"Cold" Orthogonal Translation: Psychrophilic Pyrrolysyl-tRNA Synthetase as Efficient Tool for Expanding the Genetic Code

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

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

[2] Bioanalytics Group, Institute of Biotechnology, 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.

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

Abstract

Orthogonal translation systems (OTSs) are the most expedient way to generate unnatural proteins by adding non-canonical amino acids (ncAAs) to the genetic code. Diversifying substrate specificity typically starts from a stable enzyme that is capable to withstand the structure-perturbing effects of the required mutations. We here take a radically different position, starting from an enzyme that has evolved to cope with instability and thus may better tolerate mutations. By engineering psychrophilic ("cold") PylRS orthologs we develop "cold" OTS as an alternative to the commonly used mesophilic and thermophilic OTSs. In particular, the psychrophilic PylRS homolog from *Methanococcoides burtonii* (*Mbur*PylRS) showed remarkable properties in terms of catalytic efficiency and promiscuity, even at very low ncAA concentrations. Outstanding multi-site incorporation efficiencies were achieved with N^{ε}-tert-Butoxycarbonyl-L-lysine (BocK) and especially *S*-allyl-L-cysteine in up to five sites. Given the broad range of host organisms (archaea, bacteria, and eukaryotes) in which the system can be used, 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 targeted incorporation relies on orthogonal translation systems (OTSs) made up of engineered orthogonal aminoacyl-tRNA synthetase/tRNA pairs. More than 200 ncAAs have been incorporated into proteins via various genetic code expansion routes.[8–10] The most common site-specific incorporation method is amber stop codon suppression (SCS).[1–7,11] 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).[12] Unlike all other OTSs, the PylRS system is particularly useful because its natural orthogonality allows to be used in all three domains of life and even permits anticodon mutations to other target codons.[8,13]

To shift substrate recognition of PylRS to a desired ncAA, the enzyme must be engineered. In general, top down implementation of new substrate recognition for an enzyme requires at least low levels of the desired new activity[14,15] and sufficient stability to buffer destabilizing mutations for active site remodelling.[16,17] Fortunately, the unique structure of Pyl led to no evolutionary pressure for the PylRS to acquire a highly specific recognition mechanism and consists of relatively nonspecific hydrophobic interactions.[8] During engineering, 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). This property eliminates the need for an extremely stable enzyme.[8,18] In contrast, the commonly used hyperthermophilic *Mj*TyrRS has highly specific substrate interactions demanding significantly more mutations, i.e. up to 10 mutations to acquire new substrate specificities.[19] Unfortunately, this complicates engineering which then requires computational models and also the likelihood and ease of finding the desired variants is reduced.[17,19] While the hyperthermophilic origin provides the necessary stability for these engineering campaigns, these enzymes often exhibit low catalytic activity under standard cultivation conditions, making this approach not generally transferable.[20]

The ideal OTS has high catalytic efficiency and is versatile or promiscuous enough to be easily engineered with only a few mutations. The known PylRS variants fulfil the latter point. However, their low *in vivo* efficiency is a major impeding factor for the usefulness of this OTS, and is even exacerbated in engineered variants.[21] Many efforts have been made to progressively increase the PylRS-based OTS efficiency for wild-type and engineered systems. These include optimizing OTS plasmid copy number and promotor strength of aminoacyl-tRNA synthetases (aaRSs) and/or tRNA genes[22], engineering of $tRNA^{Pyl}$ [23], rational and semi-rational engineering of aaRS variants [24], optimizing sequence context around the target codon[25], host cell engineering[26], engineering of parts of the translational machinery (e.g., elongation factor TU)[27] and increasing the solubility of the aaRS[28]. Nevertheless, the yield of target protein production falls short of that of the wild-type reference protein. Thus, to fully exploit all the valuable properties of the PylRS system, a fundamental shift away from the prevalent approach to genetic code expansion of using mesophilic and thermophilic aaRS to tolerate necessary mutations may be required.

Here, we use psychrophilic homologs of the PylRS system to achieve higher activity under standard cultivation conditions.[29,30] We show that the *Methanococcoides burtonii* (*M. burtonii*) PylRS (*Mbur*PylRS) exhibits superiority not only in terms of *in vivo* incorporation efficiency but also in incorporatable substrate diversity

