TECHNISCHE UNIVERSITÄT MÜNCHEN

Max-Planck-Institut für Biochemie

Expanding the Toolkit of Protein Engineering:
Towards Multiple Simultaneous *In Vivo* Incorporation of Noncanonical Amino Acids

Michael G. Hösl

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. Michael Groll
Prüfer der Dissertation:	1. Univ.-Prof. Dr. Nediljko Budisa, Technische Universität Berlin
	2. Univ.-Prof. Dr. Thomas Kiefhaber
	3. Univ.-Prof. Dr. Johannes Buchner

Die Dissertation wurde am 01.02.11 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 03.03.11 angenommen.

To Teresa Ariadna García-Grajalva Lucas
who influenced the idea of TAG → AGG
switch just by the existence of her name

Sleeping is giving in,
no matter what the time is.
Sleeping is giving in,
so lift those heavy eyelids.

People say that you'll die
faster than without water.
But we know it's just a lie,
scare your son, scare your daughter.

People say that your dreams
are the only things that save ya.
Come on baby in our dreams,
we can live on misbehavior.

The Arcade Fire

Parts of this work were published as listed below:

Hoesl, MG, Budisa, N. Expanding and engineering the genetic code in a single expression experiment. *ChemBioChem* **2011**, *12*, 552-555.

Further publications:

Hoesl, MG*, Staudt, H*, Dreuw, A, Budisa, N, Grininger, M, Oesterhelt, D, Wachtveitl, J. Manipulating the eletron transfer in Dodecin by isostructual noncanonical Trp analogs. **2011**, [in preparation]. *authors contributed equally to this work

Nehring, S*, **Hoesl, MG***, Acevedo-Rocha, CG*, Royter, M, Wolschner, C, Wiltschi, B, Budisa, N, Antranikian, G. Effects of additives on the activity of TTL congeners. **2011**, [in preparation]. *authors contributed equally to this work

Hoesl, MG*, Acevedo-Rocha, CG*, Nehring, S*, Royter, M, Wolschner, C, Wiltschi, B, Budisa, N, Antranikian, G. Lipase Congeners Designed by Genetic Code Engineering. *ChemCatChem* **2011**, *3*, 213-221. *authors contributed equally to this work

Hoesl, MG, Larregola, M, Cui, H, Budisa, N. Azatryptophans as tools to study polarity requirements for folding of green fluorescent protein. *J. Pept. Sci.* **2010**, *16*, 589-595.

Merkel, L, **Hoesl, MG**, Albrecht, M, Schmidt, A, Budisa, N. Blue Fluorescent Amino Acids As In Vivo Building Blocks for Proteins. *ChemBioChem* **2010**, *11*, 305-314.

Lepthien, S, **Hoesl, MG**, Merkel, L, Budisa, N. Azatryptophans endow proteins with intrinsic blue fluorescence. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 16095-16100.

Review articles:

Hoesl, MG, Merkel, L, Budisa, N. Synthetic biology of autofluorescent proteins, in *Fluorescent Proteins – from Fundamental Research to Bioanalytics* (Ed.: G. Jung), Springer Verlag, Heidelberg, **2011**, p. [in press].

Hoesl, MG, Budisa, N. In vivo incorporation of multiple non-canonical amino acids into proteins. *Angew. Chem. Int. Ed. Engl.* **2011**, [epub, DOI: 10.1002/anie.201005680].

Hoesl, MG, Budisa, N. Nicht-kanonische Aminosäuren in der Synthetischen Biologie. *BIOspektrum* **2010**, *16. Jahrgang*, 309-311.

Poster presentations:

Hoesl, MG, Lepthien, S, Merkel, L, Budisa, N, (2009). Azatryptophans endow proteins with intrinsic blue fluorescence. 3rd European Conference on Chemistry for Life Sciences, Sep 2 - 5, **2009**, Frankfurt/Main, Germany.

Oral presentations:

Kongenere Lipasen mit verbesserter katalytischer Aktivität und Substratzugang. 3. Annual meeting of the Biokatalyse2021 Cluster, Oct 18 – 19, 2010, Kiel, Germany.

Azatryptophans endow Proteins with Intrinsic Blue Fluorescence (oral poster presentation). 3rd European Conference on Chemistry for Life Sciences, Sep 2 – 5, 2009, Frankfurt/Main, Germany.

Department Retreat, Trp Codon Reassignment in the Flavoprotein Dodecin, Nov 09 – 12, 2008, Ringberg Castle, Ringberg, Germany.

8. Graduate Retreat, Design of Blue Fluorescent Proteins with an Expanded Genetic Code, Jun 28 - 30, 2008, Ringberg Castle, Ringberg, Germany.

7. Graduate Retreat, *In vivo* and *in vitro* incorporation of tryptophan analogs into ECFP and Annexin V, Jun 15 – 18, 2007, Ringberg Castle, Ringberg, Germany.

I Table of contents

II Summary

The aim of this study was to extend and improve the currently available tools for protein engineering using noncanonical amino acids in *Escherichia coli*. To date, two complementary approaches exist for incorporating noncanonical amino acids: (i) global residue-specific reassignment of sense codons by the supplementation-based incorporation method, and (ii) site-specific genetic code expansion using orthogonal aminoacyl-tRNA synthetase:suppressor tRNA pairs (aaRS:$tRNA_{Sup}$ pairs). First, it was explored whether these two complementary approaches can be successfully combined in a single expression experiment. Second, it was analyzed whether the reassignment of rare codons instead of currently used stop codons can be exploited for the site-specific incorporation of noncanonical amino acids into target proteins. Finally, a new vector system for the coexpression of aaRS:tRNA pairs was developed to facilitate the site-specific labeling of proteins in the future.

To achieve these goals, two established systems for site-specific incorporation were characterized for their substrate tolerance and stop codon suppression efficiency: The orthogonal pairs derived from (i) *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase (*mj*TyrRS) and (ii) *Saccharomyces cerevisiae* phenylalanyl-tRNA synthetase (*sc*PheRS). The *mj*TyrRS derived system only efficiently incorporated 4-benzoylphenylalanine while the engineered *sc*PheRS was capable of introducing a great variety of tryptophan and phenylalanine analogs into proteins, including 3-(1-azulenyl)alanine which had never been incorporated into a protein before.

First, the simultaneous incorporation of 4-benzoylphenylalanine with 4*S*-fluoroproline into enhanced green fluorescent protein as well as with norleucine into *Thermoanaerobacter thermohydrosulfuricus* lipase was successfully achieved. In case of the lipase, the thermal activation independency introduced by norleucine was successfully combined with the possibility for crosslinking or conjugation provided by 4-benzoylphenylalanine. This clearly indicates that it is indeed possible to combine beneficial effects engrafted by global residue-specific recoding of sense codons with site-specific labeling of a protein with a reactive handle.

Second, the first steps towards the use of rare codons for the incorporation of noncanonical amino acids were successful. In the enhanced green fluorescent protein, AGG, a rare arginine codon, was read to 35% with tryptophan instead of arginine. This was possible by challenging the endogenous translation system with the engineered *sc*PheRS in combination with three copies of its cognate tRNA, whose anticodon was mutated to recognize AGG. This result represents an interesting starting point for the development of rare codon based incorporation systems for noncanonical amino acids.

Third, the development of a new modular vector family termed ‘pMEc’ was successful. This will not only enhance the efficiency of stop codon suppression and site-specific sense codon reassignment, but also facilitate the implementation and evaluation of new orthogonal aaRS:tRNA pairs.

In summary, the presented results provide a wide range of new methodological possibilities for protein engineering with noncanonical amino acids, and also highlight the possibilities for future developments in the field of synthetic biology.

III Zusammenfassung

Das Ziel dieser Arbeit war die Erweiterung und Verbesserung der derzeit verfügbaren Methoden des *Protein Engineering* mit nichtkanonischen Aminosäuren in *Escherichia coli*. Gegenwärtig gibt es zwei komplementäre Ansätze: den aminosäurerestspezifischen Einbau durch Sense-Codon-Neuzuordnung einerseits und den ortsspezifischen Einbau durch orthogonale Aminoacyl-tRNA Synthetase:Suppressor tRNA Paare (AaRS:$tRNA_{Sup}$ Paare) andererseits. In dieser Arbeit sollte zum einen untersucht werden, ob diese komplementären Ansätze in einem einzigen Expressionsexperiment erfolgreich miteinander kombiniert werden können. Zum anderen sollte analysiert werden, ob sich statt Stop- auch selten auftretende *Sense*-Codons für den ortsspezifischen Einbau nichtkanonischer Aminosäuren in Zielproteine eignen. Darüber hinaus sollte ein neues Vektorsystem für die Koexpression von AaRS:tRNA Paaren entwickelt werden, um den ortsspezifischen Einbau in Proteine künftig zu erleichtern.

Um diese Vorhaben anzugehen, wurden zuerst zwei bereits bestehende, von *Methanocaldococcus jannaschii* Tyrosyl-tRNA Synthetase (*mj*TyrRS) und *Saccharomyces cerevisiae* Phenylalanyl-tRNA Synthetase (*sc*PheRS) abgeleitete AaRS:$tRNA_{Sup}$ Paare auf Substrattoleranz und Effizienz in der Suppression von Stop-Codons untersucht. Während das von *mj*TyrRS abgeleitete System spezifisch 4-Benzoylphenylalanin einbaute, war es mit der mutierten *sc*PheRS möglich, eine ganze Reihe Tryptophan and Phenylalanin Analoga in Zielproteine einzubringen. Dazu zählte auch das 3-(1-Azulenyl)alanin, das bisher nicht erfolgreich in Proteine eingebaut werden konnte.

Das erste Ziel wurde durch erfolgreiche Kombination des 4-Benzoylphenylalanin Einbaus mit 4*S*-Fluoroprolin im *Enhanced Green Fluorescent Protein*, und mit Norleucin in einer Lipase aus *Thermoanaerobacter thermohydrosulfuricus* erreicht. Im Fall der Lipase konnte die durch Norleucin eingeführte Unabhängigkeit von thermischer Aktivierung erfolgreich mit der Möglichkeit zur Quervernetzung oder Konjugation am 4-Benzoylphenylalanin verknüpft werden. Dieses Ergebnis zeigt, dass es möglich ist, positive Effekte, die durch aminosäurerestspezifische Sense-Codon-Neuzuordnung erzielt wurden, mit ortspezifischem Einbringen von reaktiven Gruppen zu verbinden.

Zum zweiten waren erste Schritte zur Verwendung von selten auftretenden *Sense*-Codons für den ortsspezifischen Einbau nichtkanonischer Aminosäuren erfolgreich. Im *Enhanced Green Fluorescent Protein* wurde ein seltenes Arginincodon, AGG, zu 35 % zu Tryptophan umgewidmet. Dies wurde erreicht, indem das endogene Translationssystem der mutierten *sc*PheRS in Kombination mit drei AGG-Anticodon tragenden tRNAs ausgesetzt wurde. Das Ergebnis stellt einen interessanten Startpunkt für die Entwicklung von Einbausystemen auf Basis von selten auftretenden *Sense*-Codons dar.

Nicht zuletzt konnte ein neues „pMEc“ genanntes Vektorsystem erfolgreich eingeführt werden. Es wird nicht nur die Effizienz der Stop-Codon Suppression und der ortsspezifischen Sense-Codon Neuzuordnung erhöhen, sondern auch die Implementierung und Evaluierung neuer orthogonaler AaRS:tRNA Paare vereinfachen.

Die präsentierten Ergebnisse stellen eine große Bandbreite an neuen Möglichkeiten für das *Protein Engineering* mit nichtkanonischen Aminosäuren bereit und zeigen Möglichkeiten für zukünftige Entwicklungen im Bereich der Synthetischen Biologie auf.

IV Abbreviations and definitions

IV.I Definitions

Canonical amino acids are the 20 amino acids of the standard genetic code. Therefore, they are also named **proteinogenic** amino acids. Pyrrolysine and Selenocysteine are denoted as **special proteinogenic** because they are genetically encoded but not part of the canonical amino acids.

Noncanonical amino acids are all amino acids which do not participate naturally in protein translation. In literature, the terms nonnatural amino acid, unnatural amino acid and synthetic amino acid are often incorrectly used as equivalent to noncanonical amino acids. However, the usage of these terms is misleading because a fraction of the noncanonical amino acids is also natural but not taking part in protein translation, e.g. ornithine. If a noncanonical amino acid is isostructural to a canonical one, it can also be denoted as canonical amino acid **analog**.

Genetic code engineering denotes the *in vivo* residue-specific reassignment of sense codons and refers to incorporation of noncanonical amino acids by the supplementation-based incorporation method.

Genetic code expansion denotes the *in vivo* site-specific incorporation of noncanonical amino acids via orthogonal aminoacyl-tRNA synthetase:suppressor tRNA pairs.

Orthogonal pairs are 'insulated' exogenous aminoacyl-tRNA synthetase:suppressor tRNA pairs which do not cross-react with any endogenous host components, i.e. the suppressor tRNA must not be recognized by the endogenous host aminoacyl-tRNA synthetases and the exogenous aminoacyl-tRNA synthetase does not recognize any of the endogenous host tRNAs.

Variants are all proteins which contain noncanonical amino acids. They can also be called **alloproteins**.

Congeners denote proteins which are produced by genetic code engineering, i.e. they originate from the same gene sequence, but contain a small fraction of amino acids exchanged by related noncanonical amino acids.

A **mutant** is a protein in which the wild-type sequence is changed by site-directed mutagenesis in the frame of canonical or noncanonical amino acids, i.e. changes at the DNA level were introduced into the gene of the target protein.

IV.II Abbreviations

All canonical amino acids were abbreviated according to the general three or one letter amino acid code. For noncanonical amino acids, the respective most common abbreviation in the literature was used.

The following abbreviations were used throughout the thesis:

A	Adenine
aa	Amino acid
Aal	3-(1-azulenyl)alanine
aaRS	Aminoacyl-tRNA synthetase
aa-tRNA	Aminoacyl-tRNA
^{a}C	Activated charcoal
Cm	Chloramphenicol
Aha	Azidohomoalanine
Amp	Ampicillin
AMP	Adenosine-5'-monophosphate
AmPhe	4-aminophenylalanine
APS	Ammonium persulfate
Approx.	Approximately
ATP	Adenosine-5'-triphosphate
*av*GFP	*Aequorea victoria* green fluorescent protein
au	Arbitrary units
AzPhe	4-azidophenylalanine
bp	Base pair
Bpa	4-benzoylphenylalanine
BpyAla	2,2′-bipyridin-5-yl)alanine
6BrTrp	6-bromotryptophan
BT	3-(3-benzothienyl)alanine
C	Cytosine
caa	Canonical amino acid
cal	Calculated
CuAAC	Copper(I)-catalyzed azide–alkyne Huisgen cycloaddition
6ClTrp	6-chlorotryptophan
CV	Column volume
Da	Dalton
ddH_20	Double-distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DSB	Double strand breakage

DTT	Dithiothreitol
ECFP	Enhanced cyan fluorescent protein
E. coli (*ec*)	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EF-Tu	Elongation factor Tu
EGFP	Enhanced green fluorescent protein
ε_M	Molar extinction coefficient
ESI-MS	Electrospray ionization mass spectrometry
Fal	3-(2-furyl)alanine
FACS	Fluorescence activated cell sorting
G	Guanine
GTP	Guanosine-5'-triphosphate
GlcNAc	*N*-acetylglucosamine
Hpg	Homopropargylglycine
HPLC	High performance liquid chromatography
IPhe	4-iodophenylalanine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
λ	Wavelength
Kan	Kanamycin
LB	Lysogeny broth
LC-ESI-MS	Liquid chromatography electrospray ionization mass spectrometry
M	Molar
*m*DHFR	*Murine* dihydrofolate reductase
meas	Measured
M. jannaschii (*mj*)	*Methanocaldococcus jannaschii*
mRNA	Messenger RNA
MTyr	*O*-methyltyrosine
M_w	Molecular weight
MWCO	Molecular weight cut-off
Nal	3-(2-naphthyl)alanine
ncaa	Noncanonical amino acid
nd	Not detected
ni	Non-induced
Nle	Norleucine
NME	N-terminal methionine excision
NMM	New minimal medium
NTA	Nitrilotriacetic acid
OD_{600}	Optical density at λ = 600 nm
PAGE	Polyacrylamide gel electrophoresis

PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PIPES	Piperazine-N,N′-bis(2-ethanesulfonic acid)
PP_i	Pyrophosphate (inorganic)
PoxPhe	4-propargyloxyphenylalanine
ψ-b*	Cysteine-free pseudo wild-type barstar
*p*NPP	*p*-nitrophenyl palmitate assay
PTM	Posttranslational modification
RE	Restriction enzyme
RF	Release factor
RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal RNA
RT	Room temperature
SC-medium	Synthetic Complete medium
S. cerevisiae (*sc*)	*Saccharomyces cerevisiae*
SCS	Stop codon suppression
SDS	Sodium dodecyl sulfate
SeMet	Selenomethionine
SPI	Supplementation-based incorporation
ssDNA	Single stranded DNA
T	Thymine
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
TIC	Total ion current
T_M	Melting temperature
t_R	Retention time
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
$tRNA_{Sup}$	Suppressor tRNA
TTL	*Thermoanaerobacter thermohydrosulfuricus* lipase
U	Uracil
U	Enzymatic units
Ura	Uracil
YPD	Yeast Peptone Dextrose

1 Introduction

1.1 The Genetic Code

In all biological systems the 'blueprint of life' is encoded by a polymeric macromolecule, the deoxyribonucleic acid (DNA). The biopolymer was first isolated by Miescher in 1871[1] and identified as carrier of the genetic information by Avery in 1944.[2] Based on early works of Levene, in the late 40s Todd elucidated that the basic DNA building blocks are Adenine (A), Cytosine (C), Guanine (G), and Thymine (T) linked by pentose-phosphates as scaffold. Shortly after, Chargaff found out that DNA always contains equal amounts of A and T, and C and G[3] and thus enabled the structure determination of DNA as double helix by Franklin, Watson and Crick.[4]

The concept of proteins as family of 'macromolecular nitrogen-containing substances' was likewise established in the late 18th century. In 1838, Mulder found that these substances had an uniform atomic composition.[5] To account for their fundamental significance, Berzelius created the name 'protein' (from the greek *proteios*, meaning 'primary'*)*.[6] Mulder assumed that proteins were composed of one fundamental species of building blocks ('*Grundstoffe'*), now known as amino acids. Leucine (Leu) was the first amino acid (aa) to be isolated by hydrolysis of a protein already in 1820 and threonine (Thr) was the last of the 20 canonical (proteinogenic) amino acids identified in 1935.[7] By this time, the central role of proteins as 'molecular machines' of the cells was widely accepted, not least because of Sumner who proved that the enzyme urease was a protein.[8]

In the 1960s, Crick was the first to link DNA and protein sequences in his 'central dogma of molecular biology'.[9] He postulated a unidirectional flow of genetic information from 'nucleic acids to nucleic acids or nucleic acids to protein' with no transfer being possible back from protein to nucleic acids. The dogma defined (i) the transcription of DNA to ribonucleic acid (RNA, herein T is exchanged by uracil (U)) and (ii) the translation of messenger RNA (mRNA) to protein by the ribosome (see chapter 1.2). Based on this theoretical framework, it was shown that a triplet base sequence (codon) corresponds to one amino acid in a protein[10] leading to 64 possible coding units which Nirenberg and Khorana subsequently assigned to the different amino acids.[11-13]

From code deciphering, it became clear that 61 sense codons encode the 20 amino acids while three stop codons (UAA, ochre; UGA, opal; and UAG, amber) are used to terminate translation (**Figure 1**). Thus, the genetic code is degenerated as all amino acids except methionine (Met) and tryptophan (Trp) are assigned to more than one codon. In this respect, the most extreme amino acids are Leu, arginine (Arg), and serine (Ser) which are encoded by six codons each. Further, the genetic code is not random since, for example, related codons (i.e. codons which differ only at one position) tend to code for the same or two chemically related amino acids.

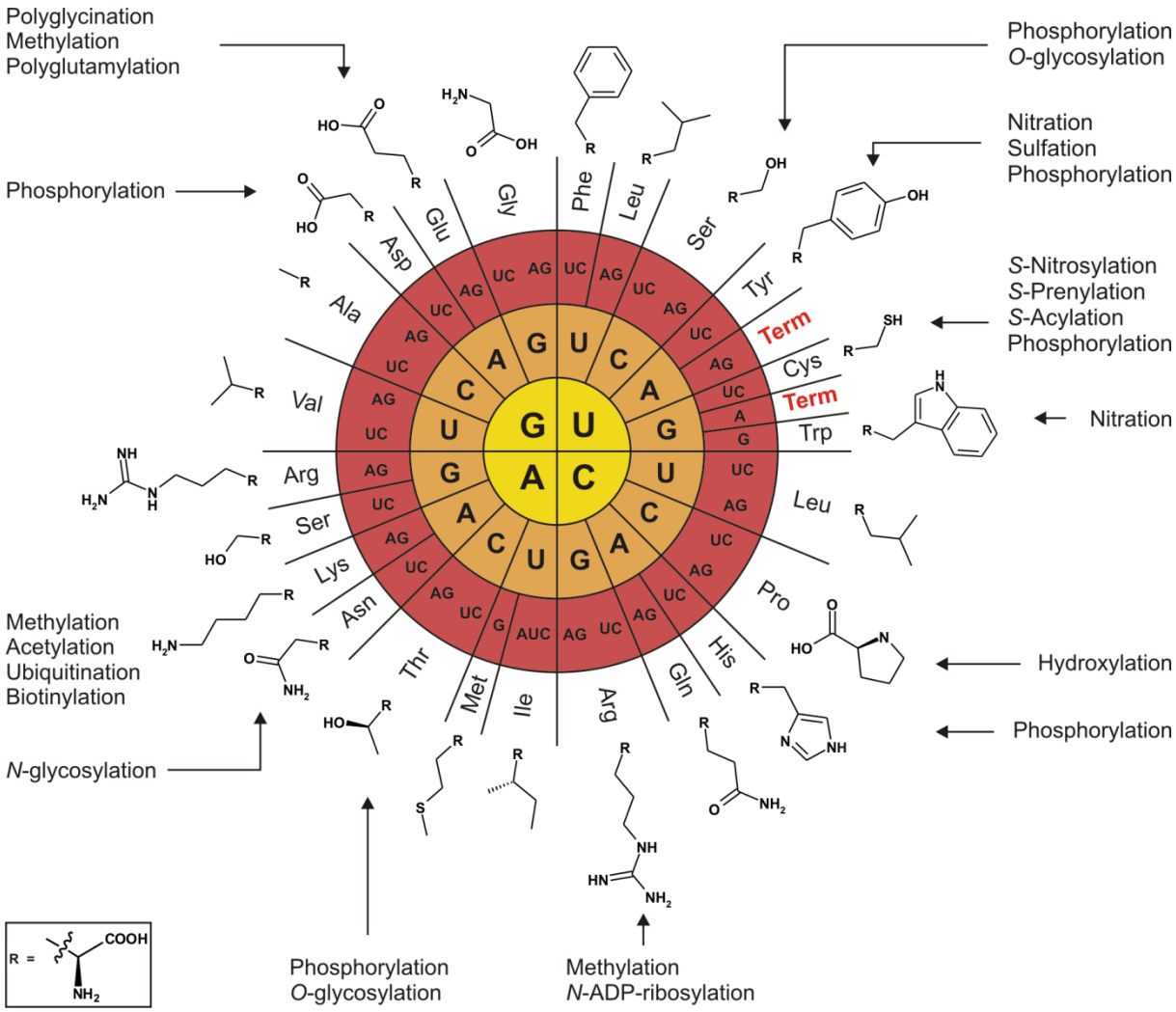


Figure 1. Presentation of the Genetic Code in RNA format. 61 of the 'three-letter' codons are assigned to 20 canonical amino acids (caa). The three residual triplets are used as stop codons (denoted as Term). Yellow circle: 5'-nucleobase of the triplet, orange circle: middle nucleobase of the triplet and red circle: 3'-nucleobase of the triplet. Beyond the 20 canonical amino acids nature expands its chemical repertoire with posttranslational modifications (PTM) of different amino acids. Some frequently appearing PTMs are depicted.

This high intrinsic non-randomness was attributed to a selective evolutionary process, which minimized the damage of mutations, and errors of transcription and translation on protein level (known as adaptive theory).[14]

Early after the deciphering of the genetic code, Crick proposed a first theory on its evolution.[15] Taking into account the limited number of only 20 canonical amino acids (caa) he suggested that after an initial period of expansion, the evolution of the code reached a dead end. At that point, the code congealed in the so called 'frozen accident' which precluded every further expansion of the amino acid pool due to the disruptive effect of new amino acids on the proteome.[16]

This theory gave a good explanation of the general universality of the genetic code within all life forms but did neither account for its obvious non-randomness (*vide supra*) and experimentally evident flexibility on DNA level nor for its receptivity to new amino acids on protein level (*vide infra*). These facts clearly point to a nonaccidental evolution.

Subsequently, additional theories have been developed trying to link the pattern of amino acid assignments to physicochemical or biological factors. Beside the adaptive theory mentioned above, the coevolution theory of genetic code and amino acids was most successful since it gained strong statistical support and left space for a further ongoing evolution of the genetic code.[17] According to this theory, additional amino acids that appeared as a result of newly evolved metabolic pathways were integrated into a pool of primordial proteinogenic amino acids by reassigning codons that originally encoded other amino acids.[18]

It is now clear that not all 20 canonical amino acids were prebiotically synthesized. However, to date, no theory could give a logic and comprehensive answer on why there is only a pool of exactly these 20 canonical amino acids.[19] The question is particularly interesting because on the one hand, much more amino acids were prebiotically available but did not end up in the standard alphabet; and on the other hand, many others are metabolic products (e.g. ornithine, citrulline or homoserine) but do not participate in protein synthesis either. This is remarkable because the side-chain functionalities of the canonical amino acids are obviously not sufficient for the many tasks proteins fulfill in living organisms since nature evolved many posttranslational modifications (PTM) to extend the chemical repertoire of proteins.[20]

Interestingly, two special amino acids, selenocysteine (Sec)[21] and pyrrolysine (Pyl),[22] exist which are not canonical, i.e. they are not encoded by sense codons, but still proteinogenic. Sec is incorporated in response to the opal stop codon, but only when a distinct structural element is present on the mRNA. In contrast, Pyl is incorporated via amber stop codon suppression (SCS) without a specific sequence context. However, organisms using Pyl seem to have evolved such that UAG is mostly avoided as stop codon.[23] The mode of Sec and Pyl incorporation can be seen as a proof of the ‘frozen accident’ concept because nature obviously circumvented deleterious effects of both amino acids on the proteome by special coding strategies. Furthermore, their existence in the code provides evidence for its ongoing evolution since their integration clearly occurred after the manifestation of the canonical amino acids and did not ubiquitously take place in all organisms.[23]

In addition to these special proteinogenic amino acids, several ‘dialects’ of the standard genetic code exist in different species or organelles. For example, UGA encodes Trp in some ciliates and the mitochondria of a number of species, and AGA got lost in the GC-rich genomes of some *Micrococcus* species.[24] To explain the evolution of these deviations, two complementary theories have been developed: the ‘ambiguous intermediate’ and the ‘codon capture theory’.[25] The ambiguous intermediate theory postulates that codon reassignment takes place through an intermediate stage in which a codon is ambiguously decoded followed by a subsequent loss of its original meaning.[26] Codon capture is believed

to occur under evolutionary pressure to decrease e.g. genomic AT-content leading to a disappearance of certain AT-rich codons. These codons would eventually reappear by random genetic drift and would subsequently be reassigned with a different meaning.[27] In both cases, the crucial events for reassignment are mutational changes in the transfer RNA sequence (tRNA, see chapter 1.2.3).

Both, the 'special' proteinogenic amino acids Sec and Pyl, as well as the described 'dialects' of the genetic code on DNA level rationalize attempts for synthetic genetic code evolution. This objective includes the development of strategies for the incorporation of noncanonical amino acids (ncaas, amino acids which are not naturally participating in protein translation) into proteins and proteomes (see chapter 1.3).

1.2 Protein translation

1.2.1 General aspects of protein translation

As briefly outlined above, the genetic information is conveyed from DNA to RNA by transcription (see **Figure 2**). At this stage, several different types of RNA with various functions are produced. Ribosomal RNAs (rRNA) are the main catalytic components of the ribosome, tRNAs act as the adapter molecules between codon and amino acid and mRNAs serve as templates for protein translation.

Translation begins with the two components of the translational machinery that carry the major responsibility for the interpretation of the genetic code: The aminoacyl-tRNA synthetase (aaRS) and the tRNA. Both are described in more detail in chapter 1.2.2 and 1.2.3, respectively. Briefly, the aaRS selects its cognate amino acid from the pool of canonical amino acids and loads it onto the acceptor stem of the cognate tRNA, thereby linking the amino acid directly to its encoding nucleic acid sequence via the anticodon. The resulting aminoacyl-tRNA (aa-tRNA) is guided to the ribosome by the GTPase (guanosine-5'-triphosphate hydrolase) elongation factor Tu (EF-Tu). Aa-tRNA binds to EF-Tu yielding a ternary EF-Tu-aa-tRNA-GTP complex. A full accommodation of aa-tRNA in the ribosome only takes place if codon-anticodon base pairing is correct, leading to aa-tRNA liberation from the ternary complex. The ribosome has three tRNA binding sites, the aminoacyl-site (A-site), the peptidyl-site (P-site) and the exit-site (E-Site). First, the aa-tRNA is accommodated at the A-site via codon-anticodon base pairing, followed by transpeptidation of the already synthesized polypeptide from the peptidyl-tRNA at the P-site to the aa-tRNA at the A-site. Then, the elongated peptidyl-tRNA at the A-site is translocated to the P-site and the deacylated tRNA leaves the ribosome via the E-site.[28] This ribosomal cycle is repeated until a stop codon resides at the ribosomal A-site allowing the binding of a release factor (RF) to the ribosome. The RF subsequently triggers the release of the nascent peptide chain from the peptidyl-tRNA at the P-site (see chapter 1.2.4).

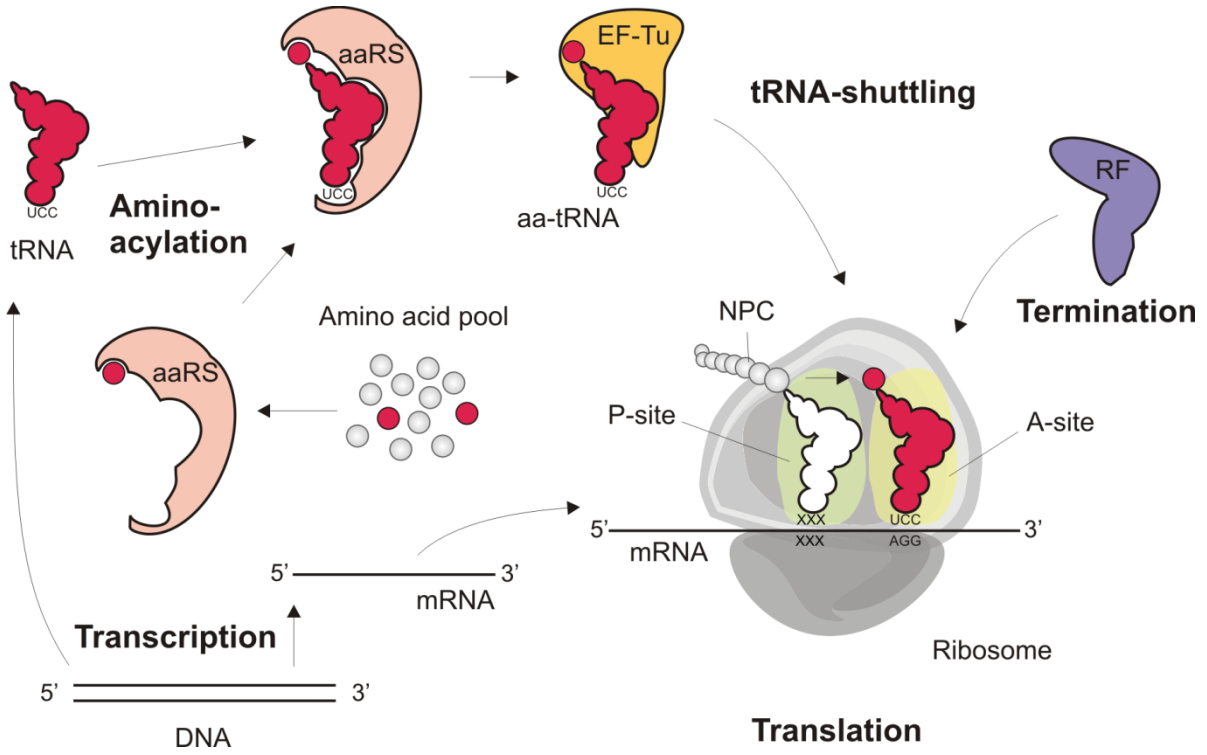


Figure 2. The cellular protein translation system. Amino acids are activated by aminoacyl-tRNA synthetases (aaRS) and loaded onto the cognate tRNA yielding an aminoacyl-tRNA (aa-tRNA). Elongation factor Tu (EF-Tu) recruits aa-tRNA and guides it to the ribosomal aminoacyl-site (A-site) where aa-tRNA undergoes codon-anticodon base pairing with the mRNA. The nascent polypeptide chain (NPC) is subsequently transferred from the peptidyl-tRNA (white) located at the peptidyl-site (P-site) to the aa-tRNA at the A-site. Thereby, the polypeptide chain extended by one amino acid. When a stop codon resides in the ribosomal A-site, a release factor (RF) enters and triggers the release of the nascent polypeptide chain.

Various proofreading checkpoints in the whole translation process assure high fidelity protein synthesis. In this control system, aaRSs are by far the most important gatekeepers of the genetic code. Their accuracy in loading the correct amino acid onto a tRNA is the crucial step, since mischarging will mostly lead to introduction of a wrong amino acid in response to a certain codon (see chapter 1.3.2). However, also the ribosome has some options for monitoring fidelity; most of them are mediated by EF-Tu. First, as mentioned above, correct codon reading is assured by EF-Tu because aa-tRNA can only fully enter at the ribosomal A-site when the codon–anticodon interaction is correct and thus GTP hydrolysis liberates aa-tRNA from EF-Tu (proofreading). Second, misacylated tRNAs stimulate the GTPase activity of EF-Tu much more slower than cognate tRNAs (initial selection).[28] As a consequence, correctly acylated tRNAs have a selective advantage at the ribosome because they have more time to expel misacylated tRNAs from the A-site. Third, tRNA misacylation negatively affects the initial binding to EF-Tu reducing the chance to be transferred to the ribosomal A-site.[29] Despite these proofreading checkpoints, the ribosome even accepts very bulky and chemically extraordinary amino acids on misacylated tRNAs for polypeptide chain elongation (see chapter 1.3).

1.2.2 Aminoacyl-tRNA synthetases

As mentioned before, aminoacyl-tRNA synthetases are the most important interpreters of the genetic code because they link the amino acid to the respective codon/anticodon by loading it onto its cognate tRNA. AaRSs perform this task in a two-step aminoacylation reaction (see **Figure 3**):

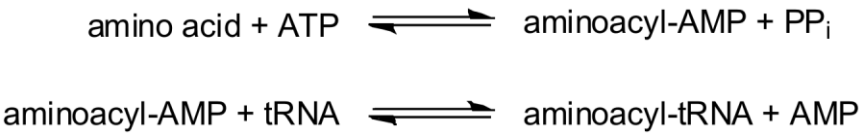


Figure 3. The aminoacylation reaction.

In the first step, the amino acid is activated. In the reaction center of the aaRS, the high-energy leaving group adenosine-5'-monophosphate (AMP) from adenosine-5'-triphosphate (ATP) is attached to the amino acid yielding the mixed anhydride aminoacyl-AMP (aa-AMP). Subsequently, the aminoacyl moiety is transferred to the tRNA.[30] In the resulting aminoacyl-tRNA, an ester bond links the amino acid and either the 3'- or 2'-OH group of the ribose at the 3'-end of the tRNA. Beside their participation in protein translation, aaRSs have many other functions in the cell (reviewed in [31]).

Although catalyzing the same reaction, aaRSs can be divided into two evolutionary very distinct structural classes (class I and II).[32] While class I aaRSs have a dinucleotide-binding Rossmann fold as catalytic domain, the active center of class II enzymes consists of an antiparallel β-fold with seven β-strands and three flanking helices. The Rossmann fold contains two consensus motifs KMSKS and HIGH which are responsible for ATP binding and stabilization of the reaction intermediate (aminoacyl-AMP:aaRS). In contrast, the antiparallel β-fold of class II aaRSs is made up of the conserved motifs 1, 2, and 3 which are important for dimerization (motif 1) as well as ATP and amino acid recognition (motif 2 and 3).[32] Except for differences in the catalytic domains, the main difference between both classes of aaRSs is the tRNA binding mode. Class I synthetases recognize the acceptor stem of the tRNA via its minor groove side with the variable loop facing the solvent, while class II enzymes bind the major groove side and the variable loop is oriented towards the aaRS. In addition, both classes have different specificities regarding the aminoacylation site at the ribose of the 3'-tRNA nucleoside. Class I aaRSs link the amino acid to the 2'-OH residue whereas class II enzymes use the 3'-OH.[32]

According to sequence and structure comparison, both aaRS classes can be further subdivided into three subclasses (I/IIa, I/IIb and I/IIc; see **Figure 4A**). Enzymes within each subgroup and across the subclasses (e.g. Ib – IIb) tend to load amino acids with related chemical properties. For example, charged amino acids like glutamate (Glu), aspartate (Asp) or lysine (Lys) group in subclasses Ib and IIb, and amino acids with aromatic moieties (Tyr, Trp, and Phe) belong to the subclasses Ic and IIc. Furthermore, class I enzymes tend to recognize the bulkier and more hydrophobic amino acids (e.g. Glu vs. Asp or Trp vs. Phe; see **Figure 4A**).[32]

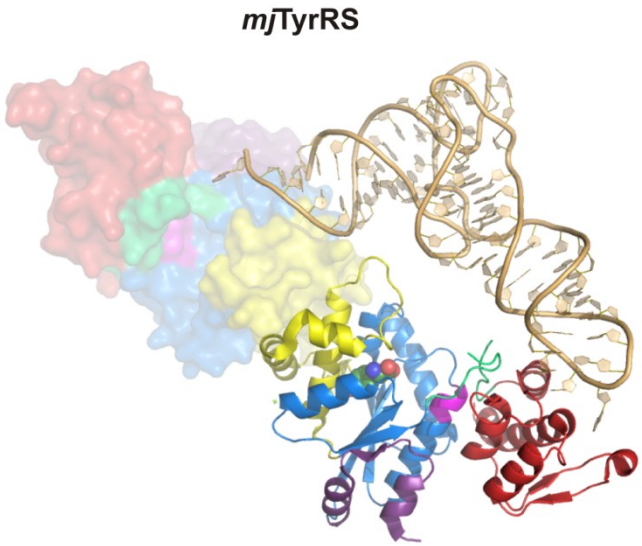
A

Class I		Class II	
Ia	MetRS ValRS LeuRS IleRS CysRS ArgRS	SerRS ThrRS AlaRS GlyRS ProRS HisRS	IIa
Ib	GluRS GlnRS LysRS-I	AspRS AsnRS LysRS-II	IIb
Ic	TyrRS TrpRS	PheRS PylRS	IIc

B

	Class I	Class II
Catalytic domain	Rossmann fold	Antiparallel β-fold
Amino-acylation site	2'-OH	3'-OH
tRNA binding	minor groove	major groove
Anticodon recognition	most	few
Quaternary structure	chiefly monomeric	oligomeric

C



D

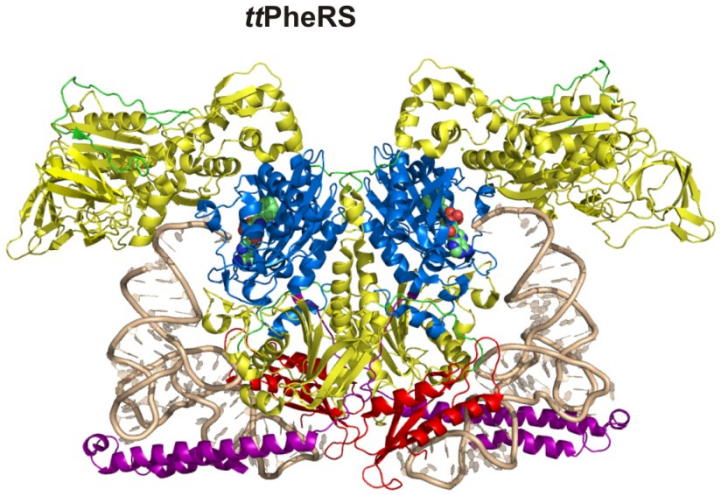


Figure 4. Overview of the two classes of aminoacyl-tRNA synthetases. A: The aaRSs can be divided into class I and class II which exhibit particular properties (**B**). **C:** The crystal structure of *Methanocaldococcus jannaschii* Tyrosyl-tRNA synthetase (*mj*TyrRS, PDB code 1J1U) is given as an example for a class I aaRS. This enzyme is a functional homodimer (α_2). One subunit is represented as ribbon plot in complex with the cognate tRNA (light brown). The other subunit is depicted as surface plot. It is visible that both subunits participate in the recognition of one tRNA molecule. The different domains of the enzyme are colored as follows: red: C-terminal domain (anticodon recognition); yellow: CP1 domain (connective peptide 1 domain, editing); blue: Rossmann fold (catalytic domain); violet: N-terminal domain; pink: HIGH motive and green: KMSKS motive (ATP binding). On the other hand, the crystal structure of *Thermus thermophilus* phenylalanyl-tRNA synthetase (*tt*PheRS, PDB code 2IY5) is an example of a class II aaRS (**D**). This enzyme is a functional heterotetramer ($\alpha_2\beta_2$). *Tt*PheRS is colored as follows: blue: catalytic part of α-subunit; pink: N-terminal coiled coil of α-subunit; yellow: β-subunit; red: anticodon recognition domain (β-subunit) and light brown: tRNA. As in *mj*TyrRS, it can be seen that the tRNA is recognized by α- and β-subunits in parallel. The α-subunit contains the catalytic site, while the β-subunit is involved in tRNA recognition and editing.

The quaternary structures of distinct aaRSs are mostly conserved within the different species. Class I synthetases are chiefly monomeric (α_1) whereas class II enzymes are always oligomeric (α_2, $\alpha_2\beta_2$ or α_4). The oligomeric states are usually crucial for enzyme activity. For instance, the TyrRS from *Methanocaldococcus jannaschii* (*M. jannaschii*, *mj*) is only active as α_2-dimer because the tRNA binds across the two subunits (see **Figure 4C**).[33]

Amino acid binding pockets of the aaRSs are generally well conserved among species while RNA recognition sites do not show significant consensus sequences. In addition, class II enzymes often do not use the anticodon loop of the tRNA (see **Figure 5**) as identity element, whereas most aaRSs from class I do so (see chapter 1.2.3 for details).[30]

Typically, cells contain one aaRS for each of the 20 canonical amino acids and a specific aaRS is only capable of recognizing its cognate tRNA isoacceptors (set of tRNAs which are aminoacylated with the same amino acid) constituting unique aaRS:tRNA pairs.[34] The minimal number of aaRSs found in autotroph species, however, has been found to be 17 because CysRS, GlnRS and AsnRS can be sometimes missing. To compensate for their absence, other strategies were evolutionary developed for correct aminoacylation of their corresponding tRNAs with the cognate amino acids.[35] For example, AspRS can aminoacylate both, $tRNA^{Asp}$ and $tRNA^{Asn}$. The misacylated Asp-$tRNA^{Asn}$ is then converted to Asn-$tRNA^{Asn}$ by aspartyl-$tRNA^{Asn}$ amidotransferase.[36] Another example for such an alternative tRNA charging mechanism is Sec incorporation. In contrast to Pyl, no specific SecRS exists in nature. Sec-$tRNA^{Sec}_{UCA}$ is produced by enzymatic modification of a Ser charged onto $tRNA^{Sec}_{UCA}$ by the endogenous SerRS.[23]

Since some amino acids differ only very little in their structure (e.g. a single methyl group between valine (Val) and isoleucine (Ile)), it is difficult for aaRSs to discriminate them effectively. For example, IleRS also substantially loads Val onto $tRNA^{Ile}$. Nevertheless, to achieve high translational fidelity, many synthetases have developed editing mechanisms towards noncognate canonical amino acids. IleRS hydrolyzes misactivated Val-AMP (pretransfer editing) and misacylated Val-$tRNA^{Ile}$ (posttransfer editing).[37,38] In class I synthetases, hydrolysis of misacylated tRNAs is performed by the CP1 domain (see for instance **Figure 4C**). In contrast, class II enzymes do not have a conserved editing domain. For example, PheRS is active as an $\alpha_2\beta_2$-heterotetramer and has a unique editing domain which is located on the β-subunit while the α-subunit harbors the active center (see **Figure 4D**).[38] Such editing mechanisms could only be evolved against amino acids which are present in nature and in sufficient concentration to make discrimination necessary. Consequently, a great number of aaRSs exhibit significant tolerance towards noncanonical amino acids making genetic code engineering possible (see chapter 1.3.2).

1.2.3 Transfer RNAs

As already mentioned in chapter 1.2.2, tRNAs act as the interface between nucleic acids and proteins during translation. They mostly have 76 nucleobases but their length can vary between 60 and 95 nucleotides depending on identity and organism. tRNAs adopt a cloverleaf secondary structure that comprises five characteristic structural domains: acceptor stem, dihydrouridine (D) stem and loop, anticodon (AC) stem and loop, variable loop and thymidine (T) stem and loop (see **Figure 5A**). 15 conserved and 8 semi-conserved residues throughout the tRNA assure an L-shaped tertiary structure (see **Figure 5B**) with an amino acid acceptor branch (acceptor stem stacked over the T-stem-loop) linked to an anticodon branch (D–stem-loop stacked over the AC–stem-loop).[39]

Many bases in tRNA can be modified by tRNA modification enzymes (see **Figure 5A**).[40] These modifications are obviously not crucial for tRNA recognition by its cognate aaRSs;[41] however, they appear to play a major role in another core principle of genetic code interpretation called wobbling. Genomes contain only around 40 tRNA genes, which are not enough to read all 61 sense codons. Thus, already in 1966, Crick hypothesized that some tRNAs could read more than one codon by non-Watson-Crick base pairing between the base located at position 34 of the tRNA sequence and the third nucleobase of a codon.[42]

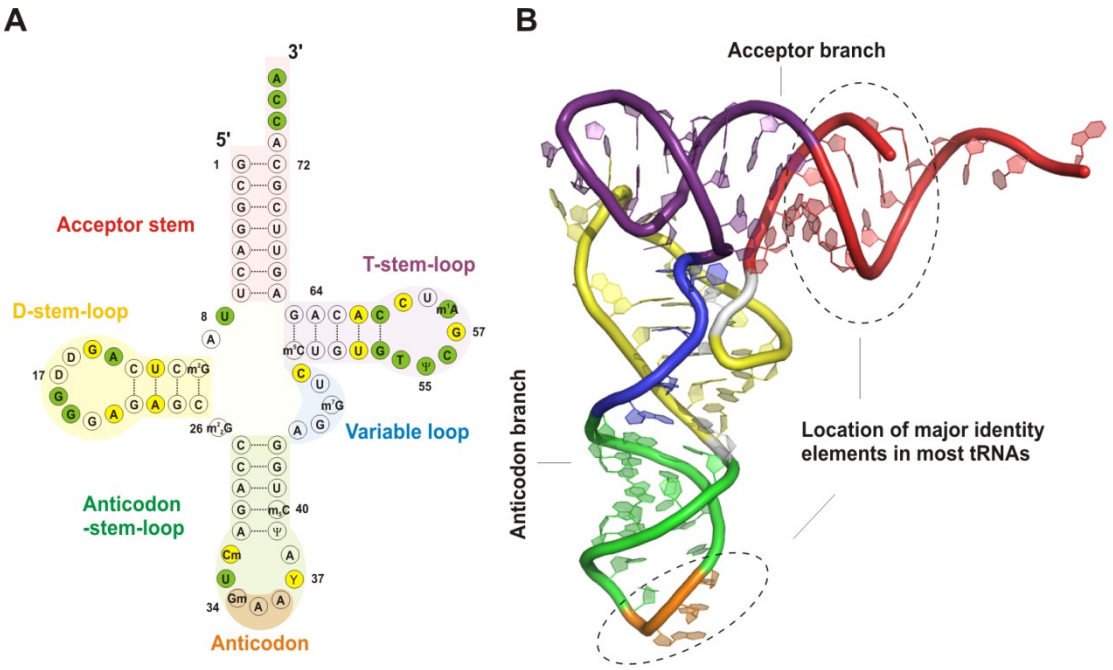


Figure 5. Sequence and structure of tRNA^{Phe}. A: Cloverleaf presentation of tRNA secondary structure. The different structural regions are labeled in color code to identify the corresponding structural elements in the crystal structure shown in **B**. Conserved and semi-conserved residues are depicted in green and yellow circles, respectively. Modified nucleobases are written in their standard nomenclature: m^{2}G: N^2-methylguanosine; D: dihydrouridine; m$^2_2$G: N^2,N^2-dimethylguanosine; Cm: 2-*O*-methylcytidine; Gm: 2-*O*-methylguanosine; Y: wybutosine; Ψ: pseudouridine; m^{5}C: 5-methylcytidine; m^{7}G: 7-methylguanosine and m^{1}A: 1-methyladenosine.[40] **B:** L-shaped tertiary structure of tRNA^{Phe} (PDB code 4TNA).

This process is usually facilitated by post-transcriptional modifications at the wobble position 34.[43] For example, $tRNA^{Arg}_{UCU}$ can only efficiently recognize the codon AGG because the uracil located at the position 34 in the anticodon is modified to 5-methylaminomethyluridine (mnm^5U).[44,45] In general, modifications at the wobble position 34 and purine 37 (see **Figure 5A**)[46] not only pre-structure the anticodon domain, but also ensure a correct codon binding, and maintain the translational reading frame.[47] The loss or change of such modifications[48] together with mutations in the anticodon of the tRNA are thought to be a main reason for species specific differences in genetic code interpretation.[41]

Due to wobbling, sometimes a single tRNA is sufficient to read all synonymous codons of a specific amino acid. However, most codons are read by different isoaccepting tRNAs with sometimes overlapping codon specificity. Not all of these isoacceptors are present in equal amounts in a cell. While major tRNAs are always highly expressed in the cell, minor, rare tRNAs are only present in tiny amounts. In fact, the concentration of a tRNA isoacceptor can often be directly correlated to the frequency of its specific codon.[49] This codon bias in the genome can vary from species to species, and is denoted as the distinct codon usage of a certain organism. For example, the Arg codons CGC and CGU are very frequent in *Escherichia coli* (*E. coli*, *ec*) strain MG1655 (approx. 22% of all Arg codons) while the rare codons AGA and AGG are only present at 1.9% and 1.0% of all Arg positions, respectively.[50] In absolute numbers, AGG and AGA are also the least used sense codons in the whole *E. coli* genome. tRNA abundance plays several regulatory roles. On the one hand, highly expressed genes generally contain codons that are decoded by highly abundant tRNAs.[51] On the other hand, clusters of rare codons are often involved in a process called ribosomal pausing. This phenomenon has been proposed to be needed for a temporal separation of the translation of protein domains to facilitate their cotranslational folding.[52]

Unique identity elements within the tRNA structure make highly specific interactions with the cognate synthetase and assure that aaRSs can discriminate between cognate and noncognate tRNAs. These identity rules are often denoted as the 'second genetic code' because they are key to the correct translation of a DNA sequence into protein. Identity determinants are nucleobases mainly located at the distal end of the acceptor and the anticodon branch. They can lead to a productive interaction with the aaRS (determinant) or prevent aminoacylation by a noncognate aaRS (antideterminant). In the acceptor stem, the first four base pairs and the 'discriminator base' N73 are the most important discriminators for all aaRSs. Furthermore, most aaRSs use the anticodon for identity verification.[53] For example, TyrRS loses most of its affinity towards $tRNA^{Tyr}$ when the anticodon is mutated to an amber stop codon.[54,55] The same mutation leads to partial misacylation of $tRNA^{Phe}$ by LysRS.[56] The anticodon is not used as discriminator by alanyl-tRNA synthetase (AlaRS), LeuRS, and SerRS. Instead, for instance, SerRS specifically discriminates for an extra long variable loop found in all $tRNA^{Ser}$.[57] Last but not least, all aaRS:tRNA pairs have some special idiosyncratic determinants.[41]

Phylogenetically distant organisms have deviations in the universal identity rules, such as changes within weak determinants and other sequence variations in the tRNA. Thus, cross-aminoacylation between aaRSs and tRNAs from evolutionary distant organisms is

normally impaired. This is clearly illustrated in the example of $tRNA^{Tyr}$. The C1-G72 identity pair defines the kingdom specificity in Archaea and Eukarya while the reverse G1-C72 pair is used in bacterial $tRNA^{Tyr}$.[58] These phylogenetic differences in the identity rules are the prerequisites for genetic code expansion (see chapter 1.3.3) since they allow for the introduction of an aaRS:tRNA pair from one phylum into another without generation of mutual cross-reactivities (orthogonality).

1.2.4 Translation termination versus termination suppression

Bacterial and archaeal protein synthesis is terminated by the release factors RF1, RF2 and RF3 (see chapter 1.2.1). While the genes of RF1 and RF2 are essential in *E. coli*, RF3 can be knocked out without being deleterious.[59] The three dimensional structures of RF1 and RF2 mimic the L-shaped form of tRNAs. Upon positioning of a nonsense (stop) codon at the ribosomal A-site, the release factors enter the ribosome and trigger polypeptide chain liberation. UAA and UAG are recognized by RF1, whereas RF2 provides translational stop in response to UAA and UGA. RF3 is a GTPase and does not have termination function *per se* but rather stimulates the dissociation of the ribosomal complex by GFP hydrolysis, thereby enhancing termination efficiency.[30] The frequency of the different stop codons in the genome is highly strain dependent. For example, the genome of *E. coli* strain K12 subtype W3100 contains 2765 UAA, 1249 UGA, and 321 UAG stop codons. Some methanogenic archaea, however, seem to avoid the amber codon totally as the translation termination signal (see chapter 1.2.1).[23] Moreover, the termination rate is highly codon dependent with a decreasing efficiency in the order UAA > UAG > UGA.[60]

Besides causing translation termination, the ‘stop’ meaning of nonsense codons can be suppressed, leading to translational read-through. Such suppression is achieved by tRNAs which have their anticodon mutated to an anti-stop codon.[61,62] The process of suppression has a natural significance. For example, the special proteinogenic amino acids Pyl and Sec are incorporated into proteins in response to stop codons (see chapter 1.1). Furthermore, suppressor tRNAs ($tRNA_{Sup}$) are needed in the production of essential genes of bacteriophages.[63] In the life sciences, isolated *E. coli* suppressor strains (e.g. with *supE* ($tRNA^{Gln}_{CUA}$)) have been used for expression of antibody-phage envelope fusion proteins for phage display.[64]

Most importantly, suppressor tRNAs can be used for genetic code expansion with noncanonical amino acids (see chapter 1.3.3). However, the process of suppression is rather inefficient as every tRNA has to compete with release factors for entry into the ribosomal A-site. The ratio between termination and suppression at a nonsense codon is highly influenced by positive or negative structural interactions of the $tRNA_{Sup}$-EF-Tu complex or the RF at the ribosomal A-site.

