

Incorporation of novel noncanonical amino acids in model proteins using rational and evolved variants of *Methanosarcina mazei* pyrrolysyl-tRNA synthetase

vorgelegt von
M. Sc. in Chemischer Biologie

Matthias P. Exner

geb. in Hagen

Von der Fakultät II – Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften
Dr. rer. nat.

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Thomas Friedrich

Gutachter: Prof. Dr. Nediljko Budisa

Gutachterin: Prof. Dr. Zoya Ignatova

Tag der wissenschaftlichen Aussprache: 08.02.2016

Berlin 2016

HC SVNT DRACONES

Hunt-Lenox globe, ca. 1510

Parts of this work were published as listed below:

Al Toma RS, Kuthning A, **Exner MP**, Denisiuk A, Ziegler J, Budisa N, Süssmuth, RD. Site-Directed and Global Incorporation of Orthogonal and Isostructural Noncanonical Amino Acids into the Ribosomal Lasso Peptide Capistruin. *ChemBioChem* **16**, 503–509 (2015).

Worst EG*, **Exner MP***, De Simone A, Schenkelberger M, Noireaux V, Budisa N, Ott A. Cell-free expression with the toxic amino acid canavanine. *Bioorg. Med. Chem. Lett.* **25**, 3658–60 (2015).
*equal contributions

Albrecht M, Lippach A, **Exner MP**, Jerbi J, Springborg M, Budisa N, Wenz G. Site-specific conjugation of 8-ethynyl-BODIPY to a protein by [2 + 3] cycloaddition. *Org. Biomol. Chem.* **13**, 6728–6736 (2015).

Table of contents

Table of contents.....	IV
Abbreviations and definitions	VI
1. Zusammenfassung.....	1
2. Summary.....	2
3. Introduction.....	4
3.1 The genetic code and protein biosynthesis.....	4
3.2 Incorporation of noncanonical amino acids into proteins	9
3.3 Engineered aminoacyl-tRNA synthetases: rational and evolved variants	16
3.4 Synthetic biology: Employing noncanonical amino acids and new-to-nature chemical reactions in biochemical pathways	19
3.5 Model proteins	23
4. Aim of this study.....	24
5. Results and discussion.....	25
5.1 Establishment of a stop codon suppression system	25
5.2 Rational mutagenesis of <i>MmPyIRS</i>	33
5.3 Selection of <i>MmPyIRS</i> variants.....	38
5.4 Misincorporation of canonical amino acids	50
5.5 Excursion: Protein engineering with supplementation pressure incorporation.....	51
6. Conclusion and outlook.....	54
7. Materials and methods	56
7.1 Materials.....	56
7.2 Molecular biology methods.....	66
7.3 Microbiological methods.....	70
7.4 Biochemical methods	74
7.5 Analytical methods.....	75
7.6 Biophysical methods.....	77
8. References.....	78
Appendices	A
Plasmid sequences	A
Gene sequences	F
Protein sequences	G
List of schemes and figures	H
List of tables	L

Abbreviations and definitions

Danksagung M

Abbreviations and definitions

Abbreviations and definitions

aa	amino acid	<i>Mm</i>	<i>Methanosarcina mazei</i>
aaRS	aminoacyl-tRNA synthetase	mRNA	messenger-RNA
ADP	adenosine diphosphate	MS	mass spectroscopy
AGE	agarose gel electrophoresis	ncaa	noncanonical amino acid
AMP	adenosine monophosphate	NMM	new minimal medium
amp	ampicillin	OAHSS	O-acetylhomoserine sulfhydrylase
ATP	adenosine triphosphate	OASS	O-acetylserine sulfhydrylase
AZT	azidothymidine	PAGE	polyacrylamide gel electrophoresis
b*	barstar	PCR	polymerase chain reaction
bp	base pair	PDB-ID	protein data bank id number
caa	canonical amino acid	PP _i	inorganic pyrophosphate
cam	chloramphenicol	pylB	methylornithine synthase
cat	chloramphenicol-acetyl transferase	pylC	lysine-methylornithine pseudopeptide synthase
CD	circular dichroism	pylD	pyrrolysine synthase
ddH ₂ O	ultrapure water	PYLIS	pyrrolysine insertion signal
<i>Dh</i>	<i>Desulfitobacterium hafniense</i>	PyIRS	pyrrolysyl-tRNA synthetase
dH ₂ O	desalted water	PyIRS_AF	<i>MmPyIRS</i> (Y306A/Y384F)
DNA	deoxyribonucleic acid	pylS	pyrrolysyl-tRNA synthetase (gene)
EF	elongation factor	pylT	tRNA ^{Pyl} (gene)
EF1/EF2	archaeal/eukaryotic elongation factors	RF	release factor
EF-Sec	elongation factor selenocysteine	RF1/RF2	bacterial release factors
EF-Tu	elongation factor thermo-unstable	RNA	ribonucleic acid
EGFP	enhanced green fluorescent protein	<i>Sa</i>	<i>Sulfolobus acidocaldarius</i>
EYFP	enhanced yellow fluorescent protein	SacRS	<i>MmPyIRS</i> (C348W/W417S)
GFP	green fluorescent protein	SCS	stop codon suppression
IEX	ion exchange chromatography	SDS	sodium dodecyl sulfate
IMAC	immobilized-metal affinity chromatography	SECIS	selenocysteine insertion signal
IPTG	isopropyl-β-D-thiogalactosid	SOB	super optimal broth
IVT	<i>in vitro</i> translation	SPI	supplementation pressure incorporation
kan	kanamycin	tdk	thymidine kinase
LB	Luria-Bertani medium	thyA	thymidylate synthase
LC-ESI-	liquid chromatography –	TMEDA	tetramethylethylenediamine
MS	electrospray ionization –mass spectrometry	<i>Tp</i>	<i>Thermincola potens</i>
<i>Mb</i>	<i>Methanosarcina barkeri</i>	tRNA ^{aa}	transfer-RNA with aa-identity
<i>Me</i>	<i>Methanohalobium evestigatum</i>	TTL	<i>Thermoanaerobacter thermohydrosulfuricus</i> lipase
<i>Mj</i>	<i>Methanocaldococcus jannschii</i>	VIS	visible light
		wt	wild-type
		ψ-b*	pseudo-wild type barstar

Abbreviations and definitions

Canonical: Subject to the limitation of 20 standard amino acids and three stop signals encoded by the consensus genetic code.

Amino acid analog: Having structural resemblance with the respective amino acid.

Amino acid surrogate: Having strong structural and/or electronic resemblance to the respective amino acid and accepted (with lower efficiency) by the amino acid's aminoacyl-tRNA synthetase.

Semicanonical: Employing canonical-like processes in a very limited set of organisms (e.g. incorporation of pyrrolysine, synthesis of cysteine from phosphoserine on-tRNA) OR widespread employment of noncanonical processes (e.g. incorporation of Sec).

Noncanonical: Not part of canonical or semicanonical processes (e.g. natural or synthetic amino acids not normally involved in translation).

Orthogonal: Not interfering with and not interfered by natural structures and processes.

Canonical/semicanonical amino acid abbreviations: Alanine (Ala/A); cysteine (Cys/C); aspartic acid (Asp/D); glutamic acid (Glu/E); phenylalanine (Phe/F); glycine (Gly/G); histidine (His/H); isoleucine (Ile/I); phosphoserine (Sep/J); lysine (Lys/K); leucine (Leu/L); methionine (Met/M); asparagine (Asn/N); pyrrolysine (Pyl/O); proline (Pro/P); glutamine (Gln/Q); arginine (Arg/R); serine (Ser/S); threonine (Thr/T); selenocysteine (Sec/U); valine (Val/V); tryptophan (Trp/W); tyrosine (Tyr/Y).

Noncanonical amino acids abbreviations: Pyl: pyrrolysine; Bok: N ϵ -butyloxycarbonyllysine; Zk: N ϵ -Cbz-lysine; Pok: N ϵ -propargyloxycarbonyllysine; Pnk: N ϵ -pentenoyllysine; Hnk: N ϵ -heptenoyllysine; Nbk: N ϵ -norbonyloxycarbonyllysine; Ayk: N ϵ -acryloyllysine; Bct: biocytin; Nuk: N ϵ -norbonylurea-lysine; Nap: N δ -norbonyl-aminopentanoic acid; Cok: N ϵ -Cyclooctynyloxycarbonyllysine; Edo: 2,7-diamino-4-enedioic acid; Ops: O-pentenylserine; Npa: N-pentenyl-N-methyl- β -amino-alanine; Sac: S-allylcysteine; Sbc: S-butenylcysteine; Spc: S-pentenylcysteine; Aha: azidohomoalanine; Can: canavanine, Bpa: benzoylphenylalanine.

Standard DNA/RNA nucleotide bases: Adenine (A); cytosine (C), guanine (G), thymine (T), uracil (U).

Degenerate DNA nucleotide bases: A/C/G/T (N); A/C/G (V); A/C/T (H); A/G/T (D); C/G/T (B); A/C (M); G/T (K); A/T (W); G/C (S); C/T (Y); A/G (R).

1. Zusammenfassung

Der Einsatz von nichtkanonischen Aminosäuren zur Erforschung und Modifikation von Proteinfunktionen basiert auf etablierten Methoden. Dennoch gibt es einen Bedarf zur Verbesserung und Diversifizierung des "nichtkanonischen Werkzeugkastens".

Diese Arbeit trägt zu diesem nichtkanonischen Werkzeugkasten in mehrfacher Hinsicht bei:

Ein Suppressionssystem basierend auf *Methanosarcina mazei* Pyrrolysyl-tRNA-Synthetase und der zugehörigen tRNA^{Pyl} wurde adaptiert, optimiert und verwendet, um bekannte Substrate in Proteine und Peptide zu einzubauen. Zusätzlich wurden die Möglichkeiten der Suppression in einem zellfreien Expressionssystem untersucht, wobei allerdings nur Vorarbeiten durchgeführt werden konnten.

Die nichtkanonische Aminosäure S-Allylcystein (Sac) wurde erstmalig in Proteine eingebaut mittels Stoppcodonsuppression. Dazu wurde eine Aminosäurebindetaschenbibliothek von *Methanosarcina mazei* Pyrrolysyl-tRNA-Synthetase (*MmPylRS*) erzeugt und durch iterative Positiv-Negativ-Selektion eine neuartige *MmPylRS* Variante mit völlig veränderter Spezifität selektiert, die tRNA^{Pyl} mit S-Allylcystein belädt. Die neue Variante *MmPylRS*(C348W/W417S), als SacRS bezeichnet, wurde auf Substratspezifität und Effizienz untersucht und scheint hochspezifisch für Sac zu sein. S-Allylcystein kann posttranslational durch diverse bioorthogonale chemische Reaktionen modifiziert werden aufgrund der aktivierten Doppelbindung, einer funktionellen Gruppe, die im natürlichen Proteom nicht vorhanden und im Cytosol der meisten Organismen sehr selten ist, und hat somit potentielle Anwendungen als vielseitiger und spezifischer chemischer Angelpunkt. Die Möglichkeit der biosynthetischen Herstellung von Sac *in situ*, durch Fermentation oder Extraktion aus Pflanzen erhöht die Anwendbarkeit im industriellen Maßstab.

Eine weitere *MmPylRS*-Bibliothek wurde nach einer Variante für die Aktivierung von Biocytin durchsucht, hat aber noch keinen Treffer erbracht.

Zusätzlich wurden durch rationales Engineering der *MmPylRS* Varianten gesucht für den Einbau von langkettigen olefinischen Aminosäuren für posttranslationale Proteinmarkierung. Nε-Heptenoyllysine (Hnk) wurde zum ersten Mal in Proteine inkorporiert unter Verwendung der bereits beschriebenen polyspezifischen Variante *MmPylRS*(Y306A/Y384F), und der Einbau von Nε-Pentenoyllysine (Pnk) wurde zum ersten Mal mit einer signifikanten Effizienz durch Verwendung der Variante *MmPylRS*(C348A/Y384F) erreicht. Dabei wurde die Position C348 erstmalig rational modifiziert und die Polyspezifität der Variante *MmPylRS*(C348A/Y384F) ist zu vermuten.

Darüber hinaus wurden durch die seitenkettenspezifische SPI-Methode die Aminosäuresurrogate Azidohomoalanin und Canavanin in Modellproteine eingebaut zur posttranslationalen chemischen Modifikation bzw. als Methode zur Untersuchung des Einflusses auf Proteinstrukturen.

Eine grafische Zusammenfassung aller Einbauversuche ist Figure 1 auf Seite 3 zu entnehmen.

2. Summary

While the use of noncanonical amino acids to probe, modify or expand protein function comprises well established methodologies, there is still a demand for improvement and diversification of the "noncanonical toolkit".

This work contributes to this noncanonical toolkit in several ways:

A suppression system based on *Methanosarcina mazei* pyrrolysyl-tRNA synthetase and its cognate tRNA^{Pyl} was established, optimized and used to incorporate known substrates into proteins and peptides. Additionally, the possibilities of suppression in a cell-free expression system were explored, but no efficient system could be derived yet.

The noncanonical amino acid S-allylcysteine (Sac) was incorporated into proteins for the first time via stop codon suppression. For this, an active-site library of *Methanosarcina mazei* pyrrolysyl-tRNA synthetase (*MmPylRS*) was generated and an iterative positive-negative selection was set up to select a novel *MmPylRS* variant that charges tRNA^{Pyl} with S-allylcysteine, exerting an entirely altered reactivity compared to the natural enzyme. The novel variant *MmPylRS*(C348W/W417S), dubbed SacRS, was screened for substrate specificity and incorporation rate and was found to be highly specific for Sac. S-allylcysteine can be modified posttranslationally by diverse bioorthogonal chemical reactions due to the activated carbon double bond, a group absent from the natural proteome and with very low abundance in the cytosol of most organisms, and thus has potential applications as a versatile chemical handle. The possibility of biosynthetically producing Sac *in situ*, by fermentation or extracting it from plants can facilitate applications on an industrial scale.

Another *MmPylRS* library was screened for a variant that activates biocytin, but has yet to yield a hit.

Additional rational engineering of *MmPylRS* enabled the incorporation of long-chain olefinic amino acids for protein posttranslational decoration. *N*_ε-heptenoyllysine was incorporated for the first time using the known polyspecific variant *MmPylRS*(Y306A/Y384F), and incorporation of *N*_ε-pentenoyllysine was achieved for the first time with significant efficiency using the variant *MmPylRS*(C348A/Y384F). This is the first instance of rationally engineering *MmPylRS* position C348, which may also prove useful for incorporation of other novel noncanonical amino acids.

Furthermore, the residue-specific SPI method was employed to incorporate the known surrogates azidohomoalanine or canavanine into model proteins, for chemical modification or as a method to study the impact on protein structures, respectively.

A graphical summary of all attempted incorporations is shown in Figure 1 on page 3.

Summary

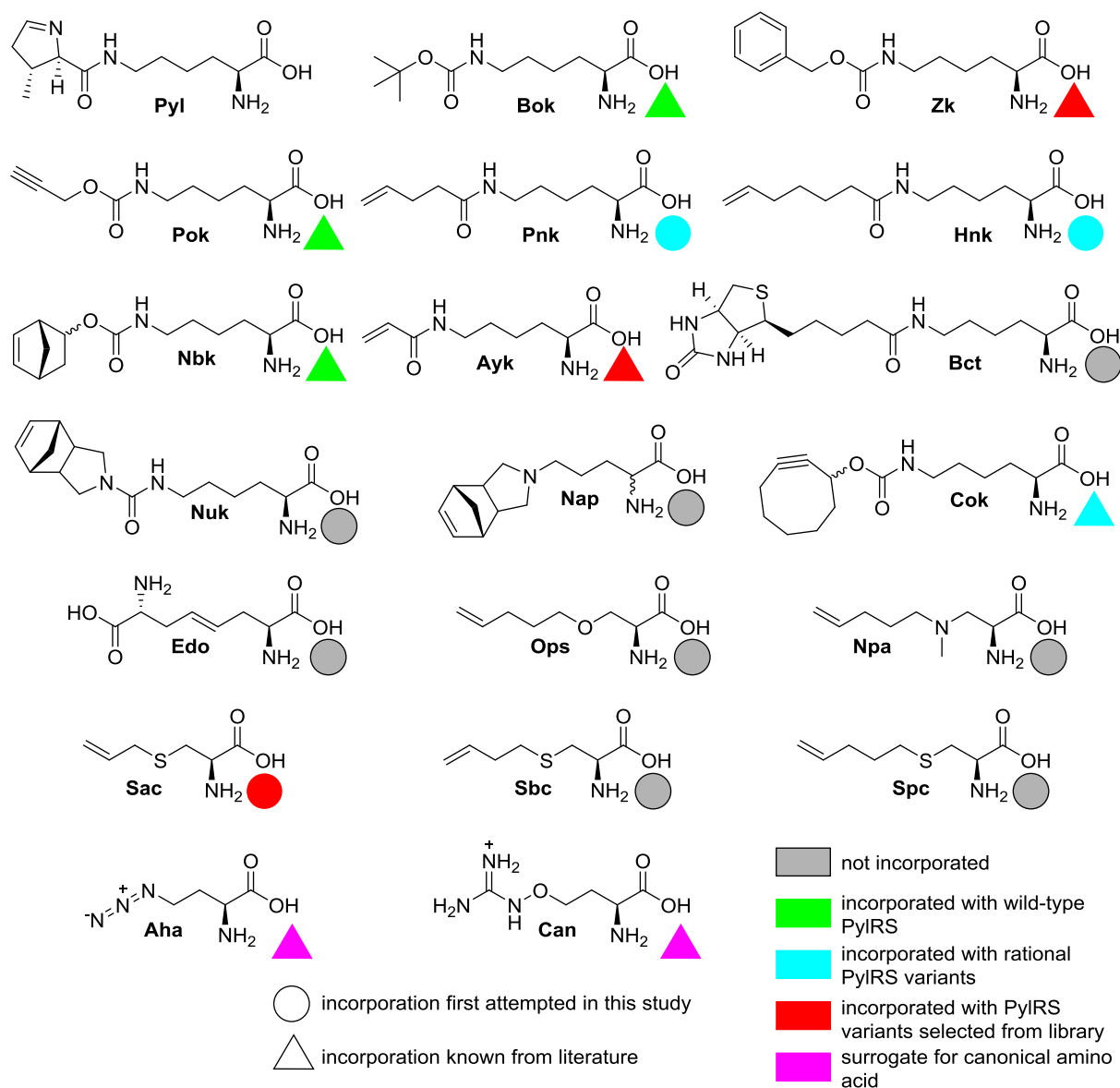


Figure 1: Amino acids relevant for this study. Pyl: pyrrolysine (natural substrate); Bok: N_ε-butyloxycarbonyllysine (known excellent surrogate for Pyl); Zk: N_ε-CBZ-lysine (not incorporated in this work); Pok: N_ε-propargyloxycarbonyllysine; Pnk: N_ε-pentenoyllysine; Hnk: N_ε-heptenoyllysine; Nbk: N_ε-norbonyloxycarbonyllysine; Ayk: N_ε-acryloyllysine; Bct: biocytin; Nuk: N_ε-norbonylurea-lysine; Nap: N_δ-norbonyl-aminopentanoic acid; Cok: N_ε-Cyclooctynyloxycarbonyllysine; Edo: 2,7-diamino-4-enedioic acid; Ops: O-pentenylserine; Npa: N-pentenyl-N-methyl-β-amino-alanine; Sac: S-allylcysteine; Sbc: S-butenylcysteine; Spc: S-pentenylcysteine; Aha: azidohomoalanine (methionine surrogate); Can: canavanine (arginine surrogate).

3. Introduction

3.1 The genetic code and protein biosynthesis

3.1.1 Discovery of the genetic code

The genetic code determines how protein primary structures (amino acid sequences) of proteins are linked to genes (nucleotide sequences) in the DNA.

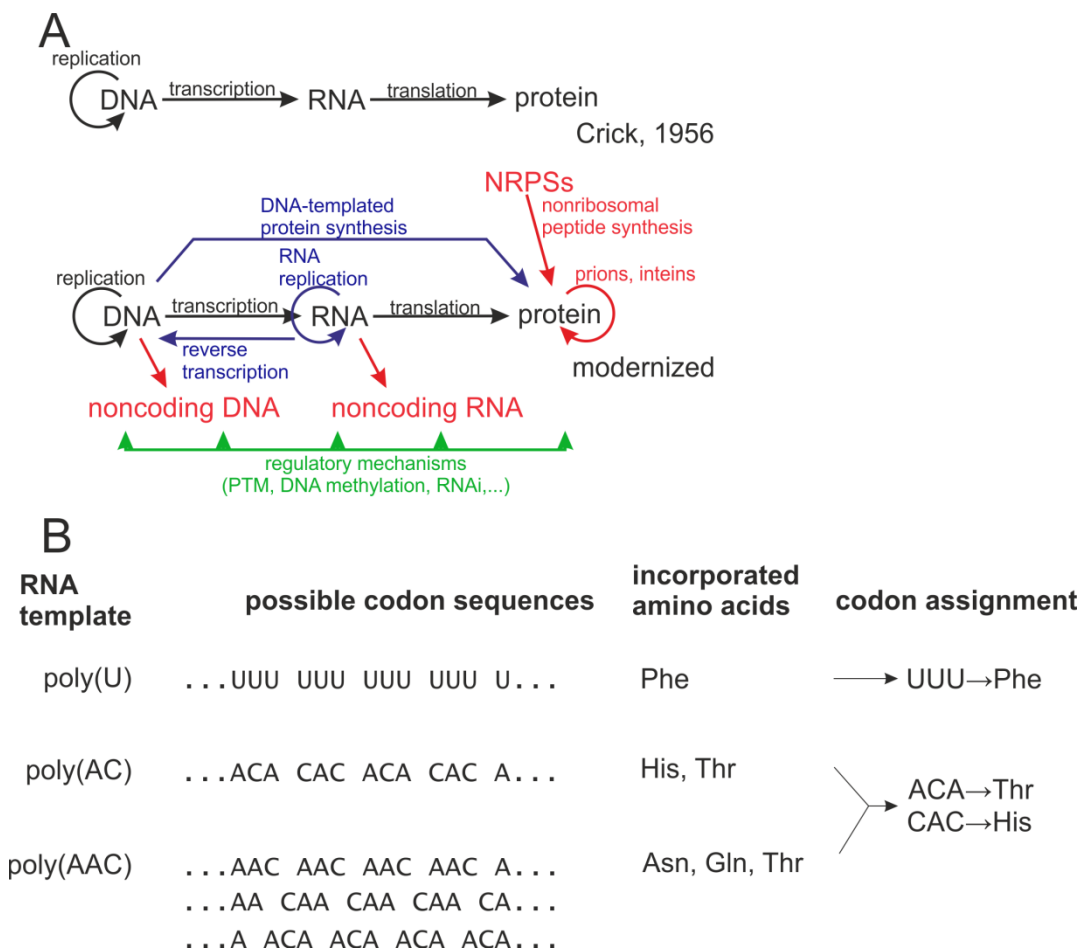


Figure 2: Postulation and experimental verification of the genetic code. A: The "Central Dogma" of molecular biology, as stated by Crick in 1956, postulated the unidirectional flow of sequential information from DNA via RNA to proteins, and an expanded modern version. The blue arrows indicate additional flow of sequential information, the red arrows indicate nonsequential informational flow not directly covered by the postulate, and the green arrows indicate external regulatory mechanisms regulating informational flow. B: To elucidate the assignment of codons, cell-free protein synthesis was primed with synthetic RNA, and the incorporation rate was determined for each amino acid. The combined results of several experiments allowed the deduction of codon assignments. In this example, the assignment UUU→Phe can be directly determined, while elucidation of the assignments ACA→Thr and CAC→His requires two experiments with different RNA templates.

Introduction

While the link between DNA and proteins has been known since the 1950s^[1], the mechanistic nature of this link remained unknown. Early propositions of a direct sequence-dependent on-DNA synthesis ("Diamond Code")^[2] were discarded after the role of RNA was elucidated. This led to the proposition of the "Central Dogma" of molecular biology, assuming a unidirectional flow of information with DNA as a storage molecule, RNA as a transfer molecule and proteins representing the decoded information^[3] (see Figure 2A). While further research discovered several violations of this Central Dogma (e.g. retroviruses using RNA as information storage molecule, and prions exerting protein-to-protein structural information flow), the majority of organisms follow the proposed flow scheme. The exact assignment of DNA/RNA sequence to protein sequence, however, could only be elucidated after major breakthroughs in nucleotide synthesis and *in vitro* protein synthesis^[4,5] (see Figure 2B). The 20 proteinogenic amino acids are encoded by 64 different possible three-nucleotide units (codons), leading to multiple codons for each amino acid, with TAA ("ochre"), TGA ("opal") and TAG ("amber") signaling the translation termination. This unexpected redundant encoding is called degeneracy of the genetic code (Figure 3).

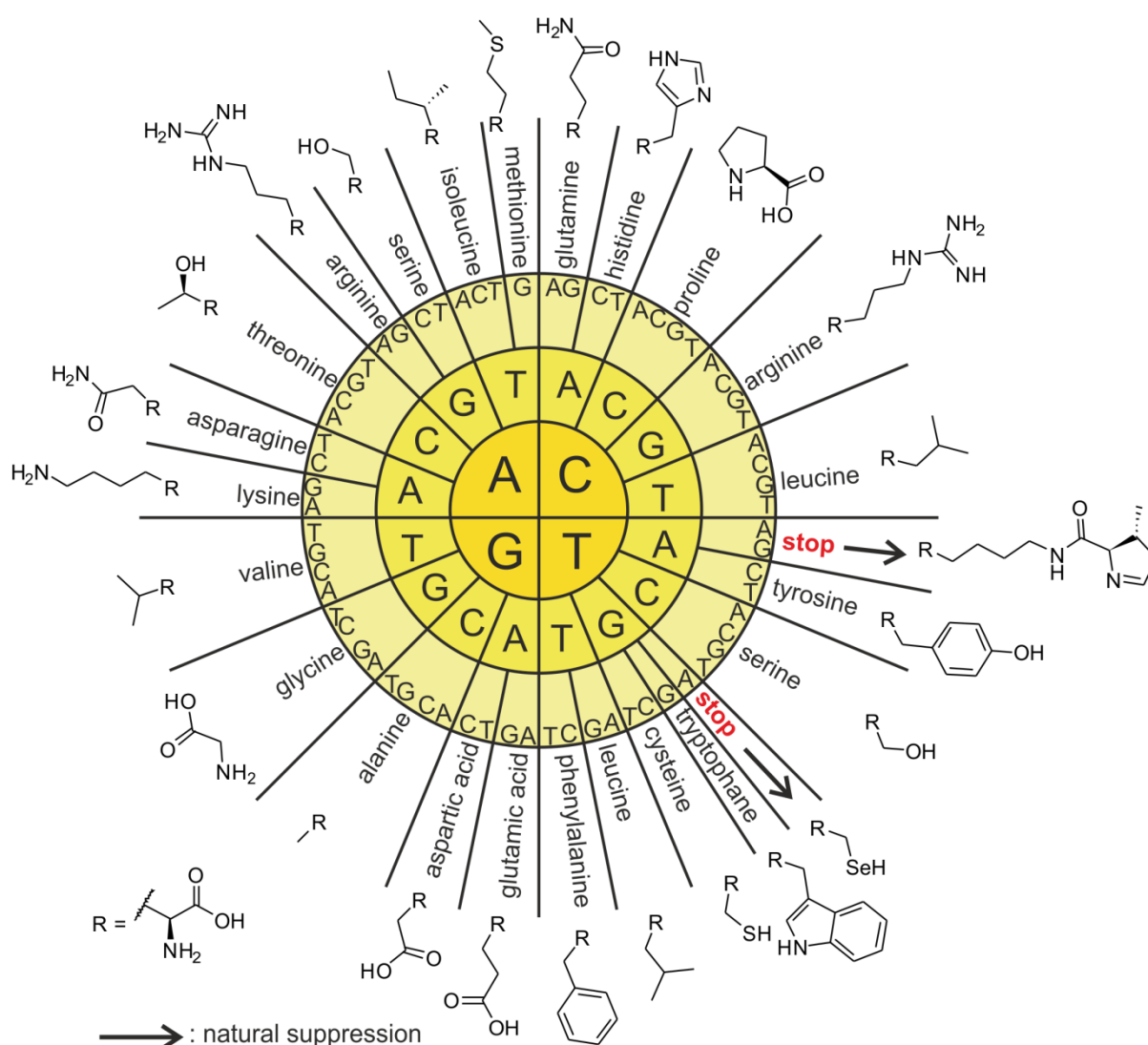


Figure 3: The genetic code. The consensus genetic code (DNA notation) used by the majority of organisms with semicanonical amino acids encoded by reassigned stop codons.

3.1.2 Linkage of DNA and protein sequence: tRNA aminoacylation and protein translation

Protein synthesis takes place at the ribosome (see Figure 4), an RNA-protein complex, and utilizes adaptor molecules to physically connect a codon to an amino acid. These adaptor molecules are called tRNAs, small RNA molecules (approx. 70 bp) assuming a defined three-dimensional structure that presents three nucleotides (anticodon) for codon recognition and are aminoacylated with the respective amino acid at the 3' terminus^[6–8]. Aminoacylated tRNAs are delivered to the ribosomal binding site for aminoacylated tRNAs (A-site) via an elongation factor (EF-Tu in bacteria) and can be accommodated in the site if a matching mRNA codon is also present. The nascent peptide chain attached to the tRNA currently in the peptide transfer site (P-site) is transaminoacylated and the ribosome translocates, ejecting the P-Site tRNA and accommodating the recent tRNA in the P-Site, freeing the A-site for the next tRNA.

A number of checkpoints exist to ensure high fidelity of translation. The most basic control mechanism lies in the ribosomal A-site and the tRNA-binding site of EF-Tu, which reject overly large, highly charged, non- α or D-amino acids due to their structural configuration^[9–12]. Recent findings indicate that EF-Tu binding and ribosomal accommodation are finely tuned, preferring binding of aminoacylated tRNAs within a narrow affinity window: Naturally, tRNA bodies with low binding affinity are aminoacylated with high-affinity amino acids and *vice versa*, ensuring similar average binding constants for all natural aminoacylated tRNAs, whereas misacylated tRNAs have a higher chance to be rejected for binding too weakly or too strongly. This principle is known as thermodynamic compensation^[13]. The major control checkpoint for translational fidelity, however, lies in the aminoacylation reaction itself. The correct linkage of tRNA anticodon to amino acid is ensured by aminoacyl-tRNA synthetases (aaRSs) that specifically recognize the amino acid and the tRNA body, making these enzymes critical determinants for the integrity of the genetic code^[14–16] (Figure 5). AaRSs fall into one of two structural classes (compared in Table 1) with various subclasses, adopting specific folds and many aaRSs possess editing domains to deacetylate near-cognate amino acids. Nevertheless, it could be shown that misacylated tRNAs are directed to the ribosome and used in translation in many cases^[9,17].

Introduction

aminoacylation

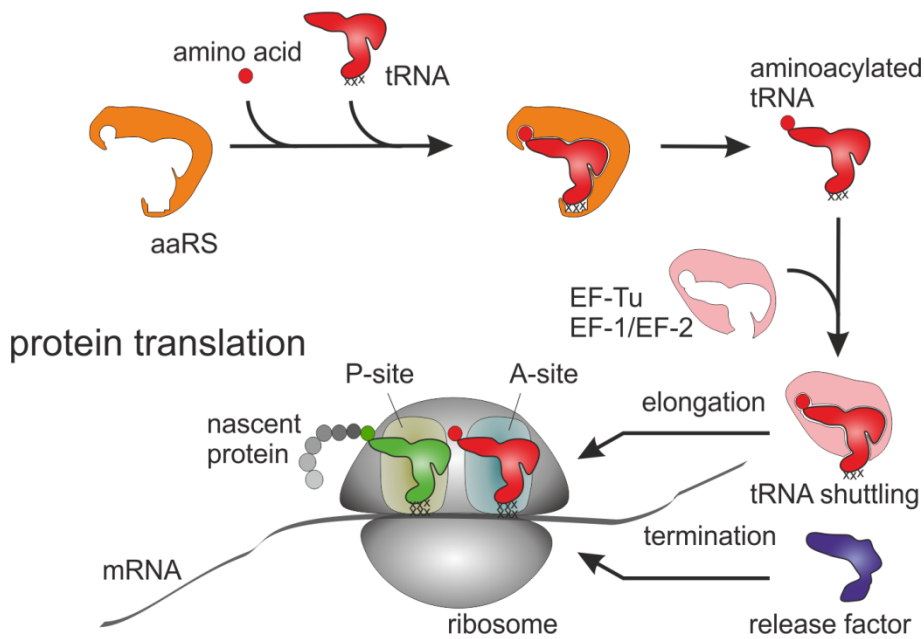


Figure 4: Protein translation: Aminoacyl-tRNA synthetases (aaRS) aminoacylate cognate tRNAs with the corresponding amino acids. Aminoacylated tRNAs bind to elongation factor EF-Tu (bacteria) or archaeal/eukaryotic elongation factors for trafficking to the ribosome. Aminoacylated tRNAs accommodate the A-site, recognizing the presented codon on the mRNA by base pairing. In the next step, the nascent protein chain is transferred from a tRNA located in the ribosome's P-site, elongating the chain by one amino acid. Finally, the ribosomal subunits briefly dissociate and migrate on the mRNA, ejecting the tRNA from the P-site and placing the tRNA formerly in the A-site, now carrying the nascent protein chain, in the P-site, freeing the A-site. A stop codon in the A-site triggers accommodation of a release factor instead of a tRNA, leading to dissociation of the ribosome and release of the protein chain (termination).

Despite all mechanisms for ensuring high translational fidelity, the genetic code has proven to be quite malleable, and mistranslation occurs frequently, raising speculations about contributions to evolutionary adaptation processes^[12,18,19].

Introduction

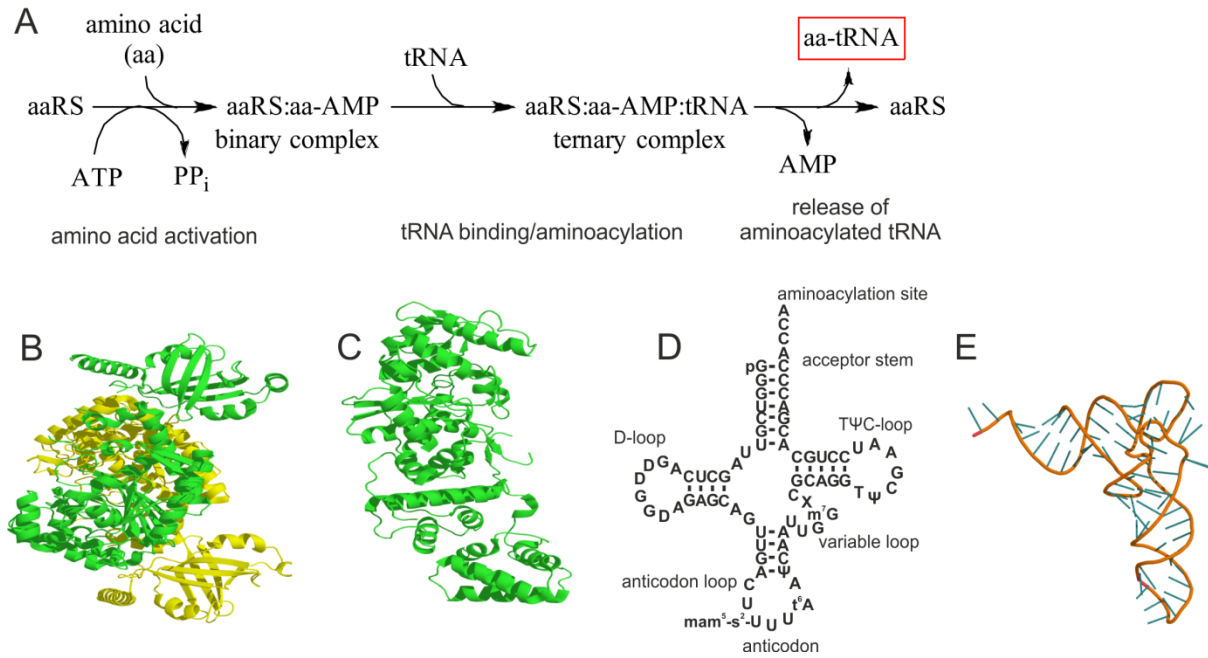


Figure 5: Aminoacylation of tRNAs. A: Schematic representation of the aminoacylation process. The three major steps are the activation of the amino acid using ATP, the tRNA binding and aminoacylation and the release of aminoacylated tRNA. Each of these steps can be rate-limiting. B: Structure of a homodimeric class 2 LysRS (from PDB-ID 1E24). C: Structure of a monomeric class 1 LysRS (from PDB-ID 1IRX). D: Schematic representation of a tRNA^{Lys}, revealing the cloverleaf structure and a high ratio of modified nucleotides^[6]. Unusual nucleotides are thymine (T), dihydrouridine (D), pseudouridine (Ψ), 7-methylguanosine (m⁷G), threonylcarbamoyladenosine (t⁶A), 3-(3-amino-3-carboxypropyl)uridine (X), 5-methoxyaminomethyl-2-thiouridine (mam⁵-2^s-U). E: Structure of tRNA^{Lys} (from PDB-ID 1FIR).

Table 1: Properties of class 1 and class 2 aminoacyl-tRNA synthetases.

	class I	class II
active-site fold	parallel β-sheet	antiparallel β-sheet
tRNA binding	acceptor stem minor groove side, variable loop faces solvent, usually with anticodon recognition	acceptor stem major groove side, variable loop faces protein, generally no anticodon recognition
quaternary structure	predominantly α, rarely α ₂	predominantly α ₂ , rarely α ₄ , (αβ) ₂
aminoacylation site	2'-OH	3'-OH
Enzymes	CysRS GluRS IleRS LysRS-I LeuRS MetRS GlnRS ArgRS ValRS TrpRS TyrRS	AlaRS AspRS PheRS GlyRS HisRS SepRS LysRS-II AsnRS PylRS ProRS SerRS ThrRS

3.2 Incorporation of noncanonical amino acids into proteins

3.2.1 Posttranslational modification and misaminoacylation of canonical tRNAs

Naturally, non-proteinogenic amino acids are introduced posttranslationally by enzymatic modification of canonical amino acids. Important posttranslational modifications (PTMs) of amino acids are hydroxylation (Pro), phosphorylation (Ser/Thr, Tyr, Asp, His), methylation/acetylation (Lys), modification with small molecules, cofactors or proteins (glycosylation, ubiquitination, prenylation, etc. of nucleophilic amino acids), and complex modification of ribosomal peptides (macrocyclisation, reduction/oxidation, etc.)^[20–22]. Additionally, several organisms and organelles use a slightly altered version of the genetic code with several reassigned codons^[23,24].

Direct incorporation of noncanonical amino acids into nascent proteins has also been observed. While most aaRSs show remarkable specificity for their cognate amino acid, misacylation can happen with structurally similar amino acids^[25]. This does not in all cases lead to incorporation of noncanonical amino acids, but to misincorporation of canonical amino acids^[9,22]. Leu and Ile, for example, are frequently swapped during the aminoacylation step. However, some noncanonical amino acids are structural surrogates to canonical ones, leading to misincorporation, often exerting toxic effects in organisms. The principal mechanisms are outlined in Figure 6. The most prominent example for a faultily activated amino acid is canavanine, an arginine surrogate produced by several plant species as a toxin^[26,27]. This inability of aaRS to distinguish close structural analogs from their cognate substrates can be utilized for amino acid replacement, as shown in Figure 7A.

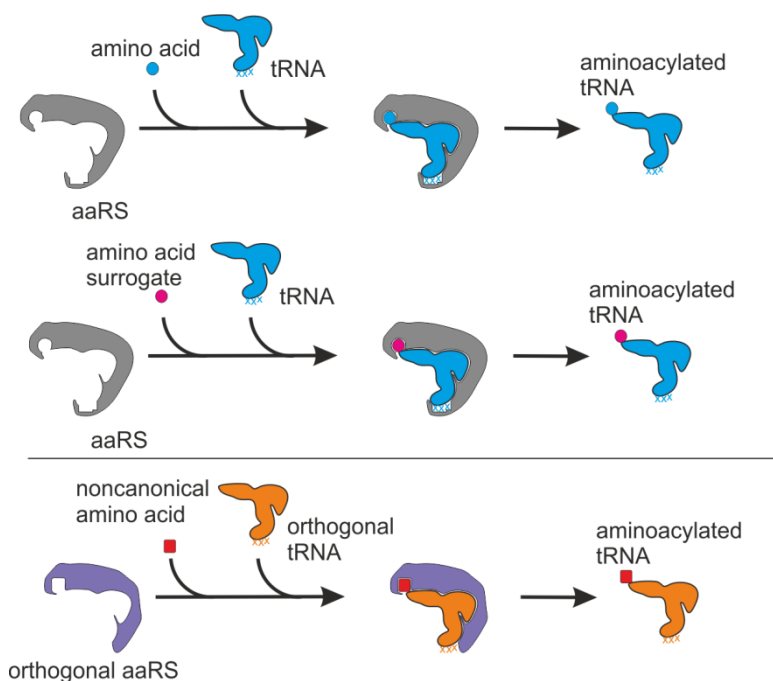


Figure 6: Incorporation of noncanonical amino acids into target proteins. A: Scheme of aminoacylation reactions. Top row: natural aminoacylation process with a canonical aaRS/tRNA pair charging the cognate amino acid. Center row: in the absence of a cognate amino acid, an isostructural surrogate can be charged via a canonical aaRS/tRNA pair. Bottom row: a non-isostructural noncanonical amino acid is not accepted by the canonical aminoacylation machinery and requires a distinct additional aaRS/tRNA pair. This aminoacylation does not interfere with canonical aaRSs, tRNAs and amino acids, and is thus orthogonal.