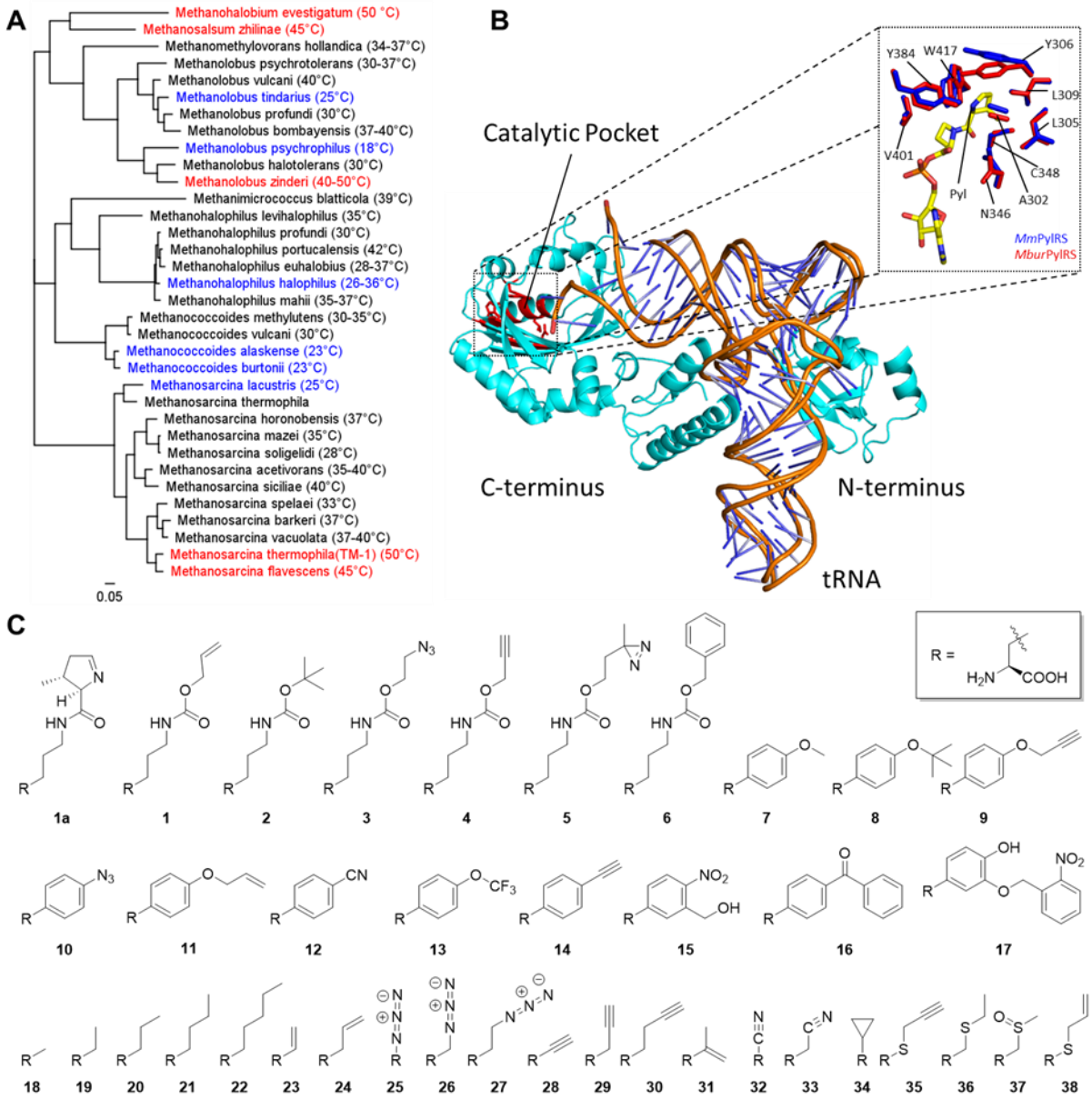

Figure 1. Coupling evolutionary phylogeny with structural biology and non-canonical amino acid chemistries led to the discovery of psychrophilic orthogonal translation ("Cold-OTS") of PylRS. **A**) Phylogenetic tree based on all N-terminus containing PylRS enzymes of *Methanosarcinales* and *Methanomassiliicoccales* 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.[31] 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.[18,32] 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**).

Results

Our goal was to find a more efficient PylRS variant based on a psychrophilic enzyme scaffold. We hypothesized that organisms with a lower optimal growth temperature (OGT) have experienced some adaption to cold environments, which would be reflected in their enzyme activity (**Figure 1A**). Detailed considerations for phylogenic analysis and construct inclusion can be found in the Supplementary Material. In short, PylRS variants from host organism with an OGT of 26 °C or less were included to the

group. To elucidate the selected PylRS activities, six different Pyl-analogs (**1**, **2**, **3**, **4**, **5**, **6**) with different recognition efficiencies were chosen (**Figure 1C**). The known catalytic activities of PylRS for these substrates range from good (**1**, **2**) to moderate (**3**, **4**, **5**[33]) to poor (**6**; very low activity with 3.5 mM supplied[34]). Because even wild-type *Mb*PylRS with an SmbP-tag showed significant improvements, the wild-type PylRS set and one mutant set with known destabilizing mutations (for S-allyl-L-cysteine (Sac, (**38**) recognition) was tested with and without this tag (**Figures S17-S34** and **S51-S56**).[28] Unfortunately, no improvements were observable besides for *Mb*PylRS which was therefore included with the tag in **Figure 2**.

