary the best in another, although mutations are transferable between the different PylRS systems of different species. The C313W:W382T variant was also reported to incorporate *S*-propargyl-L-cysteine (27) [35]. Therefore, we also tested the top four Sac (1) incorporating constructs also with substrate 27 and found that the (C313W:W382T/Y) constructs gave comparable results (Figure S5). All gathered data for the W382 mutations suggest that this residue only needs to be smaller than Trp to restore orthogonality. However, the efficiency of Sac incorporation differs for different residues at this position. In particular, variants with a polar OH- or SH-group at W382 position perform best. Examining the effect of the C313W mutation in our system, the data displayed in Figure 3B clearly indicates that the size of the residue at this position is the most prominent factor, albeit not the only one. No variant with a small amino acid at this position was able to incorporate Sac. By contrast, the bulkier Phe allows Sac incorporation, though at a lower level compared with all C313W mutants. Even with Met at this position, very low incorporation is detectable. The inactive C313H variant, comparable in size to C313F, suggests that a non-polar residue is necessary to be present in this microenvironment. Overall, the data collected suggest that a reduced size of the binding pocket containing the C313W mutation is essential for Sac incorporation.

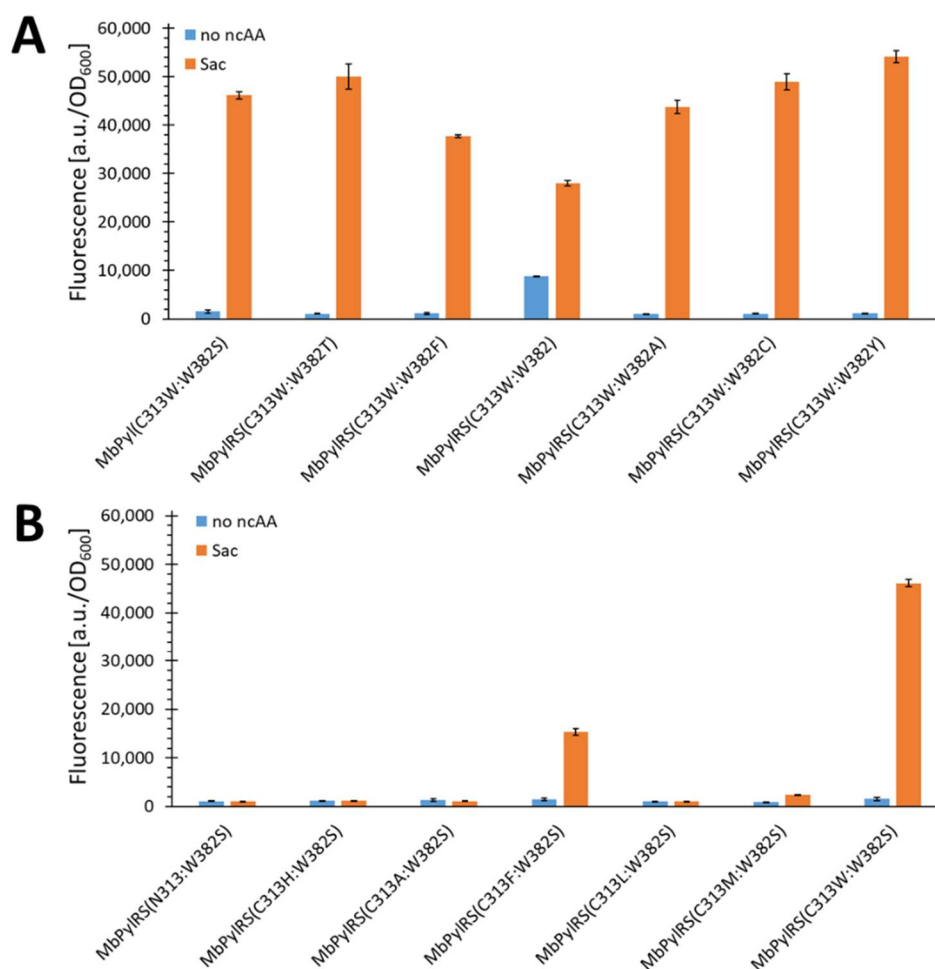


Figure 3. Comparison of Sac incorporation efficiency for MbPylRS constructs (A) MbPylRS(C313W) and variants mutated at position W382 and (B) MbPylRS(W382S) with variants mutated at position C313. The fluorescence was measured for intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2amber) reporter protein. The data (incl. standard deviation) represent the mean of three biological replicates.

2.4. Rationalizing Sac Incorporation Data and Creating Aliphatic Substrate Activating MbPylRS Variants

It was previously proposed that the C313W mutation is critical for activation of smaller substrates [32,33]. This hypothesis was also fully recapitulated in this work. Having thus established that the C313W mutation is critical for the incorporation of Sac (1) and probably also smaller aliphatic Sac (1) analogs, all these mutants were tested for incorporation of substrates 2, 3, 5, 7, and 10 (Figure 4). These fluorescence data showed that all variants, except the C313W mutant, did not incorporate the tested substrates.

Unfortunately, the construct that incorporated the aliphatic ncAAs also exhibited a considerable level of cAA background incorporation. As mentioned previously, bulky C313 mutations of MbPylRS lead to background Phe incorporation [20,34]. We hypothesized that the key to restoring the lost orthogonality associated with the C313W mutation would lie in the mutation of MbPylRS at position N311, which is also known to be one of the two gatekeeper residues for Pyl activation. Our working hypothesis was to restore the orthogonality by increasing the size of the amino acid at this position to interfere with the Phe substrate accommodation while creating a catalytic pocket suitable for the short-chain aliphatic ncAAs. As shown in Figure 5A, the orthogonality also increases by increasing the AA size at the N311 position, until it is fully restored for the N311L:C313W variant.

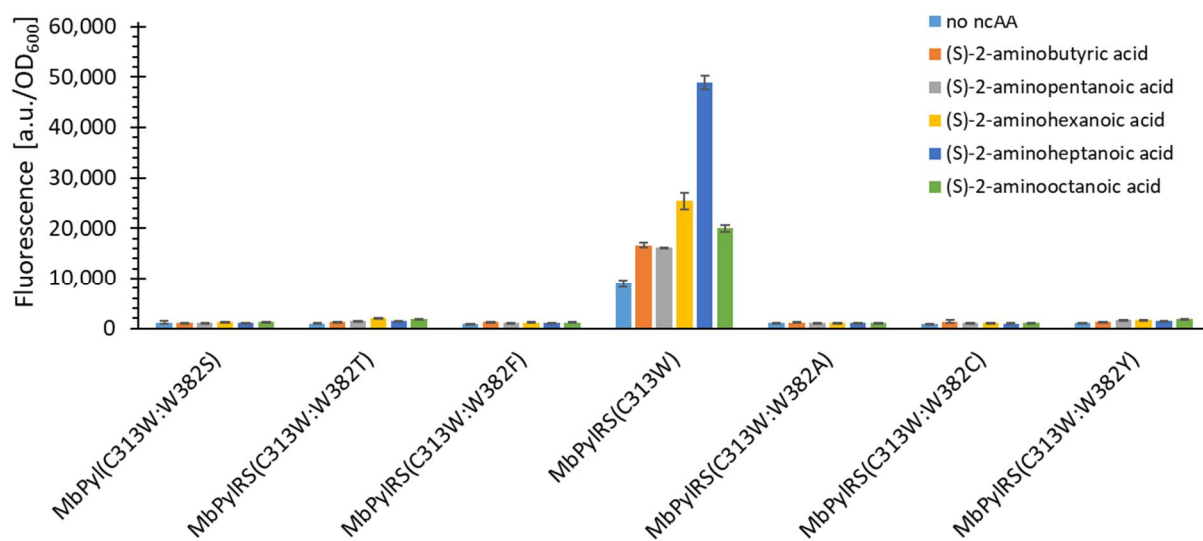


Figure 4. Comparison of the efficiency of aliphatic ncAA incorporation for *MbPyIRS(C313W)* constructs mutated at position W382. The fluorescence was measured for intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. The data (incl. standard deviation) represent the mean of three biological replicates.