Introduction

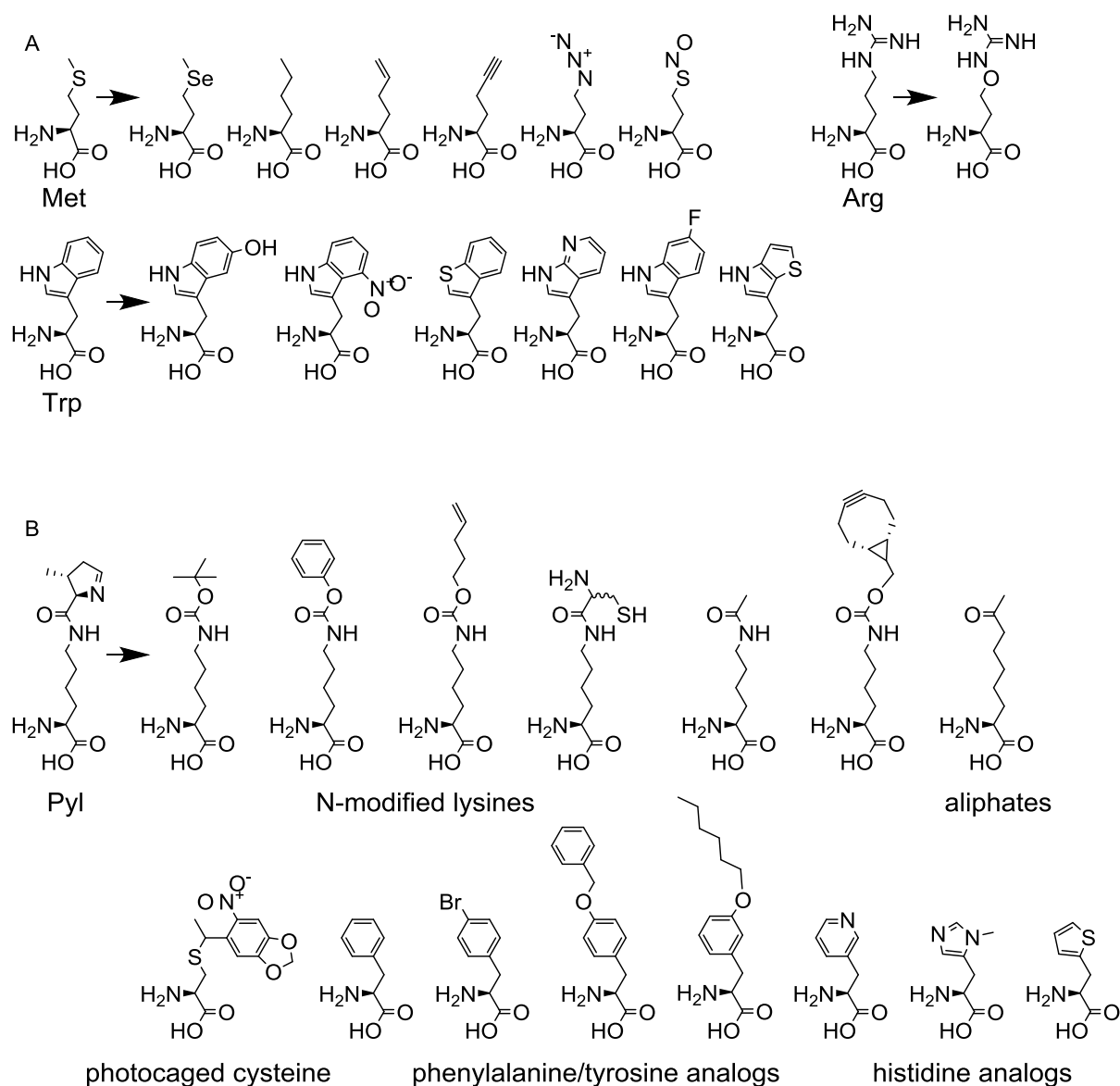


Figure 7: Some noncanonical amino acids. A: Selection of surrogate amino acids of methionine, arginine and tryptophan, which have been successfully incorporated into proteins, replacing their corresponding canonical amino acid. B: Noncanonical amino acids that are not accepted by any known canonical aaRS/tRNA pair, which have been incorporated into proteins using native or engineered variants of pyrrolysyl-tRNA synthetase and tRNA^{Pyl}.

3.2.1 Codon suppression

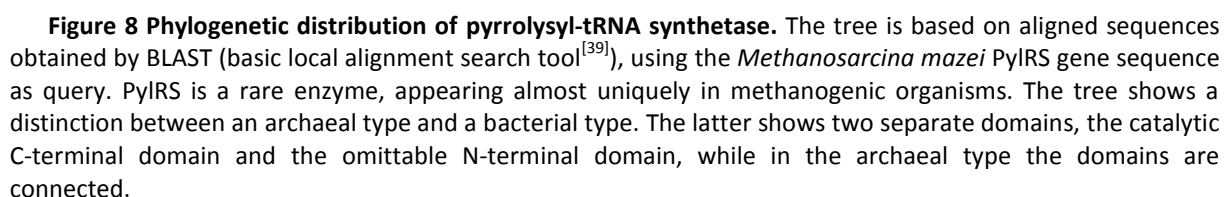
Suppression events occur when a codon is not decoded according to its canonical assignment. Natural stop codon suppressor tRNAs exist, assigning nonsense codons with canonical amino acids instead of acting as a stop signal^[28]. The most prominent example is supE, a tRNA^{Gln} with a CUA anticodon, efficiently suppressing the UAG stop signal in bacteria^[29,30]. This suppression is also naturally used to incorporate the semi-canonical amino acids selenocysteine (Sec) at opal (UGA) stop codons and pyrrolysine (Pyl) at amber (UAG) stop codons. However, these amino acids employ different decoding strategies. Pyrrolysine is incorporated into a family of methylamine demethylases in some phyla of methanogenic archaea and bacteria^[31] (Figure 8). It is activated by a distinct pyrrolysyl-tRNA synthetase (PylRS) and charged to a special tRNA^{Pyl} with a CUA anticodon, for suppression of the amber stop codon^[32,33]. While certain mRNA context elements (pyrrolysine insertion signal, "PYLIS") have been proposed and could be shown to enhance the suppression over termination events at amber stop codons, no consensus sequence between different species could be found. As a confirmation, no unambiguous amber stop codon exists in species with a native PylRS/tRNA^{Pyl} pair, and upon transplantation of the orthogonal pair to other host species, efficient stop codon suppression occurs at all amber stop codons^[34,35]. PylRS and its engineered variants have been used for incorporation of a wide array of noncanonical amino acids (Figure 7B).

In contrast, tRNA^{Sec} suppresses the ochre stop codon UGA and is initially charged with serine by SerRS, which is then used to synthesize Sec on-tRNA by serine phosphorylation, elimination and selenide addition to the resulting dehydroalanine. No distinct SecRS exists, and decoding UGA as Sec requires both a downstream signal on the tRNA and delivery of Sec-tRNA^{Sec} by a specialized additional elongation factor^[36]. The SECIS element on the mRNA is recognized by the Sec-tRNA^{Sec} binding elongation factor EF-Sec which promotes incorporation of Sec at the opal codon^[37,38].

Sec-containing proteins have been found ubiquitously in every biological phylum, making it a *de-facto* canonical amino acid with an unusual incorporation pathway. In contrast, while Pyl is incorporated via a more canonical pathway, it should not be regarded as fully canonical due to its limited distribution^[40].

A similar system exists for the generation of Cys-tRNA^{Cys}, where phosphoserine (Sep) is loaded onto tRNA^{Cys} by a distinct SepRS, and the resulting Sep-tRNA^{Cys} is used to generate Cys-tRNA^{Cys}^[41–43]. Sep is not incorporated into proteins, and this pathway solely comprises an alternative or redundant route to Cys-tRNA^{Cys} in some archaea^[42,43]. Nevertheless, this system has been adapted for Sep incorporation^[44].

The required structure for stop codon suppression with Sec and Pyl are shown in Figure 9C and D.



3.2.2 Deliberate incorporation of noncanonical amino acids

For technological purposes, novel noncanonical amino acids can be incorporated by two different general methodologies, supplementation pressure incorporation (SPI) or codon suppression (refer to Figure 6).

SPI utilizes the loose substrate recognition of a canonical aaRS for incorporation of isostructural amino acid surrogates^[45–47]. As non-cognate substrates are efficiently discriminated against, the concentration of the cognate canonical amino acid must be reduced, making auxotroph expression hosts and chemically defined media a necessity, as well as a starvation step before expression with the added isostructural amino acid^[48,49]. Advantages of this method are the easy application to an array of diverse noncanonical amino acid, easy adaptation for multiple different amino acids and often high efficiencies. The incorporation is residue-specific, fully replacing a canonical amino acid with the noncanonical one. This means that site-directed replacement is only possible when the canonical amino acid to be replaced occurs only once in the target protein sequence, which can generally be achieved for rare amino acids, especially Met, which is always found at the N-terminus. The disadvantages are the aforementioned demands of host organism and medium, reducing the general applicability, as well as the limitation to close analogs to canonical amino acids. Finally, it should be noted that multiple incorporations of amino acid analogs into the protein may not be tolerated. The surrogate amino acid usually cannot functionally replace the canonical one in all instances, prohibiting a proteome-wide replacement^[48]. The only exceptions are the long-known replacement of Met by selenomethionine^[50], and full replacement of Trp by L-β-(thieno[3,2-bi]pyrrolyl)alanine ([3,2]TPA) in an engineered *E. coli* strain^[51]. To overcome limitations of the amino acid recognition by the corresponding aaRSs, engineered aaRS with a broadened substrate spectrum^[52] and relaxed-specificity variants from natural sources^[53] have been utilized in some cases.

Suppression methods, on the other hand, rely on the establishment of an orthogonal pair of aaRS and tRNA in the host system^[54–56]. An orthogonal system exhibits no cross-reactivity with the host system, i.e. the orthogonal aaRS does exclusively activate the noncanonical amino acid and aminoacylates only the cognate tRNA. An orthogonal tRNA is not recognized by any canonical aaRS, allowing the defined incorporation of a noncanonical amino acid. This system also generally requires an unassigned codon, usually a stop codon, as naturally prescribed elongation at sense codons is greatly favored over suppression. With natural and evolved suppression systems established, this allows the suppression of stop codons and quadruplet codons in a large number of organisms including bacteria, yeast, nematodes and mammalian cells^[57–59].

While this method has a greater versatility than SPI, the noncanonical amino acids of interest are in many cases not accepted by natural or first-generation engineered orthogonal systems, making extensive directed evolution procedures necessary to generate aaRS/tRNA systems that activate the desired ncaa without interfering with the host's natural translation processes^[55,60].

As an alternative path to aminoacylated suppressor tRNAs, chemical and chemo-enzymatic aminoacylation strategies have been explored^[61,62]. However, despite considerable advances in this area^[63], the additional organic synthesis effort and the fact that chemically aminoacylated tRNAs are consumed stoichiometrically thus far prevented widespread applications of the method.

Introduction

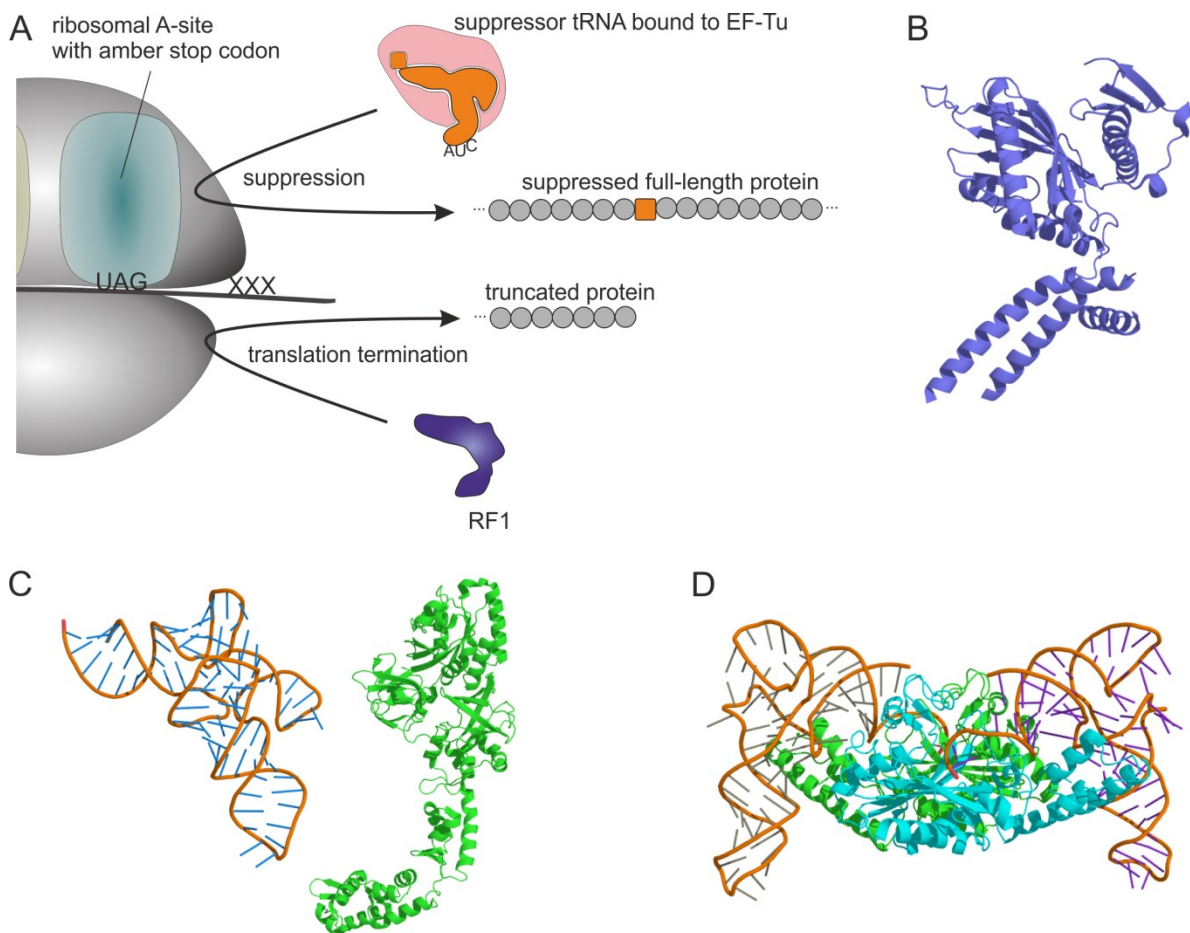


Figure 9: Stop codon suppression versus termination. A: A stop codon resting in the ribosomal A-site usually triggers binding of a release factor and dissociation of the protein biosynthesis complex. Alternatively, cytosolic aminoacylated suppressor tRNAs can decode the stop codon analogously to decoding of a sense codon, thus suppressing the termination event. In this case, protein biosynthesis continues until a non-suppressed stop codon (or the end of the mRNA) is reached. The schematic example shows the competition for an amber stop codon (UAG) between a charged amber suppressor tRNA and the bacterial release factor RF1, that triggers translation termination at amber codons. The ratio of suppression and termination event largely depends on the binding affinity of the aminoacylated suppressor tRNA to EF-Tu and to the ribosome. B: Structure of bacterial release factor RF1 (from PDB-ID 1RQ0). The overall protein structure mimics the L-shaped tRNA fold. C: Structure of tRNA^{Sec} (from PDB-ID 3LOU) and the specialized elongation factor EF-Sec (from PDB-ID 4ZU9) required for trafficking of the aminoacylated tRNA to the ribosome. tRNA^{Sec} is initially aminoacylated with serine by SerRS, which is converted to Sec on-tRNA. Delivery to the ribosome is mediated by EF-Sec and a downstream mRNA recognition element (SECIS). D: Homodimeric pyrrolysyl-tRNA synthetase with two molecules of tRNA^{Pyl} (from PDB-ID 2ZNI). PylRS and tRNA^{Pyl} constitute a system for suppression of amber stop codons independent of specialized RNA structures or proteins.

The major drawback of suppression methods is the competition with canonical processes. As no truly unassigned codon exists naturally, codon suppression with noncanonical amino acids competes with canonical translation (sense codon suppression), translation termination (stop codon suppression, shown in Figure 9A) or frameshift (frameshift/quadruplet codon suppression). This limits the efficiency of amber stop codon suppression, the most efficient form, to app. 50 % protein yield compared to normal expression for suppression of a single stop codon^[47,55,57–59]. As every release event prevents the production of full-length protein, suppression of more than one stop codon becomes exponentially less likely. The suppression of sense and frameshift codons is even more disfavored^[64,65]. A final limitation to suppression methodologies is a dependency on codon context. While it has been shown that defined downstream elements (PYLIS) do not play a significant role for

Introduction

stop codon suppression using PylRS/tRNA^{Pyl} [8,34,35], the efficiency of suppression varies greatly dependent on the position of the suppressed codon. Usually, stop codons near the N-terminus of a protein are favorable, but codon context can fully prevent suppression events of both stop and sense codons^[66–68]. The mechanistic background of this phenomenon is largely unknown, but efforts have been made to find an optimal codon context for stop codon suppression^[68]. For practical applications, a screening of stop codon positions is necessary for each desired target protein. Nevertheless, the possibility of site-directed incorporation of virtually any ncaa makes codon suppression an important and versatile tool for protein engineering. Table 2 compares the two incorporation techniques SPI and codon suppression.

Table 2: Comparison of incorporation techniques

	SPI	Suppression
mode of incorporation	residue-specific	site-specific
amino acid composition of yielded protein	canonical amino acid replaced by ncaa	additional ncaa
amino acid limitations	isostructural surrogates	undetermined, limited by evolvability of orthogonal aaRS
target gene	unchanged	mutated (insertion of suppression codons)
amino acid activation	natural aaRS of host	additional orthogonal aaRS
ncaa limitation	isostructural surrogate for natural amino acid, must bind EF-Tu/ Ribosome	activation by engineered aaRS, must bind EF-Tu/ Ribosome
competition with canonical processes	outcompeted by activation of standard aaRS substrate amino acid	competition with translation termination and natural suppressors (stop codon suppression), outcompeted by elongation with canonical aa-tRNA (sense codon suppression)
host requirements	auxotroph for amino acid to be replaced by surrogate	absence of strong natural suppressors, additional strain engineering to reduce competing processes advantageous
multiple incorporations	possible	difficult, requires strain engineering
multiple incorporations of several different ncaa	possible, requires multiple auxotrophies	difficult, requires mutually orthogonal aaRS/tRNA pairs and two or more suppressible codons

3.3 Engineered aminoacyl-tRNA synthetases: rational and evolved variants

The first engineered orthogonal pair for amber suppression was presented in 2001, based on an engineered *Methanocaldococcus jannaschii* TyrRS (*MjTyrRS*) and its cognate tRNA^{Tyr} with the anticodon changed to CUA, evolved for incorporation of O-methyltyrosine^[56]. However, it was not before the discovery of the orthogonal pair for pyrrolysine in 2004^[32] that incorporation of noncanonical amino acids via stop codon suppression became widely used. With few exceptions, engineered aaRSs for incorporation of noncanonical amino acids are based on *MjTyrRS* or *PylRS* from *Methanosarcina mazei* (*MmPylRS*) or *barkeri* (*MbPylRS*). Exceptions include engineered LeuRS or AspRS and are orthogonal only in selected host organisms (recently reviewed^[57,58,69]).

Engineered *MjTyrRS*s exist for a activation of a large number of phenylalanine derivatives, although some proposed variants have been shown to be only residually active or inactive^[70].

Engineered *PylRS*s have been developed mainly for incorporation of lysine derivatives, with a large number of noncanonical amino acids already accepted by the wild-type enzyme. The low substrate specificity and ability of the system to be directly transplanted between different host organisms allowed incorporation of a range of lysine derivatives for protein modification^[58]. The development of a *PylRS* activating Phe led to second-generation variants activating an array of Phe derivatives^[71–73]. Recently, novel *PylRS* variants were found for His analogs^[74] and Cys derivatives^[75]. So far, most incorporated analogs show structural similarity to either Pyl or Phe, which can be explained by the evolutionary connection between these aminoacylation activities^[76].

PylRS is the predominantly used orthogonal suppression system due to its high evolvability and general applicability^[54]. The crystal structure shows very few specific interactions with the natural substrate Pyl, exerted mainly by the residue N346, forming a hydrogen bridge to the ring-N of the Pyl head group. This residue has thus become known as gatekeeper residue, as it limits substrates to lysine analogs modified with *N*_ε-carbonyl groups. Other than that, the amino acid binding pocket of *PylRS* is lined with hydrophobic amino acids. This allows even for the wild-type *PylRS* to accommodate a wide range of noncanonical amino acid that show the recognized structural features (*N*_ε-carbonyl modified lysine analogs with a hydrophobic head group)^[77]. Based on these findings, rational engineering efforts have been undertaken to enlarge or reduce the size of the binding pocket, or to remove the gatekeeper residue to change the amino acid specificity. These efforts resulted in highly promiscuous variants activating classes of noncanonical amino acids that could not previously be incorporated^[72,78]. The structure of *PylRS* is shown in Figure 10A.

Generally, for incorporation of a given noncanonical amino acid it is impossible to rationally deduce which residues to change within the amino acid binding pocket, making a selection from a library necessary. The most common method to create such a gene library is site saturation mutagenesis, that is using PCR-based methods to replace codons encoding amino acids to be randomized by the degenerate NNK (or NNN) codon. Compared to NNN mutagenesis, NNK halves the amount of possible library members and eliminates two stop codons, while still allowing all 20 amino acids. This enables smaller libraries with the same diversity, making more randomization possible. The number of different library members is 32ⁿ, 32 being the number of possible codons expressed in NNK and n being the number of NNK positions. While libraries can become very large, one has to make sure that the entire library can be screened, that means that every library member has to be transformed into a cell for selection^[79]. This limits the library size, as only a limited amount of transformants can be generated per mg of library DNA. For *E. coli*, the maximum transformation efficiency (depending on

Introduction

strain and transformation method) is $10^9 - 10^{10} \text{ mg}^{-1}$ for small amounts of plasmid DNA used for transformation^[80].

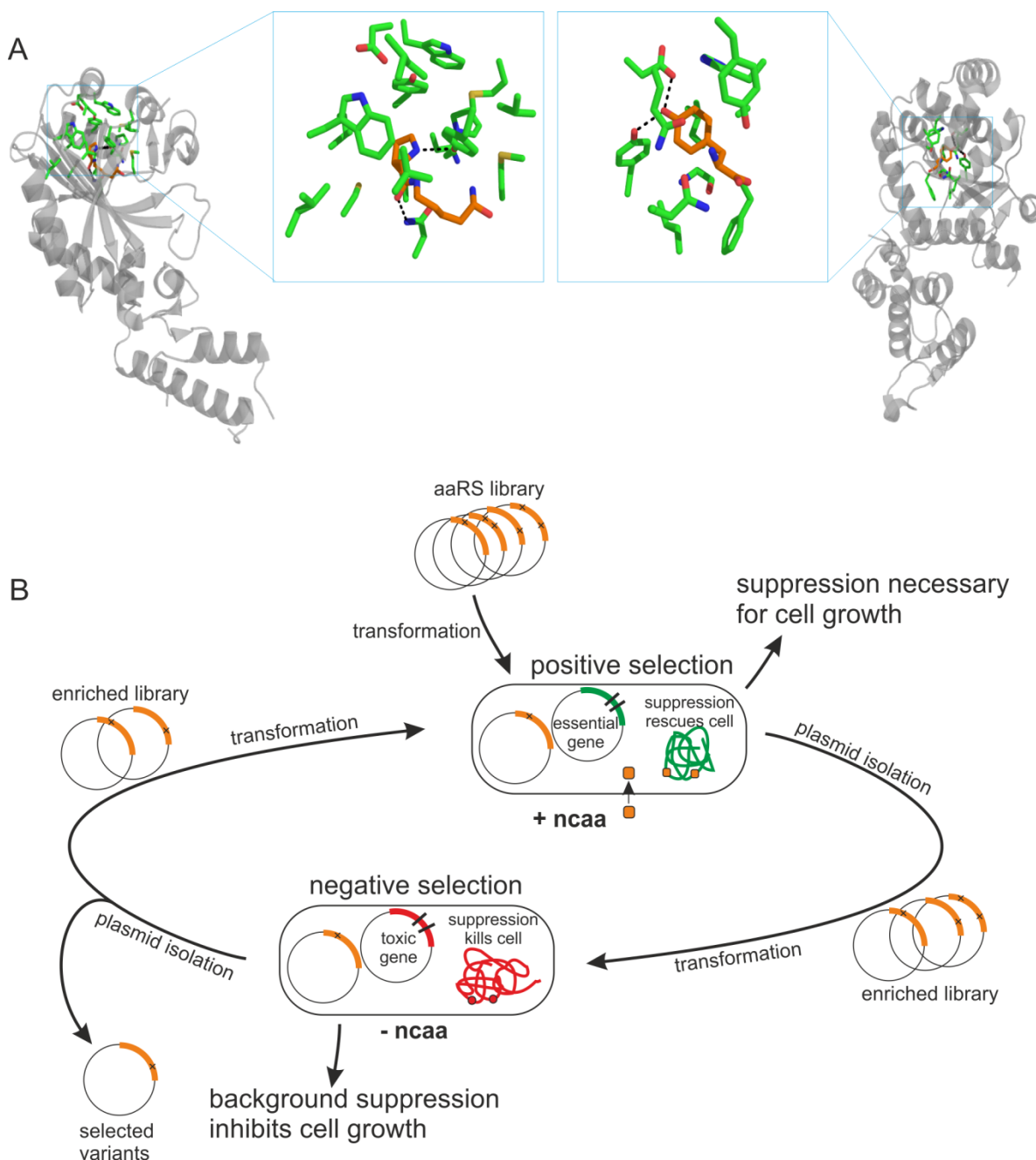


Figure 10: Engineering of orthogonal aaRS. A: Structures of the most commonly used orthogonal aaRS, *MmPylRS* (left, from PDB-ID 2ZIM) and *MjTyrRS* (right, from PDB-ID 1J1U) shown as monomers. The amino acid binding pockets are shown as green sticks, the bound amino acid substrates are shown in orange. B: Schematic selection procedure relying on positive and negative selection. For positive selection, the aaRS library is co-transformed with an amber-disrupted essential gene (positive selector). In the presence of the noncanonical amino acid of interest, successful suppression allows cell growth. The isolated enriched library is then co-transformed with an amber-disrupted toxic gene. In the absence of the desired ncaa, only aaRS variants that accept natural amino acids allow amber suppression, inactivating the transformants. The two selection steps are iteratively cycled to enrich the desired aaRS variants with aminoacylation activity for the ncaa of interest.

Introduction

Taking into account realistic conditions, i.e. uneven distribution of the 32 possible codons due to differences in synthesis efficiency, codon usage and realistic transformation procedures, a maximum of app. 10^8 total library members can be screened efficiently, correlating to a maximum of five NNK randomizations. Alternative randomization strategies, like the use of NDT codons do not allow for all amino acid combinations, while novel approaches, such as the “22C-trick”^[81], require elaborate primer combinations.

The selection of aaRSs with novel activities from a library usually requires an iterative positive and negative selection process, schematically depicted in Figure 10B. During positive selection, colonies can survive if they can suppress a disrupted essential gene in the presence of the desired noncanonical amino acid, selecting all colonies expressing active aaRS variants. For negative selection, the cells carry a disrupted toxic gene, eliminating all variants that show codon suppression in the absence of the desired noncanonical amino acid, while colonies expressing orthogonal aaRS variants can grow. Active, orthogonal variants are enriched in iterative cycles of positive and negative selection^[55,56,82].

To mitigate the limitations of competing termination processes, extensive strain engineering feats have been undertaken: as translation termination at amber stop codons are mediated by RF1, a logical conclusion is to remove RF1 from the bacterial genome. However, since the release factor is essential, this also requires replacement of amber stop codons throughout the genome, either entirely, or after essential genes, making it a challenging endeavor. Nevertheless, both routes successfully led to strains with greatly increased suppression efficiency and the ability to suppress several stop codons in a row^[83–86]. As alternative strategies, conditional inactivation of RF1^[87] or making it nonessential by “fixing” release factor 2 (RF2) for translation termination at amber stop codons^[88]. For quadruplet suppression, orthogonal ribosomes have been developed, preventing translation termination^[89,90]. Sense codon suppression is the least efficient suppression method and has not been fully achieved, although recently initial successes have been realized in the suppression of the rare arginine codon^[91–93] and interesting methodologies have been proposed to remove the Ile assignment of the AUA codon^[94].

As the limiting step for amino acid activation is the ability to accommodate the amino acid, most engineered aaRS variants are highly promiscuous, as structural features needed for specificity towards the natural substrate are removed^[71,95].

Using both rational and evolutionary methods, a large number of structurally diverse amino acids have been incorporated into proteins via codon suppression for different applications.

3.4 Synthetic biology: Employing noncanonical amino acids and new-to-nature chemical reactions in biochemical pathways

3.4.1 Combination of biosynthesis and translational incorporation of noncanonical amino acids

A strong limitation to widespread applicability of proteins containing ncaas is the laborious, often small-scale amino acid production. Large-scale industrial applications require a more economical production by biosynthetic processes, which so far could be installed for few noncanonical amino acids.

While the mechanism of Pyl biosynthesis from two lysine residues has only been recently elucidated^[96–98], it has been known that heterologous expression of Pyl biosynthesis enzymes enables *E. coli* to produce Pyl, which can subsequently be incorporated at amber stop codons, if the orthogonal pair PylRS/tRNA^{Pyl} is intracellularly present as well^[34]. Pyl biosynthesis, however, does not occur very fast and produces only low amounts of intracellular Pyl. It has been observed that addition of ornithine apparently increases the amount of Pyl produced, leading to the initial hypothesis that ornithine is a substrate in Pyl biosynthesis. In fact, an elevated intracellular ornithine concentration enables the production of pyrroline-carboxyllysine via the Pyl biosynthesis machinery, replacing one of the two substrate lysines with ornithine^[34,99,100]. These systems have been used to generate the first instances of target proteins suppressed with *in situ* biosynthesized amino acids. Recently, the Pyl biosynthesis machinery has been employed for the production of a more elaborate and functional Pyl analog, 3S-ethinylpyrrolysine for subsequent translational incorporation and protein decoration via photoclick reaction^[101]. All pathways are depicted in Figure 11.

A similar approach has been done employing the SPI methodology for the production of proteins with Met replaced by azidohomoalanine (Aha) produced intracellularly from azide and native O-acetylhomoserine by hijacking the Met biosynthesis pathway^[102]. It had been previously shown that the Cys and Met biosynthesis pathways can be exploited to generate diverse alanine and homoalanine derivatives^[103]. Despite tremendous efforts to optimize every component of the translation machinery for stop codon suppression, both *in vivo* and *in vitro*^[104], the combination of biosynthetic production of noncanonical amino acids with their translational incorporation remains a rare feat, that opens new possibilities in the feasible production of modified proteins for research and technological applications.

Diametric to such self-contained microbial factories for modified proteins is the combination of noncanonical amino acids and engineering of biochemical pathways to install a “genetic firewall”. By prescribing the incorporation of noncanonical amino acids to restore or maintain essential cell functions, it seems possible to generate organisms that are inescapably dependent on this noncanonical amino acid as a medium additive. These organisms would live with an altered genetic code, fully preventing the exchange of genetic information with the environment^[105–107]. Initial successes were made with cells adapted to the thymine analog chlorouracil^[108,109], but only very recently, cells adapted to or crucially dependent on noncanonical amino acids were reported^[51,110,111].

Introduction

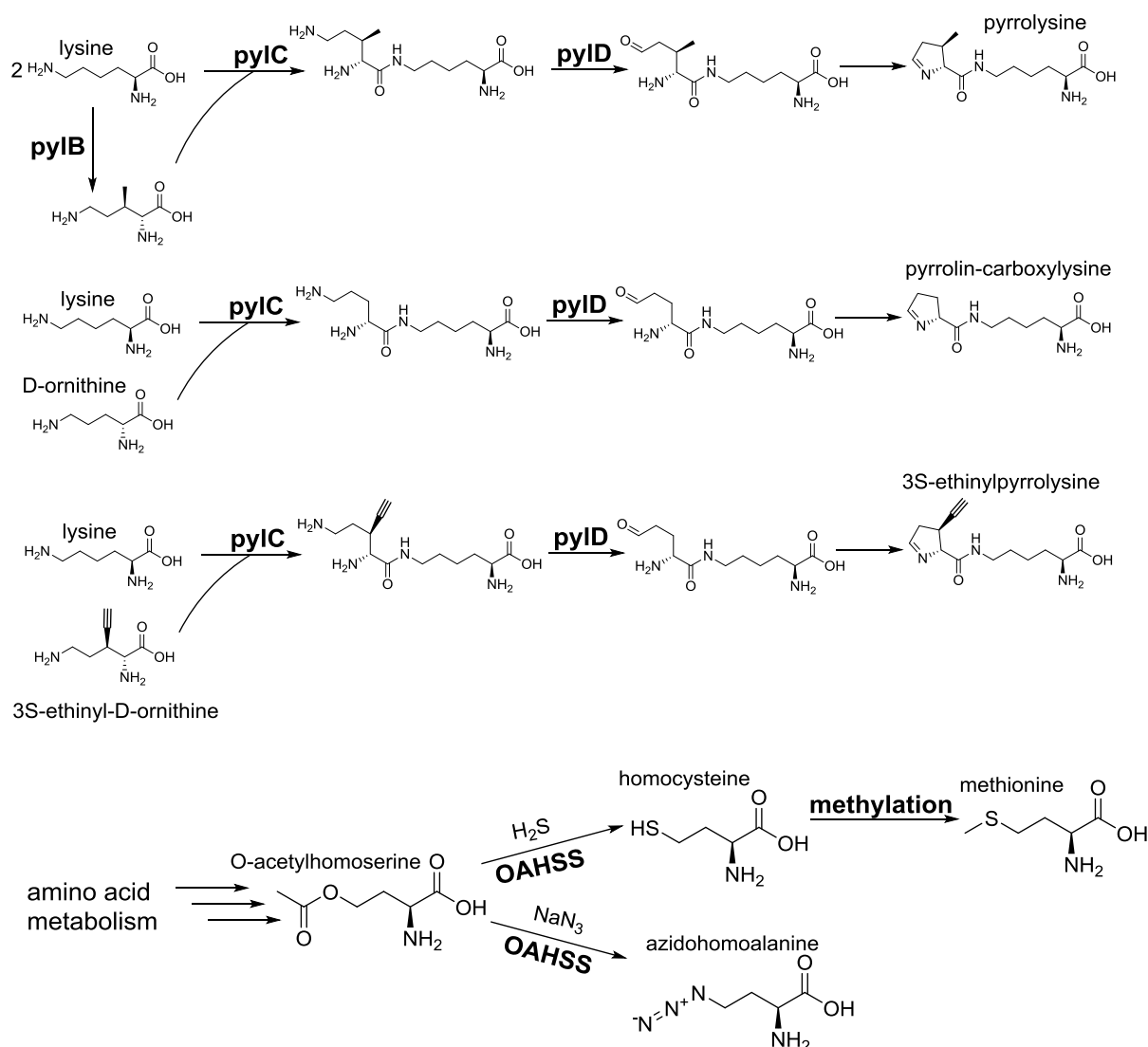


Figure 11: Biosynthetic pathways to noncanonical amino acids. Top: Synthesis of pyrrolysine from two lysines via the action of methylornithine synthase (pylB), methylornithine-lysine pseudopeptide synthase (pylC) and pyrrolysine synthase (pylD), synthesis of pyrroline-carboxyllysine from lysine and ornithine, omitting the pylB-catalyzed rearrangement step and synthesis of 3S-ethynylpyrrolysine from 3S-ethynylornithine and lysine. Bottom: Biosynthesis of methionine via homocysteine generated from direct sulphydrylation of O-acetylhomoserine by O-acetylhomoserine sulphydrylase (OAHSS). OAHSS accepts a wide range of nucleophiles as substrate, enabling the biosynthesis of diverse homoalanine derivatives including azidohomoalanine. The Cys biosynthesis via sulphydrylation of O-acetylserine by O-acetylserine sulphydrylase (OASS) can be analogously exploited as a biosynthetic route to alanine derivatives.

3.4.2 Bioorthogonal chemistry

While no artificial biosynthesis pathway exists for chemical modification of proteins containing noncanonical amino acids yet, major efforts have been made to bring chemical reactions into a biological context. In many cases, noncanonical amino acids are incorporated into proteins for subsequent posttranslational chemical modification. The use of stop codon suppression allows precise installation of the desired modification, as long as a bioorthogonal reaction type is chosen. Bioorthogonality requires a very low abundance of the involved functional groups in natural environments. The most prominent examples are noncanonical amino acids with alkene or alkyne functional groups. Both are nearly absent in the cytoplasm of bacteria and eukaryotes and highly stable in the cytosol and enable a variety of modification reactions, including thiol-ene^[112] and thiol-yne^[113] reaction, copper-induced^[114] and ring strain promoted^[115] Huisgen cycloaddition, Sonogashira^[116] and Glaser-Hay^[117] coupling, olefin metathesis^[118–120] as well as tetrazine-^[121] and photoclick^[121–123] reactions. (see Figure 12)

Noncanonical amino acids greatly expand the possibilities of enzyme engineering, as they can be incorporated to modulate activity and stability, leading to more feasible catalysts^[124], and can modulate the properties of proteins^[125,126] and peptides^[127], with an additional potential for immobilization, fluorescent labeling and other applications. As ncaas can endow proteins with functional groups not accessible within the natural range of side-chains and cofactors, the *de novo* generation of catalytic activities is possible. Of special interest are ncaa that chelate metal ions, leading to novel artificial metalloenzymes and peptides with unique properties, including an enzyme for catalysis of Friedel-Crafts acylations^[128–130].

Additionally, recent attempts to generate enzymes with new-to-nature catalytic functions yielded protein-based catalysts for olefin metathesis and other formerly purely chemical reactions^[131–133].