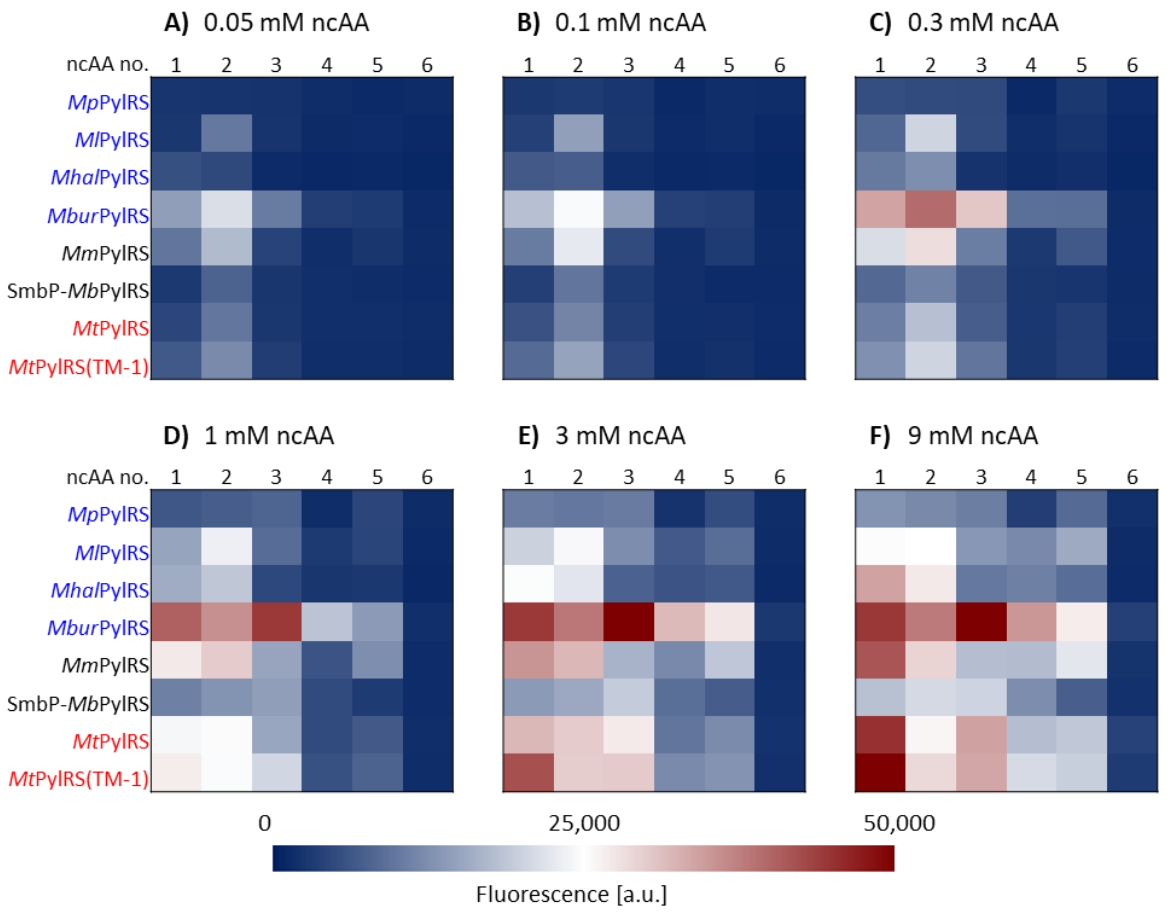

Figure 2. Heatmaps based on fluorescence of sfGFP(1x amber) expression with *E. coli* BL21(DE3) for the eight selected PylRS as a function of different ncAA concentrations; **A)**, **B)**, **C)**, **D)**, **E)**, **F)** are 0.05, 0.1, 0,3, 1, 3, 9 mM ncAA supplied. Bar charts with the number of replicates 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.

Figure 2 shows six heatmaps depicting the incorporation efficiencies of the selected PylRS variants as a function of substrate concentration. As a rough guide, successful ncAA incorporation is detectable via ESI-MS when the signal with ncAA is twice as high as the signal without ncAA. **Figure 2** shows that all of the generated PylRS OTS are functional under standard cultivation conditions and recognize substrates to varying degrees, with the exception of **6**. Encouragingly, none of the psychrophilic PylRSs show significantly higher background incorporation, indicating that they can discriminate against canonical amino acids (cAAs) with the same fidelity as the commonly used PylRSs. The *Mbur*PylRS shows superior OTS efficiency for any substrate at almost any concentration. Surprisingly, the *Mp*PylRS, derived from the most psychrophilic organism, has the lowest incorporation efficiency.

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.[24] On the other hand, SmbP-*Mb*PylRS can even outperform the *Mm*PylRS at high concentrations (3 and 9 mM) of **3**. The thermophilic *Mt*PylRS(TM-1) is on par with the mesophilic enzymes and surprisingly has higher OTS efficiency than the SmbP-*Mb*PylRS at intermediate or high ncAA concentrations for substrates **1**,

3, **4**, **5**. In contrast to highly thermophilic enzymes, which often show no or very low activity at 37 °C, this seems plausible since *Mt*PylRS(TM-1) originates from an organism that is only mildly thermophilic.

For good substrates such as BocK (**2**) 1 mM concentration is often used.[34,35] For other substrates in combination with engineered PylRSs up to 10 mM have to be used to achieve satisfactory incorporation performance.[35,36] Interestingly, the efficiency of *Mbur*PylRS for BocK (**2**) does not increase at concentrations above 0.3 mM. In general, at 0.05 mM ncAA concentration, the *Mbur*PylRS shows 30% (BocK (**2**)) to 280% (ProK (**3**)) higher OTS efficiency than the second-best variant. At 0.3 mM ProK (**3**), the difference is as high as 440%. The absolute advantage of *Mbur*PylRS decreases with increasing ncAA concentration, but it is still by far the most efficient enzyme. Considering that some of the ncAAs used are very expensive and therefore hinder research projects, increasing the incorporation efficiency at low ncAA concentrations could boost these endeavors. **Figure 2** indicates that *Mbur*PylRS could be more versatile than the other variants. Therefore, we used a method to estimate the extent of promiscuity (**Table 1**). Although this is usually done *in vitro* to exclude effects that normally bias reaction rates in living cells, we derived estimates from our *in vivo* data. The magnitude of promiscuity is usually described by kinetic parameters for analogous substrates compared with those of the native substrate.[37] Because we do not have the latter, we used the best performing substrate as a baseline. Detailed considerations of this approach can be found in the Supplementary Material. **Table 1** shows that the psychrophilic PylRSs are more promiscuous than the other PylRSs, except *Mhal*PylRS. Only SmbP-*Mb*PylRS is at the same level as *Mbur*PylRS. *Mt*PylRS(TM-1) has the second lowest promiscuity coefficient, which is within the expected range for a thermophilic enzyme. It is not surprising that the *Mm*PylRS and *Mt*PylRS have similar promiscuity values since their catalytic domain sequence is 97% identical. The substrate promiscuity trends shown in **Table 1** are roughly consistent with our model of their thermal origin.[30,38] Since the promiscuity of an enzyme is still unpredictable,[37] our promiscuity coefficients can potentially serve as an indicator of the engineering potential of a PylRS variant with usefulness in regard to guiding future engineering projects.