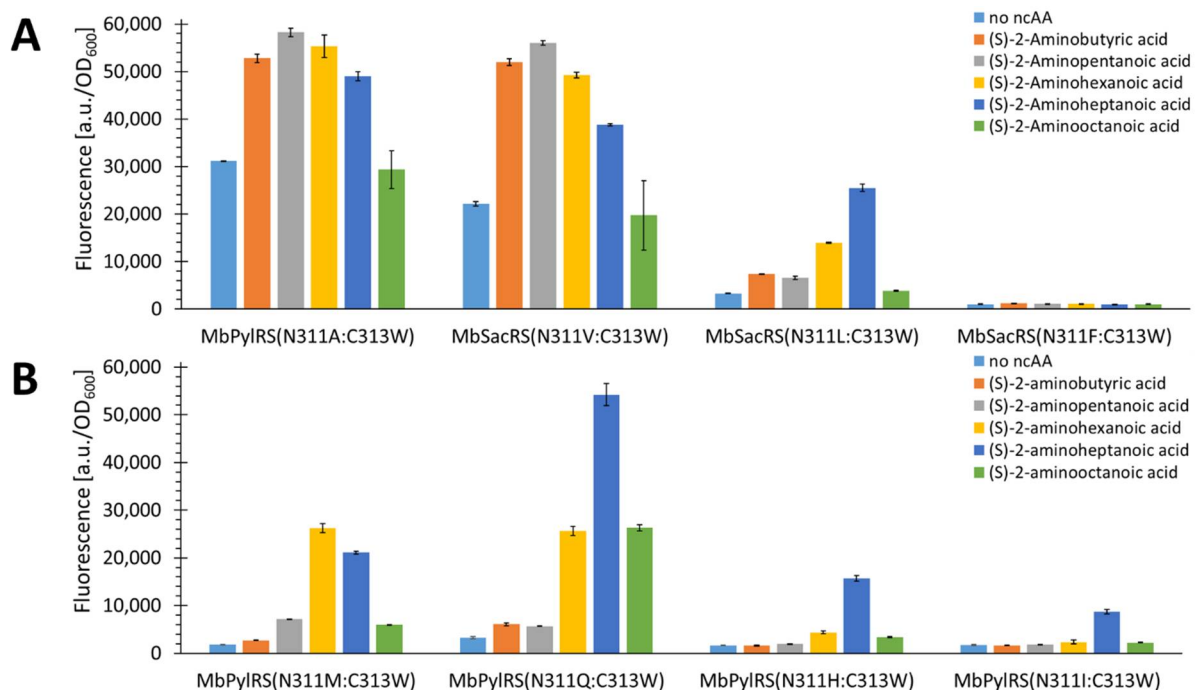


Figure 5. Comparison of aliphatic ncAA incorporation efficiency for *MbPyIRS(C313W)* constructs mutated at position N311. (A) Construct N311A with increased binding pocket size and additional mutations that stepwise decrease the size of the binding pocket. (B) Four additional active variants, found after screening of all 19 (all cAAs besides glycine) possible constructs. The fluorescence was measured for intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. The data (incl. standard deviation) represent the mean of three biological replicates.

An incorporation pattern of the tested aliphatic substrates was noticeable, most likely corresponding to the size of the newly created catalytic pocket. The signal intensity for the N311L:C313W variant suggests that substrate 2 is the optimal size for this pocket. This substrate is the aliphatic equivalent to Sac (1). The data supported the hypothesis that it was possible to restore orthogonality by replacing the residue at position N311 to bulkier AAs. This finding encouraged us to test all amino acids larger than valine at this position. The

additional functional variants found are shown in Figure 5B. Figure 5 shows all the obtained active variants. Two of them (N311M/Q:C313W) can incorporate some of the substrates even better than the N311L:C313W construct. Interestingly, these two additional variants exhibit a different ncAA incorporation profile. The N311M construct favors substrates with carbon chain length C6, while the N311Q favors C7. The incorporation profile of the N311Q variant is similar to that of the N311L.

The key residues found here, which are important for maintaining the orthogonality of the PylRS system will facilitate future enzyme engineering efforts for the incorporation of structural analogs. The information will enable the creation of smarter PylRS-libraries; in particular, now a reduction in library size is feasible and will increase the likelihood of finding desired enzymes. This will facilitate the establishment of ready-made enzyme arsenals to test a wide variety of natural and synthetic ncAAs for translational activity. In addition, for close structural analogues, the number and type of selection cycles (e.g., negative selection) can be reduced, making the overall selection process simpler and more feasible.

2.5. Semi-Rational Engineering of PylRS Constructs for Small Aliphatic Substrate Incorporation

The previously described effort led to five constructs with efficient incorporation of aliphatic ncAAs. We selected two of these constructs (N311M:C313W and N311Q:C313W) for further engineering. The goal was to improve the incorporation efficiency and/or specificity of the aliphatic substrates. For this reason, we targeted residues potentially in close proximity to them. After inspecting the crystal structure (Figure 2), we planned to randomize residues A267, V366, Y349, and W382 in the active site by site-saturation mutagenesis (SSM) with NNK primers (N = A/T/C/G; K = G/T). We started with position V366 because it is located opposite to the N311 residue. Thereby we hypothesized that altering V366 might have the strongest tuning effect with respect to aliphatic ncAA recognition. The randomization of this position resulted in two new variants for each parent construct, with mutations V366A/K (Figure 6).

Figure 6A shows the OTS performance of these constructs compared to the starting N311M:C313W construct. The V366K variant did not show a strong difference in the incorporation profile compared to the starting enzyme. Interestingly, the V366A mutation resulted in a specificity shift toward long-chain aliphatic ncAAs in the incorporation profile. This phenomenon seems plausible based on the *Mm*PylRS crystal structures (Figure 2), since this mutation potentially increases the space of the binding pocket, which should facilitate the incorporation of longer substrates. Figure 6B shows that the two variants found, based on the N311Q:C313W construct, performed comparably to the parent enzyme. Similar to the N311M:C313W:V366A construct, there is also a slight shift in the incorporation profile toward longer ncAAs observable for the V366A mutant. Since this variant already favors ncAAs with C7 chain length over C6, the shift is smaller than for the N311M:C313W construct. Unfortunately, screening of randomizations A267, Y349, and W382 did not yield better performing variants. However, a variant with a markedly different incorporation profile was found (N311M:C313W:W382H). This mutant preferably incorporated the longer chained substrate 3 (Figure S6).

