"Cold" Orthogonal Translation by Psychrophilic Pyrrolysyl-tRNA Synthetase Boosts Genetic Code Expansion

Nikolaj G. Koch[1,2,6], Peter Goettig[3], Juri Rappsilber[2,5]*, Nediljko Budisa[4]*

[1] Bioanalytics Group, Institute of Biotechnology, Technische Universität Berlin, 10623 Berlin, Germany

[2] Biocatalysis Group, Institute of Chemistry, Technische Universität Berlin, 10623 Berlin, Germany

[3] Structural Biology Group, Department of Biosciences, University of Salzburg, 5020 Salzburg, Austria

[4] Chemical Synthetic Biology Group, Department of Chemistry, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

[5]Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom.

[6]Current address: Department of Chemistry, Institute of Physical Chemistry, University of Basel, 4058 Basel, Switzerland; Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland

*Correspondence to: Nediljko.budisa@tu-berlin.de or Nediljko.Budisa@umanitoba.ca

Abstract

Using orthogonal translation systems (OTSs) is the most efficient way to produce unnatural proteins by adding non-canonical amino acids (ncAAs) to the genetic code. In the quest to expand substrate specificity the conventional approach begins with a (hyper-)stable enzyme capable of withstanding the structural changes resulting from necessary mutations. However, we here take a radically different position by starting with an enzyme that has evolved to cope with instability, and thus may exhibit greater resilience to mutations. By engineering a psychrophilic ("cold") OTS from *Methanococcoides burtonii* we developed an alternative to the commonly used mesophilic and thermophilic systems. This OTS showed remarkable properties in terms of *in vivo* efficiency and promiscuity, even at very low ncAA concentrations. Given the broad range of applicable host organisms, we anticipate that Cold-OTS will greatly facilitate the transformation of the expanded genetic code from an academic discipline into a high-value chemistry-driven biotechnology.

Introduction

New-to-nature non-canonical amino acids (ncAAs) expand the functional repertoire of proteins. [1–7] Their specific incorporation relies on orthogonal translation systems (OTSs) made up of engineered orthogonal aminoacyl-tRNA synthetase/tRNA pairs. More than 400 ncAAs have been incorporated into proteins via various genetic code expansion routes.[8–11] The most common site-specific incorporation method is amber stop codon suppression (SCS).[1–7,12] For SCS, the ncAA is incorporated in response to an in-frame stop codon placed at a predefined position in the target protein coding sequence, expressed ribosomally either *in vivo* or *in vitro*. Most OTSs used for SCS are derived from the *Methanosarcina mazei/barkeri* pyrrolysyl-tRNA synthetases (*Mm*PylRS/*Mb*PylRS) or the *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS).[8–10]

The ideal OTS exhibits both high catalytic efficiency and sufficient versatility or promiscuity, making it amenable to easy engineering with only a few mutations. The known PylRS variants fulfil the latter point but lack the former one.[13] Many efforts have been made to progressively increase the PylRS-based OTS efficiency for wild-type and engineered systems. These attempts include optimizing OTS plasmid copy number and promotor strength of aaRSs and/or tRNA genes[14], engineering of $tRNA^{Pyl}$ [15], rational and semi-rational engineering of aaRS variants [16], optimizing sequence context around the target codon[17], host cell engineering[18], engineering of parts of the translational machinery (e.g., elongation factor TU)[19] and increasing the solubility of the aaRS[20]. Despite these approaches, the yield of target protein production always falls short of that of the wild-type reference protein.

To shift the natural substrate recognition of PylRS, or any other OTS, to a desired ncAA the enzyme must be engineered. A general strategy in enzyme engineering, including the GCE field, is to use (hyper-)thermophilic enzymes as starting scaffold because they provide the required stability to tolerate the necessary mutations to remodel the active site. For example the commonly used hyperthermophilic *Mj*TyrRS has highly specific substrate interactions demanding significant number of mutations, i.e. up to 10 mutations to acquire new substrate specificities.[21] Unfortunately, these circumstances complicate engineering and reduce the likelihood and ease of finding the desired variants.[21,22] Another drawback of these thermophilic enzymes is that they often exhibit lower catalytic activity under standard cultivation conditions in comparison to their mesophilic homologs, requiring additional engineering for efficiency.[22,23]

Fortunately, the unique structure of Pyl led to little evolutionary pressure for the PylRS to acquire a highly specific recognition mechanism and consists of relatively nonspecific hydrophobic interactions.[8] During engineering campaigns, this recognition mode often leads to a high probability of encoding new substrates with very few mutations, usually requiring only one to four mutations (never more than three in this study).[8,24] This unique property among aaRSs eliminates the need for an extremely stable starting scaffold. Given these circumstances we hypothesized that the PylRS OTS could enable a fundamental shift away from the prevalent engineering approach to genetic code expansion. This advancement entails moving away from using mesophilic and thermophilic aaRSs.

Here, we use psychrophilic homologs of the PylRS system to achieve higher activity under standard cultivation conditions.[23,25,26] We show that the *Methanococcoides burtonii* (*M. burtonii*) PylRS (*Mbur*PylRS) exhibits superiority by outperforming other variants in terms of both *in vivo* incorporation efficiency and scope of substrate recognition.

Results

To overcome the current PylRS efficiency limitations, we turned our attention towards utilizing psychrophilic enzyme scaffolds. We reasoned that enzymes derived from cold adapted organisms would exhibit higher enzyme activity when used as an OTS in species with ambient temperatures. To determine which PylRS to selected we looked at the optimal growth temperature (OGT, **Figure 1A**) for the selected set of organisms (for a detailed discussion, see supplement section 1). Four variants from host organisms with an OGT of 26°C or lower were encompassed in the analysis for the psychrophilic PylRS set (additional information to the organisms can be found in the supplements, **Table S3**). We

selected six different Pyl analogs (**1**, **2**, **3**, **4**, **5**, **6**) with known differences in their recognition efficiencies to analyze their PylRS activities (**Figure 1C**). These substrates are classified into three groups and range from good (**1**, **2**)[27,28] to moderate (**3**, **4**, **5**[29]) to poor (**6**; very low activity with 3.5 mM supplied[28]). We have previously shown that adding an SmbP-tag to wild-type *Mb*PylRS improves its activity, therefore this construct was used throughout this study.[20]