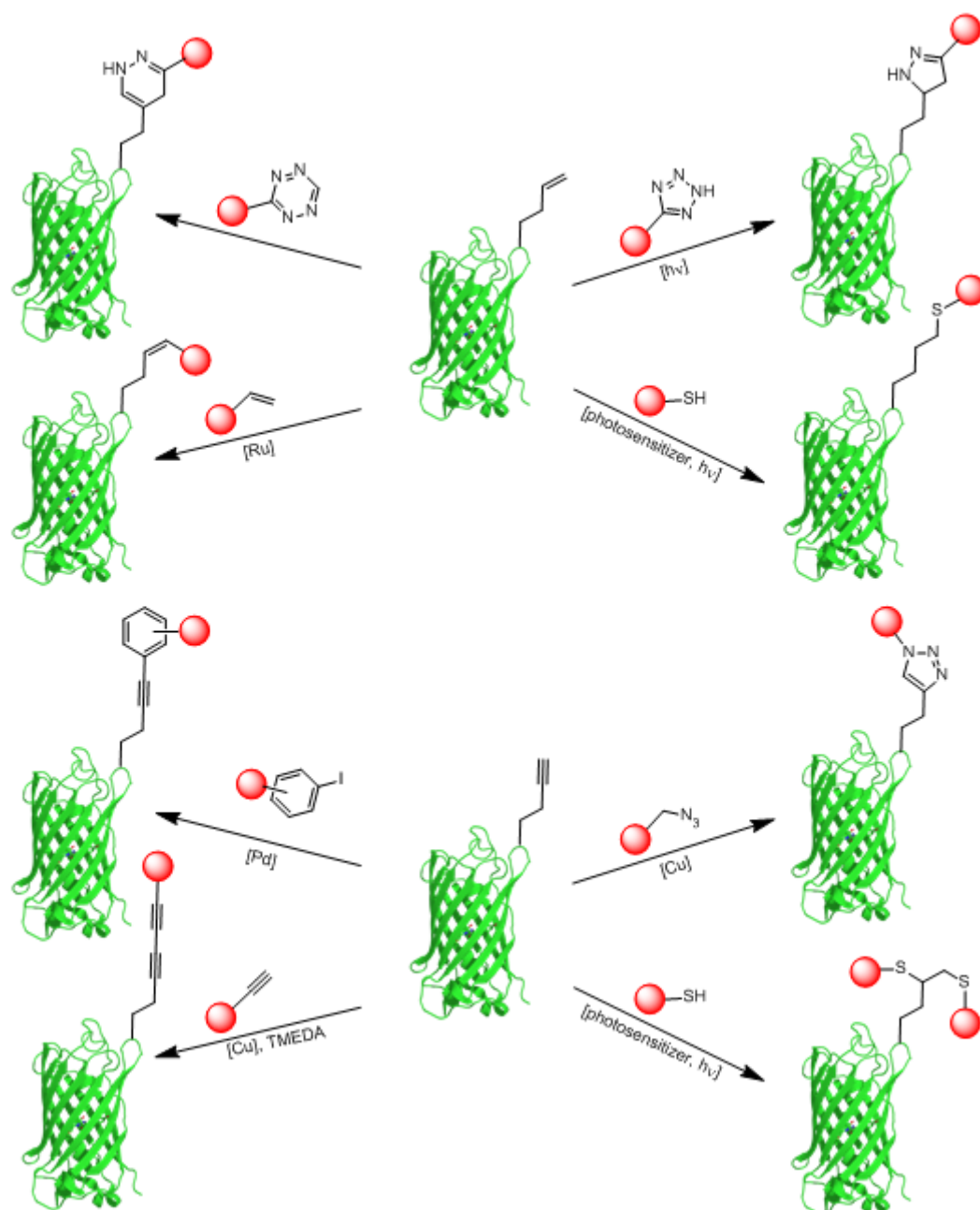


Figure 12: Posttranslational chemical reactions at noncanonical amino acid with alkene and alkyne moieties (schematic representations). Clockwise, starting top left: tetrazine-click reaction (a Diels-Alder variant), tetrazole photoclick reaction, radical thiol-ene conjugation, ruthenium-catalyzed olefin metathesis (at alkenes); Sonogashira coupling, Huisgen-Sharpless cycloaddition, radical thiol-yne reaction, Glaser-Hay reaction (at alkynes). See text for references. Some reactions require activated groups or restrictive reaction conditions.

3.5 Model proteins

In this work, three different model proteins were used (see Figure 13):

The green fluorescent protein (GFP), originally found in the jellyfish *Aequorea victoria*^[134], and its engineered derivatives are often used as reporter and model proteins due to their unique spectroscopic features^[135,136]. The protein forms a barrel structure of β -sheets interconnected by loop regions, with a strand running through the barrel. Three amino acids on this strand (S65/Y66/G67) form the fluorophore by autocatalytic dehydration and oxidation with environmental oxygen. Being the first example of genetically encoded fluorescence, GFP became widely used as reporter protein and fusion tag for localization monitoring in living cells and organisms, with many derivative fluorescent proteins engineered for different spectroscopic properties. Enhanced green fluorescent protein (EGFP) has been engineered for improved fluorescence quantum yield and can be expressed in good yields in most host organisms^[137]. Position N150 has been established as easily suppressible and does not have a significant impact on the overall protein structure, being located on the protein surface in the transition region of β -sheet to loop^[92]. These features make EGFP(N150amber) an ideal model protein for suppression experiments, as the formation of full-length protein can be easily monitored online by detection of the green color. Protein engineering also led to several variants with altered spectral characteristics, including enhanced yellow fluorescent protein (EYFP)^[138,139].

Lipases are the most versatile catalysts used for biotechnological or organic chemistry applications. *Thermoanaerobacter thermohydrosulfuricus* lipase (TTL) shows extraordinary thermal and solvent stability, and has been extensively probed for the influence of ncaas replacing canonical amino acids^[124]. With an already established suppressible site D221, TTL is a useful model for the influence of codon suppression on catalytic properties and withstands a wide range of conditions used in post-expression chemical protein modification^[92].

Barstar is a small ribonuclease inhibitor consisting of 90 amino acids and is widely used for protein folding studies^[140]. Specifically, an engineered cysteine-free "pseudo-wild type" barstar (ψ -b*), P28A/C41A/C83A with only one Met residue at the N-terminus (M1) is useful, as the incorporation of methionine analogs and subsequent coupling reaction is site-specific at the N-terminus. Its N-terminal modification can generally be expected to retain a functional protein structure while introducing novel functions^[141].

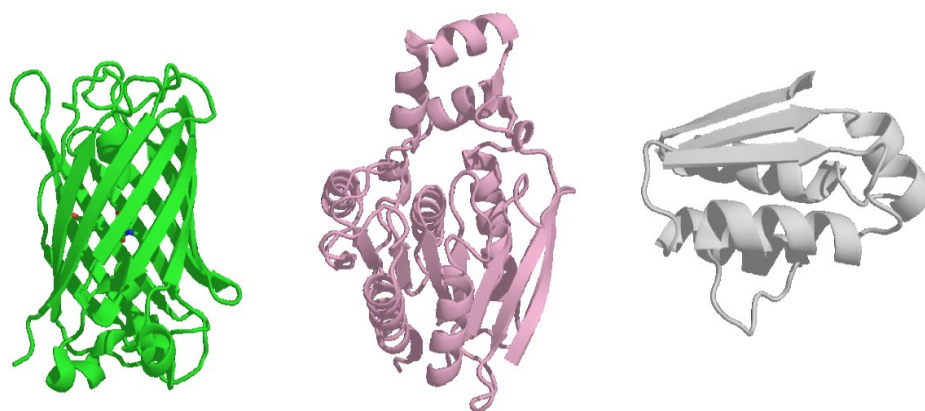


Figure 13: Model proteins used in this study. A: Enhanced green fluorescent protein (EGFP) is an autofluorescent protein often used as reporter (from PDB-ID 2YOG). B: *Thermoanaerobacter thermohydrosulfuricus* lipase (TTL, see reference^[124]) is a model enzyme already probed for acceptance of noncanonical amino acids. C: Barstar is a small protein often employed in folding studies (from PDB-ID 1BTA).

4. Aim of this study

The aim of this study was to employ variants of aminoacyl-tRNA synthetases, particularly *Methanosarcina mazei* pyrrolysyl-tRNA synthetase, to cotranslationally incorporate novel noncanonical amino acids. The amino acids were chosen with a focus on two properties: Of major interest were amino acids with unsaturated side chains, which after incorporation can undergo specific chemical reactions, including olefin metathesis, to decorate proteins after expression and purification. Other target amino acids were chosen for the possibility of biosynthetic production *in situ* in *E. coli*, allowing the establishment of a multistep artificial biosynthetic pathway leading to production of modified proteins from simple chemical media additives.

This was achieved in three steps. First, a suitable and efficient suppression system was adapted and tested for *in vivo* incorporation. Having this system in hand, it was used to incorporate noncanonical amino acids for peptide applications by cooperation partners. Additionally, attempts were made to achieve stop codon suppression *in vitro*. Second, rational variants of the orthogonal suppressor aaRS were generated to incorporate novel noncanonical amino acids into model proteins. Finally, a PylRS library was created and positive and negative selection systems were adapted and tested for selection of variants.

Additionally, SPI methodology was employed to generate modified proteins for specific applications.

5. Results and discussion

5.1 Establishment of a stop codon suppression system

5.1.1 General considerations

To establish an alternative genetic code with translation of the amber stop codon TAG as a noncanonical amino acid, three components have to be present. First, a suppressor tRNA with the CUA anticodon, that is exclusively charged with the desired noncanonical amino acid exclusively. Second, an aminoacylation enzyme that activates only the desired amino acid and is selective for charging the suppressor tRNA. The third component is a target protein gene with a stop codon inserted at a desired position. In the course of this work, several vector systems for stop codon suppression were assessed.

5.1.2 The pTRP-Duet1 system

The vector pTrp-Duet1+*MmPyl*ST-1mutEYFP was received as a gift from Professor T. Carrel^[121], harboring *Methanosarcina mazei* PylRS with an N-terminal hexahistidin tag, the corresponding tRNA^{Pyl} and EYFP with a strep-tag and an internal amber stop codon at position N114. To assess the feasibility for suppression, EYFP was expressed as previously described in aforementioned reference in the presence of 2 mM Bok. The yield from 1 L of cell culture was app. 1.5 mg and the correct mass was found (see Figure 14).

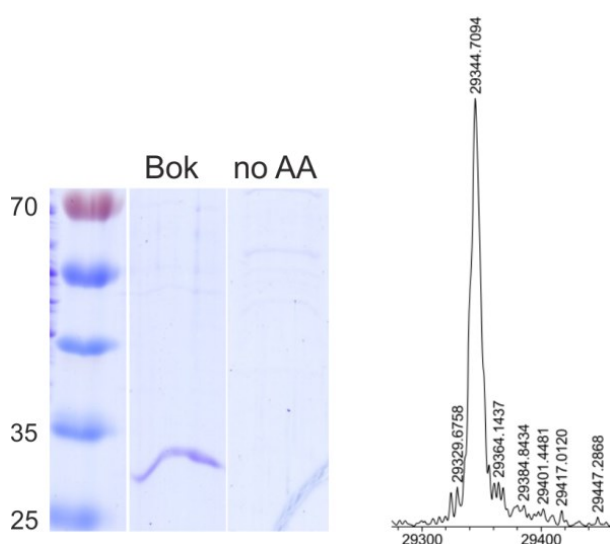


Figure 14: Expression of EYFP(114Bok). Left: Full length protein could be purified only after addition of 2 mM Bok. Right: Incorporation of Bok could be confirmed by mass spectrometry (calculated mass: 29345.5 Da; found mass: 29344.7 Da).

Due to the low expression yield of EYFP(114Bok), it was replaced by the known highly suppressible EGFP(N150TAG) with a C-terminal hexahistidin tag for purification, creating vector pTRP-Duet1+*MmPyl*ST-EGFP(150TAG). The co-detection and co-purification of *MmPyl*RS and EGFP due to

Results and discussion

an identical tag was not deemed problematic at this stage. However, the suppression of the internal stop codon proved to be unreliable with a high batch-to-batch variation. Furthermore, EGFP purified after expression with this vector system showed a high tendency toward precipitation. In fact, even wild-type EGFP, otherwise known to be a very stable protein, showed significant precipitation overnight. It can only be speculated that this happens due to a co-precipitation process with the co-purified *MmPylRS*, although a co-purification could not be confirmed in all cases. Other reactions may play a role. All results are depicted in Figure 15.

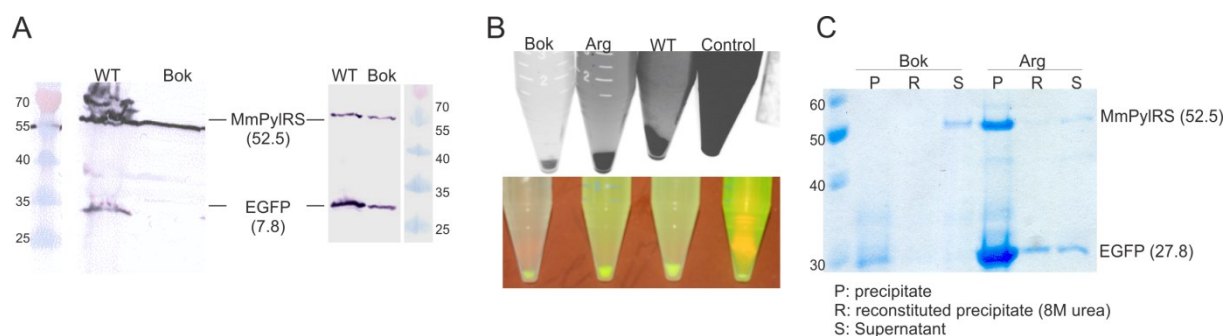


Figure 15: Expression of EGFP. A: Western blot detection of *MmPylRS* and EGFP in two different expression batches. Full-length EGFP(Bok) can only be detected in one batch, illustrating the lack of reliability of the expression system. B: UV fluorescence (top) and normal (bottom) image of EGFP(150Bok), EGFP(150Arg) expressed for control purposes employing the rare arginine codon AGG, and EGFP after hexahistidin tag purification (with co-purification of *MmPylRS*), after storage at 4 °C over night. All samples show significant precipitation. A control sample of pure EGFP after storage at 4 °C for >7d was added to illustrate the stability of the fluorescent protein. C: SDS-PAGE analysis of precipitated EGFPs, with pellet material (P), solubilized pellet material (R) and supernatant (S) applied to the gel. The pellet of EGFP(Bok) shows only a band corresponding to the mass of EGFP, while the pellet of EGFP(150Arg) shows co-precipitation with *MmPylRS*.

The low reliability of suppression and the low yield and stability of the resulting suppressed proteins made a change of the expression system necessary. As it is desirable to apply the suppression system to a variety of target proteins, further vector systems only harbor the orthogonal pair and have to be combined with a second expression plasmid for the target gene.

5.1.3 The pMEc1 μ C system

The pMEc vector series, recently established in the group^[92], is a set of expression vectors employable in both yeast and *E. coli* hosts. Particularly, the vector pMEc1 μ C was expected to show superior performance for several reasons: The p15A low copy number origin of replication reduces the intracellular amount of PylRS which decreases potential toxic effects, especially when combined with the tight, inducible propionate promoter for PylRS expression. Additionally, the URA3 selection marker allow efficient recombination cloning in yeast while the chloramphenicol resistance is compatible with common expression plasmids, e.g. the pQE (Qiagen, Hilden, D) and pET (Novagen/Merck, Darmstadt, D) plasmid series. Finally, the plasmid was already used successfully to express a MetRS/tRNA^{Met} pair from *Sulfolobus acidocaldarius*^[142].

Vector pMEc1 μ C-prp-strep-PylST was constructed by removal of *Sa*MetRS and *Sa*-tRNA^{Met} genes from vector pMEc1 μ C-prp-strep-*Sa*MetRS-SatRNA-Met and replacement via yeast recombination with an *MmPylRS*-*MmtRNA*^{Pyl} cassette generated via PCR from pTrp-Duet1+PylST-EGFP. When used for stop codon suppression in combination with a pQE-based target gene, the suppression efficiency was

Results and discussion

indeed improved over the pTRP system, allowing purification of EGFP(150Bok) with a yield of app. 2.5 mg per liter of cell culture (see Figure 16).

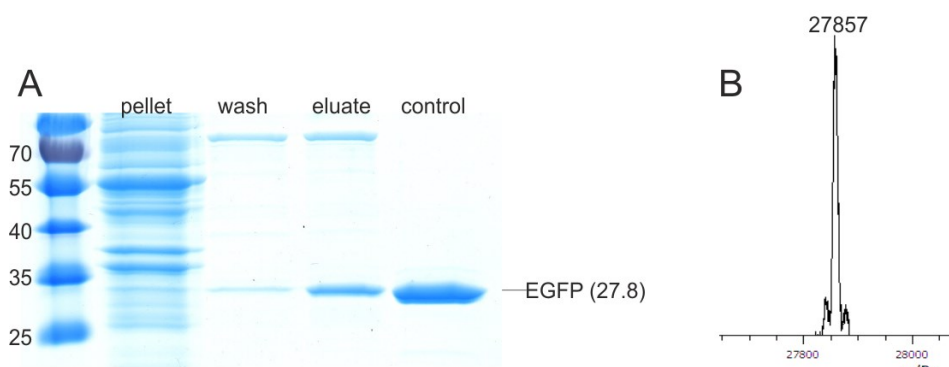


Figure 16: Analysis of EGFP(150Bock). EGFP(150amber) was co-induced with *MmPylRS/MmtRNA^{Pyl}* from plasmid pMEc1 μ C-prp-strep-PylST in the presence of 2 mM Bock. **A:** SDS-PAGE of EGFP(150Bock) purified by Ni-affinity chromatography. The eluate shows a band with the same apparent molecular weight as EGFP control. **B:** Deconvoluted mass spectrum shows a peak at 27857 Da, consistent (± 1 Da) with EGFP(150Bock) (calculated: 27858).

However, the initial success of the suppression system could not be reproduced in all later expression batches. It was concluded that the lack of a transcription termination region led to an accumulation of mRNA, which decreased translation efficiency. To increase vector and mRNA stability, a glnS' transcription termination region was inserted downstream of the tRNA-cassette. The resulting plasmid pMEc1 μ C-prp-strep-PylST-glnS' showed stable amber suppression at low levels (see Figure 17). Suppression was tested with N ϵ -Boc-lysine, N ϵ -Z-lysine, N ϵ -acryloyl-lysine, N ϵ -propargyl-lysine and an unsaturated di-amino acid, of which Bok and propargyl-lysine were accepted as substrates by *MmPylRS*. EGFP(150Bok) was purified and analyzed by mass spectrometry, confirming suppression. While the system works consistently, both the total protein yield and analog acceptance required improvement.

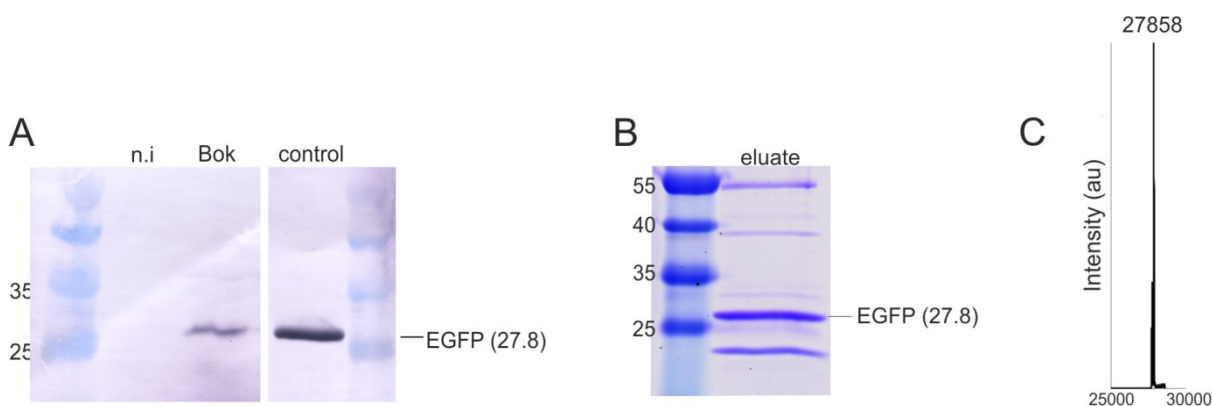


Figure 17: Suppression of EGFP(150TAG) and purification of EGFP(150Bok). EGFP was expressed in the presence of 2 mM Bok in cells carrying plasmid pMEc1 μ C-prp-strep-PylST-glnS'. **A:** Western blot of EGFP suppression with Bok shows a band at the expected position. 1 μ g of wild-type EGFP is used as a blot control. **B:** Affinity purification of EGFP(150Bok). EGFP is the main portion of eluate protein content with several impurities. **C:** Deconvoluted mass spectrum shows a peak at 27858 Da, consistent with EGFP(150Bok).

To further improve suppression efficiency and to allow for flexible combination of enzyme and tRNA, a vector series with a monocistronic transcript PylRS and tRNA^{Pyl} was constructed.

5.1.4 The pJZ system

The construction of the vector pJZ-P_{trp}-*MmPyl*3TS has been previously described^[127]. Briefly, PCR fragments of promoter, tRNA cassette and PylRS were cloned subsequently into plasmid pEVL648 (kindly provided by Volker Döring, CEA, DSV, IG, Genoscope, Evry, France) via restriction-ligation, replacing the lac-promoter with the trp-promoter. The resulting plasmid allows monocistronic transcription of three tRNA copies and synthetase gene. A plasmid with a single tRNA copy (pJZ-P_{trp}-*MmPyl*TS) was constructed the same way^[143].

The Mutation Y384F in the amino acid binding pocket frequently occurs in PylRS variants engineered for aminoacylation of noncanonical amino acids and has also been shown to enhance aminoacylation of Bok and Pyl^[78]. It was inserted in vector pJZ-P_{trp}-*MmPyl*3TS(Y384F) and vector pJZ-P_{trp}-*MmPyl*TS(Y384F) to increase the incorporation of noncanonical amino acid.

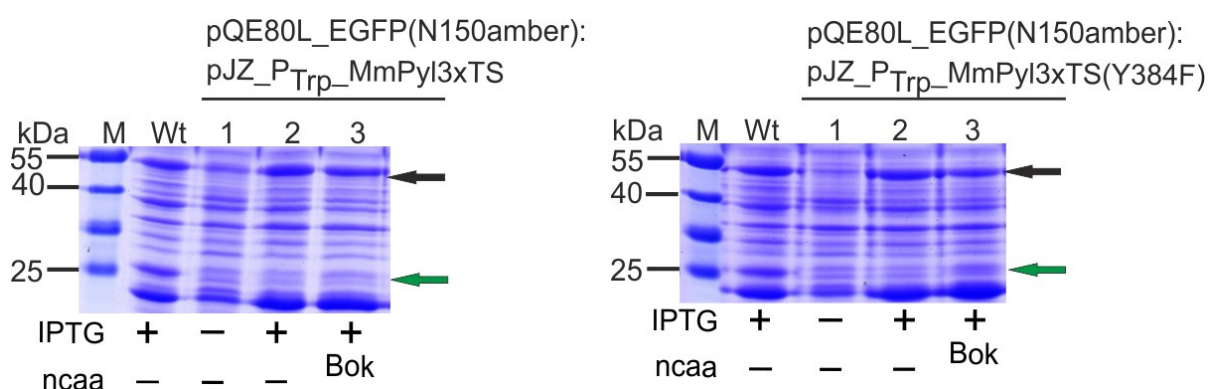


Figure 18: Expression tests with the pJZ vector system (adapted from Al Toma *et. al*^[127]). EGFP(150TAG) was suppressed with *MmPyl*RS or *MmPyl*RS(Y384F). The *MmPyl*RS band at 52 kDa (black arrow) is visible in all induced samples, indicating stable expression. The full-length EGFP at 27.8 kDa (green arrow) is faintly visible only in wild-type controls and samples with Bok added after induction, and more pronounced for suppression with *MmPyl*RS(Y384F).

The pJZ plasmid series shows stable expression of *MmPyl*RS and yields good suppression results, which could be further enhanced by app. 50% by the mutation Y384F (Figure 18). The plasmid variant with a single tRNA copy does not show significantly lower performance and was chosen for most suppression experiments.

5.1.5 Activity screening of *MmPyl*RS

With a working suppression system in hand, *MmPyl*RS and *MmPyl*RS(Y384F) were screened for activation of diverse noncanonical amino acids

As a first test of the feasibility of incorporating the Pyl analogs Ayk and Edo (doubly N-acetylated), these were assayed for toxicity in *E. coli* BL21Gold(DE3) as shown in Figure 19. Cell growth was not significantly impaired by addition of Edo or Ayk relative to the addition of Bok. All tested amino acids can be used for further incorporation experiments.

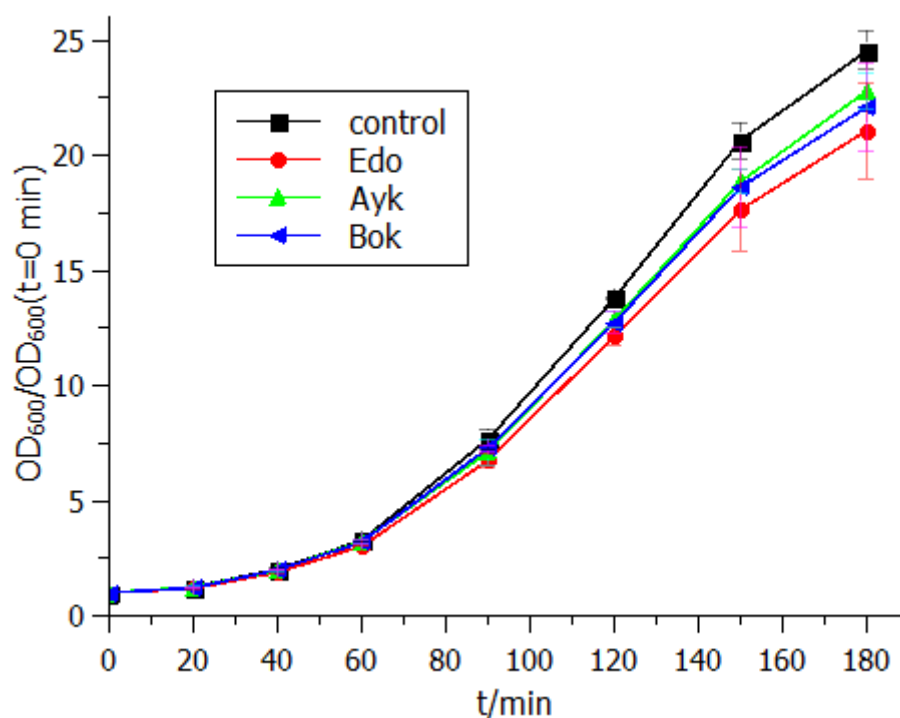


Figure 19: Growth curves of strain BL21Gold(DE3) in LB medium. Control: no aa; Edo/Ayk/Bok: 1 mM of respective amino acid added. Curves were normalized to the initial cell density and measured in duplicate.

Incorporation studies were conducted by suppression of a stop codon in enhanced green fluorescent protein (EGFP). BL21(DE3) cells co-transformed with pQE80L-EGFP(150amber)-H6, expressing a C-terminally his-tagged EGFP with the mutation N150amber and either pMEc1 μ C-*Mm*PylST, pJZ_P_{trp}-*Mm*Pyl3TS or pJZ_P_{trp}-*Mm*Pyl3TS(Y384F) harboring *Mm*-tRNA^{Pyl} and *Mm*PylRS or *Mm*PylRS(Y384F), respectively, were induced in the presence of 1 mM ncaa. Full-length EGFP was visualized via western blot. Incorporation was detected for Bok, Pok and Nbk. Additionally, residual activity was found for incorporation of Pnk, an amino acid that has only recently been identified as a weak substrate for PylRS^[112] (see Figure 20).

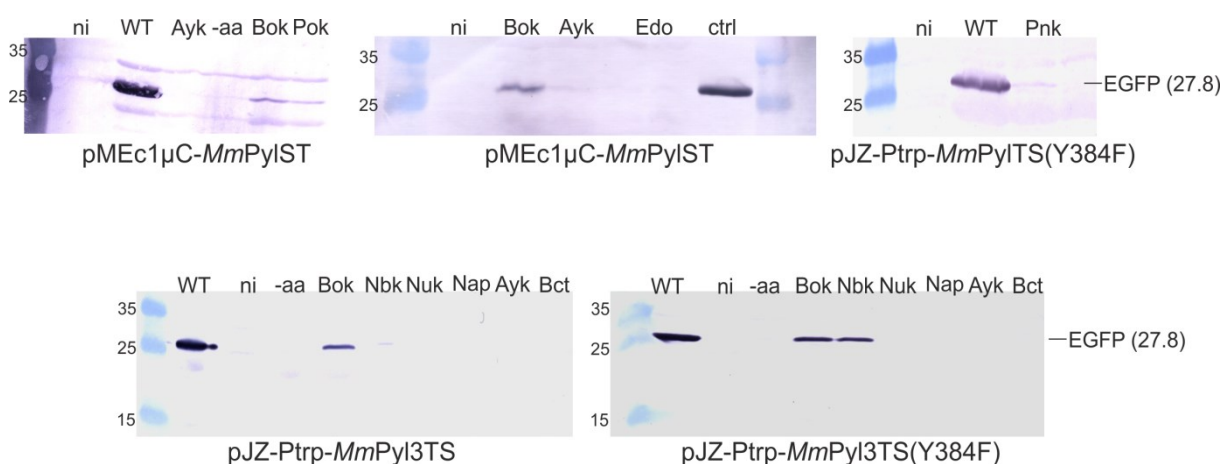


Figure 20: Incorporation tests with wild-type *Mm*PylRS (partially adapted from Al Toma *et al.* ^[127]). Two suppression systems were tested with the pJZ plasmid system showing superior performance. Additionally, the efficiency of the Y384F variant was assessed. Incorporation was found for Bok, Pok, Nbk (only residual without Y384F) and Pnk (residual).

Results and discussion

To confirm the incorporation, proteins were purified and analyzed with mass spectrometry, the results are shown in Figure 21.

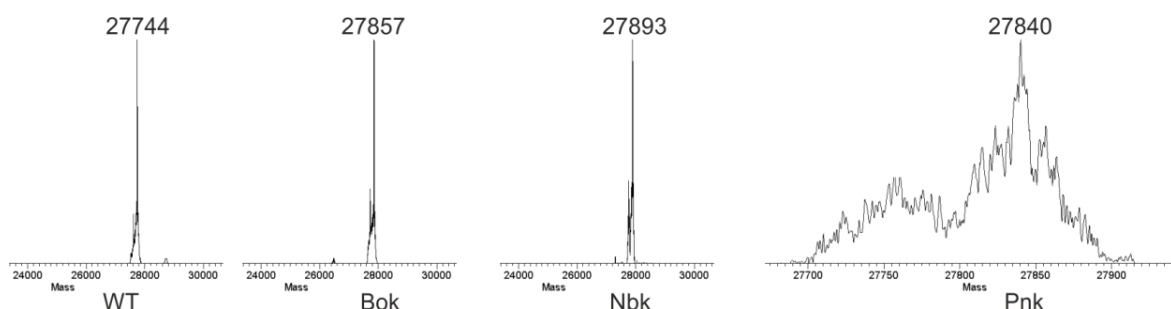


Figure 21: Deconvoluted mass spectra of EGFP(150TAG) suppressed with different amino acids. Calculated masses are 27744 Da (EGFP(WT), N150); 27858 Da (EGFP(Bok)); 27895 Da (EGFP(Nbk)); 27841 (EGFP(Pnk)). The high noise level of EGFP(Pnk) is due to impurities and low amount of protein. No mass spectrum was measured for EGFP(Pok).

With a functional suppression system in hand, rational and evolutionary alterations of the amino acid specificity can be introduced to incorporate a wider range of noncanonical amino acids.

5.1.6 Additional experiments

To optimize the yield of proteins with incorporated noncanonical amino acids, the suppression was tested in the commercially available medium EnPresso B (Biosilta, St. Ives, UK), which greatly reduces hands-on time during expressions due to flexible inoculation and induction times.

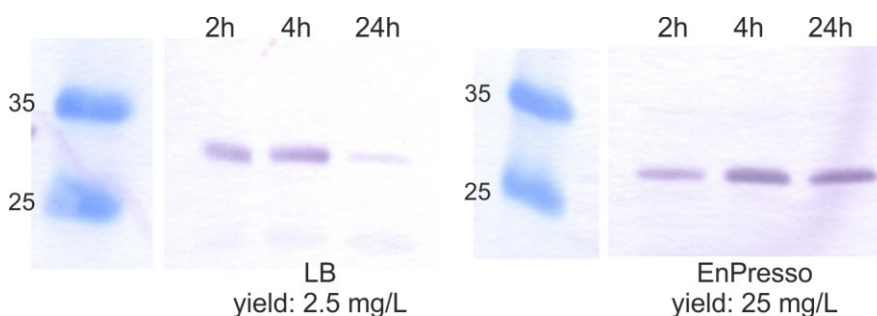


Figure 22: Detection of full-length EGFP (27 kDa) suppressed with Bok from cell lysates. Cells were harvested 2, 4 and 24 hours after induction with 1 mM IPTG and 2 mM Bok. Yields were determined after purification.

Indeed, EnPresso could be shown to significantly increase the yield of full-length protein, allowing for more efficient production of suppressed proteins (see Figure 22). It also proved to be less time-dependent. However, due to the elevated cost, it may be advisable to employ the medium only for large-scale production or with valuable noncanonical amino acids.

While the established orthogonal pairs have been used to incorporate noncanonical amino acids of astonishing diversity, alternative aaRSs may offer novel specificities and routes for the incorporation of target amino acids which cannot currently be incorporated. Two alternative PylRS from the

Results and discussion

archaeal organisms *Thermincola potens* (TpPylRS) and *Methanohalobium evestigatum* (MePylRS) were tested for the incorporation of Bok into model proteins in *E. coli*, depicted in Figure 23.

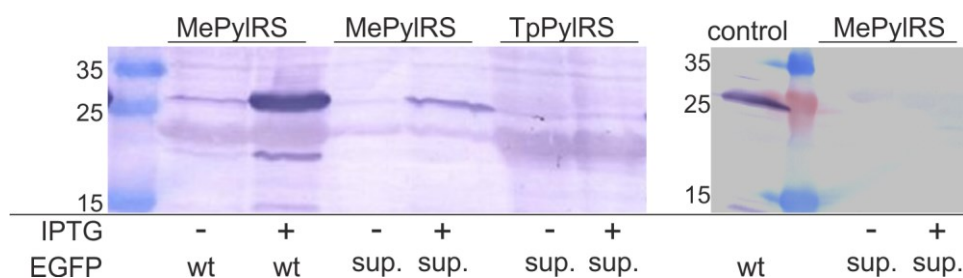


Figure 23: Test suppressions with novel orthogonal systems MePylRS/Me-tRNA^{Pyl} and TpPylRS/Tp-tRNA^{Pyl}. Full-length EGFP(Bok) was detected with MePylRS (left blot), but the finding could not be reproduced (right blot). No activity was found for TpPylRS.

While TpPylRS did not show stop codon suppression in the presence of Bok, initial results with MePylRS looked promising. These results, however, could not be reproduced in a second experiment, leaving the suitability of a MePylRS/Me-tRNA^{Pyl} suppression system questionable. Reasons for inactivity of the enzymes may lie in the adaption to vastly different environments: While all PylRS successfully adapted in *E. coli* so far originate from mesophilic methanogenic organisms, *T. potens* is thermophilic and *M. evestigatum* is extremely halophilic. Adaptions to these extreme environments often cause enzymes to be inactive under physiological conditions of *E. coli* [144,145]. Another possible reason may be a higher selectivity of both PylRSs, discriminating against Bok. This is very unlikely, however, as despite the differences in structure, all PylRS successfully used in *E. coli* so far have shown activity towards Bok [57].

To overcome limitations of cellular protein synthesis due to amino acid transport and toxicity issues, experiments with *in vitro* protein expression were carried out, using protocols by Kim *et al.* [146,147] with the addition of purified MmPylRS and *in vitro* transcribed Mm-tRNA^{Pyl}. The results are depicted in Figure 24.

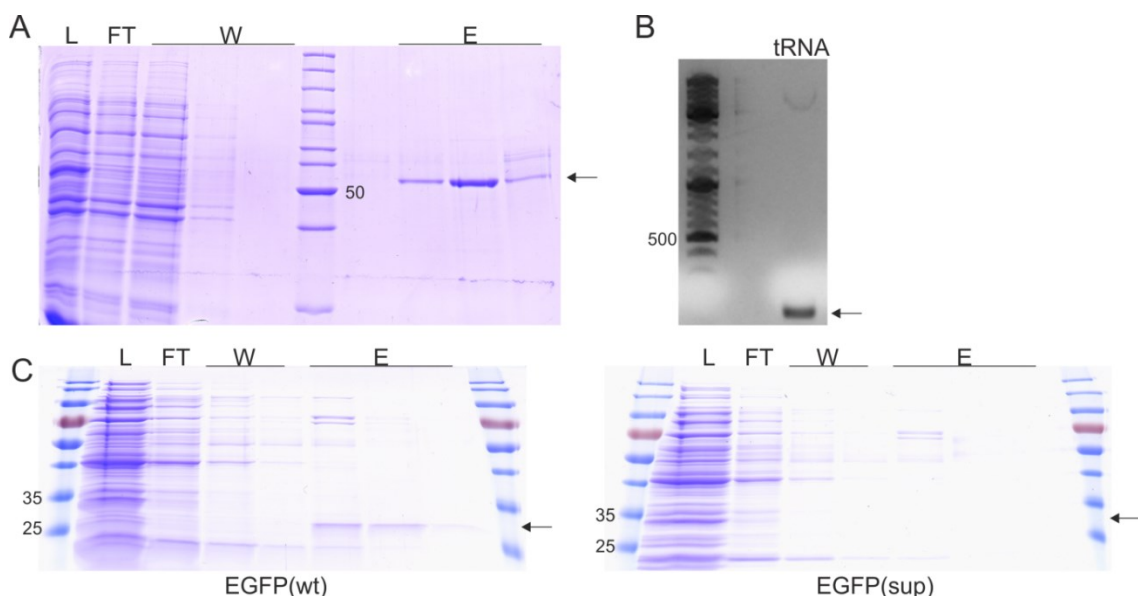


Figure 24: In vitro suppression: A: Purification of strep-tagged MmPylRS. B: Analysis of *in vitro* transcribed Mm-tRNA^{Pyl}. C: Purification of *in vitro* expressed EGFP(wt) (left) and suppressed EGFP(150Bok) (right). Only EGFP(wt) was found. Lane labels: Lysate (L), flow through (FT), wash fractions (W), eluate fractions (E).

Results and discussion

While the purification of suppression system components was successful, no suppression could be observed. Further experiments concerning *in vitro* suppression were performed by cooperation partners (see Figure 25).

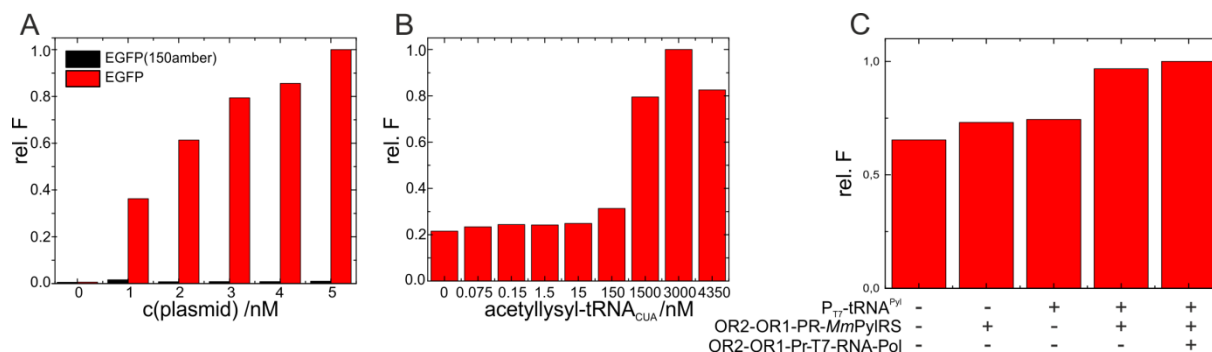


Figure 25: *In vitro* suppression with an advanced cell-free expression system (single experiment, data generated by E. Worst^[148]). A: Normalized EGFP-fluorescence after 16 h of cell-free expression. Expression of EGFP(WT) leads to increased fluorescence dependent on the plasmid concentration, while expression batches for EGFP(150amber) do not show significantly elevated fluorescence levels. This indicates a tight system with minimal background suppression. B: Normalized EGFP-fluorescence in cell-free expression reactions of EGFP(150amber) with pre-aminoacylated acetyllysine-tRNA_{CUA}. Nearly quantitative production of EGFP(150acetyllysine) is observed at 1.5 μ M of tRNA, with a maximum at 3 μ M tRNA. C: Cell-free suppression of EGFP(150amber) in the presence of 2 mM Bok. EGFP(150amber), *Mm*PylRS and T7 RNA polymerase are expressed *in vitro* using the *E. coli* transcription/translation machinery and the optimized promoter OR2-OR1-PR. tRNA^{Pyl} is transcribed from linear DNA under the control of a T7 promoter, employing the highly efficient T7 transcription. An increase in fluorescence of 36% compared to a negative without templates for tRNA^{Pyl}, *Mm*PylRS and T7 RNA polymerase could be observed. Additional controls omitting one or two components showed moderate increases of EGFP fluorescence.

An established cell-free expression system^[149,150] showed no significant background suppression and good utilization of pre-aminoacylated tRNA. Experiments with *in vitro* produced *Mm*PylRS and tRNA^{Pyl} showed an increase in fluorescence of 36% over the negative control. While the suppression efficiency is lower than achieved *in vivo*, these results indicate a good possibility for efficient *in vitro* suppression of target proteins, albeit for to fully prove the feasibility, a repetition of the experiments and optimization is needed. The efficient suppression with commercial N_ε-acetyllysyl-tRNA_{CUA} shows that the aminoacylation reaction is likely to be the limiting factor. Low activity of *Mm*PylRS *in vitro*, especially for prolonged durations has been observed previously and can possibly be alleviated by optimizing environmental conditions and tuning *Mm*PylRS expression *in situ*.

5.2 Rational mutagenesis of *MmPyIRS*

5.2.1 Rationale for mutagenesis

With an increasing number of published synthetases and the regularly discovered promiscuity of variants^[72,78] it has become a feasible method to screen existing synthetases for activities towards novel amino acids. In this study, two PyIRS variants were screened for novel activities: A PyIRS incorporating acryloyllysine^[151], one of the shortest-chain amino acids incorporated so far, was tested with short-chain non-lysine amino acids. PyIRS_AF (*MmPyIRS*(Y306A/Y384F), a promiscuous variant with an enlarged binding pocket^[78,121,152] was screened with lysine-derivatives with bulky head groups.