Table 1. Relative catalytic PylRSs efficiencies and promiscuity estimations (values are from fluorescence data with 9 mM ncAA concentration, **Figure 2F**)

Construct / substrate	**1** [%]	**2** [%]	**3** [%]	**4** [%]	**5** [%]	**6** [%]	Average efficiency with **6** [%]	Average efficiency without **6** [%]
*Mp*PylRS	100	90	81	18	61	5	59	70
*Ml*PylRS	99	100	52	45	60	2	60	71
*Mhal*PylRS	100	80	29	32	25	4	45	49
*Mbur*PylRS	84	71	100	66	50	5	63	74
*Mm*PylRS	100	70	41	41	52	3	51	61
SmbP-*Mb*PylRS	86	100	97	57	31	5	63	74
*Mt*PylRS	100	56	75	37	40	6	52	62
*Mt*PylRS(TM-1)	100	55	64	39	36	4	50	59

To clarify how the promiscuity score correlates to recognition of substrates that are not Pyl analogs, further experiments were performed. Mutant sets were created for the most efficient psychro-, meso- and thermophilic variants and tested with corresponding substrates (**Figure 3**). The PylRS double alanine variant formed a new class of PylRS enzymes specific for the incorporation of tyrosine and phenylalanine analogs.[8] In one of our previous studies, the corresponding double glycine mutant incorporated some of these ncAAs and was therefore investigated as well (**Figure 3A-E**).[28] SmbP-*Mb*PylRS and *Mbur*PylRS have the highest incorporation efficiencies for all substrates, which agrees well with the promiscuity coefficients. Interestingly, the double Gly mutants prefer the slightly shorter but

bulkier substrate **8,** whereas the much larger substrate **16** is exclusively incorporated with the double Gly variant. This result is plausible since these mutations create more space in the catalytic pocket. To obtain more information, a series of ncAAs were measured in a concentration-dependent manner (**Figure 3B-E**). High fluorescence signals were obtained only at concentrations of 5 mM and above except for **13**, with *Mbur*PylRS and SmbP-*Mb*PylRS performing best. Substrate **13** is well incorporated with *Mbur*PylRS at 1 mM. These efficiencies are at least one order of magnitude lower than for the wild-type substrate analogs.