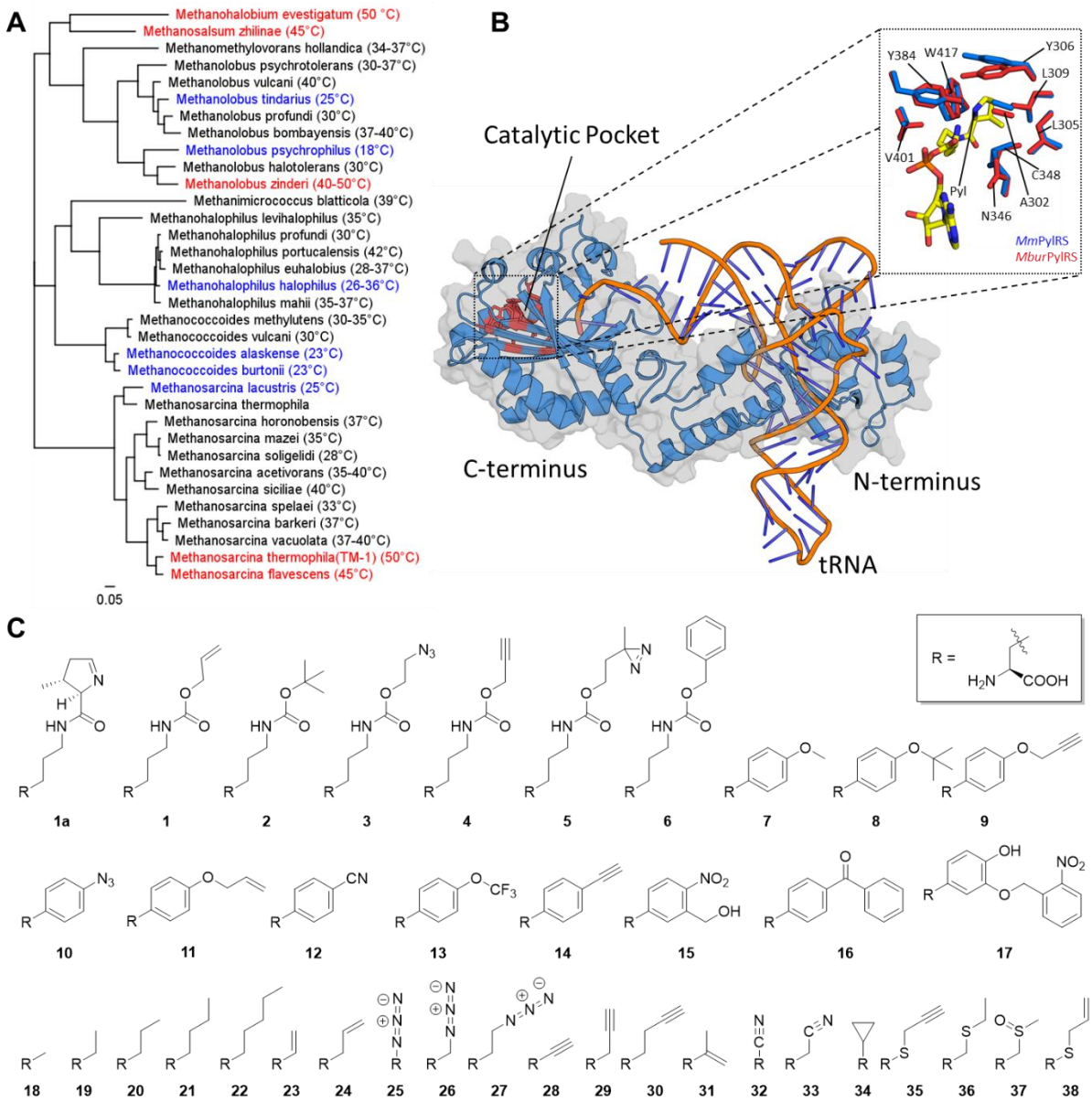

Figure 1. Coupling evolutionary phylogeny with structural biology and non-canonical amino acid chemistries led to the discovery of a novel psychrophilic orthogonal translation system ("Cold-OTS") based on PylRS. **A**) Phylogenetic tree based on all N-terminus containing PylRS enzymes of *Methanosarcinales* and *Methanomassiliicoccales* generated using the Approximate-Maximum-Likelihood method. The FastTree algorithm (version 2.1.11) was used to build the tree based sequences available in NCBI and *M. alaskense* from the JGI database.[30] The number(s) in parentheses represents the optimal growth temperature (OGT). Coloring indicates the origin of the psychrophilic (blue, equal to and below 26°C) and thermophilic (red, equal to and above 45°C) PylRSs. The scale bar indicates 5% sequence divergence. **B**) The 3D model structure (shown as cartoon) of *M. burtonii* PylRS (cyan, without linker) was created by aligning the N- and C-termini with a superposition of the corresponding termini of *M.mazei* (PDB ID 5UD5) and *Desulfitobacterium hafniense* (PDB ID 2ZNI) (see Supplementary Material for detailed discussion). The catalytic pocket with bound Pyl-AMP (yellow) is from *M. mazei* (PDB ID: 2Q7H) while the *M. burtonii* structure was calculated with ColabFold.[24,31] The conserved active site residues are shown as sticks, with *M. burtonii* PylRS in red and *M. mazei* PylRS in blue (*M. mazei* notation). For clarity and because the relative position is not clear, the linker region has been omitted but can be found in the Supplementary Material. **C**) Non-canonical amino acid substrates used in this study. The systematic and trivial names of these ncAAs are listed in the Supplementary Material (**Table S1**). The compounds **1**, **3**, **4**, **9**-**11**, **14**, **23**-**30**, **35**, and **38** are suitable for click reactions either with 1,3 dipolar cycloaddition, inverse electron-demand Diels–Alder or the photo-catalyzed thiol-ene coupling reaction.[32]

All the selected PylRS OTS were functional under standard cultivation conditions with varying degrees of activity (**Figure 2**). As expected, the bad substrate **6** lead to minimal activity at the highest ncAA concertation (9 mM, see **Figure S90**). Encouragingly, none of the psychrophilic PylRSs showed significantly higher background incorporation, indicating that they can discriminate against canonical amino acids (cAAs) with the same fidelity as the commonly used PylRSs. Interestingly, within the four tested psychrophilic PylRS the *Mbur*PylRS is clearly the outlier, since the others perform worse. *Mp*PylRS was hardly functional, with signals barley above the background suppression. Notably, *Mbur*PylRS consistently demonstrated superior OTS efficiency for any substrates at any concentration, this can also be seen from the promiscuity scores, which are discussed later (vide infra). For the two mesophilic PylRS, a mixed picture emerged. The *Mm*PylRS exhibits higher efficiency than the SmbP-*Mb*PylRS at low ncAA concentrations, which is consistent with the literature.[16] On the other hand, SmbP-*Mb*PylRS outperformed the *Mm*PylRS at high concentrations (3 and 9 mM) of ncAA **3**. This interesting behavior is picked up later. The thermophilic *Mt*PylRS(TM-1) performs similar to the *Mm*PylRS, at ncAA concentrations of 1 mM and above. In contrast to highly thermophilic enzymes, which often show no or very low activity at 37 °C, this finding is plausible since *Mt*PylRS(TM-1) originates from an organism that is only mildly thermophilic.