2.6. Evaluating the Incorporation of Biochemically Useful Aliphatic ncAA Analogs

We screened all generated *Mb*PylRS constructs for incorporation of the aliphatic ncAA/AA analogs listed in Figure 1 (besides 2, 3, 5, 7 and 10). Substrates 1–9, 20–26, and 28 were highly incorporated (Figure 5, Figures S7–S9) and could be expressed in a standard *E. coli* BL21(DE) protein production strain to higher levels (up to 21 mg/L protein per culture, Table 1). Interestingly, several constructs also incorporated Sac (1) and substrate 27, with the best Sac (1) incorporating construct being N311M:C313W:V366A (Figures S10–S12). This result clearly illustrates that multiple substrate-recognizing enzyme topologies can be based on the same scaffold. Since this variant was also able to incorporate 26, we were encouraged to perform another round of randomization to find out if its performance is

enhanced or if other interesting Cys-based amino acid derivatives were incorporated. This approach led to the identification of N311M:C313W:V366A:W382N/T/Y mutants that were able to successfully incorporate substrate 28.

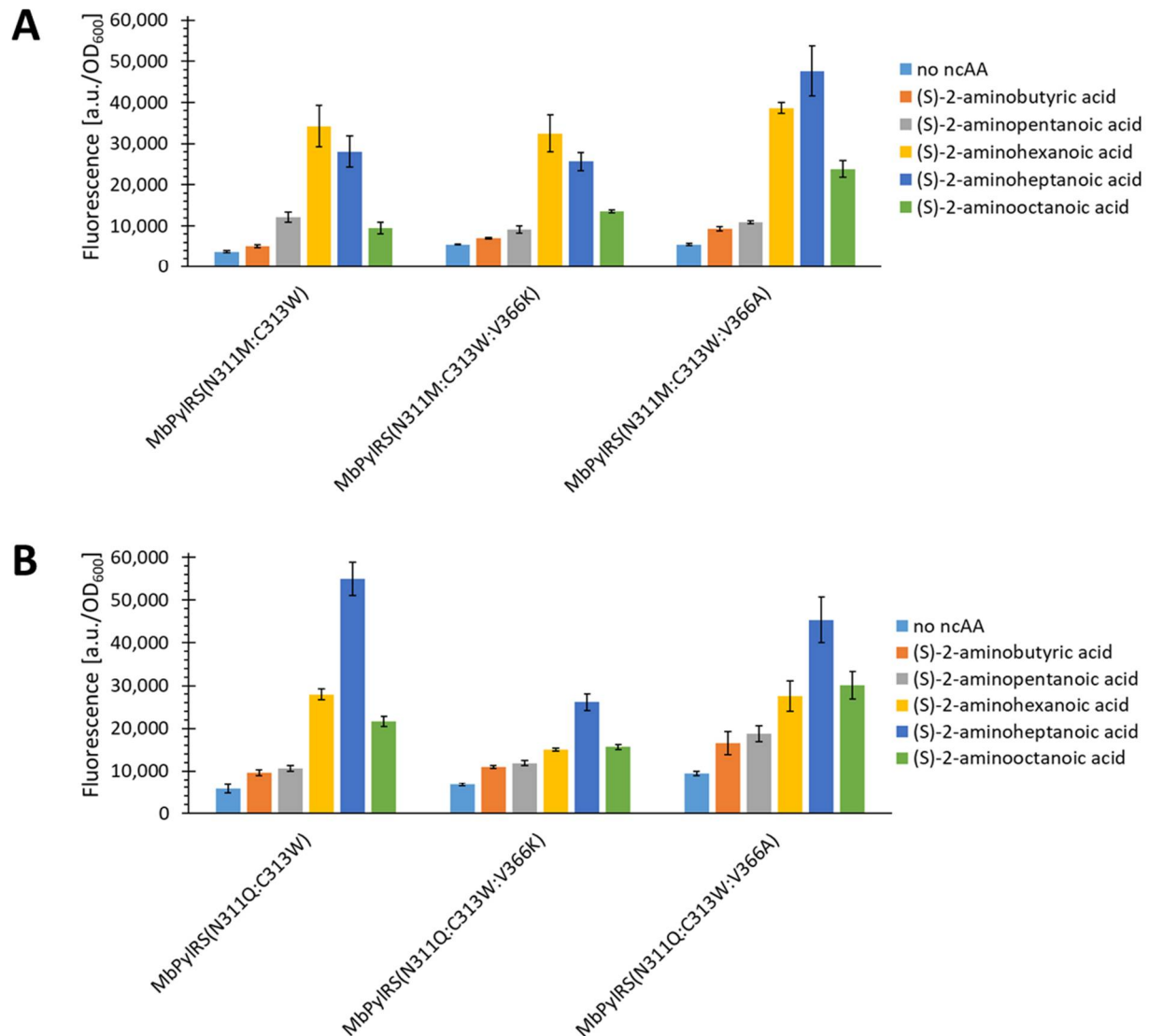


Figure 6. Comparison of aliphatic ncAA incorporation efficiency for (A) *MbPyIRS(N311M:C313W)* and (B) *MbPyIRS(N311Q:C313W)* constructs both mutated at position V366. The fluorescence was measured for intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. The data (incl. standard deviation) represent the mean of three biological replicates.

Substrates 11–19 with very low incorporation signals were additionally screened in release factor 1 (RF1) knock-out strains JX33, B-95.ΔA, and C321.ΔA.exp (Figures S13–S16) [36–38]. Usually, strains lacking RF1 produce higher amounts of full-length protein by amber suppression. All of these RF1 knock-out strains were engineered in previous work of our group to possess the lambda DE3 lysogen encoding the T7 polymerase compatible with our reporter protein setup. Although background incorporation increased in all RF1 knockout strains (as previously observed see [33]), some setups resulted in an increased ratio of ncAA/AA incorporation compared to background incorporation levels. The best performing strain and *MbPyIRS* combinations were selected for larger scale protein production (Table 1).

Table 1. Optimal reporter protein production setup, calculated and observed molecular weights of the reporter proteins His₆-SUMO-sfGFP(R2AA)-strep (a) /SUMO-sfGFP(R2AA)-His₆ (b) and protein production yields per liter of culture. The masses were determined by ESI-MS of intact proteins.