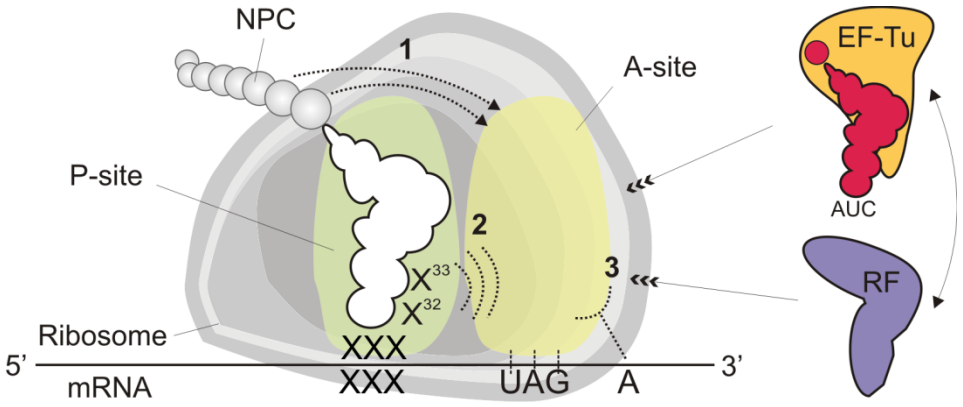


Figure 6. Translation termination *versus* suppression. Structural effects in the ribosome influence the ratio between termination by release factors and suppression by suppressor tRNAs. Interactions between the C-terminal amino acids of the nascent peptide chain and the ternary EF-Tu-aa-tRNA-GTP complex (**1**), the composition of the 5'-side of the peptidyl-tRNA (**2**), and the 3'-mRNA context (**3**) play a role in modulating the ratio between termination and suppression.

On the one hand, such interactions can take place with the peptidyl-tRNA at the P-site.[60] In brief, the 5'-side of the peptidyl-tRNA anticodon loop interacts with the tRNA at the A-site. Thus, small bases (A and U) at position 32 and 33 promote suppression by leaving the A-site more open allowing an easier entry of $tRNA_{Sup}$ (see **Figure 6**).[65] On the other hand, the competition between $tRNA_{Sup}$ and RF is influenced by the functional interaction between the C-terminal amino acids in the nascent polypeptide chain and EF-Tu.[66-68] First, if the ultimate amino acid of the polypeptide at the peptidyl-tRNA has a high propensity for an ordered structure (e.g. α-helical, 3-strand or reverse-turn motives), termination is favored. Instead, a decrease in its van der Waals volume seems to increase the efficiency of suppression.[67,68] Second, negative charge or hydrophobicity at the penultimate amino acid position provokes better suppression. The effects of the ultimate and penultimate amino acid are cooperative which indicates that the last amino acid might have only indirect influence, maybe by positioning of the interacting penultimate residue (by secondary structure propensity).[67,68] All these steric effects between tRNAs, RFs and EF-Tu are manifested in the genetic code because a selection pressure towards the avoidance of unfavorable interactions biases the 5'-context of a certain sense or nonsense codon (codon pair bias).[65,69-72] This selective pressure is so strong that certain sense codon - sense codon and sense codon - stop codon pairs (e.g. Pro - amber and Thr - amber) are totally absent from the *E. coli* genome.[69,73,74] In addition to these 5'-structural effects, also the 3'-flanking nucleobase in the mRNA influences the suppression – termination ratio. Although not all studies on this topic are compliant, the general trend is that an adenosine as 3'-base of an amber stop codon highly promotes read-through while uracil supports termination. G and C influence only moderately, but G shows a more positive effect on suppression than C.[71,73,75-77] The rationale behind this observation is that the RF affinity for a stop codon is probably enhanced by specific interactions with a 3'-uracil.[78] In contrast, higher suppression of stop codons followed by a 3'-A nucleobase is most probably due to preferential stacking between this base and the anticodon-codon complex of the decoding tRNA.[67,79]

1.3 Synthetic evolution of the genetic code

In the emerging field of synthetic biology, researchers are seeking to design biology inspired, well characterized complex devices, new metabolic pathways or even completely synthetic cells. One driving force of this emerging field is the need for the biological production of valuable and useful industrial products from renewable resources, in order to substitute petrochemical processes. For this purpose, whole engineered cells, cell lysates or purified enzymes are used, but limited to the chemical possibilities already prescribed by the 22 proteinogenic amino acids. The use of noncanonical amino acids, however, could by large extend the variety of chemical functionalities. These could endow proteins with a vast amount of new conjugation moieties, and give rise to novel functions and folds beyond those available to traditional biochemistry.

To incorporate noncanonical amino acids into proteins, the existing genetic code has to be redefined or recoded. This is necessary since coding units are needed for the ncaa to be introduced. Basically, three different methods exist for the incorporation of ncaas: *in vitro* semi-synthetic incorporation, *in vivo* residue-specific replacement (genetic code engineering), and *in vivo* site-specific incorporation via orthogonal aaRS:tRNA pairs (genetic code expansion). Proteins containing ncaas are generally called variants or alloproteins. If produced by genetic code engineering, the proteins can also be called congeners since they originate from the same gene sequence, but contain a small fraction of amino acids exchanged by the related ncaas. In contrast, for genetic code expansion, target genes have to be mutated prior to incorporation. In the following chapters, mainly the history and state of the art of *in vivo* genetic code engineering and expansion technologies will be discussed. Furthermore, some successful applications of different methods will be explained in more detail.

1.3.1 *In vitro* semi-synthetic incorporation

In vitro semi-synthetic incorporation of ncaas uses chemically or enzymatically misacylated tRNAs. The technique is based on a seminal work of Lipmann and coworkers who incorporated *in vitro* alanine instead of cysteine in response to the UGU codon using misacylated Ala-tRNA^{Cys} generated by chemical reduction of Cys-tRNA^{Cys}.[34] The first *in vitro* incorporation of ncaas into proteins was performed by Chamberlin[80] and Schultz in 1989.[81] To misacylate the amber suppressor tRNA, they used a T4 RNA ligase based approach developed by Hecht.[82] Over the years, this method has been used for the incorporation of numerous ncaas ranging from amino acid analogs like 4-aminophenylalanine[83] to very exotic and bulky compounds like boron-dipyrromethenes (BODIPY).[84] This illustrates not only the flexibility of the ribosome towards a great variety of very different ncaas, but also the big advantage of the approach: *In vitro*, the aaRSs — the genetic code gatekeepers (see chapter 1.2.1) — are bypassed and the mischarged tRNAs are directly recruited by the ribosome. Furthermore, in an *in vitro* system, cell toxicity and impaired uptake of ncaas do not have to be taken into account. In addition to stop codon suppressing tRNAs, the group of Sisido developed quadruplet frame-shift suppressor tRNAs[85] which proved to be highly useful also

for the simultaneous *in vitro* incorporation of multiple different amino acids.[84] In the last 20 years, over 100 different ncaas have been incorporated *in vitro*. However, the method was only used extensively in a few laboratories worldwide, and never found its way to the standard tool-kit of protein engineering. This is mainly due to the laborious preparation of misacylated tRNAs which requires chemistry, as well as biochemistry and molecular biology facilities often not available in one and the same laboratory.[86] Finally, the protein yield in such *in vitro* systems is generally much lower when compared to in *in vivo* systems. A detailed overview of the studies performed by the *in vitro* semi-synthetic incorporation method and the ncaas used therein can be found in the reviews of England, Dougherty and Forster.[87-89]

1.3.2 Genetic code engineering

Genetic code engineering takes advantage of the intrinsic substrate tolerance of endogenous aaRSs towards noncanonical amino acids. Since aaRSs evolved their discrimination mechanisms only in the natural amino acid context, they are not able to distinguish between their natural substrate and structurally related (isostructural) analogs (see chapter 1.2.2) not abundantly available in nature. Indeed, Szostak discovered that over 90 different amino acid analogs are substrates for wild-type aaRSs *in vitro*.[90,91] Ncaa incorporation by genetic code engineering is not considered an expansion but rather a substitution of a canonical amino acid leading to global recoding of a certain sense codon. Yet the canonical amino acid will be preferred as substrate by the cognate aaRS when in competition with an ncaa. Therefore, auxotrophic expression strains — impaired in biosynthesizing a specific amino acid — and controlled fermentation conditions are needed for genetic code engineering. The supplementation-based incorporation method (SPI) provides such conditions. First, cells are grown until mid-log phase, then the amino acid to be exchanged is depleted from the auxotrophic expression strain, the ncaa is added and target protein expression is induced. Cognate amino acid depletion can be achieved by (i) growth in minimal medium which is calibrated such that the respective amino acid is exhausted at mid-log phase[92] or (ii) harvesting, washing and resuspension of the cells in minimal medium lacking the respective amino acid.[93] For high incorporation efficiency, a tightly regulated inducible expression system is crucial, which efficiently represses target protein expression until the canonical amino acid is exchanged by the noncanonical analog in the expression medium.

1.3.2.1 History and field development

Genetic code engineering started already in the 1950s with the seminal work of Cowie and Cohen who performed a proteome-wide replacement of Met by selenomethionine (SeMet) using Met auxotrophic cells grown in minimal medium.[94] SeMet was found to be highly tolerated by the cells and even supported their growth. From today's point of view, their findings can be regarded rather as an exception since most ncaas do not support growth or are even toxic to the cells. The first study on an isolated protein congener was performed by Schlesinger who studied the effects of triazolealanine incorporation on alkaline

phosphatase.[95] However, only with its rediscovery and refinement in the 1990s particularly by Tirrell[93] and Budisa,[92] genetic code engineering by SPI found broader application in protein engineering. Since then, SPI has been used for incorporation of SeMet and ^{19}F-labeled amino acids facilitating structure determination by X-ray crystallography and nuclear magnetic resonance spectroscopy, respectively. Furthermore, the effects of ncaa incorporations on biophysical, spectroscopic, enzymatic and folding properties of target proteins have been extensively studied.[96-102] In many cases, global reside-specific incorporation of ncaas introduced beneficial synergistic effects on protein folding, stability or activity (see chapter 1.3.2.2).

Traditionally, genetic code engineering was limited by the substrate tolerance of the native aaRSs. Many interesting analogs were only very poorly loaded or could not be incorporated at all. Tirrell overcame these limitations by heterologous coexpression of native[103], editing impaired,[104,105] or active site mutated[106,107] endogenous aaRSs. In parallel, he applied these strategies on genetic code expansion (see chapter 1.3.3).

1.3.2.2 Synergistic and combined effects of ncaas on proteins

Various biological phenomena, such as protein folding and stability, arise from synergistic effects of amino acids at multiple positions in a protein structure. Residue-specifically incorporated noncanonical amino acids often act in the same way in a protein structure.

One example is the superior refolding and stability of enhanced green fluorescent protein (EGFP) upon global 4*S*-fluoroproline ((4*S*-F)Pro) incorporation.[98] In this study, (4*S*-F)Pro and (4*R*)-fluoroproline ((4*R*-F)Pro) were incorporated into EGFP but only the congener EGFP[(4*S*-F)Pro] was soluble and fluorescent (see **Figure 7A**). Refolding studies revealed that EGFP[(4*S*-F)Pro] additionally exhibited a significantly enhanced refolding capacity, fluorescence recovery and stability upon chemical denaturation ('Superfolding').

Proline is a crucial amino acid in protein folding processes. Due to its endocyclic nature, the energy barrier of peptide bond *cis*-*trans* isomerization is high and correct allocation within a protein structure often the rate limiting step in protein folding.[108] In addition, the C^{γ}-position of Pro can adopt either *endo* or *exo* pucker conformation as a consequence of its cyclic nature (see **Figure 7B**). Budisa and coworkers demonstrated that this puckering was responsible for the 'superfolding' properties of EGFP[(4*S*-F)Pro]. Nine out of ten prolines in EGFP adopt *endo* pucker conformation (e.g. Pro58, see **Figure 7B**). This conformation is stabilized by (4*S*)-fluorination,[109] which leads to a pre-organization of the prolines' pyrrolidine rings during protein folding. Only Pro56 was found to be in *exo* pucker conformation, whereby (4*S*)-fluorination provoked a repulsive interaction with a carbonyl oxygen (see **Figure 7B**). This negative interaction, however, is overcompensated by the nine positive contributions of the fluorines in *endo* conformation. Taken together, the synergistic action of the globally incorporated (4*S*-F)Pros was responsible for the enhanced folding and stability properties of EGFP[(4*S*-F)Pro].

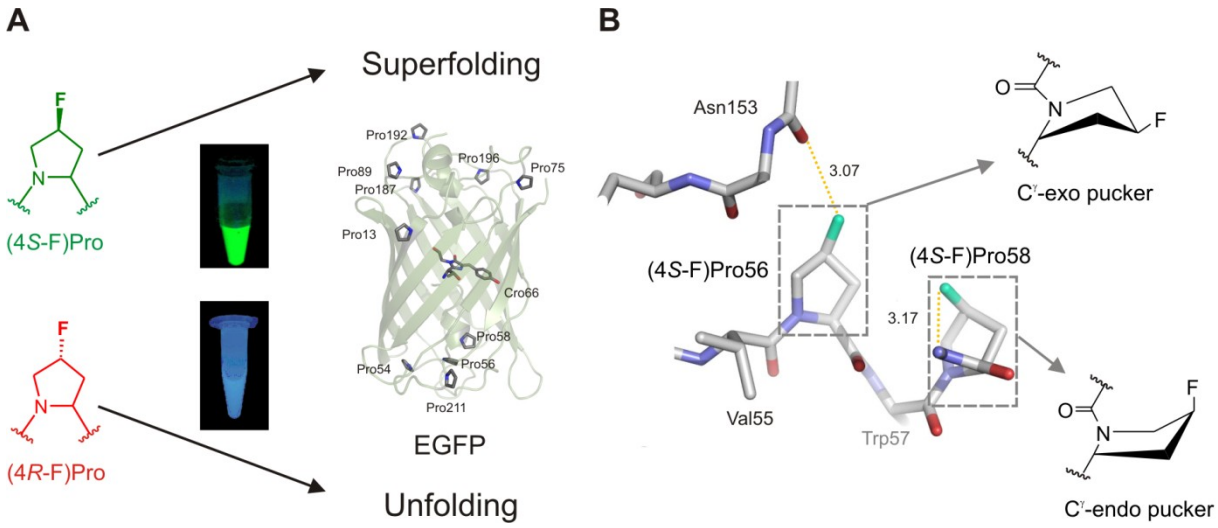


Figure 7. Incorporation of (4*S*-F)Pro yields a 'superfolder' EGFP. A: Superfolding *versus* unfolding state. While the incorporation of 4*R*-fluoroproline ((4*R*-F)Pro) leads to unfolded protein, the incorporation of 4*S*-fluoroproline ((4*S*-F)Pro) yields a 'superfolder' EGFP. **B**: Detail of two kinds of newly generated interactions upon (4*S*-F)Pro incorporation based on the crystal structure of EGFP[(4*S*-F)Pro] (PDB code 2Q6P). Nine out of ten fluorine atoms in EGFP[(4*S*-F)Pro] adopt a C^{γ}-*endo* conformation causing stabilizing interactions between fluorine atoms and other residues (as presented for (4*S*-F)Pro58 and the backbone nitrogen of residue 59). Only one fluorine adopts C^{γ}-*exo* conformation giving rise to a negative interaction ((4*S*-F)Pro56 and Asn153). This negative effect is discarded by the nine positive interactions that act synergistically to enhance EGFP[(4*S*-F)Pro] stability and folding.[98]

In the present study, the synergistic effects provided by genetic code engineering will be combined with genetic code expansion to analyze potentially combinatory effects (see chapter 1.4 and 2.4.2).[110,111]

Another example of such beneficial synergistic effects provided by the global incorporation of ncaas has been discovered in *Thermoanaerobacter thermohydrosulfuricus* lipase (TTL, see **Figure 8A**). Different ncaas, including Pro, Met and Phe analogs were incorporated into TTL and the effects on enzyme activity, stability and tolerance towards different additives were studied.[99,100] Among other things, it was discovered that incorporation of norleucine (Nle) generated a TTL congener that was, in contrast to its parent protein, highly active in the aqueous phase without the need for thermal activation (see **Figure 8B**).

The need for interfacial or thermal activation is a typical feature of many lipases in aqueous solution.[112] To be active, the active center has to adopt a conformation which is accessible to the hydrophobic substrate. TTL has a flexible lid domain (see **Figure 8A**) which closes the access to the active site. Most probably, parent TTL is only highly active when this lid is opened upon thermal activation.[99] Strikingly, TTL[Nle] exhibits already high activity without thermal activation. This is most likely attributed to an altered hydrophobicity/hydrophilicity balance of the lid domain upon Nle incorporation resulting in an 'always open' conformation. An altered hydrophobicity at the dimerization interface of TTL (see **Figure 35B** in results and discussion) might additionally play a role giving rise to a coupled network of beneficial effects that promote enzymatic activity.[112,113]

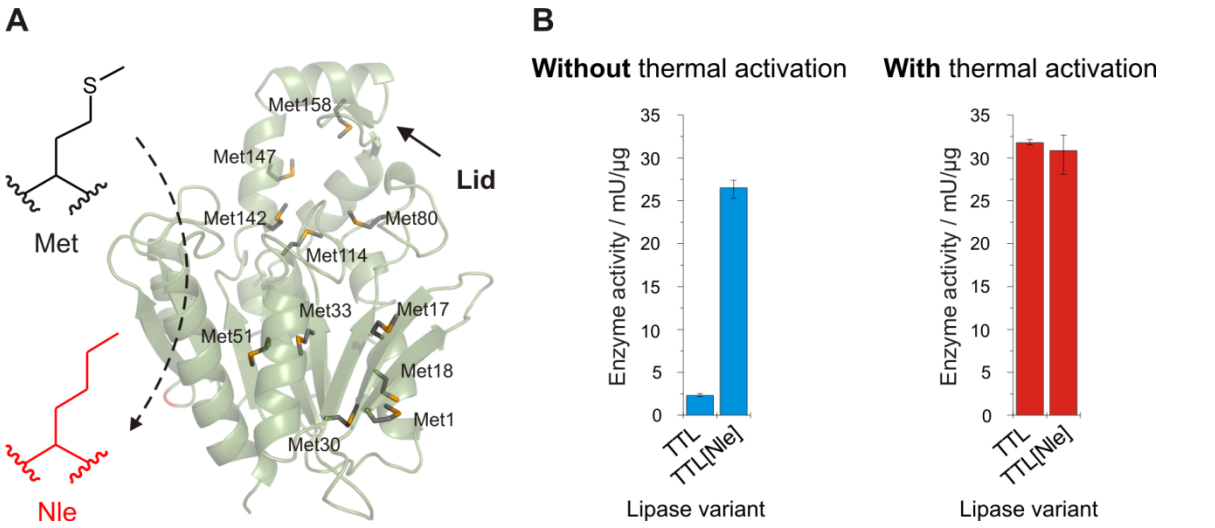


Figure 8. Effect of norleucine incorporation on *Thermoanaerobacter thermohydrosulfuricus* lipase (TTL). **A**: Modeled structure of TTL[99] with Met residues highlighted throughout the TTL structure. The active site is covered by a flexible 'lid' domain B: TTL activity of parent TTL and TTL[Nle] with and without heat activation.

The possibility of provoking beneficial synergistic effects in proteins is not limited to protein-wide incorporation of one single ncaa. Budisa and coworkers recently showed that desirable features of different ncaas can be successfully combined in one target protein by SPI of various different ncaas simultaneously.[114] In previous studies, homopropargylglycine (Hpg) or azidohomoalanine (Aha) containing cysteine-free pseudo wild-type barstar (ψ-b*)[115] had been used for individualized modifications through copper(I)-catalyzed azide–alkyne Huisgen cycloaddition (CuAAC)[116,117] with azide or alkyne-containing ligands (see **Figure 9**).[118] Independently, the isostructural Trp analog 4-azatryptophan (4AzaTrp) had been introduced as a noninvasive probe for biological imaging studies due to its strong blue fluorescence.[97] To combine both features in ψ-b* without loss of protein stability, Hpg and 4AzaTrp were coincorporated with (4*S*-F)Pro because its stabilizing effect on ψ-b* was already known.[119] In this way, a bioorthogonal handle was combined with an imaging probe in a stable ψ-b*. The reactive handle was used for posttranslational conjugation with a sugar moiety.[120]

The successful combination of these three features by genetic code engineering, however, most probably was possible because the parent ψ-b* contains only one Met, one Pro and three Trps. This makes the global residue-specific incorporation rather site-specific. In the present study, genetic code engineering will be conjoined with genetic code expansion to be able to combine the described synergistic effects of ncaas at multiple positions with real site-specific incorporation of chemical handles.

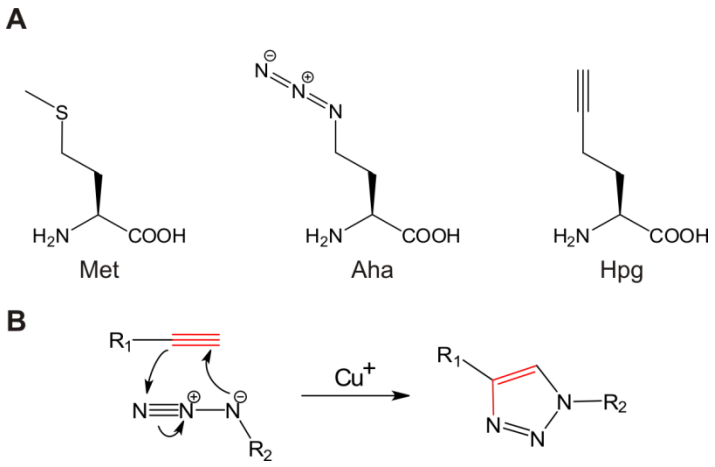


Figure 9: Methionine analogs used for copper(I)-catalyzed azide–alkyne Huisgen cycloaddition (CuAAC). **A**: Structures of methionine (Met), azidohomoalanine (Aha) and homopropargylglycine (Hpg). Aha and Hpg bear azide and alkyne moieties, respectively. **B**: The Cu^{+} catalyzed [3+2] cycloaddition takes place at physiological conditions and is highly regio-specific, i.e. only the 1, 4-regioisomer of the [1,2,3]-triazole product is formed.[116,117]

1.3.3 Genetic code expansion

For genetic code expansion, ncaas are not incorporated by replacing caas but rather by an *in vivo* addition to the standard amino acid pool. This is realized by the decoding of nonsense codons with aaRS:tRNA pairs. The minimum requirements for such a system are: (i) a $tRNA_{Sup}$ that efficiently reads the nonsense codon of choice; (ii) the $tRNA_{Sup}$ must not be recognized by the endogenous host aaRSs; and (iii) an aaRS that specifically aminoacylates the cognate $tRNA_{Sup}$ but not the endogenous host tRNAs. This crucial lack of cross-reactivity between the alien aaRS:$tRNA_{Sup}$ pair and the endogenous host components is known as orthogonality. Lack of orthogonality is the main problem when using aaRS:$tRNA_{Sup}$ pairs for ncaa incorporation *in vivo*.

1.3.3.1 Orthogonal aaRS:$tRNA_{Sup}$ pairs

In 1997, Schultz and coworkers published efforts to make an endogenous *E. coli* GlnRS:$tRNA^{Gln}_{CUA}$ pair orthogonal to the host system by rational design and directed evolution.[121,122] A $tRNA^{Gln}_{CUA}$ was designed that was not anymore recognized by the endogenous GlnRS but efficiently aminoacylated by the mutant GlnRS. However, Schultz never succeeded in evolving a mutant GlnRS not anymore capable of recognizing the endogenous host $tRNA^{Gln}$. This result suggested that it would not be possible to completely overcome the specific aaRS recognition of the phylogenetically conserved identity elements of the tRNAs (see chapter 1.2.3).[86] A successful strategy to circumvent the problem was discovered by Furter in 1998. He used a *Saccharomyces cerevisiae* (*S. cerevisiae*, *sc*) PheRS:$tRNA^{Phe}_{CUA}$ pair in *E. coli* to site-specifically incorporate 4-fluorophenylalanine. Thereby, he took advantage of the phylogenetic differences in the identity elements used for aaRS-tRNA recognition (see chapter 2.1.2 for more information on this orthogonal pair).[56] Schultz followed this approach and successfully established an orthogonal

*sc*GlnRS:*sc*tRNA$^{Gln}_{CUA}$ pair.[123] To date, ten orthogonal pairs for use in *E. coli* have been described (see **Table 1**). Most of them include class I/IIc aaRSs (PheRS, TyrRS, TrpRS, PylRS). This is most probably due to the fact that aromatic amino acids are much more efficiently incorporated with suppressor tRNAs than hydrophilic ones.[124]

Table 1. Orthogonal aaRS:tRNA$_{Sup}$ pairs for use in *E. coli*. Some orthogonal pairs are hybrids of aaRS and tRNAs form different species. The components are labeled accordingly. When tRNAs are changed at other positions different from the anticodon, they are designated as 'mutated'.

aaRS	Origin	tRNA	Codon	Reference
PheRS	*S. cerevisiae*	tRNA^{Phe}, mutated	UAG	Furter[56] and Tirrell[125]
TyrRS	*S. cerevisiae*	tRNA^{Tyr}	UAG	Nishikawa[126]
GlnRS	*S. cerevisiae*	tRNA^{Gln}	UAG	Schultz[123]
AspRS	*S. cerevisiae*	tRNA^{Asp}	UAG	Schultz[127]
TyrRS	*Methanocaldococcus jannaschii*	tRNA^{Tyr}, mutated	UAG	Schultz[128]
*sc*TyrRS	*S. cerevisiae* and *E. coli*	*ec*tRNAi^{Met}, mutated	UAG	RajBhandary[129]
*mt*LeuRS	*Methanobacterium thermoautotrophicum* and *Halobacterium sp. NRC-1*	NRC-1 tRNA^{Leu}	UAG, UGA, AGGA	Schultz[130]
GluRS	*Pyrococcus horikoshii*	Archaeal consensus tRNA^{Glu}	UAG	Schultz[131]
LysRS	*Pyrococcus horikoshii*	Archaeal consensus tRNA^{Lys}	AGGA	Schultz[132]
PylRS	*Methanosarcina barkeri*	tRNA^{Pyl}	UAG	Krzycki[133]
PylRS	*Methanosarcina mazei*	tRNA^{Pyl}	UAG, UAA, UGA, UAGA[134]	Sakamoto and Yokoyama[135]
PylRS	*Desulfitobacterium hafniense*	tRNA^{Pyl}	UAG	Söll and Nureki[136]
TrpRS	*S. cerevisiae*	tRNA^{Trp}, mutated	UAG	Ellington[137]

To date, only *mj*TyrRS and different PylRS pairs have gained importance. Especially, PylRS:tRNA^{Pyl} pairs have brought a the real breakthrough since they are 'naturally evolved' and thus *per se* orthogonal and highly efficient in suppression. In addition, tRNA^{Pyl} efficiently suppresses a large variety of codons without tedious tRNA engineering (see **Table 1**) and already recognizes different Pyl and Lys analogs — but none of the canonical amino acids — due to its intrinsic substrate tolerance.[138]

In contrast, the *mj*TyrRS:*mj*tRNA$^{Tyr}_{CUA}$ pair had to be greatly engineered to be considered as a real orthogonal pair and efficient in incorporating ncaas. First, *mj*tRNA^{Tyr} was mutated to become more orthogonal because it was still recognized by the endogenous *E. coli* aaRSs to some extent. A tRNA library was constructed where non-conserved positions in the D-, T- and anticodon-stem-loop were mutated,[139] and the library was screened by positive and negative selection, subsequently (*vide infra*).[140] A negative selection removed tRNAs that were aminoacylated by endogenous aaRS, whereas a positive selection yielded tRNAs that were efficiently aminoacylated by *mj*TyrRS. The result was a completely orthogonal *mj*TyrRS:*mj*tRNA$^{Tyr}_{CUA}$ pair. Second, a *mj*TyrRS amino acid binding pocket library[123,141] was screened by positive and negative selection for mutants that did not anymore recognize its original cognate amino acid Tyr and adopted a new specificity towards the ncaa of choice. The selection works as follows: The positive selection is based on a β-lactamase or chloramphenicol acetyl transferase gene containing a stop codon. Cells are grown in the presence of all canonical amino acids and the ncaa of choice. If the aaRS is capable of aminoacylating the tRNA$_{Sup}$ with any amino acid (canonical or noncanonical), full-length protein is produced providing antibiotic resistance. Positive clones actively aminoacylate the tRNA$_{Sup}$, but do not necessarily exhibit an abolished Tyr recognition. To achieve this, negative selection is conducted in absence of the ncaa with the toxic barnase gene harboring two or three stop codons. Herein, only cells survive which contain *mj*TyrRS:*mj*tRNA$^{Tyr}_{CUA}$ pairs not capable of incorporating canonical amino acids because any caa-tRNA$_{Sup}$ will lead to cell death since cells are killed already at very low concentrations of full-length barnase. The remaining aaRSs actively aminoacylate *mj*tRNA$^{Tyr}_{CUA}$, do not recognize caas, and have new specificity for the ncaa of choice. Using the described screening strategy, over 40 different *mj*TyrRS:*mj*tRNA$^{Tyr}_{CUA}$ pairs for incorporation of different ncaas have been evolved.[142] The incorporated amino acids provide a great spectrum of different side chains, for example, selective coupling functionalities (e.g. 4-azidophenylalanine[143]), photo-reactive handles (e.g. 4-benzoylphenylalanine,[144] see chapter 2.1.1) or metal chelators (e.g. (2,2'-bipyridin-5-yl)alanine, BpyAla[145]). In addition, the described screen has been applied to PylRS, for instance, to select mutants capable of cotranslationally incorporating the posttranslational modification 6-*N*-acetyllysine.[146,147]

Although the positive and negative selection system has proved to be highly useful, a 'dominant negative' screen like the barnase system is not always the method of choice because it can be too stringent for successful isolation of aaRS clones for difficult ncaas in the first place. This is nicely demonstrated in the example of BpyAla. Direct screening for *mj*TyrRS mutants capable of incorporating BpyAla did not yield any positive clones. However, after 'prescreening' the library for incorporation of its hydrophobic isosteric analog biphenylalanine, a mutant capable of incorporation BpyAla could be isolated. To circumvent this highly time consuming procedure, Hughes,[86] and Schultz and coworkers[148] recently designed a tunable selection system on the basis of uracil phosphoribosyl transferase to vary selection pressure.

Despite the sophisticated selection process, many evolved synthetases still are bad enzymes, which greatly limits their use in potential applications.[149] Therefore, Skerra and coworkers recently published an aaRS refinement strategy based on fluorescence activated cell sorting (FACS), which highly improved the incorporation efficiency of a *mj*TyrRS derived orthogonal pair incorporating *O*-methyltyrosine.[150] In addition, depending on the aaRS:tRNA pair used, a mutation in the tRNA anticodon can lead to tremendous decrease in tRNA acceptor activity because aaRSs often use the anticodon as identity element (see chapter 1.2.3). In case of *mj*TyrRS:*mj*tRNA$^{Tyr}_{CUA}$, for example, the GUA → CUA anticodon mutation led to a 300fold drop-down in acceptor activity.[151] To compensate this, *mj*TyrRS was readapted to the new anticodon and thereby, its acceptor activity was partially restored (discussed in detail in chapter 2.2.3.3).[55] Another interaction which has an influence on incorporation efficiency is the binding of aminoacyl-tRNA to EF-Tu (see chapter 1.2.1). The canonical aminoacyl-tRNAs bind to EF-Tu with nearly the same affinity. Ncaa-tRNA$_{Sup}$ binding, however, is often highly impaired.[152] Therefore, Schultz optimized the *mj*tRNA$^{Tyr}_{CUA}$ in every orthogonal pair with respect to EF-Tu binding of the corresponding ncaa-tRNA.[153] These and other optimization efforts made it possible only recently to simultaneously incorporate two different noncanonical amino acids into one protein by the use of two different orthogonal pairs.[134,154] A detailed overview and discussion of this currently most advanced genetic code expansion studies can be found elsewhere.[155] In spite of these improvements, however, the biggest problem of the genetic code expansion technology remains in the competition between the suppressor tRNA and the release factor for the stop codon (see chapter 1.2.4).

1.3.3.2 Engineering of the translational apparatus

Therefore, also other parts of the translational apparatus have been targeted in order to improve stop codon suppression. Chin and coworkers designed orthogonal ribosomes with reduced affinity to RF1. In the first step, orthogonally operating ribosome – mRNA pairs were developed by screening libraries of mutated Shine-Dalgarno (mRNA) and anti-Shine-Dalgrano sequences (16S rRNA) for pairs which highly interacted with each other but not longer with the endogenous host translation system. Thereby, they could separate the synthesis of the cellular proteome from target mRNA translation.[156] In the second step, they mutated crucial A-site nucleobases located in the 16S rRNA such that the interaction between the large subunit of the ribosome and RF1 was highly reduced.[157] The result was an enhanced amber stop codon suppression in the target mRNA. An analog effect was obtained in a study with the ribosomal protein L11. This protein plays an important role in the RF1-mediated polypeptide release from the ribosome and overexpression of its C-terminal domain led to significantly higher suppression.[158]

1.3.3.3 Vector systems for coexpression of orthogonal pairs

Beyond the intrinsic properties of all involved translation components, the coexpression systems for aaRS:tRNA pairs also have a high impact on suppression efficiency and overall yield of labeled protein. Many different systems based on one or two plasmids are used for the coexpression of aaRS:tRNA pairs. Two variants of two-vector-systems exist: one where

the tRNA gene is located on one plasmid, and the aaRS and target gene on the other (see **Figure 12**);[56] and a second one where the tRNA and aaRS genes are located on the same vector (denoted as tandem vector) but the target protein on another plasmid (see **Figure 11**).[159] One-vector-systems have been described in the literature with all three components on the same plasmid.[148,150,160] In general, the experimental process of ncaa incorporation by stop codon suppression is divided into two steps. First, permissive stop codon sites for efficient read-through have to be identified in the target gene (see chapter 2.4.3.1). Second, the target protein containing the ncaa at the selected stop codon position(s) has to be expressed in high yields (see chapter 2.4.3.2 and 2.4.3.3).

For the first step, it is very convenient to have a plasmid which contains both the aaRS and the tRNA on a single vector because this plasmid can then be easily combined with standard expression vectors encoding the target protein for permissive site screening. Since most standard expression vectors (particularly pET, pGEX and pQE[161]) contain the high copy ColE1/pMB1 origin of replication[162,163] in combination with an ampicillin (Amp) resistance gene, the tandem vector needs an origin of replication (Ori) from another compatibility group plus another antibiotic resistance gene. In *E. coli*, most vectors for cotransformation use the p15A Ori,[164] and a chloramphenicol (Cm) resistance gene. RSF,[165] ColA,[166] or F plasmid derived origins of replication like in pSCANS[167] are also compatible with ColE1/pMB1.

In the second step, the highest possible yield of full-length protein is to be achieved. For this purpose, an one-vector-system including all necessary expression components is more convenient since high amounts of aaRS and tRNA are required to guarantee satisfactory stop codon suppression and consequently high yield of full-length protein (see discussion in chapter 2.2.3). Such high yields can be more easily obtained if the different genes are transcribed from high copy plasmids. Since no Ori exists today that has a comparable copy number and compatibility to ColE1/pMB1 based plasmids, all components have to be combined on a single vector with this Ori. Furthermore, maintenance of two plasmids in one cell is consuming cellular resources which could otherwise be used for protein expression.

The already existing tandem plasmids and one-vector-systems for SCS experiments have several disadvantages. Mostly, they do not contain various multiple cloning sites (MCS) for convenient cloning of the different expression parts into the vector backbone. Moreover, they are only suitable for traditional ligation-based cloning procedures. In particular, this restricts the exchange of promoter sequences. These limitations make the assembly and introduction of new components time consuming and work intensive. Additionally, for historical reasons many current vectors contain relic sequences which are not needed for their proper functionality and even may have disturbing effects in some cases. For example, the pQE80L vectors contain a non-functional Cm resistance gene. Although not containing a promoter it can provide basal Cm resistance in certain genetic contexts.[168]

1.3.4 Comparison of synthetic evolution methods of the genetic code

The *in vitro* semi-synthetic incorporation approach is clearly the most versatile with respect to ncaas which can be incorporated. It bypasses the aaRS aminoacylation by directly introducing misacylated tRNAs and thereby is only limited by aminoacyl-tRNA discrimination at the ribosome. In addition, cellular uptake, toxicity and metabolic stability of ncaas — critical points concerning the *in vivo* methods — do not play a role *in vitro*. The biggest drawback for the general use of *in vitro* methods, however, is the usually low protein amount produced. In this respect, *in vivo* methods have clear advantages. In addition, for most of the relevant ncaas cellular uptake is also warranted *in vivo* since the amino acid transporters of the cells are relatively nonspecific.[169] In mammalian cells, it has been shown that certain ncaas are highly accumulated in the cells.[170] It is currently under investigation, whether this also holds true for bacterial systems.

The power of genetic code expansion is its site-specificity, orthogonality and easy handling. New chemical functionalities can be precisely introduced into proteins with specifically evolved orthogonal pairs without the need of special fermentation strategies. However, this method also has several drawbacks: (i) the competence between the release factor and the $tRNA_{Sup}$ still impede the suppression of more than three stop codons in one mRNA sequence. The final overall suppression rate decreases multiplicatively as the number of nonsense codons in an mRNA increases;[158,160] (ii) for every ncaa, a new orthogonal aaRS:$tRNA_{Sup}$ pair has to be evolved which is highly laborious and time consuming; and (iii) code expansion is generally not applicable to isostructural ncaas as it relies on the use of orthogonal aaRSs, which are developed by positive and negative selection (see chapter 1.3.3.1). These screens are not capable of isolating enzymes which discriminate efficiently between 'atomic mutations' like -H/-F, -H/-CH_3, -S-/-CH_2- or -CH=/-N=.[97] With orthogonal pairs, such small modifications can only be cotranslationally introduced when they are 'masked' with bulkier protective groups. These can be posttranslationally cleaved off.[171]

In contrast, the genetic code engineering approach is capable of introducing isostructural amino acids because the natural substrate tolerance of the endogenous aaRSs is exploited. Moreover, a high number of ncaas can be incorporated into a target protein without tremendous loss of protein yield.[99] This allows introducing coupled networks of multiple ncaas which can give rise to synergistic beneficial effects (see chapter 1.3.2.2). Furthermore, the technique does not require extensive engineering of orthogonal aaRS:$tRNA_{Sup}$ pairs and site-directed mutagenesis of target genes before use. However, genetic code engineering does not provide site-specific but residue-specific incorporation. This also implies that ncaas are not included into the standard amino acid repertoire but rather substitute a canonical amino acid. Homogeneous protein preparations can only be achieved if the fermentation conditions are tightly controlled. A comparison between the two main methods of synthetic genetic code evolution is depicted in **Figure 10**.

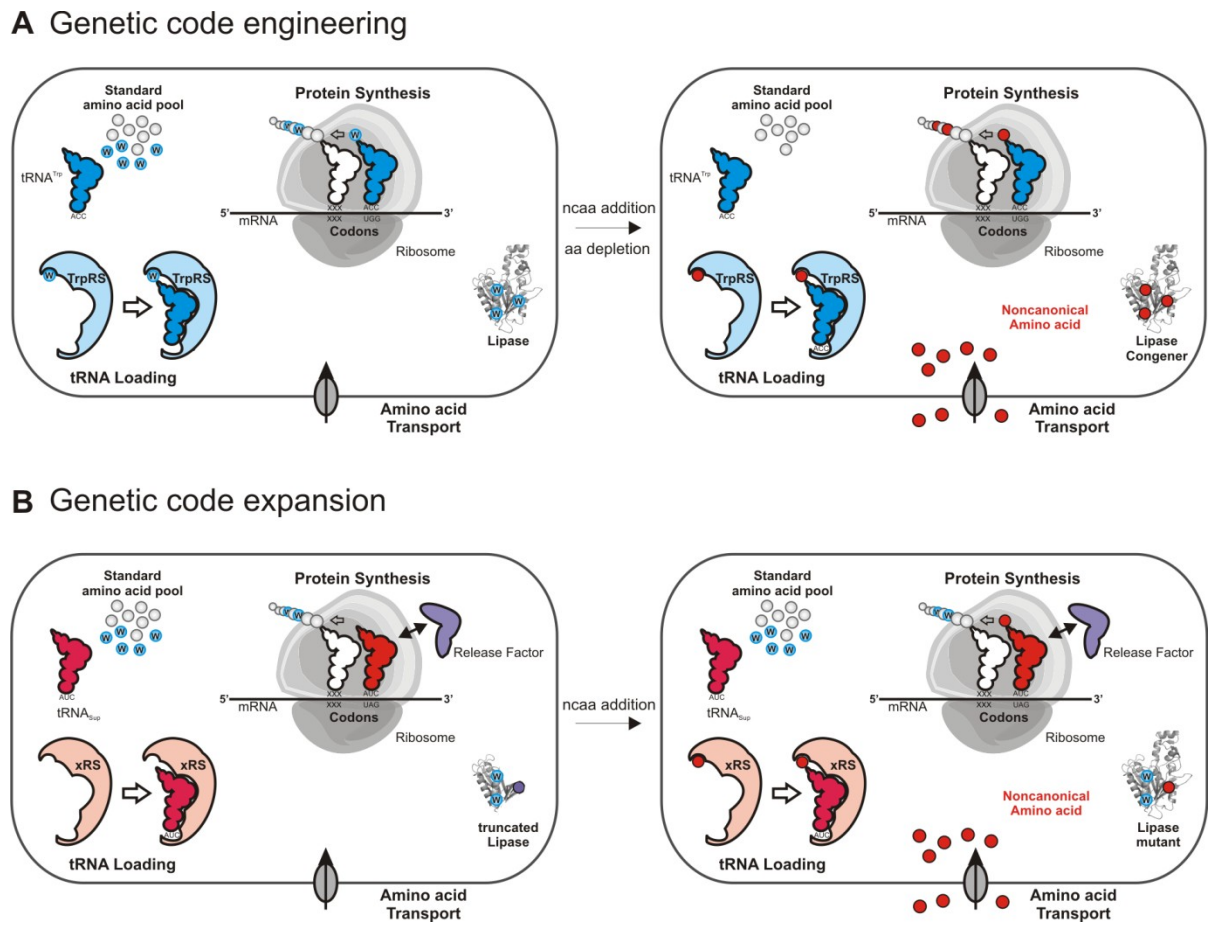


Figure 10. Comparison of genetic code engineering (A) and genetic code expansion (B). A: Genetic code engineering does not require engineering of translational components. It is realized by the supplementation-based incorporation method (SPI) in which a specific ncaa is incorporated in response to sense codons in a residue-specific manner. Fermentation conditions have to be tightly controlled because the canonical amino acid whose codons are used for ncaa incorporation has to be depleted from the cell prior to induction of target protein expression. **B**: Genetic code expansion is mostly achieved by stop codon suppression (SCS). Therefore, orthogonal aaRS:$tRNA_{Sup}$ pairs have to be evolved for each ncaa to be introduced. The method is site-specific but limited by the competition between suppression ($tRNA_{Sup}$ mediated) and termination (RF mediated). Termination events at the gene internal stop codons highly decrease the overall yield of full-length protein in favor of truncated protein species. The final full-length protein amount decreases multiplicatively as the number of nonsense codons in an mRNA increases.

1.4 Aim of this study: Expanding the toolkit of protein engineering

The main goal of the present study is the combination of genetic code engineering and expansion in a single expression experiment in *Escherichia coli*. This should be achieved because both methods are in several respects mutually complementary. Most importantly, genetic code engineering introduces coupled networks of multiple ncaas which can give rise to synergistic beneficial effects while genetic code expansion provides site-specific labeling with reactive handles. EGFP, TTL, and ψ-b* are used as target proteins since previous studies already revealed beneficial synergistic effects of noncanonical amino acids incorporated residue-specifically by genetic code engineering. To combine these features with the genetic code expansion approach, first, two established orthogonal aaRS:$tRNA_{CUA}$ pairs are characterized with respect to substrate tolerance and suppression efficiency. In this way, the best performing orthogonal pair can be chosen for the combination of sense codon reassignment and stop codon suppression.

To date, site-specific incorporation methods mainly exploit stop codons to incorporate noncanonical amino acids. The biggest problem of these systems is the competition between the suppressor tRNA and the release factor. Therefore, another objective of this study is, to make a basic survey whether the rare sense codon AGG can be used for incorporation of noncanonical amino acids via an orthogonal aaRS:tRNA pair. To this end, one of the characterized aaRS:$tRNA_{Sup}$ pairs is mutated in the anticodon to decode AGG. Expression experiments in an *Escherichia coli* strain without special genetic background are performed to evaluate whether the rare arginine codons can be at least partially reassigned.

All site-specific incorporation approaches need efficient and versatile plasmid systems for coexpression of orthogonal aaRS:tRNA pairs. To enhance the efficiency of stop codon suppression and site-specific sense codon reassignment, a new modular plasmid system is developed. The system should likewise facilitate the implementation and evaluation of new orthogonal aaRS:tRNA pairs and provide a basis for simultaneous site-specific incorporation of several noncanonical amino acids via different orthogonal aaRS:tRNA pairs, together with residue-specific substitution of canonical amino acids by means of genetic code engineering.

2 Results and Discussion

2.1 The aminoacyl-tRNA synthetase:suppressor tRNA pairs

2.1.1 *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$

The *Methanocaldococcus jannaschii* BpaRS:tRNA$^{Tyr}_{CUA}$ system is derived from *mj*TyrRS:*mj*tRNA$^{Tyr}_{GUA}$. The tRNA was mutated at nonconserved positions to generate robust orthogonality in *E. coli* (see **Figure 11A**).[139] *Mj*BpaRS is capable of aminoacylating *mj*tRNA$^{Tyr}_{CUA}$ with 4-benzoylphenylalanine (Bpa, see **Figure 13**). The enzyme was evolved by negative and positive selection in presence of Bpa, yielding the mutations indicated in **Figure 11B**.[144] Both parts of the orthogonal pair are located on the tandem vector pSUP-BpaRS-6TRN (see **Figure 11C**).[159] The target protein is expressed from a second, pQE80L decendant plasmid. The BpaRS system is one of the best established *mj*TyrRS derived orthogonal pairs. For example, it was used for site-directed photo-coupling of proteins onto surfaces coated with β-cyclodextrins[172] as well as for studying K^+ channel inactivation.[173]

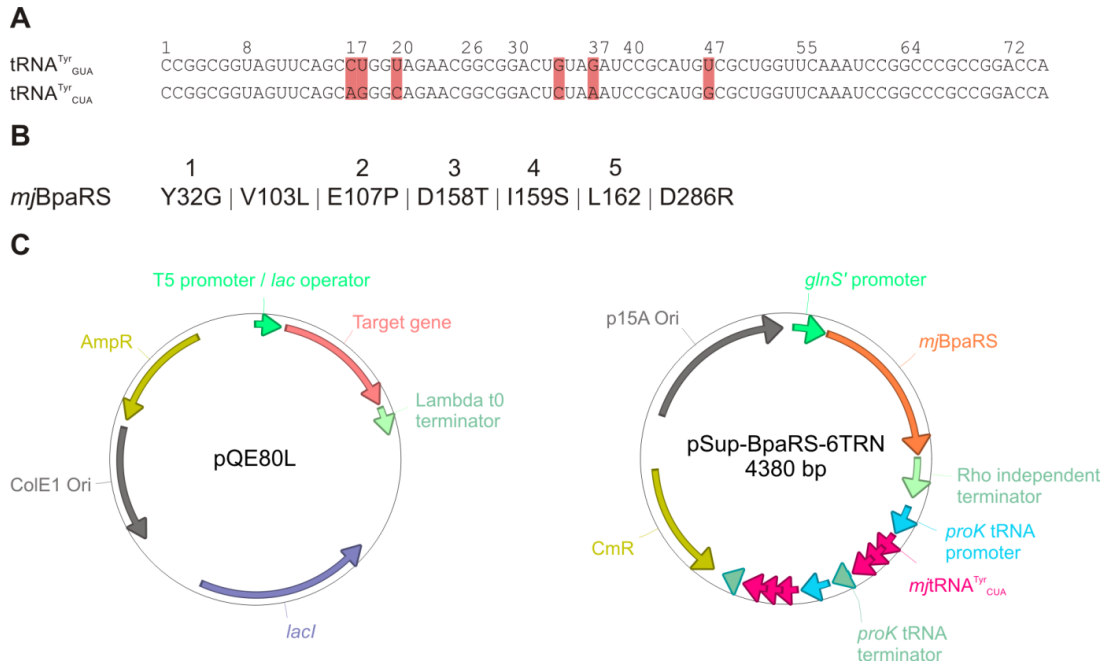


Figure 11. The orthogonal *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ pair. A: Comparison between wild-type *mj*tRNA$^{Tyr}_{GUA}$ and the evolved *mj*tRNA$^{Tyr}_{CUA}$. Mutated nucleobases are highlighted in red. **B**: Introduced mutations into *mj*TyrRS to make the synthetase specific for Bpa. The mutant library for the positive and negative selection was randomized at the positions labeled with 1 – 5.[144] While the mutation D286R was introduced to enhance tRNA recognition,[55] V103L was discovered in this study after sequencing and is not documented in the literature. Position 103 is outside of the amino acid binding pocket and the mutation Val → Leu is very conservative. Thus, it can be assumed that it does not alter the enzyme action. **C**: Two-plasmid-system for site-specific incorporation of Bpa into target proteins. pQE80L carries the target gene, while pSUP-BpaRS-6TRN provides coexpression of the *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ pair. *Mj*BpaRS expression is controlled by the strong constitutive glutaminyl-tRNA synthetase promoter *glnS'*. For high suppression efficiency, the plasmid harbors six copies of *mj*tRNA$^{Tyr}_{CUA}$ in two triple tRNA arrays under the control of the strong constitutive proline tRNA promoter *proK*.

2.1.2 *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$

The *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$ pair is a modified version of the first orthogonal pair designed by Furter (see chapter 1.3.3.1).[56] In contrast to *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ all mutations in the tRNA and aaRS were rationally designed. The original version suffered from Lys misincorporation because the uridine at position 35 is an important identity element for *E. coli* LysRS. Hence, Furter mutated the tRNA at position 37 (see **Figure 12A**), thereby increasing translational efficiency and reducing Lys misincorporation to 5%.[56] Tirrell and coworkers reported to have removed the residual Lys misacylation by additionally introducing the wobble pair 30U-40G in the anticodon-stem of *sc*tRNA$^{Phe}_{CUA}$.[125] In contrast to *mj*BpaRS, *sc*PheRS(T415G) contains only one single mutation which enlarges the amino acid binding pocket. This mutation gives rise to a highly substrate tolerant enzyme capable of incorporating a variety of Phe and Trp analogs.[174] Similar to the *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ pair, a two-plasmid-system is used for site-specific incorporation of ncaas in response to an amber stop codon. Here, however, the aaRS gene is cloned into the same plasmid as the target gene but the tRNA is translated from a separate plasmid (see **Figure 12B**). On the original expression construct of Tirrell and coworkers, the target protein is a *murine* dihydrofolate reductase (*m*DHFR) with an amber stop codon mutation at position 38. They used the *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$ pair to generate a wide variety of enhanced cyan fluorescent protein (ECFP) variants by the incorporation Phe and Trp analogs into the chromophore.[174]

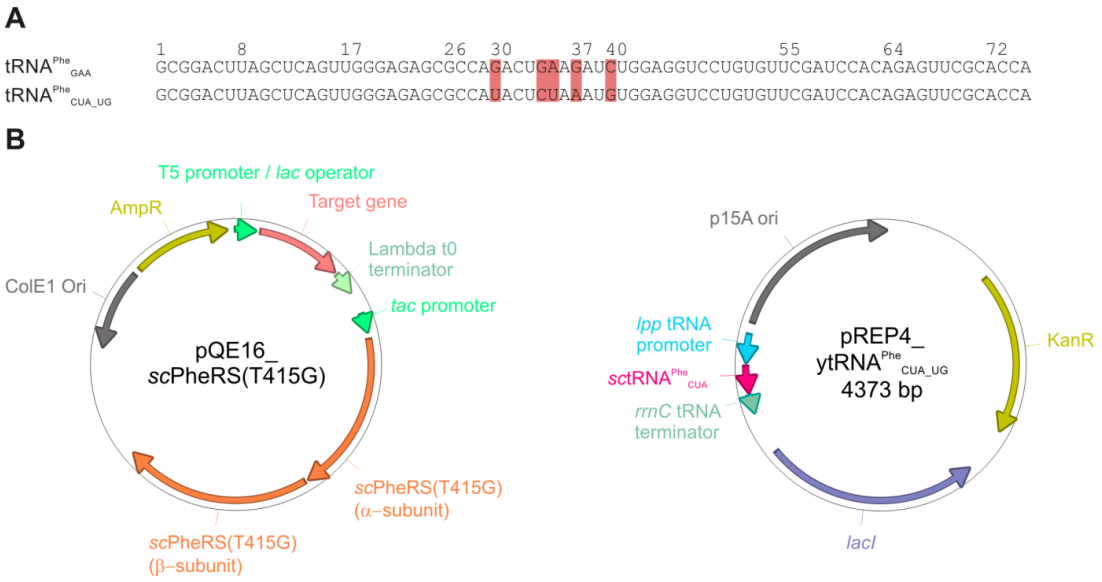


Figure 12. The orthogonal *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$ pair. A: Comparison between the wild-type *sc*tRNA$^{Phe}_{GAA}$ sequence and the mutated *sc*tRNA$^{Phe}_{CUA}$. Mutated nucleobases are highlighted in red. **B**: Two-plasmid-system for site-specific incorporation of Phe and Trp analogs into target proteins. pQE16_*sc*PheRS(T415G) carries the genes for target protein and *sc*PheRS(T415G) expression, while pREP4_ytRNA$^{Phe}_{CUA_UG}$ provides coexpression of *sc*tRNA$^{Phe}_{CUA_UG}$. The strong constitutive *tac* promoter, which is a regulation deprived β-galactosidase promoter, controls the expression of both subunits of *sc*PheRS(T415G). The tRNA is constitutively expressed from the strong lipopolyprotein promoter (*lpp*).