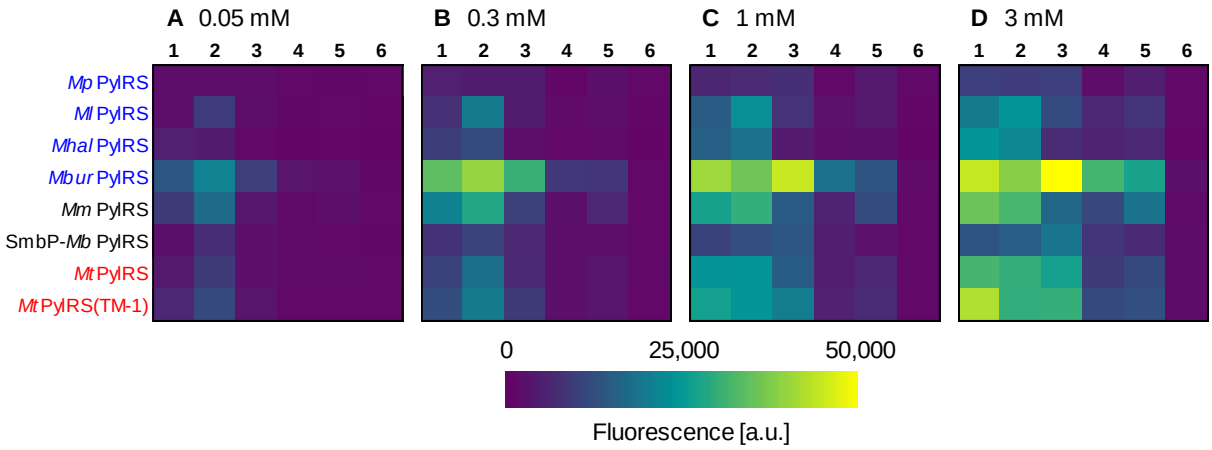

Figure 2. Heatmaps based on fluorescence of sfGFP(1x amber) expression with *E. coli* BL21(DE3) for the eight selected PylRS as a function of different ncAA concentrations; **A)**, **B)**, **C)**, **D)**, are 0.05, 0,3, 1, 3 mM ncAA supplied. The maps for 0.1 and 9 mM are in the supplements (**Figure S90**). Bar charts with the number of replicates and error bars with the standard deviation can be found in the Supplementary Material (**Figures S17-S34**). Substrate numbering corresponds to that in **Figure 1**. Background suppressions were between 900 and 1,300 [a.u.] depending on the construct.

The maximal efficiency of *Mbur*PylRS for substrate **2** is reached already at 0.3 mM. In general, *Mbur*PylRS shows superior performance at lower ncAA concentration indicating a more efficient OTS. For example, for substrate **2** the performance at 0.05 mM is 30% higher, while at 0.3 and 1 mM the *Mbur*PylRS shows between 300% (for **3**) and 400% (for **4**) higher efficiency than the second-best variant.

To gain insight on how the performance of *Mbur*PylRS compares to the currently most efficient OTS known for *E. coli*, a side-by-side comparison with the *Mj*TyrRS OTS was performed (**Figure 3**).[21,33,34] The comparison was performed with two *Mj*TyrRS variants after a substrate pre-selection to see which were the best substrates (**Figure S12** and **S13**).The *Mj*ONBYRS was developed by our group and was characterized to be extremely efficient.[33] The other one, *Mj*pCNFRS, was engineered to recognize *para*-cyano-phenylalanine and turned out to recognize various para-substituted phenylalanine-derivatives[35] For the PylRS set, we selected the two most efficient *Mbur*PylRS (**Figure 3C** and **G**) to showcase the maximum efficiency. We also show a representative pair of *Mbur*PylRS and *Mm*PylRS for substrate **3** as an example for the efficiency differences (**Figure 3D** and **H**). All OTSs were tested in the same pTECH plasmid setup. For reference, the reporter signals were normalized to the corresponding wild-type sfGFP signal (sfGFP without in-frame stop codons). Overall, the best *Mj*TyrRS (**Figure 3A** and **B**) and *Mbur*PylRS (**Figure 3C** and **G**) systems a performing on par. While it was occasionally observed that wild-type PylRS systems can perform on this level (here judged by proteins yields[36]) we are the first

to report a mutated PylRS OTS that incorporates a non-Pyl derivate, like **38**, with this high efficiency (**Figure 3G**), even at very low ncAA concentrations (0.3 mM **38**, **Figure 4E-F**). We are also the first to report a true side-by-side comparison of *Mj*TyrRS and PylRS in the identical plasmid setup, a setup especially designed for maximum performance due to the constitutive lpp promotor.