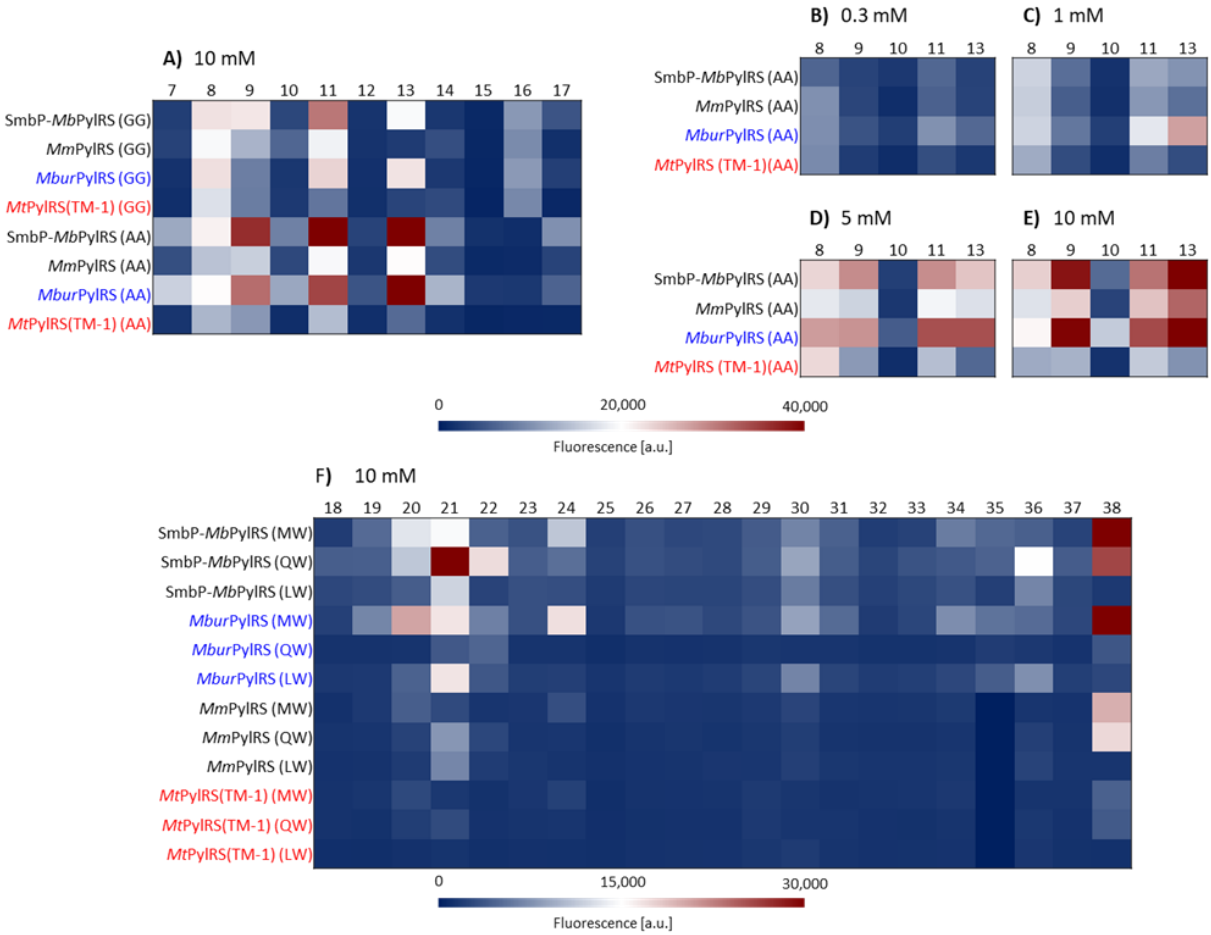

Figure 3. Heatmaps based on fluorescence of the sfGFP(1x amber) expression with *E. coli* BL21(DE3). **A**) and **F)** are selected PylRS mutants with 10 mM supplied ncAA.Concentration dependent ncAA incorporation are **B)**, **C)**, **D)**, **E)** with 0.3, 1, 5 and 10 mM ncAA, respectively. Bar charts showing the number of replicates and error bars showing standard deviation are provided in the supplements; for **A) Figure S35** and **S36**, for **B)**, **C)**, **D)**, **E)**, **Figure S37-S41** and for **F) Figure S42-S50**. Letters in brackets indicate the mutations in *M. barkeri* notation. The first letter indicates the mutation at position N311 and the second letter at C313. 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.].

In a previous study, we engineered a new class of PylRS that can incorporate aliphatic ncAAs with smaller side chains.[10] To further explore the PylRS promiscuities, the key mutations found there were transferred to the earlier set (see **Figure 3F**). In general, not all mutations that are functional in SmbP-*Mb*PylRS variants lead to the same results in other PylRSs, but there is a good overlap. The *Mbur*PylRS and SmbP-*Mb*PylRS variants exhibited a wide range of incorporation activity, whereas the other constructs did not, which correlates well with the promiscuity coefficients for the wild-type enzymes. For all ncAAs that can be incorporated well with the SmbP-*Mb*PylRS constructs, the *Mbur*PylRS (MW)/(LW) variants performed better, except for the incorporation of substrate **21**, **22** and **36**. Interestingly *Mbur*PylRS (MQ) was very inactive. Compared with the *M. barkeri* constructs, the *Mbur*PylRS(N308M:C310W) showed a tendency to incorporate ncAAs with shorter chains than C7. In general, the activities for the *Mm*PylRS and *Mt*PylRS(TM-1) constructs were low.

The last PylRS set tested contains the mutations of the PylRS variant incorporating **38**, called SacRS. The two mutations required are C313W and W382S (*M. barkeri* notation). The four PylRS tested are shown in **Figure S5**, with *Mbur*SacRS having a 120% higher performance (at 0.6 mM Sac concentration) in comparison to the second best performing. The previously found S382T (*M. barkeri* notation) mutation which further increased Sac (**38**) incorporation efficiency, was also tested for the previous constructs (**Figure S6**). Although *Mbur*SacRS was already very active, this mutation increased the *in vivo* efficiency by 80% (at 0.6 mM Sac (**38**)) and 120% for SproC (**35**) (at 1 mM) (**Figure S7**). Nearly 90% of the maximum suppression signal is obtained at 0.3 mM Sac (**38**), indicating very high activity on par with wild-type PylRS and BocK (**2**) as substrate.

After finding that the *Mbur*PylRS OTS also performed best with ncAA incorporation at multiple sites (**Figure S8-S11**), we compared it with one of the most efficient OTS for *E. coli* described in the literature (**Figure 4**). The *Mj*TyrRS system, along with the TyrRS system from *Archaeoglobus fulgidus* (*Af*TyrRS), is one of the known OTS that can produce target proteins with in-frame stop codons at wild-type level.[19,39,40] Two *Mbur*PylRS and two *Mj*TyrRS variants were used for comparison. The *Mj*ONB-Dopa OTS was selected because its high efficiency is well documented in our group.[19,41,42] The second variant was selected based on efficiency considerations and experiments described in the Supplementary Materials (**Figure S12** and **S13**). The *Mj*TyrRS variants were in a different plasmid system (pUltra) than the PylRS OTS (pTECH). To estimate the differences in efficiency despite different OTS setups, fluorescence signals were normalized to the corresponding wild-type target protein signal (sfGFP without in-frame stop codons, **Figure S57 and S58**). In general, the efficiency of suppressing a single in-frame stop codon is comparable. Surprisingly, the *Mbur*PylRS-based OTSs suppressed multiple in-frame stop codons with higher efficiency than the *Mj*TyrRS OTS (at 1 mM). The efficiency of *Mbur*PylRS suppressing five in-frame stop codons is 2.5 to 3 times higher than of *Mj*TyrRS. The *Mj*TyrRS performance was consistent with the literature.[43]