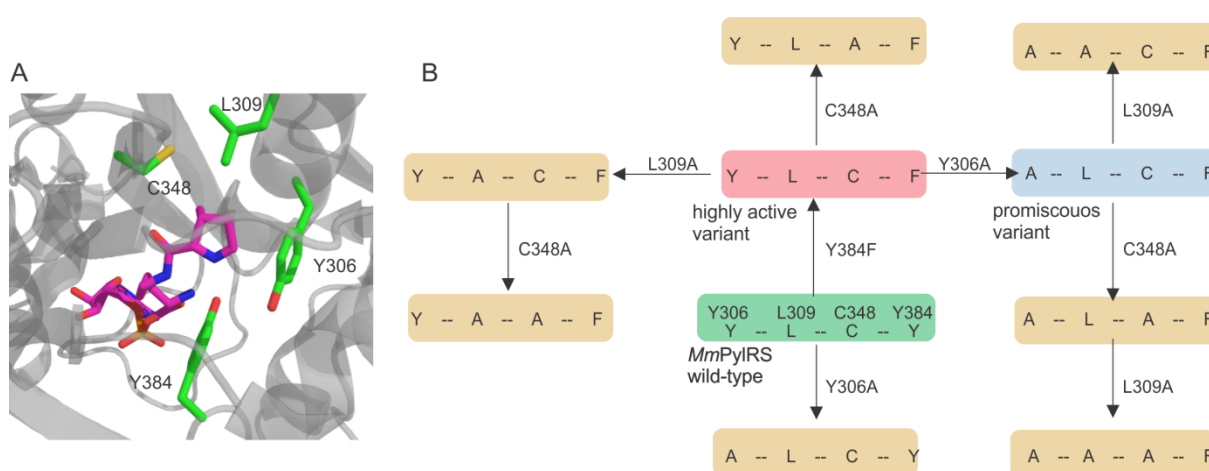


Figure 26: Rational mutagenesis of the amino acid binding pocket of *MmPyIRS*. A: The amino acid binding pocket of *MmPyIRS*. Four amino acids lining the pocket have been highlighted. These residues limit the size of the binding pocket and are frequently altered in mutants accepting larger amino acids. B: Mutagenesis scheme for PyIRS variants with rationally enlarged binding pockets. Starting from the highly active variant *MmPyIRS*(Y384F), the amino acids Y306, L309 and C348 have been exchanged by alanine in all combinations. Additionally, wild-type *MmPyIRS* and the Y306A variant of wt-PyIRS have been used.

Inspired by the versatility of the PyIRS_AF variant, additional rational variants were tested. The residues Y306, L309 and C348 were identified as size limitators of the amino acid binding pocket from the crystal structure and are indeed frequently mutated in variants accepting bulky amino acids. Rational variants with enlarged binding pockets were generated by mutating the limiting amino acids to alanines in all possible combinations as schematically shown in Figure 26B.

5.2.2 Screening of AykRS

AykRS, known from literature^[151], was tested for incorporation of the linear noncanonical amino acids Edo, Spc, Sbc, Sac, Ops, and Npa (see Figure 27).

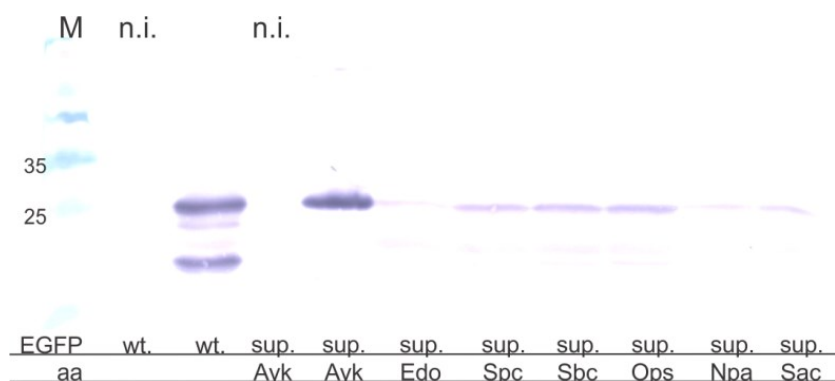


Figure 27: Test incorporation with AykRS and diverse amino acids. Good suppression was found only for addition of Ayk, bands for other amino acids are results of artifacts. wt: wild-type EGFP; sup: EGFP(150TAG); aa: 2 mM of denoted noncanonical amino acid added.

Although initial western blot analysis showed faint bands for all amino acids tested, indicating residual activity, no protein could be purified from the expressions. The bands were identified as false-positives, possibly due to background suppression, and it was concluded that no novel amino acid was incorporated by AykRS. Intriguingly, although green protein could be purified from expression batches containing Ayk, no mass could be found via MS analysis. The reason for this may be the high reactivity of the acryloyl group, leading to protein degradation before the MS measurement.

5.2.3 MmPylRS variants with enlarged amino acid binding pockets

With the promiscuous rational variant PylRS_AF in hand^[78], the incorporation of Cok into the model protein EGFP was attempted for further modification via ring-strain induced alkyne-azide cycloaddition as previously reported^[115]. Full-length EGFP was purified from a 1 L expression batch and analyzed with MS, as shown in Figure 28. Intriguingly, the expected mass (27909 Da) was not found. Instead, a lower mass (27758 Da) was measured, indicating the incorporation of lysine, glutamine or glutamic acid in the stop codon position (the masses of Lys, Gln and Glu are too similar to be distinguishable in full-protein MS). The reasons for this could be the direct incorporation of either amino acid by the PylRS_AF/tRNA^{Pyl} orthogonal pair or native *E. coli* suppression pathways, or the decomposition of the incorporated Cok to Lys. The latter possibility was deemed more probable due to the instability of the carbamate bond. However, attempts to modify the purified protein with azides were unsuccessful^[153], indicating the absence of the cyclooctyne head group.

Results and discussion

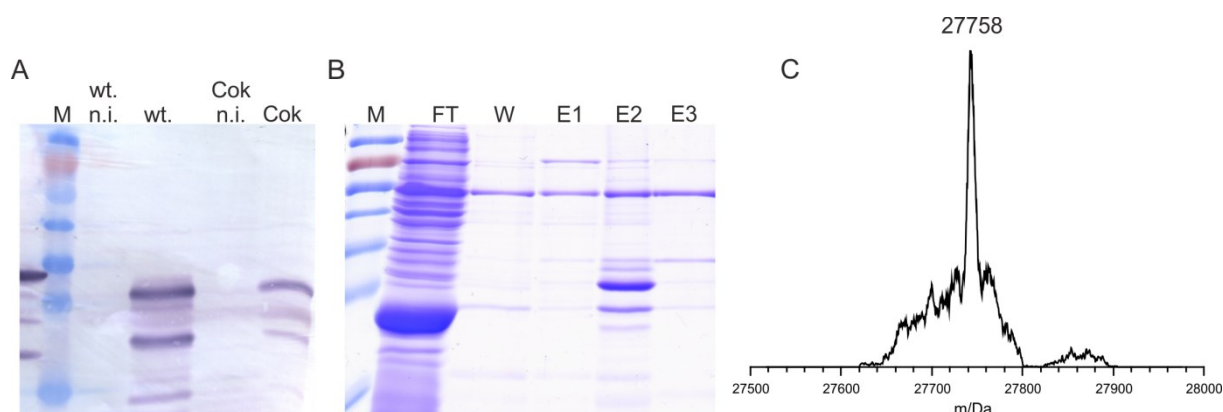


Figure 28: Incorporation of Cok with PylRS_AF. A: Western blot detection of full-length EGFP in the presence of Cok. B: Purification of EGFP(150Cok). Flow-through (FT), wash (W) and eluate (E1-E3) fractions have been applied to the gel. EGFP(150Cok) is found in fraction E2. C: Mass analysis of EGFP(Cok) (expected mass 27909 Da). The measured protein mass matches EGFP with incorporated Glu, Gln or Lys (expected masses 27759.5 Da, 27758.5 Da, or 277758.5 Da, respectively).

Novel and known rational variants were tested for incorporation of Bct. However, while most variants retained activity towards Bok, Bct could not be incorporated by any variant (Figure 29).

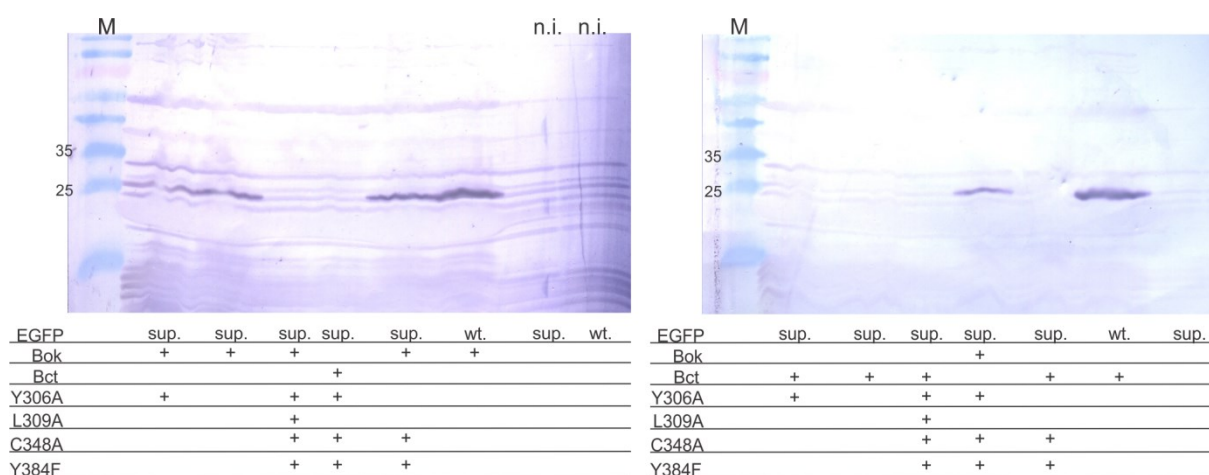


Figure 29: Screening of PylRS variants for incorporation of Bok and Bct by western blot analysis of raw cell lysate. While most variants accept Bok (only *MmPylRS*(Y306A/L309A/C348A/Y384F) is not active), Bct is not incorporated by any variant.

Hnk and Pnk are highly stable lysine-based noncanonical amino acids with terminal double bonds that can be addressed via bioorthogonal reactions like protein metathesis or thiol-ene coupling^[112]. Pnk has recently been identified as a weak substrate for *MmPylRS*^[112], confirming our unpublished findings. Both amino acids were tested for incorporation by novel rational *MmPylRS* variants (Figure 30).

Results and discussion

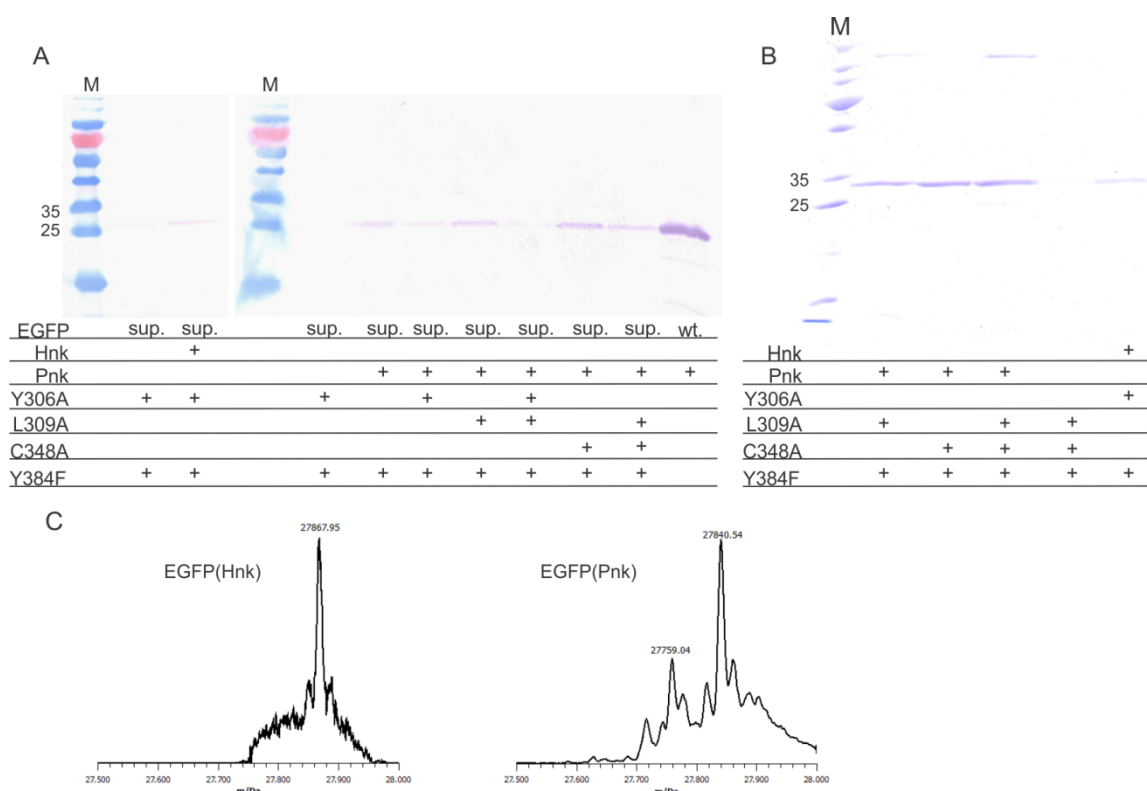


Figure 30: Screening of PyIRS variants for incorporation of Hnk and Pnk. A: Western blot detection of full-length EGFP after suppression with Hnk or Pnk. Hnk is a substrate for PyIRS_AF (*MmPyIRS*(Y306A/Y384F)), Pnk is incorporated by several variants with *MmPyIRS*(C348A/Y384F) seemingly being the most efficient. B: Purification of EGFP(Hnk) and EGFP(Pnk) incorporated with different synthetases. C: Mass analysis of EGFP(Hnk) (expected mass: 27868.5 Da; found mass: 27869 Da) and EGFP(Pnk) (expected mass: 27840.5 Da; found mass: 27840.5 Da). An additional Peak at 27759 Da may correspond to EGFP with incorporated Glu, Gln or Lys (expected masses 27759.5 Da, 27758.5 Da, or 27758.5 Da, respectively).

Hnk was identified as a novel substrate for *MmPyIRS*(Y306A/Y384F) while Pnk is accepted and incorporated into EGFP by most variants, *MmPyIRS*(C348A/Y384F) being the most efficient, allowing purification and MS analysis of Pnk containing proteins for the first time. However, besides a main mass peak corresponding to EGFP(Pnk), an additional mass peak corresponding to EGFP with Lys, Gln or Glu was found. The amide bond of the amino acids is very stable, making decomposition to Lys unlikely.

For chemical modification studies, Pnk was incorporated into TTL(D221amber), as shown in Figure 31.

Results and discussion

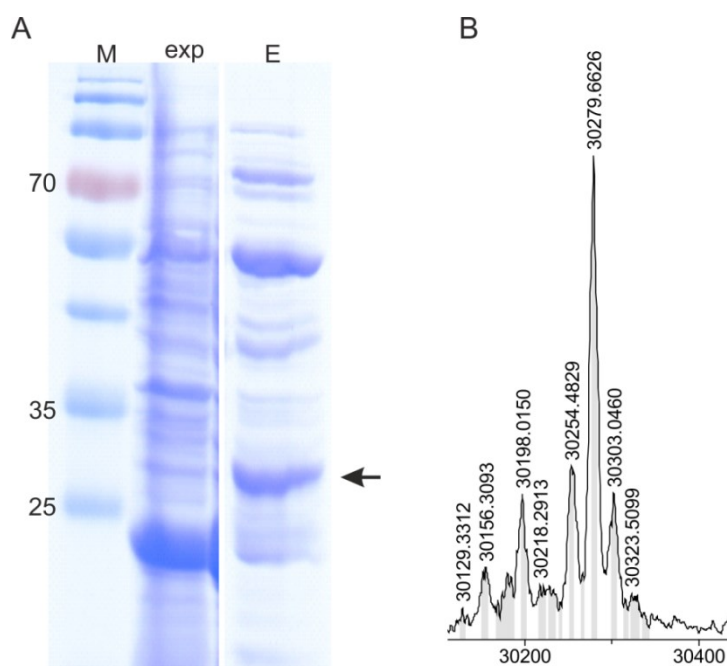


Figure 31: Incorporation of Pnk in TTL(D221TAG). A: SDS-PAGE analysis of TTL expression shows good production of TTL(150Pnk). The eluate fraction contains significant impurities, which do not interfere with subsequent labeling processes. B: Confirmation of incorporation by MS (expected: 30280 Da; found: 30279.67 Da).

While the reactivity of Hnk and Pnk can be concluded from similar amino acids^[112,120], the reactions with sulfides in a radical thiol-ene reaction or with alkenes in an olefin metathesis have yet to be shown.

Both amino acids allow posttranslational chemical decoration of proteins far away from the protein surface, as a long-chain tether is provided. This may increase the acceptance and reduce the structural impact of such modifications. In contrast to similar amino acids, both Hnk and Pnk exclusively comprise stable bonds besides the double bond, allowing applications under a broad spectrum of conditions.

5.3 Selection of *Mm*PyIRS variants

5.3.1 Design and construction of a *Mm*PyIRS variant library

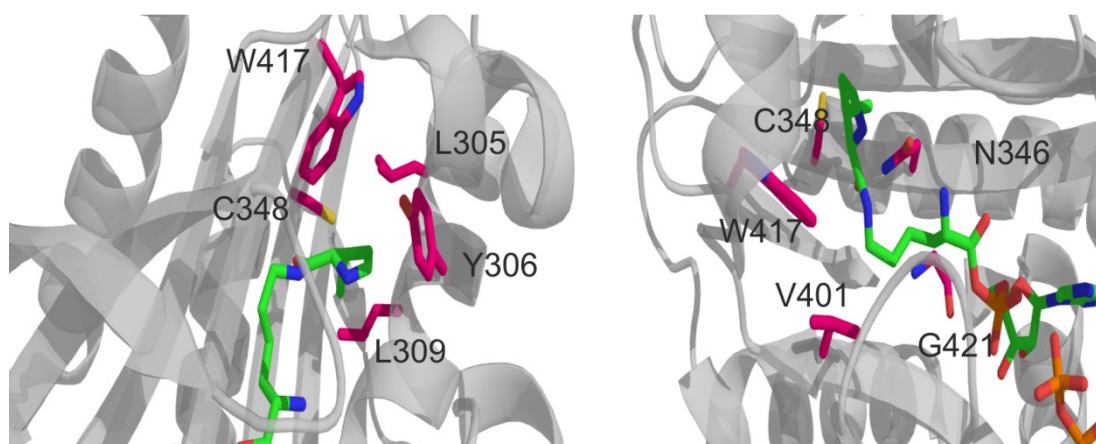


Figure 32: Positions of randomizations of libraries B (left) and S (right).

The generation of a gene library requires finding a compromise between the library variability (as a higher number of different library members increase the chance of finding potential hits) and the practical limitations of screening or selection procedures.

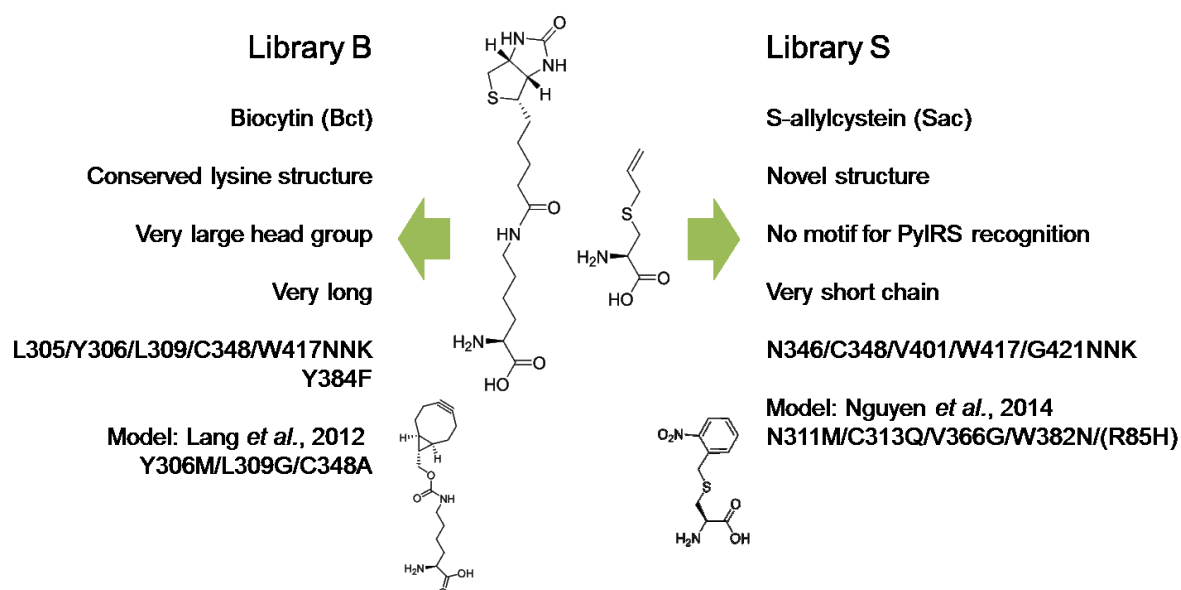


Figure 33: Rationale for library design. Due to structural differences between Bct and Sac, the corresponding libraries have to be designed differently. For Bct screening, a PyIRS library with randomizations of bulky side chains and the fixed Y384F mutation was devised, containing candidates with enlarged binding pockets to accommodate the bulky Bct. This library has successfully been used for incorporation of a large cyclooctyne-containing amino acid. For Sac, on the other hand, the randomizations concentrate on the lining of the front to middle part of the amino acid binding pocket and include a randomization of the gate keeper position N346, which may otherwise fix substrate specificity as lysine derivatives. This library (equivalent positions in *Mb*PyIRS) has been successfully screened for a PyIRS activating a photocaged cysteine, which, while largely different in steric and electronic properties, nevertheless shares substructures with Sac.

Results and discussion

For libraries to be transformed into *E. coli*, the limiting factor is the transformation efficiency of the bacteria, which typically does not exceed 10^9 transformed cells per μg of plasmid DNA. This transformation rate allows screening of libraries with up to 10^8 members with 95% chance of sampling all different members under realistic circumstances. A five-position NNK randomization library contains $3.4 \cdot 10^7$ unique members and thus is the maximum number of randomizations usable for selection^[154,155].

The choice of residues to randomize was made according to the structure of the amino acid desired to be activated and taking into account the known features of the amino acid binding pocket (Figure 32). Selection experiments were planned with biocytin and S-allylcysteine using previously described variants for structurally related ncaas as guidance (Figure 33). For Bct, the residues L305, Y306, L309, C348 and W417 were chosen, along with a fixed mutation Y384F (library B). These residues are crucial determinants of size of the amino acid binding pocket, evidenced by the crystal structure. A PylRS for an amino acid comprised of a lysine with a very bulky head group, similar to Bct, has been recently selected from a library with the same randomizations^[156]. Sac, in contrast, shows very little structural similarity with Pyl and Pyl analogs. Therefore, a library was conceived, which includes a randomization of the gate keeper residue N346, which normally restricts the substrate range to N-carbonyl modified lysine residues. The residues N346, C348, V401, W417 and G421 line the bottom of the amino acid binding pocket and may tune the recognition of short-chain substrates (library S). They have recently been randomized for the selection of PylRSs activating photo-caged cysteines^[75].

Generation of the libraries was carried out by iterative PCR cycles with degenerate primers by first amplifying the upper and lower part of the gene separately, using one degenerate and one flanking primer each, followed by another PCR reaction to join the overlapping fragments. The full-length gene with randomized positions was used as template for consecutive rounds of randomization. Closely spaced randomizations were introduced on a single primer, L305/Y306/L309 for library B, N346/C348 and W417/G421 for library S. The PCR scheme is outlined in Figure 34, the construction of library S is depicted in Figure 35. This way, both libraries were constructed in three rounds of three PCR cycles. The success of each step was confirmed by DNA sequencing.

Results and discussion

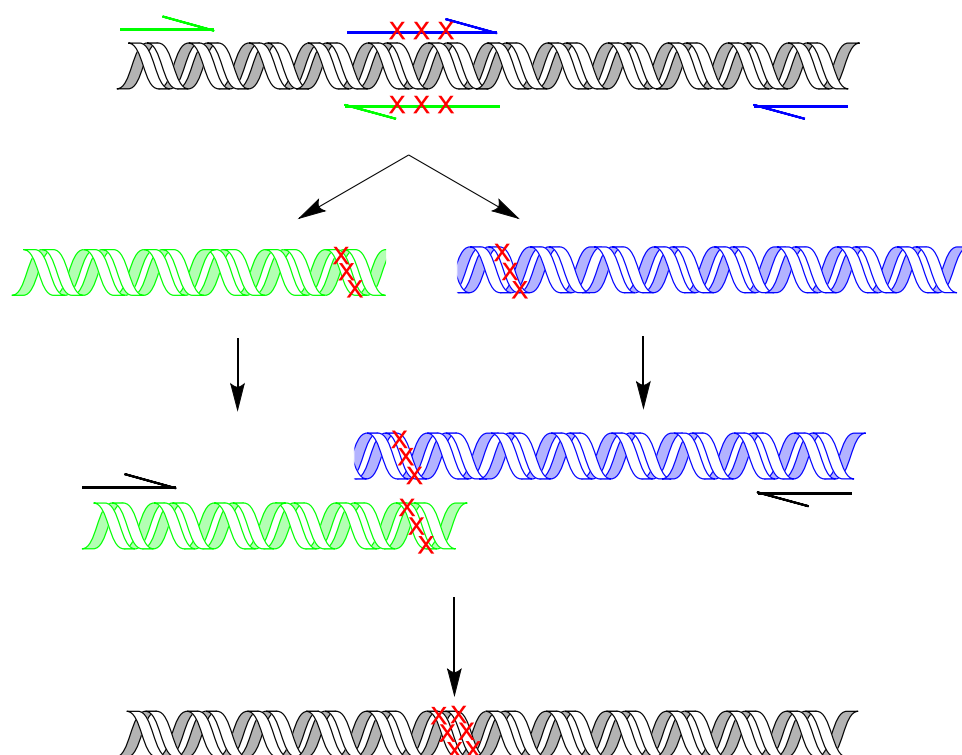


Figure 34: PCR scheme for library generation. Two PCRs are carried out using one flanking and one randomization primer, producing an upstream (green) and downstream (blue) fragment. The fragments have homologues regions due to the use of complementary randomization primers, and can be joined via overlap extension PCR with flanking primers. The resulting full-length gene contains the randomization and can be used as PCR template to generate upstream/downstream fragments with more randomizations. When all randomizations have been introduced, the final full-length product is cloned into an expression plasmid.

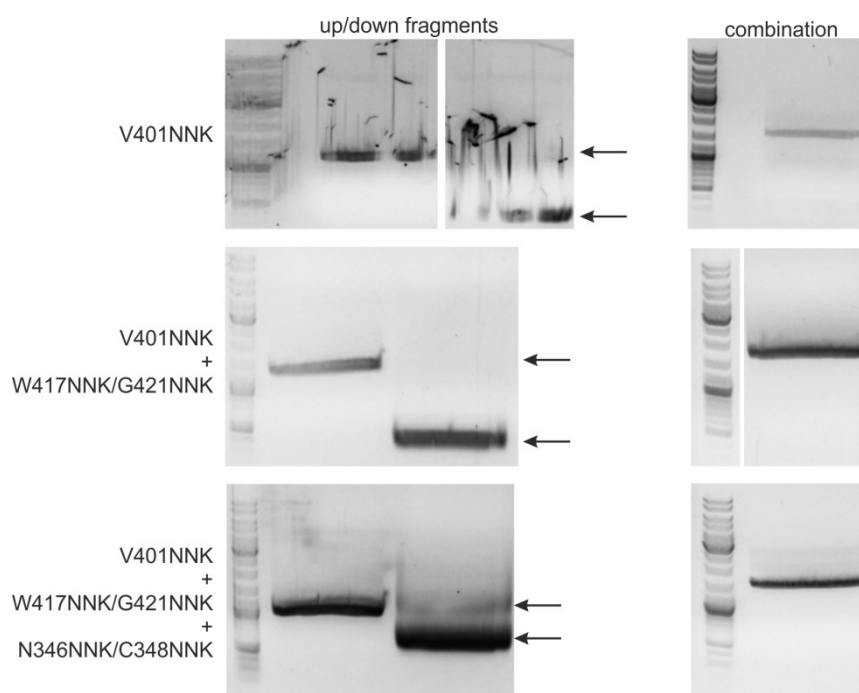


Figure 35: Construction of Library S. Shown are iterative PCR cycles to generate randomized upstream and downstream fragments containing the randomizations and overlapping sequences (left row) and the formation of full-length randomized genes via overlap extension PCR of the fragments (right row). Five randomized positions were introduced in three cycles using the product of the latest combination PCR as template.

Results and discussion

The final PCR products, containing all desired randomizations (see Figure 36), were cloned into the pSEVA derivative pPAB18_glnS' via standard restriction-ligation procedure. To ensure the conservation of the library during the transformation for plasmid propagation, commercial electrocompetent NEB5 α cells were used and directly after transformation grown in liquid media for plasmid preparation. The quality of the libraries was assessed by library sequencing as well as peer group sequencing of plasmids derived from individual transformants.

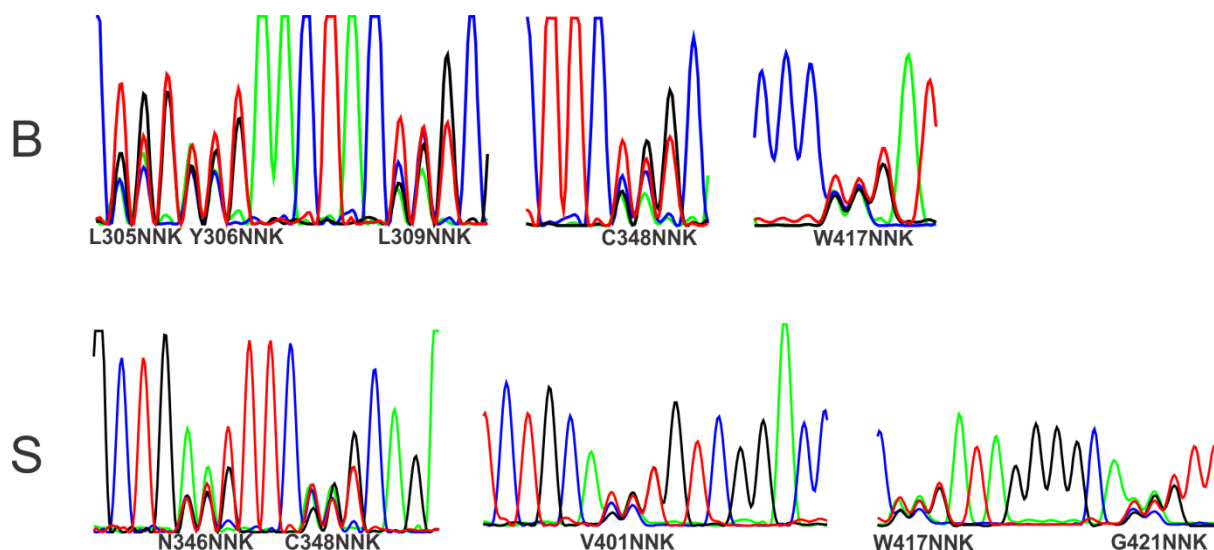


Figure 36: Sequencing of libraries B and S. A, C, G and T are represented as green, blue, black and red lines, respectively. All expected randomizations are found.

All desired positions show NNK randomization in the sequencing reactions and the peer group sequencing results (exemplary shown for library S in Table 3) do not indicate strong sequence bias. It should be noted that the peer-group sequencing reactions yielded a large number of overlay sequences, indicating a transformation of two or more different library members in a single cell. Both libraries were identified as suitable for selection experiments.

Table 3: Peer-group sequencing of six clones from a transformation of library B. Some undefined positions were expected due to transformation of multiple library members in the same cell. All sequenced clones show different genotypes and no bias towards the wild-type amino acids.

WT	CTT105 (L)	TAC106 (Y)	CTG309 (L)	TGC 348 (C)	TGG 417 (W)
1	AAG (K)	TAT (Y)	TCG (S)	CTT (L)	GAG (E)
2	MTK (M/I/L)	TRS (Y/C)	MWK (K/N/M/I/Q/H/L)	TSS (S/C/W)	AWG (K/M)
3	GTG (V)	GCT (A)	ACG (T)	TGG (W)	CGT (R)
4	CCT (P)	ATT (I)	CCG (P)	GAG (E)	AAG (K)
5	GWK (D/E/V)	TWG (F/W)	YMG (Q/S/P)	AGC (S)	TTG (L)
6	CTG (L)	GGT (G)	GCG (A)	TTT (F)	TGG (W)

5.3.2 Conception and setup of an iterative positive-negative selections system

For iterative selection of PylRS variants, established selection markers were used. Positive selection was performed with chloramphenicol acetyltransferase (cat) gene with one or two stop codons. Cells suppressing the stop codons survive on LB medium with added chloramphenicol (cam). A vector for cat expression under its native promoter was obtained in-group and adapted for positive selection. Addition of an expression cassette for tRNA^{Pyl} yielded vector pPNB2_cat_pylT, which was subsequently subjected to site-directed mutagenesis for the introduction of stop codons. Several combinations of stop codons inserted into cat were tested for suppressibility and growth behavior in combination with the orthogonal pair *MmPylRS*(Y384F)/tRNA^{Pyl} and Bok as a model amino acid. The most promising results were found for cat(Q98TAG/D181TAG) and cat(D112TAG), positions known from literature^[56,60,157].

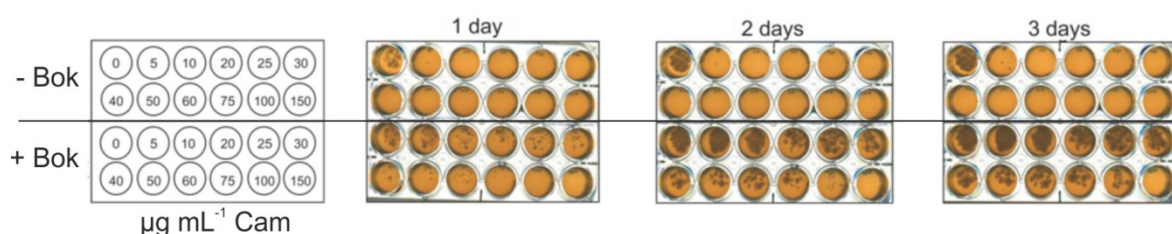


Figure 37: Positive selection system test with cat(98/181TAG) and *MmPylRS*. Cells were spotted on media containing increasing concentrations of chloramphenicol. In the presence of Bok, cells grew at up to 150 µg mL⁻¹ cam within three days at 30 °C. No growth was observed in the absence of Bok.

Cat(98/181TAG) does not allow bacterial growth on cam concentrations as low as 5 µg/ml without the addition of Bok for stop codon suppression. In the presence of Bok, cell growth is retained up to 150 µg/ml cam after three days. In contrast, cat(112TAG) shows significantly lower stringency and allows cell growth without addition of Bok, but at a slower rate compared to instances with suppression. Both genes have been successfully used for selection of PylRS variants (Figure 37). A test selection with cat(98/181TAG) and library B on selective medium with 2 mM Bok yielded a single transformant, which was identified as carrying the wild-type *MmPylRS* sequence (Figure 38). While this was the expected result, as *MmPylRS* is known to incorporate Bok, a higher number of growing transformants would be favorable to avoid arbitrary elimination of desired variants. For this reason, the cat(98/181TAG) was deemed too strict for positive selection.

Results and discussion

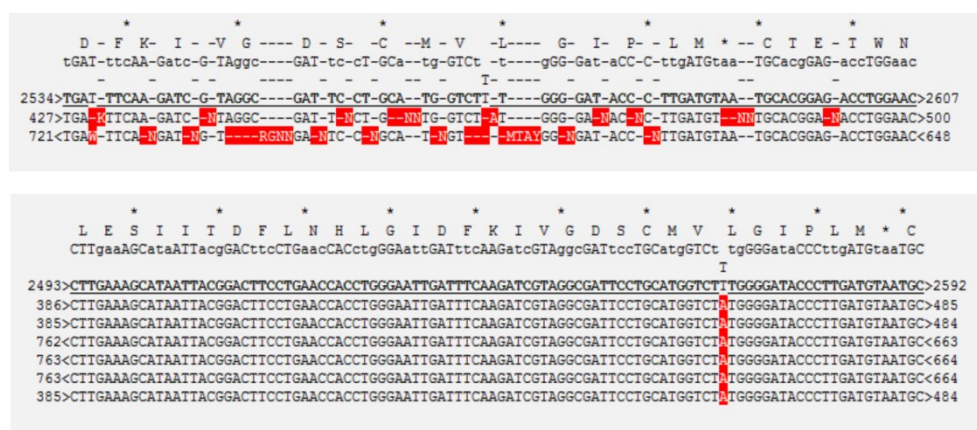


Figure 38: Test positive selection with cat(98/181TAG) from library B for activation of Bok. Top: After 3 d, a single colony was obtained, which, after plasmid isolation, showed an overlay of signal for bidirectional sequencing. Bottom: After removal of the positive selector plasmid and retransformation of the obtained PyIRS plasmid, all six tested colonies showed wild-type sequence with the retained fixed mutation Y384F (TTT-->TAT), as expected.

As more relaxed selection conditions may be favorable for selecting novel activities which may initially have very low efficiencies, cat(112TAG) was chosen as positive selection marker.

For negative selection, the established gene barnase(Q2TAG/D44TAG) was used^[157,158], encoding a toxic nuclease that inhibits cells in minimal intracellular concentrations. A plasmid for negative selection with barnase with a different orthogonal pair was obtained in-group and adapted for selection of PyIRS by exchanging the orthogonal suppressor tRNA^{Tyr} with tRNA^{Pyl}. The resulting monocistronic setup with barnase(2/44TAG) and tRNA^{Pyl} already showed good performance on the pJZ vector series. Tests with Bok revealed that this setup did not stop *E. coli* growth under suppression conditions, but significantly impaired it (Figure 39).

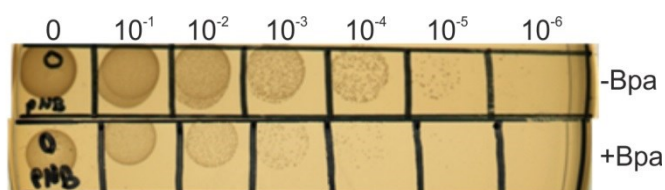


Figure 39: Negative selection system test with barnase and the BpaRS system (data generated by M. Heuchel^[159]). Dilutions of an overnight culture were spotted on medium with or without the cognate noncanonical amino acid benzoylphenylalanine (Bpa). In the presence of Bpa, cell growth is impaired by a factor of 10² to 10³.

The tested system shows relaxed selection conditions with model amino acids and was deemed suitable for use in a selection experiment.

As an alternative negative selection system, the metabolic marker thymidine kinase (tdk) was explored for its influence on cell growth under selective and nonselective conditions.

In tdk knock-out bacteria, thymidylate is phosphorylated in a complementary pathway, making tdk a non-essential gene (see Figure 40). However, the toxic thymidine analog azidothymidine (AZT) is only activated by tdk, leading to lethal AZT incorporation into the DNA. This allows the setup of selective

Results and discussion

conditions by utilizing thymine-free minimal medium with added AZT, making *tdk* an interesting alternative for barnase as negative selection marker. The main advantage over barnase is the more thorough removal of unwanted variants, in that under selective conditions cells expressing *tdk* are killed by destruction of their DNA, as opposed to just inhibited by RNA destruction by barnase. Also, the dependence of an additive allows direct control of selection stringency.