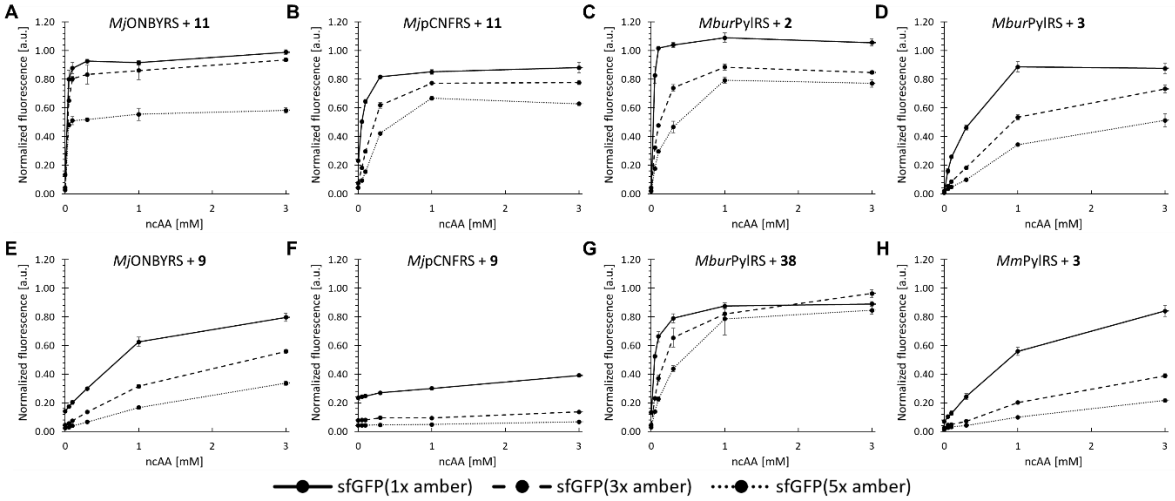

Figure 3. Concentration-dependent unnatural protein production by **A**) *Mj*ONBYRS with substrate **11**, **B**) *Mj*pCNFRS with substrate **11**, **C**) *Mbur*PylRS with substrate **2**, **D**) *Mbur*PylRS with substrate **3**, **E**) *Mj*ONBYRS with substrate **9**, **F**) *Mj*pCNFRS with substrate **9**, **G**) *Mbur*PylRS variants with tRNA from *M. alaskense* and substrate **38**, **H**) *Mm*PylRS with substrate **3**. The host for protein production was RF1 deficient *Escherichia coli* B-95.ΔA^{37}. Endpoint measurements for ncAA concentrations of 0.05, 0.1, 0.3, 1 and 3 mM. Fluorescence values were normalized to those for wild-type sfGFP reporter constructs (without in-frame stop codon), error bars represent the standard deviation of three replicates.

Towards the end of our study, the genome of the psychrophilic organism *M. alaskense* became available, prompting us to test its use as an OTS (**Figure S15**). Remarkably, pairing *Mbur*PylRS with *M. alaskense* $tRNA^{Pyl}$ increased efficiency up to 40% for some substrates (**Figure S15** and **S16**). Therefore, we also used this combination in **Figure 3G**.

To validate the observations made in small-scale fluorescent assays, we conducted scaled-up protein production experiments in shake flasks with subsequent protein purification (**Figure 4**). The protein yields obtained agreed well with the trends observed in the small-scale fluorescence experiments, confirming their reliability. However, one exception was observed for incorporation of **35**, where the yield of target protein with *Mbur*PylRS exceeded our expectations (**Figure S11** and **Figure 4E-4G**). Still within this dataset the fluorescence values and protein yields correlate well. While the improvements for the incorporation of one in-frame ncAA ranged between 25% (for 1 mM **2**) and 200% (for 1 mM **3**) (**Figure 4A-4D**), the most remarkable improvements were observed in multi-site incorporation of five **38** and **35**, with an improvement of over 400% and 300%, respectively (**Figure 4G**). A clear trend is observable that the difference in efficiency between SmbP-*Mb*PylRS and *Mbur*PylRS increases the more ncAAs are incorporated within a target protein (**Figure 4E-4G**).

To estimate the robustness of the 96-well plate assays, we analyzed the small scale 96-well plate sfGFP fluorescence data in correlation to the isolated protein yields of the shaking flask cultures and derived the R^2 values (see supplements **Figure S89**). The analysis reveals that the correlation between the 96-well plate assays for each ncAA is notably strong, with R^2 values consistently exceeding 0.97. Even when combining the data of cultivations with different ncAAs, the R^2 values remain high, reaching at least 0.85, indicating a strong correlation. The combined R^2 values are lower in comparison to values from just one ncAA, this suggests that protein yields are influenced by factors beyond simply ncAA concentration, potentially ncAA toxicity leading to growth impairment. The high correlation is not unexpected, given that our 96-well plate experiments consistently operated within the expected linear range of the potential measurement space, with cell densities above OD600 = 1 never observed. Collectively, this data robustness indicates that protein yields should be translatable to shake flaks cultivations.