AA	<i>E. coli</i> Strains ¹	<i>MbPylRS</i> Construct	Reporter Construct	Calculated Mass [Da]	Observed Mass [Da]	Δ Mass [Da]	Protein Yield [mg·L ⁻¹] ²
1	BL21	N311M:C313W:V366A	a	40,194.9	40,196	1.1	10.8
2	BL21	N311Q:C313W	a	40,178.8	40,180	1.2	5.1
3	BL21	N311M:C313W:V366A	a	40,192.9	40,194	1.1	1.6
4	BL21	N311M:C313W:V366A	a	40,176.8	40,179	2.2	1.7
5	BL21	N311M:C313W	a	40,164.8	40,166	1.2	1.9
6	BL21	N311M:C313W:V366K	a	40,162.8	40,164	1.2	1.4
7	BL21	N311M:C313W	a	40,150.8	40,153	2.2	1.2
8	BL21	N311M:C313W	a	40,148.8	40,150	1.2	0.7
9	BL21	N311M:C313W	a	40,162.8	40,164	1.2	1.4
10	BL21	N311M:C313W:V366A	a	40,136.8	40,195	58.2	0.8
11	C321.ΔA.exp	N311M:C313W	b	38,990.9	38,992	1.1	5.1
12	C321.ΔA.exp	N311M:C313W	b	38,976.9	38,979	2.1	4.9
13	C321.ΔA.exp	N311M:C313W	b	38,962.8	38,965	2.2	11.3
14	JX33	N311M:C313W	b	38,979.8	38,996	16.2	4.3
15	C321.ΔA.exp	N311M:C313W	b	38,993.8	38,994	0.2	14.2
16	C321.ΔA.exp	N311M:C313W	b ⁴	39,007.9	39,007	0.9	5.3
17	C321.ΔA.exp	N311M:C313W	b	39,021.9	38,997	24.9	6.4
18	C321.ΔA.exp	N311Q:C313W:V366K	b	38,963.8	39,015	51.2	19
19	C321.ΔA.exp	N311Q:C313W:V366K	b	38,977.8	39,014	36.2	19.9
20	BL21	N311M:C313W:V366K	a	40,190.8	40,194	3.2	4.8
21	BL21	N311M:C313W	b	38,978.9	38,982	3.1	2.6
22	BL21	N311M:C313W	b	38,998.9	38,998	0.9	3.6
23	BL21	N311Q:C313W	b	39,014.9	39,012	2.9	4.6
24	BL21	N311Q:C313W:V366K	b	39,030.9	39,015	15.9	10.7
25	BL21	N311Q:C313W	b	39,013	39,014	1	9.8
26	BL21	N311M:C313W:V366A	a	40,210.9	40,211	0.1	21
27	BL21	N311M:C313W:V366A	a ³	-	-	-	-
28	BL21	N311M:C313W:V366A:W382N	b	39,061	39,063	2	1.1

¹ all DE3, ² yield per liter of cell culture, ³ was not purified, ⁴ not the main peak

2.7. Analytics of Canonical/Non-Canonical Amino Acids Incorporation

Larger scale protein production was performed to confirm the results from the fluorescence assays in small-scale 96-well plates. The protein yields are in good agreement with the trends observed in the fluorescence experiments. We acquired the mass spectra of the intact reporter proteins via electrospray ionization mass spectrometry (ESI-MS). To facilitate MS data evaluation for AAs with a low incorporation efficiency, a reporter protein with a C-terminal His₆-tag was used. Table 1 shows the optimal setup of reporter protein production, the ESI-MS data, and protein yields for each substrate. The corresponding deconvoluted ESI-MS-spectra are shown in the supplementary information (Figures S17–S42). The incorporation of all substrates but five (14, 17, 18, 19, and 24) was confirmed in the MS analytics. For substrate 14, the molecular weight of incorporated AA is 146.3 g/mol which is equivalent to that of glutamine (146.2 g/mol). It is known that near-cognate suppressor tRNAs, like tRNA^{Gln}, read amber codons to some extent [39]. This means Gln is incorporated at amber sites when the OTS is not working and then frequently observed in MS analytics. The high reporter protein yield for a non-functioning OTS is most likely the result of the general higher background suppression observed in RF1 knock-out strains (Figures S13–S16). Gln incorporation is also observed for the setup of substrate 16, 17. For substrates 18, 19, and 24, the molecular weight of the incorporated AA is 165.3/164.3 g/mol, which is equivalent to phenylalanine (165.2 g/mol), indicating that this PylRS leads to Phe incorporation when no or inefficient substrate is present. The observed high protein yield for the setup of substrates 18 and 19 is probably caused by a combination

of the Phe incorporation activity and the use of a RF1 knock-out strain, although further analytics would be required here. The same is true for substrate 24, though on a lower level since no RF1 knock-out strain was used. Nonetheless, the fluorescence data with different *MbPylRS* constructs clearly show that the incorporation is possible (Figure S14).

In summary, in this study, we generated 28 *MbPylRS* variants which can incorporate 23 ncAAs and one cAA. To the best of our knowledge, 17 of these ncAAs (besides 1, 3, 4, 17, 27, and 28) were not ribosomally incorporated by amber suppression before and 20 of them were not previously incorporated with the PylRS system [33,35,40,41].

3. Materials and Methods

3.1. Canonical and Non-Canonical Amino Acids

Canonical amino acids were purchased from Carl Roth. Non-canonical amino acids were obtained from Fluorochem, Iris Biotech, Chempur, Sigma-Aldrich (Merck), Chiralix, Toronto Research Chemicals, Carl Roth, Thermo Fisher Scientific and TCI Deutschland (see Table S1).

3.2. Plasmid Vector Construction

All plasmids were assembled by Golden Gate cloning and confirmed by DNA sequencing. Plasmids harboring the OTS (aaRS/tRNA^{Py1}) were constructed by cloning the target aaRS gene into the pTECH vector (Addgene plasmid #104073) [42].

3.3. Site-Directed and Site-Saturation Mutagenesis

Point mutations were introduced by non-overlapping inverse PCR [43]. Focused *MbPylRS* gene libraries were also created with non-overlapping inverse PCR, but randomization was performed using mutagenic primers (with NNK, whereby N = A, T, G, or C; K = G or T) at designated positions (A267, V366, F349, and W382).

3.4. Analysis of SUMO-sfGFP Expression by Intact Cell Fluorescence.

For the small-scale expression of reporter constructs, *E. coli* BL21(DE) cells were used. Electrocompetent cells were transformed with the orthogonal translation system and reporter plasmids. LB agar plates for plating contained 1% glucose and corresponding antibiotics. Single colonies of clones were used for inoculation of 2 mL LB (in 14 mL tubes) with 1% glucose and appropriate antibiotics and were grown to saturation overnight. Assays were conducted in 96-well plate format. Cultures were added to each well at 1:100 dilution in ZYP-5052 auto induction medium to a final volume of 100 μ L supplemented with antibiotics and ncAAs. Cells were grown in black μ -plates (Greiner Bio-One, Kremsmünster, Austria) covered with a gas permeable foil (Breathe-Easy[®], Diversified Biotech, Dedham, MA, USA) with orbital shaking for 24 h at 37 °C. For endpoint measurements (Tecan M200, Männedorf, Switzerland), the plate foil was removed and fluorescence was measured with an 85 gain setting. For OD₆₀₀ measurements, 50 μ L of ZYP-5052 medium was pipetted into clear 96-well μ -plates and 50 μ L of culture was added. Excitation and emission wavelengths for fluorescence measurements were set to 481 nm and 511 nm, respectively. Fluorescence values were normalized to the corresponding OD₆₀₀. Biological triplicates were used for measurements of each aaRS construct. Relative fluorescence was normalized to the highest value. The data including standard deviation represent the mean of three biological replicates.