2.2 Characterization of *mj*BpaRS and *sc*PheRS(T415G)

Tirrell and coworkers showed that *sc*PheRS(T415G) exhibits a broad substrate tolerance by studying the ability to activate and incorporate 3-benzothienylalanine (BT), 6-chlorotryptophan (6ClTrp), 6-bromotryptophan (6BrTrp), 5-bromotryptophan, or 3-(2-naphthyl)alanine (Nal) into *m*DHFR, respectively.[174,175] In contrast, *mj*BpaRS was evolved by positive and negative selection[144] (see chapter 2.1.1 for detailed information on mutated residues) to prohibit the recognition of any canonical amino acid and was said to only recognize 4-benzoylphenylalanine (Bpa); thus, making it a counterpart to *sc*PheRS(T415G). However, it was unclear whether the exclusive incorporation would work since recent reports demonstrated that other evolved synthetases (e.g. for incorporation of *N*-acetylgalactosamine-α-*O*-threonine) from the same laboratory did not work at all or not as efficient as reported.[176-178] To explore if the mutated synthetases *sc*PheRS(T415G) and *mj*BpaRS were capable of incorporating the reported amino acids and to investigate as well other ncaas, *in vitro* activation and *in vivo* incorporation experiments were performed. The canonical and noncanonical amino acids used in this study are depicted in **Figure 13**.

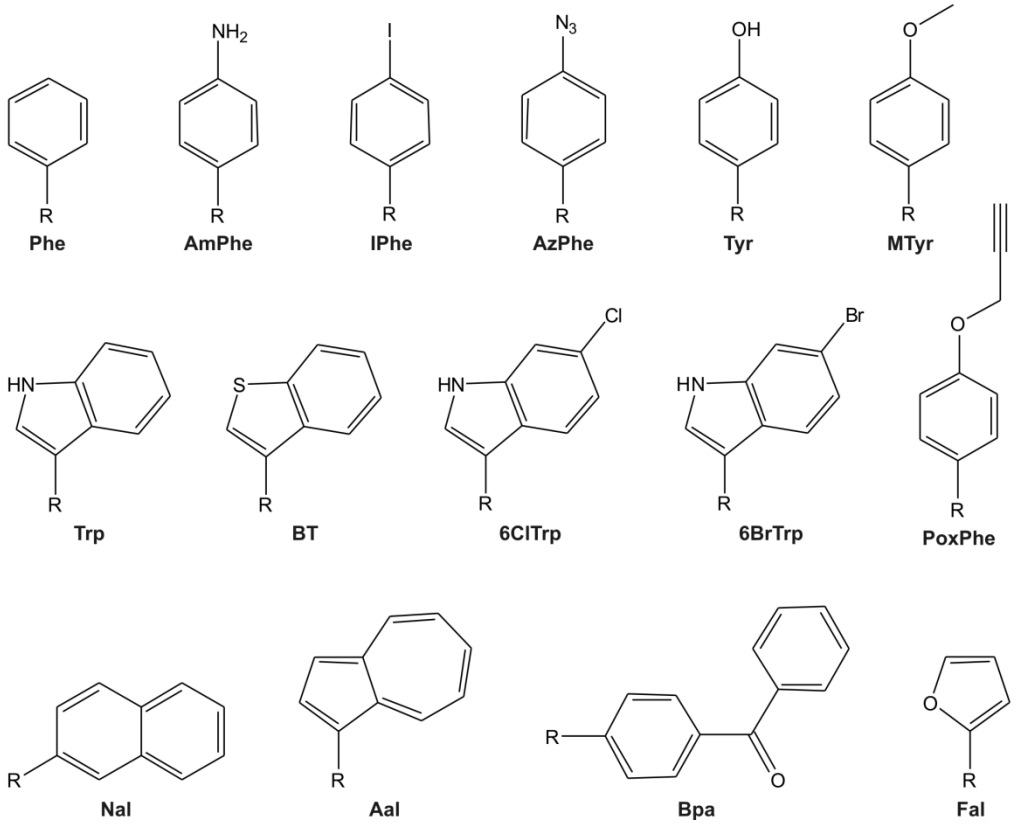


Figure 13. Chemical structures of canonical and noncanonical amino acids used to characterize *sc*PheRS(T415G) and *mj*BpaRS. Phe: phenylalanine; IPhe: 4-iodophenylalanine; AmPhe: 4-aminophenylalanine; AzPhe: 4-azidophenylalanine; Tyr: tyrosine; MTyr: *O*-methyltyrosine; PoxPhe: 4-propargyloxyphenylalanine; Trp: tryptophan; BT: 3-(3-benzothienyl)alanine; 6ClTrp: 6-chlorotryptophan; 6BrTrp: 6-bromotryptophan; Nal: 3-(2-naphthyl)alanine; Aal: 3-(1-azulenyl)alanine; Bpa: 4-Benzoylphenylalanine and Fal: 3-(2-furyl)alanine. R = $-CH_2CH(COOH)NH_2$.

2.2.1 *In vitro* analysis of noncanonical amino acid activation

*Sc*PheRS(T415G) and *mj*BpaRS were expressed heterologously as N-terminally His-tagged proteins in *E. coli* B834{pQE80L-H6-*sc*PheRS(T415G)} and *E. coli* B834{pET28a-H6-BpaRS}, respectively (see chapter 5.3.1). Enzymes were purified by Ni-NTA chromatography (see chapter 5.3.2.3) and molecular masses were determined by electrospray ionization mass spectrometry (ESI-MS, see chapter 5.5.1).

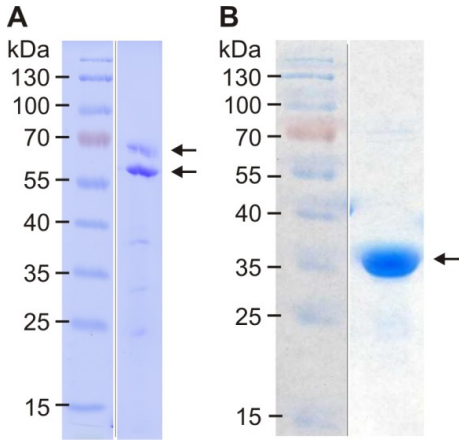


Figure 14. SDS gel of Ni-NTA purified *sc*PheRS(T415G) (A) and *mj*BpaRS (B). Purified proteins are highlighted by arrows. Note that the *sc*PheRS(T415G) consists of two subunits α and β while *mj*BpaRS is a homodimer. The β-subunit of *sc*PheRS(T415G) and *mj*BpaRS undergo N-terminal Met excision (NME). Protein preparations were analyzed with respect to ESI masses. (**A**) α-subunit, $M_{w,\,cal}$: 58874 Da, $M_{w,\,meas}$: 58882 Da; β-subunit, $M_{w,\,cal}$(-Met): 67309 Da, $M_{w,\,meas}$(-Met): 67307 Da. (**B**) $M_{w,\,cal}$(-Met): 38615 Da, $M_{w,\,meas}$: 38615 Da.

In both cases, protein purity was close to 95% (see **Figure 14**) and the correct molecular masses were detected for the β-subunit of *sc*PheRS(T415G) and *mj*BpaRS (see legend in **Figure 14**). The α-subunit of *sc*PheRS(T415G), however, exhibited a 8 Da difference between calculated and found mass. Regardless of this, *sc*PheRS(T415G) proved to be active (see **Figure 15**).

The *in vitro* activation of canonical and noncanonical amino acids by both aaRSs was performed using the ATP:PP_i exchange assay (see chapter 5.4.1). Endpoint measurement with the substrate in high access was performed. Such an experiment will not lead to directly comparable rate constants for every amino acid. Instead, it is possible to obtain an estimation on how capable is a particular aaRS of activating an amino acid. It can be argued that this 'end point scenario' can reflect the situation *in vivo* because there are hints that ncaas are highly accumulated in the cell.[170] In general, *sc*PheRS(T415G) was capable of activating all tested amino acids. While the most efficient activation was found for the caas Phe and Trp as well as the ncaas AzPhe, AmPhe, MTyr, IPhe, and Aal, only Bpa and Fal, which structurally differ very much from the original substrate of the *sc*PheRS, were not activated significantly (see **Figure 15A**). In contrast, *mj*BpaRS displayed only negligible activation capacity close to background level for all tested amino acids (see **Figure 15B**).

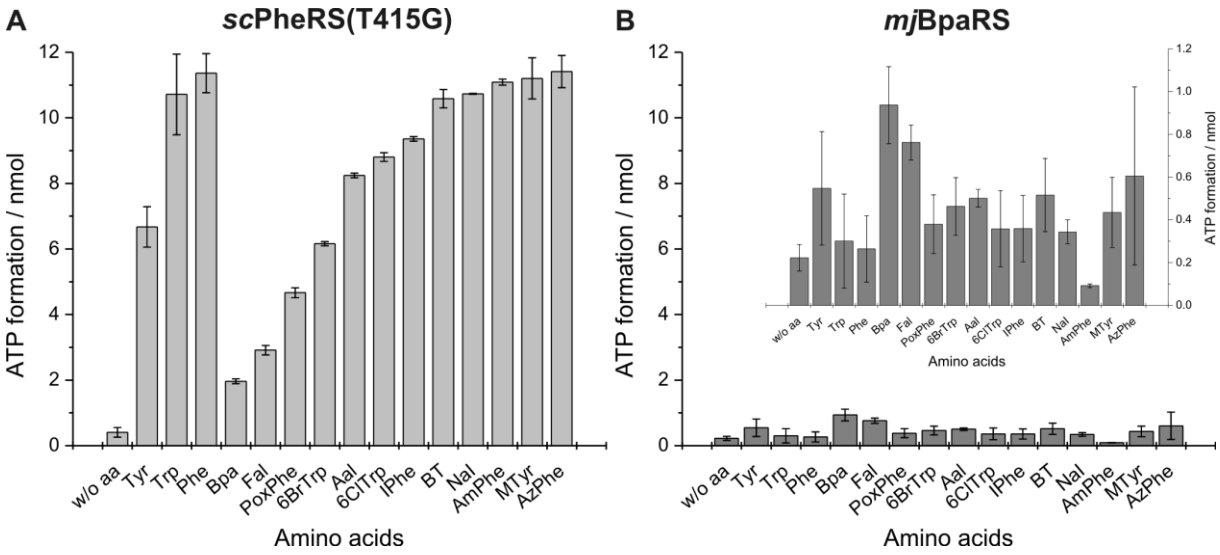


Figure 15. Activation of different canonical and noncanonical amino acids by *sc*PheRS(T415G) (A) and *mj*BpaRS (B). The reaction mixture contained 5 mM amino acids and 3 µM *sc*PheRS(T415G) or 5 µM *mj*BpaRS. The insert in **B** is a tenfold magnification of the original figure. *In vitro* activation assays were kindly performed by Sebastian Nehring.

Nonetheless, a closer look at the small activation values (see **Figure 15B, insert**) at least identifies Bpa as the ncaa which is activated to the highest extent. Most surprisingly, Fal was the only other amino acid activated over background level beside Bpa. This is remarkable since it greatly differs from Bpa in structure and van der Waals volume.

Based on these results, *in vivo* incorporation experiments with *sc*PheRS(T415G) were performed with all tested ncaas. In case of *mj*BpaRS, only Bpa, Fal, AzPhe, IPhe and Nal were selected for a further characterization because the results from the *in vitro* assays were not so promising.

2.2.2 *In vivo* analysis of noncanonical amino acid incorporation

2.2.2.1 *sc*PheRS(T415G)

For incorporation experiments with *sc*PheRS(T415G), the two-plasmid-system of Tirrell and coworkers (see chapter 2.1.2 for details) was adopted in the reported conditions to validate their results. However, the incorporation protocol itself was slightly modified. For efficient incorporation of ncaas with *sc*PheRS(T415G), Phe and Trp, which are also loaded by *sc*PheRS(T415G), have to be removed from the medium before expression induction. Tirrell and coworkers grew cells in full medium and then washed them thoroughly to remove Phe and Trp. In this study, fermentations were performed according to the SPI method (see chapter 5.3.2.3) to avoid time and material consuming washing steps. For this purpose, limiting Phe and Trp concentrations were determined for the *E. coli* AFW strain (see chapter 5.3.2.1). 40 µM each of Phe and Trp sufficed for *E. coli* AFW to reach an OD_{600} of 0.6 - 0.7. However, if Phe and Trp were completely removed from the medium, no target protein expression would take place if it contains codons for Phe and Trp in its sequence in addition

to the amber stop codon. Thus, just before induction, small amounts of Trp and Phe (10 and 30 µM, respectively) were added together with the tested ncaa in great access (3 mM) to ensure a high level of incorporation. As model protein, *m*DHFR containing an amber stop codon at position 38 was used (see chapter 2.1.2). *m*DHFR variants were purified by Ni-NTA chromatography and analyzed by ESI-MS. The resulting protein preparations are denoted as *m*DHFR(F38X)+caa/ncaa because position 38 can be occupied by different amino acids in a single expression experiment while '+caa/ncaa' refers to a particular preparation with either caa or ncaa added in excess. *m*DHFR(F38caa/ncaa) is used if the homogeneity of the preparation is known with respect to the amber position occupancy or when to a distinct mass peak is referred. In general, *m*DHFR preparations had a purity > 90% (see **Figure 40** in the appendix). However, samples were highly heterogenic, mostly because *m*DHFR is prone to degradation at the C-terminus. For illustration, the high performance liquid chromatography (HPLC) chromatogram and mass spectrum of *m*DHFR expressed in NMM with 1 mM Trp and Phe is presented in **Figure 16**. For a detailed description refer to the legend in **Table 2**.

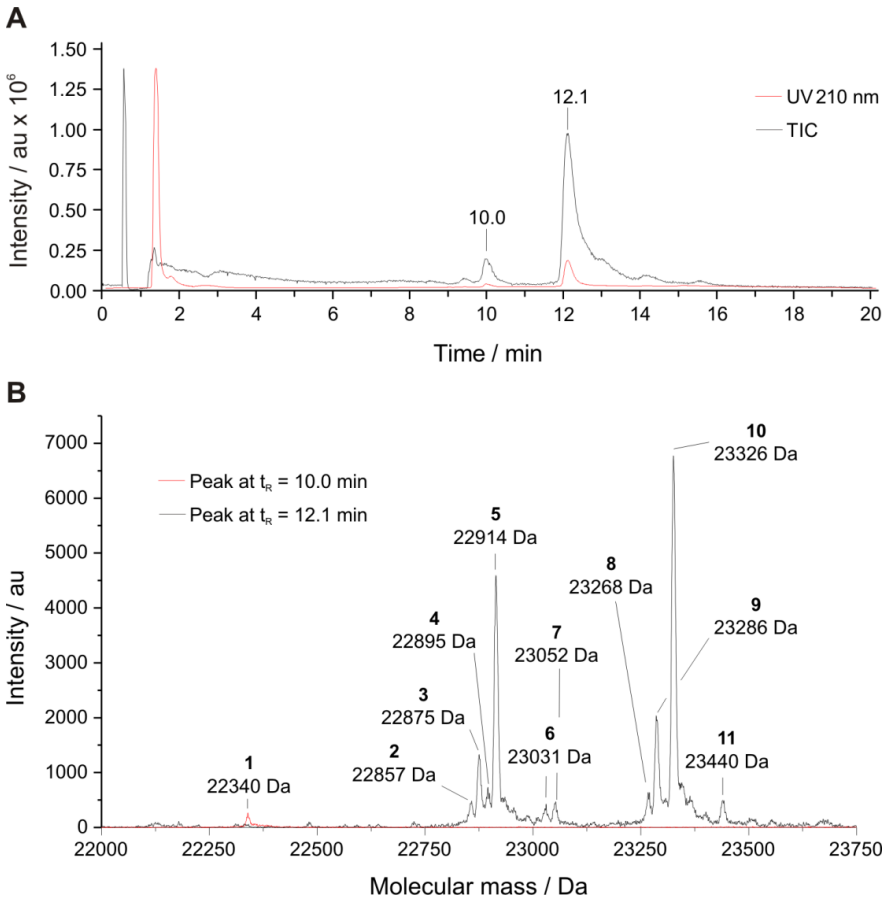


Figure 16. HPLC and ESI-MS analysis of *m*DHFR(F38X) expressed in NMM with 1 mM Trp and Phe. A: The HPLC analysis revealed two peaks at t_R = 10.0 min and t_R = 12.1 min, respectively. **B**: ESI-MS revealed that the first peak contained a non-identified *m*DHFR fragment with a mass of 22340 Da (red line). The second peak corresponded to *m*DHFR(F38X) with different amino acids at position 38 as well as different aggregates and degradation products (black line). An annotation of the mass peaks (1 – 11) can be found in **Table 2**.

Importantly, degradation was not influenced by the amino acid species (Phe, Trp or Lys) at position 38 because by comparing the peak intensities of full-length *m*DHFR with those of the correspondent degraded variants (*m*DHFR-3xHis), it could be seen that the peak intensities are proportional (see peaks 2 *vs*. 8, 3 *vs*. 9 and 5 *vs*. 10 in **Figure 16B**). Therefore, the relative occupancy of position 38 by Phe, Trp, or Lys was determined by integration of the mass peaks of full-length *m*DHFR(F38X) species (**Table 3**). The mass spectra from which these data were deduced can be found in **Figure 41** in the appendix.

The *m*DHFR(F38X) preparation in NMM with 1 mM Trp and Phe contains 73% of Trp at position 38 and only 21% of Phe indicating that *sc*PheRS(T415G) prefers Trp over Phe which, is in good agreement with the literature data.[174] A misacylation of approx. 6% of *sc*tRNA$^{Phe}_{CUA_UG}$ by the *E. coli* LysRS was detected in all *m*DHFR preparations (see Table 3). However, Tirrell and coworkers observed at the most 1% of Lys misacylation in their preparations.[174] This difference most probably arises from the different modes of analysis. While the results presented here were gained from integration of ESI-MS mass peaks, Tirrell and coworkers compared the corresponding peptide intensities of trypsin digested samples in an LC-MS/MS setting.[174] In such analysis mode, an additional Lys enhances the ionization of the corresponding peptide leading to a higher calculated Lys occupancy than actually present (*vide infra*). In the present study, Lys incorporation adds a new trypsin cleavage site to the protein which may have led to non-detection of the peptide.

Table 2. Annotation of mass peaks of *m*DHFR(F38X) expressed in NMM with 1 mM Trp and Phe. Data were deduced from **Figure 16**. The labels '– 2 x His' and '– 3 x His' correspond to *m*DHFR species where two or three His residues were C-terminally cleaved off. A cleavage of one His residue results in a mass loss of - 137 Da. Degradation most probably occurred after purification by Ni-NTA chromatography because otherwise the degraded species would most likely not anymore have bound to the Ni-NTA material. Peak 6 and 8 are most probably *m*DHFR adducts with trifluoroacetic acid (TFA, M_w = 114 Da), a component of the MS solvent.

Peak number	Detected mass / Da	Assigned protein species	Calculated mass / Da	Difference of detected and calculated mass
1	22340	Not assigned	-	-
2	22857	*m*DHFR(F38K) – 3 x His	22856	+ 1 Da
3	22875	*m*DHFR(F38) – 3 x His	22875	no
4	22895	Not assigned	-	
5	22914	*m*DHFR(F38W) – 3 x His	22914	no
6	23031	*m*DHFR(F38W) – 3 x His + TFA	23028	+ 3 Da
7	23052	*m*DHFR(F38W) – 2 x His	23051	+ 1 Da
8	23268	*m*DHFR(F38K)	23268	no
9	23286	*m*DHFR(F38)	23287	- 1 Da
10	23326	*m*DHFR(F38W)	23326	no
11	23440	*m*DHFR(F38W) + TFA	23440	no

Expressions with the different ncaas in NMM with limiting Phe and Trp concentrations showed highest capability of *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$ for incorporation of 6BrTrp, 6ClTrp, BT and Nal. These amino acids have been already reported to be well incorporated using this system. Aal, MTyr, AzPhe, and IPhe were also incorporated, however, with lower efficiency. AmPhe, PoxPhe, Fal and Bpa were not incorporated by *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$ (see **Table 3**). It was already previously assumed that suppression of the amber stop codon with Bpa and Fal would not work because the ncaas were poorly activated by the synthetase in the *in vitro* assays (see chapter 2.2.1). This was readily confirmed *in vivo*.

Table 3. Occupancy of position 38 in different *m*DHFR(F38X) preparations. All protein variants were expressed according to chapter 5.3.2.2. The mass spectra from which these data were deduced are shown in **Figure 41** in the appendix.

Incorporated amino acid	Protein yield / mg L^{-1} OD_{600}^{-1}	Occupancy of amber site / %			
		Lys	Phe	Trp	ncaa
Trp/Phe[1]	1.9	6	21	73	-
Phe[2]	2.9	6	61	33	-
Trp analogs[3]					
6-bromotryptophan[4]	-	nd	nd	nd	> 90
6-chlorotryptophan	6.8	2	2	nd	96
3-(3-benzothienyl)alanine	3.8	nd	4	6	89
3-(2-naphthyl)alanine	2.1	7	6	6	81
3-(1-azulenyl)alanine	1.5	5	21	22	51
Phe/Tyr analogs					
O-methyltyrosine	0.7	14	13	23	49
4-azidophenylalanine	2.2	10	6	42	42
4-iodophenylalanine	0.7	39	9	20	33
4-aminophenylalanine	4.0	5	17	79	nd
4-propargyloxyphenylalanine	3.2	4	25	71	nd
other ncaas					
4-benzoylphenylalanine	0.8	9	32	59	nd
3-(2-furyl)alanine	3.5	6	17	77	nd

[**1**] Protein was expressed in presence of 1 mM Trp and Phe.
[**2**] Protein was expressed in presence of 3 mM Phe and limiting amounts of Trp.
[**3**] Expressions were performed with 3 mM ncaas and limiting concentrations of 40 µM for Phe and Trp.
[**4**] Incorporation was carried out in only 200 mL culture volume due to high costs of 6BrTrp. No reasonable protein amounts for yield calculation were obtained. Furthermore, it was not possible to determine the occupancy of position 38 with Lys, Phe and Trp because the ESI-MS signal intensity was very low.

Remarkably, it was found that PoxPhe, AmPhe and AzPhe were poorly or not at all incorporated although they were highly activated by *sc*PheRS(T415G) *in vitro*; in contrast to 6BrTrp, 6ClTrp, BT and Nal which were efficiently incorporated *in vivo*. The same applies to Tyr which was well activated *in vitro* but never found in any *m*DHFR(F38X) preparation, even though the medium was supplied with Tyr. These results show that other steps in protein translation, for example loading of activated amino acids onto the tRNA, interaction of aminoacyl-tRNA and elongation factor EF-Tu or acceptance of aminoacyl-tRNA by the ribosomal machinery, may also play a crucial role for successful incorporation of ncaas (see chapter 1.2.1).

With Aal, an amino acid with very special spectral properties was incorporated into a protein for the first time. Until now, it was only used in peptide studies[114] and efforts for its incorporation by chemical acylation of suppressor tRNAs in the group of Sisido were not successful (personal communication of Prof. Moroder).

In general, the Trp and Phe contaminations were higher in the results presented here than reported elsewhere.[174] A reason for this could be the usage of the amino acid 'depletion approach' instead of the 'washing approach'. Obviously, stop codon suppression with *sc*PheRS(T415G) is much more sensitive to small residual amounts of Trp and Phe in the medium at protein expression induction than the residue specific SPI approach. Here, the parallel depletion of three amino acids in the medium and subsequent simultaneous incorporation of three different ncaas does not lead to detectable amounts of contamination with caas.[114,179]

In addition to ESI-MS, it was checked if an Orbitrap analysis (see chapter 5.5.7) would be suitable for quantification of amber site occupancy as well. For this purpose, the *m*DHFR(F38X)+Phe preparation was digested with the endoproteinase GluC and the resulting peptide fragments were applied to Orbitrap mass analysis. The results are shown in **Table 4**. Unfortunately, the results of ESI-MS are not in line with the results from Orbitrap analysis. For example, while ESI-MS data revealed around 6% of misacylation with Lys, Orbitrap analysis suggested around 19% when only peptide intensities are taken into account.

Table 4. Orbitrap analysis of *m*DHFR(F38X)+Phe preparation. *m*DHFR(F38X)+Phe was digested with GluC, and the resulting peptides were separated by HPLC and analyzed via MS as well as MS/MS. Only the fragments containing position 38 are shown. Peptide count: number of specific peptides sequenced in the Orbitrap; MS/MS count: number of times a specific peptide was selected for a sequencing event in the Orbitrap. A more detailed version of this table can be found in chapter 6.1.3 in the appendix.

Peptide sequence	Intensity / %	Peptide count ratio / %	MS/MS count ratio / %	Occupancy from ESI-MS data / %
WKYFQRMTTTSSVE	22	26	25	33
FKYFQRMTTTSSVE	58	68	72	61
KKYFQRMTTTSSVE	19	6	3	6

Also the ratios between Trp and Phe incorporation varied significantly from 1:2 for ESI-MS (33 to 61%) and approx. 1:3 for Orbitrap data (22 to 58%). The differing results of the Orbitrap analysis most probably result from a significantly different ionization efficiency of the peptides when containing Trp, Phe or Lys at position 38. In fact, it is known that peptides containing positively charged Lys or Arg residues are ionized more effectively than their uncharged counterparts. Thus, their intensity values will also be higher, thereby explaining the much higher intensity for Lys containing peptides in the present Orbitrap data (see **Table 4**). Therefore, an accurate quantification of amber site occupancy cannot be performed by comparing intensities from Orbitrap and ESI-MS data when the peptides contain chemically very different amino acids. A more accurate picture can be drawn when the peptide count ratio (number of specific peptides sequenced in Orbitrap) and MS/MS count ratio (number of times a specific peptide has been selected for a sequencing event in Orbitrap) are taken into account because they also reflect the abundance of a peptide to a certain extent (see **Table 4**). However, the analysis with Orbitrap can only give semi-quantitative estimations as long as no calibration with synthesized peptides is performed.

In summary, it could be confirmed that the *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$ pair can efficiently incorporate 6BrTrp, 6ClTrp, BT and Nal in response to amber stop codons. It might be that the incorporation rate can be further improved using the 'washing approach' used by Tirrell and coworkers instead of the 'depletion' methodology employed here. Furthermore, the basic capacity of *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$ for activating and incorporating MTyr, IPhe and AzPhe was demonstrated. More importantly, Aal incorporation, which was not successful so far, was also achieved in reasonable amounts. Based on this, it will be now possible to design and characterize new *Aequorea victoria* green fluorescent protein (*av*GFP) variants with fluorophores containing Aal. Moreover, the great substrate tolerance of the *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$ pair gives the possibility for making incorporation trials with *N*-methylated azatryptophans. These enhanced small molecule fluorophores were recently characterized[180] but the absence of an incorporation system hampered further investigations on their applicability in protein engineering.

2.2.2.2 *mj*BpaRS

The m*j*BpaRS substrate tolerance *in* vivo was only tested for with Bpa, Fal, AzPhe, IPhe and Nal as in the *in vitro* assays *mj*BpaRS was not capable of activating efficiently any of the tested amino acids (see chapter 2.2.1). For *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ coexpression, the plasmid pSup-BpaRS-6TRN was used (see chapter 2.1.1).[159] In this case, TTL with a C-terminal His-tag and an amber stop codon at position 37 was the target protein (see chapter 2.4.3.1 for details on stop codon position) because it is much more stable with respect to degradation than *m*DHFR making ESI-MS spectrum analysis easier. Expression was performed in the *E. coli* strain AFW according to chapter 5.3.2.3, followed by purification and ESI-MS analysis. Fermentation with Bpa led to a clearly visible expression band (black arrow, **Figure 17A**) leading to the purification of TTL(F37X)+Bpa up to > 90% (**Figure 17B**). ESI-MS analysis showed complete incorporation of Bpa at the expected stop codon position 37 without visible contaminations of caas (**Figure 17C**).

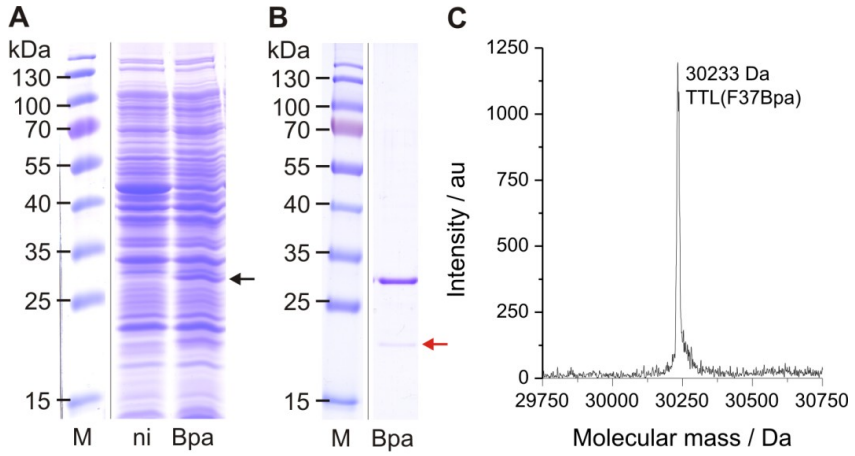


Figure 17. Expression (A), purity (B), and ESI-MS analysis (C) of TTL(F37X) expressed in *E. coli* AFW{pQE80L-TTL-H6(F37amber)/pSUP-BpaRS-6TRN} with Bpa. A: Expression of TTL(F37X) in presence of 1 mM of Bpa (0.15 OD_{600} of cell lysates per lane were applied to a 12% SDS gel). A black arrow indicates the TTL expression band (ca. 30 kDa). **M**: marker; **ni**: non-induced sample. **B**: TTL preparation containing Bpa after Ni-NTA purification. 3 µg of protein were applied to a 12% SDS gel. An additional protein species was co-purified with full-length TTL(F37X) (red arrow). This species appears in all TTL preparations and is most probably an N-terminal truncation product of TTL with 21183 Da (aa 81-267, see **Figure 42A** in the appendix) since it generates a signal in Western blot analysis (see **Figure 31**). **C**: ESI-MS analysis of TTL(F37X)+Bpa proves incorporation of Bpa at amber position 37 without significant caa contamination. TTL(F37Bpa): $M_{w,\,cal}$ = 30234 Da, $M_{w,\,meas}$ = 30233 Da.

In case of the TTL preparations with Fal, AzPhe, IPhe and Nal no overexpression could be detected by SDS gel analysis (**Figure 18A**). Also after purification, samples were highly heterogenic with no clear enrichment of a protein of the correct size (black arrow, **Figure 18B**). Due to the high impurity, instead of ESI-MS, Orbitrap analyses were performed (see chapter 5.5.7) using trypsin digested samples from the gel as highlighted in **Figure 18B**, and with TTL(F37Bpa) for comparison.

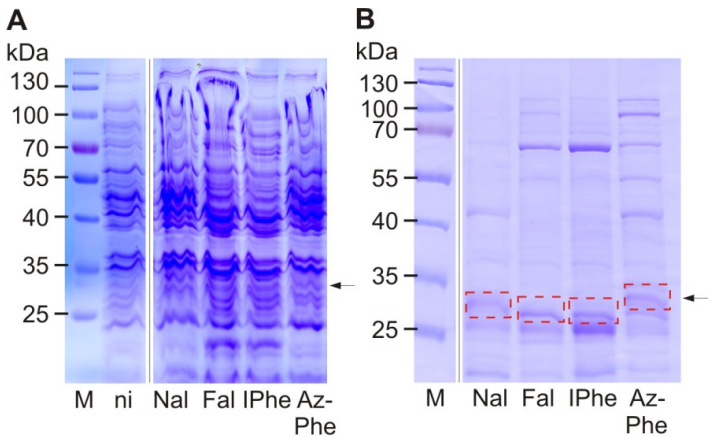


Figure 18. Expression (A) and purity (B) of TTL(F37X) expressed in *E. coli* AFW{pQE80L-TTL-H6(F37amber)/pSUP-BpaRS-6TRN} with different ncaas. A: Expression of TTL(F37X) in presence of 1 mM of respective ncaa (0.15 OD_{600} of cell lysates per lane were applied to a 12% SDS gel). Black arrows indicate the height of the expected TTL expression bands (ca. 30 kDa). **M**: marker; **ni**: non-induced sample; names of ncaas can be found in **Figure 13**. **B**: Purity of all TTL preparations after Ni-NTA affinity chromatography. 3 µg of protein were applied to a 12% SDS gel. The dotted, red squares indicate the gel regions cut out for Orbitrap analysis.

It was always taken into account that an occupancy quantification based Orbitrap data is only semi-quantitative (see chapter 2.2.2.1). Nevertheless, a rough estimation was possible, as presented in **Table 5** on the next page. In the TTL(F37Bpa) sample, the peptide VPMVIMFHG**Bpa**TGNK was found with very high intensity. Besides, a peptide with Phe at position 37 was detected, which could mean that *mj*BpaRS has a very low background activity for Phe — not detectable in ESI-MS. However, it could also be possible that VPMVIMFHG**F**TGNK was protracted from the sample run directly before, corresponding to a TTL without amber mutation at position 37. Unfortunately, such contaminations sometimes occurred when the samples were not separated by at least five foreign samples in the analysis process. In case of Nal, Fal, IPhe and AzPhe preparations such contamination can be excluded since they were analyzed subsequently and no TTL without the amber mutation at position 37 was run before. Small amounts of VPMVIMFHG**Nal**TGNK and VPMVIMFHG**AzPhe**TGNK were found in TTL(F38X)+Nal and TTL(F38X)+AzPhe, respectively (see **Table 5**). In contrast, no peptides with Fal or IPhe could be identified in the respective samples. Bpa, Nal, IPhe and AzPhe preparations, however, are dominated by VPMVIMFHG**F**TGNK suggesting that *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ indeed incorporates Phe to a small extent.

Table 5. Orbitrap analysis of TTL(F37X) expressed with different ncaas. The TTL(F37X)+ncaa variants were digested with trypsin, and the resulting peptides were separated by HPLC and analyzed via MS as well as MS/MS. Only the peptides bearing position 37 are shown. Note that only peptides with complete methionine oxidation to methionine sulfoxide are presented. The oxidation is an intrinsic artifact of the protocol used for Orbitrap analysis (see chapter 5.5.7). Partially and non-oxidized peptides were also detected and reflected the same trend as the presented completely oxidized forms. A more detailed version of this table can be found in chapter 6.2 in the appendix.

Peptide sequence	intensity / au	Intensity / %	Peptide count	MS/MS count
TTL(F37X)+Bpa				
VPM(ox)VIM(ox)FHG**F**TGNK	2247800	2	10	11
VPM(ox)VIM(ox)FHG**Bpa**TGNK	88309000	98	90	89
TTL(F37X)+Nal				
VPM(ox)VIM(ox)FHG**F**TGNK	59145000	87	67	67
VPM(ox)VIM(ox)FHG**Nal**TGNK	9193000	13	33	33
TTL(F37X)+Fal				
VPM(ox)VIM(ox)FHG**F**TGNK	nd	nd	nd	nd
VPM(ox)VIM(ox)FHG**Fal**TGNK	nd	nd	nd	nd
TTL(F37X)+IPhe				
VPM(ox)VIM(ox)FHG**F**TGNK	1484400	100	100	100
VPM(ox)VIM(ox)FHG**IPhe**TGNK	nd	nd	nd	nd
TTL(F37X)+AzPhe				
VPM(ox)VIM(ox)FHG**F**TGNK	46496000	50	75	64
VPM(ox)VIM(ox)FHG**AzPhe**TGNK	46765000	50	25	36

In summary, these data verify that the *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ pair incorporates Bpa at the expected amber site as stated by Schultz and coworkers.[144] Nal, AzPhe and Phe incorporation could be detected only by Orbitrap analysis in the respective preparations. Based on these results, the order of *mj*BpaRS *in vivo* amino acid preference was annotated as Bpa >>> Phe > Nal > AzPhe > IPhe = Fal. Nonetheless, a reasonable expression for protein preparation in high purity could only be obtained in presence of Bpa.

Taken together, *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ can be regarded as a highly specific orthogonal pair for Bpa. Other amino acids are incorporated only in trace amounts. Indeed, this result was not expected according to the *in vitro* assays because Bpa was not significantly better activated than e.g. Fal (see **Figure 15B**). Although it can be speculated that *mj*BpaRS activates Bpa by a special mechanism different from native synthetases, the structure of *mj*BpaRS with co-crystallized Bpa suggests a rather classical recognition of Bpa by *mj*BpaRS.[181] More research has to be performed to clarify this issue.

2.2.3 Evaluation of amber suppression rates

To evaluate the basic amber suppression efficiencies of the two aaRS:tRNA$_{CUA}$ pairs in use, EGFP with an amber stop codon at position 150 (EGFP(N150Amber)) was chosen as target protein because it is known that this position is efficiently suppressed in *av*GFPs.[153] Both, EGFP and EGFP(N150Amber) were cloned into pQE16_*sc*PheRS(T415G) and pQE80L (see chapter 5.1.3.2). A C-terminal His-tag was added in order to purify and selectively stain full-length protein by Western blot analysis using an anti-His-tag antibody. For suppressor tRNA supply, pREP4_*sc*tRNA$^{Phe}_{CUA_UG}$ was introduced into cells with all pQE16_*sc*PheRS(T415G) constructs. pSUP-BpaRS-6TRN was used in combination with pQE80L constructs (see chapter 2.1 for details).

2.2.3.1 Suppression efficiency of scPheRS(T415G):sctRNA$^{Phe}_{CUA_UG}$

To determine suppression efficiency, in a first step small scale expression was performed with *E. coli* strain AFW in LB medium (see chapter 5.3.1), followed by Western blot analysis (see **Figure 19**).

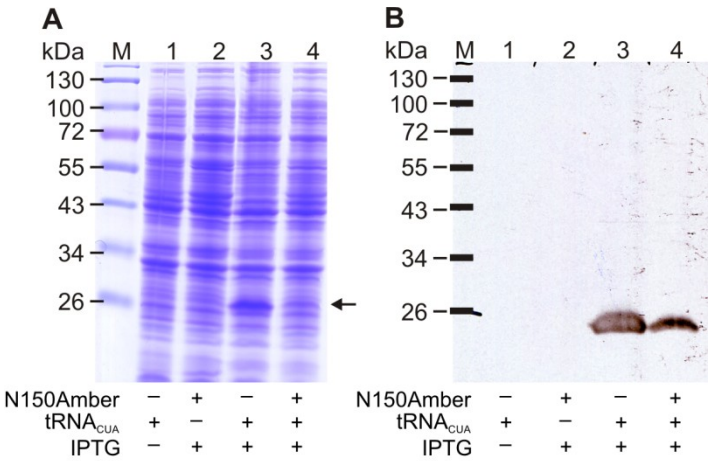


Figure 19. Evaluation of suppression efficiency of the *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$ pair in LB medium using EGFP. SDS gel (**A**) and corresponding Western blot (**B**): 0.15 OD$_{600}$ of cell lysates per lane were applied to a 12% gel. **M**: marker; **1**: non-induced sample; **2**: control expression of EGFP(N150Amber) without tRNA$_{Sup}$; **3**: parent EGFP expression; **4**: EGFP(N150Amber) expression with tRNA$_{Sup}$.

Expectedly, no EGFP expression was detected without induction of protein expression (**Figure 19**, lane 1) or without a suppressor tRNA in case of EGFP(N150Amber) (**Figure 19**, lane 2). Western blot analysis revealed high full-length protein amounts for parent protein (EGFP without amber stop codon) and amber suppression samples (**Figure 19**, lane 3 and 4). Suppression efficiency was estimated to be around 68% by integration of target protein band intensities on the Western blot using the program ImageJ (see chapter 5.5.4).

Since expression in LB medium leads only to Trp and Phe incorporation at position 150, a second expression experiment was performed in NMM to estimate the read-through rate of *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$ with ncaas (see chapter 5.3.2.3). The results using Nal are highlighted in **Figure 20**. The estimated EGFP(N150Nal) band intensity on the Western blot (**Figure 20**, lane 4) was only around 38% when compared to EGFP (**Figure 20**, lane 2). Thus, using NMM and an ncaa instead of expression in LB medium decreased the read-through rate by 44%.

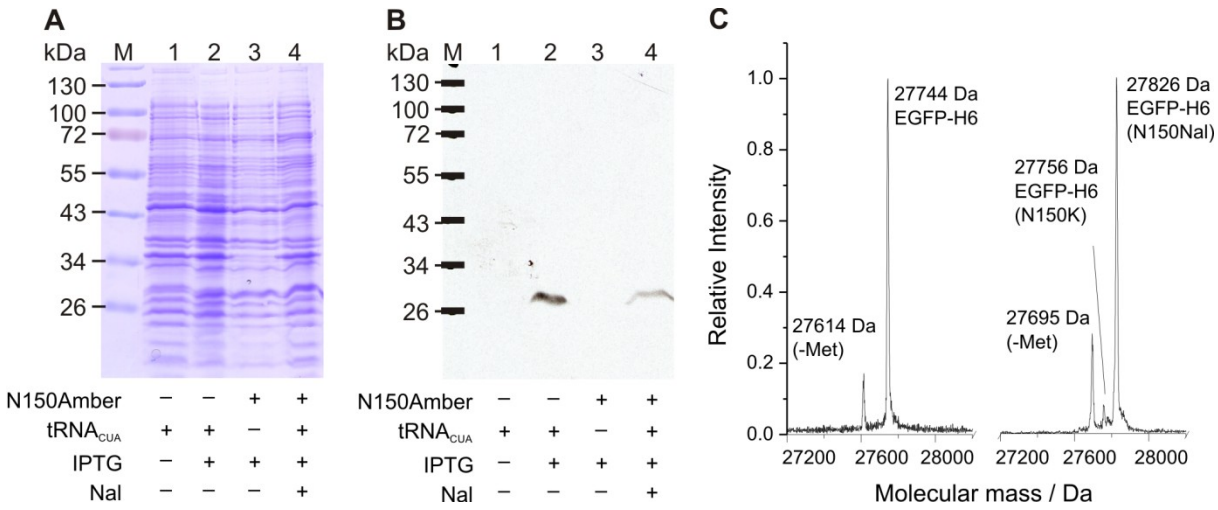


Figure 20. Evaluation of suppression efficiency of the *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$ pair in NMM using EGFP and Nal. SDS gel (**A**) and corresponding Western blot (**B**): 0.15 OD_{600} of cell lysates per lane were applied to a 12% SDS gel. **M**: marker; **1**: non-induced sample; **2**: parent EGFP expression; **3**: control expression of EGFP(N150Amber) without $tRNA_{Sup}$; **4**: EGFP(N150Amber) expression with $tRNA_{Sup}$ and Nal. **C**: ESI-MS analysis of purified EGFP and EGFP(N150Nal). EGFP: $M_{w,\,cal}$ = 27744 Da, $M_{w,\,meas}$ = 27744 Da; EGFP(N150Nal): $M_{w,\,cal}$ = 27827 Da, $M_{w,\,meas}$ = 27826 Da. The used EGFPs undergo partial NME. In EGFP(N150Nal), 6% contamination with EGFP(N150K) was found due to residual misacylation of *sc*tRNA$^{Phe}_{CUA_UG}$ by the endogenous LysRS.

2.2.3.2 Suppression efficiency of *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$

The *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ pair was tested via small scale expression in NMM as described above for *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$. ImageJ analysis of the Western blot revealed a suppression efficiency of around 46% (see **Figure 21**).

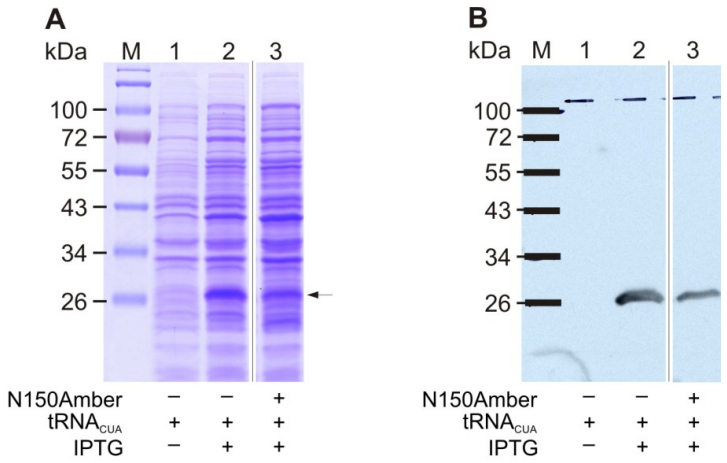


Figure 21. Evaluation of suppression efficiency of the *mj*BpaRS:mjtRNA$^{Tyr}_{CUA}$ pair in NMM using EGFP and Bpa. For the SDS gel (**A**) and corresponding Western blot (**B**), 0.15 OD_{600} of cell lysates per lane were applied to a 12% SDS gel. **M**: marker; **1**: non-induced sample; **2**: parent EGFP expression; **3**: EGFP(N150Amber) expression with $tRNA_{Sup}$ and Bpa.

2.2.3.3 Comparison of *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$ and *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$

By comparing the two different amber suppression systems, it can be seen that the *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ pair is moderately more efficient in incorporating ncaas than the rationally designed *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$ pair (46% *vs.* 38%, respectively). However, both systems are designed differently. The genes of the *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ pair are located on only one single plasmid with a low copy number (p15A Ori). In contrast, the *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$ pair is split up: The *sc*PheRS(T415G) and the target protein are located on the high copy number plasmid pQE16 (ColE1 Ori) while the *sc*tRNA$^{Phe}_{CUA_UG}$ is on a separate low copy plasmid (pREP4, p15A Ori). Furthermore, pSup-BpaRS-6TRN harbors six copies of *mj*tRNA$^{Tyr}_{CUA}$ but pREP4_*sc*tRNA$^{Phe}_{CUA_UG}$ contains only one *sc*tRNA$^{Phe}_{CUA_UG}$ gene copy.

Here, the question arises how these two differently designed aaRS:$tRNA_{CUA}$ coexpression systems show a comparable efficiency at the end. It is already known from the literature that in the *mj*BpaRS system an increase in tRNA copy number from one to six in pSup-BpaRS leads to a nearly 70% higher suppression efficiency.[159] Thus, the suppression rate with only one *mj*tRNA$^{Tyr}_{CUA}$ would most probably yield a lower read-through rate than in case of the *sc*PheRS(T415G) system (around 32%). Recently, a new plasmid was described which further enhanced amber suppression rate.[182] It contains two copies of a modified *mj*TyrRS for *O*-methyltyrosine incorporation: one associated with the constitutive *glnS*' promoter and the other with an arabinose inducible promoter. In this light, it is clear that a higher aaRS level

also leads to better suppression. In the systems discussed herein, *sc*PheRS(T415G) is most probably present in a much higher concentration in the cell than *mj*BpaRS since more templates for transcription are available due to a higher plasmid copy number. The different promoters should not play a significant role since both are strong constitutive promoters.[56,159]

Thus, in an optimal case, an aaRS:$tRNA_{CUA}$ coexpression system should provide highest possible concentrations of aaRS and $tRNA_{CUA}$ in the cell. As of the pQE16_*sc*PheRS(T415G)/ pREP4_*sc*$tRNA^{Phe}_{CUA\_UG}$ system, the $tRNA_{Sup}$ is most likely impeding a higher suppression rate while pSup-BpaRS-6TRN does not provide sufficient amounts of aaRS for a more efficient read-through.

It has to be kept in mind, however, that a very high aaRS and tRNA expression might withdraw cellular resources for target gene expression. Hence, an increase in amber suppression by a further increase of aaRS and tRNA expression could lead to an overall decrease in full-length target protein yield. This fact also has to be considered when evaluating the utility of the orthogonal ribosomes developed by Chin and coworkers (see chapter 1.3.3.2). From their publications, it is clear that the use of orthogonal ribosomes — characterized by a decreased functional interaction with the release factor 1 — increases amber suppression rates. To date, however, only SDS gels with purified proteins have been published. No data of overall yield of labeled full-length protein (neither relative nor in $mg\ mL^{-1}\ OD_{600}^{-1}$) were presented in comparison to an amber suppression system without orthogonal ribosomes. Another important issue is the quality of the components employed. Recently, Schultz and coworkers introduced an *in vitro* evolution approach to optimize aminoacyl-$tRNA_{CUA}$ for interaction with elongation factor EF-Tu.[153] Based on these results, an optimized vector was assembled, which contained only one copy of an optimized $tRNA_{CUA}$ in combination with two copies of aaRS (*vide supra*). This vector provided a threefold higher suppression efficiency when compared to the vector used in this study.[149] Furthermore it was shown that the switch from six conventional to one optimized tRNA led to better strain viability. This was also detected in the present study. Cells transformed with pSup-BpaRS-6TRN grew significantly slower than clones with less suppressor tRNA gene copies. Obviously, not the presence of an efficiently suppressing aaRS:$tRNA_{CUA}$ system, which also efficiently suppresses endogenous amber stop codons, but rather the scrambling of the cellular machinery by a high load of, perhaps, non-maturated suppressor tRNAs negatively influences the growth rate. Indeed, Ellington and coworkers could provide evidence that the mature tRNA species is only present at around 20 - 25% in the cell in their single *sc*$tRNA^{Trp}_{CUA}$ system.[137] In addition, orthogonal aaRSs still bear a lot of potential for improvement with respect to the $tRNA_{Sup}$ binding and ncaa recognition. For example, Yokoyama and coworkers reported that a single mutation of Asp286 to Arg in the C-terminal domain of *mj*TyrRS, responsible for anticodon recognition (see **Figure 4C** in the introduction), leads to eightfold higher *in vitro* aminoacylation activity (this mutation is also present in the *mj*BpaRS used in this study).[55,151] Most probably, a comprehensive study on the effects of certain amino acid residues in tRNA recognition will generate further optimized orthogonal aaRSs with enhanced activity for $tRNA_{Sup}$ aminoacylation. Recently, Skerra and coworkers refined the

amino acid binding pocket of the engineered *mj*TyrRS for MTyr incorporation originally designed by Schultz and coworkers.[139,150] With the refined TyrRS a tenfold increase in MTyr incorporation was achieved, showing that binding pocket remodeling also bears a lot of potential for improving current orthogonal pairs.

2.3 A new vector system for aaRS:tRNA coexpression

2.3.1 Design strategy for the pMEc vector family

The results and considerations from chapter 2.2.3 led to the design of a new modular vector system to facilitate cloning and expression of aaRS:tRNA pairs in combination with varying target proteins. The new vector was termed pMEc (plasmid of Modular parts for use in *E. coli*) as it was assembled from different standardized modules. The idea was inspired by the Biobrick concept of standard biological parts,[183,184] which only uses well defined and characterized building blocks for vector assembly. To facilitate cloning, the pMEc is equipped with a yeast shuttle module. Thereby, the restriction site independent homologous recombination in *S. cerevisiae* (see chapter 5.1.3.3) can be exploited for assembling of the basic vector and later exchange of certain modules or module parts. In the first step, only the tandem vector system was assembled which can be used in combination with all standard expression vectors (see chapter 1.3.3.3). In a later step, the herein designed pMEc vectors can be used to construct a one-vector-system for high yield production of ncaa labeled proteins. The basic architecture of the vector is depicted in **Figure 22**. All modules are separated by unique restrictions sites for linearization of the vector prior to recombination and to exchange single modules by ligation-based cloning. The *Avr*II, *Kpn*I, *Sma*I, *Not*I and *Apa*I sites were chosen due to their absence in the first designed parts (see chapter 4.6 and 6.5). Any of these restriction sites can later be easily exchanged when new parts are introduced into the vector by recombination. If desired, the yeast shuttle module can be removed from the vector by *Pau*I restriction and blunt-end ligation. All open reading frames on the different modules contain transcription terminators to assure enhanced mRNA and plasmid stability.[161]

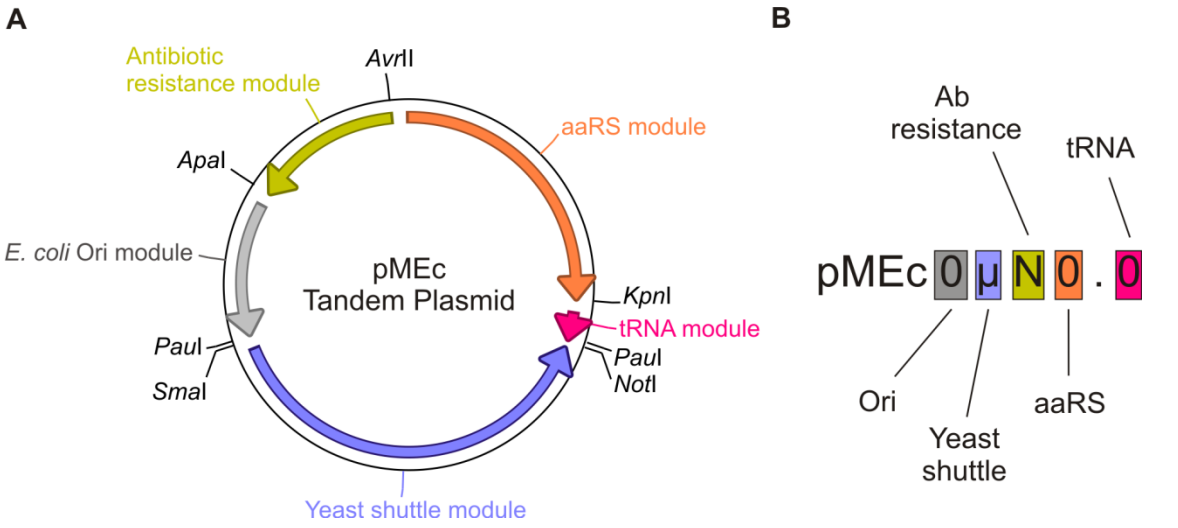


Figure 22. Basic architecture of the pMEc vector family. The plasmid backbone (**A**) and nomenclature (**B**) indicate the most important modules: *E. coli* origin of replication, yeast shuttle module, antibiotic resistance marker, aaRS expression cassette and tRNA expression cassette.

As nomenclature of the new vector family, a numeric system was chosen as depicted in **Figure 22B**. The first number after the prefix pMEc defines the Ori followed by an optional µ which represents the optionally present yeast shuttle. The character after µ gives information on the antibiotic resistance. The two numbers at the end specify which aaRS and tRNA modules are contained in the vector. In chapter 6.5 in the appendix, all defined modules are listed.

2.3.2 Assembly of pMEc vectors with aaRS:tRNA$_{CUA}$ pairs

To investigate whether such a PCR product based vector assembling by homologous recombination in yeast could successfully work, the most basic vector pMEc1µC containing only the p15a Ori, the Cm resistance marker and the yeast shuttle module was assembled from PCR product I (yeast shuttle), II (p15a) and V (Cm) according to chapter 5.1.3.3 (see **Figure 23** and chapter 4.6).