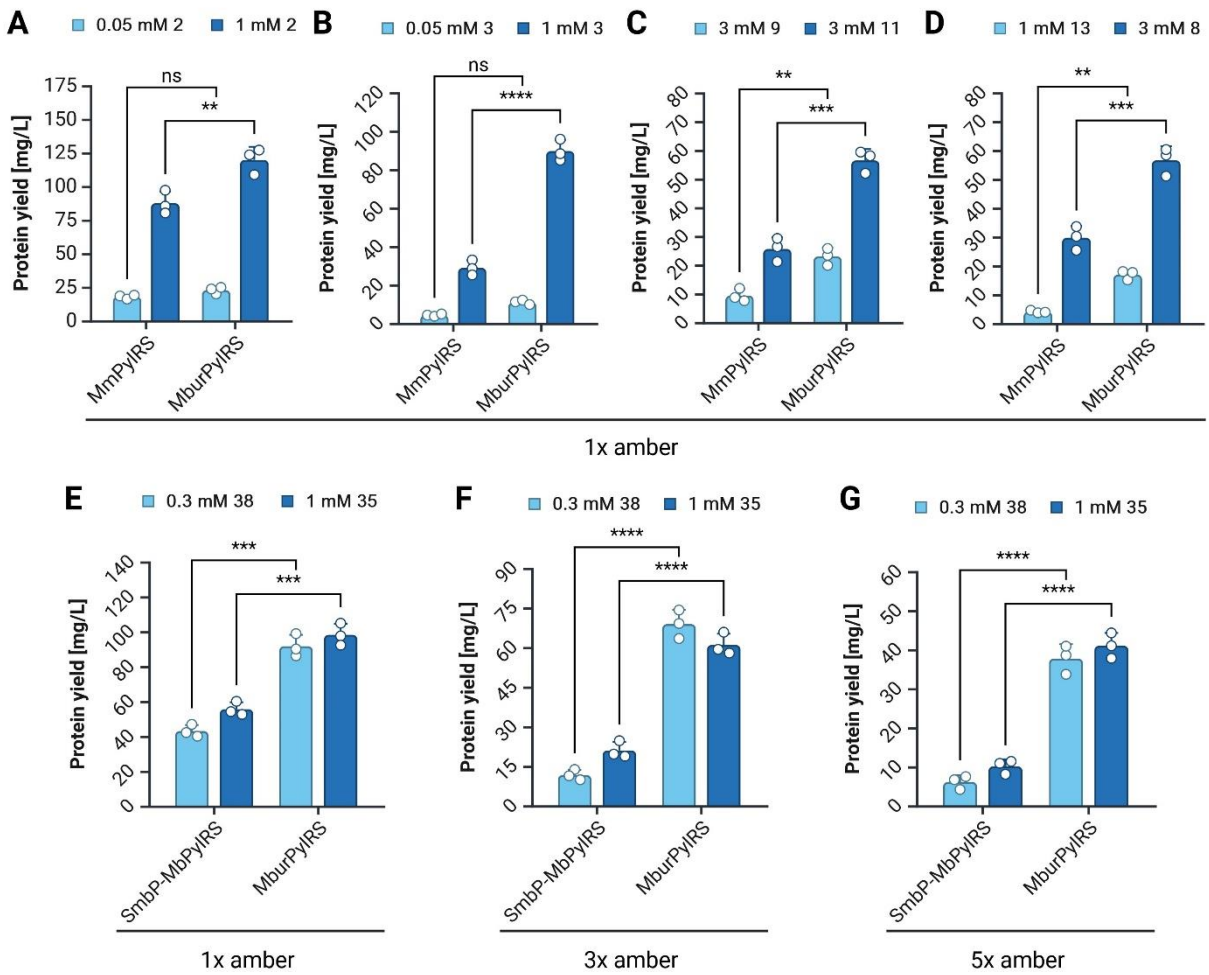

Figure 4. Protein yields of sfGFP from small-scale shake flask cultures. The protein yield is given in protein per liter of cell culture. In the upper row BL21(DE3) was used with indicated reporter. In the bottom row B-95.ΔA was used with indicated reporter. Substrate concentration is in the legends and numbering corresponds to that in **Figure 1**. For detailed information about the PylRS variants used and ESI-MS of purified proteins see **Table S1** and **S2**. Statistical analysis performed with a Two-way ANOVA with Tukey multiple comparisons test, significance is represented as stars.

Our fluorescence heatmaps (**Figure 2**) indicated that the *Mbur*PylRS exhibits higher versatility to efficiently incorporate the given set of ncAAs compared to other variants. To quantify this promiscuity, we derived corresponding promiscuity scores. Traditionally, enzyme promiscuity is determined by comparing *in vitro* kinetic parameters for new analogous substrates with those of the native substrate.[38] However, since OTSs are way to complex systems that *in vitro* parameters could give clear-cut information in regard to *in vivo* efficiency[15], we transferred the concept of *in vitro* enzyme promiscuity determination to cell-based *in vivo* scores (**Figure 5A**). The incorporation efficiencies for all substrates were correlated (normalized) to the best-performing substrate for each ncAA concentration and then mapped to a scale from 0 to 100 to yield the substrate promiscuity scores. A perfectly promiscuous enzyme would attain a value of 100, meaning that all ncAAs are incorporated to the same extent. Conversely, for a perfectly specific enzyme recognizing only one substrate, the lowest possible score is 0. The lowest score in Figure 3 is 14.5 attained by the *Ml*PylRS at 0.1 mM. While this score may just give an apparent promiscuity score biased by different in-cell ncAA concentrations, the score still gives meaningful information for this PylRS-ncAA set, since all other conditions are the same for each ncAA setup. Data on in-cell ncAA concentrations are sparse but available for **2** and **3**. Both show over 90% cell uptake from the medium.[39] With **2** being very hydrophobic and **3** very hydrophilic this gives a good estimate how the rest of the ncAA set is most probably taken up.

*Mp*PylRSs showed the highest promiscuity score for low ncAA concentrations, but was unfortunately also barely functional at these concentrations (**Figure 2**). There were no clear trends were observed regarding the promiscuity and thermal origin of the different PylRSs. Expectedly, it was observed that

the promiscuity score increases with the increase of ncAA concentrations (**Figure 5A**). This is logical as it can be observed in our data (**Figure 2** and **Figure S90**) that good and bad substrates reach the activity plateau at different increasing concentrations. So, in our setup a high promiscuity score at low ncAA concentrations is indicates higher general promiscuity since this plateau is not reached yet, skewing the signal. We observed that *Mbur*PylRS and SmbP-*Mb*PylRS showed the highest scores at low ncAA concentration (0.3 mM, **Figure 5A**, red box). As enzyme promiscuity remains unpredictable but nevertheless a very important property in the enzyme engineering field[40,41] our promiscuity coefficients might be an applicable indicator of the engineering potential of a PylRS variant. To see if there is merit to this assumption, we took the two OTS with the highest promiscuity score at 0.3 mM ncAA concentration, *Mbur*PylRS and SmbP-*Mb*PylRS, and compared their performance to the *Mm*PylRS, which is a very commonly used system, and the *Mt*PylRS(TM-1) as an example for a thermophilic PylRS.