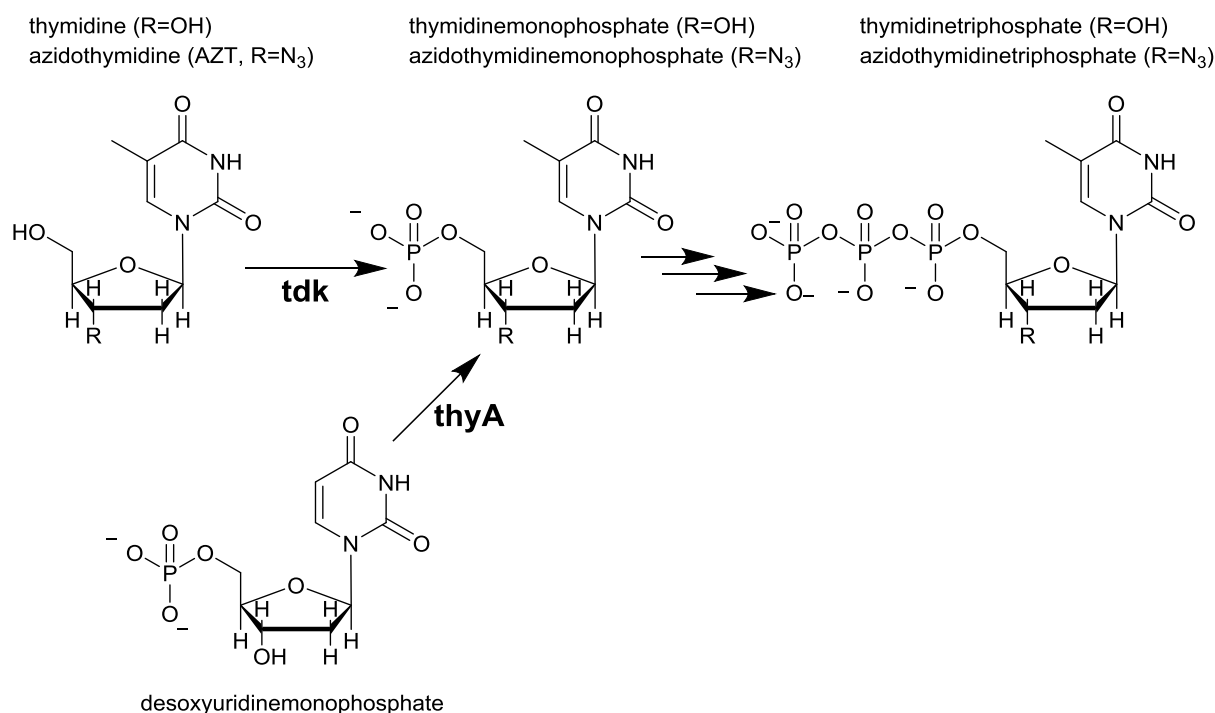


Figure 40: Negative selection scheme with *tdk* as selector. *tdk* is an enzyme involved in the TTP production pathway, catalyzing the monophosphorylation of thymidine to TMP. The enzyme is nonessential due to *thyA* providing an alternative pathway to TMP. While not inherently toxic, *tdk* cannot discriminate against AZT and is a crucial enzyme for the production of toxic AZT triphosphate from AZT. This conditional toxicity together with its other properties makes *tdk* a promising candidate for a negative selection marker.

Initial experiments showed promising results. A screening of different stop codon positions yielded variants that are not impaired in the presence of AZT without suppression, while cells with active wild-type *tdk* could not grow in the presence of AZT. However, suppression with Bok and the orthogonal pair *MmPyIRS*(Y384F)/*MmtRNA*^{Pyl} either did not impair cell growth, or fully prevented it (Figure 41) indicating too low or too high selection stringency, respectively. In conclusion, three promising variants with the stop codon combinations 2*/R29, 2*/Q79 and R29/Q79 were identified, but have yet to be assessed in a model selection experiment and the effect of AZT concentration has to be elucidated. Potentially, an optimal combination of stop codons and the ideal AZT concentration can yield an exceptionally efficient negative selection system. However, a clear limitation of this selection methodology is the dependency on both engineered knock-out strains and a chemically defined selective medium, making selection steps more time-consuming and laborious.

Results and discussion

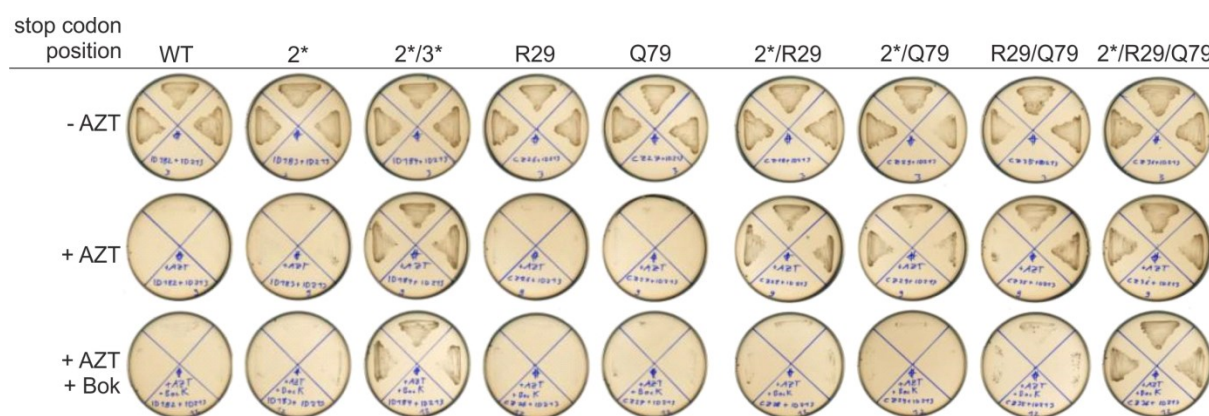


Figure 41: Negative selection system tests with tdk and PylRS. Several combinations of stop codons were tested for growth impedance in the presence of AZT and Bok in triplicate. While the combinations 2*/R29TAG, 2*/Q79TAG and R29/Q29TAG looked promising, the results could not yet be confirmed. *: The stop codon has been inserted at this position after the start codon. Each experiment was performed in triplicate by streaking three independent clones on the same test plate.

In this study the barnase system is applied for selection of PylRS variants. For further development of a tdk-based system it may be a feasible alternative route to use wild type tdk with an N-terminal suppression tag, as proposed by Pott *et al.*^[68].

Cat and barnase were successfully adapted for positive and negative selection and the performance was assessed with standard ncaas.

5.3.3 Selection of *MmPylRS* variants for activation of biocytin and S-allyl cysteine

With all components established and individually tested, selection experiments were carried out as described below to find PylRS variants for the incorporation of biocytin (Bct, from library B) and S-allylcysteine (Sac, from library S). A total of app. 3 µg library plasmid was transformed in 27 aliquots of DH10B pPNB_2_cat(112TAG)_pyIT ultra competent cells with an additional three transformations serving as controls. Transformants were grown on LB agar plates supplemented with ampicillin and kanamycin for plasmid maintenance, 37 µg/mL chloramphenicol as positive selector and 2 mM of either Bct or Sac. Control plates without noncanonical amino acids were prepared the same way. Colonies appeared on all plates after 48 h at 30 °C, but cells were harvested after 60 h to ensure growth of all positive variants. Cells from later positive selection steps were harvested after 48 h. Plasmids were isolated from 100 mL overnight culture, plasmids were split and the re-ligated library plasmid was used for transformation in the next selection round. For negative selection, DH10B pBU_pylT_barnase(2/44TAG) ultracompetent cells were transformed with the library plasmid and grown for 10-12 h at 30 °C on LB agar supplemented with ampicillin and kanamycin for plasmid maintenance. Cells were selected for two positive-negative cycles with a final positive selection round to iteratively enrich desired variants. The number of transformations was reduced each cycle to five for the final positive selection.

Results and discussion

After one positive-negative selection cycle, the next positive selection showed a slight increase in colony numbers compared to the control plates without noncanonical amino acid for SacRS selection. This difference was much more pronounced after the second positive-negative selection cycle, which was a strong indication for successful selection (Figure 42). No difference between selection and control plates was observed for selection of a BctRS, so no further selection experiments with Bct were performed.

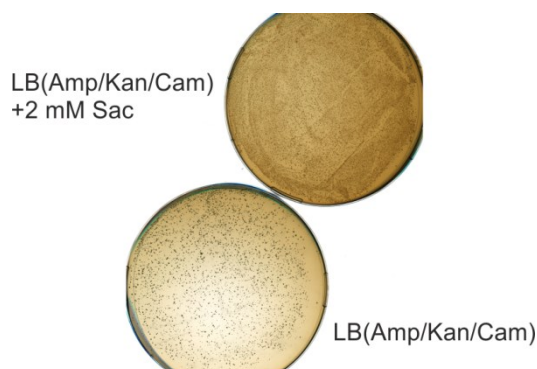


Figure 42: Plates of final positive selection step for Sac. After two positive and negative selection steps, cells showed a significant growth difference, growing to a meadow in the presence of Sac. This indicates a successful selection.

The final enriched library S was transformed in DH10B pPNB_2_cat(98/181TAG)_pyIT for a more stringent selection setup. Individual clones were spotted on selective medium with or without Sac and the growth behavior was monitored for 48 h. 11 of 12 clones tested showed growth dependent on the presence of Sac. Sequencing revealed that 9 of the 12 clones express a PylRS variant with the mutations C348W/W417S (the remaining three clones did not return a defined sequence). This result was confirmed by sequencing of 8 additional clones, indicating 75% library conversion to this genotype (Figure 43).

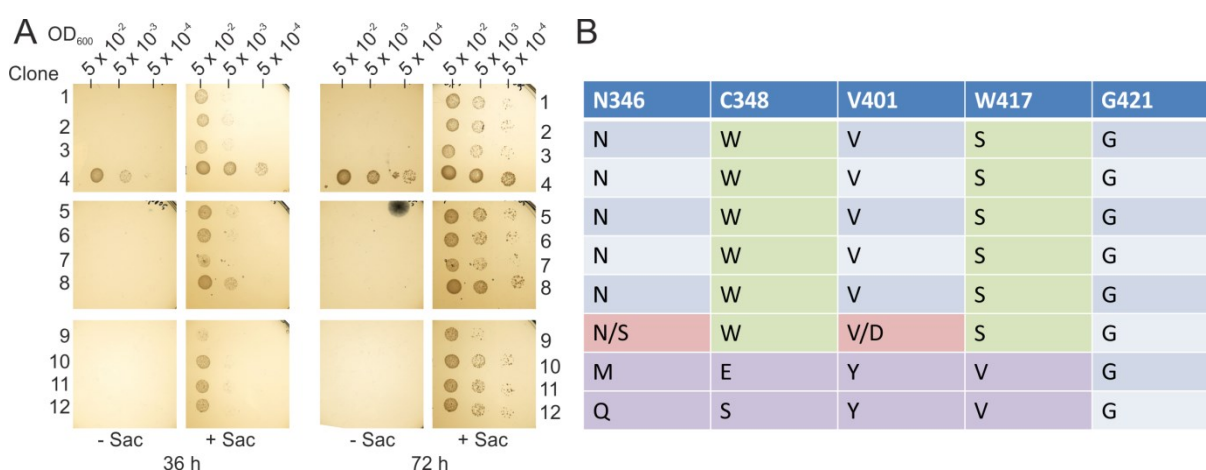


Figure 43: Peer-group tests. A: Sac-dependent growth of *E. coli* NEB10β transformed with a selector plasmid with cat(98/181TAG) and the *MmPylRS* library after 5 selection steps. 11 out of 12 colonies show chloramphenicol survival dependent on the presence of Sac. Clones 2, 3, 5 and 9-12 have the genotype *MmPylRS*(C348W/W417S) (SacRS). B: Peer-group sequencing of 8 clones of the final enriched library after 5 selection steps. Six of eight clones show the consensus sequence C348W/W417S, one of those has not fully defined positions 346 and 401, probably due to multiple plasmid transformation into the same cell.

Results and discussion

To confirm activity towards Sac, *MmPylRS*(C348W/W417S) was cloned into the suppression vector pJZ_P_{trp} and suppression was tested in the model protein EGFP(N150TAG) in the strain C321.ΔA.exp^[86], an amber-free RF1 knock-out strain. Full-length protein was detected via western blot and the presence of Sac was confirmed via mass spectrometry. Intriguingly, without addition of Sac to the medium a residual band at the expected migration length for full-length EGFP was still found. Mass analysis showed the incorporated amino acid to be Lys, Gln or Glu (Figure 44), indicating background suppression activity. The same background activity, however, was found with *MmPylRS*(Y384F), a variant known to exert very low background suppression. These findings indicate an influence of the bacterial strain.

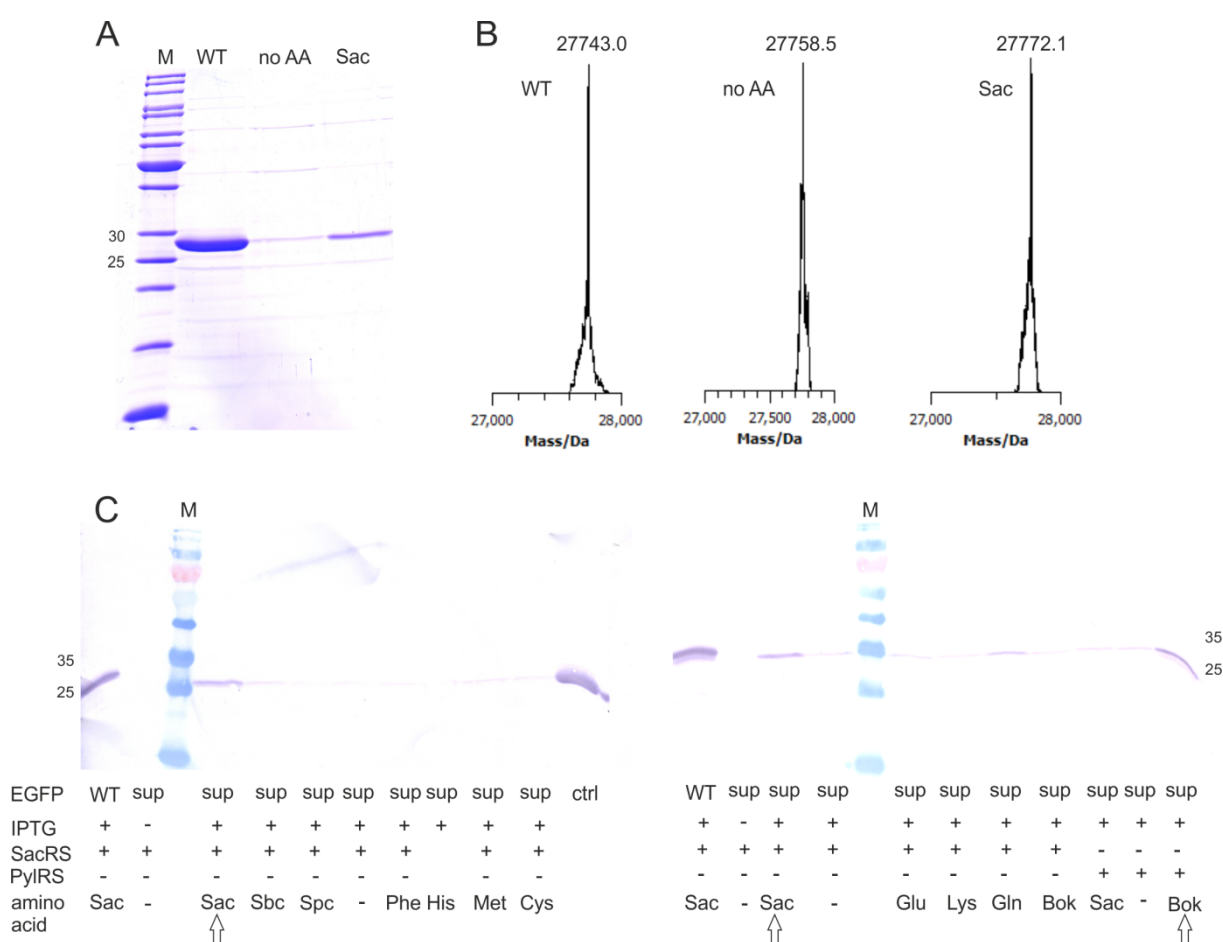


Figure 44: Confirmation of SacRS activity and specificity test. A: SDS-PAGE of purified EGFP without internal stop codon (WT), EGFP(150TAG) suppressed without addition of noncanonical amino acid (no AA) and EGFP(150TAG) suppressed with Sac to yield EGFP(150Sac) (Sac). B: Deconvoluted HPLC-ESI mass spectra of purified proteins. Expected masses are 27744.3 for EGFP(WT), 27758.4, 27759.3 or 27758.3 for suppression with Lys, Glu or Gln in absence of Sac and 27773.3 for EGFP(150Sac). C: Western blot analysis of suppression test lysates. Lysate from the same amount of cells was loaded on each lane. Note that the apparent difference in migration length between both blots is due to the fact that two different gel percentages (15 % and 20 %) were used. WT: EGFP without internal stop codon; sup: EGFP(150TAG); ctrl: purified EGFP. EGFP was expressed with either SacRS or *MmPylRS* in the presence of different amino acids (5 mM) in LB medium. The arrows indicate expected bands. All induced samples show a faint band due to background suppression of the used strain (C321.ΔA.exp).

Results and discussion

In conclusion, no PylRS variant for the activation of Bct was found and no enrichment of the library was observed. There are two possible explanations for this: Selection from a library is a difficult procedure, as the desired variants can be arbitrarily lost during early selection steps and a full sampling of all library members is not guaranteed. This can only partly be circumvented by optimization of the selection process. With regard to the transformation efficiency maximum of *E. coli*, approximately 10^{10} unique transformants can be generated in a standard experiment, which allows for a probability above 95 % of sampling every member. This number cannot feasibly be increased in a manual laboratory setup. Additionally, very few variants of all sequence possibilities may show the desired activity, leading to potentially very few true positive clones, which may be overgrown or be missed in the harvest step. Another possibility is that no variant from the sequence space of library B can potentially show activity for Bct. The choice of randomizations in library B was made according to libraries which have been successfully screened for activation of noncanonical amino acids with dimensions similar to Bct. This is, however, no guarantee that a variant for Bct activation is potentially present in the library. While the specificity of PylRS variants is mainly limited by the ability to accommodate the amino acid, other determinants do play a role. For example, a PylRS variant for propionyllysine was found inactive towards acryloyllysine despite the very similar size and shape of the amino acids ^[151]. Taking these findings into account, there is a possibility that despite the similarity to already incorporated amino acids, Bct is not accepted by any member of library B. A third explanation would be a rejection of Bct-tRNA^{Pyl} by EF-Tu or the ribosome, however, this could be ruled out as chemoenzymatically biocytinylated suppressor tRNA was shown to participate in translation ^[160,161].

In contrast, during the selection process of an *MmPylRS* variant for Sac activation, an enrichment of positive variants was observed from the second positive selection step on. Already after the first selection round, all colonies showed the wild-type glycine at position 421, indicating a strong influence of this residue to general PylRS folding or activity. After five selection steps, *MmPylRS*(C348W/W417S) was identified as a variant for Sac activation. This variant was dubbed SacRS. While the exact positioning of the amino acid during activation can only be speculated about, it can be hypothesized that the replacement of C348 by the bulky aromatic amino acid Trp closes off most of the amino acid binding pocket. This is supported by the restoration of the gate keeper residue N346. This residue, normally crucial for the N-carboxyl recognition and frequently mutated in PylRS variants activating noncanonical amino acids other than Lys derivatives ^[72,162,163], is shielded off by W348 in SacRS and its restoration may be favorable for overall protein expression or folding. The second mutation W417S may allow better accommodation of Sac in the lower binding pocket. SacRS shows good specificity for its substrate, efficiently discriminating against Sbc, Met and Aha, and does not retain wild-type activity towards Bok. Sac is the shortest-chain amino acid incorporated via stop codon suppression so far and constitutes the first member of a novel class of PylRS substrates.

Additionally, a SacRS can be used for establishing a synthetic pathway leading to metathesis-active proteins, since Sac has been identified as an excellent metathesis substrate in a protein context and is readily biosynthesized from allylthiol ^[164].

To show general applicability, Sac was incorporated into another model protein, TTL, which is better suitable for post-expression chemical reactions. Moreover, constructs with a TEV protease recognition site upstream of the hexahistidin-tag are available. Removal of the his-tag may prove advantageous for modification reactions, especially metathesis.

Results and discussion

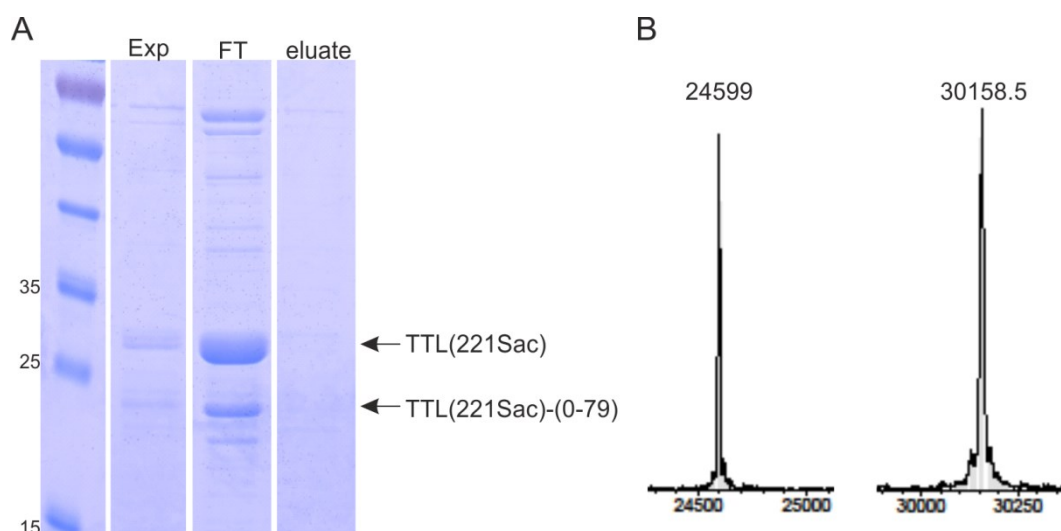


Figure 45: Incorporation of Sac into TTL and removal of the hexahistidin-tag: A: Monitoring of his-tag removal via SDS-PAGE. Exp: Expressed TTL(D221Sac) after affinity purification; FT: column flow-through after his-tag removal (contains the modified protein); eluate: column eluate after his-tag removal. The TTL gene contains a Met codon (UAG) at position 80, which can act as a secondary translational start codon, leading to a truncated product missing the amino acids at positions 1-79. B: Confirmation of Sac incorporation by mass spectrometry. Calculated masses are 30159.5 Da and 24599 Da for TTL(221Sac) and TTL(221Sac)-(0-79), respectively.

The incorporation was successful (see Figure 45) and TTL(D221Sac) is further tested for modification reactions. However, while Sac has been shown to be highly reactive in olefin metathesis and other reactions^[165,166], the reactivity of Sac incorporated into proteins remains to be proven.

5.4 Misincorporation of canonical amino acids

During several stop codon suppression experiments, background suppression in the absence of ncaa was observed. Protein purification and mass analysis was used to identify the misincorporated amino acid as Lys, Glu or Gln. These amino acids are too similar in mass to be distinguishable in a deconvoluted full-protein mass spectrum. Direct lysine incorporation at a stop codon is very unlikely, as no known pathway for this suppression exists, and PylRS variants are specifically evolved to not activate natural amino acids. However, since many noncanonical amino acids are lysine derivatives, hydrolysis of the head groups may be the reason for the apparent lysine incorporation. This head group loss is possible for suppression with Cok, but unlikely in experiments involving Pnk or Hnk, which comprise lysine modified via the more stable amide bond, and impossible for Sac and amino acids screened with AykRS, which are not lysine derivatives. Additionally, head group hydrolysis cannot explain background suppression in the absence of noncanonical amino acids.

Gln or Glu, on the other hand, are known natural suppressors for amber stop codons^[30]. While the strains used for suppression were devoid of natural suppressor tRNAs, glutamate has been found inserted at amber stop codons amber codon in an RF1-free strain^[167], indicating a higher than normal background suppression, potentially as a stress response. Additionally, tRNA^{Gln}_{CUG} recognizes the UAG stop codon with sufficient efficiency to promote suppression events^[168]. This specificity is also an explanation for the growth of control groups during positive selection. It should be noted that strong background suppression was observed with SacRS, but was easily outperformed by suppression in the presence of Sac, while *MmPylRS* variants incorporating Pnk and Hnk could not fully eliminate background incorporation, with residual protein suppressed with canonical amino acids found after purification. This is an indication of a competition event during translation.

While no quantitative data were gathered, background suppression was more pronounced when the strain C321.ΔA.exp was used, casting a doubt on its versatility for suppression experiments. While no direct manipulation in this strain can explain the enhanced background suppression, the removal of RF1 and all TAG stop codons may have an impact on regulatory processes or trigger stress responses which, along with off-target mutations throughout the strain's genome, may lead to a higher tendency of misincorporation. Further experiments were done using established suppression strains based on BL21. Competition with endogenous tRNA^{Gln} is inevitable but can be ignored if the desired suppression with a noncanonical amino acid is sufficiently efficient.

5.5 Excursion: Protein engineering with supplementation pressure incorporation

5.5.1 Incorporation of azidohomoalanine for protein modification

While the usefulness of azidohomoalanine (Aha) for protein modification with azide-alkyne cycloaddition has been demonstrated since 2003^[169,170], novel uses emerge regularly. A novel bodipy dye that shifts fluorescence to lower wavelengths upon conjugation has been recently developed in the group of Gerhard Wenz. This dye was used to modify ψ -barstar with Aha replacing Met1 via a Huisgen alkyne-azide cycloaddition^[171].

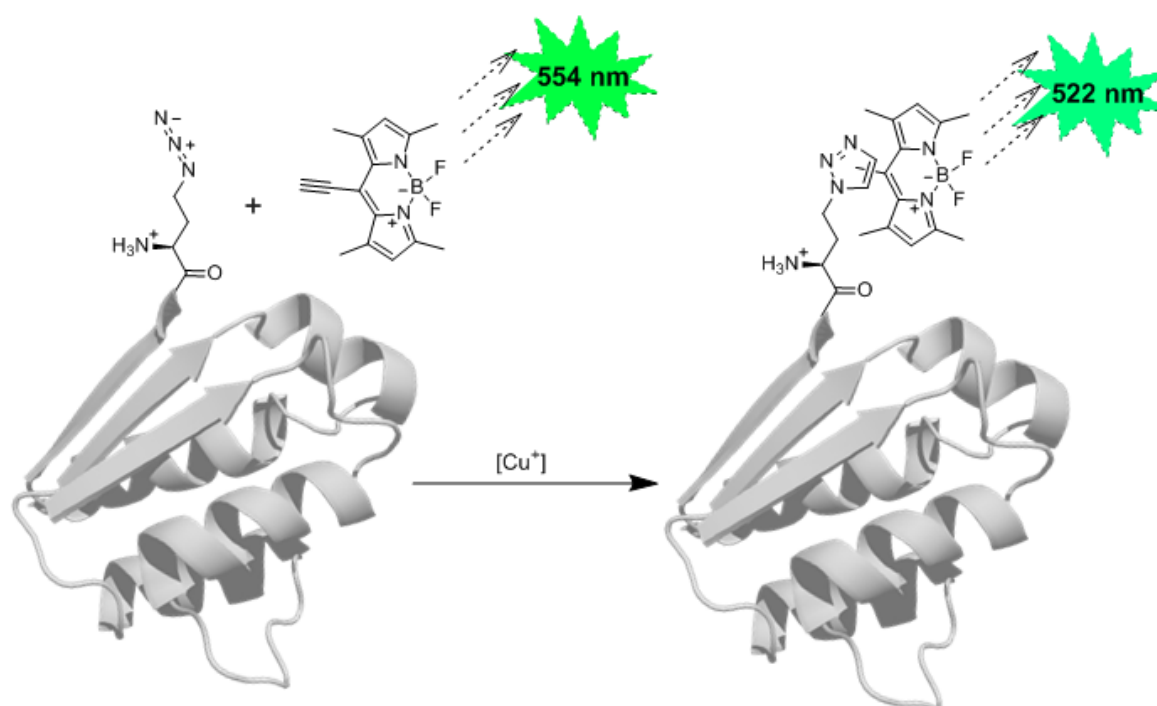


Figure 46: Modification scheme for posttranslationally modify barstar via Huisgen's cycloaddition. Conjugation of barstar(1Aha) and ethynylated bodipy via copper-induced [3+2] cycloaddition leads to a blue-shift in bodipy fluorescence.

The incorporation of Aha and bodipy conjugation was confirmed by SDS-PAGE and mass spectrometry. UV/VIS and fluorescence spectroscopy confirmed the expected shift in absorption and emission, while the protein melting point was elevated by app. 5 °C (all results are shown in Figure 47).

Results and discussion

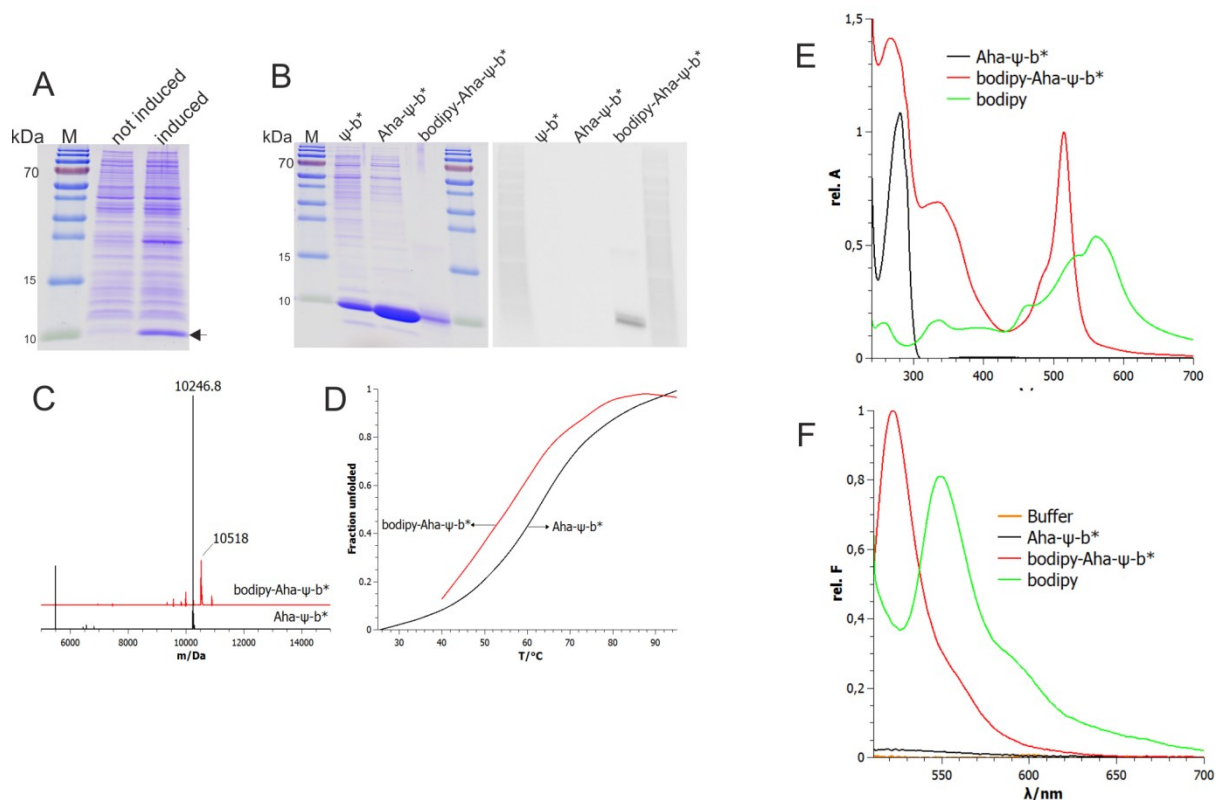


Figure 47: Purification and characterization of barstar(Aha) conjugated with ethynylated bodipy (partially adapted from [171]). A: Expression of Aha-ψ-b*. The induced culture shows a distinct band at the expected migration length ($M(\text{Aha-}\psi\text{-b}^*) = 10247.9$ Da, indicated by the arrow). B: SDS-PAGE gel of ψ-b* (3 μg), Aha-ψ-b* (5 μg) and bodipy-Aha-ψ-b* (0.75 μg). Bands were visualized at 365 nm (right) and with Coomassie staining (left). Only the band corresponding to bodipy-Aha-ψ-b* exhibits fluorescence. C: Deconvoluted ESI-MS spectrum of Aha-ψ-b* and bodipy-Aha-ψ-b*. The mass spectrum of Aha-ψ-b* shows a single peak at 10246.8 Da (theoretical mass 10247,9 Da), the spectrum of bodipy-Aha-ψ-b* exhibits a pronounced peak at 10518 Da (theoretical mass 10520 Da). D: Thermal denaturation curves of Aha-ψ-b* and bodipy-Aha-ψ-b*. Circular dichroism was monitored at 222 nm (characteristic minimum for α-helices). Melting points were calculated to be 62.2 °C for Aha-ψ-b* and 55.2 °C for bodipy-Aha-ψ-b*. E: Absorption spectra of unconjugated bodipy, ψ-b*(Aha) and conjugated Barstar(Aha bodipy). Both bodipy-Aha-ψ-b* and Aha-ψ-b* show a peak at 280 nm, as expected for proteins. bodipy-Aha-ψ-b* however, shows an additional peak at 508 nm, which is in correspondence to conjugated bodipy and clearly differs from the spectrum of unconjugated bodipy, which shows a maximum at 537 nm. F: Fluorescence emission spectra of unconjugated bodipy, Aha-ψ-b* and conjugated bodipy-Aha-ψ-b*. Fluorescence was determined with an excitation wavelength of 500 nm. A background spectrum of Tris buffer pH 8 is also shown. Aha-ψ-b* shows minimal emission signal due to light scattering. bodipy-Aha-ψ-b* exhibits a maximum at the expected wavelength (522 nm) that is blue-shifted from the emission of unconjugated bodipy at 554 nm.

The clear discrimination of free and conjugated bodipy and the minimal impact on protein stability and integrity make this method an interesting tool for conjugation method development, as it enables online monitoring of reaction turnovers.

5.5.2 Cell-free expression of proteins containing canavanines

Canavanine (Can) has been known to as an arginine analog since the 1960s^[172–174], but due to its toxicity, model protein preparations with full and confirmed arginine replacement have rarely been achieved. Canavanine incorporation in *E. coli* in particular has been demonstrated to be inefficient and unreliable^[26,172,175,176], sometimes producing a lower-mass protein due to deguanidylation of Can to homoserine^[177], or to require very sophisticated expression system^[178]. With the group of Albrecht Ott, a very simple cell-free expression has been applied for Can incorporation into a truncated EGFP, replacing six arginines^[150].

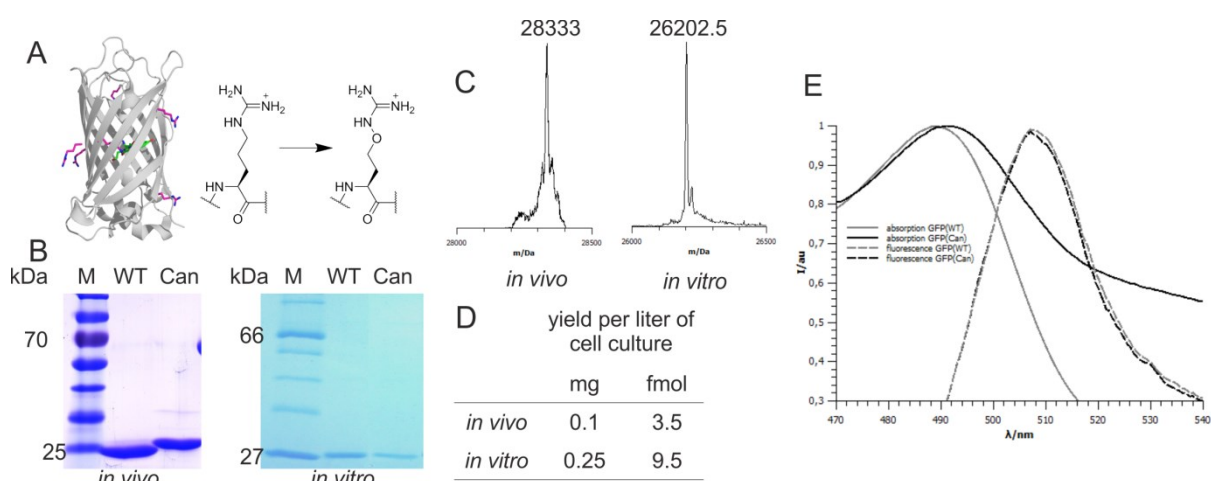


Figure 48: Replacement of Arg by Can in EGFP *in vitro* and *in vivo* (partially adapted from^[150]). A: Distribution of arginine residues (highlighted in magenta) in a protein model EGFP (structure from PDB-ID 2YOG). All residues are located in β -sheets. Five side chains face outwards, and only one is in contact with the fluorophore, localized in the barrel structure. Also shown are chemical structures of arginine and canavanine. B: Gel of purified GFP model proteins. Left: *In vivo* expressed EGFP(WT) and EGFP(Can) after affinity and size exclusion chromatography; right: *In vitro* expressed dEGFP(WT) and dEGFP(Can) after affinity chromatography. C: Confirmation of full incorporation of Can by ESI-MS; expected masses are 28333 Da (*in vivo*) and 26204 Da (*in vitro*). The mass difference of 1.5 Da for dEGFP(Can) is within the error of spectrum deconvolution. D: Typical yields of EGFP(Can) (*in vivo*) and dEGFP(Can) (*in vitro*). *In vitro* yield is a prediction based on cell-free expression using 90 μ L of cell extract derived from approximately 50 mL cell culture. E: Absorption and fluorescence spectra of dEGFP(WT) and dEGFP(Can). Five spectra were accumulated for each sample and normalized. After the incorporation of canavanine, the absorption maximum is slightly red-shifted (491 nm vs. 489 nm) with unchanged emission maximum at 508 nm.

It could be shown, that the expression of model proteins *in vitro* can be done with equal or better efficiency compared to standard expression, fully replacing all arginine residues and avoiding the deguanidylation reaction (see Figure 48). Additionally, the applied simple system greatly reduces laborious steps, making this method a viable tool for canavanine research.

6. Conclusion and outlook

The aim of this work was to develop novel tools to study and modify proteins via ncaas.

With several iterations, a reliable and efficient suppression system was developed, optimized and successfully applied to incorporate known substrates into ribosomal peptides for activity modification^[127]. Additionally, first steps were undertaken to establish efficient *in vitro* suppression. The suppression system was successfully engineered by rationally enlarging the amino acid binding pocket. Finally, an iterative positive-negative selection system was successfully adapted, tested and used to select novel variants from a library with five saturation mutations in the amino acid binding pocket. Additionally, the SPI method was applied to develop novel tools to study protein structure and function with noncanonical amino acids.

An unexpected additional finding is the apparent elevated background suppression observed in the strain C321.ΔA.exp. As this strain was explicitly developed to increase the efficiency of stop codon suppression with ncaas, these results warrant a systematic evaluation of the strain's performance and a full elucidation of the mechanistic background.

In conclusion, this work contributes of novel noncanonical amino acids and novel applications for ncaas already described in the literature, as well as two unique new variants of PyIRS:

Using the SPI method, new applications were found for the amino acids azidohomoalanine and canavanine. While the incorporation of canavanine into proteins is well-known, targeted full incorporations have been rarely achieved^[26,27,175]. In this study, an innovative *in vitro* method is presented, circumventing many problems occurring during cellular expression^[150]. As canavanine is not only an interesting structural destabilizer for proteins but also a potential anticancer agent^[176] and autoimmune stimulant^[179], this method can be expected to facilitate subsequent studies in many fields, including structural biology and cancer research. The incorporation of Aha into barstar allowed the characterization of a novel bodipy dye that shifts fluorescence upon conjugation. This provides an easy way to monitor the conjugation efficiency and will be employed for method optimization^[171].

In this work, three novel olefinic amino acids are added to the ever-growing number of noncanonical amino acids incorporated via stop codon suppression. Hnk and Pnk are exceptionally stable amino acids due to the modification of lysine via the inert amide bond. This superior stability allows applications in a wide spectrum of environments. The long chain locates the modification site away from the protein surface which may increase tolerance and availability for bioconjugations, with an additional choice of tether length between Pnk and Hnk. These features make Hnk and Pnk interesting targets for chemical modification. While Hnk was incorporated with a well-known promiscuous rational PyIRS variant, a family of PyIRS derivatives with rationally enlarged amino acid binding pockets was developed for Pnk incorporation. It is of interest to assess the substrate specificity of these variants, to find novel unspecific variants for the activation of until now unincorporated ncaas, and to gain additional insights on the mode of binding between engineered PyIRSs and their cognate amino acids.

Conclusion and outlook

S-allylcysteine is incorporated with a novel selected variant of pyrrolysyl-tRNA synthetase. The allylic sulfur greatly enhances the reactivity of the double bond^[118,180,165], making it a uniquely desirable amino acid for subsequent bioconjugations. Furthermore, Sac is a natural product abundant in garlic^[181], making direct extraction or fermentation by engineered microorganisms a possibility^[182], greatly enhancing the opportunities for applications on an industrial scale, feat that has not been achieved with any other ncaa. A simple pathway to *in situ* production from allylthiol by employment of the cysteine biosynthetic pathway in *E. coli* has recently been demonstrated^[164], making this a rare example for the combination of biosynthetic production and translational incorporation of noncanonical amino acids. Interestingly, the selected SacRS only shows two mutations relative to PylRS, none affecting the gatekeeper residue. This has implications for structure-function relationships in engineered PylRS. An interesting follow-up will be the determination of the crystal structure of SacRS, ideally with bound Sac, to gain insight on the mode of interactions. This information should lead to novel conclusions enabling further design of a new generation of engineered PylRS variants.