3.5. Library Screening

After library creation and transformation, 96 clones were picked and grown overnight in a 96-well plate in 100 μ L LB with 1% glucose and appropriate antibiotics. The next day a 96-well plate with 100 μ L ZYP-5052, appropriate antibiotics, and ncAAs was inoculated with 1 μ L culture, grown for 24 h and measured afterwards as stated above. The 96-plate which was used for inoculation was sealed with aluminum foil and stored at 4 °C. From this plate, desired clones were analyzed via PCR gene amplification and sequencing of

this PCR product afterwards. Calculations with the Toplib tool estimate the probability of finding the best performing variant to be 96% (using a yield of 85%, which is the lower limit of primer purity and therefore also the lower yield limit of created DNA constructs) [44].

3.6. Protein Expression

For expression of the SUMO-sfGFP variants, *E.coli* strains were used in 10 mL ZYP-5052 medium supplemented with 10 mM ncAA and appropriate antibiotics. The expression medium was inoculated with a starter culture (1:100). Shake flasks were incubated for 24 h at 37 °C while shaking at 200 rpm. Cells were harvested by centrifugation and stored at −80 °C or directly used for protein purification.

3.7. Protein Purification

Harvested cell pellets were resuspended (50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 8.0) and lysed with B PER[®] Bacterial Protein Extraction Reagent (Thermo Scientific, Waltham, MA, USA) according to their protocol, with addition of phenylmethanesulfonyl fluoride (PMSF, 1 mM final concentration), DNase, and RNase. Cleared lysates were loaded onto a equilibrated Ni-NTA column and purified via the P-1 peristaltic pump (Pharmacia Biotech, now Cytiva, Marlborough, MA, USA). After washing with 10 column volumes of resuspension buffer, elution buffer (50 mM sodium phosphate, 300 mM NaCl, 500 mM imidazole, pH 8.0) was applied to elute the his-tagged target proteins. The first 2 mL covering the void volume was discarded. Afterwards, the eluate (1 mL) was collected and dialyzed in cellulose film tubings against 1 L buffer (50 mM sodium phosphate, 300 mM NaCl, pH 8.0) for at least 2 h with three buffer changes. Concentrations of purified reporter proteins were determined by measuring the sfGFP chromophore absorption at 488 nm.

3.8. ESI-MS

Intact protein mass measurements of purified SUMO-sfGFP variants were performed by electrospray LC-MS on a Waters H-class instrument with a Waters Acquity UPLC protein BEH C4 column (300 Å, 1.7 µm, 2.1 mm × 50 mm). The following gradient used a flow rate of 0.3 mL/min: A: 0.01% formic acid in H₂O; B: 0.01% formic acid in MeCN. 5–95% B 0–6 min. Mass analysis was conducted with a Waters Xevo G2-XS QToF analyzer. Proteins were ionized in positive ion mode applying a cone voltage of 40 kV. Raw data were analyzed employing the maximum entropy deconvolution algorithm. The data were exported and plotted with QtiPlot (version 0.9.9.7).

4. Conclusions

We were motivated by previous works implying that the PylRS, naturally specialized for large bulky substrates, could be redesigned for small substrates. Therefore, we elucidated the structure–activity relationship of a specific *MbPylRS* (*MbSacRS*), designed to incorporate the shortest non-bulky ncAA (*S*-allyl-L-Cysteine, 1) possible to date. Based on this knowledge, we have designed *MbPylRS*s for the incorporation of aliphatic amino acids and various derivatives thereof, which are useful in biochemistry and structural biology. Some of the incorporated substrates (allylglycine, 8 and propargylglycine, 13) were recently synthesized in vivo in *E.coli* and would, therefore, open up the possibility of coupling metabolic engineering and ncAA incorporation [45]. This approach could eliminate the need to add ncAAs to the cultivation medium, which would drastically decrease costs and simplify associated applications.

In addition, valuable information was gathered about the role of specific residues in the active site that are responsible for the orthogonality of PylRS. We were able to determine key residues that maintain orthogonality upon engineering the *MbPylRS* enzyme to accommodate smaller substrates. This achievement will facilitate future enzyme engineering efforts for the incorporation of structural analogs. To this end, PylRS libraries can now be created with a smaller size, which simultaneously increases the likelihood of finding

the desired enzymes and reduces the amount of work required to do so by eliminating negative selection.

Supplementary Materials: The following tables and figures are available online at <https://www.mdpi.com/article/10.3390/ijms222011194/s1>.

Author Contributions: Conceived the project and conceptualization, N.B. and N.G.K.; methodology, N.G.K.; software, N.G.K.; validation, N.G.K.; formal analysis, N.G.K.; investigation, N.G.K.; resources, N.B.; data curation, N.G.K. and N.B.; writing—original draft preparation, N.G.K.; writing—review and editing, N.G.K., N.B., P.G. and J.R.; visualization, N.G.K.; supervision, P.G., J.R. and N.B.; project administration, P.G., J.R., N.B.; funding acquisition, P.G., J.R. and N.B. All authors have read and agreed to the published version of the manuscript.

Funding: N.G.K. acknowledges funding by the “IBÖM04: XenoGlue”-Consortium supported by the Federal Ministry of Education and Research of Germany (Grant Number: FKZ 031B0584A). This work was also performed as part of the “Site-directed cross-linking of KLK proteases from prostate” Project funded by the Lead Agency FWF (I 3877-B21) with DFG-D-A-CH (BU1404/12-1) (P. G., N.G.K.). N.B. thanks Canada Research Chairs Program (Grant Nr. 950-231971) for support. This work was partially supported by the Cluster of Excellence “Unifying Systems in Catalysis” (UniSysCat), funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy–EXC 2008/1–390540038 (N.B. and J. R.)”.

Acknowledgments: We are very grateful to Philipp Ochtrop (Leibniz-Forschungsinstitut für Molekulare Pharmakologie, Hackenberger Group) for his support in ESI-MS measurements. N.B. thanks to Christian Thomsen, President of the Technical University of Berlin, for his continuous support.