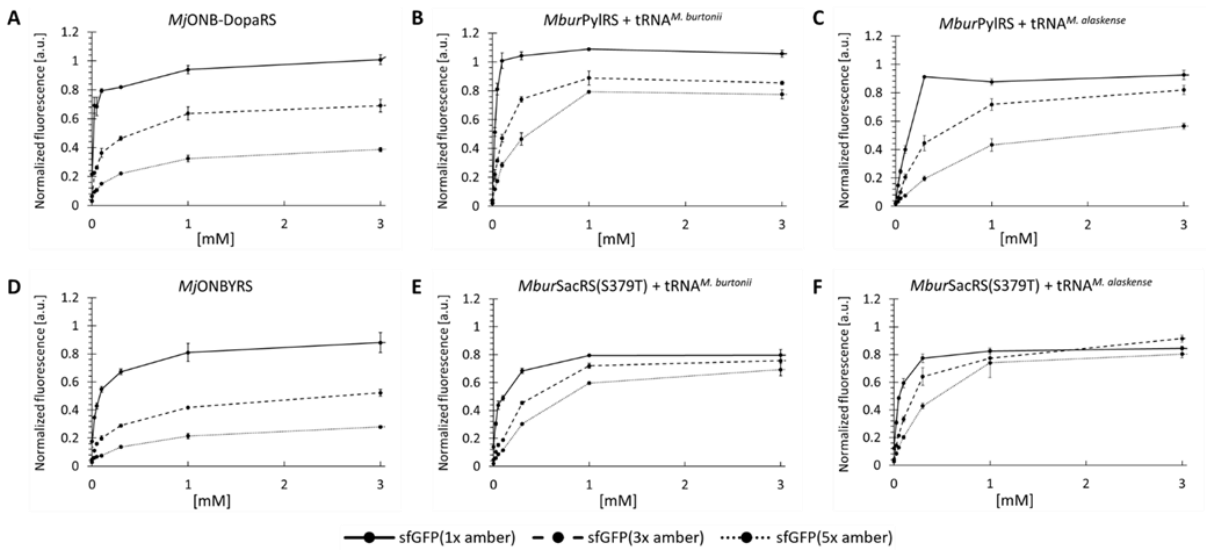

Figure 4. Concentration-dependent unnatural protein production by **A**) *Mj*ONB-DopaRS with substrate **17**, **B**) *Mbur*PylRS with substrate **2**, **C**) *Mbur*PylRS with tRNA from *M. alaskense* and substrate **2, D**) *Mj*ONBYRS with substrate **11**, and **E**) *Mbur*SacRS(S379T) with substrate **38**, **F**) *Mbur*SacRS(S379T) with tRNA from *M. alaskense* and substrate **38.** The host for protein production was RF1 deficient *Escherichia coli* B-95.ΔA[44]. Endpoint measurements for ncAA concentrations of 0.025, 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).

A scale-up of protein production was performed to see how the small-scale fluorescent assays translated into larger-scale recombinant protein production. (**Table 2**). Protein yields agreed well with the trends observed in the small-scale fluorescence experiments, except for SproC (**35**). Here, the yield of target protein with *Mbur*SacRS(S379T) was well above what was expected in **Figure S11**. Finally, $tRNA^{Pyl}$ was also examined as its engineering may have a non-negligible impact on OTS performance.[23] **Figure S59** shows a selection of $tRNA^{Pyl}$. The $tRNA^{Pyl}$ of *M. burtonii* is approximately 30% less stable (at 37°C) than

the variant of *M. mazei*. To determine whether the stability differences affect the OTS performance, the $tRNA^{Pyl}$ of *M. burtonii* was replaced with the $tRNA^{Pyl}$ from *M. mazei* to create orthogonal hybrid pairs (**Figure S14**). The differences in OTS performance are small and mostly within the standard deviations. Moreover, it was reported towards the end of our study that the genome of the psychrophilic organism *M. alaskense* is available. The OTS (*Mala*PylRS and the corresponding tRNA) was tested in combination with *Mbur*PylRS (**Figure S15**). While *Mala*PylRS did not perform better with either tRNA, *Mbur*PylRS increased by up to 50% with AzidoK (**3**) in combination with $tRNA^{Pyl}$ from *M. alaskense*. The increase in performance for Sac (**38**) and SproC (**35**) was in a similar range (**Figure S16**). Interestingly, pairing *Mbur*PylRS with the *M. alaskense* tRNA led to a different picture for the multi-site ncAA incorporation. While the efficiency seems slightly enhanced with **38** (**Figure 4F**), it is decreased for **2** (**Figure 4C**). This highlights the fact that tRNA stability is just one factor in a complex pathway which can be optimized but only in combination with the substrate, as EF-TU binding affinity also plays a role.

Table 2. Setup of protein production platform with calculated and observed molecular weights of reporter proteinscorresponding to a) SUMO-sfGFP(1x amber), b) SUMO-sfGFP(3x amber), c) SUMO-sfGFP(5x amber), d) SUMO-sfGFP(wild-type). The masses were determined by ESI-MS of the intact proteins (**Figure S60-S88**).