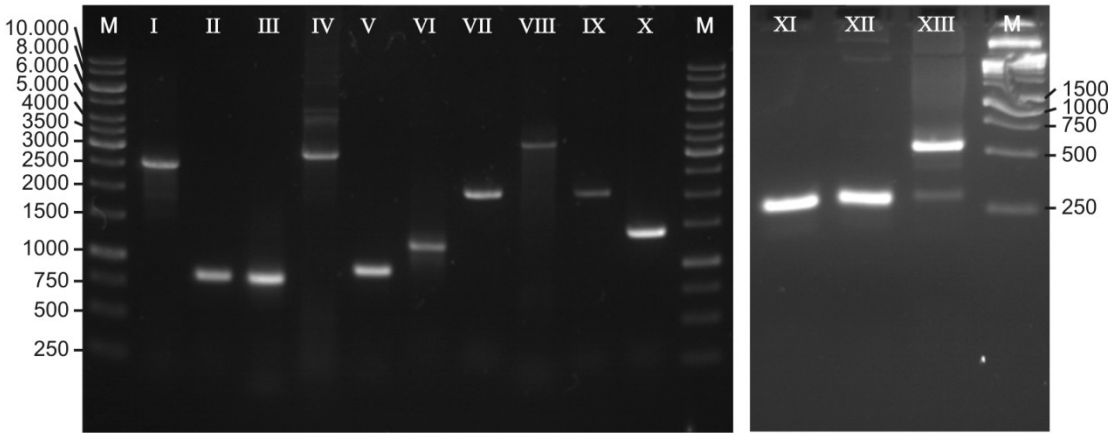


Figure 23. PCR products of amplified modules prepared for the pMEc vector series. M: marker; **I**: yeast shuttle; **II**: p15a Ori; **III**: RSF Ori; **IV**: IPTG inducible F-Ori; **V**: Cm resistance; **VI**: *glnS'*-BpaRS; **VII**: *prpR*-Strep-BpaRS; **VIII**: *tac-sc*PheRS; **IX**: *prpR*-Strep-*sc*PheRS; **X**: *araC*; **XI**: glnS-term; **XII**: *lpp*-tRNA$^{Phe}_{CUA}$ and **XIII**: *proK*-3xtRNA$^{Tyr}_{CUA}$. Left: 1% agarose gel, right: 2% agarose gel. 3 µl of PCR products were applied to the gels. All DNA fragments showed the expected size (see chapter 4.6). Note that PCR product **XIII** only contained the 3xtRNA$^{Tyr}_{CUA}$ array although primers were designed to produce the 6xtRNA$^{Tyr}_{CUA}$ and 3xtRNA$^{Tyr}_{CUA}$ array.

After yeast recombination, the plasmid could be successfully isolated and was stably maintained in *E. coli* at copy numbers comparable with other p15a based plasmids. Control restriction digests of pMEc1µC were performed with all restriction enzymes (RE) for which cleavage sites were introduced (see **Figure 24**). The correct band pattern after digestion and additional DNA sequencing proved the correct assembly of the vector. However, these analyses revealed a second *Apa*I site in the yeast shuttle module which was not reported in the pYES2 sequence from which the module was amplified. This site was later removed by site-directed mutagenesis (see chapter 5.1.2.2). This basic experiment proved the general functionality of the minimal vector backbone.

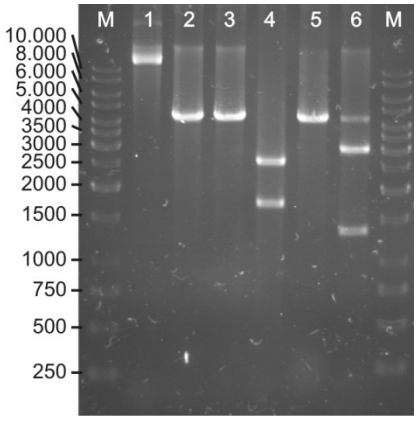


Figure 24. Analysis of pMec1µC digestion products. M: marker; **1**: undigested vector; **2**: *Avr*II; **3**: *Not*I; **4**: *Pau*I; **5**: *Sma*I and **6**: *Apa*I. 300 ng of plasmid were digested and applied to a 1% agarose gel. *Avr*II, *Not*I and *Sma*I cut only once, which resulted in a single DNA fragment with the expected size of 4342 bp. *Pau*I cuts two times and thereby removes the yeast shuttle module (1728 bp) leaving the p15a with the Cm module (2612 bp). *Apa*I should also cut only once but sequencing of pMEc1µC revealed a second, not reported *Apa*I site in the yeast shuttle module (expected band sizes: 3008 and 1334 bp).

To introduce the *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ and *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA}$ pairs into the basic pMEc1µC, the genes of both synthetases including their promoter and terminator sequences were PCR amplified from pSup-BpaRS-6TRN (**Figure 23**, lane **VI**) and pQE16am_T415G (**Figure 23**, lane **VIII**), respectively. For *sc*PheRS(T415G) no terminator sequence could be identified in the initial vector. Thus, a synthetic terminator from the Biobricks collection was selected and fused to the C-terminus via the 3'-primer. The tRNA modules were amplified from pREP4_ytRNA$^{Phe}_{CUA_UG}$ (**Figure 23**, lane **XII)** and pSup-BpaRS-6TRN (**Figure 23**, lane **XIII)**. The PCR primers for amplification of the *mj*tRNA$^{Tyr}_{CUA}$ tRNA module were designed such that the 6xtRNA$^{Tyr}_{CUA}$ and 3xtRNA$^{Tyr}_{CUA}$ arrays could be produced. However, in the PCR reaction only 3xtRNA$^{Tyr}_{CUA}$ could be amplified. The PCR products of the aaRS and tRNA modules were used to assemble the basic pMEc tandem vectors pMEc1µC1.3 (*glnS'-mj*BpaRS + 3xtRNA$^{Tyr}_{CUA}$) and pMEc1µC2.4 (*tac-sc*PheRS + *lpp*-tRNA$^{Phe}_{CUA}$). For direct comparison of the suppression efficiency between the initial vector pSup-BpaRS-6TRN and the newly designed pMEc, the 6xtRNA$^{Tyr}_{CUA}$ array was introduced into pMEc1µC1.3. This was achieved by ligation-based cloning using the unique *Sph*I site at the end of the 3xtRNA$^{Tyr}_{CUA}$ array, yielding pMEc1µC1.1.

The arabinose inducible P_{BAD} promoter has proved to be useful for obtaining high stop codon suppression rates.[182] However, without protein engineering of the P_{BAD} regulator protein AraC,[185] the induction works only in an 'on or off' manner.[149] To modulate aaRS expression, a gradually inducible promoter would be more suitable. The propionate inducible promoter P_{prpB} was reported to fulfill these prerequisites together with an even higher end promoter strength than P_{BAD}.[186] Hence, both promoters were introduced upstream of the *sc*PheRS and *mj*BpaRS genes together with the genes for their regulator proteins. In parallel, a STREPII-tag[187] was fused to the N-terminus of the aaRSs to enable the detection of different expression levels by Western blot analysis. Thereby, pMEc1µC3.4 (*prpR*-Strep-*sc*PheRS + *lpp*-tRNA$^{Phe}_{CUA}$), pMEc1µC8.4 (*araC*-Strep-*sc*PheRS + *lpp*-tRNA$^{Phe}_{CUA}$), pMEc1µC4.3 (*prpR*-Strep-BpaRS + 3xtRNA$^{Tyr}_{CUA}$) and pMEc1µC6.3 (*araC*-Strep-BpaRS + 3xtRNA$^{Tyr}_{CUA}$) were

assembled. Since it was not totally clear whether the synthetic Biobrick terminator downstream of the *sc*PheRS gene works efficiently, a pMEc1µC2.4 variant with the terminator from the *mj*BpaRS (pMEc1µC5.4) was assembled as well. Additionally, two vectors were generated, where (i) the p15a Ori in pMEc1µC2.4 was substituted for the RSF Ori (pMEc2µC2.4) or (ii) the F replicon-based Ori from pSCANS[167] (pMEc4µC2.4). The later is a fusion of F Ori and the IPTG inducible P1 lytic replicon. Under standard conditions plasmids containing this Ori are maintained at 1 - 2 copies in the cell. Upon IPTG addition, however, the plasmid is amplified to a high copy number. Co-transformation with pMEc4µC2.4 could yield aaRS and tRNA loads in the cell comparable to one-plasmid-systems but with the high flexibility of a two-plasmid-system regarding the exchange of the target protein.

2.3.3 Evaluation of assembled pMEc vectors

The assembled pMEc vectors were evaluated in a standard expression experiment using GFPuv as reporter protein and the constructs containing the *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ pair in different configurations (see chapter 5.3.2.3). For comparison, expressions with the original pSup-BpaRS-6TRN were made in parallel. Cell lysates were analyzed by Western blotting with an anti-His-tag antibody and by fluorescence spectroscopy (see **Figure 25**). Comparable parent GFPuv expression (no amber stop codon in the mRNA sequence, black arrow in **Figure 25A**) was detected with both the pSup and the pMEc vector background (see samples 1 and 3 in **Figure 25**). pMEc1µC1.1, which exhibits the same *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ configuration as the original pSup-BpaRS-6TRN vector, namely a constitutive aaRS promoter and six copies of tRNA, gave comparable amounts of full-length protein in the amber suppression experiment (see samples 2 and 5 in **Figure 25**) when compared to pSup-BpaRS-6TRN. From this result, it can be concluded that the newly designed modular pMEc vector family is as stable and efficient as the original pSup plasmid and can be used as novel standard platform for coexpression of aaRS:tRNA pairs. Expectedly, the reduction of tRNA copies from six to three yielded a significant loss of full-length GFPuv (see samples 4 and 5 in **Figure 25**). With three tRNA copies, only a very faint band can be detected on the Western blot. The use of the propionate inducible promoter for aaRS expression, however, highly compensated for tRNA reduction and gave a significantly higher full-length GFPuv signal on the Western blot as well as in the fluorescence assay (see sample 7 in **Figure 25**). Under the control of the propionate inducible promoter, *mj*BpaRS expression can be observed in the SDS gel (see red arrow in **Figure 25A**), while all other constructs did not produce a significant synthetase overexpression.

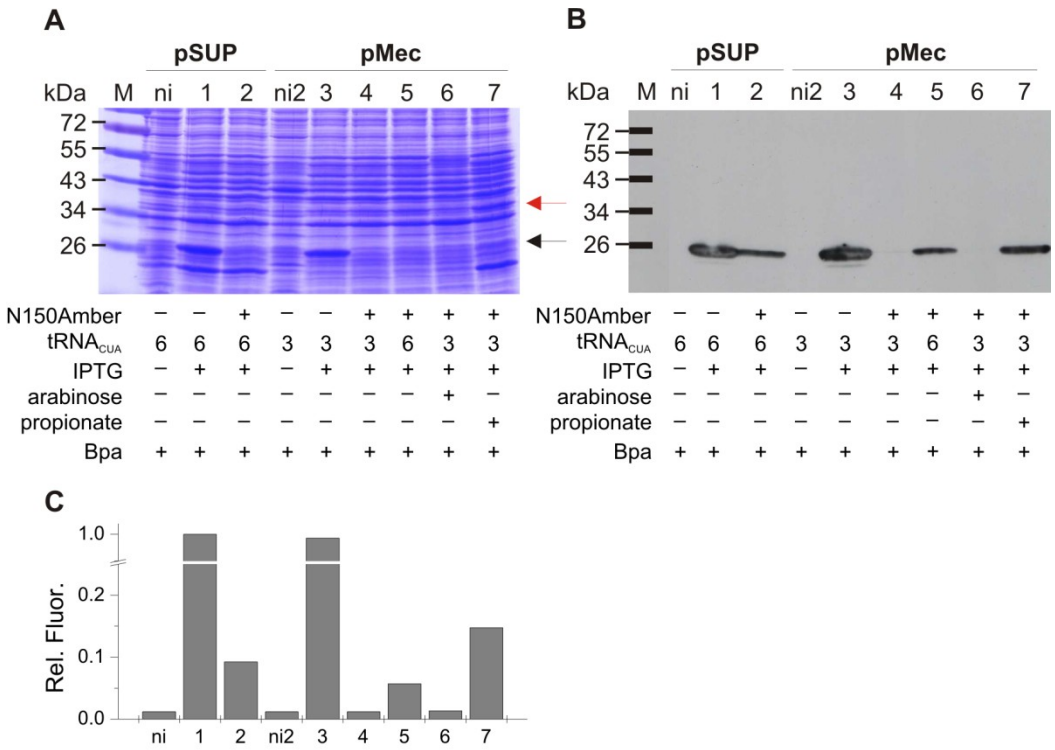


Figure 25. Evaluation of pMEc vectors containing different configurations of the *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ pair using GFPuv and Bpa. A and **B**: SDS gel and corresponding Western blot of 0.15 OD_{600} of *E. coli* strain AFW lysate per lane applied to a 12% SDS gel. **M**: marker; **ni**: non-induced sample with pSup-BpaRS-6TRN; **1**: parent GFPuv expression with pSup-BpaRS-6TRN; **2**: GFPuv(N150Amber) expression with pSup-BpaRS-6TRN; **ni2**: non-induced sample with pMEc1µC1.3 (3 copies of $tRNA_{Sup}$); **3**: parent GFPuv expression with pMEc1µC1.3 (3 copies of $tRNA_{Sup}$); **4**: GFPuv(N150Amber) expression with pMEc1µC1.3 (3 copies of $tRNA_{Sup}$); **5**: GFPuv(N150Amber) expression with pMEc1µC1.1 (6 copies of $tRNA_{Sup}$); **6**: GFPuv(N150Amber) expression with pMEc1µC6.3 (3 copies of $tRNA_{Sup}$ and P_{BAD} promoter) and **7**: GFPuv(N150Amber) expression with pMEc1µC4.3 (3 copies of $tRNA_{Sup}$ and P_{prpB} promoter). Expression of GFPuv was induced with 1 mM IPTG. In case of pMEc1µC6.3 and pMEc1µC4.3, the inducible *mj*BpaRS expression was induced by addition of 0.02% arabinose and 50 mM propionate, respectively, together with the IPTG. In the expression gel, full-length GFPuv ($M_{w, cal}$ = 27700 Da) is clearly visible in lanes 1, 3, and 5 (see black arrow). *Mj*BpaRS ($M_{w, cal}$ = 36384 Da) expression can only be detected in lane 7 (see red arrow). **C**: Relative fluorescence (Rel. Fluor.) of cleared cell lysates (see chapter 5.5.9).

Surprisingly, the use of pMEc1µC6.3 bearing *mj*BpaRS under the P_{BAD} promoter in combination with three tRNAs did not lead to *mj*BpaRS overexpression nor to amber stop codon suppression (see sample 6 in **Figure 25**). Most likely, the construct is corrupted, for example, by a not identified deleterious mutation, or the plasmid is not stable in the *E. coli* expression host. Further experiments and analysis will have to be performed on this issue. Cultures of *E. coli* that were transformed with the pSup vector reached a normal OD_{600} of 3 after overnight expression (sample 1 and 2 in **Figure 25**) while most pMEc transformed cell cultures only had an OD_{600} of 2.2 – 2.5 (sample 3 - 6 in **Figure 25**). Thus, pMEc plasmids might have a slight negative effect on cell growth, possibly by the yeast shuttle module. However, at the same time, cell cultures transformed with pMEc1µC6.3 reached an OD_{600} of 7 (sample 7 in **Figure 25**). This is remarkable because thereby pMEc1µC6.3 was not only the most efficient tandem plasmid with respect to stop codon suppression, but also the cultures transformed with it grew to two- to threefold higher density. This significantly increases the

overall yield per liter of target protein when compared to the other cultures. The observed effect is most probably not directly connected to the plasmid but rather to the aaRS promoter P_{prpB}. *E. coli* normally does not grow to more than OD_{600} = 3 – 3.5 in LB medium. The main reason for this low end density is the very low amount of utilizable carbon sources. Instead of carbohydrates, amino acids have to be used as carbon source by *E. coli* for growth.[188,189] However, by induction of aaRS expression with propionate, a new carbon source was added which can be easily used by *E. coli* as carbon and energy source.[190] This makes the P_{prpB} promoter system even more valuable because it provides a high degree of induction and the inducer can be used by *E. coli* to grow on much higher densities.

Taken together, the design, assembly and evaluation of the pMEc vector family can be regarded as very successful. The designed plasmids are highly modular and greatly facilitate the cloning of new aaRS:tRNAs pairs and/or the modulation of aaRS and tRNA gene dosages by exchanging the promoter sequences. This was best illustrated by the outperforming amber suppression rate of pMEc1µC6.3 bearing *mj*BpaRS with a propionate inducible promoter.

In combination with other antibiotic resistance markers like for kanamycin (Kan), pMEc2µ plasmids containing the RSF Ori would be compatible with pMEc1µC as well as with all standard expression vectors. This is a prerequisite for coexpression of different combinations of orthogonal pairs in parallel. Such a coexpression enables simultaneous genetic code expansion with two different noncanonical amino acids like it was recently reported by Chin[154] and Liu.[134] While these groups combined amber stop codon with frame-shift (Chin) or ochre stop codon suppression (Liu), it should basically also be possible to combine amber suppression with rare codon reassignment for such simultaneous genetic code expansions. A basic evaluation of such a reassignment will be presented in chapter 2.5.

2.4 Simultaneous genetic code expansion and engineering

As already outlined in the introduction, a combination of genetic code engineering (SPI) and expansion (SCS) would be highly desirable because both methods have certain advantages and drawbacks. While SPI only provides residue-specific incorporation, SCS is capable of introducing ncaas in a site-specific manner. However, to date, SCS is not applicable for incorporation at multiple positions to provide synergistic effects in a target protein. The simultaneous suppression of only three amber stop codons has been reported *in vivo*.[160] Furthermore, code expansion is generally not applicable to isostructural ncaas since guided evolution approaches used for orthogonal aaRSs development are not able to generate enzymes which discriminate between subtle structural variations. However, the incorporation of isostructural amino acids at multiple positions within a protein is the particular advantage of genetic code engineering.

To test whether a synergistic combination of SPI and SCS is possible, and to analyze the generality of this approach, three different model proteins (ψ-b*, EGFP and TTL) were chosen for such combination of genetic code engineering and expansion. For all three proteins a beneficial application of SPI had already been reported (see chapter 1.3.2.2). Based on these former experiments, Hpg was selected for experiments with ψ-b* (provides reactive handle), (4*S*-F)Pro for EGFP (provides 'superfolding' properties), and Nle for TTL (provides thermal activation independency). As SCS system, *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ (provided by the plasmid pSup-BpaRS-6TRN) was chosen since it proved to be highly selective for Bpa and exhibited better amber stop codon suppression rates than *sc*PheRS(T415G):*sc*tRNA$^{Phe}_{CUA_UG}$. Furthermore, Bpa equips target proteins with crosslinker and conjugation functionality.

2.4.1 Simultaneous incorporation of Hpg and Bpa into ψ-b*

For ψ-b*, no permissive sequence positions for amber stop codon suppression are known. Thus, in the first step, the crystal structure of ψ-b* was examined for solvent exposed positions occupied by aromatic amino acids to minimize a negative structural effect of the bulky Bpa residue upon incorporation. The residues Y30 and W45 were selected to introduce amber codons by site-directed mutagenesis (see **Figure 26A**). Protein expression experiments revealed that only small amounts of full-length ψ-b*(W45Bpa) were produced (**Figure 26B**, lane 2) while ψ-b*(Y30Bpa) apparently was not expressed at all (**Figure 26B**, lane 3). Nevertheless, expression cultures were subjected to protein purification, however, without success. This might indicate that not every mRNA is permissive for read-through with amber suppressor tRNAs. Nevertheless, more sequence positions will have to be screened to exactly verify this hypothesis. Still, it cannot be totally excluded that Bpa had a negative effect on ψ-b* stability, which caused rapid protein degradation. Based on these negative results, no further attempts to find other permissive sites in ψ-b* mRNA were made, and a parallel incorporation of Hpg and Bpa was not performed.

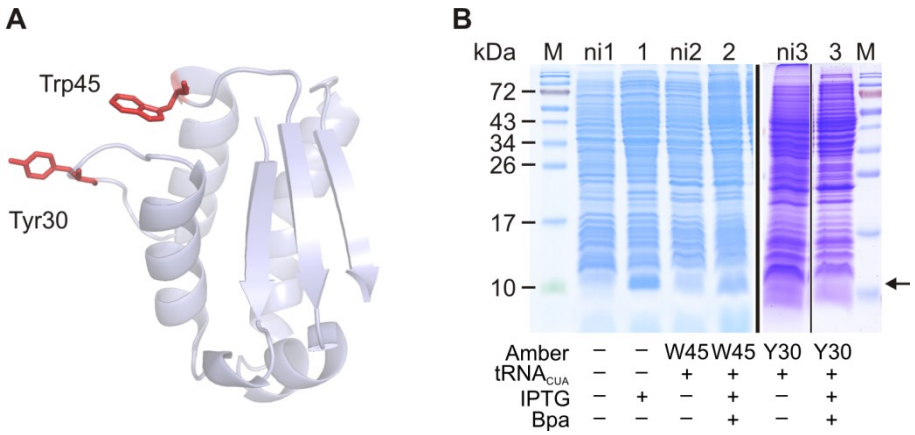


Figure 26. Screening of permissive sites for amber stop codon suppression in ψ-b* with Bpa. A: Structure of ψ-b* (PDB code 2HXX) with the solvent exposed residues Tyr30 and Trp45 highlighted in red. The residues are expected to be suitable Bpa incorporation sites with minimal steric effects of Bpa **B**: Expression of ψ-b* and its mutants in *E. coli* CAG18491 transformed with the respective pQE80L constructs and where needed additionally with pSup-BpaRS-6TRN. **1**: expression of parent ψ-b*; **2**: expression of ψ-b*(W45Bpa); and **3**: expression of ψ-b*(Y30Bpa). 0.15 OD_{600} of cell lysate per lane were applied to a 17% SDS gel. **M**: marker; **ni1**, **ni2**, and **ni3**: non-induced samples.

2.4.2 Simultaneous incorporation of (4*S*-F)Pro and Bpa into EGFP

In case of EGFP, the permissive site N150 was already known as efficient for read-through (see chapter 2.2.3). Therefore, incorporation of Bpa at position 150 was combined with a global exchange of the ten Pro residues by (4*S*-F)Pro. This was performed to test if site-specific incorporation of the photo-crosslinker Bpa could be combined with the enhanced refolding provided by the presence of (4*S*-F)Pro in the structure of EGFP.[98] The analysis results are summarized in **Figure 27**. Expression of all target proteins was detected except for the sample containing 1 mM Tyr instead of Bpa (**Figure 27**, sample c) evidencing again that BpaRS is not longer capable of charging its original substrate Tyr onto *mj*tRNA$^{Tyr}_{CUA}$. The full-length target protein ratio of the different cell lysates was calculated by ImageJ analysis of the Western blot (**Figure 27B**). A summary can be found in **Table 7**. Briefly, EGFP(N150Bpa) was expressed to approx. 46% when compared to EGFP. This reduction in full-length protein amount is due to competition of $tRNA_{Sup}$ with RF 1 and is outlined in detail in chapter 2.2.3. EGFP[(4*S*-F)Pro] exhibited only 20% of expression when compared to EGFP. Obviously, the incorporation of ten (4*S*-F)Pro into EGFP is not easily coped with by the translation machinery leading to a significant loss of protein yield. Surprisingly, EGFP(N150Bpa)[(4*S*-F)Pro] expression was likewise reduced to around 20% relative to EGFP. Rather the full-length protein amount would have been expected to drop further due to insufficient stop codon suppression. However, it might be that a simultaneous incorporation of (4*S*-F)Pro into endogenous host proteins is 'rescuing' the protein yield by facilitating the amber stop codon suppression by *mj*tRNA$^{Tyr}_{CUA}$. This issue will be discussed in more detail in chapter 2.4.6.

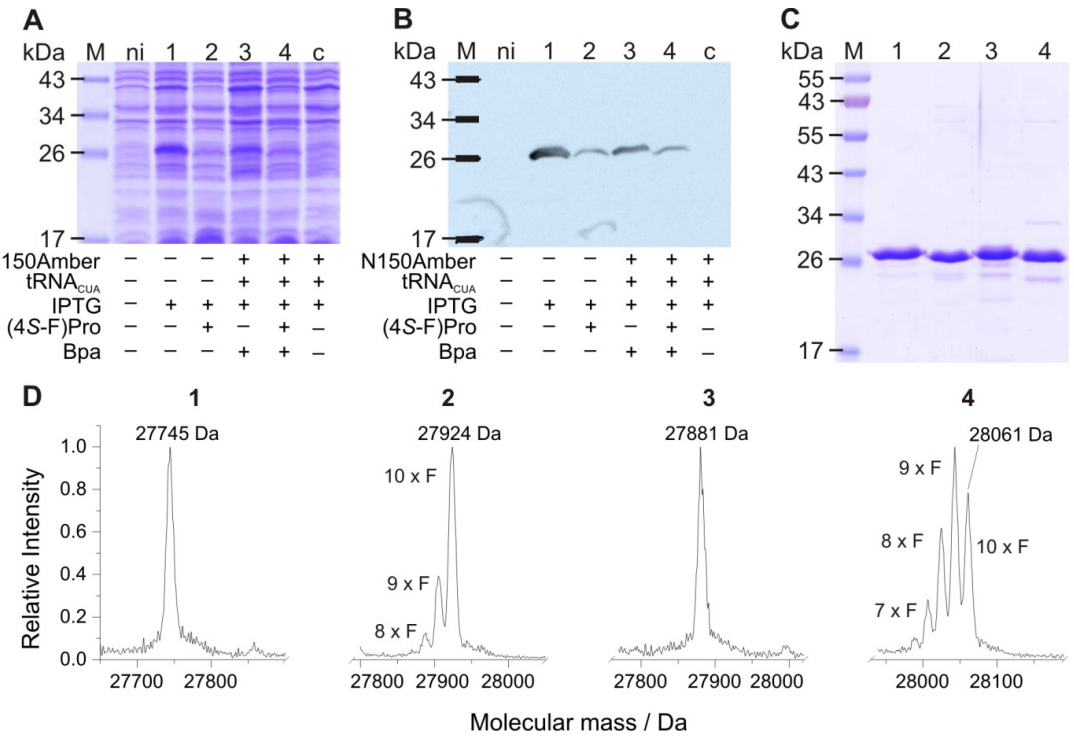


Figure 27. Expression, purity, and mass spectrometric analysis of parallel incorporation of (4*S*-F)Pro and Bpa into EGFP. A and **B**: SDS gel and corresponding Western blot of the parallel expression experiment. 0.15 OD_{600} of cell lysates were applied to a 12% SDS gel. Expression of EGFP and its variants was performed in in *E. coli* CAG18491 transformed with the respective pQE80L constructs and where needed additionally with pSup-BpaRS-6TRN. **M**: marker; **ni**: non-induced sample; **1**: parent EGFP expression; **2**: expression of EGFP[(4*S*-F)Pro]; **3**: expression of EGFP(N150Bpa); **4**: expression of EGFP(N150Bpa)[(4*S*-F)Pro], and **c:** control expression EGFP(N150Amber) with 1 mM Tyr instead of Bpa. **C**: purity gel of expressed protein variants. 3 µg of protein were applied to the SDS gel. **D**: Mass spectrometric profiles revealed a very high level of incorporation. Calculated and detected masses were as follows: **1:** $M_{w,\,cal}$ = 27745 Da, $M_{w,\,meas}$ = 27745 Da; **2:** $M_{w,\,cal}$ = 27925 Da, $M_{w,\,meas}$ = 27924 Da; **3:** $M_{w,\,cal}$ = 27882 Da, $M_{w,\,meas}$ = 27881 Da; and **4:** $M_{w,\,cal}$ = 28062 Da, $M_{w,\,meas}$ = 28061 Da. Note that the used EGFP constructs undergo partial N-terminal Met excision (NME) which complicates the mass spectra. These additional mass peaks are not shown for clarity reasons. Full mass spectra of all purified proteins can be found in the supplementary of reference [191].

All proteins were purified by Ni-NTA chromatography (**Figure 27C**) and yields correlated well with the full-length protein ratios determined from the Western blots (see **Table 7**). The analytical quality of incorporation was analyzed by ESI-MS. Expectedly, calculated and determined masses correlated very well as is shown in **Figure 27D**. In preparations containing (4*S*-F)Pro, partially labeled species with 7 - 9 fluorine atoms were also detected. However, proteins with a very high degree of substituted prolines (9 - 10 (4*S*-F)Pro) dominated in the protein preparations.

All variants were subjected to photophysical characterization. The results are summarized in **Figure 28A** and **Table 6.** The general shape of UV- and fluorescence profiles remained unchanged upon ncaa incorporation (see **Figure 28A**). The maximum excitation and emission wavelengths of the chromophore were 488 nm and 510 nm in all protein variants. In Bpa containing EGFPs, the Trp + Tyr absorbance maximum was slightly blue-shifted and the molar extinction coefficient (ε_M) at $\lambda_{max,\,Trp\,+\,Tyr}$ was increased by 30 - 40% while the chromophore absorbance remained mainly unaffected (see **Figure 28A**).

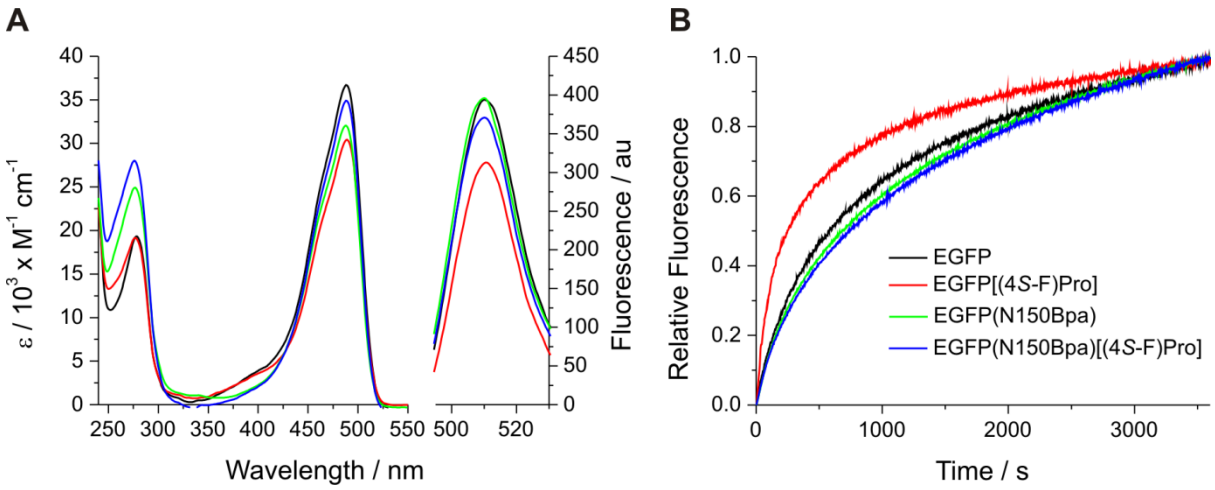


Figure 28. Photophysical characterization of EGFP variants. A: Absorbance and fluorescence profiles. **B**: Refolding studies of parent EGFP, its congeners and mutated congeners.

This effect is caused by the intrinsic absorbance of Bpa in this spectral region[192] and leads to a highly altered signal ratio between the chromophore and the Trp + Tyr absorbance (1.9 for parent EGFP, 1.3 for Bpa containing variants, see **Table 6**). However, ncaa incorporation did not affect chromophore absorbance and fluorescence. Refolding was monitored by fluorescence increase upon excitation at 488 nm (**Figure 28B**). Protein samples were denatured by boiling in 8 M urea. Refolding was induced by 1:100 dilution in refolding buffer (see chapter 5.5.10). Expectedly, Pro → (4*S*-F)Pro substitution yielded in a notably increased refolding rate as shown in **Figure 28B**. The origin of this effect had already been attributed to the favorable pre-organization of the C^{γ}-*endo* puckering by 4*S*-fluorine atoms in the related Pro residues in the EGFP structure.[98] EGFP(N150Bpa) refolded only slightly slower when compared to parent EGFP showing that Bpa incorporation at position 150 does not significantly alter the refolding of EGFP. However, the beneficiary effect of fluorinated prolines in EGFP was lost in EGFP(N150Bpa)[(4*S*-F)Pro]. In fact, the combination of Bpa and (4*S*-F)Pro was successful but the presence of Bpa could not be combined with the enhanced refolding that is induced by the presence of (4*S*-F)Pro in the structure of EGFP.

Table 6. Absorbance and fluorescence properties of EGFP, EGFP(N150Bpa), EGFP[(4*S*-F)Pro], and EGFP(N150Bpa)[(4*S*-F)Pro] analyzed under identical conditions.

	Trp+Tyr absorbance		Chromophore absorbance (Cro)			Fluorescence	
Protein	λ_{max} / nm	ε_M / M^{-1} cm^{-1}	λ_{max} / nm	$\varepsilon_{M(Cro)}$ / M^{-1} cm^{-1}	$\varepsilon_{M(Cro)}$ ε_M^{-1}	λ_{max} / nm	Intensity / au
EGFP	278	19325	488	36687	1.9	510	394
EGFP[(4*S*-F)Pro]	278	19142	488	30416	1.6	510	312
EGFP(N150Bpa)	277	24911	488	32044	1.3	510	396
EGFP(N150Bpa)[(4*S*-F)Pro]	276	28003	488	34900	1.2	510	371

Generally, it would be possible that these effects are caused by the marginally lower incorporation efficiency of (4*S*-F)Pro in EGFP(N150Bpa)[(4*S*-F)Pro] (see **Figure 27C**). However, in previous experiments[98] it was observed that a global labeling > 60% with (4*S*-F)Pro induces better refolding. Thus, the comparably slower refolding of EGFP(N150Bpa)[(4*S*-F)Pro] cannot be attributed to the presence of these, not fully substituted, species.

Position 150 in *av*GFPs is solvent exposed at the surface of a β-sheet near the chromophore. In general, the substitution of Asn with hydrophilic amino acids such as lysine was found to increase the fluorescence intensity. These results were attributed to a higher local hydrophilicity at the surface and enhanced hydrogen-bonding around the chromophore.[193] It can be speculated that the presence of a hydrophobic Bpa at position 150 impairs EGFP folding. However, following this line of argumentation, EGFP(N150Bpa) folding would also have to be significantly impaired. To test, whether the hydrophobic character of Bpa plays a role, pyrrolysine analogs could be introduced via PylRS derived orthogonal pairs.[134,160,194] If the chemical character of Bpa negatively influences folding of EGFP(N150Bpa)[(4*S*-F)Pro], it is reasonable to expect that Pyl analogs together with globally incorporated (4*S*-F)Pro may yield proteins with enhanced refolding capacity and a site-specifically introduced chemical handle (e.g. photo-crosslinking or bio-orthogonal conjugation via Huisgen cycloaddition). In this study, position 150 was chosen for Bpa incorporation into EGFP because it had been reported as highly permissive site for amber stop codon suppression in *av*GFPs before.[153] A more detailed study on combined site-specific incorporation and enhanced folding by (4*S*-F)Pro will also have to include the search for and subsequent evaluation of other permissive positions in EGFP.

In summary, the parallel genetic code engineering and expansion of EGFP was successful but the beneficial effects of Bpa and (4*S*-F)Pro could not be successfully combined.

2.4.3 Simultaneous incorporation of Nle or Aha and Bpa into TTL

To date, TTL has not yet been used for site-specific incorporation of ncaa and no permissive amber stop codon positions were known. Thus, prior to SPI and SCS combination, the TTL sequence was screened for effective read-through positions.

2.4.3.1 Test for permissive amber stop codon positions

The aromatic residues Y9, F37, W136, H233 and W240 were chosen for the stop codon screen to minimize the impact of the bulky Bpa residue on TTL structure upon incorporation (see **Figure 29**). Only Y9 and W240 are partially surface exposed. All other amino acids are buried in the TTL hydrophobic core. D221 was also included in the screen because it is a completely surface exposed residue, though not aromatic.

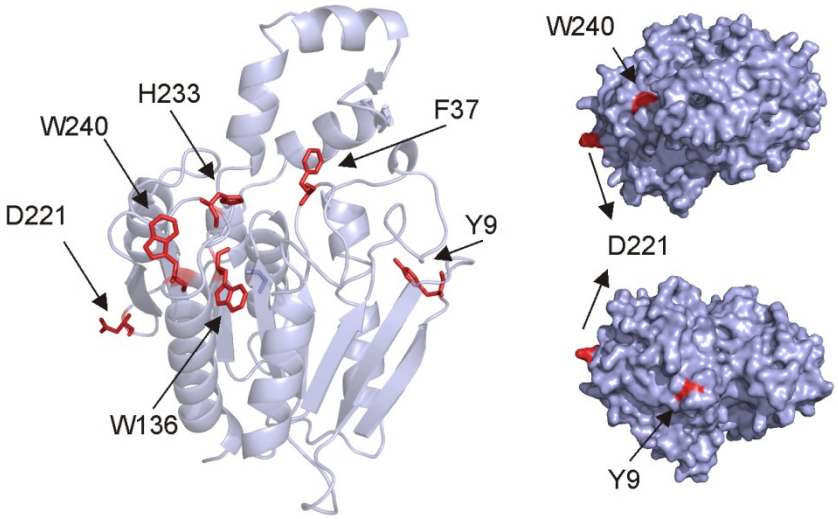


Figure 29. Structure of TTL with highlighted amino acids mutated to stop codons. Left: Cartoon presentation of TTL with highlighted amino acid positions (red) mutated to amber stop codons. **Right:** Surface presentation of TTL. Lower image is a 180° vertical rotation of the upper image. The depicted structure was kindly provided by Pinotsis (unpublished results).

Amber stop codon mutations were introduced into pQE80L-TTL-H6 by site-directed mutagenesis (see chapter 5.1.2.2), and the generated mutants were expressed in 200 mL scale in *E. coli* AFW transformed with the respective pQE80L construct and pSup-BpaRS-6TRN according to chapter 5.3.2.3. Full-length protein expression was monitored by SDS-PAGE and subsequent Western blot analysis (see **Figure 30**). Band intensities were calculated using ImageJ as previously indicated. Full-length protein was detected in all TTL preparations but the band intensities varied significantly among the different mutants. As mentioned above, mutation positions were selected such that the disturbing effect of the bulky Bpa in the TTL structure was minimized. Thereby, misfolded protein species and the possibility of their immediate degradation was tried to be prevented. In this way, it should be possible to attribute the differences in full-length protein amount directly to the position and the sequence context of the stop codon in the mRNA.

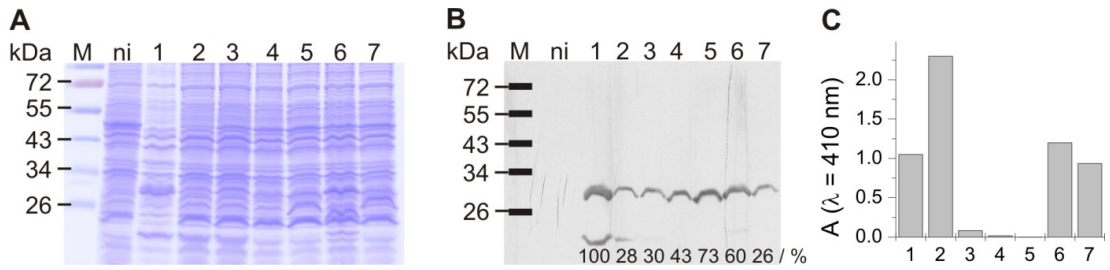


Figure 30. Expression and activity of TTL containing Bpa at different positions in the protein. The TTL mutants were expressed in *E. coli* AFW transformed with the respective pQE80L construct and pSup-BpaRS-6TRN. **A** and **B**: SDS gel and corresponding Western blot of cell lysates. 0.15 OD_{600} per lane were applied to a 12% SDS gel. **M**: marker; **ni**: non-induced sample; **1**: TTL; **2**: TTL(Y9Bpa); **3**: TTL(F37Bpa); **4**: TTL(W136Bpa); **5**: TTL(H233Bpa); **6**: TTL(D221Bpa); **7**: TTL(W240Bpa). Full-length protein amounts are given as percent of parent protein amount (at the bottom of **B**). **C:** Enzyme activity of TTL and its variants is displayed as *p*NP absorbance at λ = 410 nm. Activity was determined by the *p*-nitrophenyl palmitate (*p*NPP) assay at 75 °C for 20 min. Samples 3 – 5 were incubated for 1 h.

Lowest amber suppression took place at the very 5'- and 3'-end of the mRNA with Y9amber being suppressed by around 28% and W240amber by around 26% (see **Figure 30B**). Read-through increased gradually with the sequence position of the stop codon from Y9amber to H233amber which exhibited 73% suppression, the highest value of all tested stop codons. These data show that amber suppression is strongly dependent on the stop codon position in the mRNA sequence and/or the sequence context. A detailed analysis of the sequence context of stop codon suppression can be found in chapter 2.4.5.

All variants were purified and lipase activity was subsequently assessed with the *p*-nitrophenyl palmitate assay (*p*NPP, see chapter 5.4.3). While the TTL variants with surface exposed Bpa positions (positions 9, 211, and 240; see **Figure 29**) were active, the presence of Bpa at positions 37, 136, and 233 impaired enzyme activity (see **Figure 30C**). It was expected that lipase activity would be absent from the F37Bpa and H233Bpa mutants because both residues are catalytically very important. In particular, H233 is the catalytic His in the Ser-Asp-His hydrolysis triade and F37 is part of the oxyanion hole which is important for substrate binding and stabilization of the intermediate state during catalysis.[195]

In summary, all tested stop codon positions were suppressed but to a significantly different extent. For combination of genetic code expansion and engineering, the TTL mutant with the amber stop codon at position 221 was chosen as the read-through rate was high, the resulting TTL(D221Bpa) was still active and Bpa was surface-exposed upon incorporation.

2.4.3.2 Combining Bpa and Nle incorporation in TTL

For combination of genetic code engineering and expansion in TTL, Bpa was used in combination with Nle since its global incorporation generates an 'always active' TTL congener where no thermal activation is needed anymore.[99,112] Expression experiments were performed in the Met auxotrophic *E. coli* CAG18491. The analysis results are summarized in **Figure 31**.

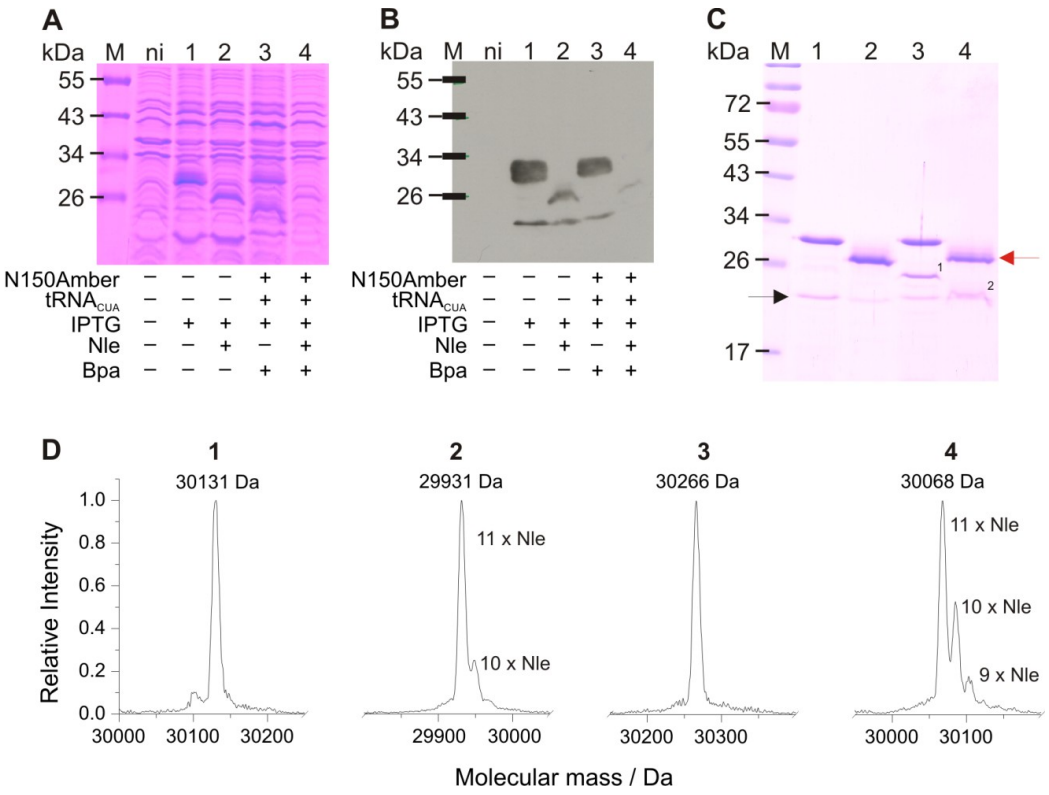


Figure 31. Expression, purity, and mass spectrometric analysis of parallel incorporation of Nle and Bpa into TTL. A and **B**: Expression gel and corresponding Western blot of the parallel expression experiment. Parent protein and congeners were expressed in *E. coli* CAG18515 which were transformed with the respective pQE80L construct and mutants with the respective pQE80L construct and pSup-BpaRS-6TRN. 0.15 OD_{600} of cell lysates per lane were applied to a 12% SDS gel. **M**: marker; **ni**: non-induced sample; **1**: TTL; **2**: TTL[Nle]; **3**: TTL(D221Bpa); **4**: TTL(D221Bpa)[Nle]. **C**: purity gel of expressed protein variants. 3 µg of protein were applied to the SDS gel. **D**: The mass spectra show a very high degree of incorporation. Calculated and detected masses were as follows: **1**: $M_{w, cal}$ = 30131 Da, $M_{w, meas}$ = 30131 Da; **2**: $M_{w, cal}$ = 29932 Da, $M_{w, meas}$ = 29931 Da; **3**: $M_{w, cal}$ = 30266 Da, $M_{w, meas}$ = 30266 Da; and **4**: $M_{w, cal}$ = 30068 Da; $M_{w, meas}$ = 30068 Da. Complete mass spectral data with truncated TTL species can be found in **Figure 42** in the appendix.

In accordance with the EGFP experiments, full-length protein expression was detected for TTL, its congener TTL[Nle], the mutant TTL(D221Bpa) and the mutated congener TTL(D221Bpa)[Nle] (see **Figure 31A** and **B**). Interestingly, Nle incorporation led to a change in the electrophoresis characteristics of TTL. TTL[Nle] and TTL(D221Bpa)[Nle] migrated faster in the SDS gel (see red arrow in **Figure 31C**) although their M_w is only marginally decreased by incorporation.

The amounts of full-length protein on the gel and of purified protein varied greatly, as summarized in **Table 7**. All preparations were purified by Ni-NTA chromatography (see chapter 5.3.3) and purity was analyzed by SDS-PAGE (see **Figure 31C**). In contrast to EGFP, TTL preparations contained other significant bands beside the full-length protein. An N-terminal truncation was present in all preparations (black arrow in **Figure 31C**, see 2.2.2.2). The truncated TTL part arising from premature translation termination at the amber stop codon (bands numbered 1 and 2 in **Figure 31C** and respective mass spectra in **Figure 42** in the appendix) was also observed. In general, the copurification of the stop codon truncation product should be avoided by the use of a C-terminal His-tag because only full-length protein

species carry the His-tag and can be purified by Ni-NTA chromatography. TTL, however, was crystallized as a dimer (unpublished data, see **Figure 35**) and is most probably also active as a dimer. Thus, a stable TTL heterodimerization between full-length and truncated produt is most probably the main reason for the observed copurification. Indeed, denaturing Ni-NTA chromatography led to a removal of the truncated version, but the subsequently refolded full-length lipase was only marginally active. ESI-MS revealed correct masses for all TTL preparations (see **Figure 31D**). However, small amounts of TTL with only ten or nine incorporated Nle residues were also detected. As for EGFP, the combination of genetic code expansion and engineering worked perfectly with respect to noncanonical amino acid incorporation.

The activity of all TTL preparations was tested using the *p*NPP assay. **Figure 32** demonstrates that a global Met → Nle substitution yielded TTL[Nle] with an enhanced activity under assay conditions when compared to TTL. In contrast, TTL(D221Bpa) activity was not significantly different from that of parent TTL; however, TTL(D221Bpa)[Nle] again exhibited a significantly higher activity in comparison to TTL but lesser than TTL [Nle].

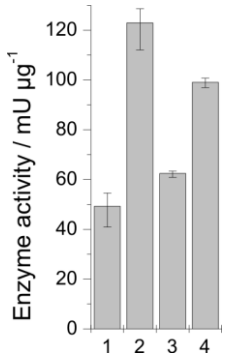


Figure 32. Activity of TTL preparations. Lipase activity was determined using 5 µg enzyme in a *p*NPP assay at 70 °C for 10 min after preincubation for 20 min at 70 °C. **1**: TTL; **2**: TTL[Nle]; **3**: TTL(D221Bpa); **4**: TTL(D221Bpa)[Nle].

Thus, unlike EGFP(N150Bpa)(4*S*-F)Pro, the parallel genetic code expansion and engineering led to a combination of two advantageous characteristics, namely enhanced enzyme performance under assay conditions which can be attributed to the incorporation of Nle and the possibility of photo-crosslinking or coupling reactions provided by Bpa.

2.4.3.3 Simultaneous incorporation of Aha and Bpa into TTL

To further explore the possibilities of double incorporations by parallel genetic code expansion and engineering, TTL was also expressed with Bpa and Aha. The product TTL(D221Bpa)[Aha] would allow a double conjugation by copper(I)-catalyzed azide–alkyne Huisgen cycloaddition (as an example for click chemistry) at Aha residues and oxime ligation on Bpa.[196] The protein was expressed as described above for TTL(D221Bpa)[Nle]. The expression gel and ESI-MS spectrum are shown in **Figure 33**. The sample numbering was continued from chapter 2.4.3.2. Basic Huisgen cycloaddition experiments were performed to evaluate if the protein is suitable for such conjugations. Further conjugation studies are performed in cooperation with Dr. Hackenberger at Freie Universität Berlin, Germany.

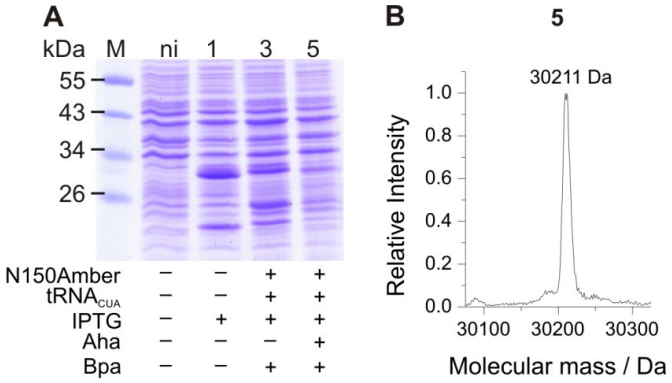


Figure 33. Expression and mass spectrometric analysis of TTL(D221Bpa)[Aha]. **A**: Parent protein and its congeners were expressed in *E. coli* CAG18515 transformed with the respective pQE80L construct and mutants with the respective pQE80L construct and pSup-BpaRS-6TRN. 0.15 OD_{600} of cell lysates per lane were applied to a 12% SDS gel. **M**: marker; **ni**: non-induced sample; **1**: expression of parent TTL; **3**: expression of TTL(D221Bpa); **5**: expression of TTL(D221Bpa)[Aha]. **B**: The determined M_w of TTL(D221Bpa)[Aha] was in full agreement with the calculated mass. No partially labeled species were identified. Complete mass spectral data with truncated TTL species can be found in **Figure 42** in the appendix.

TTL(D221Bpa)[Aha] was expressed in higher amounts than TTL(D221Bpa)[Nle] (see **Figure 31A** and **Figure 33A**). Also the yield of purified full-length protein was higher (see **Table 7**) and the ESI-MS analysis revealed full labeling with Aha without any detectable amounts of partially labeled species. Cycloadditions of TTL(D221Bpa)[Aha] with the sugar conjugates propargyl lactoside (Lac, + 380 Da per residue) and propargyl *N*-acetylglucosaminide (GlcNAc, + 259 Da per residue) were performed, leading to SDS gel bands with respectively higher M_w (see **Figure 34A**). These results indicate an at least partially successful sugar conjugation. The protein recovery from the cycloaddition batch was around 30%. An ESI-MS analysis of the potentially artificially glycosylated TTL(D221Bpa)[Aha] samples was not successful probably because highly glycosylated proteins do not effectively ionize under the standard ESI-MS settings used in the MPI core facility (personal communication with Mrs. Weyher-Stingl). Indirectly, nevertheless, this finding is an indication for a high degree of conjugation. **Figure 34B** shows that TTL(D221Bpa)[Aha]-{triazole}Lac and TTL(D221Bpa)[Aha]-{triazole}GlcNAc had no enzymatic activity.

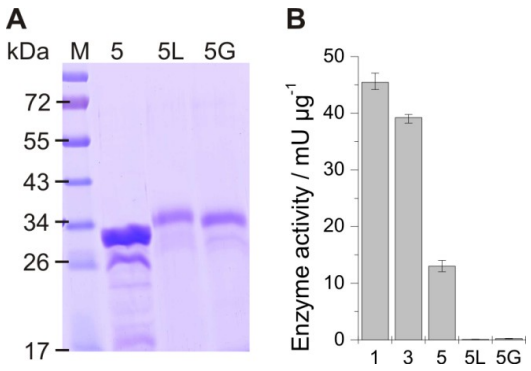


Figure 34. Purity and activity of TTL(D221Bpa)[Aha] and its Huisgen cycloaddition products. A: Purified TTL conjugation products: **1**: TTL; **3**: TTL(D221Bpa); **5**: TTL(D221Bpa)[Aha]; **5L**: TTL(D221Bpa)[Aha] conjugated with propargyl lactoside (Galβ(1,4)-Glcβ-O-CH_2-CCH); and **5G**: TTL(D221Bpa)[Aha] conjugated with propargyl *N*-acetylglucosaminide (GlcNAcβ-O-CH_2-CCH). Note that the increase in mass generated by the cycloaddition reaction can clearly be seen on the SDS gel. **B**: Enzymatic activity of TTL conjugation products: activity was determined using 4 µg enzyme in a *p*NPP assay at 70 °C for 15 min.

The enzyme activity seemed to be completely abolished by sugar conjugation. These results suggest that cycloaddition may have taken place not only at solvent exposed Aha positions but also in the protein hydrophobic interior (see **Figure 35B**). It is known that TTL is a highly flexible enzyme.[99] Thus, a partial solvent exposition and subsequent labeling of some normally buried Aha residues is possible. However, conjugation of sugar moieties to interior Aha residues may have led to protein unfolding. Alternatively, it is possible that the cycloaddition conditions resulted in (partial) unfolding of the lipase leading to cycloaddition at Aha positions in the protein interior and subsequent protein inactivation. To gain more insight about the residues modified, Orbitrap analysis was performed. The results are summarized in the appendix (see chapter 6.3.2).

As already discussed in chapter 2.2.2.1, Orbitrap data only provides semi-quantitative information about ncaa incorporation. Nevertheless, the analysis of TTL(D221Bpa)[Aha] supports the ESI-MS result of a very high substitution of Met by Aha residues. However, certain positions retain more residual Met than others. Interestingly, there seems to be a very vague correlation between the proximity of two Met positions and a residual Met retention. The closer two positions are in the primary sequence the more residual Met was found in the detected peptides. This is best illustrated for Met17 and Met18 where > 30% of the detected peptides contained at least one position occupied by a Met (see chapter 6.3.2.1). It is known from literature that the ultimate amino acid in the nascent protein chain has an influence on the recruitment of the next aminoacyl-tRNA in the ribosomal A-site (see chapter 1.2.4).[68] Thus, it is possible that an incorporated Aha negatively influences the recruitment of a second one in favor of a Met. However, the trend for more Met incorporation at closer AUG codon positions in the mRNA is not statistically significant and detailed studies should be performed to elucidate this issue.