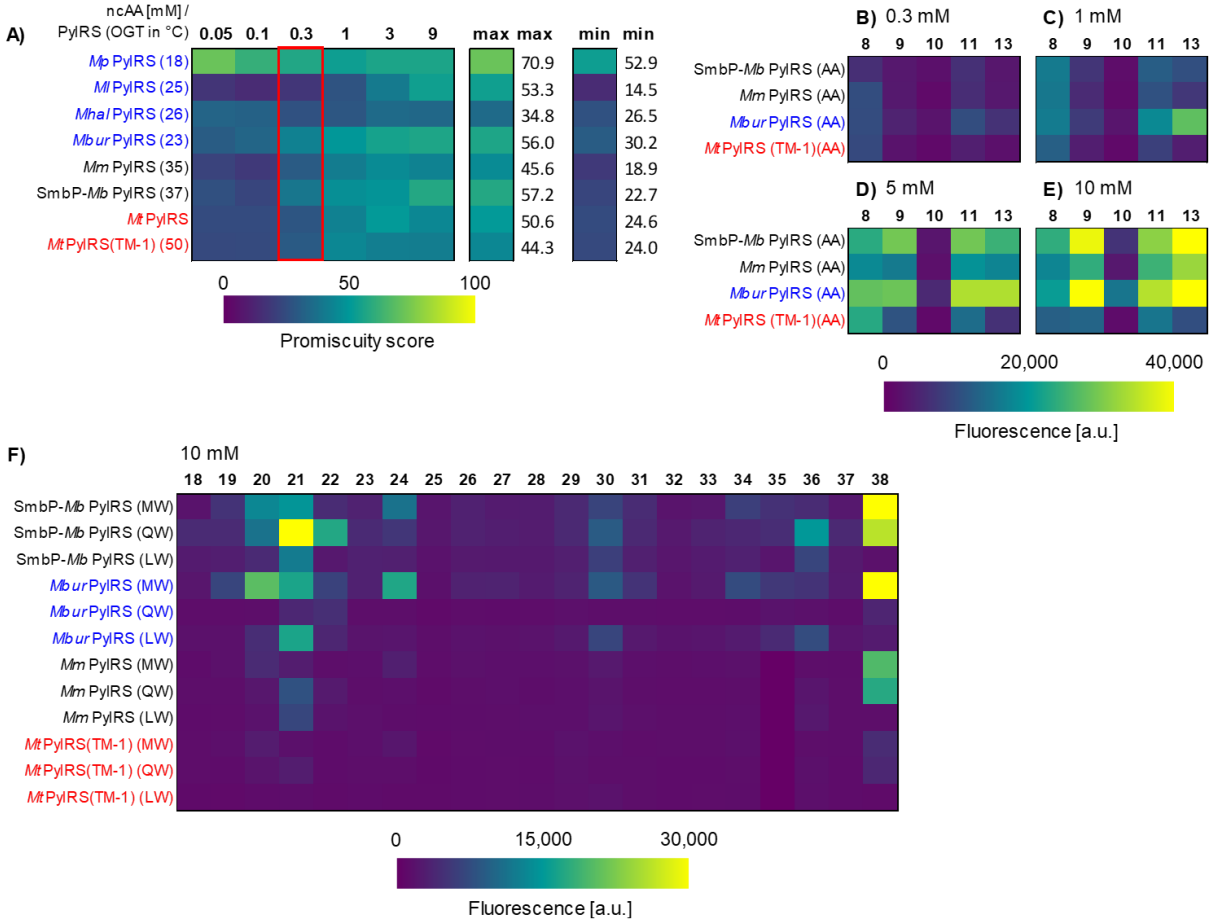

Figure 5. A) Heat maps showing the promiscuity score for each PylRS construct at given ncAA concentrations and the maximum (max) and minimum (min) promiscuity scores in an extra column. In brackets of the constructs is the optimal growth temperature (OGT) from which this PylRS was derived. The ncAA concentration of 0.3 is highlighted, see text for detailed explanation. **B)**, **C)**, **D)**, **E)** are heatmaps based on fluorescence of the sfGFP(1x amber) expression with *E. coli* BL21(DE3). with 0.3, 1, 5 and 10 mM ncAA concentrations. **F)** Are selected PylRS mutants supplied with 10 mM ncAA.Figures showing the number of replicates and error bars are provided in the supplements; for **B)**, **C)**, **D)**, **E)**, **Figure S37-S41** and for **F) Figure S42-S50**. Letters in brackets indicate the N311 and the C313 mutations in *M. barkeri* notation. Substrate numbering can be found in **Figure 1**. Growth for the *Mm*PylRS and *Mt*PylRS(TM-1) constructs with substrate **35** was extremely low, so these values were excluded. Background suppressions were construct dependent and ranged from 1,100 to 2,500 [a.u.].

We tested different PylRS mutants for two additional ncAA classes, the one class being Phe-analogs (**7**-**17**) and the other class being small aliphatic ncAAs (**18**-**38**) we identified in previous works.[10] After a pre-screen of the Phe-analogs (**Figure S91**) the ncAA with good activity were screened in a concentration depend manner to gain further insights (**Figure 5B**-**E**). Here, *Mbur*PylRS and SmbP-*Mb*PylRS showed the best performance with *Mbur*PylRS being better at lower ncAA concentrations. *Mbur*PylRS even incorporated **10**, were none of the other PylRS were able to incorporate this substrate. A similar picture emerged for the second class (**Figure 5F**). While most of the tested ncAAs tested did not show super high activity, for the ones that did the *Mbur*PylRS and SmbP-*Mb*PylRS performed clearly

better than the *Mm*PylRS and *Mt*PylRS(TM-1) variants. In summary, our data suggests that the promiscuity score could be useful in guiding future engineering projects, which is also in line with the literature.[40,41]

Discussion

To overcome the low *in vivo* efficiency of the PylRS orthogonal translation system, we conducted a comparative study exploring PylRS homologs derived from extremophilic organisms. Among them, we discovered a highly active variant called *Mbur*PylRS, which originates from the psychrophilic (cold-active) organism *Methanococcoides burtonii*. *Mbur*PylRS exhibited exceptional efficiency in incorporating various ncAAs into proteins.

The superior incorporation efficiency was consistently observed for all major substrate classes, demonstrating the robustness of the results. Particularly remarkable was the multi-site ncAA incorporation with BocK (**2**) and especially Sac (**38**). To the best of our knowledge, no previous PylRS mutant has achieved such high ncAA incorporation efficiency for non-Pyl-analogs, approaching that of the wild-type reference. This finding is significant as it can potentially address the challenges posed by the high cost of certain ncAAs, thus facilitating and accelerating research projects in the field.

Considering that only two OTSs are known to approach wild-type recombinant protein production yields (*Mj*TyrRS[33,34,42] and *Af*TyrRS[42]), the addition of a third OTS to this group is highly desirable. The inclusion of a third orthogonal translation system enhances the potential for simultaneous incorporation of a greater variety of ncAAs, both identical and different ones, with high efficiency. Particularly, the most promising improvements are expected in organisms with liberated codons as there is a high degree of freedom in selection of $tRNA^{Pyl}$ anticodon for orthogonal translation.[43,44] When the *Mm*PylRS is used in a recoded *E. coli* strain, OTS efficiency is higher for some of these newly created codons, but to a limited extent.[44] This result suggests that the catalytic efficiency of the PylRS OTS remains a limiting factor even when free codons are available. *Mbur*PylRS, on the other hand, shows promising potential in overcoming this limitation, as demonstrated by its high performance in a RF1 deficient strain (**Figure 3** and **4**).