7. Materials and methods

7.1 Materials

7.1.1 Technical equipment

Centrifuges

Centrifuge 5810 R	Eppendorf AG, Hamburg, D
Centrifuge 5418 R	Eppendorf AG, Hamburg, D
MiniSpin plus	Eppendorf AG, Hamburg, D
Heraeus Fresco 17	Thermo Scientific, Waltham, USA
Avanti J-26 XP	Beckman Coulter, Krefeld, D

Incubators, mixers & shakers

Ecotron	Infors HT, Einsbach, D
Multitron	Infors HT, Einsbach, D
Incubator series B, KB	Binder, Tuttlingen, D

Gel electrophoresis and blotting

Horizontal agarose gel system	Custom made by Max-Planck Institute for Biochemistry, Martinsried, D
Electrophoresis unit	Hoefer Scientific Instruments, Holliston, USA
Vertical SDS-gel system	Custom made by Max-Planck Institute for Biochemistry, Martinsried, D
Fastblot B43	Biometra, Jena, D
Power supply Power Pack P25 T	Biometra, Jena, D
Power supply Consort E 143	Sigma-Aldrich, Taufkirchen, D

Thermocyclers

Mastercycler Gradient	Eppendorf AG, Hamburg, D
Peqstar 2x Gradient	Peqlab, Erlangen, D

Spectroscopy

Ultrospec 6300 pro	Amersham Biosciences (now: GE Healthcare, München, D)
BioPhotometer plus	Eppendorf AG, Hamburg, D
CD spectrometer J-815	Jasco Deutschland, Groß-Umstadt, D
Fluorescence spectrometer LS 55	Perkin Elmer, Rodgau, D

Materials and methods

Liquid Chromatography

Äkta purifier	GE Healthcare Life Sciences, München, D
Peristaltic pump P1	Pharmacia Biotech (now: GE Healthcare Life Sciences, München, D)

Mass spectrometry

MS Exactive	Thermo Scientific, Waltham, USA
QTOF 6530	Agilent, Santa Clara, USA

Thermal shakers

Thermomixer compact	Eppendorf AG, Hamburg, D
Thermomixer 5437	Eppendorf AG, Hamburg, D
Mixing Block MB-102	Bioer Technology, Binjiang, PRC

Balances

TE 1502S	Sartorius, Göttingen, D
GR-120	A&D, San Jose, USA
Mettler PE 3600 Deltarange	Mettler Toledo, Gießen, D

Miscellaneous

Sonopuls HD 3200	Bandelin, Berlin, D
Sonotrodes MS72, KE76	Bandelin, Berlin, D
Microfluidizer M-110L	Microfluidics, Newton, USA
Orbital shaker Rotamax 120	Heidolph, Schwabach, D
Vortex Genie	Bender & Hobein AG, Zürich, CH
Ice machine Scotsman AF 80	Scotsman, Vernon Hills, USA
pH-Meter S20-SevenEasy	Mettler Toledo, Gießen, D
Gel-documentation system Felix 2050	Biostep, Jahnsdorf, D
Scanner ViewPix 700	Biostep, Jahnsdorf, D
Water bath VWB 12	VWR, Darmstadt, D
Elektroporator MicroPulser™	Bio-Rad Laboratories GmbH, Munich, D

7.1.2 Chemicals

Standard chemicals were purchased from Carl Roth GmbH (Karlsruhe, D), Merck (Darmstadt, D), VWR International GmbH (Darmstadt, D) or Sigma-Aldrich (Taufkirchen, D) unless stated otherwise. Chemicals were purchased "p. a." or similar grade, buffer substances and media components were purchased in cell-culture grade or similar.

Materials and methods

Noncanonical amino acids were obtained from the following vendors/academic partners: N_ε-Boc-lysine and N_ε-Cbz-lysine were obtained from the group of Professor L. Moroder (Max Planck Institute for Biochemistry, Martinsried, D); Azidohomoalanine and N_ε-propargyloxycarbonyllysine were synthesized in-group by Dr. P. Durkin; N_ε-norbornylurea-lysine, N_δ-norbornyl-aminopentanoic acid and 2,7-diaminooct-4-enedioic acid were synthesized in the group of Professor Siegfried Blechert (TU Berlin, D); Canavanine was purchased from Sigma-Aldrich (Taufkirchen, D); N_ε-norbornoyllysine was obtained from the group of Professor R. Suessmuth (TU Berlin, D); N_ε-heptenoyllysine was synthesized in the group of Marie-Pierre Heck (CEA, Gif-sur-Yvette, F); N_ε-acryloyl-lysine and biocytin were purchased from Syntheval (Caen, F); S-allylcysteine was purchased from TCI (Eschborn, D); N_ε-Cyclooctynyloxycarbonyllysine was purchased from SiChem (Bremen, D); all other noncanonical amino acids were synthesized in the group of Professor P. Herdewijn (KU Leuven, B).

7.1.3 Media and supplements

Liquid media were autoclaved for 20 min at 121 °C at 1.5 bar; for agar plates 1.5% agar was added before autoclaving. Antibiotics were sterilized by filtration and then added to media not above 50 °C.

LB	10 g L ⁻¹ peptone/tryptone
	5 g L ⁻¹ yeast extract
	10 g L ⁻¹ NaCl
NMM	7.5 mM (NH ₄) ₂ SO ₄
	8.5 mM NaCl
	22.5 mM KH ₂ PO ₄
	50 mM K ₂ HPO ₄
	20 mM glucose
	1 mM MgSO ₄
	1 mg L ⁻¹ CaCl ₂
	1 mg L ⁻¹ FeCl ₂
	10 mg L ⁻¹ thiamine
	10 mg L ⁻¹ biotin
	0.01 mg L ⁻¹ trace elements (CuSO ₄ , ZnCl ₂ , MnCl ₂ (NH ₄) ₂ MoO ₄)
	10 mg L ⁻¹ of each amino acid
SOB	5 g L ⁻¹ yeast extract
	20 g L ⁻¹ tryptone
	0.6 g L ⁻¹ NaCl
	0.2 g L ⁻¹ KCl
	10 mM MgCl ₂
	10 mM MgSO ₄

Materials and methods

SOC	20 mM glucose in SOB
EnPresso B	Tablets (BioSilta, St. Ives, UK) dissolved according to manufacturer's instructions
supplements	ampicillin ($100\ \mu\text{g mL}^{-1}$) kanamycin ($50\ \mu\text{g mL}^{-1}$) chloramphenicol ($37\ \mu\text{g mL}^{-1}$) IPTG ($0.5 - 1\ \text{mM}$) noncanonical amino acid ($1 - 10\ \text{mM}$)

7.1.4 Strains

Strain name	Genotype	Source/Reference
<i>E. coli</i> DH5α	<i>E. coli</i> K12 C> F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ-	Life Technologies, Darmstadt, D
<i>E. coli</i> BL21 Gold	<i>E. coli</i> B C> F- ompT hsdS(rB- mB-) dcm+ Tetr gal λ (DE3) endA Hte	Life Technologies, Darmstadt, D
<i>E. coli</i> BL21 (DE3)	<i>E. coli</i> B C> F- ompT gal dcm lon hsdSB(rB- mB-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Life Technologies, Darmstadt, D
<i>E. coli</i> B834	<i>E. coli</i> B C> F- ompT gal dcm lon hsdSB (rB- mB-) I(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) met	Life Technologies, Darmstadt, D
DH10b	<i>E. coli</i> K12 C> F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara) 7697 galE15 galK16	Life Technologies, Darmstadt, D
MG1655	<i>E. coli</i> K12 F- λ- <i>ilvG- rfb-50 rph-1</i>	M. S. Guyer, CGSC collection ^[95]
MG1655Δtdk	<i>E. coli</i> MG1655 Δtdk::cat+, FRT+	AK Budisa
C321.ΔA.exp	<i>E. coli</i> MG1655 C> Δ (ybhB-bioAB)::zeoR ΔprfA, all TAG stop codons removed	George Church, Addgene plasmid # 49018 ^[86]

Materials and methods

7.1.5 Plasmids

All plasmids were stored at -20 °C.

Name	Function
pTRP-Duet1+ <i>MmPyl</i> ST-EYFP(114amber)	suppression of EYFP(114amber)
pQE80L-EGFP_H6	expression of EGFP
pQE80L-EGFP(150amber)-H6	expression of EGFP(150amber)
pTRP-Duet1+ <i>MmPyl</i> ST-EGFP	suppression system
pMEc1 μ C-prp-Strep-PylST	suppression system
pMEc1 μ C-prp-Strep-SaMetRS-SatRNA-Met	suppression system
pMEc1 μ C-prp-Strep-PylST-glnS'	suppression system
pPAB26'_cat(98/181TAG)	positive selection
pPAB26'_cat(98/181TAG)_ <i>MmPyl</i> Trev	positive selection
pPAB1_glnS'_ <i>MmPyl</i> S(Y384F)	library generation
pQE80L-psiBarstar(1M)	expression of Barstar
pJZ_P _{trp} _MmPylITS(Y384F)	general suppression
pJZ_P _{trp} _MmPylITS(Y384F/Y306A)	general suppression
pJZ_P _{trp} _AykRS	general suppression
pJZ_P _{trp} _TpPylITS	general suppression
pJZ_P _{trp} _MmPylITS(Y384F/C348A)	general suppression
pJZ_P _{trp} _MmPylITS(Y384F/L309A/C348A)	general suppression
pPAB26'_cat	positive selection
pPAB26'_cat_MmPylTrev	positive selection
pPAB26'_cat(112TAG)_MmpylTrev	positive selection
pNB26'2_pylT_barnase(2/44TAG)	negative selection
pJZ_P _{trp} _MePylITS	general suppression
pJZ_P _{trp} _MmPylITS	general suppression
pPAB18_glnS'_SacRS(C348W/W417S)	selected from library
pJZ_P _{trp} _SacRS(C348W/W417S)	general suppression

Materials and methods

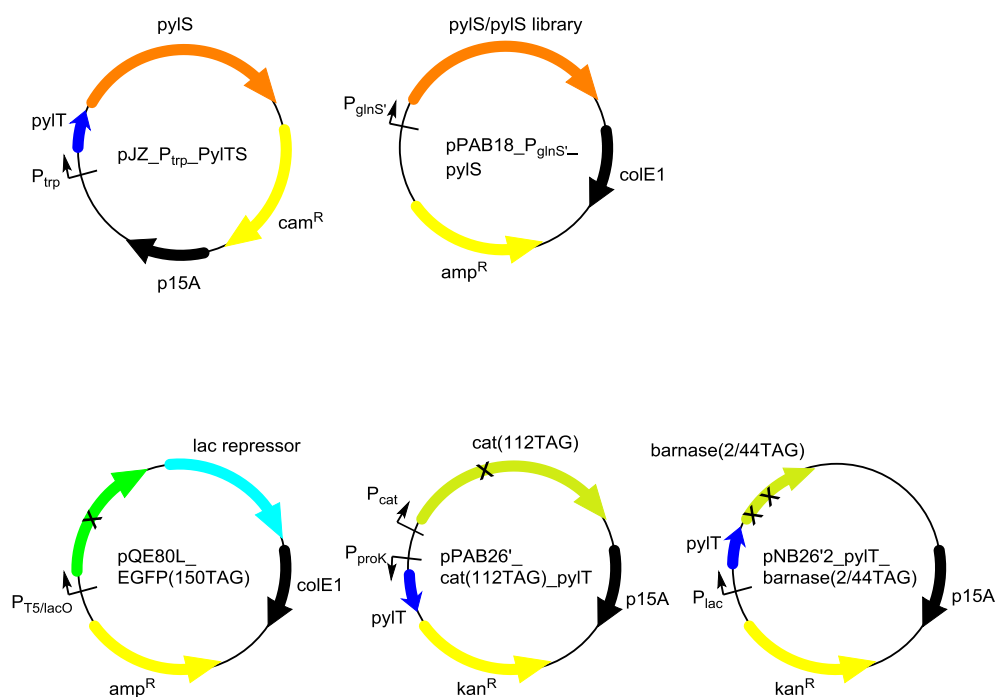


Figure 49: Finalized plasmids used in this work: Top row: general suppression plasmid (variable pylS variant); pylS library expression plasmid. Bottom row: general expression plasmid (variable target gene), positive selection plasmid (variable selector gene), negative selection plasmid (variable selector gene).

7.1.6 Synthetic oligonucleotides

All oligonucleotides were purchased from Metabion (Martinsried, D), Biomers (Ulm, D) or Sigma-Aldrich (Taufkirchen, D) at 100 μ M in ddH₂O and stored at -20°C. Oligonucleotides up to 40 bp were purchased in desalted quality without further purification. Longer molecules were ordered with HPLC purification.

7.1.7 Enzymes

Restriction enzymes and ligases were used for cloning while different polymerases were used in PCR and site directed mutagenesis.

Name	Supplier
Phusion DNA Polymerase	Thermo Scientific, Waltham, USA
Phusion PCR Master Mix	Thermo Scientific, Waltham, USA
Taq DNA Polymerase	Budisa group
Turbo Pfu-DNA-Polymerase	Agilent Technologies (Waldbronn)
FastDigest restriction enzymes	Thermo Scientific, Waltham, USA
Restriction enzymes	Thermo Scientific, Waltham, USA
T4 DNA Ligase	Thermo Scientific, Waltham, USA
DNase	Carl Roth (Karlsruhe)
RNase	Carl Roth (Karlsruhe)

7.1.8 Kits

Name	Supplier
GeneJET™ Plasmid Mini-prep Kit	Thermo Scientific, Waltham, USA
GeneJET™ Plasmid Midi-prep Kit	Thermo Scientific, Waltham, USA
GeneJET™ PCR Purification Kit	Thermo Scientific, Waltham, USA
GeneJET™ Gel Extraction Kit	Thermo Scientific, Waltham, USA

7.1.9 Buffers and Solutions

All buffers for SDS-PAGE, protein purification and agarose gel electrophoresis were prepared with desalted water (dH₂O). Buffers for protein purification with the peristaltic pump P1 or Äkta purification systems were additionally filtered in order to remove particles

Name	Composition
Coomassie staining solution	0.1% Coomassie Brilliant Blue R-250 in 50% EtOH/10% AcOH
5X SDS loading dye	80 mM Tris pH 6.8 10% SDS 12.5% glycerol 4% (v/v) mercaptoethanol 0.2% (w/v) bromophenol blue
10X SDS running buffer	62 g Tris 20 g SDS 288 g glycine ad 2 l dH ₂ O
resolving gel buffer	1.5 M Tris-HCl pH 8.8
stacking gel buffer	0.5 M Tris-HCl pH 6.8
resolving gel	15% or 20% acrylamide (37.5:1) 0.1% SDS 0.05% APS 0.05% TEMED 25% resolving gel buffer

Materials and methods

stacking gel	4% acrylamide (37.5:1) 0.1% SDS 0.05% APS 0.17% TEMED 25% stacking gel buffer
transfer buffer	25 mM Tris 192 mM glycine 20% methanol
10X TBS	50 mM Tris-HCl 300 mM NaCl pH 7.8
TBST	1x TBS 0.1% (v/v) Tween 20
Blocking solution	3% (w/v) BSA in TBST
AP buffer	100 mM Tris 100 mM NaCl 5 mM MgCl ₂ pH 9.5
NBT solution	50 mg/mL nitro blue tetrazolium chloride in 70% DMF
BCIP solution	50 mg/mL 5-Bromo-4-chloro-3-indolyl phosphate in DMF
50X TAE buffer	2 M Tris 2 M acetic acid 10% (v/v) 0.5 M EDTA pH 8.0
6X DNA loading dye	0.25% bromophenol blue 0.25% xlenecyanole 30% glycerol
binding buffer (IMAC)	50 mM NaH ₂ PO ₄ 300 mM NaCl 20 mM imidazole 0 – 10% glycerol pH8

Materials and methods

wash buffer (IMAC)	50 mM NaH_2PO_4 300 mM NaCl 30 mM imidazole 0 – 10% glycerol pH 8
elution buffer (IMAC)	50 mM NaH_2PO_4 300 mM NaCl 500 mM imidazole 0 – 10% glycerol pH 8
IVT Buffer	85 mM HEPES-KOH pH 8.3 450 mM potassium glutamate 400 mM NH_4OAc 60 mM $\text{Mg}(\text{OAc})_2$ 10 % (m/v) PEG 8000 10 mM DTT 10 mM of each amino acid
Bradford solution	0.01% Coomassie Brilliant Blue R-250 in 0.5% EtOH/8.5% H_3PO_4
protein storage buffer	50 mM tris-Cl pH 8 100 mM NaCl (10% glycerol)
buffer W	100 mM tris pH 8 150 mM NaCl 1 mM EDTA
buffer E	100 mM tris pH 8 150 mM NaCl 1 mM EDTA 2.5 mM desthiobiotin
buffer R	100 mM tris pH 8 150 mM NaCl 1 mM EDTA 1 mM HABA

7.2 Molecular biology methods

7.2.1 Isolation of plasmid DNA

Small-, medium- and large-scale plasmid DNA preparations from *E. coli* cells were performed with either GeneJET Plasmid Mini-prep Kit, GeneJET Plasmid Midi-prep Kit and GeneJET Plasmid Maxi-prep Kit (all Thermo Scientific, Waltham, USA) according to the manufacturer's instructions. Typically, 5 mL or 50 mL cell culture were used for small or medium-scale plasmid extraction, respectively. The DNA eluate was stored at -20 °C.

7.2.2 Polymerase chain reaction

PCR was used for amplification of DNA sequences for cloning, insert verification and for site-directed mutagenesis.

7.2.2.1 Standard PCR

Primers for standard PCR were designed manually or using the ApE software. For all DNA amplifications within a standard PCR, Phusion High-Fidelity DNA polymerase was used.

Compound	Phusion Polymerase	Phusion PCR Master Mix
ddH ₂ O	Add to 25 µL	Add to 25 µL
5X Phusion HF buffer	5 µL	-
10 µM dNTPs	0.5 µL	-
10 µM primer forward	0.5 µL	0.5 µL
10 µM primer reverse	0.5 µL	0.5 µL
Template	10 – 50 ng	10 – 50 ng
DMSO (optional)	3 – 5%	3 – 5%
Phusion DNA Polymerase	0.25 µL	-
2X Phusion PCR Master Mix	-	12.5 µL

PCR step	Temperature	Time	Cycles
Denaturation	95 °C	30 s	1
Denaturation	95 °C	15 s	30
Annealing	primer mp	20 s	
Extension	72 °C	20 s/kb	
Extension	72 °C	5 min	1
End	8 °C	hold	1

Materials and methods

After thermal cycling PCR products were analyzed for purity on agarose. The PCR product purification was performed with the GeneJET PCR Purification Kit (Thermo Scientific, Waltham, USA) whenever analytical agarose gel electrophoresis indicated that only fragments of the desired size were produced. Otherwise, preparative gel electrophoresis was performed using the GeneJET Gel Extraction Kit (Thermo Scientific, Waltham, USA). Randomizations were introduced using degenerate primers.

7.2.2.2 Colony PCR

Colony PCR was carried out to prove the presence of a certain DNA sequence on a plasmid in *E. coli* cells.

A PCR mastermix was prepared by adding the components for n+1 reactions and aliquoted in 0.2 mL reaction tubes (20 µL each). A single colony was picked with a sterile tip from a plate, briefly dipped into a reaction tube with mastermix and then streaked on a new LB replica plate. The replica plate was incubated at 37 °C to recover positive colonies.

Compound	Composition
ddH ₂ O	Add to 20 µL
10x Dream Taq Polymerase Buffer	2 µL
25 mM MgCl ₂	1.2 µL
10 µM dNTPs	0.5 µL
10 µM primer forward	0.5 µL
10 µM primer reverse	0.5 µL
Boiled colony	1 µL
Taq DNA Polymerase	0.6 µL

PCR step	Temperature	Time	Cycles
Denaturation	95 °C	3 min	1
Denaturation	95 °C	30 s	30
Annealing	primer mp -5 °C	30 s	
Extension	68 °C	30 s/kb	
Extension	68 °C	10 min	1
End	8 °C	hold	1

7.2.2.3 Site-directed mutagenesis

All mutations were introduced using the QuickChange Site-Directed Mutagenesis procedure (Agilent Technologies, Waldbronn, D). PfuTurbo DNA polymerase (Agilent Technologies, Waldbronn) was used in these PCR-like reactions. Mutagenic primers were designed according to the QuickChange protocol with the help of the online tool PrimerX (<http://www.bioinformatics.org>) with the mutation to be introduced in the middle of their sequence. In a linear amplification scheme, nicked circular plasmids with the desired mutations were created.

Compound	Composition
ddH ₂ O	Add to 25 µL
10X Pfu buffer	2.5 µL
DMSO	0.75 µL
10 µM dNTPs	0.5 µL
10 µM primer forward	0.5 µL
10 µM primer reverse	0.5 µL
Template	app. 50 ng
PfuTurbo DNA Polymerase	0.5 µL

PCR step	Temperature	Time	Cycles
Denaturation	95 °C	3 min	1
Denaturation	95 °C	30 sec	25
Annealing	55 °C	60 sec	
Extension	68 °C	2 min/kb	
Extension	68 °C	20 min	1
End	8 °C	hold	1

After thermal cycling, the PCR product was digested with 0.5 µL DpnI for 5 h to remove the template DNA, which is supercoiled and has a superior transformation efficiency, thus interfering with the transformation process. The digested product was transformed into *E. coli* competent cells for nick repair, and recovered via standard plasmid purification.

7.2.3 Cloning

All enzymes for cloning were purchased from Thermo Scientific, Waltham, USA.

7.2.3.1 *Ligation-based cloning*

Purified PCR products were used for ligation-based cloning. PCR products were purified by the use of PCR purification or gel extraction.

The purified DNA fragment and target vector were digested with appropriate restriction enzymes. Generally, app. 1 µg plasmid DNA or 0.5 µg PCR product were restricted with 0.5 µL restriction enzyme in the enzyme's buffer for 1 – 16 h. Digested DNA was spin-column purified or agarose gel purified.

For DNA ligation, 1 µL T4 DNA Ligase (5U, Thermo Scientific, Waltham, USA) was used to ligate 50 ng of vector DNA with the insert DNA in a threefold to tenfold molar excess in a total volume of 20 µL. The ligation reaction was carried out for 1 h at RT or overnight at 16 °C. Subsequently, 5 µL of the reaction mix was used to transform chemically competent *E. coli*. For transformation of electrocompetent *E. coli*, only 2-3 µL of ligation mix were used. Alternatively, for better transformation efficiency, the ligation reaction was desalted by dialysis on a filter plate against ddH₂O. App. 15 µL could be recovered from the filter plate and were used for transformation entirely.

7.3 Microbiological methods

7.3.1 Cultivation and storage of *E. coli* cells

E. coli strains were cultivated in different liquid and solid media supplemented with the appropriate antibiotics for selective growth or other supplements like ncaas. Solid medium was prepared in 12 or 18 cm agar plates or 24-well plates. Liquid medium cultivation was performed either in 14 mL cultivation tubes or different sized flasks (baffled or plain).

Short term storage of agar plates was done at 4 °C. For long-term storage, liquid cultures were supplemented with 7% sterile DMSO (1:1) and stored at -80 °C.

7.3.2 Preparation of competent cells

7.3.2.1 *Electro competent E. coli cells*

An *E. coli* overnight culture grown at 37 °C and 200 rpm was used to inoculate 200 mL LB medium. Cells were grown at 37°C and 200 rpm until the OD₆₀₀ reached 0.3 – 0.5. The cultures were immediately chilled on ice and cells were harvested by centrifugation (4 °C, 15 min, 3000x g). Afterwards, the supernatant was discarded and the cells washed twice with 20 mL ice-cold and sterile 10% glycerol. The washed cells were resuspended in 0.5 – 1 mL ice-cold sterile 10% glycerol and incubated on ice for 1 h. Finally, the cells were aliquoted to 50 µL, shock frozen in liquid nitrogen and stored at -80 °C. For ultracompetent cells, a culture of 500 mL was cultivated to reach an OD₆₀₀ of 0.3. Cells were harvested, washed three times in 10% glycerol (app. 300 mL total), resuspended in 1 mL of 10% glycerol. After spin-down and supernatant removal, the total volume was brought to 500 µl by adding app. 200 µL of 10% glycerol. Cells were aliquoted after incubation on ice for 1 h.

7.3.2.2 *CaCl₂ competent E. coli cells*

400 mL LB media was inoculated with 4 mL (1:100) of an overnight *E. coli* culture. Cells were incubated at 37 °C and 200 rpm until an OD₆₀₀ of 0.6 – 0.7 was reached. The culture was cooled on ice and portioned in 50 mL tubes. Subsequently, the cells were harvested by centrifugation at 4 °C (5000xg, 10 min). The supernatant was discarded and the pellets were cooled on ice. Each pellet was resuspended in 10 mL of ice-cold 100 mM MgCl₂ and further incubated on ice for 20 – 30 min. Cell suspension of two tubes were combined and centrifuged (4 °C, 10 min, 4000 rpm). Again, the supernatant was discarded and the cell pellet resuspended in 2 mL ice-cold CaCl₂ solution (100 mM CaCl₂, 15% glycerol). Finally, the competent cells were aliquoted to 50, shock frozen in liquid nitrogen and stored at -80 °C.

7.3.3 Transformation

7.3.3.1 Transformation of electro competent cells

Electro competent *E. coli* cells were thawed on ice for approximately 10 min. Plasmid DNA and electroporation cuvettes (1 mm gap width) were also prechilled on ice. 50 – 100 ng plasmid DNA was added to the cells and applied to the electroporation cuvette. Electroporation was performed in a BioRad Gene Pulser Xcell using the EC1 program (1.8 kV, one pulse). The time constant was monitored. Subsequently, 950 μ L LB or SOC medium were immediately added and the suspension was transferred to a sterile 1.5 mL reaction tube. For cell recovery, the culture was incubated at 37 °C and 450 rpm for 1 h. Finally, the cells were plated on agar plates with appropriate supplements and incubated.

7.3.3.2 Transformation of CaCl_2 competent cells

Chemically competent *E. coli* cells and plasmid DNA were thawed on ice for approximately 10 min. 50 – 100 ng of plasmid DNA were added to the cells and the mixture was incubated on ice for 30 min. Afterwards a heat shock at 42 °C was applied for 2 min and 950 μ L LB or SOC medium (pre-warmed to 42 °C) was immediately added. Cells were incubated for 45 min at 37 °C and 450 rpm. Subsequently, the cells were plated on agar plates with appropriate supplements and incubated overnight at 37 °C.

7.3.4 Standard gene expression in *E. coli*

For recombinant protein expression, the plasmid with the target gene was transformed into an *E. coli* expression strain (mostly *E. coli* BL21Gold(DE3) or BL21(DE3)). Standard gene expression was performed in LB medium. A preculture (1 - 50 mL LB with the appropriate antibiotics) was inoculated with a single colony or from a cryostock and grown overnight. For expression, a main culture was inoculated 1:100 to 1:50 from the preculture. Expression cultures were shaken at 37 °C up to $0.6 < \text{OD}_{600} < 0.8$. A non-induced sample was taken for reference and gene expression was induced as appropriate (0.5 - 1 mM IPTG or 20 mM sodium propionate). Protein expression was performed at 30 °C or 37 °C for 4-16 h. Another induced sample was taken for analysis by SDS-PAGE. Finally, the cells were harvested by centrifugation (20 min, 4 °C, 8000xg). The cell pellet was stored at -20 °C until further usage.

7.3.4.1 Test expression in *E. coli*

Small scale test expressions were carried out to evaluate soluble and insoluble expression of different proteins (i.e. different GFP constructs).

2 - 20 mL of LB medium with the appropriate antibiotics were inoculated 1:50 from an overnight culture and shaken at 37 °C until $0.6 < \text{OD}_{600} < 0.8$. After induction of gene expression the cultures were shaken at 30 °C for 4 h. Subsequently, the cells were harvested by centrifugation at 4 °C for 15 min at 3200xg or 1 min at 7000xg. The pellet was stored at -20 °C. Samples of the non-induced and induced culture were taken for analysis via SDS-PAGE.

7.3.4.2 Stop codon suppression

For stop codon suppression, an orthogonal pair of aaRS and its cognate suppressor tRNA was expressed alongside the amber-disrupted target gene, usually requiring a two-vector system. Expression was performed with the following alterations to the standard protocol: For non-constitutively expressed suppression systems, the expression of the orthogonal pair was induced 5 – 30 min prior to the target gene expression. The noncanonical amino acid was added directly before the target gene induction. In some cases, induction was delayed 5 – 10 min to allow dissolving and cellular uptake of the ncaa.

7.3.5 Selective-pressure incorporation (SPI) of noncanonical amino acids

The incorporation of ncaas with the SPI method requires *E. coli* strains that are auxotrophic for the *caa* that has to be exchanged. Here, methionine was exchanged for azidohomoalanine (Aha). Therefore, the methionine auxotrophic strain *E. coli* B834 (DE3) was used. Competent *E. coli* B834(DE3) cells were transformed with the plasmid containing the target gene under a T5 promotor. A 5 mL preculture in LB medium was grown overnight at 37 °C and 200 rpm. 1 L of New Minimal Medium (NMM) with the appropriate antibiotic and 45 µM methionine was inoculated 1:1000 from the overnight culture. Cells were grown overnight at 30 °C until depletion of methionine. The defined concentration of the canonical amino acid allows the production of cell mass up to an OD₆₀₀ value of app. 0.7. After 12 – 14 h, 1 mM Aha was added and the target gene expression was induced by addition of 0.5 mM IPTG. Protein expression was performed for 4 h at 30 °C. Subsequently, the cells were harvested by centrifugation at 8000xg, 4 °C for 20 min and the pellet was stored at -20 °C. Canavanine was incorporated at arginine positions analogously, utilizing the Arg auxotrophic strain AH5 JE7094.

7.3.5.1 SPI in an *in vitro* system

For incorporation of Can into target proteins replacing Arg *in vitro*, the cell-free expression system developed in the group of Vincent Noireaux^[149] was modified by preparing the reaction buffer with Can instead of Arg.

7.3.6 Cell lysis

7.3.6.1 Cell lysis of test expressed proteins

The cell pellet from was resuspended in 500 – 1000 µL lysis buffer. 100 µg mL⁻¹ lysozyme, 0.2 mM PMSF, and 100 µg mL⁻¹ of each DNase I and RNase A were added. The mixture was incubated on ice for 30 -60 min. To enhance cell lysis, the mixture was subjected to freeze-thaw cycles for 1-3 times. Cell debris was pelleted by centrifugation at 13200xg for 15 min at 4 °C. The cell lysate was immediately stored on ice until purification.

7.3.6.2 *Standard cell lysis for soluble proteins*

Cells were thawed on ice and resuspended in 5 ml binding buffer per gram cell pellet. 100 $\mu\text{g mL}^{-1}$ lysozyme, 0.2 mM PMSF, and 100 $\mu\text{g mL}^{-1}$ of each DNase I and RNase A were added. The mixture was incubated on ice for 30 -60 min.. Afterwards, the cells were lysed via a pressure-based homogenizer (Microfluidizer M-110L) or by sonication (Sonopuls HD3200) for 15 min, 73% amplitude at an interval of 1 sec on and 4 sec off. Cellular debris was removed by centrifugation at 18000xg for 30 min at 4 °C. For protein purification on liquid chromatography systems the lysate was cleaned by ultrafiltration using a syringe filter with a pore size of 0.22 μm .

7.3.6.3 *Cell lysis and purification of proteins from inclusion bodies*

Barstar is expressed in inclusion bodies and needs refolding. Therefore, pellets were resuspended in 50 mM Tris pH 8 before lysis. After cell lysis and lysate clearing, the supernatant was discarded and the debris pellet, containing the inclusion bodies, was resuspended in 20 mL 50 mM Tris pH 8 with 8 M urea. After 15 min of incubation, the lysate was cleared again. The supernatant, now containing the barstar, was dialyzed three times against 5 L 50 mM tris to refold the protein. The first dialysis was carried out for 4 h, the second overnight and the third again for 4 h. After refolding, the solution was centrifuged again at 15000xg for 45 min at 4 °C. The supernatant was filtered (0.22 μm) and stored on ice for purification.

7.4 Biochemical methods

7.4.1 Purification of His₆-tagged proteins by immobilized metal affinity chromatography

Immobilized metal ion chromatography (IMAC) was carried out for EYFP and EGFP variants and TTL. The following formats were used: FPLC (HisTrap FF crude columns, 1 or 5 mL, GE Healthcare, Munich, D) was performed on an ÄKTA FPLC system for large-scale purification or a on a free-standing peristaltic pump with optical control of binding and elution for large and medium-scale purifications. Gravity flow columns were loaded with Ni-NTA sepharose (Qiagen, Hilden, D) and used for medium-scale purification of fluorescent proteins. Small-scale purification was carried out using midi cobalt resin spin columns (HIS buster, amocol, Teltow, D) or mini nickel resin spin columns (HisPur, Pierce, Darmstadt, D). First, the respective column was equilibrated with binding buffer followed by application of the cell lysate onto the column. The column was washed with 10 – 20 column volumes of wash buffer and the protein was eluted with elution buffer in 3 – 5 fractions of 1 column volume each. Fractions containing the desired protein were pooled. The buffer was exchanged to storage buffer either by dialysis, diafiltration or a desalting column (Nap5 or Nap10, GE Healthcare, Munich, D). Protein purity was analyzed by SDS-PAGE.

7.4.2 Purification of strep-tagged proteins

Strep-tagged *MmPyIRS* and EYFP were purified using gravity flow columns packed with strep-tactin resin (IBA, Göttingen, D). Columns were equilibrated with buffer W. After application of the lysate, columns were washed with 10 column volumes buffer W. Elution was performed with 5 applications of 1 column volume of buffer E. Fractions containing the desired protein were pooled, the buffer was exchanged for storage buffer and protein purity was analyzed by SDS-PAGE. Columns were regenerated with buffer R and re-equilibrated.

7.4.3 Purification by ion exchange chromatography

Purification by ion exchange chromatography (IEX) was done for barstar. An anion exchange column (Resource Q 1 mL, GE Healthcare, Munich, D) was used. The column was equilibrated with 50 mM tris-Cl pH 8/100 mM NaCl and the protein solution was applied to the column. Proteins were eluted from the column with a NaCl gradient (0.1 M – 1 M NaCl). Fractions containing the desired protein were pooled, the buffer was exchanged for storage buffer and protein purity was analyzed by SDS-PAGE.

7.4.4 Size-exclusion chromatography

For enhanced purity, some samples were purified on a HiLoad 16/60 Superdex 75 size exclusion column using 50 mM tris-Cl pH 8 as mobile phase on an ÄKTA purifier FPLC system. Fractions containing the desired protein were pooled, the buffer was exchanged for storage buffer and protein purity was analyzed by SDS-PAGE.

7.5 Analytical methods

7.5.1 Agarose gel electrophoresis

DNA fragments were separated by agarose gel electrophoresis. DNA samples were run on 1% agarose gels in TAE buffer. Fragments above 5000 bp were visualized on 0.7% agarose, fragments smaller than 1000 bp on 2% agarose. To visualize DNA, gels contained 0.5 μ L of 10% ethidiumbromide (EtBr). DNA samples were mixed with 6X loading buffer (Thermo Scientific, Waltham, USA) and applied to the gel. Electrophoresis was performed at 100 mV for 20 -30 min. The GeneRuler Mix Ladder or GeneRuler 1 kb DNA Ladder (Thermo Scientific, Waltham, USA) was used as standard.

7.5.2 Polyacrylamide gel electrophoresis

Proteins were separated by polyacrylamide gel electrophoresis on 15% or 20% discontinuous polyacrylamide gels with 4% polyacrylamide stacking gel and PageRuler Prestained Protein Ladder or PageRuler unstained protein ladder (Thermo Scientific, Waltham, USA) as size standards. Samples were mixed with 5X SDS loading dye and boiled at 95 °C for 10 min prior to loading onto the gel. Electrophoresis was performed between 80 V for 15 min for protein stacking and then at 180 V for 30 min for resolving in SDS running buffer. Gels were used for western blotting or stained with Coomassie staining solution for 30 min at RT. For destaining, gels were boiled with water and a paper towel and agitated overnight at RT. Gels were documented by transmissive scanning.

7.5.3 Western Blot

His₆-tag containing proteins were detected via Western. First, the proteins were separated by SDS-PAGE and the unstained gel was soaked with transfer buffer for 2 min to avoid irregularities due to later shrinking. Four layers of Whatman paper and the PVDF membrane were cut to size. The membrane was activated by dipping in methanol, washed with dH₂O and soaked in transfer buffer together with the Whatman paper for a minimum of 15min at 4°C. The SDS gel was placed on the membrane and both were covered with two buffer soaked Whatman paper slides from both sides. Blotting was performed for 25 min per blotting stack at 300 mA. After blotting, the membrane was incubated for 1.5 h in blocking solution to block unspecific binding sites. The blocked membrane was subsequently incubated with rabbit monoclonal anti-polyHistidine antibody (Abcam, Cambridge, UK) (1:3000 in 1x TBST/3% BSA) for 2 h at RT. The membrane was washed three times for 5 min with TBST buffer, and incubated for 1 h with donkey anti-rabbit IgG conjugated with alkaline phosphatase (Abcam, Cambridge, UK) (1:100000 in 1x TBST, 3% BSA). After washing three times with TBST for 5 min each, the membrane was soaked in 15 mL AP buffer. Finally, 135 μ L NBT solution and 60 μ L BCIP solution were added. Staining was performed until maximum band intensity was reached (1 - 10 min). The reaction was stopped by substrate solution removal and the membrane was washed in dH₂O for 10 min. The membrane was scanned with in reflective mode. For semi-quantitative expression assessment, densitometric measurements were performed using the program ImageJ.

7.5.4 Determination of protein and DNA concentration

Protein concentrations were determined using the Bradford assay. 990 µL of Bradford solution were mixed with 10 µL of sample, incubated for 5 min and the absorption was measured at 595 nm. The protein concentration was calculated using a standard curve derived from measurements of standard samples of bovine serum albumin (BSA) solutions of known concentrations.

DNA concentration and potential contaminations with proteins were evaluated by absorbance measurements at 260 nm and 280 nm with background measurement at 320 nm. Concentrations of dsDNA were determined based on the assumption that an A_{260} of 1 equates to 50 µg/mL. The formula

$$c_{\text{DNA}} = (A_{260} - A_{320}) \cdot 50 \frac{\mu\text{g}}{\text{mL}} \cdot d$$

c_{DNA} = DNA concentration in µg/mL

d: dilution factor

was used. For purity assessment, the ratio

$$Q = \frac{A_{260} - A_{320}}{A_{280} - A_{320}}$$

was determined. It should be above 1.7 for pure DNA.

7.6 Biophysical methods

7.6.1 LC-ESI-MS

Purified proteins containing noncanonical amino acids were analyzed using LC-ESI-MS. 5 – 15 μL of protein solution (0.2 – 0.5 mg/mL) were injected. Proteins were separated on a C5 column (Supelco Analytical, Sigma-Aldrich, Taufkirchen) by reversed-phase chromatography with a gradient of 20% - 90% acetonitrile over 25 min. HPLC-purified proteins were injected into a coupled ESI-MS machine with orbitrap (Exactive) or quadrupole/time-of-flight (QTOF) analyser. The obtained spectrum was deconvoluted with the program MagTran or the software supplied with the instrument to obtain the zero-charge mass.

7.6.2 Circular dichroism spectroscopy for melting curves

Melting curves of proteins were measured at a concentration of 20 μM by monitoring the changes in residual ellipticity (i.e. unfolding) at 220 nm, the characteristic wavelength for α -helix formation. The protein solutions were heated from 25 °C or 40 °C to 95 °C with a rate of 30 °C h^{-1} in 110-QS Hellma quartz cells with an optical path length of 0.1 cm. These experiments were performed on a JASCO J-715 dichrograph equipped with a Peltier-type FDCD and water bath attachment (model PTC-423S/15 and F250).

7.6.3 Fluorescence spectroscopy

Fluorescence spectra of samples were measured on a LS 55 (PerkinElmer Life Sciences, Boston, USA) with an excitation/emission slit of 5 nm. The concentrations of the samples was 0.2 - 1 μM . At least three spectra were accumulated for each sample.

Absorption spectra of samples were measured on a Lambda 35 UV/VIS spectrophotometer (PerkinElmer Life Sciences, Boston, USA) in a range of 250 to 700 nm at 20 °C.