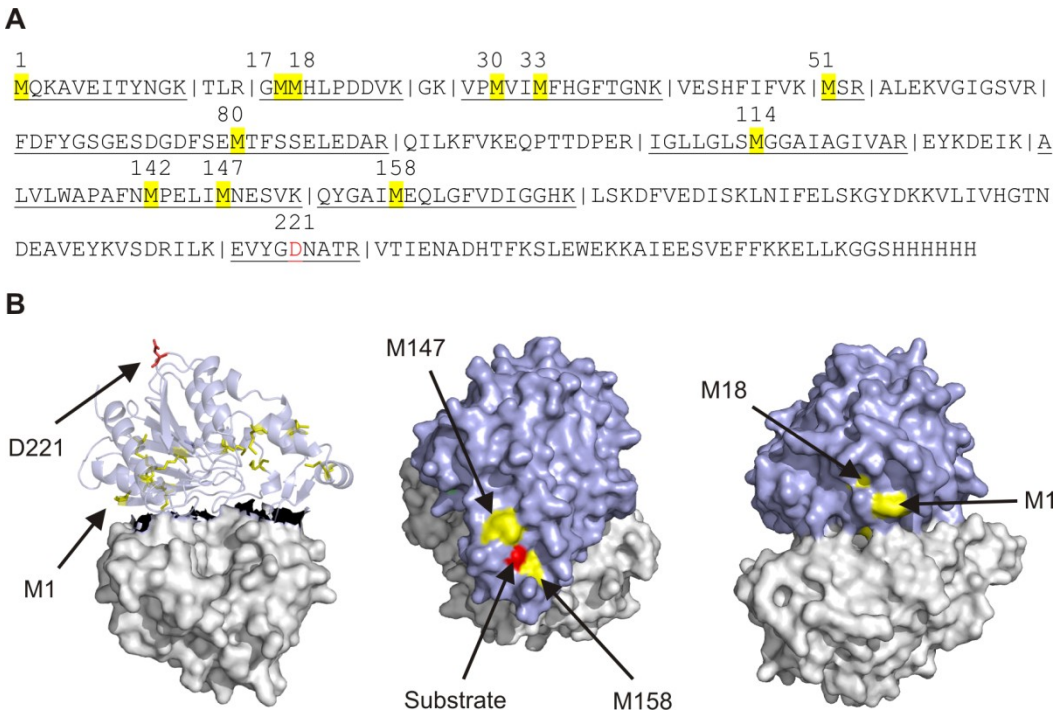


Figure 35. Distribution of Met residues in the primary sequence (A) and crystal structure (B) of TTL. Met residues are highlighted in yellow, D221 in red. **A**: underlined sequences are peptides containing Met residues after trypsin digestion. **B**: TTL is presented with both subunits to see the dimer interface. **Left**: monomer presentation as ribbon plot; **Middle** and **Right**: Surface presentation of TTL dimer. It is clearly visible that only M1, M18, M147 and M158 are surface exposed. The depicted structure was kindly provided by Pinotsis (unpublished results).

Orbitrap analysis of sugar conjugated TTL(D221Bpa)[Aha] detected clicked GlcNAc at positions 1, 17, 18, 30/33, 80, 114, 142/147 and 158. Lac was found only at positions 17/18 and 142/147, but the peptide counts and protein coverage were much lower than for TTL(D221Bpa)[Aha]-{triazole}GlcNAc (see chapter 6.3.2.2 and 6.3.2.3). Any statement on the conjugation efficiency at each single Aha position is speculative because the peptide count was generally low and it is totally unclear how particularly glycosylated peptides behave under the standard Orbitrap conditions used. However, taking into account the issues discussed, it is reasonable to suppose a rather efficient conjugation at most Aha residues. This was not expected in the first place because normally only surface exposed residues are modified (unpublished results of other proteins). These issues were out of the focus of this work and are currently more deeply investigated in the group of Dr. Hackenberger at Freie Universität Berlin, Germany. Within this cooperation, a potential triple conjugation at TTL(D221Bpa, S261C)[Aha] will be investigated as well (see **Figure 43** in the appendix for analytics of the preparation). This mutant TTL congener contains a Cys at the C-terminus for 'classical' coupling e.g. with *N*-ethylmaleimide derived ligands. In addition, various cycloaddition conditions will be tested to clarify if the reaction buffer used in this study already caused partial denaturation of TTL. Furthermore, a mutational screen will be performed to identify and remove cycloaddition positions which might impair enzyme activity when coupling takes place.

2.4.4 Evaluation of protein yield in combined SPI and SCS experiments

Table 7 gives an overview of full-length protein amount in the cell lysates determined from the Western blots and actually purified full-length protein from the two successful SPI + SCS combination experiments. Generally, the yield of purified protein correlates well with the full-length band intensities, except for TTL(D221Bpa)[Nle] where more protein was purified than expected from gel analysis. This is most probably due to slightly lower cell lysate amounts applied to the SDS gel when compared to the other samples (see **Figure 31A**). The protein yield generally decreases with higher complexity of the incorporation experiment. Variants containing SCS and SPI incorporated amino acids were expressed in lowest amounts. In EGFP and TTL with Nle expression experiments, the double incorporation resulted in a slightly lower incorporation efficiency of the amino acid introduced by SPI (see **Figure 27D-4** and **Figure 31D-4**). However, this does not seem to be a general rule since experiments with TTL and Aha yielded a very homogeneous TTL(D221Bpa)[Aha] sample (see **Figure 33**). Obviously, the incorporation efficiency depends on specific amino acid combinations. Furthermore, incorporation yields do not seem to be combinatorial, i.e. tendencies of yields from single incorporation experiments cannot be extrapolated to double incorporations. For example, TTL is expressed in comparable amounts with Nle or Aha.[99] In combination with Bpa, however, experiments using Nle yielded less protein than with Aha (see **Table 7**).

Table 7. Summary of full-length protein amounts and purified protein yields in SPI + SCS combination experiments.

Protein	Full length protein amount / % [a]	Protein yield / $mgL^{-1}OD_{600}^{-1}$ [b]	Relative protein yield / %
EGFP	100	17.0	100
EGFP[(4*S*-F)Pro]	20	1.7	10
EGFP(N150Bpa)	46	5.5	32
EGFP(N150Bpa)[(4*S*-F)Pro]	20	0.6	4
TTL	100	26.4	100
TTL[Nle]	27	10.0	38
TTL(D221Bpa)	62	18.3	69
TTL(D221Bpa)[Nle]	5	6.4	24
TTL(D221Bpa)[Aha]	-[c]	13.1	50

[**a**] The full-length protein amount in whole cell lysates was determined by integrating band intensities on the Western blots (Figure 27B and Figure 31B) using the software ImageJ.

[**b**] Protein yield refers to the amount of full-length protein effectively purified from 1 L of culture. For mutant TTL(D221Bpa) and its congeners the amount of co-purified truncated protein (which arose from the translational termination at the amber stop codon) was subtracted from the protein yield. To account for culture growth differences during the expression phase, the yield was referenced to 1 OD_{600}.

[**c**] Full length protein amount was not determined for TTL(D221Bpa)[Aha].

In general, the ncaas selected for SPI incorporation ((4*S*-F)Pro in EGFP and Nle in TTL) yielded comparatively low protein amounts. This is rather an exception than a rule because normally SPI produces high yields of labeled protein, often comparable to the expression of the parent protein. For example, in our recent study about global incorporation effects on TTL, ten different ncaas were tested and only two caused significantly lower yields than parent TTL. The incorporation of two others, strikingly, yielded as much as twice the amount of parent protein.[99]

2.4.5 Dependence of suppression rate on mRNA sequence context

As previously discussed in chapter 2.2.3, the stop codon suppression rate significantly depends on the intrinsic properties of the orthogonal aaRS:$tRNA_{CUA}$ system employed, e.g. quality of the orthogonal pair, tRNA copy number and promoter strengths of the different components. Moreover, the presented experiments provide further evidence that the stop codon suppression rate is target sequence dependent and varies highly between different positions in the same mRNA and among different mRNAs. The question now is whether a rule can be identified to predict — at least to a certain degree — the suppression rate when a stop codon is placed into a certain context on an mRNA sequence. Many studies are known from the literature investigating the importance of sequence context on translation termination (see chapter 1.2.4).[63,65,68,76] In summary, the base in 3'-position of the stop codon as well as the ultimate and penultimate amino acid of the nascent protein chain have significant impact on the stop codon suppression efficiency. Adenine as 3'-base is believed to promote the highest read-through, whereas uracil supports termination. G and C seem to play only a moderate role. In case of amber stop codons, the penultimate amino acid of the nascent polypeptide chain is less important than the ultimate. At this position, amino acids with low van der Waals radii and perturbing effect on ordered secondary structures (e.g. α-helices) have a positive effect on read-through. Especially, Thr and Pro at the ultimate position in the nascent protein chain have a positive effect on suppression.[67] Interestingly, the combinations Thr-amber stop and Pro-amber stop are absent in the whole *E. coli* genome, suggesting a strong selection against such combinations.[69] However, it is difficult to apply these rules to the results gained in this study because all aspects are combinatorial and it is not clear to which extent the different parameters influence the overall read-through rate.

Figure 36 depicts the 5'- and 3'-sequence context of the stop codons used in the present study and from investigations recently published by Schultz and coworkers. They also used a *mj*TyrRS:*mj*$tRNA^{Tyr}_{CUA}$ derived orthogonal pair to study suppression efficiency at different amber stop codon positions in GFPuv and whale sperm myoglobin.[149] Most of the investigated positions contain A or G at the 3'-position, except for position 30 in ψ-b* where T is the 3'-vicinal nucleobase. Thus, it could be speculated that this context hampered the suppression at this position in ψ-b*. For real conclusions or predictions based on the above described indications, however, the effects of all possible 5'- and 3'-context combinations would have to be evaluated in a single experiment. This would result in a score matrix which could then be applied in a bioinformatic program for read-through prediction.

The question now is whether the read-through determinants discussed above are in fact the critical factors for stop codon suppression with ncaas during heterologous gene expression. The data obtained from the presented study and from reference [149] suggest a totally different scenario which may be more important and thus superimposed over the above discussed determinants. Generally, read-through efficiency seems to peak at approx. 80% of a protein sequence and decreases towards the 3'- and 5'-ends of the mRNA. In **Figure 36** all tested stop codon positions are aligned with respect to their relative position in the target gene sequence. In GFPuv, TTL and myoglobin read-through is low in the 5'-region of the mRNA. Best suppression is found at around 75 - 85% of mRNA length before a harsh decrease at the very 3'-end takes place. Moreover, it could be that some mRNAs are generally not permissive for amber stop codon read-through as suggested by the results of ψ-b*.

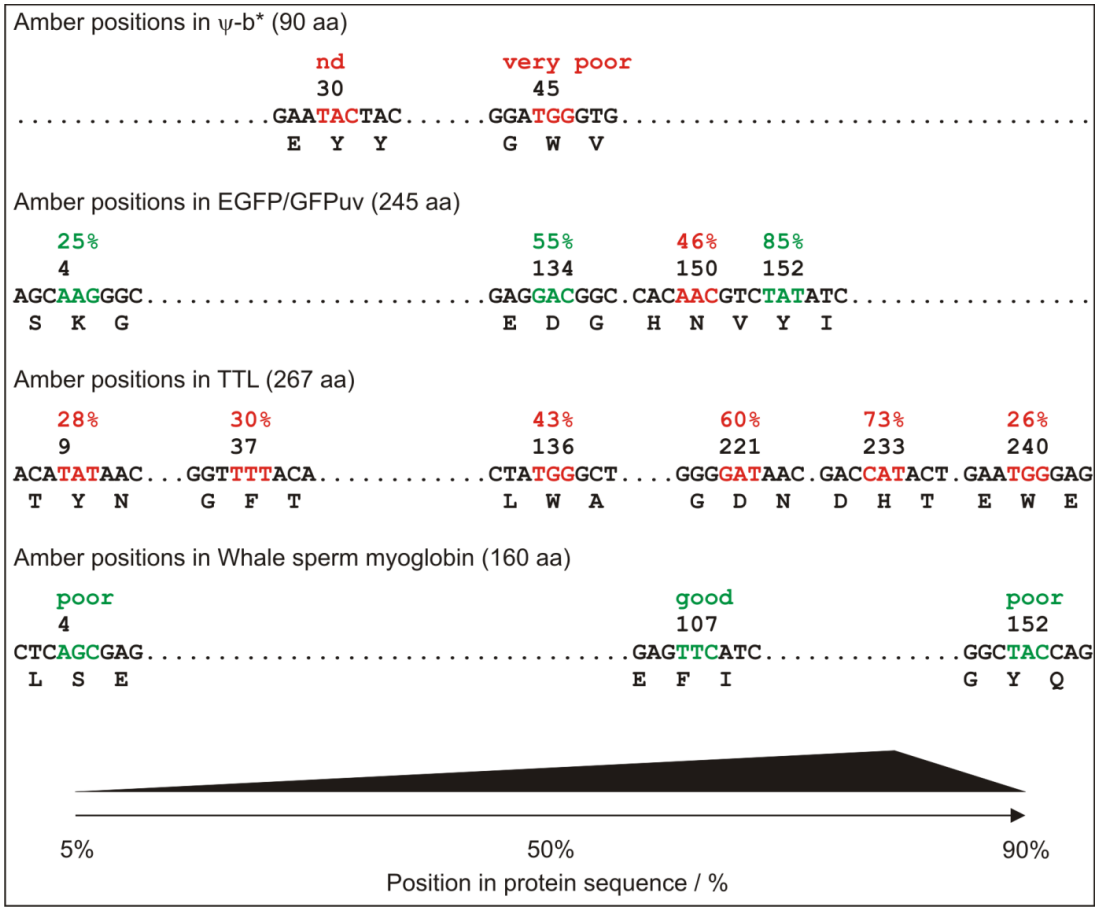


Figure 36. Suppression efficiency depending on amber stop codon location in the mRNA sequence. The sequence position and 5' - 3' context of introduced stop codon sites of the present study (red) and from the literature[149] (green) are depicted. The number above the DNA sequence is the protein sequence position of the stop codon. Suppression efficiencies or general suppression tendencies are given above the stop codon position. Note that suppression experiments of reference [149] were performed with the newly evolved pEVOL plasmid series being approx. threefold more effective in suppression. In TTL, stop codon position 221 (after 80% of the sequence) and 233 (after 87% of the sequence) are not exactly within the scale of the x-axis.

A real conclusion cannot be drawn with a dataset of three proteins; however, the proposed theory would be a good hypothesis for the design of more general experiments, which could certainly shed more light on the important issue of stop codon read-through rates at different protein positions.

2.4.6 Does SPI promote amber suppression?

Having a closer look at the protein preparations from parallel SPI and SCS incorporation experiments, the question arises whether the stop codon suppression rate maybe increases when an amino acid is incorporated by SPI in parallel. There are some indications supporting this hypothesis. First, in the purified protein samples of TTL(D221Bpa)[Nle] and TTL(D221Bpa)[Aha], the truncation product resulting from premature translation termination at the internal amber stop codon is present in lower amounts (see **Figure 37**, red and green arrow). This trend is also affirmed by the relative intensities of full-length and truncated protein species in the ESI-MS spectra (see **Figure 42**). In the cell lysates of TTL(D221Bpa, S261C)[Aha] expression, the truncated TTL also seems to be present in significantly lower amount than in TTL(D221Bpa) lysates (compare **Figure 43A**, lane 1 with **Figure 31A**, lane 3).

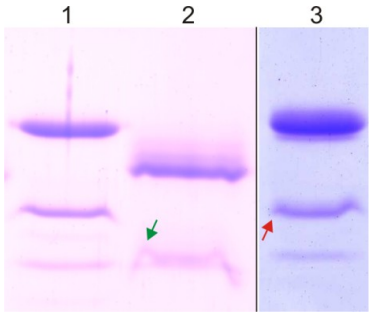


Figure 37. Possible increase of stop codon suppression by parallel SPI. Comparison of purified TTL(D221Bpa) (**1**), TTL(D221Bpa)[Nle] (**2**) and TTL(D221Bpa)[Aha] (**3**). The preparations from combined SCS + SPI experiments contain lower amounts of truncated protein species from premature translation termination at the internal amber stop codon.

Another hint supporting this hypothesis could be that the full-length protein amount of EGFP[(4*S*-F)Pro] was the same as EGFP(N150Bpa)[(4*S*-F)Pro] on the expression gel (see **Figure 27**). Normally, the full-length protein level should decrease because the additional Bpa incorporation results in truncated product.

A possible explanation for this finding can be found in the nature of the SPI procedure. Since SPI exploits the substrate promiscuity of the endogenous aaRSs, not only the target protein but also all other proteins in the *E. coli* proteome will be labeled with ncaa. Thus, also RF1 — responsible for translation termination at the amber stop codon — will be modified when an SPI experiment is performed. However, the incorporation of ncaas in RF1 might lead to its partial or full inactivation making the stop codon suppression by *mj*tRNA$^{Tyr}_{CUA}$ more efficient. Indeed, such an effect triggered by ncaas on the endogenous translation system was recently shown for T7 RNA polymerase.[197] In T7 RNA polymerase based expression

systems Aha can be incorporated with high yields but expression with Hpg did not yield target protein at all. Lee and coworkers were able to show that these observations result from a selective inactivation of T7 RNA polymerase by the incorporation of Hpg. Such an inactivation by ncaa incorporation could be possible also for RF1.[197]

2.4.7 Summary and concluding remarks

The presented thesis aimed to combine genetic code engineering and expansion. In general, the parallel use of both methods for protein modification is indeed possible. In addition, the example of TTL(D221Bpa)[Nle] proves that this can also lead to a highly modified protein in which two desired features (in this case enhanced activity and possibility of crosslinking/conjugation) are successfully combined. However, the present study elucidated several critical issues that have to be considered when using the SCS method in general and, in particular, when combining SCS with SPI. Extensive pre-evaluation has to be performed prior to such experiments since not all proteins seem to be suitable for SCS experiments as shown for ψ-b*. Furthermore, there are great differences in 'permissivity' of stop codon positions within an mRNA sequence. Experimental and literature data suggest that read-through peaks at around 80% of a protein sequence and decreases towards 3' and 5' ends of the mRNA. This pattern, however, has to be verified by studying SCS in several other proteins.

Another critical point is the nature of the orthogonal aaRS:$tRNA_{CUA}$ pair employed for incorporation. The Bpa system works well with respect to stop codon suppression but other orthogonal pairs are much less efficient.[149] Their use may yield a significantly smaller amount of labeled full-length protein. In any case, protein yields vary significantly with the different combined amino acids although prior single incorporations give equal yields. Moreover, not all ncaa features can be successfully combined as shown in case of EGFP(N150Bpa)[(4*S*-F)Pro].

2.5 Towards AGG codon reassignment for ncaa incorporation

As already outlined in chapter 2.3.2, a simultaneous genetic code expansion with various orthogonal pairs in parallel would be also highly desirable. However, to date, only very few 'free' coding units are available for encoding ncaas *in vivo* (amber, ochre and a few quadruplet codons). Furthermore, amber stop codon suppression will always be limited until a viable strain without RF1 is available (see chapter 1.3.3). Quadruplet codon based systems, in theory, could provide plenty of unassigned coding units but, to date, suffer from a high frame-shifting frequency. Thus, it will pay off to explore the possibilities of a sense codon reassignment for site-specific incorporation of ncaas. The Arg codon AGG is a good candidate for such reassignment because it is the rarest sense codon available in the genetic code of *E. coli* (see chapter 1.2.3) and *argW*, the gene encoding the Arg isoacceptor $tRNA^{Arg}_{CCU}$, is not essential in *E. coli*.[198] The isoacceptor $tRNA^{Arg}_{UCU}$ compensates for the loss of $tRNA^{Arg}_{CCU}$ since it can decode AGG via wobble base pairing.[199] However, the Gibbs free energy for melting the G-U base pairing is 4.1 kcal mol^{-1} while 6.3 kcal mol^{-1} are needed for melting of A-U.[200] Hence, the wobble base pair G-U is less stable than the Watson-Crick base pair A-U. On this basis, it might be possible to displace the $tRNA^{Arg}_{UCU}$ from reading AGG by re-introducing a tRNA which is capable of Watson-Crick base pairing with AGG and can be charged with an ncaa.

*Sc*PheRS(T415G):*sc*$tRNA^{Phe}_{CUA\_UG}$ was used to make a first attempt towards the reassignment of AGG because the pREP4_*sc*$tRNA^{Phe}_{CUA\_UG}$ contains only one tRNA copy which can easily be mutated to harbor a CCU instead of a CUA anticodon. Furthermore, a synthetase with high substrate tolerance can have certain advantages in proteome wide AGG reassignment attempts. Some genetic manipulations in the *E. coli* tRNA pool could lead to lethal phenotypes because AGG cannot be read anymore (see chapter 3.2 in the outlook section). The introduction of an exogenous aaRS:$tRNA_{CCU}$ pair could help to prevent this lethal phenotype. The sole possibility for AGG reading might not be enough, however, because in certain proteins the mutation of Arg to, for example, Trp may lead to a non-functional protein. Using *sc*PheRS(T415G), Trp, Phe and a large pool of ncaas can be used in the same expression experiment for substituting Arg in different cellular proteins. This might lead to a higher compensation rate of otherwise lethal mutations at the AGG positions.

Accordingly, the *sc*$tRNA^{Phe}_{CUA\_UG}$ anticodon was mutated to CCU by site-directed mutagenesis (see chapter 5.1.2.2). EGFP with the mutation N150AGG was used as target protein. In a first attempt, expressions were performed in LB medium without ncaa according to chapter 5.3.1. *E. coli* AFW and the *argW* deletion strain *E. coli* JD27920 were used in the presence and absence of *sc*$tRNA^{Phe}_{CCU\_UG}$. If a certain amount of codon reassignment takes place, position 150 of EGFP will be occupied by Trp and Phe in addition to Arg. **Figure 38** shows the resulting expression gel. Neither the presence of AGG at position 150 nor the presence of *sc*$tRNA^{Phe}_{CCU\_UG}$ affected the EGFP expression level in both strains.

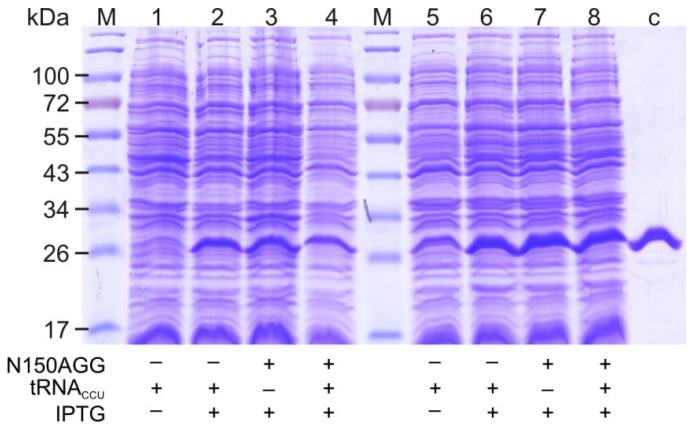


Figure 38. Reassignment of AGG at position 150 in EGFP with *sc*tRNA$^{Phe}_{CCU_UG}$ in LB medium. Fermentations were performed in *E. coli* AFW (lanes 1 – 4) and *E. coli* JD27920 (lanes 5 – 8). 0.15 OD_{600} of cell lysates per lane were applied to a 12% SDS gel. **M**: marker; **1** and **5**: non-induced samples; **2** and **6**: parent EGFP expression; **3** + **7**: expression of EGFP(N150R) (expression of EGFP(N150AGG) in absence of *sc*tRNA$^{Phe}_{CCU_UG}$); **4** and **8**: test for reassignment of AGG (expression of EGFP(N150AGG) in presence of *sc*tRNA$^{Phe}_{CCU_UG}$), and **c**: control (3 µg purified EGFP).

All preparations were purified by Ni-NTA chromatography and analyzed by ESI-MS. Unfortunately, no reassignment could be observed. Expressions of EGFP(N150AGG) in presence and absence of *sc*tRNA$^{Phe}_{CCU_UG}$ revealed equal masses of 27786 Da which correspond exactly to EGFP(N150R) ($M_{w,\,cal}$ = 27786 Da). However, from this experiment it is difficult to say if deletion of *argW* has an effect on AGG reassignment because the strain *E. coli* JD27920 did not efficiently repress target protein expression before induction (see **Figure 38**, lane 5). In contrast to *E. coli* AFW, the non-induced *E. coli* JD27920 was clearly green fluorescent after cell harvest (data not shown). An extensive literature search revealed that the *argW* deletion in *E. coli* JD27920 was achieved by insertion of a transposon bearing the Kan resistance marker. Thus, the pREP4 based plasmid harboring *sc*tRNA$^{Phe}_{CCU_UG}$ as well as *lacI* might have been lost during growth on Kan medium. For a detailed study on the effects of *argW* deletion, a marker-free knock-out should be made in *E. coli* AFW.

In order to assess whether AGG can be partially reassigned in minimal medium, the above described experiment was repeated with *E. coli* AFW in NMM according to chapter 5.3.2.2. 40 µM Phe and 15 µM Trp were used as limiting amino acid concentrations to grow cells up to an OD_{600} 0.6 – 0.7. Subsequently, 3 mM Trp and 15 µM Phe were added and target protein expression was induced. Trp was added in great excess to get *sc*tRNA$^{Phe}_{CCU_UG}$ loaded only with Trp and not with a mixture of Trp and Phe. Interestingly, in this setup AGG was reassigned to Trp at about 15% (**Figure 39C**) whereas in LB medium no reassignment was detected. In the following, it was analyzed whether the Arg/Trp ratio at position 150 of EGFP could be further triggered by a higher concentration of *sc*tRNA$^{Phe}_{CCU_UG}$ in the cell. Therefore, a 3 x *sc*tRNA$^{Phe}_{CCU_UG}$ array under the strong *proK* promoter and terminator — as it was used in pSup-BpaRS-6TRN — was designed and produced by gene synthesis.

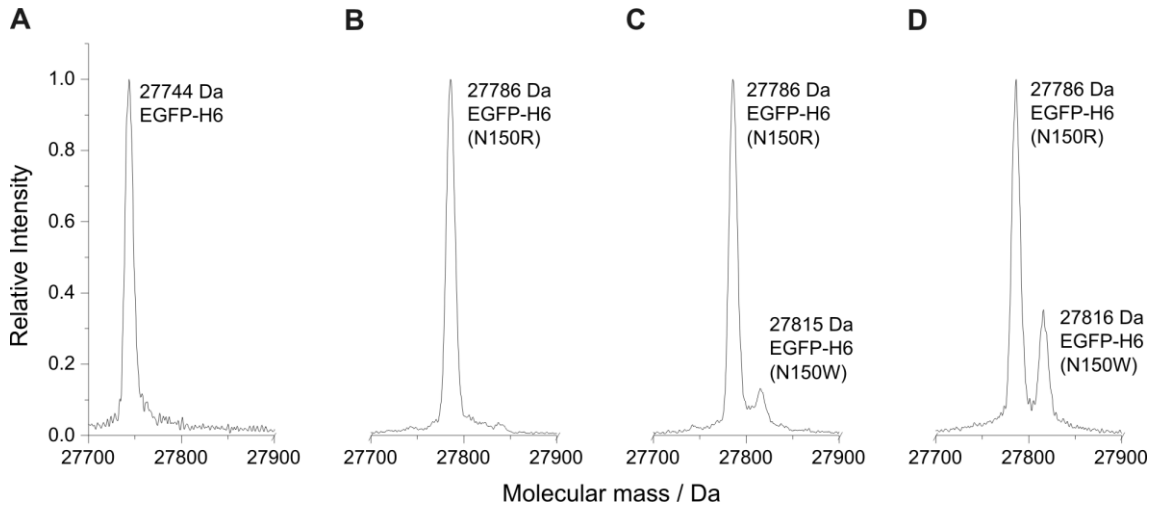


Figure 39. Mass spectral analysis of AGG codon reassignment. Masses of EGFP (**A**), EGFP(N150R) (**B**), EGFP(N150X)+*sc*tRNA$^{Phe}_{CCU_UG}$ (**C**), and EGFP(N150X)+3x*sc*tRNA$^{Phe}_{CCU_UG}$ (**D**) are presented. Calculated masses: EGFP = 27744 Da; EGFP(N150R) = 27786 Da; EGFP(N150W) = 27816 Da. The ratios between N150R and N150W were calculated by ESI-MS peak integration. Using one *sc*tRNA$^{Phe}_{CCU_UG}$, AGG was reassigned to Trp at around 15% (**C**) while three *sc*tRNA$^{Phe}_{CCU_UG}$ triggered 35% reassignment (**D**).

The synthetic construct was cloned into pREP4 by *in vitro* homologous recombination (see chapter 5.1.3.2) and the resulting construct was called pREP4_3x*sc*tRNA$^{Phe}_{CCU_UG}$. The described experiment was repeated with the new construct and the occupancy of position 150 by Trp could be shifted to approx. 35% (**Figure 39D**). Expectedly, a higher tRNA load was able to positively influence the AGG reassignment. Thus, a tripling of *sc*tRNA$^{Phe}_{CCU_UG}$ led to a 130% increase of Trp at position 150. Most probably, the six tRNAs as used in pSup-BpaRS-6TRN for amber stop codon suppression would further shift the Trp/Arg ratio towards Trp. It remains questionable why the partial reassignment worked in NMM but not in LB medium. NMM contains 50 mg L^{-1} ($\triangleq$ 0.3 mM) of Arg, whereas its concentration in LB medium was estimated to be around 2 – 3 mM although LB contains oligopeptides which have to be digested by *E. coli* in the first place.[188,189] Thus, the higher Arg concentration in the LB medium should have resulted in a higher overall concentration of Arg-tRNA$^{Arg}_{CCU}$ and Arg-tRNA$^{Arg}_{UCU}$ competing for the AGG that could also be read by *sc*tRNA$^{Phe}_{CCU_UG}$. Following this line of reasoning, the AGG decoding *sc*tRNA$^{Phe}_{CCU_UG}$ could possibly be further enhanced by lowering the Arg concentration in the medium.

These data depict an interesting starting point for the use of the AGG sense codon for incorporation of ncaas. 35% of reassignment was achieved by only introducing *sc*tRNA$^{Phe}_{CCU_UG}$ into an *E. coli* strain with an *argW* wildtype background. Already in this configuration, the newly introduced tRNA$_{CCU}$ generates a cellular 'ambiguous intermediate' status in which one codon has two different meanings (see chapter 1.1). By strain engineering (e.g. knock-out of *argW*) the reassignment rate will surely be increased because *sc*tRNA$^{Phe}_{CCU_UG}$ does not have to compete with the also Watson-Crick base pairing tRNA$^{Arg}_{CCU}$. Furthermore, is can be expected that the anticodon mutation reduced the aminoacylation of *sc*tRNA$^{Phe}_{CCU_UG}$ by *sc*PheRS(T415G). By optimizing the new *sc*tRNA$^{Phe}_{CCU_UG}$ for enhanced aminoacylation, higher reassignment rates will also be achieved. A sketch of

future work that could be performed based on these preliminary results will be presented in the following outlook section (see chapter 3.2). The final goal is to remove the old codon meaning completely and thereby attaining a free coding unit which can be used for incorporation of any amino acid of choice.

3 Conclusions and outlook

Noncanonical amino acids endow proteins with new reaction partners or give rise to novel functions and folds beyond those available to traditional biochemistry. Integrated into industrially relevant bio-production problems such developments are essential for the transition of the traditional chemical to a biochemical/bioengineering industry — a prerequisite to tackle the challenges of the 21st century.

In the present study, the toolkit of protein engineering with noncanonical amino acids was further expanded by simultaneous genetic code engineering and expansion. However, to implement such advanced methodologies in industrial processes, many efforts still have to be made. Matter of factly, currently available systems for code expansion are typically special rather than broadly applicable solutions. For example, the *mj*BpaRS:*mj*tRNA$^{Tyr}_{CUA}$ orthogonal pair is only capable of incorporating Bpa. In this way, every new amino acid to be incorporated needs novel engineering efforts at least at the level of the aaRS. Industry, however, needs stably working generalist tools.

Recent progresses in this direction include: (i) the generation of two mutually orthogonal pairs from one parent *mj*TyrRS:*mj*tRNA$^{Tyr}_{CUA}$ pair.[201] This development provides the possibility to combine different already available *mj*TyrRS based orthogonal pairs for simultaneous incorporation of various ncaas into a target protein; (ii) the discovery of generalist aaRS:tRNA$_{Sup}$ pairs which were impaired in incorporating their original substrate but capable of incorporating a variety of different ncaas with comparable efficiency.[202,203] In the present work, *sc*PheRS(T415G) and *mj*BpaRS were also evaluated for such promiscuity (see chapter 2.2). *Mj*BpaRS was found to be highly specific for Bpa and did not incorporate any of the other tested ncaas in significant manner. *Sc*PheRS(T415G), however, is such a generalist aaRS because it is capable of incorporating a great variety of ncaas. Nevertheless, due to its capacity to incorporate also the canonical amino acids Phe and Trp, it is not a candidate for industrial application until further optimized.

Even more importantly, a complete orthogonalization of the whole ncaa incorporation system in the cell would be highly desired to get towards an industrial application as current systems still cross-react significantly with the host cell by also suppressing stop codons in the host cell genome. To avoid this, codon emancipation from in the genome is needed.

3.1 Emancipation of the amber stop codon

As discussed extensively in the present study, the process of suppression is highly inefficient in *E. coli* because every suppressor tRNA has to compete with release factors for entry in the ribosomal A-site. In this context, it has been shown that the reduction of RF1 levels (e.g. by using temperature sensitive mutants) enhanced amber suppression.[204,205] A total removal of RF1 from the genome would be highly desirable because then the UAG codon could be used as an additional sense codon for ncaa coding. However, it was generally thought that a complete removal of RF1 is not possible because the deletion of its gene was

lethal (see chapter 1.2.4). The complete removal of all amber stop codons from the *E. coli* genome was the only scenario in which a knock-out was thought to be possible. Very recently, however, Sakamoto and coworkers successfully demonstrated that RF1 can indeed be removed when some additional genetic elements are provided in *cis* and/or *trans*.[206] In brief, by screening an single gene knock-out database they found out that in *E. coli* only seven essential genes exist which are terminated by an amber stop codon. Their heterologous coexpression — the termination codon was changed from UAG to UAA — together with the presence of an amber suppressor tRNA was enough to rescue the lethal RF1 knock-out. When the *mj*TyrRS derived IPheRS:$tRNA^{Tyr}_{CUA}$ pair was introduced as an amber suppressor system, culture growth was dependent on IPhe presence in the medium. Heterologous expression of glutathione S-transferase harboring six internal amber stop codons resulted in full-length protein with six IPhe incorporated at the specific amber stop codon site. In summary, Sakamoto and coworkers could totally reassign the UAG codon to a noncanonical amino acid by knock-out of RF1. This knock-out could be achieved because apparently only seven proteins in the whole *E. coli* genome are sensitive to C-terminal peptide tails produced by amber suppression. If the reported results turn out to be repeatable, the study of Sakamoto can be denoted as the biggest break-through in the field of genetic code expansion in the last 13 years (i.e. since Furter published his work on incorporation of 4-fluorophenylalanine with the orthogonal *sc*PheRS:*sc*$tRNA^{Phe}_{CUA}$ pair). Basically, the incorporation of an unlimited number of noncanonical amino acid residues in a target protein would be possible without additional truncation product.

In addition, the results would prove the ambiguous intermediate theory for sense codon reassignment in some species (see chapter 1.1). Thereby, the current state of genetic code expansion with coexistence of RF1 and an orthogonal aaRS:$tRNA_{CUA}$ pair represents the ambiguous status in which the UAG has both, the stop and the ncaa meaning. By gradual removal of the old meaning (in this case the exchange of UAG in the seven essential genes by UAA), the reassignment gets tolerable for *E. coli*. After successful knock-out of RF1, the ambiguity is deleted and the UAG totally switched to the ncaa meaning.

The results of the Sakamoto study make it also reasonable to think about a complete reassignment of a sense codon to an ncaa. In the present study, first attempts in this direction were already made. It was discussed that for a simultaneous genetic code expansion with various noncanonical amino acids, more free coding units than amber, ochre and the until now inefficient quadruplet codons need to be available (see chapter 2.5). Thus, more research in this direction is certainly worthwhile. In the following, potential future works are briefly sketched.

3.2 Emancipation of the AGG codon

The experiments described in chapter 2.5 gave a first indication that partial AGG reassignment is possible. In the next step, experiments should be performed for exploring the influence of the medium composition on the reassignment. For example, a reassignment in NMM with a calibrated Arg concentration could be performed. First, the experiment

described in chapter 2.5 could be repeated with *E. coli* AFW in NMM without Arg. *E. coli* AFW is not auxotrophic for Arg and can produce it on its own if Arg is not supplied with the medium. This could lead to a lower overall concentration of arginylated endogenic $tRNA^{Arg}_{CCU}$ and $tRNA^{Arg}_{UCU}$ and thus a higher incorporation of Trp in response to AGG. Furthermore, reassignment experiments could be repeated in a strain which is additionally Arg auxotrophic. Here, the reassignment can be tested directly in dependence of the Arg amount supplied in the medium.

In order to get the AGG codon completely free of Arg incorporation, *E. coli* will have to be genetically modified. *ArgW* and *argU* will have to be deleted from the genome to avoid recognition of AGG by $tRNA^{Arg}$. *ArgW* is only capable of reading AGG; *argU,* however, can read AGA by Watson-Crick base-pairing and AGG by wobbling (see chapter 2.5). Therefore, *argW* can be deleted from the *E. coli* genome without major effects on strain viability[198] because *argU* can compensate its deletion. However, an *argU* knock-out would be lethal. In a first step, the influence of an *argW* deletion should be investigated similar to the experiment in chapter 2.5 but in an *E. coli* strain with a different genetic background than JD27920. In the next step, different strategies for blocking of AGG recognition by *argU* have to be tested. For example, antisense RNA expression could be induced at a certain time point to knock-down $tRNA^{Arg}_{UCU}$.[207]

The basic AGG codon reassignment experiments in this study provides further insight that sense codons can principally be used to incorporate noncanonical amino acids. The s*c*PheRS(T415G):*sc*$tRNA^{Phe}_{CCU\_UG}$ pair, however, might not be the best choice for a genome-wide reassignment of the codon AGG since the exchange of Arg to Phe, Trp or their noncanonical analogs could have lethal effects in some essential genes. Such impacts can indeed never be excluded but the choice of an orthogonal pair which is capable of incorporating amino acids which are structurally more similar to Arg could minimize such drawbacks.

In addition, orthogonal pairs could be further optimized with respect to a better recognition of the tRNA by the aaRS and a better interaction of the aminoacyl-tRNA with elongation factor EF-Tu. These issues were already discussed for amber suppressor tRNAs in chapter 2.2.3. For optimization of aaRSs or tRNAs, efficient selection systems are needed. The setup of such a screening system is relatively easy for aaRS:$tRNA_{CUA}$ orthogonal pairs due to the fact that an amber suppression based system only produces full-length protein when the amber stop codon is suppressed. Thus, present selection systems contain antibiotic resistance genes with amber stop codons in their sequence (see chapter 1.3.3.1). Only if the enzyme is produced as full-length protein, it can provide antibiotic resistance and the higher the amount of full-length protein the higher the antibiotic concentration tolerated by the cell. Thereby, the fitter aaRS:$tRNA_{CUA}$ pairs can be selected from a library when a higher antibiotic pressure is applied. Similarly, FACS based on e.g. GFP(amber) can be applied for screening because fitter aaRS:$tRNA_{CUA}$ pairs will lead to a higher GFP fluorescence.[150] In addition, more complex FACS setups — e.g. based on the cell surface display of enzymes[208] — may even lead to better screening results because the dynamic range is much higher. However, such screening setups cannot be applied to sense codon based aaRS:tRNA

pairs since full-length protein is produced in any case. Here, screens have to be based on the suppression of lethal phenotypes through amino acids introduced by a sense codon based aaRS:tRNA pair. Alternatively, the orthogonal pair could be optimized as aaRS:$tRNA_{CUA}$ with amber stop codon based methods and the anticodon later switched to the respective sense codon. In the latter case, however, the orthogonal pair would most probably not be optimally evolved for reading the sense codon efficiently.

4 Materials

4.1 Chemicals

All standard chemicals were obtained from Sigma (Sternheim, Germany), Merck (Darmstadt, Germany) or Carl Roth GmbH (Karlsruhe, Germany) unless otherwise indicated.

4-iodophenylalanine (CAS 24250-85-9), 4-azidophenylalanine (CAS 33173-53-4), *O*-methyltyrosine (CAS 6230-11-1), 3-(3-benzothienyl)alanine (CAS 72120-71-9), 4-benzoylphenylalanine (CAS 104504-45-2) and 4*S*-fluoroproline (CAS 2438-57-5) were purchased from Bachem (Weil am Rhein, Germany). 6-chlorotryptophan (CAS 17808-21-8) and 6-bromotryptophan (CAS 33599-61-0) were from Biosynth. 3-(2-naphthyl)alanine (CAS 58438-03-2) was obtained from Wako Chemicals (Neuss, Germany), 4-aminophenylalanine (CAS 2410-24-4) and norleucine (CAS 327-57-1) from Sigma, and 3-(2-furyl)alanine (CAS 127682-08-0) from Parkway Scientific (New York, USA). 4-propargyloxyphenylalanine was synthesized by CS Dong according to reference [209] and 3-(1-azulenyl)alanine by J Musiol using the protocol given in [210]. L Merkel provided azidohomoalanine (CAS 120042-14-0) synthesized according to the protocol of Link.[211]

4.2 Media and supplements

LB medium

For *E. coli* growth under non limiting conditions, LB medium was used. Components were purchased from BD Biosciences (Sparks MD, USA).

10 g	Bacto Tryptone
5 g	Bacto Yeast extract
10 g	NaCl
	adjust to 1 L with H_2O

LB liquid medium was autoclaved for 20 min at 121 °C and 1.5 bars; for agar plates 1.5% agar was added before autoclaving. Antibiotics were sterilized by filtration and added to media, subsequently.

NMM

For supplementation-based incorporation of ncaas, *E. coli* cells were grown in New Minimal Medium (NMM).[92] The medium was prepared from autoclaved or sterile-filtered stock solutions of the following components in double distilled water (ddH_2O).

7.5 mM	$(NH_4)_2SO_4$	20 mM	Glucose	10 mg L^{-1}	Thiamine
8.5 mM	NaCl	1 mM	$MgSO_4$	10 mg L^{-1}	Biotin
22.5 mM	KH_2PO_4	1 mg L^{-1}	Ca^{2+}	10 mg L^{-1}	Trace elements
50 mM	K_2HPO_4	1 mg L^{-1}	Fe^{2+}	50 mg L^{-1}	Amino acids

Trace elements (Cu^{2+}, Zn^{2+}, Mn^{2+}, and MoO_4^{2-}) were supplied as 10 µg mL^{-1} stock solution. Non-limiting amino acids were added as combined 0.5 g L^1 stock solution in potassium phosphate buffer (100 mM, pH 7.2).

Yeast extract Peptone Dextrose medium (YPD medium)

10 g	Bacto Yeast extract
20 g	Bacto Peptone
20 g	Glucose
	adjust to 1 L with H_2O

YPD liquid medium was autoclaved for 20 min at 121 °C and 1.5 bars; for agar plates 1.5% agar was added before autoclaving.

Synthetic Complete medium (SC medium)

6.7 g	Difco Yeast Nitrogen Base w/o amino acids
2 g	Drop-put mix
20 g	Glucose
	adjust to 1 L with H_2O

Components of the drop-out mix:

0.5 g	Adenine	2.0 g	Glycine	2.0 g	Phenylalanine
2.0 g	Alanine	2.0 g	Histidine	2.0 g	Proline
2.0 g	Arginine	2.0 g	Inositol	2.0 g	Serine
2.0 g	Asparagine	2.0 g	Isoleucine	2.0 g	Threonine
2.0 g	Aspartic acid	10.0 g	Leucine	2.0 g	Tryptophan
2.0 g	Cysteine	2.0 g	Lysine	2.0 g	Tyrosine
2.0 g	Glutamine	2.0 g	Methionine	2.0 g	Uracil
2.0 g	Glutamic acid	2.0 g	*p*-aminobenzoic acid	2.0 g	Valine

The respective components used for selection were omitted. In this study, uracil was used as selection marker. SC liquid medium was autoclaved for 20 min at 121 °C and 1.5 bar; for agar plates 1.5% agar was added before autoclaving.

Supplements

Supplement	Stock concentration	End concentration
Ampicillin	100 mg mL^{-1} in ddH_2O	100 µg mL^{-1}
Kanamycin	50 mg mL^{-1} in ddH_2O	50 µg mL^{-1}
Chloramphenicol	34 mg mL^{-1} in ethanol	34 µg mL^{-1}
IPTG	1 M in ddH_2O	1 mM

4.3 Strains

Genotype abbreviations were used according to reference [212].

Strain name	Genotype	Source / Reference / Strain development
E. coli AFW	Hfr(PO2A) *phe*S13 *rel*-1 *ton*A22 *thi* $T2^R$, *phe*A18::Tn10, *trp*B114::Tn10	K12 → Cavalli Hfr → K10 Hfr[213]/KB → PFP-10[214] → K10-F6[56] → *E. coli* AFW[174]
E. coli B834(DE3)	F^- *dcm ompT hsdS*(r_B^- m_B^-) gal *met* λ(DE3 [*lac*I *lacUV5*-T7 gene 1 *ind*1 *sam*7 *nin*5])	B → B834[215] → B834(DE3)[216,217]
E. coli ME5355	F^- *pro*A2 *arg*E3 *phe*AI *tyr*A4 *trp*-401 *aro*B351 *thi*	National BioResource Project *E. coli* at National Institute of Genetics (Japan) Equivalent to the *E. coli* K12 derivative AN92[218]
E. coli JD27920	F^- λ^- *rph*-1 INV(*rrn*D *rrn*E) *lacIq* *lac*ZΔM15 *gal*K2 *gal*K22 *argW*::mini-Tn10(*lac*Zα-kan)	National BioResource Project E.coli at National Institute of Genetics (Japan) W3110 type A → KP7600 → JD27920[219]
E. coli CAG18491	F^- λ^- *ilv*G- *rfb*-50 *rph*-1 *metEo 3079:Tn10*	K12 → MG1655[50,213] → CAG18491[220]
E. coli CAG18515	F^- λ^- *ilv*G- *rfb*-50 *rph*-1 *proA3096:Tn10-kan*	K12 → MG1655[50,213] → CAG18515[220]
E. coli DH5α	F^- λ^- *end*A1 *gln*V44 *thi*-1 *rec*A1 *rel*A1 *gyr*A96 *deo*R *nup*G Φ80d*lacZ*ΔM15 Δ(*lacZYA*-*argF*)U169 *hsd*R17(r_K^- m_K^+)	K12 → 1100[221] → MM294[222] → DH1[223] → DH5[224] → DH5α[224]
E. coli Omnimax	F′ {*pro*AB+ *lacIq* *lac*ZΔM15 Tn10-TetR Δ(*ccd*AB)}, *mcrA* Δ(*mrr*-*hsd*RMS-*mcr*BC), Φ 80(*lac*Z) ΔM15 Δ(*lac*ZYA-*arg*F)U169 *end*A1 *rec*A1 *sup*E44 *thi*-1 *gyr*A96 *rel*A1 *ton*A *pan*D	Invitrogen, Darmstadt, Germany
S. cerevisiae INV*Sc*1	*MATa his*3D1 *leu*2 *trp*1-289 *ura*3-52	Invitrogen, Darmstadt, Germany

4.4 Synthetic DNA constructs

The synthetic DNA construct 3x*sc*tRNA$^{Phe}_{CCU_UG}$ was ordered from Mr. Gene (Regensburg, Germany). It was an exact copy of 3x*mj*tRNA$^{Tyr}_{CUA}$ but with the tRNAs exchanged to sctRNA$^{Phe}_{CCU_UG}$. 3x*sc*tRNA$^{Phe}_{CCU_UG}$ was the source DNA for the insert in pREP4_3x*sc*tRNA$^{Phe}_{CCU_UG}$.

4.5 Primers

All primers were obtained from Metabion (Martinsried, Germany).