Additionally, the high performance of *Mbur*PylRS even at very low ncAA concentrations will pave the way for efficient coupling with ncAA producing metabolic pathways that could end the need for external ncAA supply. Efficient OTS activity with intracellular ncAA concentrations below 1 mM is essential, because in most organisms intracellular canonical amino acid concentrations are below 1 mM and engineering pathways that allow for ncAA concentrations above 1 mM seems very unlikely.[45] High performance paired with low ncAA concentration has further implications. The low price and concentration required for efficient incorporation of Sac (**38**) means that recombinant proteins containing one or even multiple bioorthogonal alkene functions can now be produced at negligible additional cost at wild-type levels. The alkene function can be perfectly addressed with the most advanced bioconjugation reaction, the inverse electron demand Diels–Alder (IEDDA) reaction, using a tetrazine moiety.[46] In general, click reactions will be greatly facilitated as 17 of the 38 incorporated ncAAs (**Figure 1**) are crucial precursors for establishing covalent links in otherwise transient complexes of biomolecules. With some substrates *Mbur*PylRS also outperformed one of the best OTS known, *Mj*TyrRS, in incorporating ncAAs at multiple sites which is particularly important for applications that require multiple efficient orthogonal pairs to produce the desired target-protein.[43,47] The *Mbur*PylRS exhibits great promise as a candidate for further expansion of the ncAA repertoire with additional substrates. This expectation is attributed to its enhanced substrate promiscuity, which increases the likelihood of successful engineering and opens up new possibilities for further advancements in this field.[40,48–50]

Our discoveries highlight a paradigm shift in the search for efficient orthogonal translation systems by focusing on psychrophilic organisms. While previous research in orthogonal translation has primarily centered around engineering thermophilic and mesophilic aminoacyl-tRNA synthetases (aaRS),[51,52] our study demonstrates the potential of psychrophilic enzymes as a valuable source for highly efficient

OTSs. This paradigm shift is supported by the unique properties of psychrophilic enzymes, which offer advantages not only in terms of catalytic activity but also substrate promiscuity and tolerance to active-site mutations. Moreover, these properties extend beyond aaRSs and may have implications for other components of the host translational machinery, such as orthogonal ribosomes and other parts of the translational machinery.[47,53,54] By harnessing the unique advantages of psychrophilic macromolecules, particularly their higher promiscuity, future engineering efforts could be greatly facilitated. This increased promiscuity enhances the likelihood of discovering altered functional engineered versions, ultimately enabling applications previously limited by existing systems.

In our study, we've identified a psychrophilic PylRS enzyme displaying typical properties of psychrophilic enzymes, such as heightened enzyme efficiency and promiscuity. These traits are crucial for synthetic biology and protein engineering, yet these enzymes remain underutilized beyond protease research for industrial purposes. Additionally, enzymes with heightened promiscuity hold significant potential for facilitating enzyme engineering efforts for new substrate recognitions. Therefore, we hope our findings will encourage further exploration and application of psychrophilic enzymes in various scientific and industrial contexts.

Materials and Methods

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, ThermoFisher Scientific and TCI Deutschland (see Table S4).

Plasmid Vector Construction

The pTECH vector was a gift from Dieter Söll (Addgene plasmid #104073).[16] Genes were ordered as codon optimized fragments from Twist Bioscience. All plasmids were assembled by Golden Gate cloning and confirmed by Sanger sequencing. Point mutations were introduced by non-overlapping inverse PCR.

Analysis of SUMO-sfGFP Expression by Intact Cell Fluorescence.

E. coli BL21(DE) cells were used for the small-scale expression of reporter constructs. 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 µL 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 OD_{600} measurements, 50 µL of ZYP-5052 medium was pipetted into clear 96-well µ-plates and 50 µL 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 OD_{600}. 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.

Protein Expression and Purification

For expression of SUMO-sfGFP variants, *E. coli* strains were used in 10 mL ZYP-5052 medium supplemented with indicated ncAA concentrations 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.

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 (ThermoFisher 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) or by gravity flow. 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 1.5 mL (1 mL for gravity flow) 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. I

Intact Protein 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 H_2O; 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).

Acknowledgments

This work was 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) and the Natural Sciences and Engineering Research Council (NSERC) of Canada through the Discovery Grant (RGPIN-05669-2020) 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.). We are very grateful to Christian Stieger (Leibniz-Forschungsinstitut für Molekulare Pharmakologie, Hackenberger Group) for his support in ESI-MS measurements. A portion of the used data were produced by the US Department of Energy Joint Genome Institute (https://ror.org/04xm1d337; operated under Contract No. DE-AC02-05CH11231) in collaboration with the user community. Special thanks go to my bench neighbor Georg Johannes Freiherr von Sass for inspiration to take this journey.

References

1. Drienovská, I., Mayer, C., Dulson, C. & Roelfes, G. SI_A designer enzyme for hydrazone and oxime formation featuring an unnatural catalytic aniline residue. *Nat. Chem.* (2018) doi:10.1038/s41557-018-0082-z.
2. Tang, S. *et al.* Mechanism-based traps enable protease and hydrolase substrate discovery. *Nature* (2022) doi:10.1038/s41586-022-04414-9.
3. Mukai, T., Lajoie, M. J., Englert, M. & Söll, D. Rewriting the Genetic Code. *Annu. Rev. Microbiol.* **71**, 557–577 (2017).
4. Agostini, F. *et al.* Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology.