8. References

- [1] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 964–967.
- [2] G. Gamow, *Nature* **1954**, *173*, 318–318.
- [3] F. Crick, *Nature* **1970**, *227*, 561–563.
- [4] H. Matthaei, M. W. Nirenberg, *Biochem. Biophys. Res. Commun.* **1961**, *4*, 404–8.
- [5] P. Leder, B. F. Clark, W. S. Sly, S. Pestka, M. W. Nirenberg, *Proc. Natl. Acad. Sci. U. S. A.* **1963**, *50*, 1135–1143.
- [6] K. Chakraborty, A. Steinschneider, R. V. Case, A. H. Mehler, *Nucleic Acids Res.* **1975**, *2*, 2069–2076.
- [7] R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, A. Zamir, *Science* **1965**, *147*, 1462–1465.
- [8] B. F. C. Clark, *J. Biosci.* **2006**, *31*, 453–457.
- [9] T. Bullwinkle, B. Lazazzera, M. Ibba, *Annu. Rev. Genet.* **2014**, *48*, 149–166.
- [10] H. S. Zaher, R. Green, *Nature* **2009**, *457*, 161–166.
- [11] A. Moghal, K. Mohler, M. Ibba, *FEBS Lett.* **2014**, *588*, 4305–4310.
- [12] G. R. Moura, L. C. Carreto, M. A. Santos, *Curr. Opin. Microbiol.* **2009**, *12*, 631–637.
- [13] F. J. LaRiviere, A. D. Wolfson, O. C. Uhlenbeck, *Science* **2001**, *294*, 165–168.
- [14] S. Cusack, *Curr. Opin. Struct. Biol.* **1997**, *7*, 881–889.
- [15] J. Arnez, *Trends Biochem. Sci.* **1997**, *22*, 211–216.
- [16] L. Ribas de Pouplana, P. Schimmel, *Trends Biochem. Sci.* **2001**, *26*, 591–596.
- [17] T. Hilal, C. M. T. Spahn, *Mol. Cell* **2015**, *57*, 389–90.

References

- [18] D. W. Schultz, M. Yarus, *J. Mol. Evol.* **1996**, *42*, 597–601.
- [19] L. Ribas de Pouplana, M. a. S. Santos, J.-H. Zhu, P. J. Farabaugh, B. Javid, *Trends Biochem. Sci.* **2014**, *39*, 355–362.
- [20] H. Ryšlavá, V. Doubnerová, D. Kavan, O. Vaněk, *J. Proteomics* **2013**, *92*, 80–109.
- [21] L. Cesaro, L. A. Pinna, M. Salvi, *Curr. Genomics* **2015**, *16*, 128–138.
- [22] R. Bischoff, H. Schlüter, *J. Proteomics* **2012**, *75*, 2275–2296.
- [23] E. V. Koonin, A. S. Novozhilov, *IUBMB Life* **2009**, *61*, 99–111.
- [24] J. Ling, P. O'Donoghue, D. Soll, *Nat Rev Micro* **2015**, *advance on*.
- [25] O. Namy, J. Rousset, S. Naphine, I. Brierley, **2004**, *13*, 157–168.
- [26] M. H. Richmond, *Biochem. J.* **1959**, *73*, 261–264.
- [27] G. a. Rosenthal, D. L. Dahlman, *J. Agric. Food Chem.* **1991**, *39*, 987–990.
- [28] H. Beier, M. Grimm, *Nucleic Acids Res.* **2001**, *29*, 4767–82.
- [29] D. L. Hatfield, D. W. E. Smith, B. J. Lee, P. J. Worland, S. Oroszlan, *Crit. Rev. Biochem. Mol. Biol.* **1990**, *25*, 71–96.
- [30] H. Inokuchi, M. Kodaira, F. Yamao, H. Ozeki, *J. Mol. Biol.* **1979**, *132*, 663–677.
- [31] B. Hao, W. Gong, T. K. Ferguson, C. M. James, J. a Krzycki, M. K. Chan, *Science* **2002**, *296*, 1462–6.
- [32] G. Srinivasan, C. M. James, J. a Krzycki, *Science* **2002**, *296*, 1459–1462.
- [33] C. Polycarpo, A. Ambrogelly, A. Bérubé, S. M. Winbush, J. a McCloskey, P. F. Crain, J. L. Wood, D. Söll, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12450–4.
- [34] O. Namy, Y. Zhou, S. Gundllapalli, C. R. Polycarpo, A. Denise, J.-P. Rousset, D. Söll, A. Ambrogelly, *FEBS Lett.* **2007**, *581*, 5282–5288.
- [35] D. G. Longstaff, S. K. Blight, L. Zhang, K. B. Green-Church, J. A. Krzycki, *Mol. Microbiol.* **2007**,

References

- 63, 229–241.
- [36] Y. Zhang, P. V. Baranov, J. F. Atkins, V. N. Gladyshev, *J. Biol. Chem.* **2005**, *280*, 20740–20751.
 - [37] Z. Liu, M. Reches, I. Groisman, H. Engelberg-Kulka, *Nucleic Acids Res.* **1998**, *26*, 896–902.
 - [38] S. C. Low, M. J. Berry, *Trends Biochem. Sci.* **1996**, *21*, 203–208.
 - [39] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, *215*, 403–10.
 - [40] J. Yuan, P. O'Donoghue, A. Ambrogelly, S. Gundllapalli, R. L. Sherrer, S. Palioura, M. Simonović, D. Söll, *FEBS Lett.* **2010**, *584*, 342–349.
 - [41] C.-M. Zhang, C. Liu, S. Slater, Y.-M. Hou, *Nat. Struct. Mol. Biol.* **2008**, *15*, 507–514.
 - [42] A. Sauerwald, W. Zhu, T. A. Major, H. Roy, S. Palioura, D. Jahn, W. B. Whitman, J. R. 3rd Yates, M. Ibba, D. Soll, *Science* **2005**, *307*, 1969–1972.
 - [43] S. I. Hauenstein, J. J. Perona, *J. Biol. Chem.* **2008**, *283*, 22007–22017.
 - [44] H.-S. H.-S. Park, M. J. Hohn, T. Umehara, L.-T. Guo, E. M. Osborne, J. Benner, C. J. Noren, J. Rinehart, D. Soll, D. Söll, *Science (80-)*. **2011**, *333*, 1151–1154.
 - [45] N. Budisa, C. Minks, S. Alefelder, W. Wenger, F. Dong, L. Moroder, R. Huber, *FASEB J.* **1999**, *13*, 41–51.
 - [46] N. Budisa, C. Minks, F. J. Medrano, J. Lutz, R. Huber, L. Moroder, *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 455–9.
 - [47] N. Budisa, *Angew. Chemie - Int. Ed.* **2004**, *43*, 6426–6463.
 - [48] M. J. Pine, **1978**, *13*, 676–685.
 - [49] N. Budisa, S. Alefelder, J. A. E. H. Bae, R. Golbik, C. Minks, R. Huber, L. Moroder, **2001**, 1281–1292.
 - [50] D. B. Cowie, G. N. Cohen, *Biochim. Biophys. Acta* **1957**, *26*, 252–261.
 - [51] M. G. Hoesl, S. Oehm, P. Durkin, E. Darmon, L. Peil, H.-R. Aerni, J. Rappsilber, J. Rinehart, D. Leach, D. Söll, et al., *Angew. Chemie Int. Ed.* **2015**, *54*, 10030–10034.

References

- [52] M. Ibba, H. Hennecke, **1995**, *364*, 272–275.
- [53] J. Shao, I. V. Korendovych, J. Broos, *Amino Acids* **2015**, *47*, 213–216.
- [54] T. Fekner, M. K. Chan, *Curr. Opin. Chem. Biol.* **2011**, *15*, 387–391.
- [55] L. Wang, J. Xie, P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249.
- [56] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498–500.
- [57] A. Dumas, L. Lercher, C. D. Spicer, B. G. Davis, *Chem. Sci.* **2014**, DOI:10.103, 1–18.
- [58] W. Wan, J. M. Tharp, W. R. Liu, *Biochim. Biophys. Acta - Proteins Proteomics* **2014**, *1844*, 1059–1070.
- [59] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–44.
- [60] C. E. Melançon, P. G. Schultz, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3845–7.
- [61] D. Mendel, V. W. Cornish, P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 435–462.
- [62] C. Noren, S. Anthony-Cahill, M. Griffith, P. Schultz, *Science (80-.)*. **1989**, *244*, 182–188.
- [63] D. A. Dougherty, E. B. Van Arnem, *ChemBioChem* **2014**, *15*, 1710–1720.
- [64] R. Krishnakumar, J. Ling, *FEBS Lett.* **2014**, *588*, 383–388.
- [65] P. O'Donoghue, L. Prat, I. U. Heinemann, J. Ling, K. Odoi, W. R. Liu, D. Söll, *FEBS Lett.* **2012**, *586*, 3931–7.
- [66] R. Martin, M. K. Phillips-Jones, F. J. Watson, L. S. Hill, *Biochem. Soc. Trans.* **1993**, *21*, 846–851.
- [67] Q. Wang, T. Sun, J. Xu, Z. Shen, S. P. Briggs, D. Zhou, L. Wang, *ChemBioChem* **2014**, *15*, 1744–1749.
- [68] M. Pott, M. J. Schmidt, D. Summerer, *ACS Chem. Biol.* **2014**, *9*, 2815–2822.
- [69] J. J. Perona, A. Hadd, **2012**.
- [70] S. Nehring, N. Budisa, B. Wiltschi, *PLoS One* **2012**, *7*, e31992.

References

- [71] Y.-S. Wang, X. Fang, H.-Y. Chen, B. Wu, Z. U. Wang, C. Hilty, W. R. Liu, *ACS Chem. Biol.* **2013**, *8*, 405–415.
- [72] Y.-S. Wang, W. K. Russell, Z. Wang, W. Wan, L. E. Dodd, P.-J. Pai, D. H. Russell, W. R. Liu, *Mol. Biosyst.* **2011**, *7*, 714–717.
- [73] V. K. Lacey, G. V. Louie, J. P. Noel, L. Wang, *ChemBioChem* **2013**, *14*, 2100–2105.
- [74] H. Xiao, F. B. Peters, P.-Y. Yang, S. Reed, J. R. Chittuluru, P. G. Schultz, *ACS Chem. Biol.* **2014**, *9*, 1092–6.
- [75] D. P. Nguyen, M. Mahesh, S. J. Elsässer, S. M. Hancock, C. Uttamapinant, J. W. Chin, *J. Am. Chem. Soc.* **2014**, *136*, 2240–3.
- [76] J. Ko, Y.-S. Wang, A. Nakamura, L.-T. Guo, D. Söll, T. Umehara, *FEBS Lett.* **2013**, *587*, 3243–8.
- [77] K. Nozawa, P. O'Donoghue, S. Gundllapalli, Y. Arais, R. Ishitani, T. Umehara, D. Soll, O. Nureki, *Nature* **2009**, *457*, 1163–1167.
- [78] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Chem. Biol.* **2008**, *15*, 1187–1197.
- [79] Y. Nov, *Appl. Environ. Microbiol.* **2012**, *78*, 258–262.
- [80] H. Inoue, H. Nojima, H. Okayama, *Gene* **1990**, *96*, 23–28.
- [81] S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z. G. Zhang, D. J. Opperman, M. T. Reetz, J. P. Acevedo, *ACS Synth. Biol.* **2013**, *2*, 83–92.
- [82] H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat Chem Biol* **2008**, *4*, 232–234.
- [83] T. Mukai, a. Hayashi, F. Irah, a. Sato, K. Ohtake, S. Yokoyama, K. Sakamoto, *Nucleic Acids Res.* **2010**, *38*, 8188–8195.
- [84] W. H. Schmied, S. J. Elsasser, C. Uttamapinant, J. W. Chin, *J. Am. Chem. Soc.* **2014**, *136*, 15577–15583.
- [85] T. Mukai, H. Hoshi, K. Ohtake, M. Takahashi, A. Yamaguchi, A. Hayashi, S. Yokoyama, K. Sakamoto, *Sci. Rep.* **2015**, *5*, 9699.

References

- [86] M. J. Lajoie, A. J. Rovner, D. B. Goodman, H.-R. Aerni, A. D. Haimovich, G. Kuznetsov, J. a Mercer, H. H. Wang, P. a Carr, J. a Mosberg, et al., *Science* **2013**, *342*, 357–60.
- [87] S. Sando, A. Ogawa, T. Nishi, M. Hayami, Y. Aoyama, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1216–20.
- [88] D. B. F. Johnson, J. Xu, Z. Shen, J. K. Takimoto, M. D. Schultz, R. J. Schmitz, Z. Xiang, J. R. Ecker, S. P. Briggs, L. Wang, *Nat. Chem. Biol.* **2011**, *7*, 779–786.
- [89] K. Wang, H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Biotechnol.* **2007**, *25*, 770–7.
- [90] H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, *Nature* **2010**, *464*, 441–4.
- [91] T. Mukai, A. Yamaguchi, K. Ohtake, M. Takahashi, A. Hayashi, F. Iraha, S. Kira, T. Yanagisawa, S. Yokoyama, H. Hoshi, et al., *Nucleic Acids Res.* **2015**, *43*, 8111–8122.
- [92] M. G. Hoesl, Expanding the Toolkit of Protein Engineering: Towards Multiple Simultaneous In Vivo Incorporation of Noncanonical Amino Acids, Technische Universität München, **2012**.
- [93] Y. Zeng, W. Wang, W. R. Liu, *ChemBioChem* **2014**, *15*, 1750–1754.
- [94] N. Bohlke, N. Budisa, *FEMS Microbiol. Lett.* **2014**, *351*, 133–144.
- [95] L.-T. Guo, Y.-S. Wang, A. Nakamura, D. Eiler, J. M. Kavrán, M. Wong, L. L. Kiessling, T. A. Steitz, P. O'Donoghue, D. Söll, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 16724–16729.
- [96] M. a Gaston, L. Zhang, K. B. Green-Church, J. a Krzycki, *Nature* **2011**, *471*, 647–50.
- [97] M. A. Gaston, R. Jiang, J. A. Krzycki, *Curr. Opin. Microbiol.* **2011**, *14*, 342–9.
- [98] J. A. Krzycki, *Curr. Opin. Chem. Biol.* **2013**, *17*, 619–625.
- [99] S. E. Cellitti, W. Ou, H.-P. Chiu, J. Grunewald, D. H. Jones, X. Hao, Q. Fan, L. L. Quinn, K. Ng, A. T. Anfora, et al., *Nat Chem Biol* **2011**, *7*, 528–530.
- [100] W. Ou, T. Uno, H.-P. Chiu, J. Grunewald, S. E. Cellitti, T. Crossgrove, X. Hao, Q. Fan, L. L. Quinn, P. Patterson, et al., *Proc. Natl. Acad. Sci.* **2011**, *108*, 10437–10442.
- [101] M. Ehrlich, M. J. Gattner, B. Viverge, J. Bretzler, D. Eisen, M. Stadlmeier, M. Vrabel, T. Carell, *Chem. - A Eur. J.* **2015**, *21*, 7701–7704.

References

- [102] Y. Ma, H. Biava, R. Contestabile, N. Budisa, M. L. Di Salvo, *Molecules* **2014**, *19*, 1004–1022.
- [103] T. H. P. Maier, *Nat. Biotechnol.* **2003**, *21*, 422–427.
- [104] N. Terasaka, Y. Iwane, A.-S. Geiermann, Y. Goto, H. Suga, *Int. J. Mol. Sci.* **2015**, *16*, 6513–6531.
- [105] P. Marliere, *Syst. Synth. Biol.* **2009**, *3*, 77–84.
- [106] M. Schmidt, *BioEssays* **2010**, *32*, 322–331.
- [107] N. Budisa, *Curr. Org. Chem.* **2014**, *18*, 936–943.
- [108] P. Marliere, J. Patrouix, V. Doring, P. Herdewijn, S. Tricot, S. Cruveiller, M. Bouzon, R. Mutzel, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 7109–7114.
- [109] C. G. Acevedo-Rocha, N. Budisa, *Angew. Chemie Int. Ed.* **2011**, *50*, 6960–6962.
- [110] D. J. Mandell, M. J. Lajoie, M. T. Mee, R. Takeuchi, G. Kuznetsov, J. E. Norville, C. J. Gregg, B. L. Stoddard, G. M. Church, *Nature* **2015**, *518*, 55–60.
- [111] A. J. Rovner, A. D. Haimovich, S. R. Katz, Z. Li, M. W. Grome, B. M. Gassaway, M. Amiram, J. R. Patel, R. R. Gallagher, J. Rinehart, et al., *Nature* **2015**, *518*, 89–93.
- [112] J. Torres-Kolbus, C. Chou, J. Liu, A. Deiters, *PLoS One* **2014**, *9*, e105467.
- [113] Y. Li, M. Pan, Y. Li, Y. Huang, Q. Guo, *Org. Biomol. Chem.* **2013**, *11*, 2624–2629.
- [114] X. Li, T. Fekner, M. K. Chan, *Chem. Asian J.* **2010**, *5*, 1765–9.
- [115] T. Plass, S. Milles, C. Koehler, C. Schultz, E. A. Lemke, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 3878–81.
- [116] N. Li, R. K. V Lim, S. Edwardraja, Q. Lin, *J. Am. Chem. Soc.* **2011**, *133*, 15316–9.
- [117] J. S. Lampkowski, J. K. Villa, T. S. Young, D. D. Young, *Angew. Chemie Int. Ed.* **2015**, n/a–n/a.
- [118] Y. A. Lin, J. M. Chalker, B. G. Davis, *J. Am. Chem. Soc.* **2010**, *132*, 16805–11.
- [119] Y. A. Lin, J. M. Chalker, B. G. Davis, *Chembiochem* **2009**, *10*, 959–69.

References

- [120] H.-W. Ai, W. Shen, E. Brustad, P. G. Schultz, *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 935–7.
- [121] E. Kaya, M. Vrabel, C. Deiml, S. Prill, V. S. Fluxa, T. Carell, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 4466–9.
- [122] W. Song, Y. Wang, J. Qu, Q. Lin, *J. Am. Chem. Soc.* **2008**, *130*, 9654–9655.
- [123] J. Wang, W. Zhang, W. Song, Y. Wang, Z. Yu, J. Li, M. Wu, L. Wang, J. Zang, Q. Lin, *J. Am. Chem. Soc.* **2010**, *132*, 14812–14818.
- [124] M. G. Hoesl, C. G. Acevedo-Rocha, S. Nehring, M. Royter, C. Wolschner, B. Wiltschi, N. Budisa, G. Antranikian, *ChemCatChem* **2011**, *3*, 213–221.
- [125] L. Merkel, M. Schauer, G. Antranikian, N. Budisa, *ChemBioChem* **2010**, *11*, 1505–1507.
- [126] M. G. Hoesl, N. Budisa, *Chembiochem* **2011**, *12*, 552–5.
- [127] R. S. Al Toma, A. Kuthning, M. P. Exner, A. Denisiuk, J. Ziegler, N. Budisa, R. D. Süssmuth, *ChemBioChem* **2015**, *16*, 503–509.
- [128] J. Xie, W. Liu, P. G. Schultz, *Angew. Chemie Int. Ed.* **2007**, *46*, 9239–9242.
- [129] J. C. Lewis, *Curr. Opin. Chem. Biol.* **2015**, *25*, 27–35.
- [130] I. D. Petrik, J. Liu, Y. Lu, *Curr. Opin. Chem. Biol.* **2014**, *19*, 67–75.
- [131] C. Mayer, D. G. Gillingham, T. R. Ward, D. Hilvert, *Chem. Commun. (Camb)*. **2011**, *47*, 12068–70.
- [132] A. S. Bommarius, *Annu. Rev. Chem. Biomol. Eng.* **2015**, *6*, 319–345.
- [133] M. Dürrenberger, T. R. Ward, *Curr. Opin. Chem. Biol.* **2014**, *19*, 99–106.
- [134] O. Shimomura, F. H. Johnson, Y. Saiga, *J. Cell. Comp. Physiol.* **1962**, *59*, 223–239.
- [135] J. R. Enterina, L. Wu, R. E. Campbell, *Curr. Opin. Chem. Biol.* **2015**, *27*, 10–17.
- [136] A. S. Mishin, V. V. Belousov, K. M. Solntsev, K. A. Lukyanov, *Curr. Opin. Chem. Biol.* **2015**, *27*, 1–9.

References

- [137] B. P. Cormack, R. H. Valdivia, S. Falkow, *Gene* **1996**, *173*, 33–38.
- [138] L. Lybarger, D. Dempsey, G. H. Patterson, D. W. Piston, S. R. Kain, R. Chervenak, *Cytometry* **1998**, *31*, 147–52.
- [139] S. Delagrave, R. E. Hawtin, C. M. Silva, M. M. Yang, D. C. Youvan, *Biotechnology. (N. Y.)* **1995**, *13*, 151–154.
- [140] R. Golbik, G. Fischer, A. R. Fersht, *Protein Sci.* **1999**, *8*, 1505–14.
- [141] L. Merkel, H. S. G. Beckmann, V. Wittmann, N. Budisa, *Chembiochem* **2008**, *9*, 1220–1224.
- [142] A. de Simone, Dissertation, Technische Universität Berlin, **n.d.**
- [143] J. Ziegler, Pyrrolsyt-tRNA-Synthetasen-Basierte Orthogonale Paare Für Ortsspezifischen Einbau von Nicht-Kanonischen Aminosäuren in Escherichia Coli, Technische Universität Berlin, **2012**.
- [144] S. D. Amico, J. Marx, C. Gerday, G. Feller, **2003**, *278*, 7891–7896.
- [145] C. Schiraldi, M. Acone, M. Giuliano, I. Di Lernia, C. Maresca, M. Cartenì, M. de Rosa, *Extremophiles* **2001**, *5*, 193–8.
- [146] T.-W. Kim, J.-W. Keum, I.-S. Oh, C.-Y. Choi, C.-G. Park, D.-M. Kim, *J. Biotechnol.* **2006**, *126*, 554–561.
- [147] K. B. Lee, H.-C. Kim, D.-M. Kim, T. J. Kang, H. Suga, *J. Biotechnol.* **2013**, *164*, 330–335.
- [148] E. G. Worst, Dissertation, Universität des Saarlandes, **2015**.
- [149] J. Shin, V. Noireaux, *J. Biol. Eng.* **2010**, *4*, 8.
- [150] E. G. Worst, M. P. Exner, A. De Simone, M. Schenkelberger, V. Noireaux, N. Budisa, A. Ott, *Bioorg. Med. Chem. Lett.* **2015**, *25*, 3658–60.
- [151] Y.-J. Lee, B. Wu, J. E. Raymond, Y. Zeng, X. Fang, K. L. Wooley, W. R. Liu, *ACS Chem. Biol.* **2013**, *8*, 1664–70.
- [152] M. J. Schmidt, D. Summerer, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 4690–3.
- [153] N. Nischan, *Personal Correspondence*, **n.d.**

References

- [154] B. Levay-young, M. Olesiuk, C. Gustafsson, J. Minshull, O. B. Drive, M. Park, *Genet. Eng. Biotechnol. News* **2013**, 33, 1–3.
- [155] F. Cheng, L. Zhu, U. Schwaneberg, *Chem. Commun.* **2015**, 51, 9760–9772.
- [156] K. Lang, L. Davis, S. Wallace, M. Mahesh, D. J. Cox, M. L. Blackman, J. M. Fox, J. W. Chin, *J. Am. Chem. Soc.* **2012**, 134, 10317–10320.
- [157] P. R. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew. Chemie - Int. Ed.* **2009**, 48, 4052–5.
- [158] L. Wang, P. G. Schultz, *Chem. Biol.* **2001**, 8, 883–890.
- [159] M. J. Heuchel, Dissertation, Technische Universität Berlin, **n.d.**
- [160] T. Watanabe, N. Muranaka, I. Iijima, T. Hohsaka, *Biochem. Biophys. Res. Commun.* **2007**, 361, 794–799.
- [161] R. B. Quast, D. Mrusek, C. Hoffmeister, A. Sonnabend, S. Kubick, *FEBS Lett.* **2015**, 589, 1703–1712.
- [162] J. M. Kavran, S. Gundllapalli, P. O'Donoghue, M. Englert, D. Söll, T. A. Steitz, D. Soll, T. A. Steitz, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, 104, 11268–73.
- [163] K. Nozawa, P. O'Donoghue, S. Gundllapalli, Y. Arais, R. Ishitani, T. Umehara, D. Söll, O. Nureki, *Nature* **2008**, 457, 1163–1167.
- [164] T. Künzl, *Personal Correspondence*, **n.d.**
- [165] J. M. Chalker, Y. A. Lin, O. Boutureira, B. G. Davis, *Chem. Commun. (Camb)*. **2009**, 3714–6.
- [166] Y. A. Lin, J. M. Chalker, N. Floyd, G. J. L. Bernardes, B. G. Davis, *J. Am. Chem. Soc.* **2008**, 130, 9642–3.
- [167] S. H. Hong, I. Ntai, A. D. Haimovich, N. L. Kelleher, F. J. Isaacs, M. C. Jewett, *ACS Synth. Biol.* **2014**, 3, 398–409.
- [168] A. Ogawa, M. Hayami, S. Sando, Y. Aoyama, *J. Nucleic Acids* **2012**, 2012, 1–7.
- [169] A. J. Link, D. A. Tirrell, *J. Am. Chem. Soc.* **2003**, 125, 11164–11165.

References

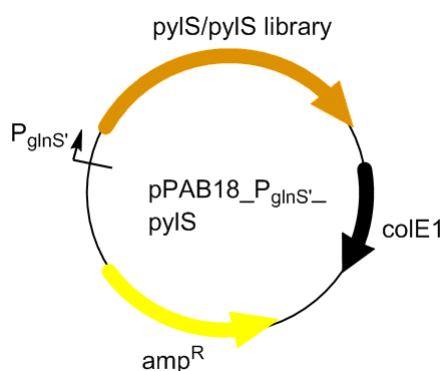
- [170] K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 19–24.
- [171] M. Albrecht, A. Lippach, M. P. Exner, J. Jerbi, M. Springborg, N. Budisa, G. Wenz, *Org. Biomol. Chem.* **2015**, *13*, 6728–6736.
- [172] P. F. Kruse, P. B. White, H. a Carter, *Cancer Res.* **1959**, *19*, 122–125.
- [173] M. H. Richmond, *Bacteriol. Rev.* **1962**, *26*, 398–420.
- [174] C. F. Schachtele, P. Rogers, *J. Mol. Biol.* **1965**, *14*, 474–489.
- [175] M. Pines, G. a Rosenthal, S. W. Applebaum, *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 5480–5483.
- [176] A. K. Bence, P. a Crooks, *J. Enzyme Inhib. Med. Chem.* **2003**, *18*, 383–394.
- [177] H. Kihara, J. M. Prescott, E. E. Snell, *J. Biol. Chem.* **1955**, *217*, 497–504.
- [178] Y. Ishida, J. H. Park, L. Mao, Y. Yamaguchi, M. Inouye, *J. Biol. Chem.* **2013**, *288*, 7564–7571.
- [179] J. Akaogi, T. Barker, Y. Kuroda, D. C. Nacionales, Y. Yamasaki, B. R. Stevens, W. H. Reeves, M. Satoh, *Autoimmun. Rev.* **2006**, *5*, 429–435.
- [180] Y. A. Lin, B. G. Davis, *Beilstein J. Org. Chem.* **2010**, *6*, 1219–1228.
- [181] J. Borlinghaus, F. Albrecht, M. Gruhlke, I. Nwachukwu, A. Slusarenko, *Molecules* **2014**, *19*, 12591–12618.
- [182] X. Xu, Y. Miao, J. Y. Chen, Q. Zhang, J. Wang, *J. Food Sci. Technol.* **2015**, *52*, 1724–1729.

Appendices

Plasmid sequences

For successful stop codon suppression with *MmPylRS*, three components are necessary: The orthogonal pair comprising pyrrolysyl-tRNA synthetase (gene *pylS*) and pyrrolysyl-tRNA (gene *pylT*), as well as the target protein gene with an internal amber stop codon. Plasmids are represented as schematic maps and important sequence parts are shown with highlighting (promoters and transcriptional terminators in lower case italics, protein genes in bold, tRNA genes underlined, and internal stop codons underlined and not bold, affinity tags in bold italics). If not stated otherwise, all plasmids used are derivatives of the pSEVA plasmid series. For selection and subsequent suppression tests, the following plasmids were used:

Library plasmid



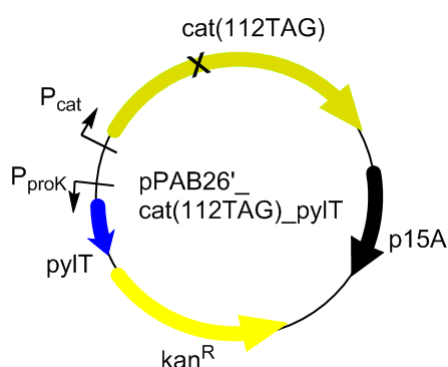
For presentation of the PylRS library, a high copy plasmid with ampicillin resistance was chosen. The PylRS gene was cloned under the control of a mutated *glnS* promoter and terminator to achieve high-level constitutive expression. The presence of *pylS* on the plasmid facilitated subsequent cloning of the *pylS* library, as an internal restriction site could be used. This allowed the cloning of a smaller fragment containing only the amino acid binding site.

Relevant sequence part (*pylS* with promoter/terminator):

```
CTAGTtcatcaatcatccccataatccttgtagattatcaattttaaaaaactaacagttgtcagcctgtcccgcttataatatcatacgccg
ttatacgttggttacgctttgaggaagaccCCTGCAGGATGGATAAAAAACCACTAAACACTCTGATATCTGCAACCGGG
CTCTGGATGTCCAGGACCGGAACAATTCATAAAATAAAACACCACGAAGTCTCTCGAAGCAAAATCTATATT
GAAATGGCATGCGGAGACCACCTTGTGTAAACAACCTCCAGGAGCAGCAGGACTGCAAGAGCGCTCAGGCA
CCACAAATACAGGAAGACCTGCAAACGCTGCAGGGTTTCGGATGAGGATCTCAATAAGTTCCTCACAAAGG
CAAACGAAGACCAGACAAGCGTAAAAGTCAAGGTCGTTTCTGCCCCTACCAGAACGAAAAAGGCAATGCCA
AAATCCGTTGCGAGAGCCCCGAAACCTCTTGAGAATACAGAAGCGGCACAGGCTCAACCTTCTGGATCTAAA
TTTTACCTGCGATACCGGTTTCCACCCAAGAGTCAGTTTCTGTCCCGGCATCTGTTTCAACATCAATATCAAG
CATTTCTACAGGAGCAACTGCATCCGCACTGGTAAAAGGGAATACGAACCCATTACATCCATGTCTGCCCCT
GTTTCAGGCAAGTGGCCCCGCACTTACGAAGAGCCAGACTGACAGGCTTGAAGTCTGTAAACCCAAAAGAT
GAGATTTCCCTGAATTCCGGCAAGCCTTTCAGGGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAGAC
CTGCAGCAGATCTACGCGGAAGAAAGGGAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTT
TGTGGACAGGGGTTTTCTGGAAATAAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGCAT
TGATAATGATACCGAACTTTCAAAACAGATCTTCAGGGTTGACAAGAACTTCTGCCTGAGACCCATGCTTGCT
```


Appendices

CCAAACCTTTACAACCTACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAAAATTTTGAATAGGCC
 CATGCTACAGAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGCTGAACTTCTGCCAGATG
 GGATCGGGATGCACACGGGAAAATCTTGAAAGCATAATTACGGACTTCTGAACACCTGGGAATTGATTTC
 AAGATCGTAGGCGATTCTGCATGGTCTATGGGGATACCCTTGATGTAATGCACGGAGACCTGGAACCTTCC
 TCTGCAGTAGTCGGACCCATACCGCTTGACCGGGAATGGGGTATTGATAAACCTGGATAGGGGCAGGTTT
 CGGGCTCGAACGCCTTCTAAAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTA
 CTATAACGGGATTTCTACCAACCTGTAACCTGCAGtttcaaacgctaaattgcctgatgcgctacgcttatcaggcctacatgatc
 tctgcaatatattgagtttgcgtgctttttagggccgataaaggcgttcacgccgcatccggcaagaacagcaaacatccaaaacgccg
 cggttcagcggcgctttttctgctttGGTACCGAG

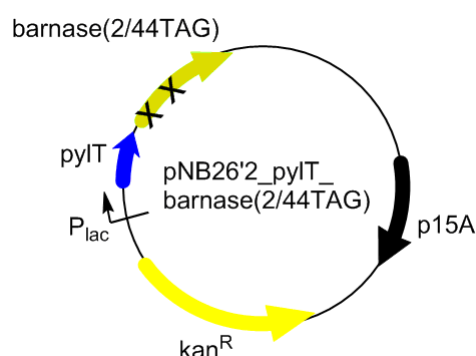
Positive selection plasmid

The positive selector chloramphenicol-acetyltransferase(112TAG) was installed on a low copy plasmid with kanamycin resistance under control of its own promoter. On the same plasmid, the orthogonal tRNA was cloned under the control of the proK tRNA promoter/terminator.

Relevant sequence part (pyIT/cat with promoters/terminators):

GTTTCATGAatgcaaaaaagcctgctcgttgagcaggcttttgaattTGGCGGAAACCCCGGGAATCTAACCCGGCTGAAC
 GGATTTAGAGTCCATTCGATCTACATGATCAGGTTTCCaatgcggggcgcatcttactgcgcagatacgcctcgtcaatccct
 taatagcaaaatgcctttgatcgcgagaaagtcagcaggccgcttagttagccTCATGATGATATATTTTATCTTGTGCAAT
 GTAACATCAGAGATTTTGAGACACAAATTTAAATCGTAATTATTGGGGACCCCTGGATTCTCACCAATAAAAAA
 CGCCCGGCGGCAACCGAGCGTTCTGAACAAATCCAGATGGAGTTCTGAGGTCATTACTGGATCTATCAACAGG
 AGTCCAAGACTAGTgatattctggcgaatgagacgttgatcggcacgtaagaggttccaactttcaccataatgaataagatcac
 taccggcgctatttttgagttatcgagatttcaggagctaaggaagctaaaATGGAGAAAAAATCACTGGATATACCACC
 GTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCACTCAGTTGCTCAATGTACCTATAACC
 AGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCT
 TTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTCGTATGGCAATGAAAGACGGTGAGCTGGT
 GATATGGGATAGTGTTACCCCTTGTTACACCGTTTTCCATGAGCAAACCTGAAACGTTTTCATCGCTCTGGAGT
 GAATACCACGACTAGTTCGGGCAGTTTCTACACATATATTGCAAGATGTGGCGTGTTACGGTGAAAACCTG
 GCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTACCAGTTT
 TGATTTAAACGTGGCCAATATGGACAACCTTCTCGCCCCGTTTTTACCATGGGCAAATATTATACGCAAGGC
 GACAAGGTGCTGATGCCGCTGGCGATTACAGTTCATCATGCCGTCTGTGATGGCTTCCATGTCCGGCAGAATG
 CTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAATTAATTAAAGGCATC

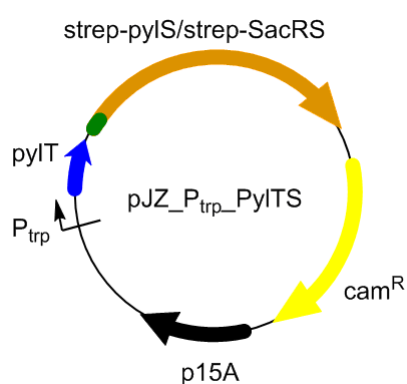
Appendices

Negative selection plasmid

To avoid false-negative selection due to the high toxicity of barnase, we chose a low-copy plasmid and limited the amount of tRNA in the cell by placing both pylT and barnase under the control of the same promoter to yield a single transcript to be processed into pylT and barnase-mRNA. The employed lac-promoter is constitutive in absence of lac-repressor, but IPTG was added to the medium regardless.

Relevant sequence part (pylT-barnase with promoter/terminator):

GGCACCCCAGGCTTTACACTttagcttccggctcgtatgttgtgtggaaTTGTGAGCGGATAACAATTCACACAGGAA
 ACAGCTATGACCATGATTACGAATTCGGAACCTGATCATGTAGATCGAATGGACTCTAAATCCGTTTCAGCCG
 GGTTAGATTCCCGGGGTTTCCGCCACTGCAGGTAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGGCATGCC
 TAGGTTTAACTAAGGAGGTTAATTAATGGCATAGGTTATCAACACGTTTGACGGGGTTGCGGATTATCTTCA
 GACATATCATAAGCTACCTGATAATTACATTACAAAATCAGAAGCACAAAGCCCTCGGCTGGGTGGCATCAAA
 AGGGAACCTTGCA¹TAGGTCGCTCCGGGGAAAAGCATCGGCGGAGACATCTTCTCAAACAGGGAAGGCAAA
 CTCCCGGGCAAAAGCGGACGAACATGGCGTGAAGCGGATTAATACTATACATCAGGCTTCAGAAATTCAGA
 CCGGATTCTTTACTCAAGCGACTGGCTGATTTACAAAACAACGGACCATTATCAGACCTTTACAAAAATCAGA
 TAAGCGGCCGCaggcatcaataaaacgaaggctcagtcgaaagactgggcctttcgttttatctgtgtgttcggtgaacgctctcc
 tgagtaggacaaatccgccGCCCTAGACAGCTGG

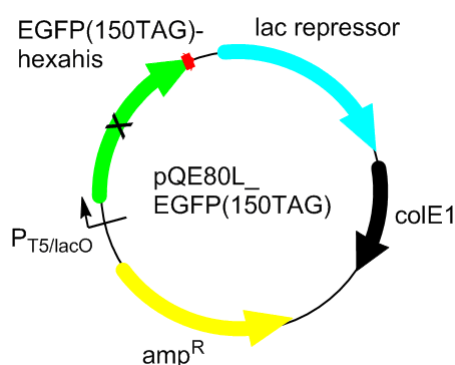
Suppression plasmid

The pJZ plasmid was recently used for efficient incorporation of noncanonical amino acids into proteins using a PylRS variant. The same setup was used for suppression of target genes with SacRS.

Relevant sequence part (pylT-pylS with promoter/terminator):

Appendices

CGCACCGGTGcttactccccatccccctgttgacaattaatcatcgaactagttaactagtacgcaggggcaGGTACCCGGGGATC
 CGGAAACCTGATCATGTAGATCGAATGGACTCTAAATCCGTTTACGCCGGGTTAGATTCCCGGGGTTTCCGCCA
 CCTAGGTTTAACTAAGGAGGTTAATTAAT**ATGAGCGCTTGGAGCCACCCGAGTTTCAAAAACTCGAGATGG**
ATAAAAAACCACTAAACACTCTGATATCTGCAACCGGGCTCTGGATGTCCAGGACCGGAACAATTCATAAAA
TAAACACCCACGAAGTCTCTCGAAGCAAAATCTATATTGAAATGGCATGCGGAGACCACCTTGTTGTAAACA
ACTCCAGGAGCAGCAGGACTGCAAGAGCGCTCAGGCACCACAAATACAGGAAGACCTGCAACGCTGCAG
GGTTTCGGATGAGGATCTCAATAAGTTCCTCACAAAGGCAAACGAAGACCAGACAAGCGTAAAGTCAAGG
TCGTTTCTGCCCCTACCAGAACGAAAAAGGCAATGCCAAAATCCGTTGCGAGAGCCCCGAAACCTCTTGAGA
ATACAGAAGCGGCACAGGCTCAACCTTCTGGATCTAAATTTTACCTGCGATACCGGTTTCCACCCAAGAGTC
AGTTTCTGTCCCGGCATCTGTTTCAACATCAATATCAAGCATTTCTACAGGAGCAACTGCATCCGCACTGGTA
AAAGGGAATACGAACCCATTACATCCATGTCTGCCCTGTTCAGGCAAGTGCCCCCGCACTTACGAAGAGC
CAGACTGACAGGCTTGAAGTCTGTTAAACCCAAAAGATGAGATTTCCCTGAATTCGGGCAAGCCTTTCAGG
GAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAGGGAGA
ATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTTCTGGAAATAAAATCCC
CGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGCATTGATAATGATACCGAACTTTCAAAACAGATCTT
CAGGGTTGACAAGAACTTCTGCCTGAGACCCATGCTTGCTCCAAACCTTTACAACTACCTGCGCAAGCTTGAC
AGGGCCCTGCCTGATCCAATAAAAAATTTTGAATAGGCCCATGCTACAGAAAAGAGTCCGACGGCAAAGA
ACACCTCGAAGAGTTTACCATGCTGAACCTTCTGCCAGATGGGATCGGGATGCACACGGGAAAATCTTGAAA
GCATAATTACGGAATTCCTGAACCACCTGGGAATTGATTCAAGATCGTAGGCGATTCTGCATGGTCTaTGG
GGATACCTTGATGTAATGCACGGAGACCTGGAACCTTCTCTGCAGTAGTCGGACCCATACCGCTTGACCG
GGAATGGGGTATTGATAAACCTGGATAGGGGCAGGTTTCGGGCTCGAACGCCTTCTAAAGGTTAAACACG
ACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTACTATAACGGGATTTCTACCAACCTGTAACTGC
 AGtttcaaacgctaaattgcctgatgcgctacgcttatcaggcctacatgatctctgcaatatattgagtttgctgcttttgtagccggata
 aggcgttcacgccgcatccggcaagaacagcaacaatccaaaacgccggttcagcggttttttctgctttGCGGCCGCAAGCT
 T

Target gene plasmid

The model target gene was cloned on a commercially available high-copy vector pQE80L (Qiagen, Hilden, D) with an additional plasmid-based lac repressor for inducible target protein production in any host strain.