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

Abbreviations

AA	amino acid
AMP	adenosine monophosphate
AMP-PNP	adenosine-5'-[(β,γ)-imido]triphosphate
aaRS	aminoacyl-tRNA synthetase
cAA	canonical amino acid
ESI-MS	electrospray ionization mass spectrometry
FTIR	Fourier-transform infrared spectroscopy
MjTyrRS	<i>Methanocaldococcus jannaschii</i> tyrosyl-tRNA synthetase
MmPylRS/MbPylRS	<i>Methanosarcina mazei/barkeri</i> pyrrolysyl-tRNA synthetase
ncAA	non-canonical amino a
NMR	nuclear magnetic resonance spectroscopy
OTS	orthogonal translation system
Pyl	pyrrolysine
PylRS	pyrrolysyl-tRNA synthetase
Sac	S-allyl-L-cysteine
sfGFP	superfolder GFP
SmbP	small metal-binding protein
RF1	release factor 1

Appendix A

Appendix A.1. General Features and Perspectives of ncAAs Used in This Study

As mentioned in the introduction, ncAAs with side chain functionality closer to the protein backbone would be highly advantageous for various applications. We used an entire library of small side chain containing ncAAs that can be structurally and functionally categorized into five classes, as shown in Figure 1. (i) Aliphatic Sac (1) analogs for structure activity elucidation (2, 3, 5, 7, 10). (ii) Site-specific bioorthogonal reaction handles that can be used for a variety of bioconjugation reactions (4, 6, 8, 9, 11, 12, 13, 14, 15, 16, 17, 27). These reactions include metal-free (e.g., click-chemistry, Staudinger ligations and strain-promoted cycloadditions) and transition metal-mediated (e.g., ruthenium-based

olefin cross-metathesis or palladium based oxidative HECK and SONOGASHIRA cross-coupling reactions) approaches [46]. (iii) ncAAs are/could be used as biophysical probes, e.g., in vibrational Stark, IR, and NMR spectroscopy (18, 19, 26) and as genetically encoded photo-crosslinker (20) [19,47,48]. (iv) Methionine analogs (23, 24, 25) as tools to elucidate the role of methionine oxidation in proteins, enzymes, and cells [49]. (v) Substrates with a terminal alkene as site-specific chemical cleavage sites (4, 6, 8, 21) [50–53]. For example, the cleavage reaction with substrate 8 proceeds presumably via iodolactonization, suggesting that this reaction could also proceed with substrates 6 and 8. The transition state would change from a five-membered iodolactone to a 6/7 membered one [54–56]. In contrast to the classical site-specific peptide cleavage with cyanogen bromide at a methionine position, these ncAAs could be cleaved with non-toxic iodine under mild conditions [57].

In yeast, substrates 3 and 4 have already been incorporated with an orthogonal *E. coli* leucyl-tRNA synthetase [40]. Although we used the five aliphatic ncAAs (2, 3, 5, 7, 10) to estimate the size of the narrowed active site, some of these ncAAs could potentially help to address certain questions regarding post-translational lipidation of proteins [58–61]. Lipidation is also a common strategy to improve the pharmacokinetic properties of biopharmaceuticals, especially to prolong the systemic half-life in patients with a corresponding increase in bioavailability [62]. This approach opens a potentially interesting biomedical application area for these ncAAs. Notably, ncAA substrates 5, 9, 12, 15, 17, 20, and 25 have been incorporated in a residue specific manner (using auxotrophy-based methods) but never in a site-specific mode, besides 17 [48,63–67]. Substrate 9 contains a cyclopropane ring with properties closely resembling to those of an olefinic double bond, which could be exploited for a wide range of site-specific bioorthogonal protein conjugation reactions [68]. These reactions include enzymatic halogenation with a haloperoxidase [69], enzymatic oxidation with a mono-oxygenase [70], nucleophilic substitutions, electrophilic ring opening reactions, and a plethora of other reactions [68]. Lastly, allylglycine (8) and propargylglycine (13) have recently been synthesized in vivo in *E. coli* and would, therefore, open up the possibility of coupling metabolic engineering and the incorporation of ncAA [45]. This procedure could eliminate the need to add ncAAs to the cultivation medium, which would drastically decrease costs and simplify associated applications.

References

1. Pagar, A.D.; Patil, M.D.; Flood, D.T.; Yoo, T.H.; Dawson, P.E.; Yun, H. Recent Advances in Biocatalysis with Chemical Modification and Expanded Amino Acid Alphabet. *Chem. Rev.* **2021**, *121*, 6173–6245. [[CrossRef](#)] [[PubMed](#)]
2. Groff, D.; Thielges, M.C.; Cellitti, S.; Schultz, P.G.; Romesberg, F.E. Efforts toward the direct experimental characterization of enzyme microenvironments: Tyrosine 100 in dihydrofolate reductase. *Angew. Chem. Int. Ed.* **2009**, *48*, 3478–3481. [[CrossRef](#)] [[PubMed](#)]
3. Baumann, T.; Hauf, M.; Schildhauer, F.; Eberl, K.B.; Durkin, P.M.; Deniz, E.; Löffler, J.G.; Acevedo-Rocha, C.G.; Jaric, J.; Martins, B.M.; et al. Site-Resolved Observation of Vibrational Energy Transfer Using a Genetically Encoded Ultrafast Heater. *Angew. Chem. Int. Ed.* **2019**. [[CrossRef](#)] [[PubMed](#)]
4. Minnihhan, E.C.; Young, D.D.; Schultz, P.G.; Stubbe, J. Incorporation of fluorotyrosines into ribonucleotide reductase using an evolved, polyspecific aminoacyl-tRNA synthetase. *J. Am. Chem. Soc.* **2011**, *133*, 15942–15945. [[CrossRef](#)] [[PubMed](#)]
5. Li, J.C.; Nastertorabi, F.; Xuan, W.; Han, G.W.; Stevens, R.C.; Schultz, P.G. A Single Reactive Noncanonical Amino Acid Is Able to Dramatically Stabilize Protein Structure. *ACS Chem. Biol.* **2019**, *14*, 1150–1153. [[CrossRef](#)]
6. Agostini, F.; Völler, J.S.; Koksche, B.; Acevedo-Rocha, C.G.; Kubyshekin, V.; Budisa, N. Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angew. Chem. Int. Ed.* **2017**, *56*, 9680–9703. [[CrossRef](#)]
7. Drienovská, I.; Mayer, C.; Dulson, C.; Roelfes, G. A designer enzyme for hydrazone and oxime formation featuring an unnatural catalytic aniline residue. *Nat. Chem.* **2018**, *10*, 946–952. [[CrossRef](#)]
8. Burke, A.J.; Lovelock, S.L.; Frese, A.; Crawshaw, R.; Ortmyer, M.; Dunstan, M.; Levy, C.; Green, A.P. Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. *Nature* **2019**, *570*, 219–223. [[CrossRef](#)]
9. Liu, C.C.; Schultz, P.G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, *79*, 413–444. [[CrossRef](#)]
10. Mukai, T.; Lajoie, M.J.; Englert, M.; Söll, D. Rewriting the Genetic Code. *Annu. Rev. Microbiol.* **2017**, *71*, 557–577. [[CrossRef](#)]
11. Chin, J.W. Expanding and reprogramming the genetic code. *Nature* **2017**, *550*, 53–60. [[CrossRef](#)] [[PubMed](#)]
12. Wan, W.; Tharp, J.M.; Liu, W.R. Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta Proteins Proteom.* **2014**, *1844*, 1059–1070. [[CrossRef](#)] [[PubMed](#)]

13. Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode N ϵ -(o-Azidobenzoyloxycarbonyl) lysine for Site-Specific Protein Modification. *Chem. Biol.* **2008**, *15*, 1187–1197. [[CrossRef](#)]
14. Yanagisawa, T.; Kuratani, M.; Seki, E.; Hino, N.; Sakamoto, K.; Yokoyama, S. Structural Basis for Genetic-Code Expansion with Bulky Lysine Derivatives by an Engineered Pyrrolysyl-tRNA Synthetase. *Cell Chem. Biol.* **2019**, *26*, 936–949. [[CrossRef](#)] [[PubMed](#)]
15. 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](#)]
16. Wang, Y.S.; Fang, X.; Wallace, A.L.; Wu, B.; Liu, W.R. A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. *J. Am. Chem. Soc.* **2012**, *134*, 2950–2953. [[CrossRef](#)]
17. Tseng, H.; Baumann, T.; Sun, H.; Wang, Y.; Ignatova, Z.; Budisa, N. Expanding the Scope of Orthogonal Translation with Pyrrolysyl-tRNA Synthetases Dedicated to Aromatic Amino Acids. *Molecules* **2020**, *25*, 4418. [[CrossRef](#)]
18. Xiao, H.; Peters, F.B.; Yang, P.Y.; Reed, S.; Chittuluru, J.R.; Schultz, P.G. Genetic incorporation of histidine derivatives using an engineered pyrrolysyl-tRNA synthetase. *ACS Chem. Biol.* **2014**, *9*, 1092–1096. [[CrossRef](#)]
19. Völler, J.S.; Biava, H.; Hildebrandt, P.; Budisa, N. An expanded genetic code for probing the role of electrostatics in enzyme catalysis by vibrational Stark spectroscopy. *Biochim. Biophys. Acta Gen. Subj.* **2017**, *1861*, 3053–3059. [[CrossRef](#)]
20. Kavran, J.M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Söll, D.; Steitz, T.A. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11268–11273. [[CrossRef](#)] [[PubMed](#)]
21. Trudeau, D.L.; Tawfik, D.S. Protein engineers turned evolutionists—the quest for the optimal starting point. *Curr. Opin. Biotechnol.* **2019**, *60*, 46–52. [[CrossRef](#)] [[PubMed](#)]
22. Arnold, F.H. Directed Evolution: Bringing New Chemistry to Life. *Angew. Chem. Int. Ed.* **2018**, *57*, 4143–4148. [[CrossRef](#)] [[PubMed](#)]
23. Bloom, J.D.; Labthavikul, S.T.; Otey, C.R.; Arnold, F.H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 5869–5874. [[CrossRef](#)] [[PubMed](#)]
24. Tokuriki, N.; Tawfik, D.S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **2009**, *19*, 596–604. [[CrossRef](#)]
25. Grasso, K.T.; Yeo, M.J.R.; Hillenbrand, C.M.; Ficarella, E.D.; Italia, J.S.; Huang, R.L.; Chatterjee, A. Structural Robustness Affects the Engineerability of Aminoacyl-tRNA Synthetases for Genetic Code Expansion. *Biochemistry* **2021**, *60*, 489–493. [[CrossRef](#)]
26. Hu, L.; Qin, X.; Huang, Y.; Cao, W.; Wang, C.; Wang, Y.; Ling, X.; Chen, H.; Wu, D.; Lin, Y.; et al. Thermophilic Pyrrolysyl-tRNA Synthetase Mutants for Enhanced Mammalian Genetic Code Expansion. *ACS Synth. Biol.* **2020**, *9*, 2723–2736. [[CrossRef](#)]
27. Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. Crystallographic Studies on Multiple Conformational States of Active-site Loops in Pyrrolysyl-tRNA Synthetase. *J. Mol. Biol.* **2008**, *378*, 634–652. [[CrossRef](#)]
28. Vargas-Cortez, T.; Morones-Ramirez, J.R.; Balderas-Renteria, I.; Zarate, X. Expression and purification of recombinant proteins in *Escherichia coli* tagged with a small metal-binding protein from *Nitrosomonas europaea*. *Protein Expr. Purif.* **2016**, *118*, 49–54. [[CrossRef](#)]
29. 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](#)]
30. Baumann, T.; Hauf, M.; Richter, F.; Albers, S.; Möglich, A.; Ignatova, Z.; Budisa, N. Computational aminoacyl-tRNA synthetase library design for photocaged tyrosine. *Int. J. Mol. Sci.* **2019**, *20*, 2343. [[CrossRef](#)]
31. 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](#)] [[PubMed](#)]
32. Takimoto, J.K.; Dellas, N.; Noel, J.P.; Wang, L. Stereochemical Basis for Engineered Pyrrolysyl-tRNA Synthetase and the Efficient in Vivo Incorporation of Structurally Divergent Non-native Amino Acids. *ACS Chem. Biol.* **2011**, *6*, 733–743. [[CrossRef](#)]
33. 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](#)]
34. Wang, Y.S.; Russell, W.K.; Wang, Z.; Wan, W.; Dodd, L.E.; Pai, P.J.; Russell, D.H.; Liu, W.R. The de novo engineering of pyrrolysyl-tRNA synthetase for genetic incorporation of l-phenylalanine and its derivatives. *Mol. Biosyst.* **2011**, *7*, 714–717. [[CrossRef](#)] [[PubMed](#)]
35. Liu, J.; Cheng, R.; Wu, H.; Li, S.; Wang, P.G.; DeGrado, W.F.; Rozovsky, S.; Wang, L. Building and Breaking Bonds via a Compact S-Propargyl-Cysteine to Chemically Control Enzymes and Modify Proteins. *Angew. Chem. Int. Ed.* **2018**, *57*, 12702–12706. [[CrossRef](#)] [[PubMed](#)]
36. 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](#)]
37. Mukai, T.; Hoshi, H.; Ohtake, K.; Takahashi, M.; Yamaguchi, A.; Hayashi, A.; Yokoyama, S.; Sakamoto, K. Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon. *Sci. Rep.* **2015**, *5*, 9699. [[CrossRef](#)]
38. 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](#)]