Number / MH#	Name	Sequence (5' → 3')
Sequencing		
1	T5 Promoter	CCCGAAAAGTGCCACCTG
2	T5 Terminator	GTTCTGAGGTCATTACTGG
3	T7 Promoter	TAATACGACTCACTATAGGG
4	T7 Terminator	GCTAGTTATTGCTCAGCGG
5	pQE16_yPheRS for	TCCAAGCTAGCTTGGCG
6	pQE16_yPheRS rev	CCGGATGAGCATTCATCA
7	pQE16_RS f2	TGCTCAAATTTTGAATGAAGGT
8	pQE16_RS r2	CTTTAAATACAACGTCACCGGT
9	pQE16_RS f3	CGATCTCGGCTTTGGTC
10	pQE16_RS r3	ACTTTCTACATCAAGGACCCAC
11	pREP4_tRNA Phe for	TCAGCTACTGACGGGGTG
12	pREP4_tRNA Phe rev	GACGCACCGGTGCA
25	mjTyrRSforward1	ATGGACGAATTTGAAATGATAAAG
26	mjTyrRSreverse1	CTAAATGTATTTTACCACTTGGTTCA
27	mjTyrRSforward2	CTTTTACCAAAAAAGGTTGTTTGTAT
28	mjTyrRSforward3	GATTTTAGAGCCAATTAGAAAGAGATTA
29	p15A ori (pRARE)	TTTACCGGTGTCATTCCG
73	pQE16_RS f4	TTTATCGAAACCAATCCTGC
74	pQE16_RS f5	GATCGAGTTTTATAAATCTCCAGAAC
93	2µ Ori_rev	TCGAGTTTAGATGCAAGTTCAAG
96	MCS Pro24_fwd	CTAGCAGGAGGAATTCACCAT
97	MCS Pro24_rev	CATGCCTGCAGGTCG
102	pPro24_rev2	CCTCTTCTATCGGCGTTGT
118	RSF_rev	CCTTGGTTCAAAGAGTTGGTAG
121	PPro24_rev3	CGGGCATCAGCCTGT
148	2µ Ori_fwd	CGTACAGTAGACGGAGTATCTAGTATAGTC
151	pPro24_insert_fwd	CTGAAACGTTAACTGAAACGC

Number / MH#	Name	Sequence (5' → 3')
Sequencing		
154	pAra_seq	CGCTTTTTATCGCAACTCTC
163	2µ Ori_fwd2	TGTATTATAAGTAAATGCATGTATACTAAACTCAC
Ligation-based cloning		
37	pQE80L-EGFP-H6 forward	TATATCTATATCGAATTCATTAAAGAGGAGAAATTAAGCATGGTGAGCAAGGGCGAG
38	90803 pQE80L-EGFP-H6 (reverse)	CGTGCCGGCGGCTGCAGTTAGTGATGGTGATGGTGATGCTT
49	H6-PheRS forward	TATATCTATATCGGATCCATGTCTGACTTCCAATTAGAAATTCTAA
50	H6-PheRS reverse	CGTGCCGGCTGCTGCAGTTATCATAGGAAGACTTCGGCATTAACC
60	TthLip_fwd	TATATCTATATCGAATTCATTAAAGAGGAGAAATTAAGCATGCAAAAGGCTGTTGAAATTACATATAAC
61	TthLip-H6_rev	CGTGCCGGCGGCTGCAGTTATCAGTGATGGTGATGGTGATGGGATCCTCCCTTTAACAATTCCTTTTTGAAAAACT
136	pPro-GFPuvH6_fwd	AAGGAAACCGGAATTCACCATGACTAGCAAAGGAGAAGAACTTTTCA
137	pPro-GFPuvH6_rev	TTCCTTTTTTCTGCAGTCATTAGTGATGGTGATGGTGATG
Homologous recombination in *S. cerevisiae*		
75	p15A_fwd	GCCGGGCGTTTTTTATTGGTGAGAATCCAAGCGGGCCCGGATATATTCCGCTTCCTCG
76	p15A_rev	TTATTACCCATTGAAAAAGGAAGAGTCCCGGGGCGCGCAACAACTTATATCGTATGGGGCT
77	Cm_fwd	GAAGCGGTATTCGCAATGGCGCGCGCGGCCGCCCTAGGGATATCTGGCGAAAATGAGACG
78	Cm_rev	CGGAATATATCCGGGCCCGCTTGGATTCTCACCAATAAAA
79	Yeast shuttle_fwd	ATATAAGTTGTTGCGCGCCCCGGGACTCTTCCTTTTTCAATGGGTAA
80	Yeast shuttle_rev	TCGCCAGATATCCCTAGGGCGGCCGCGCGCGCCATTGCGAATACCGCTTC
81	*glnS'*_fwd	GAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAGATATCCCTAGGCATCAATCATCCCCATAATCC
82	*glnS'*_rev	CAGGCCGCTTAGTTAGCCGGTACCAAAGCAGAAAAAACGCCG
83	*proK*_fwd	GCGTTCAGCGGCGTTTTTTCTGCTTTGGTACCGGCTAACTAAGCGGCCT
84	*proK*_rev	AGAGATACTTTTGAGCAATGTTTGTGGAAGCGGTATTCGCAATGGCGCGCATGCAAAAAAGCCTGCTC
87	RSF_fwd	CGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCGGGCCCAGCGCTCTTCCGCTTC
88	RSF_rev	TAATTATATCAGTTATTACCCATTGAAAAAGGAAGAGTCCCGGGGCGCGCGTAACGGAATAGCTGTTCGTTG
89	*lpp*_fwd	AAACCCCGCCCCTGACAGGGCGGGGTTTTTTTTGGTACCGCTTCTTTGAGCGAACGAT
90	*rrnC*_rev	AGAGATACTTTTGAGCAATGTTTGTGGAAGCGGTATTCGCAATGGCGCGCAAAAAAAATCCTTAGCTTTCGC

Number / MH#	Name	Sequence (5' → 3')
Homologous recombination in *S. cerevisiae*		
91	*tac*_fwd	GAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAGATATCCCTAGG GTTGACAATTAATCATCGGCT
92	yPheRS+Term_rev	CTCAAAGAAGCGGTACCAAAAAAAACCCCGCCCTGTCAGGGGCGGG GTTTTTTTTTTTATAGGAAGACTTCGGCATTAA
127	F_ORI_fwd	CGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCGGGC CCGGCGTCCAGTTTTGGCA
128	F_ORI_rev	TAATTATATCAGTTATTACCCATTGAAAAAGGAAGAGTCCCGGGGCGCG CACCGTGG
129	prpB_fwd	GAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAGATATCCCTAGG CATGAGCCCGAAGTGGC
130	prpB_STREP_BpaRS_rev	TCGCTGATAATTTCAGATGTGTTTCTCTTTATCATTTCAAATTCGTCCATG GCGCCTTTTTCGAACTGCGGGTGGCTCCAAGCGCTCATGGTGAATTCC TCCTGCTAG
131	prpB_STREP_yPheRS_rev	TCATCCAATTCATCTAGTTTCTTTAGAATTTCTAATTGGAAGTCAGACAT GGCGCCTTTTTCGAACTGCGGGTGGCTCCAAGCGCTCATGGTGAATTC CTCCTGCTAG
140	yPheRS-*glnS'*Term_fwd	TTCCATTCGCTGCATCCTTTGTAGAGGTTAATGCCGAAGTCTTCCTATAA CTGCAGTTTCAAACGCTAAATTG
141	yPheRS-*glnS'*Term_rev	TTTTTGATGGGGCGCCACTTATTTTTGATCGTTCGCTCAAAGAAGCGGT ACCAAAGCAGAAAAAACGC
142	SphI-3xtRNATyrCUA_fwd	AAGGAAAACATGCATGCGGCTAACTAAGCGGCCTGC
143	3xtRNATyrCUA-SphI_rev	AAGGAAACATGCATGCAAAAAAGCCTGCT
157	*araC*_fwd	GAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAGATATCCCTAGG GCATAATGTGCCTGTCAAATG
158	*araC*_BpaRS_rev	TCATTTCAAATTCGTCCATGGCGCCTTTTTCGAACTGCGGGTGGCTCCA GCTAGCCATGGTTAATTCCTCCTGTTAGC
In-Fusion PCR cloning		
108	GFPuv_infu_fwd	GAGGAGAAATTAACTATGACTAGCAAAGGAGAAGAACTTTTCACTG
109	GFPuv_infu_rev	CTCAGCTAATTAAGCTCATTAGTGATGGTGATGGTGATGTTTGTATAGTT CATCCATGCCATGT
110	pQE16_PheRS_infu_fwd	GCTTAATTAGCTGAGCTTGGACTCC
111	pQE16_PheRS_infu_rev	AGTTAATTTCTCCTCTTTAATGAATTCTGTGT
123	EGFP-ECFP_infu_fwd	GAGGAGAAATTAACTATGGTGAGCAAGGGCGA
124	EGFP-ECFP-H6_infu_rev	CTCAGCTAATTAAGCTTAGTGATGGTGATGGTGATGCTT
144	H6-GFPuv_InFu_fwd	GAGGAGAAATTAACTATGAGAGGATCG
145	H6-GFPuv_InFu_rev	CTCAGCTAATTAAGCTTGGCTGC
149	GFPuv_infu_pBAD_fwd	GAGGAATTAACCATGACTAGCAAAGGAGAAGAACTTTTC
150	GFPuv_infu_pBAD_rev	CAAAACAGCCAAGCTCATTAGTGATGGTGATGGT
155	pREP43xtRNA_fwd	TTTTTTGCATGCGCGCAAGCTTGGCACTGGCC
156	pREP43xtRNA_rev	TTAGTTAGCCGGTACCCGCCAGGGTGGTTTTT

Number / MH#	Name	Sequence (5’ → 3’)
Site-directed mutagenesis		
39	LipY9AmberF	AAATTACATAGAACGGCAAAACTTTAAG
40	LipY9AmberR	TTTTGCCGTTCTATGTAATTTCAACAGC
41	LipF17AmberF	TTTCACGGTTAGACAGGCAATAAAGTAG
42	LipF17AmberR	ATTGCCTGTCTAACCGTGAAACATTATT
43	LipW136AmberF	TTGGTGCTATAGGCTCCAGCTTTTAATA
44	LipW136AmberR	AGCTGGAGCCTATAGCACCAACGCCTTT
45	LipH233AmberF	AATGCAGACTAGACTTTTAAGAGTTTAG
46	LipH233AmberR	CTTAAAAGTCTAGTCTGCATTTTCGATT
47	LipW240AmberF	AGTTTAGAATAGGAGAAAAAGGCGATTG
48	LipW240AmberR	CTTTTTCTCCTATTCTAAACTCTTAAAA
53	ECFPW79AmberF	CACCCTGACCTAGGGCGTGCAGTGCTTCAG
54	ECFPW79AmberR	ACTGCACGCCCTAGGTCAGGGTGGTCACGA
57	TthLip H233W forward	AATGCAGACTGGACTTTTAAGAGTTTAG
58	TthLip H233W reverse	CTTAAAAGTCCAGTCTGCATTTTCGATT
65	LipD221Amber_fwd	GTTTATGGGTAGAACGCTACAAGAGTGA
66	LipD221Amber_rev	TGTAGCGTTCTACCCATAAACCTCTTTT
67	b*_W45Amber_fwd	CTGACCGGATAGGTGGAGTACCCGCTCG
68	b*_W45Amber_rev	GTACTCCACCTATCCGGTCAGAGCATCC
69	TthLip-H6_S261C_fwd	AAGGGAGGATGCCATCACCATCACCATC
70	TthLip-H6_S261C_rev	ATGGTGATGGCATCCTCCCTTTAACAAT
94	b*-Y30Amber_fwd	CTTGCGGAATAGTACGGTGAAAACCTGG
95	b*-Y30Amber_rev	TTCACCGTACTATTCCGCAAGGGCAAGC
104	GFPuv_N149Amber_fwd	AACTCACACTAGGTATACATCACGGCAG
105	GFPuv_N149Amber_rev	GATGTATACCTAGTGTGAGTTATAGTTG
106	tRNAPhe_CUA-CCU_fwd	CGCCATACTCCTAATGTGGAGGTCCTGT
107	tRNAPhe_CUA-CCU_rev	CTCCACATTAGGAGTATGGCGCTCTCCC
112	EGFP_N149Amber_fwd	AACAGCCACTAGGTCTATATCATGGCCG
113	EGFP_N149Amber_rev	GATATAGACCTAGTGGCTGTTGTAGTTG
119	EGFP N149AGG_fwd	AACAGCCACAGGGTCTATATCATGGCCG
120	EGFP N149AGG_rev	GATATAGACCCTGTGGCTGTTGTAGTTG
134	prpB-*Pau*I_fwd	AACTGGCGCGTGAGTCGGCGAAAACTCC
135	prpB-*Pau*I_rev	TCGCCGACTCACGCGCCAGTTCCGGCAG
146	2µ-*Apa*I_fwd	GGTGTGGTGGGTCCAGGTATTGTTAGCG
147	2µ-*Apa*I_rev	AATACCTGGACCCACCACACCGTGTGCA

4.6 PCR products for pMEc vector assembly

#	PCR product name	Description	PCR product length / bp	Primers used for amplification
I	2µ	• 2µ origin of replication for plasmid maintenance in *S. cerevisiae* • URA3 marker for selection in yeast • amplified from pYES2	2656	MH#79, MH#80
II	p15a	• low copy origin of replication for *E. coli* • amplified from pProLar.A	873	MH#75, MH#76
III	RSF	• low copy origin of replication for *E. coli* • amplified from pRSFDuet	858	MH#87, MH#88
IV	F1(mod)	• IPTG inducible *E. coli* origin of replication • amplified from pSCANS-#4	3034	MH#127, MH#128
V	Cm	• Chloramphenicol resistance gene • amplified from pProTet.E	963	MH#77, MH#78
VI	*glnS'*-BpaRS	• BpaRS under the control of the constitutive *glnS'* promoter and glnS terminator • amplified from pSup-BpaRS-6TRN	1287	MH#81, MH#82
VII	*prpR*-Strep-BpaRS	• propionate inducible P_{prpB} promoter with regulator PrpR • Overhang to BpaRS which fuses a StrepII-tag to the N-terminus • amplified from pPro24(-*Pau*I)	2120	MH#129, MH#130
VIII	*tac*-*sc*PheRS	• *sc*PheRS under the control of the constitutive *tac* promoter and a synthetic terminator (BBa_B1006) • amplified from pQE16am_T415G	3506	MH#91, MH#92
IX	*prpR*-Strep-*sc*PheRS	• propionate inducible P_{prpB} promoter with regulator PrpR • Overhang to *sc*PheRS which fuses a StrepII-tag to the N-terminus • amplified from pPro24(-*Pau*I)	2120	MH#129, MH#131
X	*araC*	• arabinose inducible promoter P_{BAD} with regulator AraC • Overhang to StrepII-tag • amplified from pBAD His C	1421	MH#157, MH#158
XI	glnS-term	• glnS terminator with overhang to C-terminus of *sc*PheRS • amplified from pSup-BpaRS-6TRN	275	MH#140, MH#141
XII	*lpp*-$tRNA^{Phe}_{CUA}$	• $tRNA^{Phe}_{CUA\_UG}$ with *lpp* promoter and *rrnC* terminator • amplified from pREP4_ytRNA$^{Phe}_{CUA_UG}$	295	MH#89, MH#90
XIII	*proK*-3x$tRNA^{Tyr}_{CUA}$	• 3 x $tRNA^{Tyr}_{CUA}$ array with *proK* promoter and terminator • amplified from pSup-BpaRS-6TRN	539	MH#83, MH#84

4.7 Plasmids

Plasmid name	Size / bp	Resistance marker	Ori	Source
pQE80L-H6-*sc*PheRS(T415G)	8040	Amp^R	ColE1	Assembled in this study REs: *Bam*HI, *Pst*I Primers: MH#49, MH#50
pET28a-H6-BpaRS	6271	Kan^R	ColE1	Assembled by S Nehring
pREP4	4042	Kan^R	p15A	From Qiagen, Hilden, Germany; source for pREP4 backbone
pREP4_ytRNA$^{Phe}_{CUA_UG}$	4373	Kan^R	p15A	From DA Tirrell[174]
pREP4_*sc*tRNA$^{Phe}_{CCU_UG}$	4373	Kan^R	p15A	Assembled in this study Primers: MH#106, MH#107
pREP4_3x*sc*tRNA$^{Phe}_{CCU_UG}$	4631	Kan^R	p15A	Assembled in this study Primers: MH#155, MH#156 + MH#83, MH#84
pSup-BpaRS-6TRN	4380	Cm^R	p15A	From PG Schultz
pQE80L	4751	Amp^R	ColE1	From Qiagen, Hilden, Germany; source for pQE80L backbone
pETBlue-1/Lip1	4256	Amp^R	ColE1	From G Antranikian; Source for TTL
pQE80L-TTL-H6	5494	Amp^R	ColE1	Assembled in this study REs: *Eco*RI, *Pst*I Primers: MH#60, MH#61
pQE80L-TTL-H6(Y9Amber)	5494	Amp^R	ColE1	Assembled in this study Primers: MH#39, MH#40
pQE80L-TTL-H6(F37Amber)	5494	Amp^R	ColE1	Assembled in this study Primers: MH#41, MH#42
pQE80L-TTL-H6(W136Amber)	5494	Amp^R	ColE1	Assembled in this study Primers: MH#43, MH#44
pQE80L-TTL-H6(D221Amber)	5494	Amp^R	ColE1	Assembled in this study Primers: MH#65, MH#66
pQE80L-TTL-H6(H233Amber)	5494	Amp^R	ColE1	Assembled in this study Primers: MH#45, MH#46
pQE80L-TTL-H6(W240Amber)	5494	Amp^R	ColE1	Assembled in this study Primers: MH#47, MH#48
pQE80L-H6-EGFP	5435	Amp^R	ColE1	From P Birle and T Krywcun; Source for EGFP
pQE80L-H6-ECFP	5438	Amp^R	ColE1	From P Birle and T Krywcun; Source for ECFP
pQE80L-H6-GFPuv	5437	Amp^R	ColE1	From P Birle and T Krywcun; Source for GFPuv

Plasmid name	Size / bp	Resistance marker	Ori	Source
pQE80L-EGFP-H6	5415	Amp^R	ColE1	Assembled in this study Primers: MH#110, MH#111 + MH#123, MH#124
pQE80L-EGFP-H6(N150Amber)	5415	Amp^R	ColE1	Assembled in this study Primers: MH#112, MH#113
pQE80L-EGFP-H6(N150AGG)	5415	Amp^R	ColE1	Assembled in this study Primers: MH#119, MH#120
pQE80L-ECFP-H6	5415	Amp^R	ColE1	Assembled in this study Primers: MH#110, MH#111 + MH#123, MH#124
pQE80L-ECFP-H6(W67Amber)	5415	Amp^R	ColE1	Assembled in this study Primers: MH#53, MH#54
pQE80L-GFPuv-H6	5418	Amp^R	ColE1	Assembled in this study Primers: MH#110, MH#111 + MH#108, MH#019
pQE80L-GFPuv-H6(N150Amber)	5418	Amp^R	ColE1	Assembled in this study Primers: MH#104, MH#105
pQE16am_T415G	7810	Amp^R	ColE1	From DA Tirrell[174] ; Source for pQE16_*sc*PheRS(T415G)
pQE16_*sc*PheRS(T415G) -EGFP-H6	7939	Amp^R	ColE1	Assembled in this study Primers: MH#110, MH#111 + MH#123, MH#124
pQE16_*sc*PheRS(T415G) -EGFP-H6(N150Amber)	7939	Amp^R	ColE1	Assembled in this study Primers: MH#110, MH#111 + MH#123, MH#124
pQE16_*sc*PheRS(T415G) -EGFP-H6(N150AGG)	7939	Amp^R	ColE1	Assembled in this study Primers: MH#110, MH#111 + MH#123, MH#124
pQE16_*sc*PheRS(T415G) -ECFP-H6	7939	Amp^R	ColE1	Assembled in this study Primers: MH#110, MH#111 + MH#123, MH#124
pQE16_*sc*PheRS(T415G) -ECFP-H6(W67Amber)	7939	Amp^R	ColE1	Assembled in this study Primers: MH#110, MH#111 + MH#123, MH#124
pQE16_*sc*PheRS(T415G) -GFPuv-H6	7942	Amp^R	ColE1	Assembled in this study Primers: MH#110, MH#111 + MH#108, MH#109
pQE16_*sc*PheRS(T415G) -GFPuv-H6(N150Amber)	7942	Amp^R	ColE1	Assembled in this study Primers: MH#110, MH#111 + MH#108, MH#109
pQE80L-ψ-b*	5111	Amp^R	ColE1	From P Birle and T Krywcun
pQE80L-ψ-b*(Y30Amber)	5111	Amp^R	ColE1	Assembled in this study Primers: MH#94, MH#95
pQE80L-ψ-b*(W45Amber)	5111	Amp^R	ColE1	Assembled in this study Primers: MH#67, MH#68
pYES2	5856	Amp^R/URA3	ColE1/2µ	From Invitrogen, Darmstadt, Germany

Plasmid name	Size / bp	Resistance marker	Ori	Source
pProLar.A	2631	Kan^R	p15A	From Clontech-Takara Bio Europe, Saint-Germain-en-Laye, France
pRSFDuet	3829	Kan^R	RSF	From Novagen (Merck Chemicals), Nottingham, UK
pSCANS-#4	4428	Kan^R	F1	From J Dunn
pProTet.E	2151	Cm	ColE1	From Clontech-Takara Bio Europe, Saint-Germain-en-Laye, France
pPro24	5139	Amp^R	ColE1	From JD Keasling[186]
pPro24(-*Pau*I)	5139	Amp^R	ColE1	Assembled in this study
pBAD/His C	4102	Amp^R	ColE1	From Invitrogen, Darmstadt, Germany
pMEc1µC		Cm^R	p15A	Assembled in this study with PCR products **I**, **II** and **V**
pMEc1µC1.1	6469	Cm^R	p15A	Assembled in this study with pMEc1µC1.3 and PCR of 3 x $tRNA^{Tyr}_{CUA}$ with MH#142 and MH#143; RE: *Sph*I
MEc1µC1.3	6010	Cm^R	p15A	Assembled in this study with pMEc1µC(*Not*I) + PCR products **VI** and **XIII**
pMEc1µC2.4	7982	Cm^R	p15A	Assembled in this study with pMEc1µC(*Not*I) + PCR products **VIII** and **XII**
pMEc1µC3.4	9920	Cm^R	p15A	Assembled in this study with pMEc1µC2.4(*Bam*HI) + PCR product **VIII**
pMEc1µC8.4	9346	Cm^R	p15A	Assembled in this study with pMEc1µC3.4(*Avr*II) + PCR product **X**
pMEc1µ4.3	7911	Cm^R	p15A	Assembled in this study with pMEc1µC1.3(*Avr*II) + PCR product **VII**
pMEc1µ6.3	7243	Cm^R	p15A	Assembled in this study with pMEc1µC4.3(*Avr*II, *Nhe*I) + PCR product **X**
pMEc2µ2.4	7946	Cm^R	RSF	Assembled in this study with PCR products **I**, **II**, **III**, **VIII** and **XII**
pMEc4µ2.4	10122	Cm^R	F1	Assembled in this study with pMEc1µC2.4(*Sac*II) + PCR product **IV**
pMEc1µC5.4	8119	Cm^R	p15A	Assembled in this study with pMEc1µC2.4(*Kpn*I) + PCR product **XI**

5 Methods

5.1 Molecular biological methods

5.1.1 Isolation of plasmid DNA

Small and medium-scale DNA plasmid preparation from *E. coli* was performed with the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) and Qiagen Plasmid Midi Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. For plasmid preparation from yeast, the QIAprep Spin Miniprep Kit was used with the following modifications: After cell resuspension in buffer P1 50 - 100 µL of acid-washed glass beads (Sigma G-8772) were added and the suspension was vortexed for 5 min. The beads were allowed to settle and the supernatant was transferred to a fresh 1.5 mL microfuge tube. 250 µL buffer P2 was added and the tube was inverted 4 – 6 times to mix. After incubation for 5 min at RT, it was proceeded according to the manufacturer's protocol.

5.1.2 Polymerase chain reaction

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

5.1.2.1 Standard PCR

Primers were designed with the program Oligo Analyzer 1.0.2 (Teemu Kuulasmaa, Kuopio, Finland) to have a sequence-dependent melting temperature (T_M) of 60 – 64 °C. T_M is calculated by the program using the nearest neighbor thermodynamic theory. If not required differently by a special polymerase, the annealing temperature was 4 °C below the T_M. DNA fragments < 1 kb were amplified using Pfu DNA polymerase (Fermentas, St. Leon-Rot, Germany). For longer fragments, PfuUltra II (Agilent, Böblingen, Germany) or Phusion High-Fidelity DNA Polymerase (New England Biolabs, Frankfurt am Main, Germany) were applied. The used primers for each construct are listed in chapter 4.7, the respective primer sequences are listed in chapter 4.5.

A typical reaction mix consisted of:

Template DNA	0.2 – 10 ng μL^{-1}
Forward primer	0.5 pmol μL^{-1}
Reverse primer	0.5 pmol μL^{-1}
dNTPs	0.2 – 0.25 mM each
10 x polymerase buffer	1 x
Polymerase	0.02 – 0.1 U μL^{-1}

Standard PCR program for Pfu and PfuUltra II polymerase:

Initial denaturation	95 °C	5 min	
Denaturation	95 °C	30 s	
Annealing	T_M – 4 °C	1 min	30 x
Extension	72 °C	1 min kb^{-1}	
End elongation	72 °C	10 min	

General PCR program for PCRs with Phusion High-Fidelity DNA Polymerase:

Initial denaturation	98 °C	30 s	
Denaturation	98 °C	10 s	
Annealing	T_M + 3 °C	30 s	30 x
Extension	72 °C	15 s kb^{-1}	
End elongation	72 °C	10 min	

5.1.2.2 Site-directed mutagenesis

All mutations were introduced using the QuikChange Site-Directed Mutagenesis procedure (Agilent, Böblingen, Germany). PfuUltra II (Agilent, Böblingen, Germany) or Phusion High-Fidelity DNA Polymerase (New England Biolabs, Frankfurt am Main, Germany) were used in the PCR reactions. The mutagenesis primers for each construct were designed according to the QuikChange protocol and are indicated in chapter 4.7 with respective primer sequences are listed in chapter 4.5. PCR conditions were chosen according to the instructions of the DNA polymerases.

5.1.3 Cloning

All enzymes for cloning were purchased from Fermentas (St. Leon-Rot, Germany) and New England Biolabs (Frankfurt am Main, Germany).

5.1.3.1 Ligation-based cloning

For ligation-based cloning, PCR products (see chapter 5.1.2.1) were purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). The purified PCR product and target vector were digested with appropriate restriction enzymes (RE). 20 U of each RE was used for digestion of 1 µg of DNA. The reactions were performed consecutively for 1 – 2 h under optimal reaction conditions for each enzyme. Digested DNA was agarose gel purified by the JETSORB Gel Extraction Kit (Genomed, Löhne, Germany) or the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). For ligation, 5 U of T4 DNA ligase were used to ligate 50 ng of vector DNA with the insert DNA in fivefold molar excess. The reaction was carried out for 15 min to 1 h at RT or overnight at 16 °C. Subsequently, the reaction mix was used to transform electro-competent *E. coli* DH5α or *E. coli* Omnimax (see chapter 5.2.2.1).

5.1.3.2 *In vitro* homologous recombination by In-Fusion PCR cloning

All constructs containing EGFP-H6, ECFP-H6 and GFPuv-H6 were assembled by *in vitro* homologous recombination using the In-Fusion Advantage PCR Cloning Kit (Takara Bio Europe/Clontech, Saint-Germain-en-Laye, France) according to the manufacturer's instructions. Genes and vector backbones were PCR amplified (see chapter 5.1.2.1) with the respective primers as outlined in chapter 4.7.

5.1.3.3 *In vivo* homologous recombination in *S. cerevisiae*

S. cerevisiae is extremely efficient in homologous recombination. Double strand breaks are immediately repaired when another linear DNA fragment with sequences homologous to the 5'- and 3'-ends of the double strand break is present. Botstein and coworkers utilized this feature for RE independent cloning of DNA fragments into target vectors containing origins of replication for *S. cerevisiae* and *E. coli*.[225] In this study, homologous recombination in *S. cerevisiae* was used for the assembling of the pMEc vector series from two to five linear DNA fragments. PCR products of the desired modules or a linearized vector were designed such that the two ends to be connected comprised a 50 bp homologous sequence overlap with each other. Recombination at all overlapping sites led to circular plasmid DNA. For propagation in *S. cerevisiae*, one linear DNA fragment always contained the 2µ Ori. An overview of the used PCR products can be found in chapter 4.6.

200 – 300 ng of the linear fragments to be connected were mixed without prior purification in a final volume of 65 µL of ddH_2O. *S. cerevisiae* was transformed with the DNA mix according to chapter 5.2.2.3. The transformed cells were plated on SC-Ura drop-out medium and transformants were recovered at 30 °C for 2 - 4 days. Plasmid DNA was isolated (see chapter 5.1.1) and used for transformation of electro-competent *E. coli* DH5α or Omnimax cells (see chapter 5.2.2.1).

5.2 Microbiological methods

5.2.1 Preparation of competent cells

5.2.1.1 Electro competent *E. coli* cells

E. coli cells were grown in LB medium at 37 °C and 220 rpm until the OD_{600} reached 0.7 - 0.8. Cells were harvested by centrifugation (4 °C, 12 min, 8000 x g) and washed twice with ice-cold 10% glycerol, subsequently. The washed cells were resuspended in ice-cold 10% glycerol to a final OD_{600} around 60 – 70, aliquoted as 50 µL portions, frozen in liquid nitrogen and stored at -80 °C.

5.2.1.2 $CaCl_2$ competent *E. coli* cells

E. coli cells were grown in LB medium at 37 °C and 220 rpm until $OD_{600} \approx 0.4$. The culture was cooled on ice for 10 min and cells were harvested by centrifugation (4 °C, 12 min, 8000 x g), subsequently. For washing, the pellet was resuspended in ice-cold $CaCl_2$ solution (60 mM $CaCl_2$, 10 mM PIPES, 15% glycerol) to 1/5 of the previous culture volume. The washing step was repeated twice, and the cells were resuspended in ice-cold $CaCl_2$ solution to a final OD_{600} around 60 – 70, subsequently. Competent cells were aliquoted as 50 µL portions, frozen in liquid nitrogen and stored at -80 °C.

5.2.1.3 Competent *S. cerevisiae* cells

5 mL of YPD medium were inoculated with *S. cerevisiae* Inv*Sc*1 from a glycerol stock and incubated for 1 – 2 days at 37 °C and 200 rpm. The stationary culture was used to inoculate 50 mL of pre-warmed YPD with a final $OD_{600} = 0.2$. For mid-log phase Inv*Sc*1 the culture was grown until an $OD_{600} \approx 0.8 – 1.2$ at 30 °C and 200 rpm. The cells were harvested by centrifugation (RT, 5 min, 3000 x g) and washed with 25 mL sterile ddH_2O, subsequently. The washed cells were resuspended in 1 mL 100 mM LiOAc and transferred to a 1.5 mL microfuge tube. Therein, cells were pelleted (RT, 15 s, 13000 x g) and resuspended in 100 mM LiOAc to a final volume of 500 µL. Competent cells were aliquoted in 50 µL portions and immediately used for transformation. Alternatively, 50 µL of 70% glycerol was added to an aliquot, cells were incubated on ice for 2 h, then frozen in liquid nitrogen and stored at -80 °C until further use.

5.2.2 Transformation

5.2.2.1 Transformation of electro-competent *E. coli* cells

Plasmid DNA was dialyzed against ddH_2O for 1 h using Millipore "V" Series Membranes (0.025 µm filter pore size, Schwalbach/Ts, Germany). Electro competent *E. coli* cells (see chapter 5.2.1.1) were thawed on ice, plasmid DNA and electroporation cuvettes (GenePulser/Micropulser cuvettes, 1 mm gap width; Biorad, Hercules CA, USA) were

prechilled at 4 °C. Cells were mixed with 50 – 100 ng of plasmid DNA and applied to the electroporation cuvette. Electroporation was performed in an Electroporator 1000 (Stratagene, La Jolla, USA) by applying 1650 kV. Immediately after, 1 mL LB medium was added, the suspension was transferred to a sterile 1.5 mL microfuge tube and cells were recovered for 1 h at 37 °C and 800 rpm. Finally, the cells were plated on agar plates with appropriate antibiotics and incubated overnight at 30 or 37 °C.

5.2.2.2 Transformation of $CaCl_2$ competent *E. coli* cells

Competent cells (see chapter 5.2.1.2) were rapidly thawed in the hand. 50 – 100 ng of plasmid DNA were added and the mixture was incubated on ice for 10 min. After that a heat shock at 42 °C was applied for 2 min and 1 mL LB medium was added immediately after. Cells were recovered for 1 h at 37 °C and 800 rpm. Finally, the cells were plated on agar plates with appropriate antibiotics and incubated overnight at 30 °C or 37 °C.

5.2.2.3 Transformation of competent *S. cerevisiae* cells

For transformation of *S. cerevisiae* Inv*Sc*1, competent cells (see chapter 5.2.1.3) were pelleted in a 1.5 mL microfuge tube (RT, 15 s, 13000 x g) and access LiOAc was removed with a pipette. Single stranded DNA (ssDNA, 5 mg mL^{-1}) was boiled for 5 min and quickly chilled on ice. The pellet was overlaid with 240 µL 50% (w/v) PEG 3350. Subsequently, 36 µL 1 M LiOAc, 10 µL ssDNA and 65 µL of the solution containing the plasmid DNA and/or PCR products were carefully added on top of the PEG layer. The batch was vortexed until homogeneity (ca. 1 min) and incubated for 30 min at 30 °C. Incubated cells were heat shocked at 42 °C for 20 – 25 min. To remove the transformation mix, cells were pelleted (RT, 15 s, 5000 x g), the supernatant was removed with a pipette and the cells were resuspnded in 250 µL sterile ddH_2O. The transformed cells were plated on SC-Ura drop-out medium agar plates and transformants were recovered at 30 °C for 2 - 4 days.

5.3 Protein expression and purification

5.3.1 Standard protein expression in *E. coli*

Parent protein expression was performed at non-limiting conditions in LB medium. Transformed expression strains were always prepared freshly and five single colonies were checked for target protein expression in small scale (5 mL) prior to 1 L fermentation. The best expressing clone was selected for inoculating of 1 L of LB medium with appropriate antibiotics. Cells were grown at 37 °C and 200 rpm up to OD_{600} between 0.6 and 0.8. A non-induced sample was taken for reference and protein expression was induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), subsequently. Protein expression was performed for 4 h at 30 °C and 200 rpm for *av*GFP variants, TTL and aaRSs. ψ-b* was expressed at 28 °C for 4 h. Thereafter, cells were harvested by centrifugation (4 °C, 8000 x g, 12 min).

5.3.2 Expression of proteins containing noncanonical amino acids

For incorporation of ncaas into proteins by the SPI method *E. coli* strains have to be used which are auxotrophic for respective caas to be exchanged. Since most ncaas do not support growth, auxotrophic strains have to be provided with a certain amount of the caa to allow cell growth and only when target expression is induced the ncaa is added. However, to get homogeneous protein preparations with incorporated ncaas residual amounts of the respective caas should be avoided at the timepoint of induction. Thus, the fermentation is calibrated such that the caa is only available until mid-log phase where target protein expression is induced. To determine the concentration needed for growth until OD_{600} = 0.6 - 0.7, a limitation test has to be performed for every caa to be exchanged.

5.3.2.1 Calibration of fermentation (limitation test)

50 mL NMM with different concentrations of the respective caa were inoculated with 50 µL of LB pre-culture and incubated overnight at 37 °C and 220 rpm. On the next day, OD_{600} was measured and the amino acid concentration which allowed growth to OD_{600} = 0.6 – 0.7 was selected as limiting concentration.

5.3.2.2 Incorporation of ncaas by SPI, and SPI combined with SCS

Transformed expression strains were always prepared freshly and five single colonies were checked for target protein expression in small scale (5 mL) prior to 1 L fermentation. 1 mL of the best expressing clone was used to inoculate 1 L NMM. Transformed cells were grown in NMM containing appropriate antibiotics and limiting amounts of respective canonical amino acids as natural substrate until depletion at OD_{600} between 0.6 – 0.8. Subsequently, the different combinations of ncaas were added to the medium, and target protein expression was induced by addition of 1 mM IPTG. Bpa was added in a final concentration of 1 mM, Nle, Aha, and (4*S*-F)Pro to a concentration of 0.5 mM. Protein expression was performed for 4 h at 30°C with vigorous shaking.

5.3.2.3 Incorporation of ncaas by SCS

Incorporations with the *mj*BpaRS:$tRNA^{Tyr}_{CUA}$ pair were performed as described in chapter 5.3.1. Ncaas were added in a final concentration of 1 mM at induction of protein expression. For evaluation of pMEc plasmids, transformed cells were grown at 30 °C and 200 rpm overnight. 50 mL expression cultures were inoculated to a start OD_{600} of 0.2 and grown at 30 °C and 200 rpm up to OD_{600} between 0.6 - 0.8. Non-induced samples were taken for reference and target protein expression was induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), subsequently. When aaRS genes were under the control of inducible P_{BAD} or P_{prpB} promoters, 0.02% arabinose or 50 mM propionate were added with IPTG, respectively. Protein expression was performed overnight at 30 °C and 200 rpm.

For expressions with the *sc*PheRS(T415G):*sc*$tRNA^{Phe}_{CUA\_UG}$ pair, transformed auxotrophic cells were grown in NMM supplemented with limiting amounts of Phe and Trp (see chapter 5.3.2.1). After depletion of the canonical amino acids (which leads to growth arrest) in the

mid-log phase (OD_{600} = 0.6 – 0.7), the respective ncaa was added in a final concentration of 3 mM together with 10 µM Trp and 30 µM Phe. Target protein expression was induced by addition of IPTG in a final concentration of 1 mM, subsequently. When *m*DHFR was used as target protein, expression was performed overnight at RT and vigorous shaking. EGFP and ECFP expressions were carried out at 30 °C for 5 h and vigorous shaking.

5.3.3 Purification of soluble His-tagged proteins by Ni-NTA chromatography

For later Ni-NTA chromatography, harvested cells were resuspended in 20 mL lysis buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 10 mM imidazole). Cells were lysed using a French Press or by sonication and the lysate was cleared by high speed centrifugation (15000 x g, 4 °C, 30 min). In TTL preparations, the resuspended cells were incubated for 30 min with 0.1% Triton X-100, 0.5 mg mL^{-1} lysozyme, and 1 mg mL^{-1} each, DNase and RNase prior to sonication. Soluble target proteins were purified from the supernatant using a Ni-NTA column (GE Healthcare, Munich, Germany) equilibrated with washing buffer (50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 20 mM imidazole). The column was washed with 10 - 20 column volumes (CV) of washing buffer and proteins were eluted with imidazole gradient (20 - 250 mM, 20 - 500 mM for TTL). Fractions containing the desired protein were pooled. Buffer was exchanged to Tris-HCl buffer (50 mM, pH 8.0) or aaRS buffer (see chapter 5.4.1) by dialysis with Vivaspin centrifugal ultrafiltration devices having an appropriate MWCO (Sartorius Stedim Biotech GmbH, Göttingen, Germany). Protein purity was analyzed by SDS-PAGE and Coomassie staining.

Ni-NTA chromatography under denaturing conditions was performed accordingly with the following modifications. The purification process was carried out at RT and all buffers additionally contained 8 M urea. Protein preparations were denatured by addition of urea in a final concentration of 8 M and subsequent incubation for 30 min at RT. After chromatography, target protein containing fractions were pooled and Arg was added in a final concentration of 1 M. Protein was refolded by dialysis against a 100fold excess of Tris-HCl buffer (20 mM, pH 8.0). Precipitated protein was removed by high speed centrifugation (15000 x g, 4 °C, 30 min).

5.3.4 Purification of ψ-b* by ion exchange chromatography

The pellet was resuspended in buffer I (50 mM Tris-HCl, pH 8.0, 1 mM phenylmethylsulfonyl fluoride, 50 µg mL^{-1} lysozyme) and incubated on ice for 30 min. Cell lysis was performed by sonication and the insoluble fraction (containing the ψ-b* in inclusion bodies) was harvested by high speed centrifugation (15,000 x g, 4 °C, 30 min). The pellet was homogenized at RT in buffer II (50 mM Tris-HCl, pH 8.0, 7.5 M urea) and cleared from insoluble substances by high speed centrifugation. ψ-b* was refolded by dialysis of the supernatant against a 100fold excess of buffer III (50 mM Tris-HCl, pH 8.0, 100 mM NaCl) overnight at 4 °C. To remove residual urea, dialysis was repeated two times with fresh buffer for 4 h at 4 °C. The solution was cleared from precipitated protein by high-speed centrifugation and the supernatant was applied to a 5 mL HiTrap Q Sepharose column

(GE Healthcare, Munich, Germany) equilibrated with buffer III. ψ-b* was eluted from the column with a NaCl gradient (0.1 M – 1 M NaCl). Fractions containing ψ-b* were pooled and protein purity was analyzed by SDS-PAGE and Coomassie staining.

5.4 Biochemical methods

5.4.1 ATP:PP_i exchange assay

The ATP:PP_i exchange assay was used for determination of the amino acid activation rate by a particular aaRS. In the reaction, the AMP group of ATP is transferred to the amino acid yielding aminoacyl-adenylate (AMP-aa) and PP_i. In the reverse reaction, radioactively ^{32}P-labeled PP_i forms [^{32}P]-ATP which can be selectively isolated from the reaction by binding on activated charcoal (^{a}C) and removal of [^{32}P]-PP_i by washing. Thereby, the amino acid activation rate can be determined by evaluation of the [^{32}P]-ATP formation rate.[226]

Reaction: $aa + ATP + aaRS \rightleftharpoons aaRS(AMP\text{-}aa) + PP_i$

$aaRS(AMP\text{-}aa) + tRNA \rightleftharpoons aa\text{-}tRNA + aaRS + AMP$

Radioactive reaction:

$aa + [^{32}P]ATP\text{–}^{a}C + aaRS \rightleftharpoons aaRS(AMP\text{-}aa) + [^{32}P]PP_i + {}^{a}C$

The reaction mix contained 5 µM *mj*BpaRS or 3 µM *sc*PheRS(T415G) and 5 mM amino acid in ATP:PP_i reaction buffer (100 mM Tris-HCl pH 8.0 with 80 mM $MgCl_2$, 5 mM KF, 700 mM β-mercaptoethanol, 5.5 mM ATP, 0.1 mg mL^{-1} BSA, 2.2 mM [^{32}P]-PP_i (0.2 cpm $pmol^{-1}$)) in a total volume of 200 µL. AaRSs were stored in aaRS buffer (20 mM Tris-HCl, pH 8.0, 150 mM KCl, 15 mM $MgCl_2$, 5 mM β-mercaptoethanol) as 20 µM stock solutions. Amino acid stocks had a concentration of 12.5 mM in 0.01 M HCl. As negative control 80 µL of 0.01 M HCl instead of amino acid stock solution were added to the mixture. The reaction was carried out at 37 °C for 15 min and then stopped by mixing 100 µL of the reaction mix with 600 µL 240 mM sodium pyrophosphate solution containing 70% (v/v) perchloric acid. 200 µL of a 7.5% (w/v) solution of activated charcoal was added to absorb the [^{32}P] ATP formed. The suspension was mixed thoroughly and filtered though a Whatman GF/F filter. To remove [^{32}P]-PP_i, the filter was washed twice with 10 mL of ddH_2O and then put into scintillation solution to determine the amount of radioactivity. Each value was determined twice.

5.4.2 Copper(I)-catalyzed azide–alkyne Huisgen cycloaddition

The basic reaction mixture consisted of 300 µL TTL(D221Bpa)[Aha] solution (45 µM, 1.35 mg mL^{-1}), 50 µL 1 M Tris-HCl pH 8.0, 25 µL $CuSO_4$ (50 mM), 25 µL ascorbic acid (50 mM) and a 1 cm piece of Cu wire. Additionally, 100 µL of propargyl *N*-acetylglucosaminide (77 mM) or propargyl lactoside (53 mM) were added as reaction partners for the azido group of Aha[118]. The reaction mixture was incubated for 2 days at 4 - 8 °C. Afterwards, the buffer was exchanged to Tris-HCl buffer (20 mM, pH 8.0) with centrifugal concentrators (Vivaspin 6,

Sartorius Stedim Biotech, Aubagne Cedex, France) and proteins were analyzed by SDS-PAGE gel and NanoLC-MS/MS.

5.4.3 *p*-Nitrophenyl palmitate assay

TTL activity was assayed in a spectrophotometric assay with *p*-nitrophenyl palmitate (*p*NPP, Sigma) as the substrate.[227] A *p*NPP suspension (50 mM Tris-HCl pH 8.0, 1 mM *p*NPP, 1 mg ml^{-1} gum arabic (Acros Organics, Geel, Belgium)) was homogenized with an Ultra-Turrax (IKA Werke GmbH & Co. KG, Staufen, Germany) at 22,000 rpm for 4 min at room temperature. 4 – 8 µg of lipase were typically used for the assay. The reaction was carried out for 10 min – 1 h at 70 °C with vigorous shaking. Hydrolysis was stopped by chilling the sample on ice for 1°min and addition of 100 µL of 1 M Na_2CO_3. The sample was cleared from precipitates at room temperature for 10 min at 20,000 x g and the absorbance of the supernatant was measured at 410 nm in an Ultrospec 6300 pro spectrophotometer (GE Healthcare, Munich, Germany). To substract autohydrolysis of the substrate, a blank sample with 15 µL of 50 mM Tris-HCl pH 8.0 instead of protein solution was run in parallel. All values are the results of three independent measurements. One unit (1 U) of lipase activity is defined as the amount of enzyme needed to liberate 1 µmol of *p*NPP (extinction coefficient at pH 8.0 = 12.75 x 10^6 M^{-1} cm^{-1}) per minute under the conditions described above.

5.5 Analytical methods

5.5.1 Agarose gel electrophoresis

DNA fragments were separated by agarose gel electrophoresis. Fragments smaller than 700 bp were run on 2% agarose gels in TAE buffer (40 mM Tris-HCl pH 8.0, 0.1% acetic acid, 1 mM ethylenediaminetetraacetic acid, EDTA), whereas for longer nucleotide sequences 1% agarose gels were used. To visualize DNA, gels contained 0.8 µg mL^{-1} ethidiumbromide (EtBr) or 1 x SYBR safe (Invitrogen, Karlsruhe, Germany). DNA samples were mixed with 6 x loading buffer (0.25% bromphenol blue, 0.25% xylencyanole, 30% glycerol) and applied to the gel. Electrophoresis was performed at 70 mV for 1 - 2 hours. The GeneRuler 1kb DNA Ladder from Fermentas (St. Leon-Rot, Germany) was used as standard.

5.5.2 Polyacrylamide gel electrophoresis

Proteins are separated by polyacrylamide gel electrophoresis according to their molecular masses. 12% and 17% acrylamide gels were used in combination with a 5% stacking gel. Samples were mixed with 5 x sample buffer, boiled at 96 °C for 5 – 10 min and loaded onto the gel. Electrophoresis was performed between 110 V and 180 V with running buffer. Gels were stained with coomassie staining solution for 2 h (or overnight) at RT and destained with

destaining solution prior to gel documentation. The PageRuler Prestained Protein Ladder from Fermentas (St. Leon-Rot, Germany) was used as standard.

Polyacrylamide gel electrophoresis gels and buffers		
Resolving gel	Tris-HCl pH 8.8	380 mM
	Acrylamide/bis-acrylamide (37.5:1)	12% or 17%
	Sodium dodecyl sulfate (SDS)	0.1%
	Ammonium persulfate (APS)	0.05%
	Tetramethylethylenediamine (TEMED)	0.05%
Stacking gel	Tris-HCl pH 6.8	125 mM
	Acrylamide/bis-acrylamide (37.5:1)	5%
	SDS	0.1%
	APS	0.05%
	TEMED	0.17%
5 x sample buffer	Tris-HCl pH 6.8	100 mM
	SDS	4%
	Glycerol	30%
	β-mercaptoethanol	15%
	Bromphenol blue	0.25%
Running buffer	Glycine	190 mM
	Tris-HCl	25 mM
	SDS	3.5 mM
Staining solution	Coomassie Brilliant Blue R 250	0.1%
	Ethanol	25%
	Acetic acid	8%
Destaining solution	Ethanol	25%
	Acetic acid	8%

5.5.3 Western blot

His-tag containing proteins were immuno-stained via Western blotting in a semi-dry blotting system. Proteins were separated by PAGE and the SDS gel was soaked for 5 min with transfer buffer (25 mM Tris-HCl pH 8.0, 192 mM glycine, 20% ethanol, 0.1% SDS) together with the nitrocellulose membrane (Whatman Protran 0.1 µM, GE healthcare, Munich, Germany). The SDS gel was placed on the membrane and both were covered with buffer soaked Whatman paper from both sides. Blotting was performed for 2 h at 200 mA. After electrophoresis, the membrane was stained with Ponceau S (0.5% in 1% acetic acid) to mark lanes and the protein standard bands. The membrane was washed with ddH_2O and unspecific antibody binding sites were blocked by incubation for 1 h at RT in TTBS buffer

(50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Tween20) containing 1.5% bovine serum albumin (BSA). Thereafter, the membrane was incubated with an anti-His-tag antibody (1:2500 in TTBS buffer; 6 x His-tag [ADI.I.I0] antibody, GeneTex, Irvine, USA) overnight at 4 °C. The membrane was washed three times with TTBS for 10 min, and incubated with horseradish peroxidase (HRP) conjugated goat anti-mouse antibody (1:4000 in TTBS buffer; Goat Anti-Mouse IgG (H+L)-HRP Conjugate #172-1011, Bio-Rad, Hercules, USA) for 1 h at RT, subsequently. After washing three times with TTBS for 10 min, the chemiluminescence reaction was started by addition of 1:1 luminol:peroxide reagent (Western Lightning Western Blot Chemiluminescence Reagent, PerkinElmer, Rodgau, Germany). After 5 min of incubation, the chemiluminescence reagent was stripped off and photo paper was exposed to the membrane for 30 s – 5 min. The film was developed in a X-OMAT 1000 processor photo developer (Eastman Kodak, Rochester, USA). For StrepII-tag staining, Strep MAB-Class antibody (1:1000 in TTBS buffer; IBA BioTAGnology, Göttingen, Deutschland) was used instead of an anti-His-antibody.

5.5.4 ImageJ analysis of SDS gels and Western blots

For relative quantification of protein expression, the band intensities of SDS gels and Western blots were integrated with the program ImageJ (http://rsbweb.nih.gov/ij/index.html) according to the documentation of Luke P. Miller (http://lukemiller.org/index.php/2010/11/analyzing-gels-and-western-blots-with-image-j).

5.5.5 Determination of protein concentration

The concentrations of unmodified parent proteins were determined in their standard buffer by measurement of UV_{280} absorbance (see chapter 5.5.8) and a calculated ε_M (ProtParam, Expasy Proteomics Server, www.expasy.ch/tools/#proteome). Congener and mutated congener concentrations were measured with the Bradford protein assay.[228] 100 µL (1 – 10 µg) of protein solution were mixed with 900 µL of Bradford reagent (100 mg L^{-1} Coomassie brilliant blue G-250, 5% ethanol, 10% phosphoric acid) and incubated for 5 min. Subsequently, absorbance was measured at 595 nm. The protein concentration was calculated by comparing the absorbance with a calibration curve of the respective parent protein.

5.5.6 LC-ESI-MS

For LC-ESI-MS, 20 µL aliquots of the purified proteins were pre-separated on a Discovery Bio Wide Pore C5 column (3.5 µm particle size, 100 x 2.1 mm, Supelco, Bellefonte, USA) by eluting with a gradient from 20% B to 90% B within 15 min and 90% B to 95% B in 2 min (eluent A: 0.05% (v/v) TFA in water, eluent B: 0.05% (v/v) TFA in acetonitrile). A flow rate of 250 µL min^{-1} was used. The masses of the eluted fractions were analyzed by MicroTOF ESI-MS (Bruker Daltonics, Bremen, Germany). The absorbance was detected at 210 nm.

Analysis was performed by Mrs. Elisabeth Weyher-Stingl at the Microchemistry Core Facility of the Max Planck Institute of Biochemistry.

5.5.7 Orbitrap MS (NanoLC-MS/MS)

Analysis was performed by Dr. Cyril Boulegue at the Microchemistry Core Facility of the Max Planck Institute of Biochemistry.

In-gel digestion

Gel bands were cut into 1 mm^3 cubes and washed two times with 50 mM ammonium bicarbonate, 50% ethanol. For protein reduction, gel pieces were incubated with 10 mM dithiothreitol (DTT) in 50 mM ammonium bicarbonate for 1 h at 56 °C. Alkylation of cysteines was performed by incubating the samples with 10 mM iodoacetamide in 50 mM ammonium bicarbonate for 45 min at 25 °C in the dark. Gel pieces were washed two times with 50 mM ammonium bicarbonate, 50% ethanol, dehydrated with 100% ethanol, and dried in a vacuum concentrator, subsequently. The gel pieces were rehydrated with 12.5 ng μL^{-1} trypsin (sequencing grade, Promega, Mannheim, Germany) or endoproteinase Glu-C (exicision grade, Calbiochem, Darmstadt, Germany) in 50 mM ammonium bicarbonate and incubated overnight at 37 °C for protein digestion.[229] Supernatants were transferred to fresh tubes, and the remaining peptides were extracted by incubating the gel pieces two times in 30% acetonitrile in 3% TFA followed by dehydration with 100% acetonitrile. The extracts were combined and desalted using RP-C18 StageTip columns, and the eluted peptides were subjected to mass spectrometric analysis.

NanoLC-MS/MS

All digested peptide mixtures were separated by online nanoLC and analyzed by electrospray tandem mass spectrometry. The experiments were performed on an Agilent 1200 nanoflow system connected to an LTQ Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany) equipped with a nanoelectrospray ion source (Proxeon Biosystems, Odense, Denmark). Binding and chromatographic separation of the peptides took place in a 15 cm fused silica emitter (75 µm inner diameter from Proxeon Biosystems, Odense, Denmark), in-house packed with reversed-phase ReproSil-Pur C18-AQ 3 µm resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). Peptide mixtures were injected onto the column with a flow of 500 nL min^{-1} and subsequently eluted with a flow of 250 nL min^{-1} from 2% to 40% acetonitrile in 0.5% acetic acid, in a 100 min gradient. The precursor ion spectra were acquired in the Orbitrap analyzer (m/z 300 – 1800, R = 60,000, and ion accumulation to a target value of 1,000.000), and the five most intensive ions were fragmented and recorded in the ion trap. The lock mass option enabled accurate mass measurement in both MS and Orbitrap MS/MS mode as described previously.[230] Target ions already selected for MS/MS were dynamically excluded for 90 s.

Peptide identification via MASCOT database search

Data analysis was performed with the MaxQuant software as described elsewhere[231] and supported by Mascot as database search engine for peptide identifications. Peaks in MS scans were determined as three-dimensional hills in the mass-retention time plane. MS/MS peak lists were filtered to contain at most six peaks per 100 Da interval and searched by Mascot (Matrix Science) against a concatenated forward and reversed version of *E. coli* K12 data (extracted from NCBInr and containing the sequence of the target proteins with frequently observed contaminants like proteases and human keratins). The initial mass tolerance in MS mode was set to 7 ppm and MS/MS mass tolerance was 0.5 Da. Cysteine carbamidomethylation was searched as a fixed modification, whereas N-acetyl protein, oxidized methionine and ncaa incorporations were searched as variable modifications.

Data analysis

Desired peptides containing the sequence positions to be analyzed were extracted from the dataset. If possible, the peptide species without Met oxidation were used for the estimation of ncaa incorporation. In case of TTL(F37X), however, this procedure was not possible because the dataset mainly contained fully oxidized peptide species. Thus, fully oxidized peptides were taken for analysis. In the analysis of TTL(D221Bpa)[Aha] and its cycloaddition products, oxidized species are only listed if no unoxidized peptides are found as well. The same holds true for peptides with missed cleavage sites. For estimation of ncaa incorporation, Peptide count (number of specific peptides sequenced), MS/MS count (number of times a specific peptide was selected for a sequencing event) and peptide intensity were used.

5.5.8 UV/VIS spectroscopy

Absorbance spectra were recorded on a Lambda 19 photometer (PerkinElmer, Überlingen, Germany). Spectra were measured at 20 °C in Tris-HCl buffer (100 mM, pH 8.0). Molar extinction coefficients (ε_M) were calculated from absorbance using the Lambert-Beer-Equation $A = \varepsilon \cdot c \cdot d$ (A, absorbance; ε, extinction coefficient; c, concentration; d, path length).

5.5.9 Fluorescence spectroscopy

Fluorescence spectra of EGFP and its variants were measured at 20 °C in Tris-HCl buffer (100 mM, pH 8.0) using the luminescence spectrometer LS50B (PerkinElmer, Überlingen, Germany). Spectra were recorded from 500 – 600 nm using an excitation wavelength of 488 nm, excitation/emission slits of 5/5 and a filter cut-off at 350 nm. The protein concentration was adjusted to 0.3 µM.

GFPuv fluorescence intensity in cleared cell lysates was determined for evaluation of the pMEc vector family. 1 OD_{600} of cells was harvested and resuspended in phosphate buffer (50 mM, pH 7.5, 100 mM NaCl). Cells were opened by sonication and the lysates were cleared by high speed centrifugation (RT, 5 min, 13000 x g). GFP fluorescence in the cleared

lysates was measured at 20 °C and detected at 510 nm upon excitation at 395 nm. The excitation/emission slits were set to 2.5/2.5 and fluorescence was cut off at 350 nm.

5.5.10 Refolding of EGFP

15 µM of EGFP or one of its variants were denatured for 5 min at 95 °C in sodium phosphate buffer (50 mM, 100 mM NaCl, pH 7.4) containing 8 M urea and 100 mM DTT. Refolding was initialized by a 1:100 dilution of the sample in sodium phosphate buffer (50 mM, 100 mM NaCl, pH 7.4) supplemented with 1 mM $MgCl_2$ and 5 mM DTT. Refolding was recorded at RT in time intervals of 2 s for 60 min using the luminescence spectrometer LS50B measuring the increase of the fluorescence signal at 504 nm upon excitation at 488 nm. The excitation/emission slits were set to 5/5.

6 Appendix

6.1 Supplementary material for characterization of *sc*PheRS(T415G)

6.1.1 Expression and purification of *m*DHFR(F38X) preparations

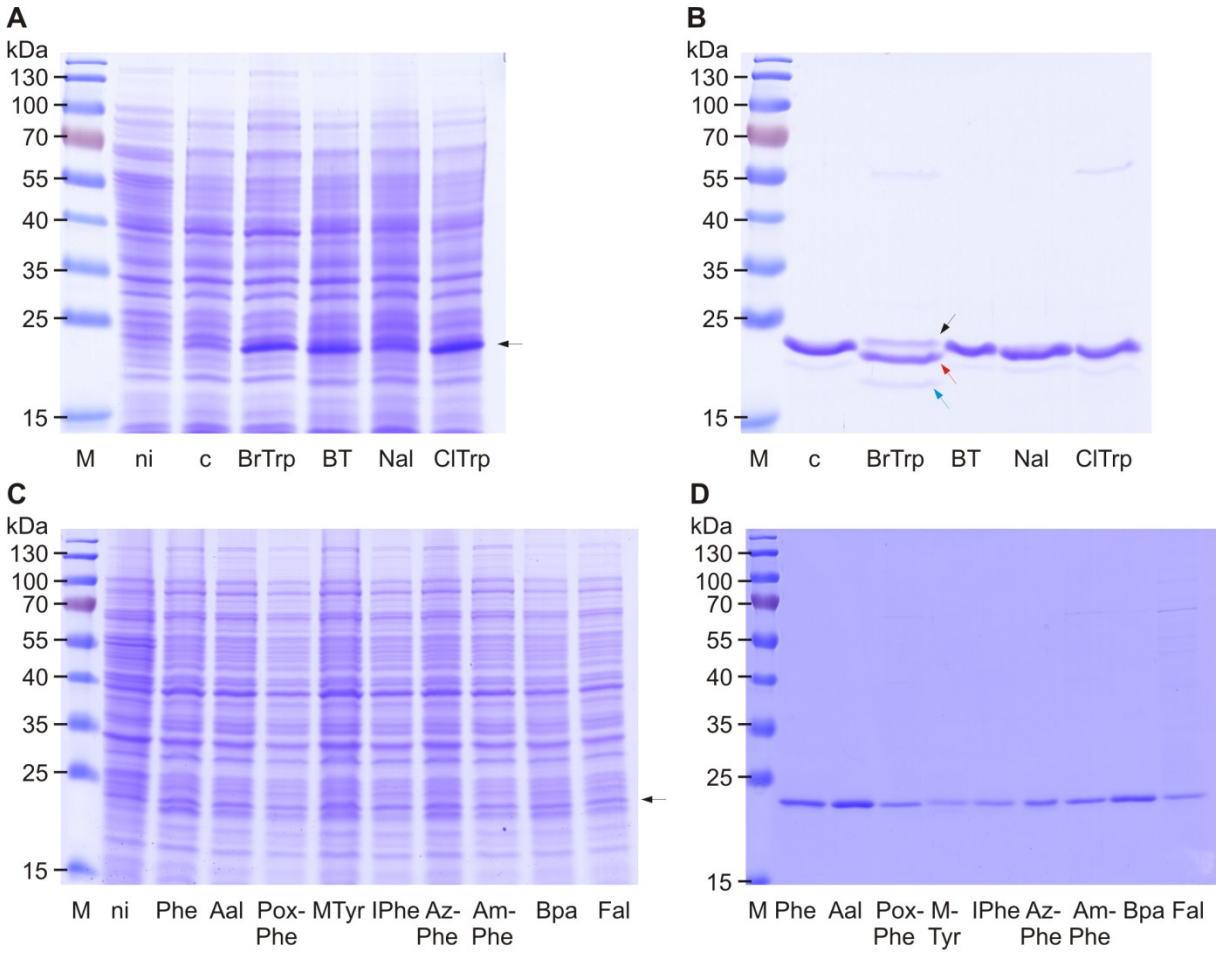


Figure 40. Expression and purity of *m*DHFR(F38X) expressed in *E. coli* AFW with different ncaas. A and **C**: The expression of *m*DHFR(F38X) in the presence of 3 mM of the respective ncaa is shown. 0.15 OD_{600} of cell lysates were applied to a 12% SDS gel. Black arrows indicate the *m*DHFR expression bands (ca. 23 kDa). **M**: marker; **ni**: non-induced sample; **c**: control expression with 10 µM Trp and 30 µM Phe as caas were added. The abbreviations of ncaas can be found in **Figure 13**. **B** and **D**: Purity of all *m*DHFR preparations. 1 – 3 µg of protein were applied to a 12% SDS gel. Preparations are prone to degradation as can best be seen in case of the 6BrTrp preparation. The original protein band (black arrow) is already strongly diminished in favor of a second (red arrow) and third (blue arrow) band. With proceeding time the red and blue arrowed bands further intensify, whereas the black arrowed band decreases (data not shown).