Relevant sequence part (EGFP with promoter/terminator):

CTCGAGaaatcataaaaaatttttgccttgtagcggtataacaattataatagattcaattgtgagcggtataacaatttcacacaGAA
 TTCATTAAAGAGGAGAAATTA**ACTATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCTG**
GTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT

Appendices

ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCTGCCCTGGCCCACCCTCGTGACCAC
CCTGACCTACGGCGTGCACTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCC
ATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA
GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGC
AACATCCTGGGGCACAAGCTGGAGTACAACACAACAGCCACTAGGTCTATATCATGGCCGACAAGCAGAA
GAACGGCATCAAGGTGAAGTTCAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACT
ACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCACAACCACTACCTGAGCACCCAGTCCG
CCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATC
ACTCTCGGCATGGACGAGCTGTACAAGCATCACCATCACCATCACTAAGCTTAATTAGCTGAGcttgactcctgtt
gatagatccagtaatatgacctcagaactccatctggatttgttcagaacgctcggttgccgcccggcggtttttattggtgagaatccaagCT
AGCTTGCGC

Gene sequence of TTL(D221TAG)

ATGAGCGGATCCCAAAAGGCTGTTGAAATTACATATAACGGCAAACTTTAAGAGGAATGATGCATTTGCCT
GATGATGTTAAGGGTAAAGTGCCTATGGTAATAATGTTTCACGGTTTTACAGGCAATAAAGTAGAGTCTCAC
TTTATTTTTGTGAAGATGTCAAGAGCTTTAGAAAAAGTAGGTATTGGGAGTGTAAGGTTTGACTTTTATGGT
TCTGGAGAAAGTGATGGGGACTTTAGTGAAGTACATTTAGCAGTGAATTGGAAGATGCAAGACAAATTTT
AAAGTTTGTGAAGAGCAACCTACGACTGACCCTGAGAGAATAGGACTACTTGGTTTTGAGTATGGGAGGAG
CTATTGCAGGGATTGTAGCAAGGGAATATAAAGATGAAATAAAGGCGTTGGTGCTATGGGCTCCAGCTTTT
AATATGCCTGAGCTTATAATGAACGAAAGTGTAAGCAATACGGAGCTATTATGGAACAATTGGGCTTTGT
AGACATAGGAGGACATAAACTGAGTAAAGATTTTGTGAGGATATTTCAAAATTAATATATTGAGCTGTC
AAAAGGATACGATAAAAAAGTGCTTATAGTTTCATGGGACAAATGATGAAGCGTTGAATATAAAGTTTCTG
ATAGAATCTTAAAAGAGGTTTATGGGAAAAACGCTACAAGAGTGACAATCGAAAATGCAGACCATACTTTT
AAGAGTTTAGAATGGGAGAAAAAGGCGATTGAGGAGTCAGTAGAGTTTTTCAAAAAGGAATTGTTAAAGG
GAGGATCCGAGAACCTGTACTTCCAATCCGCCCATCACCATCACCATCACTGA

Gene sequence of barstar

ATGGTCCAATCTGCAGCCGTCCGAGACAGGAGGACATCGTCCAGCTGAAACCGGGGCGAGAATCCGGCCATT
TCTGAAGAGAAAAATGGTAACTGATAGAATAAAATCATAAGAAAGGAGCCGCACATGAAAAAGCAGTC
ATTAACGGGGAACAAATCAGAAGTATCAGCGACCTCCACCAGACATTGAAAAAGGAGCTTGCCCTTCCGGA
ATACTACGGTGAAAACCTGGACGCTTTATGGGATTGTCTGACCGGATGGGTGGAGTACCCGCTCGTTTTGGA
ATGGAGGCAGTTTGAACAAAGCAAGCAGCTGACTGAAAATGGCGCCGAGAGTGTGCTTCAGGTTTTCCGTG
AAGCGAAAGCGGAAGGCTGCGACATCACCATCATACTTTCTTAATACGATCAATGGGAGATGAACAATATG
GAAACACAAACCCTCAGCTATCCGATTGGAGAATATCAGCCGCTGAATCG

Appendices

Gene sequences***MmPylRS***

ATGGATAAAAAACCACTAAACACTCTGATATCTGCAACCGGGCTCTGGATGTCCAGGACCGGAACAATTCATAAAATAAAA
CACCACGAAGTCTCTCGAAGCAAAATCTATATTGAAATGGCATGCGGAGACCACCTTGTTGTAAACAACTCCAGGAGCAGC
AGGACTGCAAGAGCGCTCAGGCACCACAAATACAGGAAGACCTGCAAACGCTGCAGGGTTTCGGATGAGGATCTCAATA
AGTTCCTCACAAAGGCAAACGAAGACCAGACAAGCGTAAAAGTCAAGGTCGTTTCTGCCCTACCAGAACGAAAAAGGCA
ATGCCAAAATCCGTTGCGAGAGCCCCGAAACCTCTTGAGAATACAGAAGCGGCACAGGCTCAACCTTCTGGATCTAAATTT
TCACCTGCGATACCGGTTTCCACCCAAGAGTCAGTTTCTGTCCCGGCATCTGTTTCAACATCAATATCAAGCATTCTACAG
GAGCAACTGCATCCGCACTGGTAAAAGGGAATACGAACCCATTACATCCATGTCTGCCCTGTTTCAAGCAAGTGCCCCG
CACTTACGAAGAGCCAGACTGACAGGCTTGAAGTCTGTAAACCCAAAAGATGAGATTTCCCTGAATTCGGCAAGCCTT
TCAGGGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAGGGAGAATTAT
CTGGGGAAACTCGAGCGTGAAATTACCAGTTCTTTGTGGACAGGGGTTTTCTGGAAATAAAATCCCCGATCCTGATCCCT
CTTGAGTATATCGAAAGGATGGGCATTGATAATGATACCGAAGCTTTCAAACAGATCTTCAGGGTTGACAAGAACTTCTGC
CTGAGACCCATGCTTGCTCCAAACCTTTACAACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAAAATTTTGTG
AAATAGGCCCATGCTACAGAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGCTGAATTTCTGCCAGATG
GGATCGGGATGCACACGGGAAAAATCTTGAAAGCATAATTACGGACTTCTGAACACCTGGGAATTGATTTCAAGATCGT
AGGCGATTCTGCATGGTCTATGGGGATACCTTGATGTAATGCACGGAGACCTGGAAGTTTCTCTGCAGTAGTCGGACC
CATAACGCTTGACCGGAATGGGGTATTGATAAACCTGGATAGGGGCAGGTTTCGGGCTCGAACGCCTTCTAAAGGTTA
AACACGACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTACTATAACGGGATTTCTACCAACCTGTAA

SacRS

ATGGATAAAAAACCACTAAACACTCTGATATCTGCAACCGGGCTCTGGATGTCCAGGACCGGAACAATTCATAAAATAAAA
CACCACGAAGTCTCTCGAAGCAAAATCTATATTGAAATGGCATGCGGAGACCACCTTGTTGTAAACAACTCCAGGAGCAGC
AGGACTGCAAGAGCGCTCAGGCACCACAAATACAGGAAGACCTGCAAACGCTGCAGGGTTTCGGATGAGGATCTCAATA
AGTTCCTCACAAAGGCAAACGAAGACCAGACAAGCGTAAAAGTCAAGGTCGTTTCTGCCCTACCAGAACGAAAAAGGCA
ATGCCAAAATCCGTTGCGAGAGCCCCGAAACCTCTTGAGAATACAGAAGCGGCACAGGCTCAACCTTCTGGATCTAAATTT
TCACCTGCGATACCGGTTTCCACCCAAGAGTCAGTTTCTGTCCCGGCATCTGTTTCAACATCAATATCAAGCATTCTACAG
GAGCAACTGCATCCGCACTGGTAAAAGGGAATACGAACCCATTACATCCATGTCTGCCCTGTTTCAAGCAAGTGCCCCG
CACTTACGAAGAGCCAGACTGACAGGCTTGAAGTCTGTAAACCCAAAAGATGAGATTTCCCTGAATTCGGCAAGCCTT
TCAGGGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAGGGAGAATTAT
CTGGGGAAACTCGAGCGTGAAATTACCAGTTCTTTGTGGACAGGGGTTTTCTGGAAATAAAATCCCCGATCCTGATCCCT
CTTGAGTATATCGAAAGGATGGGCATTGATAATGATACCGAAGCTTTCAAACAGATCTTCAGGGTTGACAAGAACTTCTGC
CTGAGACCCATGCTTGCTCCAAACCTTTACAACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAAAATTTTGTG
AAATAGGCCCATGCTACAGAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGCTGAATTTCTGGCAGATG
GGATCGGGATGCACACGGGAAAAATCTTGAAAGCATAATTACGGACTTCTGAACACCTGGGAATTGATTTCAAGATCGT
AGGCGATTCTGCATGGTCTATGGGGATACCTTGATGTAATGCACGGAGACCTGGAAGTTTCTCTGCAGTTGTCTGGACC
CATAACGCTTGACCGGAATGGGGTATTGATAAACCTCGATAGGGGCAGGTTTCGGGCTCGAACGCCTTCTAAAGGTTA
AACACGACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTACTATAACGGGATTTCTACCAACCTGTAA

Protein sequences

MmPylRS

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARALRHHKYRKTCKRCRVSDE
DLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSPASVST
SISSISTGATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQQI
YAEERENYLGLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRPMLAPNLYNLRKLDRA
LPDPIKIFEIGPCYRKESDGKEHLEETMLNFCQMGSGCTRENLESIITDFLNHLGIDFKIVGDSCMVYGDTLDMHMG
DLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL

MmPylRS(C348A/Y384F)

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARALRHHKYRKTCKRCRVSDE
DLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSPASVST
SISSISTGATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQQI
YAEERENYLGLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRPMLAPNLYNLRKLDRA
LPDPIKIFEIGPCYRKESDGKEHLEETMLNFAQMGSCTRENLESIITDFLNHLGIDFKIVGDSCMVFGDTLDVMHMG
DLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL

SacRS

The consensus sequence after selection is C348W/W417S relative to *MmPylRS*:

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARALRHHKYRKTCKRCRVSDE
DLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSPASVST
SISSISTGATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQQI
YAEERENYLGLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRPMLAPNLYNLRKLDRA
LPDPIKIFEIGPCYRKESDGKEHLEETMLNFWQMGSCTRENLESIITDFLNHLGIDFKIVGDSCMVYGDTLDMHMG
GDLELSSAVVGPIPLDREWGIDKPSIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL

List of schemes and figures

Figure 1: Amino acids relevant for this study. Pyl: pyrrolysine (natural substrate); Bok: N_ε-butyloxycarbonyllysine (known excellent surrogate for Pyl); Zk: N_ε-CBZ-lysine (not incorporated in this work); Pok: N_ε-propargyloxycarbonyllysine; Pnk: N_ε-pentenoyllysine; Hnk: N_ε-heptenoyllysine; Nbk: N_ε-norbornyloxycarbonyllysine; Ayk: N_ε-acryloyllysine; Bct: biocytin; Nuk: N_ε-norbornylurea-lysine; Nap: N_δ-norbornyl-aminopentanoic acid; Cok: N_ε-Cyclooctynyloxycarbonyllysine; Edo: 2,7-diamino-oct-4-enedioic acid; Ops: O-pentenylserine; Npa: N-pentenyl-N-methyl-β-amino-alanine; Sac: S-allylcysteine; Sbc: S-butenylcysteine; Spc: S-pentenylcysteine; Aha: azidohomoalanine (methionine surrogate); Can: canavanine (arginine surrogate)..... 3

Figure 2: Postulation and experimental verification of the genetic code. A: The "Central Dogma" of molecular biology, as stated by Crick in 1956, postulated the unidirectional flow of sequential information from DNA via RNA to proteins, and an expanded modern version. The blue arrows indicate additional flow of sequential information, the red arrows indicate nonsequential informational flow not directly covered by the postulate, and the green arrows indicate external regulatory mechanisms regulating informational flow. B: To elucidate the assignment of codons, cell-free protein synthesis was primed with synthetic RNA, and the incorporation rate was determined for each amino acid. The combined results of several experiments allowed the deduction of codon assignments. In this example, the assignment UUU-->Phe can be directly determined, while elucidation of the assignments ACA-->Thr and CAC-->His requires two experiments with different RNA templates. 4

Figure 3: The genetic code. The consensus genetic code (DNA notation) used by the majority of organisms with semicanonical amino acids encoded by reassigned stop codons. 5

Figure 4: Protein translation: Aminoacyl-tRNA synthetases (aaRS) aminoacylate cognate tRNAs with the corresponding amino acids. Aminoacylated tRNAs bind to elongation factor EF-Tu (bacteria) or archaeal/eukaryotic elongation factors for trafficking to the ribosome. Aminoacylated tRNAs accommodate the A-site, recognizing the presented codon on the mRNA by base pairing. In the next step, the nascent protein chain is transferred from a tRNA located in the ribosome's P-site, elongating the chain by one amino acid. Finally, the ribosomal subunits briefly dissociate and migrate on the mRNA, ejecting the tRNA from the P-site and placing the tRNA formerly in the A-site, now carrying the nascent protein chain, in the P-site, freeing the A-site. A stop codon in the A-site triggers accommodation of a release factor instead of a tRNA, leading to dissociation of the ribosome and release of the protein chain (termination)..... 7

Figure 5: Aminoacylation of tRNAs. A: Schematic representation of the aminoacylation process. The three major steps are the activation of the amino acid using ATP, the tRNA binding and aminoacylation and the release of aminoacylated tRNA. Each of these steps can be rate-limiting. B: Structure of a homodimeric class 2 LysRS (from PDB-ID 1E24). C: Structure of a monomeric class 1 LysRS (from PDB-ID 1IRX). D: Schematic representation of a tRNA^{Lys}, revealing the cloverleaf structure and a high ratio of modified nucleotides^[6]. Unusual nucleotides are thymine (T), dihydrouridine (D), pseudouridine (Ψ), 7-methylguanosine (m⁷G), threonylcarbamoyladenosine (t⁶A), 3-(3-amino-3-carboxypropyl)uridine (X), 5-methoxyaminomethyl-2-thiouridine (mam⁵-2^s-U). E: Structure of tRNA^{Lys} (from PDB-ID 1FIR)..... 8

Figure 6: Incorporation of noncanonical amino acids into target proteins. A: Scheme of aminoacylation reactions. Top row: natural aminoacylation process with a canonical aaRS/tRNA pair charging the cognate amino acid. Center row: in the absence of a cognate amino acid, an isostructural surrogate can be charged via a canonical aaRS/tRNA pair. Bottom row: a non-isostructural noncanonical amino acid is not accepted by the canonical aminoacylation machinery and requires a distinct additional aaRS/tRNA pair. This aminoacylation does not interfere with canonical aaRSs, tRNAs and amino acids, and is thus orthogonal..... 9

Figure 7: Some noncanonical amino acids. A: Selection of surrogate amino acids of methionine, arginine and tryptophan, which have been successfully incorporated into proteins, replacing their corresponding canonical amino acid. B: Noncanonical amino acids that are not accepted by any known canonical aaRS/tRNA pair, which have been incorporated into proteins using native or engineered variants of pyrrolysyl-tRNA synthetase and tRNA^{Pyl}..... 10

Figure 8 Phylogenetic distribution of pyrrolysyl-tRNA synthetase. The tree is based on aligned sequences obtained by BLAST (basic local alignment search tool^[39]), using the *Methanosarcina mazei* PylRS gene sequence as query. PylRS is a rare enzyme, appearing almost uniquely in methanogenic organisms. The tree shows a distinction between an archaeal type and a bacterial type. The latter shows two separate domains, the catalytic C-terminal domain and the omissible N-terminal domain, while in the archaeal type the domains are connected. 12

Figure 9: Stop codon suppression versus termination. A: A stop codon resting in the ribosomal A-site usually triggers binding of a release factor and dissociation of the protein biosynthesis complex. Alternatively, cytosolic aminoacylated suppressor tRNAs can decode the stop codon analogously to decoding of a sense codon, thus suppressing the termination event. In this case, protein biosynthesis continues until a non-suppressed stop codon (or the end of the mRNA) is reached. The schematic example shows the competition for an amber stop codon (UAG) between a charged amber suppressor tRNA and the bacterial release factor RF1, that triggers translation termination at amber codons. The ratio of suppression and termination event largely depends on the binding affinity of the aminoacylated suppressor tRNA to EF-Tu and to the

Appendices

ribosome. B: Structure of bacterial release factor RF1 (from PDB-ID 1RQ0). The overall protein structure mimics the L-shaped tRNA fold. C: Structure of tRNA^{Sec} (from PDB-ID 3LOU) and the specialized elongation factor EF-Sec (from PDB-ID 4ZU9) required for trafficking of the aminoacylated tRNA to the ribosome. tRNA^{Sec} is initially aminoacylated with serine by SerRS, which is converted to Sec on-tRNA. Delivery to the ribosome is mediated by EF-Sec and a downstream mRNA recognition element (SECIS). D: Homodimeric pyrrolysyl-tRNA synthetase with two molecules of tRNA^{Pyl} (from PDB-ID 2ZNI). PylRS and tRNA^{Pyl} constitute a system for suppression of amber stop codons independent of specialized RNA structures or proteins. 14

Figure 10: Engineering of orthogonal aaRS. A: Structures of the most commonly used orthogonal aaRS, *MmPylRS* (left, from PDB-ID 2ZIM) and *MjTyrRS* (right, from PDB-ID 1J1U) shown as monomers. The amino acid binding pockets are shown as green sticks, the bound amino acid substrates are shown in orange. B: Schematic selection procedure relying on positive and negative selection. For positive selection, the aaRS library is co-transformed with an amber-disrupted essential gene (positive selector). In the presence of the noncanonical amino acid of interest, successful suppression allows cell growth. The isolated enriched library is then co-transformed with an amber-disrupted toxic gene. In the absence of the desired ncaa, only aaRS variants that accept natural amino acids allow amber suppression, inactivating the transformants. The two selection steps are iteratively cycled to enrich the desired aaRS variants with aminoacylation activity for the ncaa of interest. 17

Figure 11: Biosynthetic pathways to noncanonical amino acids. Top: Synthesis of pyrrolysine from two lysines via the action of methylornithine synthase (pylB), methylornithine-lysine pseudopeptide synthase (pylC) and pyrrolysine synthase (pylD), synthesis of pyrroline-carboxyllysine from lysine and ornithine, omitting the pylB-catalyzed rearrangement step and synthesis of 3S-ethinylpyrrolysine from 3S-ethinylornithine and lysine. Bottom: Biosynthesis of methionine via homocysteine generated from direct sulfhydrylation of O-acetylhomoserine by O-acetylhomoserine sulfhydrylase (OAHSS). OAHSS accepts a wide range of nucleophiles as substrate, enabling the biosynthesis of diverse homoalanine derivatives including azidohomoalanine. The Cys biosynthesis via sulfhydrylation of O-acetylserine by O-acetylserine sulfhydrylase (OASS) can be analogously exploited as a biosynthetic route to alanine derivatives. 20

Figure 12: Posttranslational chemical reactions at noncanonical amino acid with alkene and alkyne moieties (schematic representations). Clockwise, starting top left: tetrazine-click reaction (a Diels-Alder variant), tetrazole photoclick reaction, radical thiol-ene conjugation, ruthenium-catalyzed olefin metathesis (at alkenes); Sonogashira coupling, Huisgen-Sharpley cycloaddition, radical thiol-yne reaction, Glaser-Hay reaction (at alkynes). See text for references. Some reactions require activated groups or restrictive reaction conditions. 22

Figure 13: Model proteins used in this study. A: Enhanced green fluorescent protein (EGFP) is an autofluorescent protein often used as reporter (from PDB-ID 2YOG). B: *Thermoanaerobacter thermohydrosulfuricus* lipase (TTL, see reference^[124]) is a model enzyme already probed for acceptance of noncanonical amino acids. C: Barstar is a small protein often employed in folding studies (from PDB-ID 1BTA). 23

Figure 14: Expression of EYFP(114Bok). Left: Full length protein could be purified only after addition of 2 mM Bok. Right: Incorporation of Bok could be confirmed by mass spectrometry (calculated mass: 29345.5 Da; found mass: 29344.7 Da)... 25

Figure 15: Expression of EGFP. A: Western blot detection of *MmPylRS* and EGFP in two different expression batches. Full-length EGFP(Bok) can only be detected in one batch, illustrating the lack of reliability of the expression system. B: UV fluorescence (top) and normal (bottom) image of EGFP(150Bok), EGFP(150Arg) expressed for control purposes employing the rare arginine codon AGG, and EGFP after hexahistidin tag purification (with co-purification of *MmPylRS*), after storage at 4 °C over night. All samples show significant precipitation. A control sample of pure EGFP after storage at 4°C for >7d was added to illustrate the stability of the fluorescent protein. C: SDS-PAGE analysis of precipitated EGFPs, with pellet material (P), solubilized pellet material (R) and supernatant (S) applied to the gel. The pellet of EGFP(Bok) shows only a band corresponding to the mass of EGFP, while the pellet of EGFP(150Arg) shows co-precipitation with *MmPylRS*. 26

Figure 16: Analysis of EGFP(150Bock). EGFP(150amber) was co-induced with *MmPylRS*/*MmtRNA*^{Pyl} from plasmid pMEc1μC-prp-strep-PylST in the presence of 2 mM Bock. A: SDS-PAGE of EGFP(150Bock) purified by Ni-affinity chromatography. The eluate shows a band with the same apparent molecular weight as EGFP control. B: Deconvoluted mass spectrum shows a peak at 27857 Da, consistent (±1 Da) with EGFP(150Bock) (calculated: 27858)... 27

Figure 17: Suppression of EGFP(150TAG) and purification of EGFP(150Bok). EGFP was expressed in the presence of 2 mM Bok in cells carrying plasmid pMEc1μC-prp-strep-PylST-glnS'. A: Western blot of EGFP suppression with Bok shows a band at the expected position. 1 μg of wild-type EGFP is used as a blot control. B: Affinity purification of EGFP(150Bok). EGFP is the main portion of eluate protein content with several impurities. C: Deconvoluted mass spectrum shows a peak at 27858 Da, consistent with EGFP(150Bok). 27

Figure 18: Expression tests with the pJZ vector system (adapted from Al Toma *et al.*^[127]). EGFP(150TAG) was suppressed with *MmPylRS* or *MmPylRS*(Y384F). The *MmPylRS* band at 52 kDa (black arrow) is visible in all induced samples, indicating stable expression. The full-length EGFP at 27.8 kDa (green arrow) is faintly visible only in wild-type controls and samples with Bok added after induction, and more pronounced for suppression with *MmPylRS*(Y384F). 28

Appendices

- Figure 19: Growth curves of strain BL21Gold(DE3) in LB medium. Control: no aa; Edo/Ayk/Bok: 1 mM of respective amino acid added. Curves were normalized to the initial cell density and measured in duplicate..... 29
- Figure 20: Incorporation tests with wild-type *MmPylRS* (partially adapted from Al Toma *et al.* ^[127]). Two suppression systems were tested with the pJZ plasmid system showing superior performance. Additionally, the efficiency of the Y384F variant was assessed. Incorporation was found for Bok, Pok, Nbk (only residual without Y384F) and Pnk (residual). 29
- Figure 21: Deconvoluted mass spectra of EGFP(150TAG) suppressed with different amino acids. Calculated masses are 27744 Da (EGFP(WT), N150); 27858 Da (EGFP(Bok)); 27895 Da (EGFP(Nbk)); 27841 (EGFP(Pnk)). The high noise level of EGFP(Pnk) is due to impurities and low amount of protein. No mass spectrum was measured for EGFP(Pok). 30
- Figure 22: Detection of full-length EGFP (27 kDa) suppressed with Bok from cell lysates. Cells were harvested 2, 4 and 24 hours after induction with 1 mM IPTG and 2 mM Bok. Yields were determined after purification..... 30
- Figure 23: Test suppressions with novel orthogonal systems *MePylRS/Me-tRNA^{Pyl}* and *TpPylRS/Tp-tRNA^{Pyl}*. Full-length EGFP(Bok) was detected with *MePylRS* (left blot), but the finding could not be reproduced (right blot). No activity was found for *TpPylRS*. 31
- Figure 24: *In vitro* suppression: A: Purification of strep-tagged *MmPylRS*. B: Analysis of *in vitro* transcribed *Mm-tRNA^{Pyl}*. C: Purification of *in vitro* expressed EGFP(wt) (left) and suppressed EGFP(150Bok) (right). Only EGFP(wt) was found. Lane labels: Lysate (L), flow through (FT), wash fractions (W), eluate fractions (E). 31
- Figure 25: *In vitro* suppression with an advanced cell-free expression system (single experiment, data generated by E. Worst^[148]). A: Normalized EGFP-fluorescence after 16 h of cell-free expression. Expression of EGFP(WT) leads to increased fluorescence dependent on the plasmid concentration, while expression batches for EGFP(150amber) do not show significantly elevated fluorescence levels. This indicates a tight system with minimal background suppression. B: Normalized EGFP-fluorescence in cell-free expression reactions of EGFP(150amber) with pre-aminoacylated acetyllysine-tRNA_{CUA}. Nearly quantitative production of EGFP(150acetyllysine) is observed at 1.5 μM of tRNA, with a maximum at 3 μM tRNA. C: Cell-free suppression of EGFP(150amber) in the presence of 2 mM Bok. EGFP(150amber), *MmPylRS* and T7 RNA polymerase are expressed *in vitro* using the *E. coli* transcription/translation machinery and the optimized promoter OR2-OR1-PR. tRNA^{Pyl} is transcribed from linear DNA under the control of a T7 promoter, employing the highly efficient T7 transcription. An increase in fluorescence of 36% compared to a negative without templates for tRNA^{Pyl}, *MmPylRS* and T7 RNA polymerase could be observed. Additional controls omitting one or two components showed moderate increases of EGFP fluorescence. 32
- Figure 26: Rational mutagenesis of the amino acid binding pocket of *MmPylRS*. A: The amino acid binding pocket of *MmPylRS*. Four amino acids lining the pocket have been highlighted. These residues limit the size of the binding pocket and are frequently altered in mutants accepting larger amino acids. B: Mutagenesis scheme for *PylRS* variants with rationally enlarged binding pockets. Starting from the highly active variant *MmPylRS*(Y384F), the amino acids Y306, L309 and C348 have been exchanged by alanine in all combinations. Additionally, wild-type *MmPylRS* and the Y306A variant of wt-*PylRS* have been used. 33
- Figure 27: Test incorporation with *AykRS* and diverse amino acids. Good suppression was found only for addition of Ayk, bands for other amino acids are results of artifacts. wt: wild-type EGFP; sup: EGFP(150TAG); aa: 2 mM of denoted noncanonical amino acid added..... 34
- Figure 28: Incorporation of Cok with *PylRS_AF*. A: Western blot detection of full-length EGFP in the presence of Cok. B: Purification of EGFP(150Cok). Flow-through (FT), wash (W) and eluate (E1-E3) fractions have been applied to the gel. EGFP(150Cok) is found in fraction E2. C: Mass analysis of EGFP(Cok) (expected mass 27909 Da). The measured protein mass matches EGFP with incorporated Glu, Gln or Lys (expected masses 27759.5 Da, 27758.5 Da, or 277758.5 Da, respectively).35
- Figure 29: Screening of *PylRS* variants for incorporation of Bok and Bct by western blot analysis of raw cell lysate. While most variants accept Bok (only *MmPylRS*(Y306A/L309A/C348A/Y384F) is not active), Bct is not incorporated by any variant. 35
- Figure 30: Screening of *PylRS* variants for incorporation of Hnk and Pnk. A: Western blot detection of full-length EGFP after suppression with Hnk or Pnk. Hnk is a substrate for *PylRS_AF* (*MmPylRS*(Y306A/Y384F), Pnk is incorporated by several variants with *MmPylRS*(C348A/Y384F) seemingly being the most efficient. B: Purification of EGFP(Hnk) and EGFP(Pnk) incorporated with different synthetases. C: Mass analysis of EGFP(Hnk) (expected mass: 27868.5 Da; found mass: 27869 Da) and EGFP(Pnk) (expected mass: 27840.5 Da; found mass: 27840.5 Da). An additional Peak at 27759 Da may correspond to EGFP with incorporated Glu, Gln or Lys (expected masses 27759.5 Da, 27758.5 Da, or 277758.5 Da, respectively). 36
- Figure 31: Incorporation of Pnk in TTL(D221TAG). A: SDS-PAGE analysis of TTL expression shows good production of TTL(150Pnk). The eluate fraction contains significant impurities, which do not interfere with subsequent labeling processes. B: Confirmation of incorporation by MS (expected: 30280 Da; found: 30279.67 Da)..... 37
- Figure 32: Positions of randomizations of libraries B (left) and S (right)..... 38
- Figure 33: Rationale for library design. Due to structural differences between Bct and Sac, the corresponding libraries have to be designed differently. For Bct screening, a *PylRS* library with randomizations of bulky side chains and the fixed

Appendices

Y384F mutation was devised, containing candidates with enlarged binding pockets to accommodate the bulky Bct. This library has successfully been used for incorporation of a large cyclooctyne-containing amino acid. For Sac, on the other hand, the randomizations concentrate on the lining of the front to middle part of the amino acid binding pocket and include a randomization of the gate keeper position N346, which may otherwise fix substrate specificity as lysine derivatives. This library (equivalent positions in *MbPylRS*) has been successfully screened for a PylRS activating a photocaged cysteine, which, while largely different in steric and electronic properties, nevertheless shares substructures with Sac. 38

Figure 34: PCR scheme for library generation. Two PCRs are carried out using one flanking and one randomization primer, producing an upstream (green) and downstream (blue) fragment. The fragments have homologous regions due to the use of complementary randomization primers, and can be joined via overlap extension PCR with flanking primers. The resulting full-length gene contains the randomization and can be used as PCR template to generate upstream/downstream fragments with more randomizations. When all randomizations have been introduced, the final full-length product is cloned into an expression plasmid. 40

Figure 35: Construction of Library S. Shown are iterative PCR cycles to generate randomized upstream and downstream fragments containing the randomizations and overlapping sequences (left row) and the formation of full-length randomized genes via overlap extension PCR of the fragments (right row). Five randomized positions were introduced in three cycles using the product of the latest combination PCR as template. 40

Figure 36: Sequencing of libraries B and S. A, C, G and T are represented as green, blue, black and red lines, respectively. All expected randomizations are found. 41

Figure 37: Positive selection system test with cat(98/181TAG) and *MmPylRS*. Cells were spotted on media containing increasing concentrations of chloramphenicol. In the presence of Bok, cells grew at up to $150 \mu\text{g mL}^{-1}$ cam within three days at 30°C . No growth was observed in the absence of Bok. 42

Figure 38: Test positive selection with cat(98/181TAG) from library B for activation of Bok. Top: After 3 d, a single colony was obtained, which, after plasmid isolation, showed an overlay of signal for bidirectional sequencing. Bottom: After removal of the positive selector plasmid and retransformation of the obtained PylRS plasmid, all six tested colonies showed wild-type sequence with the retained fixed mutation Y384F (TTT \rightarrow TAT), as expected. 43

Figure 39: Negative selection system test with barnase and the BpaRS system (data generated by M. Heuchel^[159]). Dilutions of an overnight culture were spotted on medium with or without the cognate noncanonical amino acid benzoylphenylalanine (Bpa). In the presence of Bpa, cell growth is impaired by a factor of 10^2 to 10^3 43

Figure 40: Negative selection scheme with tdk as selector. tdk is an enzyme involved in the TTP production pathway, catalyzing the monophosphorylation of thymidine to TMP. The enzyme is nonessential due to thyA providing an alternative pathway to TMP. While not inherently toxic, tdk cannot discriminate against AZT and is a crucial enzyme for the production of toxic AZT triphosphate from AZT. This conditional toxicity together with its other properties makes tdk a promising candidate for a negative selection marker. 44

Figure 41: Negative selection system tests with tdk and PylRS. Several combinations of stop codons were tested for growth impedance in the presence of AZT and Bok in triplicate. While the combinations 2*/R29TAG, 2*/Q79TAG and R29/Q29TAG looked promising, the results could not yet be confirmed. *: The stop codon has been inserted at this position after the start codon. Each experiment was performed in triplicate by streaking three independent clones on the same test plate. 45

Figure 42: Plates of final positive selection step for Sac. After two positive and negative selection steps, cells showed a significant growth difference, growing to a meadow in the presence of Sac. This indicates a successful selection. 46

Figure 43: Peer-group tests. A: Sac-dependent growth of *E. coli* NEB10 β transformed with a selector plasmid with cat(98/181TAG) and the *MmPylRS* library after 5 selection steps. 11 out of 12 colonies show chloramphenicol survival dependent on the presence of Sac. Clones 2, 3, 5 and 9-12 have the genotype *MmPylRS*(C348W/W417S) (SacRS). B: Peer-group sequencing of 8 clones of the final enriched library after 5 selection steps. Six of eight clones show the consensus sequence C348W/W417S, one of those has not fully defined positions 346 and 401, probably due to multiple plasmid transformation into the same cell. 46

Figure 44: Confirmation of SacRS activity and specificity test. A: SDS-PAGE of purified EGFP without internal stop codon (WT), EGFP(150TAG) suppressed without addition of noncanonical amino acid (no AA) and EGFP(150TAG) suppressed with Sac to yield EGFP(150Sac) (Sac). B: Deconvoluted HPLC-ESI mass spectra of purified proteins. Expected masses are 27744.3 for EGFP(WT), 27758.4, 27759.3 or 277758.3 for suppression with Lys, Glu or Gln in absence of Sac and 27773.3 for EGFP(150Sac). C: Western blot analysis of suppression test lysates. Lysate from the same amount of cells was loaded on each lane. Note that the apparent difference in migration length between both blots is due to the fact that two different gel percentages (15 % and 20 %) were used. WT: EGFP without internal stop codon; sup: EGFP(150TAG); ctrl: purified EGFP. EGFP was expressed with either SacRS or *MmPylRS* in the presence of different amino acids (5 mM) in LB medium. The arrows indicate expected bands. All induced samples show a faint band due to background suppression of the used strain (C321. Δ A.exp). 47

Appendices

Figure 45: Incorporation of Sac into TTL and removal of the hexahistidin-tag: A: Monitoring of his-tag removal via SDS-PAGE. Exp: Expressed TTL(D221Sac) after affinity purification; FT: column flow-through after his-tag removal (contains the modified protein); eluate: column eluate after his-tag removal. The TTL gene contains a Met codon (UAG) at position 80, which can act as a secondary translational start codon, leading to a truncated product missing the amino acids at positions 1-79. 49

Figure 46: Modification scheme for posttranslationally modify barstar via Huisgen's cycloaddition. Conjugation of barstar(1Aha) and ethynylated bodipy via copper-induced [3+2] cycloaddition leads to a blue-shift in bodipy fluorescence. 51

Figure 47: Purification and characterization of barstar(Aha) conjugated with ethynylated bodipy (partially adapted from [171]). A: Expression of Aha- ψ -b*. The induced culture shows a distinct band at the expected migration length (M(Aha- ψ -b*) = 10247.9 Da, indicated by the arrow). B: SDS-PAGE gel of ψ -b* (3 μ g), Aha- ψ -b* (5 μ g) and bodipy-Aha- ψ -b* (0.75 μ g). Bands were visualized at 365 nm (right) and with Coomassie staining (left). Only the band corresponding to bodipy-Aha- ψ -b* exhibits fluorescence. C: Deconvoluted ESI-MS spectrum of Aha- ψ -b* and bodipy-Aha- ψ -b*. The mass spectrum of Aha- ψ -b* shows a single peak at 10246.8 Da (theoretical mass 10247.9 Da), the spectrum of bodipy-Aha- ψ -b* exhibits a pronounced peak at 10518 Da (theoretical mass 10520 Da). D: Thermal denaturation curves of Aha- ψ -b* and bodipy-Aha- ψ -b*. Circular dichroism was monitored at 222 nm (characteristic minimum for α -helices). Melting points were calculated to be 62.2 °C for Aha- ψ -b* and 55.2 °C for bodipy-Aha- ψ -b*. E: Absorption spectra of unconjugated bodipy, ψ -b*(Aha) and conjugated Barstar(Aha bodipy). Both bodipy-Aha- ψ -b* and Aha- ψ -b* show a peak at 280 nm, as expected for proteins. bodipy-Aha- ψ -b* however, shows an additional peak at 508 nm, which is in correspondence to conjugated bodipy and clearly differs from the spectrum of unconjugated bodipy, which shows a maximum at 537 nm. F: Fluorescence emission spectra of unconjugated bodipy, Aha- ψ -b* and conjugated bodipy-Aha- ψ -b*. Fluorescence was determined with an excitation wavelength of 500 nm. A background spectrum of Tris buffer pH 8 is also shown. Aha- ψ -b* shows minimal emission signal due to light scattering. bodipy-Aha- ψ -b* exhibits a maximum at the expected wavelength (522 nm) that is blue-shifted from the emission of unconjugated bodipy at 554 nm. 52

Figure 48: Replacement of Arg by Can in EGFP *in vitro* and *in vivo* (partially adapted from [150]). A: Distribution of arginine residues (highlighted in magenta) in a protein model EGFP (structure from PDB-ID 2YOG). All residues are located in β -sheets. Five side chains face outwards, and only one is in contact with the fluorophore, localized in the barrel structure. Also shown are chemical structures of arginine and canavanine. B: Gel of purified GFP model proteins. Left: *In vivo* expressed EGFP(WT) and EGFP(Can) after affinity and size exclusion chromatography; right: *In vitro* expressed dEGFP(WT) and dEGFP(Can) after affinity chromatography. C: Confirmation of full incorporation of Can by ESI-MS; expected masses are 28333 Da (*in vivo*) and 26204 Da (*in vitro*). The mass difference of 1.5 Da for dEGFP(Can) is within the error of spectrum deconvolution. D: Typical yields of EGFP(Can) (*in vivo*) and dEGFP(Can) (*in vitro*). *In vitro* yield is a prediction based on cell-free expression using 90 μ L of cell extract derived from approximately 50 mL cell culture. E: Absorption and fluorescence spectra of dEGFP(WT) and dEGFP(Can). Five spectra were accumulated for each sample and normalized. After the incorporation of canavanine, the absorption maximum is slightly red-shifted (491 nm vs. 489 nm) with unchanged emission maximum at 508 nm. 53

Figure 49: Finalized plasmids used in this work: Top row: general suppression plasmid (variable pylS variant); pylS library expression plasmid. Bottom row: general expression plasmid (variable target gene), positive selection plasmid (variable selector gene), negative selection plasmid (variable selector gene). 62

List of tables

Table 1: Properties of class 1 and class 2 aminoacyl-tRNA synthetases.	8
Table 2: Comparison of incorporation techniques.....	15
Table 3: Peer-group sequencing of six clones from a transformation of library B. Some undefined positions were expected due to transformation of multiple library members in the same cell. All sequenced clones show different genotypes and no bias towards the wild-type amino acids.	41

Danksagung

Ich habe bei der Erstellung dieser Arbeit sehr viel Unterstützung erhalten von zahlreichen Personen, bei denen ich mich an dieser Stelle bedanken möchte:

Bei Professor Dr. Nediljko Budisa für die langjährige hervorragende Betreuung, die Zusammenarbeit, das Gutachten und die vielen guten Ratschläge, für die Zeit, die er sich immer für mich genommen hat und dafür, dass er mir eine ständige Quelle der Inspiration war.

Bei Professor Dr. Zoya Ignatova, für die Übernahme des Zweitgutachtens.

Bei meinen Kooperationspartnern an der TU Berlin, der Universität des Saarlandes und des METACODE-Konsortiums für die gute und in vielen Fällen erfolgreiche Zusammenarbeit.

Bei meinen Diplomanden Juliane Ziegler, Zheni Palyancheva und Clemens Zaengel und meinen Forschungsstudenten für die hervorragende Arbeit, ohne die viele Projekte nicht hätten realisiert werden können.

Bei der Chemieabteilung der Arbeitsgruppe Biokatalyse für die Versorgung mit Aminosäuren.

Bei Michael Hösl, Patrick Durkin, Jelena Jaric und Tobias Baumann für die angeregten Diskussionen, bei Waltraud Wenger und Christin Treiber für die hervorragende technische Unterstützung und erneut bei Patrick Durkin und Tobias Baumann für das kritische Lesen dieses Manuskripts.

Bei der gesamten Arbeitsgruppe Biokatalyse für das nette Arbeitsklima und die tolle Atmosphäre. Ich hätte es auch schlimmer treffen können!

Bei meiner Familie, die mich auf meinem Weg nach Kräften unterstützt hat.

Bei allen, die mir geholfen haben und noch nicht aufgeführt sind.