Angew. Chemie - Int. Ed. **56**, 9680–9703 (2017).
5. Pagar, A. D. *et al.* Recent Advances in Biocatalysis with Chemical Modification and Expanded Amino Acid Alphabet. *Chem. Rev.* **121**, 6173–6245 (2021).
6. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).
7. Young, D. D. & Schultz, P. G. Playing with the Molecules of Life. *ACS Chem. Biol.* **13**, 854–870 (2018).
8. 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 Proteomics* **1844**, 1059–1070 (2014).
9. Dumas, A., Lercher, L., Spicer, C. D. & Davis, B. G. Designing logical codon reassignment-Expanding the chemistry in biology. *Chem. Sci.* **6**, 50–69 (2015).
10. 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.* **22**, 11194 (2021).
11. Icking, L. *et al.* iNClusive: a database collecting useful information on non-canonical amino acids and their incorporation into proteins for easier genetic code expansion implementation. *Nucleic Acids Res.* 1–7 (2023) doi:10.1093/nar/gkad1090.
12. Budisa, N. Prolegomena to Future Experimental Efforts on Genetic Code Engineering by Expanding Its Amino Acid Repertoire. *Angew. Chemie Int. Ed.* **43**, 6426–6463 (2004).
13. O'Donoghue, P., Ling, J., Wang, Y.-S. & Söll, D. Upgrading protein synthesis for synthetic biology. *Nat. Chem. Biol.* **9**, 594–598 (2013).
14. Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H. & Schultz, P. G. A versatile platform for single- and multiple-unnatural amino acid mutagenesis in escherichia coli. *Biochemistry* **52**, 1828–1837 (2013).
15. Fan, C., Xiong, H., Reynolds, N. M. & Söll, D. Rationally evolving tRNA Pyl for efficient incorporation of noncanonical amino acids. *Nucleic Acids Res.* **43**, e156–e156 (2015).
16. Bryson, D. I. *et al.* Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **13**, 1253–1260 (2017).
17. Pott, M., Schmidt, M. J. & Summerer, D. Evolved sequence contexts for highly efficient amber suppression with noncanonical amino acids. *ACS Chem. Biol.* **9**, 2815–2822 (2014).
18. Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science (80-.).* **342**, 357–360 (2013).
19. Park, H. S. *et al.* Expanding the genetic code of Escherichia coli with phosphoserine. *Science (80-.).* **333**, 1151–1154 (2011).
20. Koch, N. G., Baumann, T. & Budisa, N. Efficient Unnatural Protein Production by Pyrrolysyl-tRNA Synthetase With Genetically Fused Solubility Tags. *Front. Bioeng. Biotechnol.* **9**, 1–14 (2021).
21. Hauf, M. *et al.* Photoactivatable Mussel-Based Underwater Adhesive Proteins by an Expanded Genetic Code. *ChemBioChem* **18**, 1819–1823 (2017).
22. Grasso, K. T. *et al.* Structural Robustness Affects the Engineerability of Aminoacyl-tRNA Synthetases for Genetic Code Expansion. *Biochemistry* **60**, 489–493 (2021).
23. Feller, G. & Gerday, C. Psychrophilic enzymes: hot topics in cold adaptation. *Nat. Rev. Microbiol.* **1**, 200–208 (2003).
24. Kavran, J. M. *et al.* Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11268–11273 (2007).
25. Siddiqui, K. S. & Cavicchioli, R. Cold-adapted enzymes. *Annu. Rev. Biochem.* **75**, 403–433 (2006).
26. Feller, G. Psychrophilic Enzymes: From Folding to Function and Biotechnology. *Scientifica (Cairo).* **2013**, 1–28 (2013).
27. Vatansever, E. C. *et al.* A Designed, Highly Efficient Pyrrolysyl-tRNA Synthetase Mutant Binds o-Chlorophenylalanine Using Two Halogen Bonds. *J. Mol. Biol.* **434**, 167534 (2022).
28. Yanagisawa, T. *et al.* Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode Nε-(o-Azidobenzyloxycarbonyl) lysine for Site-Specific Protein Modification. *Chem. Biol.* **15**, 1187–1197 (2008).
29. Chou, C., Uprety, R., Davis, L., Chin, J. W. & Deiters, A. Genetically encoding an aliphatic diazirine for protein photocrosslinking. *Chem. Sci.* **2**, 480–483 (2011).
30. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **5**, e9490 (2010).
31. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
32. Boutureira, O. & Bernardes, G. J. L. Advances in chemical protein modification. *Chem. Rev.* **115**, 2174–2195 (2015).

33. Baumann, T. *et al.* Computational aminoacyl-tRNA synthetase library design for photocaged tyrosine. *Int. J. Mol. Sci.* **20**, (2019).
34. Amiram, M. *et al.* Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. *Nat. Biotechnol.* **33**, 1272–1279 (2015).
35. Young, D. D. *et al.* An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **50**, 1894–1900 (2011).
36. Muzika, M. *et al.* Chemically-defined lactose-based autoinduction medium for site-specific incorporation of non-canonical amino acids into proteins. *RSC Adv.* **8**, 25558–25567 (2018).
37. Mukai, T. *et al.* Highly reproductive Escherichia coli cells with no specific assignment to the UAG codon. *Sci. Rep.* **5**, 9699 (2015).
38. KHERSONSKY, O., ROODVELDT, C. & TAWFIK, D. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* **10**, 498–508 (2006).
39. Galindo Casas, M., Stargardt, P., Mairhofer, J. & Wiltschi, B. Decoupling Protein Production from Cell Growth Enhances the Site-Specific Incorporation of Noncanonical Amino Acids in E. coli. *ACS Synth. Biol.* **9**, 3052–3066 (2020).
40. Leveson-Gower, R. B., Mayer, C. & Roelfes, G. The importance of catalytic promiscuity for enzyme design and evolution. *Nat. Rev. Chem.* **3**, 687–705 (2019).
41. Nobeli, I., Favia, A. D. & Thornton, J. M. Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol.* **27**, 157–167 (2009).
42. Cervettini, D. *et al.* Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase–tRNA pairs. *Nat. Biotechnol.* **38**, 989–999 (2020).
43. Robertson, W. E. *et al.* Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science (80-.).* **372**, 1057–1062 (2021).
44. Fischer, E. C. *et al.* New codons for efficient production of unnatural proteins in a semisynthetic organism. *Nat. Chem. Biol.* **16**, 570–576 (2020).
45. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat. Chem. Biol.* **5**, 593–599 (2009).
46. Oliveira, B. L., Guo, Z. & Bernardes, G. J. L. Inverse electron demand Diels–Alder reactions in chemical biology. *Chem. Soc. Rev.* **46**, 4895–4950 (2017).
47. de la Torre, D. & Chin, J. W. Reprogramming the genetic code. *Nat. Rev. Genet.* **22**, 169–184 (2021).
48. Zeymer, C. & Hilvert, D. Directed Evolution of Protein Catalysts. *Annu. Rev. Biochem.* **87**, 131–157 (2018).
49. Tóth-Petróczy, Á. & Tawfik, D. S. The robustness and innovability of protein folds. *Curr. Opin. Struct. Biol.* **26**, 131–138 (2014).
50. Tokuriki, N. & Tawfik, D. S. Protein Dynamism and Evolvability. *Science (80-.).* **324**, 203–207 (2009).
51. Nandanwar, S. K., Borkar, S. B., Lee, J. H. & Kim, H. J. Taking Advantage of Promiscuity of Cold-Active Enzymes. *Appl. Sci.* **10**, 8128 (2020).
52. Jin, X., Park, O.-J. & Hong, S. H. Incorporation of non-standard amino acids into proteins: challenges, recent achievements, and emerging applications. *Appl. Microbiol. Biotechnol.* **103**, 2947–2958 (2019).
53. Fricke, R. *et al.* Expanding the substrate scope of pyrrolysyl-transfer RNA synthetase enzymes to include non-α-amino acids in vitro and in vivo. *Nat. Chem.* **15**, 960–971 (2023).
54. Watson, Z. L. *et al.* Atomistic simulations of the Escherichia coli ribosome provide selection criteria for translationally active substrates. *Nat. Chem.* 1–39 (2023) doi:10.1038/s41557-023-01226-w.