6.1.2 ESI-MS analyses of *m*DHFR(F38X) preparations

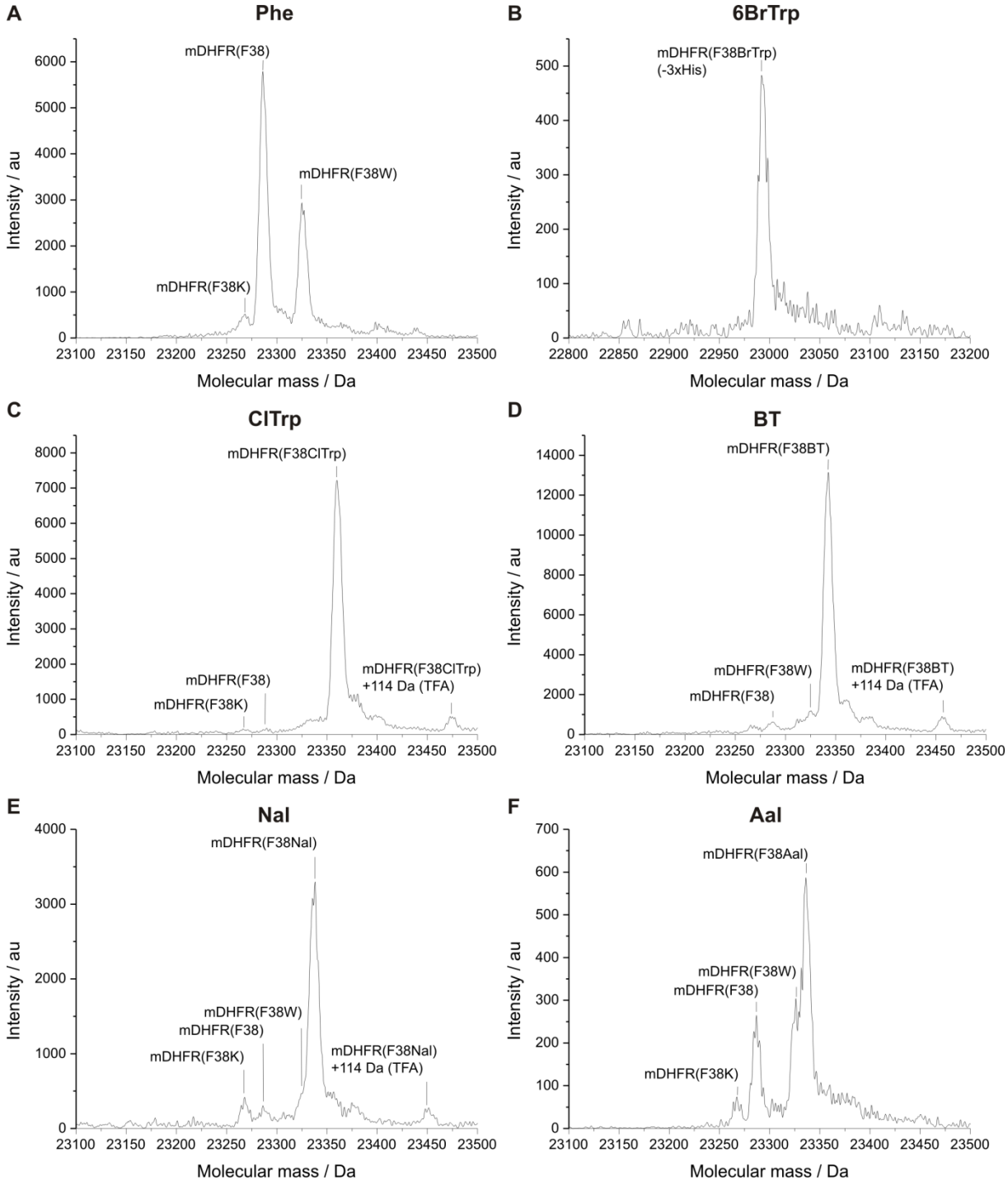


Figure 41. Mass spectra of *m*DHFR(F38X) preparations. An overview of calculated and detected masses can be found in Table 8.

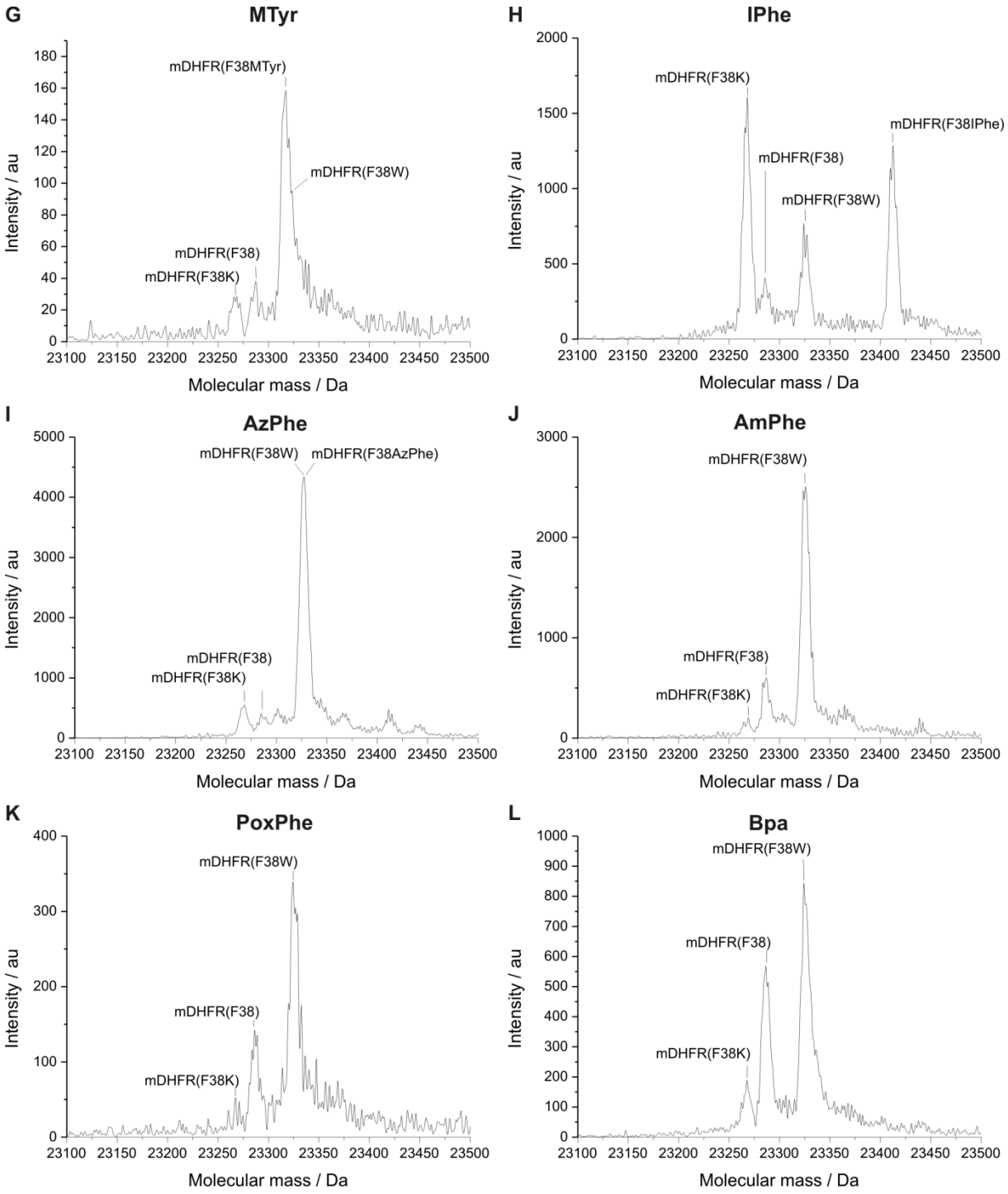


Figure 41 continued. Mass spectra of *m*DHFR(F38X) preparations. An overview of calculated and detected masses can be found in Table 8.

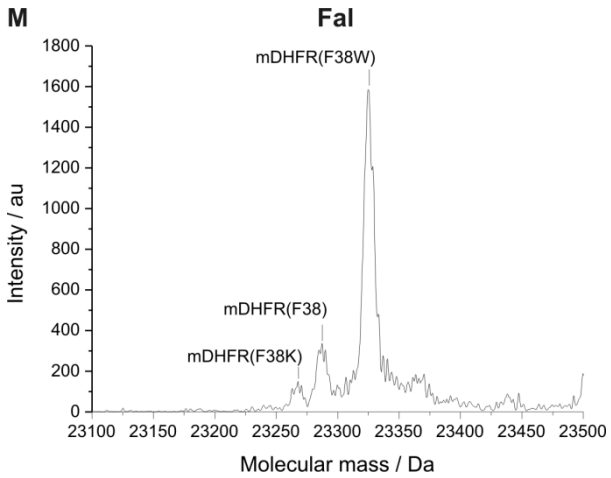


Figure 41 continued. Mass spectra of *m*DHFR(F38X) preparations. An overview of calculated and detected masses can be found in Table 8.

Table 8. Overview of calculated and respectively detected masses of *m*DHFR with a differently occupied position 38.

Incorporated amino acid	Calculated mass / Da	Detected mass / Da
Lysine	23268	23268
Phenylalanine	23287	23286
Tryptophan	23326	23325
6-bromotryptophan	23405	22992 (-3xHis)
6-chlorotryptophan	23360	23360
3-(3-benzothienyl)alanine	23343	23343
3-(2-naphthyl)alanine	23337	23338
3-(1-azulenyl)alanine	23337	23336
O-methyltyrosine	23317	23317
4-iodophenylalanine	23413	23412
4-azidophenylalanine	23328	23327
4-aminophenylalanine	23302	-
4-propargyloxyphenylalanine	23341	-
4-benzoylphenylalanine	23391	-
3-(2-furyl)alanine	23277	-

6.1.3 Orbitrap analyses of *m*DHFR(F38X)+Phe

Peptide sequence	Calculated Mass / Da	Detected Mass / Da	Mascot Score	Int. / au	Int. / %	Peptide count	MS/MS count
WKYFQRMTTTSSVE	1762.83	1762.8349	58.56	25831000	22	14	25
FKYFQRMTTTSSVE	1723.82	1723.824	32.74	77875000	58	36	72
KKYFQRMTTTSSVE	1704.85	1704.8505	44.06	29488000	19	3	3

6.2 Orbitrap analyses of TTL(F37X) with Bpa, Fal, AzPhe, IPhe and Nal

Peptide sequence	Calculated Mass / Da	Detected Mass / Da	Mascot Score	Int. / au	Int. / %	Peptide count	MS/MS count
TTL(F37X)+Bpa							
VPM(ox)VIM(ox)FHGFTGNK	1609.9128	1608.7793	14.88	2247800	2	3	4
VPM(ox)VIM(ox)FHG**Bpa**TGNK	1714.0128	1712.8055	43.84	88309000	98	26	34
TTL(F37X)+Nal							
VPM(ox)VIM(ox)FHGFTGNK	1609.9128	1608.7793	51.87	59145000	87	8	4
VPM(ox)VIM(ox)FHG**Nal**TGNK	1660.0238	1658.7949	47.72	9193000	13	4	2
TTL(F37X)+Fal							
VPM(ox)VIM(ox)FHGFTGNK	1609.9128	-	-	-	-	-	-
VPM(ox)VIM(ox)FHG**Fal**TGNK	1599.8738	-	-	-	-	-	-
TTL(F37X)+IPhe							
VPM(ox)VIM(ox)FHGFTGNK	1609.9128	1608.7793	14.16	1484400	100	2	4
VPM(ox)VIM(ox)FHG**IPhe**TGNK	1735.8098	-	-	-	-	-	-
TTL(F37X)+AzPhe							
VPM(ox)VIM(ox)FHGFTGNK	1609.9128	1608.7793	64.57	46496000	50	6	7
VPM(ox)VIM(ox)FHG**AzPhe**TGNK	1650.9238	1649.7807	11.76	46765000	50	2	4

6.3 Mass analyses of TTL, TTL(D221Bpa) and their congeners

6.3.1 ESI-MS analyses

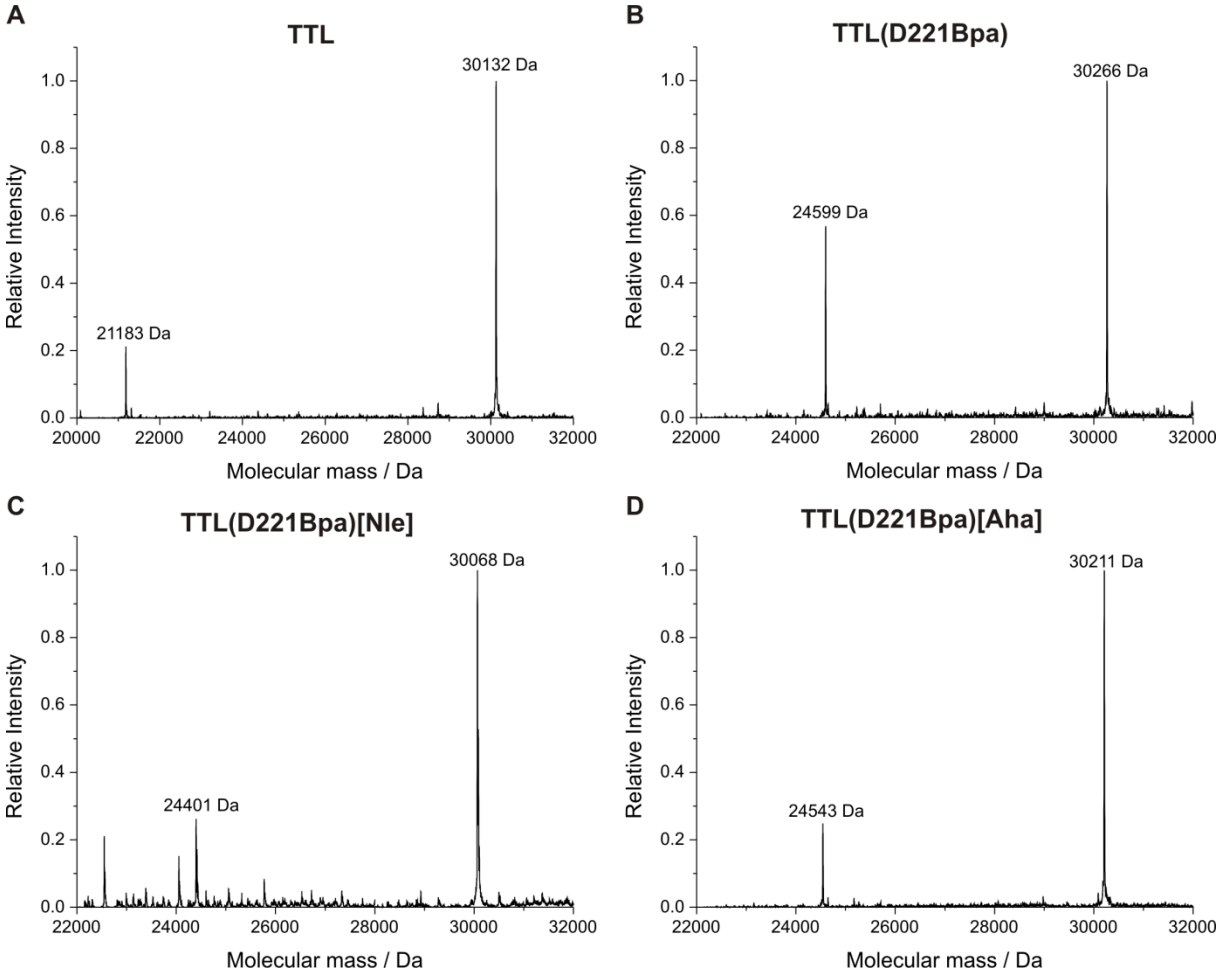


Figure 42. Full mass spectra of TTL, TTL(D221Bpa), TTL(D221Bpa)[Nle], and TTL(D221Bpa)[Aha]. A: Full-length TTL: $M_{w,\,cal}$ = 30131 Da, $M_{w,\,meas}$ = 30132 Da; N-terminal TTL truncation product: $M_{w,\,cal}$ (TTL aa 81-267) = 21183 Da, $M_{w,\,meas}$ = 21183 Da. **B**: Full-length TTL(D221Bpa): $M_{w,\,cal}$ = 30266 Da, $M_{w,\,meas}$ = 30266 Da; TTL(D221Amber) truncation product: $M_{w,\,cal}$ (TTL aa 1-220) = 24599 Da, $M_{w,\,meas}$ (TTL aa 1-220) = 24599 Da. **C**: Full-length TTL(D221Bpa)[Nle]: $M_{w,\,cal}$ = 30068 Da; $M_{w,\,meas}$ = 30068 Da; TTL(D221Amber)[Nle] truncation product: $M_{w,\,cal}$ (TTL aa 1-220) = 24401 Da, $M_{w,\,meas}$ (TTL aa 1-220) = 24401 Da. **D**: Full-length TTL(D221Bpa)[Aha]: $M_{w,\,cal}$ = 30211 Da; $M_{w,\,meas}$ = 30211 Da; TTL(D221Amber)[Aha] truncation product: $M_{w,\,cal}$ (TTL aa 1-220) = 24543 Da, $M_{w,\,meas}$ (TTL aa 1-220) = 24544 Da.

6.3.2 Orbitrap analyses of TTL(D221Bpa)Aha and its Huisgen cycloaddition products

6.3.2.1 TTL(D221Bpa)Aha

Peptide sequence	Measured Mass / Da	Mascot Score	Int. / au	Int. / %	Peptide count	MS/MS count
XQKAVEITYNGK						
Aha	1375.7208	59.18	2430600	100	2	3
Met	-	-	-	-	-	-
GXXHLPDDVK						
2 x Aha	1131.5534	35.72	37121000	47	10	8
1 x Aha, 1 x Met	1136.5397	21.12	36340000	46	2	3
2 x Met	1141.526	26.52	4876600	6	2	3
VPXVIXFHGFTGNK						
2 x Aha	1566.8168	46.66	118500000	67	44	93
1 x Aha, 1 x Met	1571.8031	53.89	55235000	31	2	4
2 x Met	1576.7894	50.32	2570300	1	2	3
XSR (as VESHFIFVKXSR)						
Aha	1473.7841	23.94	2677400	100	-	1
Met	-	-	-	-	-	-
FDFYGSGESDGDFSEXTFSSELEDAR						
Aha	2919.1744	129.29	274390000	99	22	61
Met	-	-	-	-	-	-
Met(ox)	2940.1556	82.23	814580	<1	1	8
IGLLGLSXGGAIAGIVAR						
Aha	1662.9893	131.01	709890000	>99	31	56
Met	1667.9756	35.35	1651500	<1	6	15
ALVLWAPAFNXPELIXNESVK						
2 x Aha	2362.2546	61.43	1590100000	>98	17	31
1 x Aha, 1 x Met	2367.2409	46.74	10520000	<1	1	1
2 x Met	-	-	-	-	-	-
2 x Met(1 x ox)	2388.2222	60.39	12886000	<1	2	4
QYGAIXEQLGFVDIGGHK						
Aha	1956.9806	84.06	354790000	83	31	113
Met	-	-	-	-	-	-
Met(ox)	1977.9618	91.52	71145000	17	3	15
EVYGYNATR						
Bpa	1159.5298	51.6	10590000	98	23	32
Asn	1023.4621	43	206920	2	1	1

6.3.2.2 TTL(D221Bpa)[Aha]-{triazole}Lac

Peptide sequence	Measured Mass / Da	Mascot Score	Int. / au	Int. / %	Peptide count	MS/MS count
XQKAVEITYNGK						
nd	-	-	-	-		-
GXXHLPDDVK						
2 x Aha	1131.5534	15.84	682910	21	1	2
1 x Aha(click), 1 x Met(ox)	1516.6715	3.67	2627300	79	2	2
VPXVIXFHGFTGNK						
2 x Aha	1566.8168	38.74	130870000	96	7	11
1 x Aha, 1 x Met	1571.8031	44.58	5274300	4	2	2
XSR (as VESHFIFVKXSR)						
Aha	1473.7841	12.56	1381800	100	1	1
Met	-	-	-	-	-	-
FDFYGSGESDGDFSEXTFSSELEDAR						
Met(ox)	2940.1556	46.92	1499500	100	3	3
IGLLGLSXGGAIAGIVAR						
Aha	1662.9893	71.94	6020700	67	39	129
Met	1667.9756	81.33	1877100	23	2	4
AIVIWAPAFNXPELIXNESVK						
1 x Aha(click), 1 x Met(ox)	2747.3728	11.45	650610	60	1	1
2 x Met(2 x ox)	2404.2171	8.15	444770	40	1	1
QYGAIXEQLGFVDIGGHK						
Aha	1956.9806	31.22	1735300	20	1	1
Met	1961.9669	42.87	6897700	80	2	4
EVYGYNATR						
Bpa	1159.5298	47.88	6207600	100	43	71

6.3.2.3 TTL(D221Bpa)[Aha]-{triazole}GlcNAc

Peptide sequence	Measured Mass / Da	Mascot Score	Int. / au	Int. / %	Peptide count	MS/MS count
XQKAVEITYNGK						
Aha(click)	1634,8264	34.33	2411200	100	2	2
GXXHLPDDVK						
2 x Aha(click) as +GK	1834.881	9.39	602270	14	1	1
1 x Aha(click), 1 x Met	1395.6453	6.15	5246600	77	2	3
2 x Met(ox)	1173.5158	23.04	937650	9	1	1
VPXVIXFHGFTGNK						
2 x Aha	1566.8168	47.25	152960000	95	7	10
1 x Aha, 1 x Met	1571.8031	18.49	673900	<1	1	1
1 x Met, 1 x Met(ox)	1592.7843	21.69	650040	<1	3	2
1 x Aha(click), 1 x Aha as GK +	2011.0388	6.49	7599800	5	2	2
XSR (as VESHFIFVKXSR)						
Aha	1473.7841	32.64	5717400	100	2	4
Met	-	-	-	-	-	-
FDFYGSGESDGDFSEXTFSSELEDAR						
1 x Aha(click)	3178.2799	10.83	20722000	97	5	5
Met(ox)	2940.1556	41.74	672820	3	4	4
IGLLGLSXGGAIAGIVAR						
Aha(click)	1922.0949	33,47	1218200000	100	36	90
Aha	1662.9893	42,37	756130	<1	3	4
Met	1667.9756	50,33	2391100	<1	4	6
ALVLWAPAFNXPELIXNESVK						
1 x Aha(click)	2626.3465	8.13	615330	40	1	2
2 x Met(ox)	2404.2171	6.4	930310	60	1	1
QYGAIXEQLGFVDIGGHK						
Aha(click)	2216.0862	27.81	53703000	76	3	7
Met	1961.9669	43.05	16946000	24	3	7
EVYGYNATR						
Bpa	1159.5298	47.91	6156900	95	21	93
Asn	1023.4621	38.68	348740	5	1	1

6.4 Expression, purification, and mass analysis of TTL(D221Bpa, S261C)[Aha]

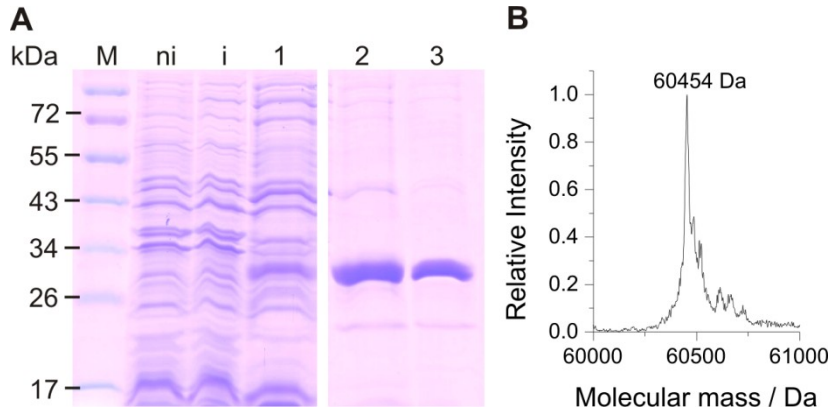
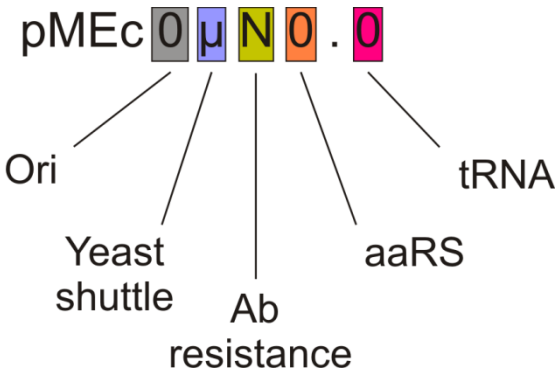


Figure 43. Expression, purity, and mass spectrometric analysis of TTL(D221Bpa, S261C)[Aha]. Expression was performed in *E. coli* CAG18515{pQE80L-TTL-H6(D221amber, S261C)/pSup-BpaRS-6TRN}. 0.15 OD_{600} of cell lysates were applied to a 12% SDS gel. The protein was purified by denaturing Ni-NTA chromatography. **A**: M: marker; ni: non-induced sample; i: induced sample; 1: TTL(D221Bpa, S261C)[Aha] in 8 M urea prior to Ni-NTA chromatography; 2: protein in 8 M urea after Ni-NTA chromatography and 3: Refolded protein in Tris-HCl (20 mM, pH 8.0). **B**: Mass spectrometric profile of TTL(D221Bpa, S261C)[Aha]. $M_{w, cal}$ = 30227 Da, $M_{w, mea}$ = 60453 Da. In Tris-HCl without further additives, the protein is present as a dimer with an intermolecular disulfide bond.

6.5 pMEc modules and sequences

6.5.1 Overview of all defined modules



Ori

Code	Ori
1	P15A
2	RSF
3	ColE1
4	F-Ori
µ	Yeast shuttle

tRNA

Code	tRNA module
1	*proK*-6xtRNA$^{Tyr}_{CUA}$
2	deleted
3	*proK*-3xtRNA$^{Tyr}_{CUA}$
4	*lpp*-tRNA$^{Phe}_{CUA_UG}$

Ab resistance

Code	Antibiotic
A	ampicillin
C	chloramphenicol
E	erythromycin
G	gentamycin
K	kanamycin
N	neomycin
Na	nalidixic acid
R	rifampicin
S	spectinomycin
St	streptomycin
T	tetracycline
Tm	trimethoprim
Z	zeocin

aaRS

Code	aaRS module
1	*glnS'*-BpaRS
2	*tac-sc*PheRS(T415G)
3	*prpR*-STREP-*sc*PheRS(T415G)
4	*prpR*-STREP-BpaRS
5	*tac-sc*PheRS(T415G)-*glnS'*
6	*araC*-STREP-BpaRS
7	deleted
8	*araC*-STREP-*sc*PheRS(T415G)

6.5.2 Sequences

6.5.2.1 Origins of replication and yeast shuttle

P15a Ori (1)

*Pau*I
GCGCGCaacaacttatatcgtatggggctgacttcaggtgctacatttgaagagataaattgcactgaaatctag
taatattttatctgattaataagatgatcttcttgagatcgttttggtctgcgcgtaatctcttgctctgaaaac
gaaaaaaccgccttgcagggcggtttttcgaaggttctctgagctaccaactctttgaaccgaggtaactggctt
ggaggagcgcagtcaccaaaacttgtcctttcagtttagccttaaccggcgcatgacttcaagactaactcctct
aaatcaattaccagtggctgctgccagtggtgcttttgcatgtctttccgggttggactcaagacgatagttacc
ggataaggcgcagcggtcggactgaacgggggttcgtgcatacagtccagcttggagcgaactgcctacccgga
actgagtgtcaggcgtggaatgagacaaacgcggccataacagcggaatgacaccggtaaaccgaaaggcaggaa
caggagagcgcacgagggagccgccaggggaaacgcctggtatctttatagtcctgtcgggtttcgccaccact
gatttgagcgtcagatttcgtgatgcttgtcaggggggcggagcctatggaaaaacggctttgccgcggccctct
cacttccctgttaagtatcttcctggcatcttccaggaaatctccgccccgttcgtaagccatttccgctcgccg
cagtcgaacgaccgagcgtagcgagtcagtgagcgaggaagcggaatatatccGGGCCC
*Apa*I

RSF Ori (2)

*Pau*I
GCGCGCgtaacggaatagctgttcgttgacttgatagaccgattgattcatcatctcataaataaagaaaaacca
ccgctaccaacggtggttttctcaaggttcgctgagctaccaactctttgaaccaaggtaagtgggttggaggac
cgcactcaccaaaatctgttcttttcagtttagccttaacaggtgcataacttcaagacaaagtcctctaaatcag
ttaccaatggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataag
gcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgaga
taccaacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggc
agggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggttt
cgccacctctggcttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgcctgcggc
gttggcttcttccggtgctttgctttttgctcacatgttctttccggctttatcccctgattctgtggataaccg
tattaccgcttttgagtgagctgacaccgctcgccgcagtcgaacgaccgagcgtagcgagtcagtgagcgagga
agcggaagagcgctGGGCCC
*Apa*I

F Ori (4)

*Pau*I

GCGCGCACCGTGGGAAAAACTCCAGGTAGAGGTACACACGCGGATAGCCAATTCAGAGTAATAAACTGTGATAAT
CAACCCTCATCAATGATGACGAACTAACCCCCGATATCAGGTCACATGACGAAGGGAAAGAGAAGGAAATCAACT
GTGACAAACTGCCCTCAAATTTGGCTTCCTTAAAAATTACAGTTCAAAAAGTATGAGAAAATCCATGCAGGCTGA
AGGAAACAGCAAAACTGTGACAAATTACCCTCAGTAGGTCAGAACAAATGTGACGAACCACCCTCAAATCTGTGA
CAGATAACCCTCAGACTATCCTGTCGTCATGGAAGTGATATCGCGGAAGGAAAATACGATATGAGTCGTCTGGCG
GCCTTTCTTTTTCTCAATGTATGAGAGGCGCATTGGAGTTCTGCTGTTGATCTCATTAACACAGACCTGCAGGAA
GCGGCGGCGGAAGTCAGGCATACGCTGGTAACTTTGAGGCAGCTGGTAACGCTCTATGATCCAGTCGATTTTCAG
AGAGACGATGCCTGAGCCATCCGGCTTACGATACTGACACAGGGATTCGTATAAACGCATGGCATACGGATTGGT
GATTTCTTTTGTTTCACTAAGCCGAAACTGCGTAAACCGGTTCTGTAACCCGATAAAGAAGGGAATGAGATATGG
GTTGATATGTACACTGTAAAGCCCTCTGGATGGACTGTGCGCACGTTTGATAAACCAAGGAAAAGATTCATAGCC
TTTTTCATCGCCGGCATCCTCTTCAGGGCGATAAAAAACCACTTCCTTCCCCGCGAAACTCTTCAATGCCTGCCG
TATATCCTTACTGGCTTCCGCAGAGGTCAATCCGAATATTTCAGCATATTTAGCAACATGGATCTCGCAGATACC
GTCATGTTCCTGTAGGGTGCCATCAGATTTTCTGATCTGGTCAACGAACAGATACAGCATACGTTTTTGATCCCG
GGAGAGACTATATGCCGCCTCAGTGAGGTCGTTTGACTGGACGATTCGCGGGCTATTTTTACGTTTCTTGTGATT
GATAACCGCTGTTTCCGCCATGACAGATCCATGTGAAGTGTGACAAGTTTTTAGATTGTCACACTAAATAAAAAA
GAGTCAATAAGCAGGGATAACTTTGTGAAAAAACAGCTTCTTCTGAGGGCAATTTGTCACAGGGTTAAGGGCAAT
TTGTCACAGACAGGACTGTCATTTGAGGGTGATTTGTCACACTGAAAGGGCAATTTGTCACAACACCTTCTCTAG
AACCAGCATGGATAAAGGCCTACAAGGCGCTCTAAAAAAGAAGATCTAAAAACTATAAAAAAAATAATTATAAAA
ATATCCCCGTGGATAAGTGGATAACCCCAAGGGAAGTTTTTTCAGGCATCGTGTGTAAGCAGAATATATAAGTGC
TGTTCCCTGGTGCTTCCTCGCTCACTCGAAATTCCCGGGGATAGCTTTATGCTTGTAAACCGTTTTGTGAAAAAA
TTTTTAAAATAAAAAAGGGGACCTCTAGGGTCCCCAATTAATTAGTAATATAATCTATTAAAGGTCATTCAAAAG
GTCATCCACCGGATCAATTCCCCTGCTCGCGCAGGCTGGGTGCCAAGCTCTCGGGTAACATCAAGGCCCGATCCT
TGGAGCCCTTGCCCTCCCGCACGATGATCGTGCCGTGATCGAAATCCAGATCCTTGACCCGCAGTTGCAAACCCT
CACTGATCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAAT
TAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAAT
TGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGGTACCTGTCGCGGCAA
CGCGCTAACAGACGTAGTAAGAACCACCAGCATTGTAATGCTGGCTAAAGTCACTTTCCTGAGCTGTATAACGAT
GAGCGATTTTACTTTTTCTGGCTATGAATTGGCCTGCTTTGTAACACACTCCGGTCTATCCCGTAGCGCCGGGCA
TATCCTGTCGCAATGTGCAAATCTCGCGGCAACAACCAGTGAATACTTCATTCACAAGCCTCACCGCCTGATCGC
GGCAGAAACTGGTTATAGCCAATCAACCGTCGTTCGTGCATTCCGTGAAGCTGTAAACAAAGGTATCCTGTCTGT
AGAGATTGTTATCGGCGATCACCGTGAACGTCGCGCTAACCTGTACCGGTTTACACCATCCTTTTTGGCCTTCGC
ACAACAAGCCAAAAATGCGCTGATAGAAAGCAAATTAAAGATCTCTTCAGCGGCAACCAAGGTTAAAGCTGTTCT
CGCTAAGACATTGGCTTTATTTAATTTTTTATCCACACCCCCATGTCAAAATGATACCCCCTCCCCCTGTCAGGA
TGACGTGGCAATAAAGAATAAGAAGTCACAAGTTAAAAAAACAAAAAGATCAGTTTCCGGCGGTGCCGGAACAAC
CAGCCTCAAAAAATTGACTTCATGGATCGCTAAGGCAAAAGCAAAGGCTGACAATCTGCGGTTATCCAAAAAACG
CACTCAAAAACATGAGTTCAAGCAGAAAGTAGAGGCGGCTGCGCGGAAATATGCTTACCTGAAGAACAAGCGTTC
GCCTGATATTGGCGGGATATCAAACTTCGATAACCTACCGCATTGCATGACGGTAAACGAAGCTCTTAATGCGGT
TTTAGCCAAAAATAAAGATAACGAACAATGGGGTATACCGGCAGGATTCAGAGGGTAATGAATTGCTCTAATTAT
AACCATGCATACTTTCAACACCTCTAGTTTGCCATGAGGCAAACTCATAGGTGTCCTGGTAAGAGGACACTGTTG
CCAAAACTGGACGCCGGGCCC

*Apa*I

Yeast shuttle module (μ)

*Pau*I

GCGCGCcattgcgaataccgcttccacaaacattgctcaaaagtatctctttgctatatatctctgtgctatatc
cctatataacctacccatccacctttcgctccttgaacttgcatctaaactcgacctctacatttttttatgttta
tctctagtattactctttagacaaaaaaattgtagtaagaactattcatagagtgaatcgaaaacaatacgaaaa
tgtaaacatttcctatacgtagtatatagagacaaaatagaagaaaccgttcataattttctgaccaatgaagaa
tcatcaacgctatcactttctgttcacaaagtatgcgcaatccacatcggtatagaatataatcggggatgcctt
tatcttgaaaaaatgcacccgcagcttcgctagtaatcagtaaacgcgggaagtggagtcaggcttttttttatgg
aagagaaaatagacaccaaagtagccttcttctaaccttaacggacctacagtgcaaaaagttatcaagagactg
cattatagagcgcacaaaggagaaaaaaagtaatctaagatgctttgttagaaaaatagcgctctcgggatgcat
ttttgtagaacaaaaaagaagtatagattctttgttggtaaaatagcgctctcgcgttgcatttctgttctgtaa
aaatgcagctcagattctttgtttgaaaaattagcgctctcgtcgcgttgcatttttgttttacaaaaatgaagc
acagattcttcgttggtaaaatagcgctttcgcgttgcatttctgttctgtaaaaatgcagctcagattctttgt
ttgaaaaattagcgctctcgcgttgcatttttgttctacaaaatgaagcacagatgcttcgttaacaaagatatg
ctattgaagtgcaagatggaaacgcagaaaatgaaccggggatgcgacgtgcaagattacctatgcaatagatgc
aatagtttctccaggaaccgaaatacatacattgtcttccgtaaagcgctagactatatattattatacaggttc
aaatatactatctgtttcagggaaaactcccaggttcggatgttcaaaattcaatgatgggtaacaagtacgatc
gtaaatctgtaaaacagtttgtcggatattaggctgtatctcctcaaagcgtattcgaatatcattgagaagctg
cagcgtcacatcggataataatgatggcagccattgtagaagtgccttttgcatttctagtctctttctcggtct
agctagttttactacatcgcgaagatagaatcttagatcacactgcctttgctgagctggatcaatagagtaaca
aaagagtggtaaggcctcgttaaaggacaaggacctgagcggaagtgtatcgtacagtagacggagtatctagta
tagtctatagtccgtggaattaattctcatctttgacagcttatcatcgataagctagcttttcaattcaattca
tcatttttttttattctttttttgatttcggtttctttgaaatttttttgattcggtaatctccgaacagaag
gaagaacgaaggaaggagcacagacttagattggtatatatacgcatatgtagtgttgaagaaacatgaaattgc
ccagtattcttaacccaactgcacagaacaaaaacctgcaggaaacgaagataaatcATGTCGAAAGCTACATAT
AAGGAACGTGCTGCTACTCATCCTAGTCCTGTTGCTGCCAAGCTATTTAATATCATGCACGAAAAGCAAACAAAC
TTGTGTGCTTCATTGGATGTTCGTACCACCAAGGAATTACTGGAGTTAGTTGAAGCATTAGGTCCCAAAATTTGT
TTACTAAAAACACATGTGGATATCTTGACTGATTTTTCCATGGAGGGCACAGTTAAGCCGCTAAAGGCATTATCC
GCCAAGTACAATTTTTTACTCTTCGAAGACAGAAAATTTGCTGACATTGGTAATACAGTCAAATTGCAGTACTCT
GCGGGTGTATACAGAATAGCAGAATGGGCAGACATTACGAATGCACACGGTGTGGTGGGCCCAGGTATTGTTAGC
GGTTTGAAGCAGGCGGCAGAAGAAGTAACAAAGGAACCTAGAGGCCTTTTGATGTTAGCAGAATTGTCATGCAAG
GGCTCCCTATCTACTGGAGAATATACTAAGGGTACTGTTGACATTGCGAAGAGCGACAAAGATTTTGTTATCGGC
TTTATTGCTCAAAGAGACATGGGTGGAAGAGATGAAGGTTACGATTGGTTGATTATGACACCCGGTGTGGGTTTA
GATGACAAGGGAGACGCATTGGGTCAACAGTATAGAACCGTGGATGATGTGGTCTCTACAGGATCTGACATTATT
ATTGTTGGAAGAGGACTATTTGCAAAGGGAAGGGATGCTAAGGTAGAGGGTGAACGTTACAGAAAAGCAGGCTGG
GAAGCATATTTGAGAAGATGCGGCCAGCAAAACTAAaaaactgtattataagtaaatgcatgtatactaaactca
caaattagagcttcaatttaattatatcagttattacccattgaaaaaggaagagtCCCGGGGCGCGC

*Sma*I *Pau*I

6.5.2.2 Chloramphenicol resistance module (C)

*Avr*II

CCTAGGgatatctggcgaaaatgagacgttgatcggcacgtaagaggttccaactttcaccataatgaaataaga
tcactaccgggcgtatttttgagttatcgagattttcaggagctaaggaagctaaaATGGAGAAAAAAATCACT
GGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGT
ACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTAT
CCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTCGTATGGCAATGAAAGACGGTGAG
CTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGG
AGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTG
GCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTT
GATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGAC
AAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGAATGCTTAAT
GAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAAtttgatatcgagctcgcttggactcctgttgat
agatccagtaatgacctcagaactccatctggatttgttcagaacgctcggttgccgccgggcgtttttttattgg
tgagaatccaagcGGGCCC

*Apa*I

6.5.2.3 tRNA modules

proK-6xtRNA$^{Tyr}_{CUA}$ (1)

*Kpn*I

GGTACCggctaactaagcggcctgctgactttctcgccgatcaaaaggcattttgctattaagggattgacgagg
gcgtatctgcgcagtaagatgcgccccgcattccggcggtagttcagcagggcagaacggcggactctaaatccg
catggcgctggttcaaatccggcccgccggaccaactactttatgtagtctccgccgtgtagcaagaaattgaga
agtccggcggtagttcagcagggcagaacggcggactctaaatccgcatggcgctggttcaaatccggcccgccg
gaccaaatttgcacggcaaatttgaagaggttttaactacatgttatccggcggtagttcagcagggcagaacgg
cggactctaaatccgcatggcgctggttcaaatccggcccgccggaccaaattcgaaaagcctgctcaacgagca
ggctttttgcatgcggctaactaagcggcctgctgactttctcgccgatcaaaaggcattttgctattaaggga
ttgacgagggcgtatctgcgcagtaagatgcgccccgcattccggcggtagttcagcagggcagaacggcggact
ctaaatccgcatggcgctggttcaaatccggcccgccggaccaactactttatgtagtctccgccgtgtagcaag
aaattgagaagtccggcggtagttcagcagggcagaacggcggactctaaatccgcatggcgctggttcaaatcc
ggcccgccggaccaaatttgcacggcaaatttgaagaggttttaactacatgttatccggcggtagttcagcagg
gcagaacggcggactctaaatccgcatggcgctggttcaaatccggcccgccggaccaaattcgaaaagcctgct
caacgagcaggctttttgcatGCGCGC

*Pau*I

proK-3xtRNA$^{Tyr}_{CUA}$ (2)

*Kpn*I
GGTACCggctaactaagcggcctgctgactttctcgccgatcaaaaggcattttgctattaagggattgacgagg
gcgtatctgcgcagtaagatgcgccccgcattccggcggtagttcagcagggcagaacggcggactctaaatccg
catggcgctggttcaaatccggcccgccggaccaactactttatgtagtctccgccgtgtagcaagaaattgaga
agtccggcggtagttcagcagggcagaacggcggactctaaatccgcatggcgctggttcaaatccggcccgccg
gaccaaatttgcacggcaaatttgaagaggttttaactacatgttatccggcggtagttcagcagggcagaacgg
cggactctaaatccgcatggcgctggttcaaatccggcccgccggaccaaattcgaaaagcctgctcaacgagca
ggctttttgcatGCGCGC
*Pau*I

lpp-tRNA$^{Phe}_{CUA_UG}$ (4)

*Kpn*I
GGTACCgcttctttgagcgaacgatcaaaaataagtggcgccccatcaaaaaaatattctcaacataaaaaactt
tgtgtaatacttgtaacgctgaattcgcggacttagctcagttgggagagcgccatactctaaatgtggaggtcc
tgtgttcgatccacagagttcgcaccactgcagatccttagcgaaagctaaggatttttttGCGCGC
*Pau*I

6.5.2.4 aaRS modules

glnS'-BpaRS (1)

*Avr*II
CCTAGGcatcaatcatccccataatccttgttagattatcaattttaaaaaactaacagttgtcagcctgtcccg
ctttaatatcatacgccgttatacgttgtttacgctttgaggaatcccatATGGACGAATTTGAAATGATAAAGA
GAAACACATCTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAAGATGAAAAATCTGCTGGTATAG
GTTTTGAACCAAGTGGTAAAATACATTTAGGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATGCTG
GATTTGATATAATTATATTGTTGGCTGATTTACACGCCTATTTAAACCAGAAAGGAGAGTTGGATGAGATTAGAA
AAATAGGAGATTATAACAAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCAAAATATCTTTATGGAAGTCCTTTCC
AGCTTGATAAGGATTATACACTGAATGTCTATAGATTGGCTTTAAAAACTACCTTAAAAAGAGCAAGAAGGAGTA
TGGAACTTATAGCAAGAGAGGATGAAAATCCAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATACGA
GTCATTATCTGGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGAGAAAAATACACATGTTAGCAAGGGAGCTTT
TACCAAAAAAGGTTGTTTGTATTCACAACCCTGTCTTAACGGGTTTGGATGGAGAAGGAAAGATGAGTTCTTCAA
AAGGGAATTTTATAGCTGTTGATGACTCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAGCATACTGCCCAGCTG
GAGTTGTTGAAGGAAATCCAATAATGGAGATAGCTAAATACTTCCTTGAATATCCTTTAACCATAAAAAGGCCAG
AAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAGTTAGAGAGTTTATTTAAAAATAAGGAATTGCATC
CAATGCGCTTAAAAAATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCAATTAGAAAGAGATTATAActgc
agtttcaaacgctaaattgcctgatgcgctacgcttatcaggcctacatgatctctgcaatatattgagtttgcg
tgcttttgtaggccggataaggcgttcacgccgcatccggcaagaaacagcaaacaatccaaaacgccgcgttca
gcggcgtttttttctgctttGGTACC
*Kpn*I

tac-scPheRS(T415G) (2)

*Avr*II

CCTAGGgttgacaattaatcatcggctcgtataatggatccaattgtgagcggaatcgattttcacacaggaaac
agaccatgaatctagagATGTCTGACTTCCAATTAGAAATTCTAAAGAAACTAGATGAATTGGATGAGATCAAGT
CCACACTGGCAACTTTCCCTCAGCACGGCTCTCAAGATGTTCTTTCCGCTTTGAACTCTTTGAAAGCCCACAACA
AGTTAGAGTTTTCCAAGGTCGACACGGTTACGTATGACTTGACCAAAGAAGGTGCTCAAATTTTGAATGAAGGTT
CGTACGAAATTAAACTAGTCAAGCTCATCCAAGAGTTGGGTCAACTTCAAATCAAAGATGTGATGTCCAAACTAG
GCCCTCAAGTTGGTAAGGTCGGTCAGGCTAGAGCTTTCAAGAACGGCTGGATCGCCAAAAACGCCTCAAACGAGC
TTGAACTCTCCGCAAAATTGCAAAATACCGATTTAAATGAGCTTACTGATGAAACGCAATCTATTCTAGCGCAAA
TCAAGAACAACTCGCATCTGGATAGCATTGACGCCAAGATTTTGAACGACTTGAAGAAAAGAAAGTTAATTGCTC
AAGGTAAAATCACAGATTTCAGTGTCACCAAAGGGCCAGAGTTCTCGACCGACCTCACCAAATTGGAAACCGATC
TTACCTCCGACATGGTCTCCACCAATGCATACAAGGACTTGAAGTTCAAGCCTTACAATTTCAATTCTCAAGGTG
TGCAAATATCTTCAGGTGCTCTTCACCCCTTAAACAAAGTCAGAGAGGAATTTAGACAAATTTTCTTTTCCATGG
GATTCACAGAGATGCCCTCGAACCAATACGTCGAGACAGGTTTCTGGAACTTCGATGCCCTTTACGTTCCACAAC
AGCATCCTGCTCGTGACCTGCAAGACACTTTCTACATCAAGGACCCACTAACCGCTGACTTGCCCGATGACAAGA
CATACATGGACAATATCAAAGCCGTTCACGAACAGGGGAGATTCGGGTCCATCGGTTATCGTTACAACTGGAAGC
CAGAAGAATGTCAAAAATTGGTCTTGAGAACTCACTCCACAGCCATCTCTGCCAGAATGCTGCACGATTTGGCCA
AAGATCCAAAGCCCACCAGATTGTTTTCTATCGACCGTGTTTTCCGTAACGAAGCAGTTGACGCCACCCATTTGG
CCGAATTCCACCAGGTGGAAGGTGTTCTTGCCGACTACAACATTACTCTGGGTGACCTGATCAAGTTCATGGAAG
AGTTTTTCGAAAGAATGGGTGTCACCGGTTTGAGATTCAAGCCTACCTACAATCCTTACGGCGAGCCATCAATGG
AAATCTTTTCTTGGCACGAAGGTTTGCAAAAATGGGTCGAAATCGGTAACTCTGGTATGTTCAGACCAGAAATGC
TCGAGTCCATGGGTCTACCAAAGGATCTAAGAGTCCTTGGTTGGGGGTTATCCTTGGAAAGACCTACCATGATCA
AATATAAGGTTCAAAACATCAGAGAACTGTTAGGTCATAAAGTCTCTTTGGACTTTATCGAAACCAATCCTGCTG
CTAGATTGGACGAAGACTTCTACGAATAAGGCAGGAATAGATTATGCCTACCGTCTCCTTGAACAAGCAGCAATT
ATTTGATCTTCTAGGCAAAGACTACACTTCCCAAGAGTTCGACGAATTATGTTTTGAATTCGGTATGGAAATGGA
CGAAGACACCACAGAAGAGGCCTTGAAAACCGGGGAGGAGCCGGAATTGAAGCTTGATATCAGTGCCAATCGTTA
CGATTTGCTTTGTATCGAAGGTATTTCACAGTCGCTGAACGAATACTTGGAACGTAAAGAAAGACCTGACTATAA
ATTAAGCAAGCCAACCACTAAGTTGATCATCGACAAATCAACGGAGCAAATTAGACCTTTTGCTACCGCTGCTGT
ATTGAGAAATATCAAGCTTAACGAAAAATCTTACGCTTCTTTTATTGCCTTGCAAGATAAATTACATGCCAATCT
ATGTAGAAACAGAAGCTTGGTTGCCATGGGTACTCACGATTTAGATTCAATTGAAGGTCCATTCCATTACAGAGC
TCTACCACCAAAGGACATCAAGTTCGTACCATTGAATCAAACCCAAGAGTTTACTGGTGACAAATTGATCGAGTT
TTATAAATCTCCAGAACAGAAAAACAACATAGGGAGATACGTTCACATTATTGAGGATTCTCCAGTCTTCCCAGT
TATTATGGACAGCAAAGATCGTGTTTGCTCCCTGCCACCATTAATCAATAGTGAACATTCGAAGATCTCTGTGAA
CACCCGTAACATTTTGATTGATATAACCGCCACCGATAAGACCAAAGCCGAGATCGTTTTGAACATATTAACTAC
AATGTTCTCACGTTATTGTGACGAACCATTCACGGTTGAGCCTGTAGAAATTGTCTCTGAACACAATGGCCAATC
CCGTTTGGCGCCAAACTTCAACGATAGAATTATGGATGTCTCCATTAAGTACATCAACTCCTGTCTTGGCCTAGA
TCAATCCGCTGATGAAATTGCTCATTGTCTAAAAAAGATGTCGTTGCATGCCGTTCAATCAAAGGAAGACAAGGA
CATCTTGCACGTTGACATTCCGGTAACTAGACCTGATATTTTGCACGCTTGTGATATAATGGAAGATGCCGCTGT
CGGTTATGGTTTCAATAATTTGCCAAAGGGTGAGAAATTATCCAATGCCAACTTCATTGCCAAACCATTACCAAT
CAACAAGGTTTCTGATATTTTCAGAGTTGCATCCTCTCAAGCCACGTGGGTTGAGGTTTTACCATTGACCTTATG
TTCGCACGATGAAAACTTTAAATTTCTAAGACAATCCGACAATGATGATTTAGCTGTCAAATTGGCCAACCCAAA
GACTTTGGAATACCAAGTTGTTAGAACCACTTTATTGCCTGGTATCTTAAAGACAGTCAAGGAAAACAGAAAACA
TTCCTTACCAATCAAAGTCTTTGAAACCGGTGACGTTGTATTTAAAGACGACAAACTAGAAAGGAAGGCGTACAA

TGAACGTCACTGGGCTGCCATCTACGTGGGTAAGAATTCTGGGTTTGAAATCATTCAAGGGTTATTGGGTAAAAT
CATGCAAACTTTTAGAACAGAGTGGATTGCAGACTACGGTGCTGCTGCTTCTGGCAGAGGTTACTGGATTGAAGA
AGACGATTCTGTGAAAACCTATTTCCCAGGTAGAGGTGCCAAGGTCATGTTCAGATCCAAAGAAGGCGCTGAGCC
AAAGCAAATCGGCCACTTGGGTGTCTTGCATCCTGAAGTCATGATGAATTTCGACGTTCCATTCGCTGCATCCTT
TGTAGAGGTTAATGCCGAAGTCTTCCTATAAaaaaaaaaaccccgcccctgacagggcggggttttttttGGTAC
C *KpnI*

prpR-STREP (3, 4 and 7)