39. O'Donoghue, P.; Prat, L.; Heinemann, I.U.; Ling, J.; Odoi, K.; Liu, W.R.; Söll, D. Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNA^{Pyl} for genetic code expansion. *FEBS Lett.* **2012**, *586*, 3931–3937. [[CrossRef](#)]
40. Ai, H.W.; Shen, W.; Brustad, E.; Schultz, P.G. Genetically encoded alkenes in yeast. *Angew. Chem. Int. Ed.* **2010**, *49*, 935–937. [[CrossRef](#)]
41. Wang, Y.; Chen, X.; Cai, W.; Tan, L.; Yu, Y.; Han, B.; Li, Y.; Xie, Y.; Su, Y.; Luo, X.; et al. Expanding the Structural Diversity of Protein Building Blocks with Noncanonical Amino Acids Biosynthesized from Aromatic Thiols. *Angew. Chem.* **2021**, *133*, 10128–10136. [[CrossRef](#)]
42. 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](#)] [[PubMed](#)]
43. Dominy, C.N.; Andrews, D.W. Site-Directed Mutagenesis by Inverse PCR. In *E. coli Plasmid Vectors*; Humana Press: Totowa, NJ, USA, 2003; pp. 209–224.
44. Nov, Y. When second best is good enough: Another probabilistic look at saturation mutagenesis. *Appl. Environ. Microbiol.* **2012**, *78*, 258–262. [[CrossRef](#)] [[PubMed](#)]
45. Marchand, J.A.; Neugebauer, M.E.; Ing, M.C.; Lin, C.-I.; Pelton, J.G.; Chang, M.C.Y. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **2019**, *567*, 420–424. [[CrossRef](#)] [[PubMed](#)]
46. Boutureira, O.; Bernardes, G.J.L. Advances in chemical protein modification. *Chem. Rev.* **2015**, *115*, 2174–2195. [[CrossRef](#)] [[PubMed](#)]
47. Chen, W.-N.; Kuppan, K.V.; Lee, M.D.; Jaudzems, K.; Huber, T.; Otting, G. O-tert -Butyltyrosine, an NMR Tag for High-Molecular-Weight Systems and Measurements of Submicromolar Ligand Binding Affinities. *J. Am. Chem. Soc.* **2015**, *137*, 4581–4586. [[CrossRef](#)]
48. Mishra, P.K.; Yoo, C.; Hong, E.; Rhee, H.W. Photo-crosslinking: An Emerging Chemical Tool for Investigating Molecular Networks in Live Cells. *ChemBioChem* **2020**, *21*, 924–932. [[CrossRef](#)]
49. Kim, G.; Weiss, S.J.; Levine, R.L. Methionine oxidation and reduction in proteins. *Biochim. Biophys. Acta Gen. Subj.* **2014**, *1840*, 901–905. [[CrossRef](#)]
50. Wang, B.; Lodder, M.; Zhou, J.; Baird, T.T.; Brown, K.C.; Craik, C.S.; Hecht, S.M. Chemically Mediated Site-Specific Cleavage of Proteins. *J. Am. Chem. Soc.* **2000**, *122*, 7402–7403. [[CrossRef](#)]
51. Baird, T.; Wang, B.; Lodder, M.; Hecht, S.M.; Craik, C.S. Generation of Active Trypsin by Chemical Cleavage. *Tetrahedron* **2000**, *56*, 9477–9485. [[CrossRef](#)]
52. Wang, B.; Brown, K.C.; Lodder, M.; Craik, C.S.; Hecht, S.M. Chemically Mediated Site-Specific Proteolysis. Alteration of Protein–Protein Interaction†. *Biochemistry* **2002**, *41*, 2805–2813. [[CrossRef](#)] [[PubMed](#)]
53. Liutkus, M.; Fraser, S.A.; Caron, K.; Stigers, D.J.; Easton, C.J. Peptide Synthesis through Cell-Free Expression of Fusion Proteins Incorporating Modified Amino Acids as Latent Cleavage Sites for Peptide Release. *ChemBioChem* **2016**, *17*, 908–912. [[CrossRef](#)] [[PubMed](#)]
54. Nolsøe, J.M.J.; Hansen, T.V. Asymmetric Iodolactonization: An Evolutionary Account. *Eur. J. Org. Chem.* **2014**, *2014*, 3051–3065. [[CrossRef](#)]
55. Kristianslund, R.; Tungen, J.E.; Hansen, T.V. Catalytic enantioselective iodolactonization reactions. *Org. Biomol. Chem.* **2019**, *17*, 3079–3092. [[CrossRef](#)] [[PubMed](#)]
56. Lodder, M.; Golovine, S.; Laikhter, A.L.; Karginov, V.A.; Hecht, S.M. Misacylated Transfer RNAs Having a Chemically Removable Protecting Group. *J. Org. Chem.* **1998**, *63*, 794–803. [[CrossRef](#)] [[PubMed](#)]
57. Gross, E.; Witkop, B. Nonenzymatic cleavage of peptide bonds: The methionine residues in bovine pancreatic ribonuclease. *J. Biol. Chem.* **1962**, *237*, 1856–1860. [[CrossRef](#)]
58. Jiang, H.; Zhang, X.; Chen, X.; Aramsangtienchai, P.; Tong, Z.; Lin, H. Protein Lipidation: Occurrence, Mechanisms, Biological Functions, and Enabling Technologies. *Chem. Rev.* **2018**, *118*, 919–988. [[CrossRef](#)]
59. Chen, J.J.; Boehning, D. Protein lipidation as a regulator of apoptotic calcium release: Relevance to cancer. *Front. Oncol.* **2017**, *7*, 138. [[CrossRef](#)]
60. Coleman, D.T.; Gray, A.L.; Kridel, S.J.; Cardelli, J.A. Palmitoylation regulates the intracellular trafficking and stability of c-Met. *Oncotarget* **2016**, *7*, 32664–32677. [[CrossRef](#)]
61. Hang, H.C.; Linder, M.E. Exploring Protein Lipidation with Chemical Biology. *Chem. Rev.* **2011**, *111*, 6341–6358. [[CrossRef](#)]
62. Bech, E.M.; Pedersen, S.L.; Jensen, K.J. Chemical Strategies for Half-Life Extension of Biopharmaceuticals: Lipidation and Its Alternatives. *ACS Med. Chem. Lett.* **2018**, *9*, 577–580. [[CrossRef](#)]
63. Acevedo-Rocha, C.G.; Geiermann, A.-S.; Budisa, N.; Merkel, L. Design of protein congeners containing β-cyclopropylalanine. *Mol. Biosyst.* **2012**, *8*, 2719. [[CrossRef](#)]
64. Budisa, N.; Steipe, B.; Demange, P.; Eckerskorn, C.; Kellermann, J.; Huber, R. High-level Biosynthetic Substitution of Methionine in Proteins by its Analogs 2-Aminohexanoic Acid, Selenomethionine, Telluromethionine and Ethionine in Escherichia coli. *Eur. J. Biochem.* **1995**, *230*, 788–796. [[CrossRef](#)]
65. Kiick, K.L.; Saxon, E.; Tirrell, D.A.; Bertozzi, C.R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 19–24. [[CrossRef](#)]

-
66. Tanrikulu, I.C.; Schmitt, E.; Mechulam, Y.; Goddard, W.A.; Tirrell, D.A. Discovery of *Escherichia coli* methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15285–15290. [[CrossRef](#)]
 67. Wang, Z.A.; Kurra, Y.; Wang, X.; Zeng, Y.; Lee, Y.; Sharma, V.; Lin, H.; Dai, S.Y.; Liu, W.R. A Versatile Approach for Site-Specific Lysine Acylation in Proteins. *Angew. Chem. Int. Ed.* **2017**, *56*, 1643–1647. [[CrossRef](#)]
 68. Salaün, J. Cyclopropane Derivatives and their Diverse Biological Activities. *Top. Curr. Chem.* **2000**, *207*, 1–67.
 69. Geigert, J.; Neidleman, S.L.; Dalietos, D.J. Novel haloperoxidase substrates. Alkynes and cyclopropanes. *J. Biol. Chem.* **1983**, *258*, 2273–2277. [[CrossRef](#)]
 70. Dalton, H.; Golding, B.T.; Waters, B.W.; Higgins, R.; Taylor, J.A. Oxidations of cyclopropane, methylcyclopropane, and arenes with the mono-oxygenase system from *Methylococcus capsulatus*. *J. Chem. Soc. Chem. Commun.* **1981**, *28*, 482. [[CrossRef](#)]