ncAA	[mM]	E. coli strain[a]	PylRS construct	reporter	calculated mass [Da]	found mass [Da]]	Δ mass [Da]	protein yield [mg/L][b]
2	0.05	BL21	*Mm*PylRS	a[c]	39096.0	-	-	18.5
2	0.05	BL21	*Mbur*PylRS	a[c]	39096.0	-	-	20
2	1	BL21	*Mm*PylRS	a	39096.0	39094	2	97.2
2	1	BL21	*Mbur*PylRS	a	39096.0	39094	2	123.6
3	0.05	BL21	*Mm*PylRS	a[c]	39109.0	-	-	4.3
3	0.05	BL21	*Mbur*PylRS	a[c]	39109.0	-	-	11.2
3	1	BL21	*Mm*PylRS	a	39109.0	39105	4	33
3	1	BL21	*Mbur*PylRS	a	39109.0	39106	3	84.9
9	3	BL21	*Mm*PylRS(N346A:C348A)	a	39069.0	39065	4	11.9
9	3	BL21	*Mbur*PylRS(N308A:C310A)	a	39069.0	39065	4	19.9
11	3	BL21	*Mm*PylRS(N346A:C348A)	a	39071.0	39068	3	26.4
11	3	BL21	*Mbur*PylRS(N308A:C310A)	a	39071.0	39068	3	52
13	1	BL21	*Mm*PylRS(N346A:C348A)	a	39098.9	39095	3.9	4.6
13	1	BL21	*Mbur*PylRS(N308A:C310A)	a	39098.9	39095	3.9	15.1
8	3	BL21	*Mm*PylRS(N346A:C348A)	a	39087.0	39084	3	30.2
8	3	BL21	*Mbur*PylRS(N308A:C310A)	a	39087.0	39084	3	51.1
38	0.3	B-95.ΔA	SmbP-*Mb*SacRS(S382T)	d	39023.9	39023	0.9	89.8
38	0.3	B-95.ΔA	SmbP-*Mb*SacRS(S382T)	a	39010.9	39010	0.9	42.2
38	0.3	B-95.ΔA	SmbP-*Mb*SacRS(S382T)	b	39055.2	39053	2.2	13.8
38	0.3	B-95.ΔA	SmbP-*Mb*SacRS(S382T)	c	39097.5	39095	2.5	6.7
38	0.3	B-95.ΔA	*Mbur*SacRS(S379T)	d	39023.9	39023	0.9	113.2
38	0.3	B-95.ΔA	*Mbur*SacRS(S379T)	a	39010.9	39011	0.1	86.2
38	0.3	B-95.ΔA	*Mbur*SacRS(S379T)	b	39055.2	39054	1.2	69.1
38	0.3	B-95.ΔA	*Mbur*SacRS(S379T)	c	39097.5	39096	1.5	38.6

35	1	B-95.ΔA	SmbP-*Mb*SacRS(S382T)	a	39008.9	39009	0.1	60.1
35	1	B-95.ΔA	SmbP-*Mb*SacRS(S382T)	b	39049.0	39047	2	19.6
35	1	B-95.ΔA	SmbP-*Mb*SacRS(S382T)	c	39087.2	39084	3.2	10.9
35	1	B-95.ΔA	*Mbur*SacRS(S379T)	a	39008.9	39008	0.9	104.9
35	1	B-95.ΔA	*Mbur*SacRS(S379T)	b	39049.0	39048	1	59.3
35	1	B-95.ΔA	*Mbur*SacRS(S379T)	c	39087.2	39086	1.2	41

[a] All DE3. [b] Yield per liter of cell culture; sfGFP absorption was measured directly in the elution fraction before dialysis. [c] Was not analyzed by ESI-MS

Discussion

The PylRS OTS is the most used system for the incorporation of ncAAs into proteins. We addressed its major drawback, low catalytic efficiency, by performing a comparative study for PylRS homologs of extremophilic origins. We found a particularly active variant, *Mbur*PylRS, derived from a psychrophilic (cold-active) organism, *Methanococcoides burtonii*. This PylRS shows remarkable efficiency in incorporating different ncAAs. The superior incorporation efficiency was observed for all major substrate classes highlighting the robustness of the results. Most striking was the multi-site ncAA incorporation with BocK (**2**) and especially Sac (**38**). To the best of our knowledge, no PylRS mutant has ever been reported to have nearly the same ncAA incorporation efficiency as the wild-type.

Considering that only two OTSs known to approach wild-type recombinant protein production (*Mj*TyrRS[39–41] and *Af*TyrRS[40]), the addition of a third OTS to this group is highly desirable. This increases the likelihood that a larger number of ncAAs (identical and different ones) can be simultaneously incorporated with good efficiency. In this context, 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.[45,46] It has been shown that 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.[46] This suggests that the catalytic efficiency of the PylRS OTS is still a limiting factor even when free codons are available. The high performance of *Mbur*PylRS in a RF1 deficient strain (**Figure 4**) suggest that this problem could be alleviated with this enzyme.

The high performance even at very low ncAA concentrations will pave the way for efficient coupling *Mbur*PylRS with ncAA producing engineered metabolic pathways that eliminates the need for 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.[47] The low price and concentration required for efficient incorporation of Sac (**38**) will mean that recombinant proteins containing one or even multiple bioorthogonal alkene functions can 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.[48]

Compared to one of the best performing OTS known, *Mj*TyrRS, *Mbur*PylRS also proved superior for incorporation of ncAAs at multiple sites. This performance is particularly important because research into other domains that require efficient orthogonal pairs urgently needs more active enzymes which are mutual orthogonal.[45,49]

In addition, we have elucidated a crucial enzyme property known as catalytic promiscuity i.e., the ability of a single active site to catalyze chemical transformations with different substrates, which is often enhanced in psychrophilic enzymes.[29,30,50,51] Encouragingly, *Mbur*PylRS shows higher promiscuity in

comparison to other PylRS which should facilitate further engineering for additional substrate recognitions.

Although white biotechnology has long benefited from promiscuous cold-active enzymes, research on orthogonal translation has focused primarily on engineering thermophilic and, to a lesser extent, mesophilic aaRS.[52] Based on our results, we would like to propose a paradigm shift to especially seek psychrophilic organisms as a valuable source for additional efficient OTSs, with Class II aaRS in particular being of highest interest.[18] Since other parts important for full emancipation of the host translational machinery, such as orthogonal ribosomes[49], could also benefit from the unique properties of psychrophilic enzymes, we want to encourage researchers in the field to pay particular attention to these organisms. We anticipate that the combination of high catalytic activity and greater substrate promiscuity will greatly facilitate future engineering efforts and enable applications previously not possible due to the aforementioned limitations of existing systems.

If the higher efficiency of our "Cold"-OTS based *Mbur*PylRS also holds true in other hosts, which is plausible, this could boost the usefulness for the production of therapeutic proteins with significance for commercial use.[53]

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 S1).

Plasmid Vector Construction

The pTECH vector was a gift from Dieter Söll (Addgene plasmid #104073).[24] 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). 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.

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) 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.

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. 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).

12. Melnikov, S. V. & Söll, D. Aminoacyl-TRNa synthetases and tRNAS for an expanded genetic code: What makes them orthogonal? *Int. J. Mol. Sci.* **20**, (2019).

13. Elliott, T. S. *et al.* Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. *Nat. Biotechnol.* **32**, 465–472 (2014).

14. Lovelock, S. L. *et al.* The road to fully programmable protein catalysis. *Nature* **606**, 49–58 (2022).

15. Almhjell, P. J., Boville, C. E. & Arnold, F. H. Engineering enzymes for noncanonical amino acid synthesis. *Chem. Soc. Rev.* **47**, 8980–8997 (2018).

16. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596–604 (2009).

17. Grasso, K. T. *et al.* Structural Robustness Affects the Engineerability of Aminoacyl-tRNA Synthetases for Genetic Code Expansion. *Biochemistry* **60**, 489–493 (2021).

18. 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).

19. Hauf, M. *et al.* Photoactivatable Mussel-Based Underwater Adhesive Proteins by an Expanded Genetic Code. *ChemBioChem* **18**, 1819–1823 (2017).

20. Sarmiento, F., Peralta, R. & Blamey, J. M. Cold and hot extremozymes: Industrial relevance and current trends. *Front. Bioeng. Biotechnol.* **3**, (2015).

21. O'Donoghue, P., Ling, J., Wang, Y.-S. & Söll, D. Upgrading protein synthesis for synthetic biology. *Nat. Chem. Biol.* **9**, 594–598 (2013).

22. 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).

23. Fan, C., Xiong, H., Reynolds, N. M. & Söll, D. Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids. *Nucleic Acids Res.* **43**, 1–10 (2015).

24. Bryson, D. I. *et al.* Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **13**, 1253–1260 (2017).

25. 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).

26. Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science (80-.).*

342, 357–360 (2013).

27. Park, H. S. *et al.* Expanding the genetic code of Escherichia coli with phosphoserine. *Science (80-.).* **333**, 1151–1154 (2011).

28. 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).

29. Siddiqui, K. S. & Cavicchioli, R. Cold-adapted enzymes. *Annu. Rev. Biochem.* **75**, 403–433 (2006).

30. Feller, G. Psychrophilic Enzymes: From Folding to Function and Biotechnology. *Scientifica (Cairo).* **2013**, 1–28 (2013).

31. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **5**, e9490 (2010).

32. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

33. 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).

34. 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).

35. 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).

36. Dunkelmann, D. L., Willis, J. C. W., Beattie, A. T. & Chin, J. W. Engineered triply orthogonal pyrrolysyl–tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids. *Nat. Chem.* **12**, 535–544 (2020).

37. Khersonsky, O. & Tawfik, D. S. *Enzyme promiscuity—evolutionary and mechanistic aspects*. (Elsevier, 2010).

38. Santiago, M., Ramírez-Sarmiento, C. A., Zamora, R. A. & Parra, L. P. Discovery, Molecular Mechanisms, and Industrial Applications of Cold-Active Enzymes. *Front. Microbiol.* **7**, (2016).

39. Baumann, T. *et al.* Computational aminoacyl-tRNA synthetase library design for photocaged tyrosine. *Int. J. Mol. Sci.* **20**, (2019).

40. Cervettini, D. *et al.* Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase–tRNA pairs. *Nat. Biotechnol.* **38**, 989–999 (2020).

41. 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).

42. Koch, N. G., Baumann, T., Nickling, J. H., Dziegielewski, A. & Budisa, N. Engineered bacterial host for genetic encoding of physiologically stable protein nitration. *Front. Mol. Biosci.* **9**, 1–16 (2022).

43. Zheng, Y. *et al.* Performance of optimized noncanonical amino acid mutagenesis systems in the absence of release factor 1. *Mol. Biosyst.* **12**, 1746–1749 (2016).

44. Mukai, T. *et al.* Highly reproductive Escherichia coli cells with no specific assignment to the UAG codon. *Sci. Rep.* **5**, 1–9 (2015).

45. Robertson, W. E. *et al.* Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science (80-.).* **372**, 1057–1062 (2021).

46. Fischer, E. C. *et al.* New codons for efficient production of unnatural proteins in a semisynthetic organism. *Nat. Chem. Biol.* **16**, 570–576 (2020).

47. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat. Chem. Biol.* **5**, 593–599 (2009).

48. 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).

49. de la Torre, D. & Chin, J. W. Reprogramming the genetic code. *Nat. Rev. Genet.* **22**, 169–184 (2021).

50. 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).

51. Zeymer, C. & Hilvert, D. Directed Evolution of Protein Catalysts. *Annu. Rev. Biochem.* **87**, 131–157 (2018).

52. 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).

53. 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).