*Avr*II
CCTAGGcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgc
acctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatctgctcatgtttgacagcttatcatcg
atTCAGCTTTTCAGCCGCCGCCAGAACGTCGTCCGGCTGATGCCTAAATAATTCGCCGCTGCTGTTTTATCGCCA
TTAAATTTCTCCAGTGCCTGTTGTGGTGTCAGTAAGCGTGGAGCGGGAGTTTTCGCCGACTCACGCGCCAGTTCC
GGCAGTAGCAGCTGCAAAAATTGCGGCGTTAAATCCGGCGTCGGTTCCACACTTAAAAATAGCGCCAGTCGCTCC
ATCATATTGCGCAGTTCACGAATATTGCCCGGCCAGTCGTAGTGCACCAGCACGGTTTCGCTTGCCTGTAATCCC
TGGCGTAATGCGGCAGAAAACGGGGTGGAGAGCGCCGCCAGAGACACTTTCAAAAAGCTTTCCGCCAGTGGCAGA
ATATCCGCCACCCGCTCGCGCAGTGGTGGCAATTGCAGACGCAAAATACTCAGCCGATAAAACAGGTCACGGCGA
AACTGCCCTTGCCGCATATCTTCTTCCAGATTGCAGTGAGTGGCGCTAATGACCCGCACATCCACCGGAACAGGC
TGATGCCCGCCGACGCGGGTGACCTCTTTTTCTTCCAGCACCCGCAGCAGCCGGGTCTGCAACGGCAGCGGCATT
TCGCCAATCTCATCGAGAAACAGCGTACCTCCGTGGGCGATTTCAAACAGCCCGGCGCGACCGCCGCGTCGCGAG
CCGGTAAACGCCCCTTCCTCATAGCCAAACAGTTCTGCTTCCAGCAGCGATTCGGCAATCGCCCCGCAGTTGACT
GCAACAAACGGATGCGACTTTTTGCCCTGTCGCGCATCGTGGCGGGCAAAATATTCCCGATGAATCGCCTGGGCC
GCCAGCTCTTTGCCCGTCCCCGTTTCCCCCTCAATCAACACCGCTGCACTGGAGCGGGCATACAGCAAAATAGTC
TGCCGTACTTGTTCCATCTGTGGTGATTGACCGAGCATATCGCCCAGCACGTAACGAGTTCTCAGGGCGTTGCGG
GTGGCATCGTGAGTGTTATGGCGTAACGACATGCGCGTCATATCCAGCGCATCGCTGAACGCCTGGCGCACGGTG
GCGGCGGAATAGATAAAAATTCCGGTCATTCCGGCTTCTTCTGCCAGATCGGTAATCAGCCCCGCGCCGACCACC
GCTTCGGTGCCGTTAGCTTTTAGCTCGTTAATCTGCCCGCGTGCGTCTTCTTCGGTAATGTAGCTACGTTGATCG
AGGCGCAAATTAAAGGTTTTTTGAAACGCCACCAGTGCCGGAATAGTTTCCTGATAAGTGACAACGCCGATAGAA
GAGGTGAGTTTTCCGGCTTTTGCCAGTGCCTGTAACACATCGTAGCCGCTCGGTTTAATCAAAATAACTGGCACT
GACAGGCGGCTTTTCAGGTACGCGCCGTTAGATCCAGCGGCGATGATGGCGTCACAGCGTTCGTTTGCCAGTTTC
TTGTGGATGTAGGTCACCGCTTTTTCAAAGCCAAGCTGGATAGGGGTAATGTTCGCCAGGTGATCAAACTCGAGG
CTGATATCGCGAAACAGCTCGAACAGGCGCGTTACAGATACCGTCCAGATAACCGGTTTGTCGTCATTAAGCCGT
GGTGGATGTGCCATagcgcaccgcaaagttaagaaaccgaatattgggtttagtcttgtttcataattgttgcaa
tgaaacgcggtgaaacattgcctgaaacgttaactgaaacgcatatttgcggattagttcatgactttatctcta
acaaattgaaattaaacatttaattttattaaggcaattgtggcacacccttgctttgtctttatcaacgcaaa
taacaagttgataacaagctagcaggaggaattcaccATGAGCGCTTGGAGCCACCCGCAGTTCGAAAAAGGCGC
CATG
Start ATG synthetase

araC-STREP (6 and 8)

*Avr*II
CCTAGGgcataatgtgcctgtcaaatggacgaagcagggattctgcaaaccctatgctactccgtcaagccgtca
attgtctgattcgttaccaaTTATGACAACTTGACGGCTACATCATTCACTTTTTCTTCACAACCGGCACGGAAC
TCGCTCGGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCAAAACCAACATTGCGA
CCGACGGTGGCGATAGGCATCCGGGTGGTGCTCAAAAGCAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAG
CTTAAGACGCTAATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGACGGCGACAAGCAAACATGCTGTGCG
ACGCTGGCGATATCAAAATTGCTGTCTGCCAGGTGATCGCTGATGTACTGACAAGCCTCGCGTACCCGATTATCC
ATCGGTGGATGGAGCGACTCGTTAATCGCTTCCATGCGCCGCAGTAACAATTGCTCAAGCAGATTTATCGCCAGC
AGCTCCGAATAGCGCCCTTCCCCTTGCCCGGCGTTAATGATTTGCCCAAACAGGTCGCTGAAATGCGGCTGGTGC
GCTTCATCCGGGCGAAAGAACCCCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTCATGCCAGTAGGCGCGC
GGACGAAAGTAAACCCACTGGTGATACCATTCGCGAGCCTCCGGATGACGACCGTAGTGATGAATCTCTCCTGGC
GGGAACAGCAAAATATCACCCGGTCGGCAAACAAATTCTCGTCCCTGATTTTTCACCACCCCCTGACCGCGAATG
GTGAGATTGAGAATATAACCTTTCATTCCCAGCGGTCGGTCGATAAAAAAATCGAGATAACCGTTGGCCTCAATC
GGCGTTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGATCATTTTGCGCTTCAGCCATA
CTTTTCATactcccgccattcagagaagaaaccaattgtccatattgcatcagacattgccgtcactgcgtcttt
tactggctcttctcgctaaccaaaccggtaaccccgcttattaaaagcattctgtaacaaagcgggaccaaagcc
atgacaaaaacgcgtaacaaaagtgtctataatcacggcagaaaagtccacattgattatttgcacggcgtcaca
ctttgctatgccatagcatttttatccataagattagcggatcctacctgacgctttttatcgcaactctctact
gtttctccatacccgtttttttgggctaacaggaggaattaaccATGGCTAGCTGGAGCCACCCGCAGTTCGAAA
AAGGCGCCATG
Start ATG synthetase

6.6 Model protein sequences

6.6.1 EGFP

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACG
GCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCA
CCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCT
ACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCT
TCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCG
AGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACA
ACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACG
GCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACC
ACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCG
TGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGCATCACCATCACCATCACTAA

6.6.2 TTL

ATGCAAAAGGCTGTTGAAATTACATATAACGGCAAAACTTTAAGAGGAATGATGCATTTGCCTGATGATGTTA
AGGGTAAAGTGCCTATGGTAATAATGTTTCACGGTTTTACAGGCAATAAAGTAGAGTCTCACTTTATTTTTGTGA
AGATGTCAAGAGCTTTAGAAAAAGTAGGTATTGGGAGTGTAAGGTTTGACTTTTATGGTTCTGGAGAAAGTGATG
GGGACTTTAGTGAAATGACATTTAGCAGTGAATTGGAAGATGCAAGACAAATTTTAAAGTTTGTGAAAGAGCAAC
CTACGACTGACCCTGAGAGAATAGGACTACTTGGTTTGAGTATGGGAGGAGCTATTGCAGGGATTGTAGCAAGGG
AATATAAAGATGAAATAAAGGCGTTGGTGCTATGGGCTCCAGCTTTTAATATGCCTGAGCTTATAATGAACGAAA
GTGTAAAGCAATACGGAGCTATTATGGAACAATTGGGCTTTGTAGACATAGGAGGACATAAACTGAGTAAAGATT
TTGTTGAGGATATTTCAAAATTAAATATATTTGAGCTGTCAAAAGGATACGATAAAAAAGTGCTTATAGTTCATG
GGACAAATGATGAAGCGGTTGAATATAAAGTTTCTGATAGAATCTTAAAAGAGGTTTATGGGGATAACGCTACAA
GAGTGACAATCGAAAATGCAGACCATACTTTTAAGAGTTTAGAATGGGAGAAAAAGGCGATTGAGGAGTCAGTAG
AGTTTTTCAAAAAGGAATTGTTAAAGGGAGGATCCCATCACCATCACCATCACTGA

6.6.3 ψ-b*

ATGAAAAAAGCAGTCATTAACGGGGAACAAATCAGAAGTATCAGCGACCTCCACCAGACATTGAAAAAGGAGC
TTGCCCTTGCGGAATACTACGGTGAAAACCTGGACGCTTTATGGGATGCTCTGACCGGATGGGTGGAGTACCCGC
TCGTTTTGGAATGGAGGCAGTTTGAACAAAGCAAGCAGCTGACTGAAAATGGCGCCGAGAGTGTGCTTCAGGTTT
TCCGTGAAGCGAAAGCGGAAGGCGCCGACATCACCATCATACTTTCTTAA

6.6.4 *m*DHFR

ATGAGAGGATCCGGCATCATGGTTCGACCATTGAACTCGATCGTCGCCGTGTCCCAAAATATGGGGATTGGCA
AGAACGGAGACCTACCCTGGCCTCCGCTCAGGAACGAGTTTAAGTACTTCCAAAGAATGACCACAACCTCTTCAG
TGGAAGGTAAACAGAATCTGGTGATTATGGGTAGGAAAACCTGGTTCTCCATTCCTGAGAAGAATCGACCTTTAA
AGGACAGAATTAATATAGTTCTCAGTAGAGAACTCAAAGAACCACCACGAGGAGCTCATTTTCTTGCCAAAAGTT
TGGATGATGCCTTAAGACTTATTGAACAACCGGAATTGGCAAGTAAAGTAGACATGGTTTGGATAGTCGGAGGCA
GTTCTGTTTACCAGGAAGCCATGAATCAACCAGGCCACCTTAGACTCTTTGTGACAAGGATCATGCAGGAATTTG
AAAGTGACACGTTTTTCCCAGAAATTGATTTGGGGAAATATAAACTTCTCCCAGAATACCCAGGCGTCCTCTCTG
AGGTCCAGGAGGAAAAAGGCATCAAGTATAAGTTTGAAGTCTACGAGAAGAAAGGTTCCAGATCTCATCACCATC
ACCATCACTAA

7 Literature

[1] Miescher, F. Über die chemische Zusammensetzung der Eiterzellen. *Hoppe-Seyler's medicinisch-chemische Untersuchungen* **1871**, *4*, 441-460.

[2] Avery, OT, MacLeod, CM, McCarty, M. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III. *J. Exp. Med.* **1944**, *79*, 137-158.

[3] Chargaff, E. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia* **1950**, *6*, 201-209.

[4] Watson, JD, Crick, FHC. Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737-738.

[5] Mulder, GJ. Sur la composition des quelques substance animales. *Bulletin des Sciences physiques et naturelles en Néerlande* **1838**, *1*, 104.

[6] Hartley, H. Origin of the Word 'Protein'. *Nature* **1951**, *168*, 244-244.

[7] Tanford, C, Reynolds, J. *Nature's Robots: A History of Proteins*, Oxford University Press, Oxford **2001**.

[8] Sumner, JB. The Isolation and Crystallization of the Enzyme Urease. Preliminary Paper. *J. Biol. Chem* **1926**, *69*, 435–441.

[9] Crick, FHC. On protein synthesis. *Symp. Soc. Exp. Biol.* **1957**, *12*, 138-163.

[10] Crick, FHS, Brenner, S, Wats-Tobin, RJ, Barnett, L. General nature of the genetic code for proteins. *Nature* **1961**, *192*, 1227-1232.

[11] Nirenberg, M, Matthaei, JH. Dependence of Cell-free Protein Synthesis in E. ColiI upon Naturally Occuring or Synthetic Polynucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **1961**, *47*, 1588-1602.

[12] Nirenberg, MW, Leder, P. RNA Codewords and Protein Synthesis - The Effect of Trinucleotides upon the Binding of sRNA to Ribosomes. *Science* **1964**, *145*, 1399-1407.

[13] Moffatt, JG, Khorana, HG. Nucleoside Polyphosphates. X. The Synthesis and Some Reactions of Nucleoside-5' Phosphoromorpholidates and Related Compounds. Improved Methods for the Preparation of Nucleoside-5' Polyphosphates. *J. Am. Chem. Soc.* **1961**, *83*, 649-658.

[14] Freeland, SJ, Wu, T, Keulmann, N. The case for an error minimizing standard genetic code. *Orig. Life Evol. Biosph.* **2003**, *33*, 457-477.

[15] Koonin, EV, Novozhilov, AS. Origin and Evolution of the Genetic Code: The Universal Enigma. *IUBMB Life* **2009**, *61*, 99-111.

[16] Crick, FHC. The origin of the genetic code. *J. Mol. Biol.* **1968**, *38*, 367-379.

[17] Wong, JTF. Coevolution theory of the genetic code at age thirty. *Bioessays* **2005**, *27*, 416-425.

[18] Wong, JTF. Co-evolution Theory of Genetic Code. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 1909-1912.

[19] Lu, Y, Freeland, S. On the evolution of the standard amino-acid alphabet. *Genome Biol.* **2006**, *7*.

[20] Walsh, CT. *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, Roberts and Company Publishers, Greenwood Village, **2005**.

[21] Böck, A, Forchhammer, K, Heider, J, Leinfelder, W, Sawers, G, Veprek, B, Zinoni, F. Selenocysteine - the 21st Amino-Acid. *Mol. Microbiol.* **1991**, *5*, 515-520.

[22] Srinivasan, G, James, CM, Krzycki, JA. Pyrrolysine encoded by UAG in Archaea: Charging of a UAG-decoding specialized tRNA. *Science* **2002**, *296*, 1459-1462.

[23] Zhang, Y, Baranov, PV, Atkins, JF, Gladyshev, VN. Pyrrolysine and selenocysteine use dissimilar decoding strategies. *J. Biol. Chem.* **2005**, *280*, 20740-20751.

[24] Knight, RD, Freeland, SJ, Landweber, LF. Rewiring the keyboard evolvability of the genetic code. *Nat. Rev. Genet.* **2001**, *2*, 49-58.

[25] Santos, MAS, Moura, G, Massey, SE, Tuite, MF. Driving change: the evolution of alternative genetic codes. *Trends Genet.* **2004**, *20*, 95-102.

[26] Schultz, DW, Yarus, M. Transfer RNA Mutation and the Malleability of the Genetic Code *J. Mol. Biol.* **1994**, *235*, 1377-1380.

[27] Osawa, S, Jukes, TH, Watanabe, K, Muto, A. Recent evidence for evolution of the genetic code. *Microbiol. Rev.* **1992**, *56*, 229-264.

[28] Nilsson, M, Ryden-Aulin, M. Glutamine is incorporated at the nonsense codons UAG and UAA in a suppressor-free Escherichia coli strain. *Biochim. Biophys. Acta* **2003**, *1627*, 1-6.

[29] Doi, Y, Ohtsuki, T, Shimizu, Y, Ueda, T, Sisido, M. Elongation Factor Tu Mutants Expand Amino Acid Tolerance of Protein Biosynthesis System. *J. Am. Chem. Soc.* **2007**, *129*, 14458-14462.
[30] Voet, D, Voet, JG, Pratt, CW. *Lehrbuch der Biochemie*, Wiley-VCH Verlag, Weinheim, **2002**.
[31] Hausmann, CD, Ibba, M. Aminoacyl-tRNA synthetase complexes: molecular multitasking revealed. *FEMS Microbiol. Rev.* **2008**, *32*, 705-721.
[32] Ibba, M, Soll, D. Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **2000**, *69*, 617-650.
[33] Meinnel, T, Mechulam, Y, Blanquet, S. *Aminoacyl-tRNA Synthetases: Occurrence, Structure, and Funcion,* in *tRNA: Structure, Biosynthesis, and Function* (Eds.: D. Söll, U. RajBhandary), American Society of Microbiology, Washington, **1995**.
[34] Chapeville, F, Ehrenstein, GV, Benzer, S, Weisblum, B, Ray, WJ, Lipmann, F. On the Role of Soluble Ribonucleic Acid in Coding for Amino Acids. *Proc. Natl. Acad. Sci. U.S.A.* **1962**, *48*, 1086-1092.
[35] Dubois, D, Lapointe, J, Sekine, S. *Glutamyl-tRNA Synthetases,* in *The Aminoacyl-tRNA Synthetases* (Eds.: M. Ibba, C. Francklyn, S. Cusack), Landes Bioscience, Georgetown, **2005**.
[36] Ibba, M, Becker, HD, Stathopoulos, C, Tumbula, DL, Söll, D. The Adaptor hypothesis revisited. *Trends Biochem. Sci.* **2000**, *25*, 311-316.
[37] Jakubowski, H, Goldman, E. Editing of errors in selection of amino acids for protein synthesis. *Microbiol. Rev.* **1992**, *56*, 412-429.
[38] Ling, JQ, Reynolds, N, Ibba, M. Aminoacyl-tRNA Synthesis and Translational Quality Control. *Annu. Rev. Microbiol.* **2009**, *63*, 61-78.
[39] Dirheimer, G, Keith, G, Dumas, P, Westhof, E. *Primary, Secondary, and Tertiary Structures of tRNAs,* in *Trna: Structure, Biosynthesis, and Function* (Eds.: D. Söll, U. RajBhandary), American Society of Microbiology, Washington, **1995**.
[40] Czerwoniec, A, Dunin-Horkawicz, S, Purta, E, Kaminska, KH, Kasprzak, JM, Bujnicki, JM, Grosjean, H, Rother, K. MODOMICS: a database of RNA modification pathways. 2008 update. *Nucl. Acids Res.* **2009**, *37*, D118-D121.
[41] Giege, R, Sissler, M, Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucl. Acids Res.* **1998**, *26*, 5017-5035.
[42] Crick, FHC. Codon-Anticodon Pairing - Wobble Hypothesis. *J. Mol. Biol.* **1966**, *19*, 548-555.
[43] Agris, PF. Wobble position modified nucleosides evolved to select transfer RNA codon recognition: A modified-wobble hypothesis. *Biochimie* **1991**, *73*, 1345-1349.
[44] Cochella, L, Green, R. Wobble during decoding: more than third-position promiscuity. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1160-1162.
[45] Sakamoto, K, Kawai, G, Watanabe, S, Niimi, T, Hayashi, N, Muto, Y, Watanabe, K, Satoh, T, Sekine, M, Yokoyama, S. NMR Studies of the Effects of the 5‘-Phosphate Group on Conformational Properties of 5-Methylaminomethyluridine Found in the First Position of the Anticodon of Escherichia coli tRNA4Arg†. *Biochemistry* **1996**, *35*, 6533-6538.
[46] Yarus, M. Translational efficiency of transfer RNA's: uses of an extended anticodon. *Science* **1982**, *218*, 646-652.
[47] Agris, PF, Vendeix, FAP, Graham, WD. tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.* **2007**, *366*, 1-13.
[48] Alfonzo, JD, Blanc, V, Estevez, AM, Rubio, MAT, Simpson, L. C to U editing of the anticodon of imported mitochondrial tRNA(Trp) allows decoding of the UGA stop codon in Leishmania tarentolae. *EMBO J.* **1999**, *18*, 7056-7062.
[49] Dong, HJ, Nilsson, L, Kurland, CG. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J. Mol. Biol.* **1996**, *260*, 649-663.
[50] Hayashi, K, Morooka, N, Yamamoto, Y, Fujita, K, Isono, K, Choi, S, Ohtsubo, E, Baba, T, Wanner, BL, Mori, H, Horiuchi, T. Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110. *Mol. Syst. Biol.* **2006**, *2*, 2006.0007.
[51] Berg, OG, Kurland, CG. Growth rate-optimised tRNA abundance and codon usage. *J. Mol. Biol.* **1997**, *270*, 544-550.
[52] Zhang, G, Hubalewska, M, Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* **2009**, *16*, 274-280.
[53] Giege, R. Toward a more complete view of tRNA biology. *Nat. Struct. Mol. Biol.* **2008**, *15*, 1007-1014.

[54] Fechter, P, Rudinger-Thirion, J, Tukalo, M, Giege, R. Major tyrosine identity determinants in Methanococcus jannaschii and Saccharomyces cerevisiae tRNA(Tyr) conserved but expressed differently. *Eur. J. Biochem.* **2001**, *268*, 761-767.

[55] Kobayashi, T, Nureki, O, Ishitani, R, Yaremchuk, A, Tukalo, M, Cusack, S, Sakamoto, K, Yokoyama, S. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Biol.* **2003**, *10*, 425-432.

[56] Furter, R. Expansion of the genetic code: Site-directed p-fluoro-phenylalanine incorporation in Escherichia coli. *Protein Sci.* **1998**, *7*, 419-426.

[57] Borel, F, Vincent, C, Leberman, R, Hartlein, M. Seryl-tRNA synthetase from Escherichia coli: implication of its N-terminal domain in aminoacylation activity and specificity. *Nucl. Acids Res.* **1994**, *22*, 2963-2969.

[58] Bonnefond, L, Giege, R, Rudinger-Thirion, J. Evolution of the tRNA(Tyr)/TyrRS aminoacylation systems. *Biochimie* **2005**, *87*, 873-883.

[59] Mikuni, O, Ito, K, Moffat, J, Matsumura, K, McCaughan, K, Nobukuni, T, Tate, W, Nakamura, Y. Identification of the prfC gene, which encodes peptide-chain-release factor 3 of Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 5798-5802.

[60] Mottagui-Tabar, S, Isaksson, LA. Only the last amino acids in the nascent peptide influence translation termination in Escherichia coli genes. *FEBS Lett.* **1997**, *414*, 165-170.

[61] Herrington, MB. *Nonsense Mutations and Suppression*, John Wiley & Sons, Ltd, **2001**.

[62] Murgola, E. *Translational Suppression: When Two Wrongs DO Make a Right,* in *tRNA: Structure, Biosynthesis, and Function* (Eds.: D. Söll, U. RajBhandary), American Society of Microbiology, Washington, **1995**.

[63] Buckingham, RH. Codon context. *Cell. Mol. Life Sci.* **1990**, *46*, 1126-1133.

[64] Coomber, DWJ. *Panning of Antibody Phage-Display Libraries,* in *Antibody Phage Display, Vol. 178* (Eds.: P. M. O'Brien, R. Aitken), Humana Press, **2002**, pp. 133-145.

[65] Smith, D, Yarus, M. tRNA-tRNA interactions within cellular ribosomes. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 4397-4401.

[66] Mottagui-Tabar, S, Isaksson, LA. Influence of the last amino acid in the nascent peptide on EF-Tu during decoding. *Biochimie* **1996**, *78*, 953-958.

[67] Bjornsson, A, Mottagui-Tabar, S, Isaksson, LA. Structure of the C-terminal end of the nascent peptide influences translation termination. *EMBO J.* **1996**, *15*, 1696-1704.

[68] Mottagui-Tabar, S, Bjornsson, A, Isaksson, LA. The second to last amino acid in the nascent peptide as a codon context determinant. *EMBO J.* **1994**, *13*, 249-257.

[69] Boycheva, S, Chkodrov, G, Ivanov, I. Codon pairs in the genome of Escherichia coli. *Bioinformatics* **2003**, *19*, 987-998.

[70] Irwin, B, Heck, JD, Hatfield, GW. Codon Pair Utilization Biases Influence Translational Elongation Step Times. *J. Biol. Chem.* **1995**, *270*, 22801-22806.

[71] Curran, JF, Poole, ES, Tate, WP, Gross, BL. Selection of aminoacyl-tRNAs at sense codons: the size of the tRNA variable loop determines whether the immediate 3' nucleotide to the codon has a context effect. *Nucl. Acids Res.* **1995**, *23*, 4104-4108.

[72] Buchan, JR, Aucott, LS, Stansfield, I. tRNA properties help shape codon pair preferences in open reading frames. *Nucl. Acids Res.* **2006**, *34*, 1015-1027.

[73] Poole, ES, Brown, CM, Tate, WP. The identity of the base following the stop codon determines the efficiency of in vivo translational termination in Escherichia coli. *EMBO J.* **1995**, *14*, 151-158.

[74] Tats, A, Tenson, T, Remm, M. Preferred and avoided codon pairs in three domains of life. *BMC Genomics* **2008**, *9*, 463.

[75] Miller, JH, Albertini, AM. Effects of surrounding sequence on the suppression of nonsense codons. *J. Mol. Biol.* **1983**, *164*, 59-71.

[76] Bossi, L, Roth, JR. The influence of codon context on genetic code translation. *Nature* **1980**, *286*, 123-127.

[77] Kopelowitz, J, Hampe, C, Goldman, R, Reches, M, Engelbergkulka, H. Influence of codon context on UGA suppression and readthrough. *J. Mol. Biol.* **1992**, *225*, 261-269.

[78] Pedersen, WT, Curran, JF. Effects of the nucleotide 3' to an amber codon on ribosomal selection rates of suppressor tRNA and release factor-1. *J. Mol. Biol.* **1991**, *219*, 231-241.

[79] Stormo, GD, Schneider, TD, Gold, L. Quantitative analysis of the relationship between nucleotide sequence and functional activity. *Nucl. Acids Res.* **1986**, *14*, 6661-6679.

[80] Bain, JD, Glabe, CG, Dix, TA, Chamberlin, AR, Diala, ES. Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J. Am. Chem. Soc.* **1989**, *111*, 8013-8014.

[81] Noren, CJ, Anthony-Cahill, SJ, Griffith, MC, Schultz, PG. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **1989**, *244*, 182-188.

[82] Heckler, TG, Chang, LH, Zama, Y, Naka, T, Chorghade, MS, Hecht, SM. T4 RNA ligase mediated preparation of novel "chemically misacylated" tRNAPhes. *Biochemistry* **1984**, *23*, 1468-1473.

[83] Kajihara, D, Hohsaka, T, Sisido, M. Synthesis and sequence optimization of GFP mutants containing aromatic non-natural amino acids at the Tyr66 position. *Protein Eng. Des. Sel.* **2005**, *18*, 273-278.

[84] Kajihara, D, Abe, R, Iijima, I, Komiyama, C, Sisido, M, Hohsaka, T. FRET analysis of protein conformational change through position-specific incorporation of fluorescent amino acids. *Nat. Methods* **2006**, *3*, 923-929.

[85] Hohsaka, T, Ashizuka, Y, Murakami, H, Sisido, M. Incorporation of nonnatural amino acids into streptavidin through in vitro frame-shift suppression. *J. Am. Chem. Soc.* **1996**, *118*, 9778-9779.

[86] Hughes, RA. PhD thesis, University of Texas (Austin), **2008**.

[87] England, PM. Unnatural Amino Acid Mutagenesis: A Precise Tool for Probing Protein Structure and Function. *Biochemistry* **2004**, *43*, 11623-11629.

[88] Dougherty, DA. Unnatural amino acids as probes of protein structure and function. *Curr. Opin. Chem. Biol.* **2000**, *4*, 645-652.

[89] Tan, Z, Blacklow, SC, Cornish, VW, Forster, AC. De novo genetic codes and pure translation display. *Methods* **2005**, *36*, 279-290.

[90] Hartman, MCT, Josephson, K, Lin, CW, Szostak, JW. An Expanded Set of Amino Acid Analogs for the Ribosomal Translation of Unnatural Peptides. *PLoS ONE* **2007**, *2*.

[91] Hartman, MCT, Josephson, K, Szostak, JW. Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4356-4361.

[92] Budisa, N, Steipe, B, Demange, P, Eckerskorn, C, Kellermann, J, Huber, R. High-Level Biosynthetic Substitution of Methionine in Proteins by Its Analogs 2-Aminohexanoic Acid, Selenomethionine, Telluromethionine and Ethionine in Escherichia-Coli. *Eur. J. Biochem.* **1995**, *230*, 788-796.

[93] Yoshikawa, E, Fournier, MJ, Mason, TL, Tirrell, DA. Genetically Engineered Fluoropolymers. Synthesis of Repetitive Polypeptides Containing p-Fluorophenylalanine Residues. *Macromolecules* **1994**, *27*, 5471-5475.

[94] Cowie, DB, Cohen, GN. Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur. *Biochim. Biophys. Acta* **1957**, *26*, 252-261.

[95] Schlesinger, S, Schlesinger, M. The Effect of Amino Acid Analogues on Alkaline Phosphatase Formation in Escherichia coli K-12. 1. Substitution of Triazolealanine for Histidine. *J. Biol. Chem.* **1967**, *242*, 3369-3372.

[96] Hoesl, MG, Merkel, L, Budisa, N. *Synthetic biology of autofluorescent proteins,* in *Fluorescent Proteins – from Fundamental Research to Bioanalytics* (Ed.: G. Jung), Springer Verlag, Heidelberg, **2011**, p. [in press].

[97] Lepthien, S, Hoesl, MG, Merkel, L, Budisa, N. Azatryptophans endow proteins with intrinsic blue fluorescence. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 16095-16100.

[98] Steiner, T, Hess, P, Bae, JH, Wiltschi, B, Moroder, L, Budisa, N. Synthetic Biology of Proteins: Tuning GFPs Folding and Stability with Fluoroproline. *PLoS ONE* **2008**, *3*, Article No.: e1680.

[99] Hoesl, MG, Acevedo-Rocha, CG, Nehring, S, Royter, M, Wolschner, C, Wiltschi, B, Budisa, N, Antranikian, G. Lipase Congeners Designed by Genetic Code Engineering. *ChemCatChem* **2011**, *3*, 213-221.

[100] Nehring, S, Hoesl, MG, Acevedo-Rocha, CG, Royter, M, Wolschner, C, Wiltschi, B, Budisa, N, Antranikian, G. Effects of additives on the activity of TTL congeners.**2011**, [in preparation].

[101] Wolschner, C, Giese, A, Kretzschmar, HA, Huber, R, Moroder, L, Budisa, N. Design of anti- and pro-aggregation variants to assess the effects of methionine oxidation in human prion protein. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 7756-7761.

[102] Hoesl, MG, Larregola, M, Cui, H, Budisa, N. Azatryptophans as tools to study polarity requirements for folding of green fluorescent protein. *J. Pept. Sci.* **2010**, *16*, 589-595.

[103] Kiick, KL, van Hest, JCM, Tirrell, DA. Expanding the scope of protein biosynthesis by altering the methionyl-tRNA synthetase activity of a bacterial expression host. *Angew. Chem. Int. Ed.* **2000**, *39*, 2148-2152.

[104] Tang, Y, Tirrell, DA. Attenuation of the editing activity of the Escherichia coli Leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. *Biochemistry* **2002**, *41*, 10635-10645.

[105] Tang, Y, Wang, P, Van Deventer, JA, Link, AJ, Tirrell, DA. Introduction of an Aliphatic Ketone into Recombinant Proteins in a Bacterial Strain that Overexpresses an Editing-Impaired Leucyl-tRNA Synthetase. *ChemBioChem* **2009**, *10*, 2188-2190.

[106] Sharma, N, Furter, R, Kast, P, Tirrell, DA. Efficient introduction of aryl bromide functionality into proteins in vivo. *FEBS Lett.* **2000**, *467*, 37-40.

[107] Tanrikulu, IC, Schmitt, E, Mechulam, Y, Goddard, WA, Tirrell, DA. Discovery of Escherichia coli methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15285-15290.

[108] Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, WH Freeman, New York, **1999**.

[109] Shoulders, MD, Satyshur, KA, Forest, KT, Raines, RT. Stereoelectronic and steric effects in side chains preorganize a protein main chain. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 559-564.

[110] Moroder, L, Budisa, N. Synthetic Biology of Protein Folding. *ChemPhysChem* **2010**, *11*, 1181-1187.

[111] Hoesl, MG, Budisa, N. Nicht-kanonische Aminosäuren in der Synthetischen Biologie. *BIOspektrum* **2010**, *16. Jahrgang*, 309-311.

[112] Acevedo-Rocha, CG. PhD thesis, Technische Universität München (Munich), **2010**.

[113] Turner, NA, Needs, EC, Khan, JA, Vulfson, EN. Analysis of conformational states of Candida rugosa lipase in solution: Implications for mechanism of interfacial activation and separation of open and closed forms. *Biotechnol. Bioeng.* **2001**, *72*, 108-118.

[114] Lepthien, S, Merkel, L, Budisa, N. In Vivo Double and Triple Labeling of Proteins Using Synthetic Amino Acids. *Angew. Chem. Int. Ed.* **2010**, *49*, 5446–5450.

[115] Nolting, B, Golbik, R, Fersht, AR. Submillisecond Events in Protein-Folding. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10668-10672.

[116] Tornøe, CW, Christensen, C, Meldal, M. Peptidotriazoles on solid phase: 1,2,3 -triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **2002**, *67*, 3057-3064.

[117] Rostovtsev, VV, Green, LG, Fokin, VV, Sharpless, KB. A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596-2599.

[118] Merkel, L, Beckmann, HSG, Wittmann, V, Budisa, N. Efficient N-terminal glycoconjugation of proteins by the N-end rule. *ChemBioChem* **2008**, *9*, 1220-1224.

[119] Renner, C, Alefelder, S, Bae, JH, Budisa, N, Huber, R, Moroder, L. Fluoroprolines as Tools for Protein Design and Engineering. *Angew. Chem. Int. Ed.* **2001**, *40*, 923-925.

[120] Van Kasteren, SI, Kramer, HB, Gamblin, DP, Davis, BG. Site-selective glycosylation of proteins: creating synthetic glycoproteins. *Nat. Protoc.* **2007**, *2*, 3185-3194.

[121] Liu, DR, Magliery, TJ, Schultz, PG. Characterization of an 'orthogonal' suppressor tRNA derived from E-coli tRNA(2)(Gln). *Chem. Biol.* **1997**, *4*, 685-691.

[122] Liu, DR, Magliery, TJ, Pasternak, M, Schultz, PG. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10092-10097.

[123] Liu, DR, Schultz, PG. Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4780-4785.

[124] Short, GF, Golovine, SY, Hecht, SM. Effects of release factor 1 on in vitro protein translation and the elaboration of proteins containing unnatural amino acids. *Biochemistry* **1999**, *38*, 8808-8819.

[125] Kwon, I, Wang, P, Tirrell, DA. Design of a bacterial host for site-specific incorporation of p-bromophenylalanine into recombinant proteins. *J. Am. Chem. Soc.* **2006**, *128*, 11778-11783.

[126] Ohno, S, Yokogawa, T, Fujii, I, Asahara, H, Inokuchi, H, Nishikawa, K. Co-expression of yeast amber suppressor tRNA(Tyr) and tyrosyl-tRNA synthetase in Escherichia coli: Possibility to expand the genetic code. *J. Biochem.* **1998**, *124*, 1065-1068.

[127] Pastrnak, M, Magliery, TJ, Schultz, PG. A new orthogonal suppressor tRNA/aminoacyl-tRNA synthetase pair for evolving an organism with an expanded genetic code. *Helv. Chim. Acta* **2000**, *83*, 2277-2286.

[128] Wang, L, Magliery, TJ, Liu, DR, Schultz, PG. A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins. *J. Am. Chem. Soc.* **2000**, *122*, 5010-5011.

[129] Kowal, AK, Kohrer, C, RajBhandary, UL. Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2268-2273.

[130] Anderson, JC, Schultz, PG. Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. *Biochemistry* **2003**, *42*, 9598-9608.

[131] Santoro, SW, Anderson, JC, Lakshman, V, Schultz, PG. An archaebacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in Escherichia coli. *Nucl. Acids Res.* **2003**, *31*, 6700-6709.

[132] Anderson, JC, Wu, N, Santoro, SW, Lakshman, V, King, DS, Schultz, PG. An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7566-7571.

[133] Blight, SK, Larue, RC, Mahapatra, A, Longstaff, DG, Chang, E, Zhao, G, Kang, PT, Church-Church, KB, Chan, MK, Krzycki, JA. Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature* **2004**, *431*, 333-335.

[134] Wan, W, Huang, Y, Wang, Z, Russell, William K, Pai, P-J, Russell, David H, Liu, WR. A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in Escherichia coli. *Angew. Chem. Int. Ed.* **2010**, *49*, 3211-3214.

[135] Yanagisawa, T, Ishii, R, Fukunaga, R, Kobayashi, T, Sakamoto, K, Yokoyama, S. Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode N-epsilon-(o-Azidobenzyloxycarbonyl) lysine for Site-Specific Protein Modification. *Chem. Biol.* **2008**, *15*, 1187-1197.

[136] Nozawa, K, O'Donoghue, P, Gundllapalli, S, Araiso, Y, Ishitani, R, Umehara, T, Soll, D, Nureki, O. Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. *Nature* **2009**, *457*, 1163-U1127.

[137] Hughes, RA, Ellington, AD. Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA. *Nucl. Acids Res.* **2010**, gkq521.

[138] Li, WT, Mahapatra, A, Longstaff, DG, Bechtel, J, Zhao, G, Kang, PT, Chan, MK, Krzycki, JA. Specificity of Pyrrolysyl-tRNA Synthetase for Pyrrolysine and Pyrrolysine Analogs. *J. Mol. Biol.* **2009**, *385*, 1156-1164.

[139] Wang, L, Brock, A, Herberich, B, Schultz, PG. Expanding the genetic code of Escherichia coli. *Science* **2001**, *292*, 498-500.

[140] Wang, L, Schultz, PG. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* **2001**, *8*, 883-890.

[141] Santoro, SW, Wang, L, Herberich, B, King, DS, Schultz, PG. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* **2002**, *20*, 1044-1048.

[142] Liu, CC, Schultz, PG. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, *79*, 413-444.

[143] Chin, JW, Santoro, SW, Martin, AB, King, DS, Wang, L, Schultz, PG. Addition of p-azido-L-phenylaianine to the genetic code of Escherichia coli. *J. Am. Chem. Soc.* **2002**, *124*, 9026-9027.

[144] Chin, JW, Martin, AB, King, DS, Wang, L, Schultz, PG. Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11020-11024.

[145] Xie, JM, Liu, WS, Schultz, PG. A genetically encoded bidentate, metal-binding amino acid. *Angew. Chem. Int. Ed.* **2007**, *46*, 9239-9242.

[146] Neumann, H, Peak-Chew, SY, Chin, JW. Genetically encoding N-epsilon-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **2008**, *4*, 232-234.

[147] Liu, WR, Wang, Y-S, Wan, W. Synthesis of proteins with defined posttranslational modifications using the genetic noncanonical amino acid incorporation approach. *Mol. Biosyst.* **2011**, *7*, 38-47.

[148] Melançon III, CE, Schultz, PG. One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3845-3847.

[149] Young, TS, Ahmad, I, Yin, JA, Schultz, PG. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *J. Mol. Biol.* **2010**, *395*, 361-374.

[150] Kuhn, SM, Rubini, M, Fuhrmann, M, Theobald, I, Skerra, A. Engineering of an Orthogonal Aminoacyl-tRNA Synthetase for Efficient Incorporation of the Non-natural Amino Acid O-Methyl-L-tyrosine using Fluorescence-based Bacterial Cell Sorting. *J. Mol. Biol.* **2010**, *404*, 70-87.

[151] Giege, R. Genetic code expansion. *Nat. Struct. Mol. Biol.* **2003**, *10*, 414-416.

[152] Dale, T, Sanderson, LE, Uhlenbeck, OC. The affinity of elongation factor Tu for an aminoacyl-tRNA is modulated by the esterified amino acid. *Biochemistry* **2004**, *43*, 6159-6166.

[153] Guo, J, III, Charles EM, Lee, HS, Groff, D, Schultz, Peter G. Evolution of Amber Suppressor tRNAs for Efficient Bacterial Production of Proteins Containing Nonnatural Amino Acids. *Angew. Chem. Int. Ed.* **2009**, *48*, 9148-9151.

[154] Neumann, H, Wang, K, Davis, L, Garcia-Alai, M, Chin, JW. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **2010**, *464*, 441-444.

[155] Hoesl, MG, Budisa, N. In vivo incorporation of multiple non-canonical amino acids into proteins. *Angew. Chem. Int. Ed.* **2011**, [epub, DOI: 10.1002/anie.201005680].

[156] Rackham, O, Chin, JW. A network of orthogonal ribosome center dot mRNA pairs. *Nat. Chem. Biol.* **2005**, *1*, 159-166.

[157] Wang, KH, Neumann, H, Peak-Chew, SY, Chin, JW. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.* **2007**, *25*, 770-777.

[158] Huang, Y, Russell, WK, Wan, W, Pai, PJ, Russell, DH, Liu, WS. A convenient method for genetic incorporation of multiple noncanonical amino acids into one protein in Escherichia coli. *Mol. Biosyst.* **2010**, *6*, 683-686.

[159] Ryu, YH, Schultz, PG. Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. *Nat. Methods* **2006**, *3*, 263-265.

[160] Kaya, E, Gutsmiedl, K, Vrabel, M, Müller, M, Thumbs, P, Carell, T. Synthesis of Threefold Glycosylated Proteins using Click Chemistry and Genetically Encoded Unnatural Amino Acids. *ChemBioChem* **2009**, *10*, 2858-2861.

[161] Sorensen, HP, Mortensen, KK. Advanced genetic strategies for recombinant protein expression in Escherichia coli. *J. Biotechnol.* **2005**, *115*, 113-128.

[162] Bazaral, M, Helinski, DR. Circular DNA forms of colicinogenic factors E1, E2 and E3 from Escherichia coli. *J. Mol. Biol.* **1968**, *36*, 185-194.

[163] Hershfield, V, Boyer, HW, Yanofsky, C, Lovett, MA, Helinski, DR. Plasmid ColE1 as a Molecular Vehicle for Cloning and Amplification of DNA. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3455-3459.

[164] Chang, ACY, Cohen, SN. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* **1978**, *134*, 1141-1156.

[165] Conrad, SE, Wold, M, Campbell, JL. Origin and direction of DNA replication of plasmid RSF1030. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 736-740.

[166] Zverev, VV, Khmel, IA. The nucleotide sequences of the replication origins of plasmids ColA and ColD *Plasmid* **1985**, *14*, 192-199.

[167] Dunn, J. http://genome.bnl.gov/Vectors/pscans.php.

[168] Qiagen. *The QIAexpressionist - A handbook for high-level expression and purification of 6xHis-tagged proteins, Vol. 5*, **2003**.

[169] Budisa, N. *Engineering the genetic code: expanding the amino acid repertoire for the design of novel proteins*, Wiley-VCH, Weinheim, **2005**.

[170] Giese, C, Lepthien, S, Metzner, L, Brandsch, M, Budisa, N, Lilie, H. Intracellular uptake and inhibitory activity of aromatic fluorinated amino acids in human breast cancer cells. *ChemMedChem* **2008**, *3*, 1449-1456.

[171] Wilkins, BJ, Marionni, S, Young, DD, Liu, J, Wang, Y, Di Salvo, ML, Deiters, A, Cropp, TA. Site-Specific Incorporation of Fluorotyrosines into Proteins in Escherichia coli by Photochemical Disguise. *Biochemistry* **2010**, *49*, 1557-1559.

[172] Jensen, RL, Stade, LW, Wimmer, R, Stensballe, A, Duroux, M, Larsen, KL, Wingren, C, Duroux, L. Direct Site-Directed Photocoupling of Proteins onto Surfaces Coated with beta-Cyclodextrins. *Langmuir* **2010**, *26*, 11597-11604.

[173] Wang, WY, Takimoto, JK, Louie, GV, Baiga, TJ, Noel, JP, Lee, KF, Slesinger, PA, Wang, L. Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat. Neurosci.* **2007**, *10*, 1063-1072.

[174] Kwon, I, Tirrell, DA. Site-specific incorporation of tryptophan analogues into recombinant proteins in bacterial cells. *J. Am. Chem. Soc.* **2007**, *129*, 10431-10437.
[175] Kwon, I, Kirshenbaum, K, Tirrell, DA. Breaking the degeneracy of the genetic code. *J. Am. Chem. Soc.* **2003**, *125*, 7512-7513.
[176] Antonczak, AK, Simova, Z, Tippmann, EM. A Critical Examination of Escherichia coli Esterase Activity. *J. Biol. Chem.* **2009**, *284*, 28795-28800.
[177] Zhang, Z, Gildersleeve, J, Yang, Y-Y, Xu, R, Loo, JA, Uryu, S, Wong, C-H, Schultz, PG. Retraction. *Science* **2009**, *326*, 1187.
[178] Xu, R, Hanson, SR, Zhang, Z, Yang, Y-Y, Schultz, PG, Wong, C-H. Site-Specific Incorporation of the Mucin-Type N-Acetylgalactosamine-α-O-threonine into Protein in Escherichia coli. *J. Am. Chem. Soc.* **2009**, *131*, 13883-13883.
[179] Merkel, L, Schauer, M, Antranikian, G, Budisa, N. Parallel Incorporation of Different Fluorinated Amino Acids: On the Way to Teflon Proteins. *ChemBioChem* **2010**, *11*, 1505-1507.
[180] Merkel, L, Hoesl, MG, Albrecht, M, Schmidt, A, Budisa, N. Blue Fluorescent Amino Acids As In Vivo Building Blocks for Proteins. *ChemBioChem* **2010**, *11*, 305-314.
[181] Liu, WS, Alfonta, L, Mack, AV, Schultz, PG. Structural basis for the recognition of para-benzoyl-L-phenylalanine by evolved aminoacyl-tRNA synthetases. *Angew. Chem. Int. Ed.* **2007**, *46*, 6073-6075.
[182] Cellitti, SE, Jones, DH, Lagpacan, L, Hao, XS, Zhang, Q, Hu, HY, Brittain, SM, Brinker, A, Caldwell, J, Bursulaya, B, Spraggon, G, Brock, A, Ryu, Y, Uno, T, Schultz, PG, Geierstanger, BH. In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* **2008**, *130*, 9268-9281.
[183] The-Biobricks-Foundation. http://bbf.openwetware.org/.
[184] Shetty, R, Endy, D, Knight, T. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **2008**, *2*, 5.
[185] Lee, SK, Chou, HH, Pfleger, BF, Newman, JD, Yoshikuni, Y, Keasling, JD. Directed evolution of AraC for improved compatibility of arabinose- and lactose-inducible promoters. *Appl. Environ. Microbiol.* **2007**, *73*, 5711-5715.
[186] Lee, SK, Keasling, JD. A Propionate-Inducible Expression System for Enteric Bacteria. *Appl. Environ. Microbiol.* **2005**, *71*, 6856-6862.
[187] Schmidt, TGM, Skerra, A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protocols* **2007**, *2*, 1528-1535.
[188] Sezonov, G, Joseleau-Petit, D, D'Ari, R. Escherichia coli Physiology in Luria-Bertani Broth. *J. Bacteriol.* **2007**, *189*, 8746-8749.
[189] Nikaido, H. http://schaechter.asmblog.org/schaechter/2009/11/the-limitations-of-lb-medium.html.
[190] Lee, SK, Newman, JD, Keasling, JD. Catabolite Repression of the Propionate Catabolic Genes in Escherichia coli and Salmonella enterica: Evidence for Involvement of the Cyclic AMP Receptor Protein. *J. Bacteriol.* **2005**, *187*, 2793-2800.
[191] Hoesl, MG, Budisa, N. Expanding and engineering the genetic code in a single expression experiment. *ChemBioChem* **2011**, *12*, 552-555.
[192] Vinette, AL, McNamee, JP, Bellier, PV, McLean, JRN, Scaiano, JC. Prompt and delayed nonsteroidal anti-inflammatory drug-photoinduced DNA damage in peripheral blood mononuclear cells measured with the comet assay. *Photochem. Photobiol.* **2003**, *77*, 390-396.
[193] Cubitt, AB, Woollenweber, LA, Heim, R. Understanding structure-function relationships in the Aequorea victoria green fluorescent protein. *Methods Cell Biol.* **1999**, *58*, 19-30.
[194] Mukai, T, Kobayashi, T, Hino, N, Yanagisawa, T, Sakamoto, K, Yokoyama, S. Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818-822.
[195] Royter, M, Schmidt, M, Elend, C, Höbenreich, H, Schäfer, T, Bornscheuer, U, Antranikian, G. Thermostable lipases from the extreme thermophilic anaerobic bacteria Thermoanaerobacter thermohydrosulfuricus SOL1 and Caldanaerobacter subterraneus subsp. tengcongensis. *Extremophiles* **2009**, *13*, 769-783.
[196] de Graaf, AJ, Kooijman, M, Hennink, WE, Mastrobattista, E. Nonnatural Amino Acids for Site-Specific Protein Conjugation. *Bioconjug. Chem.* **2009**, *20*, 1281-1295.

[197] Ayyadurai, N, Neelamegam, R, Nagasundarapandian, S, Edwardraja, S, Park, HS, Lee, SJ, Yoo, TH, Yoon, H, Lee, SG. Importance of expression system in the production of unnatural recombinant proteins in Escherichia coli. *Biotechnol. Bioprocess Eng.* **2009**, *14*, 257-265.

[198] Moritz, RL, Witte, CE, Welch, RA. Deletion of the rare codon tRNA argW causes a decrease in virulence of uropathogenic Escherichia coli strain CFT073. *Abstr. Gen. Meet. Am. Soc. Microbiol.* **2006**, *106*, 181.

[199] Spanjaard, RA, Chen, K, Walker, JR, Vanduin, J. Frameshift suppression at tandem AGA and AGG codons by cloned tRNA genes: assigning a codon to argU tRNA and T4 tRNAArg. *Nucl. Acids Res.* **1990**, *18*, 5031-5036.

[200] Meroueh, M, Chow, CS. Thermodynamics of RNA hairpins containing single internal mismatches. *Nucl. Acids Res.* **1999**, *27*, 1118-1125.

[201] Neumann, H, Slusarczyk, AL, Chin, JW. De Novo Generation of Mutually Orthogonal Aminoacyl-tRNA Synthetase/tRNA Pairs. *J. Am. Chem. Soc.* **2010**, *132*, 2142-2144.

[202] Miyake-Stoner, SJ, Refakis, CA, Hammill, JT, Lusic, H, Hazen, JL, Deiters, A, Mehl, RA. Generating Permissive Site-Specific Unnatural Aminoacyl-tRNA Synthetases. *Biochemistry* **2010**, *49*, 1667-1677.

[203] Stokes, AL, Miyake-Stoner, SJ, Peeler, JC, Nguyen, DP, Hammer, RP, Mehl, RA. Enhancing the utility of unnatural amino acid synthetases by manipulating broad substrate specificity. *Mol. Biosyst.* **2009**, *5*, 1032-1038.

[204] Kleina, LG, Masson, JM, Normanly, J, Abelson, J, Miller, JH. Construction of Escherichia coli amber suppressor tRNA genes: II. Synthesis of additional tRNA genes and improvement of suppressor efficiency *J. Mol. Biol.* **1990**, *213*, 705-717.

[205] Ryden, SM, Isaksson, LA. A temperature-sensitive mutant of Escherichia coli that shows enhanced misreading of UAG/A and increased efficiency for tRNA nonsense suppressors *Mol. Gen. Genet.* **1984**, *193*, 38-45.

[206] Mukai, T, Hayashi, A, Iraha, F, Sato, A, Ohtake, K, Yokoyama, S, Sakamoto, K. Codon reassignment in the Escherichia coli genetic code. *Nucl. Acids Res.* **2010**, *38*, 8188-8195.

[207] Kanda, T, Takai, K, Yokoyama, S, Takaku, H. Knocking out a specific tRNA species within unfractionated Escherichia coli tRNA by using antisense (complementary) oligodeoxyribonucleotides. *FEBS Lett.* **1998**, *440*, 273-276.

[208] Becker, S, Hobenreich, H, Vogel, A, Knorr, J, Wilhelm, S, Rosenau, F, Jaeger, KE, Reetz, MT, Kolmar, H. Single-cell high-throughput screening to identify enantioselective hydrolytic enzymes. *Angew. Chem. Int. Ed.* **2008**, *47*, 5085-5088.

[209] Deiters, A, Cropp, TA, Mukherji, M, Chin, JW, Anderson, JC, Schultz, PG. Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. *J. Am. Chem. Soc.* **2003**, *125*, 11782-11783.

[210] Loidl, G, Musiol, HJ, Budisa, N, Huber, R, Poirot, S, Fourmy, D, Moroder, L. Synthesis of beta-(1-azulenyl)-L-alanine as a potential blue-colored fluorescent tryptophan analog and its use in peptide synthesis. *J. Pept. Sci.* **2000**, *6*, 139-144.

[211] Link, AJ, Vink, MKS, Tirrell, DA. Presentation and Detection of Azide Functionality in Bacterial Cell Surface Proteins. *J. Am. Chem. Soc.* **2004**, *126*, 10598-10602.

[212] Berlyn, MKB. Linkage map of Escherichia coli K-12, edition 10: The traditional map. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 814-984.

[213] Bachmann, BJ. *Derivations and genotypes of some mutant derivatives of Escherichia coli K-12,* in *Escherichia coli and Salmonella: cellular and molecular biology*, second ed. (Eds.: F. C. Neidhardt, I. R. Curtiss, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, H. E. Umbarger), ASM Press, Washington, D.C, **1996**, pp. 2460-2488.

[214] Fangman, WL, Neidhardt, FC. Demonstration of an Altered Aminoacyl Ribonucleic Acid Synthetase in a Mutant of Escherichia coli. *J. Biol. Chem.* **1964**, *239*, 1839-1843.

[215] Wood, WB. Host specificity of DNA produced by Escherichia coli: Bacterial mutations affecting the restriction and modification of DNA *J. Mol. Biol.* **1966**, *16*, 118-133.

[216] Studier, FW, Moffatt, BA. Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes. *J. Mol. Biol.* **1986**, *189*, 113-130.

[217] Studier, FW, Daegelen, P, Lenski, RE, Maslov, S, Kim, JF. Understanding the Differences between Genome Sequences of Escherichia coli B Strains REL606 and BL21(DE3) and Comparison of the E-coli B and K-12 Genomes. *J. Mol. Biol.* **2009**, *394*, 653-680.

[218] Young, IG, Langman, L, Luke, RKJ, Gibson, F. Biosynthesis of the Iron-Transport Compound Enterochelin: Mutants of Escherichia coli Unable to Synthesize 2, 3-Dihydroxybenzoate. *J. Bacteriol.* **1971**, *106*, 51-57.

[219] Miki, T, Yamamoto, Y, Matsuda, H. A novel, simple, high-throughput method for isolation of genome-wide transposon insertion mutants of Escherichia coli K-12. *Methods Mol. Biol.* **2008**, 195-204.

[220] Singer, M, Baker, TA, Schnitzler, G, Deischel, SM, Goel, M, Dove, W, Jaacks, KJ, Grossman, AD, Erickson, JW, Gross, CA. A Collection of Strains Containing Genetically Linked Alternating Antibiotic Resistance Elements. *Microbiol. Rev.* **1989**, *53*, 1-24.

[221] Dürwald, H, Hoffmann-Berling, H. Endonuclease I-deficient and ribonuclease I-deficient Escherichia coli mutants. *J. Mol. Biol.* **1968**, *34*, 331-346.

[222] Meselson, M, Yuan, R. DNA Restriction from *E. coli*. *Nature* **1968**, *217*, 1110-1114.

[223] Hanahan, D. Studies on transformation of Escherichia coli with plasmids *J. Mol. Biol.* **1983**, *166*, 557-580.

[224] Grant, SGN, Jessee, J, Bloom, FR, Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 4645-4649.

[225] Ma, H, Kunes, S, Schatz, PJ, Botstein, D. Plasmid construction by homologous recombination in yeast. *Gene* **1987**, *58*, 201-216.

[226] Lawrence, F, Blanquet, S, Poiret, M, Robert-Gero, M, Waller, J-P. The Mechanism of Action of Methionyl-tRNA Synthetase. *Eur. J. Biochem.* **1973**, *36*, 234-243.

[227] Winkler, UK, Stuckmann, M. Glycogen, hyaluronate, and some other polysaccharides greatly enhance the formation of exolipase by Serratia marcescens. *J. Bacteriol.* **1979**, *138*, 663-670.

[228] Bradford, MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248-254.

[229] Shevchenko, A, Wilm, M, Vorm, O, Mann, M. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal. Chem.* **1996**, *68*, 850-858.

[230] Olsen, JV, de Godoy, LMF, Li, GQ, Macek, B, Mortensen, P, Pesch, R, Makarov, A, Lange, O, Horning, S, Mann, M. Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell Proteomics* **2005**, *4*, 2010-2021.

[231] Gruhler, A, Olsen, JV, Mohammed, S, Mortensen, P, Faergeman, NJ, Mann, M, Jensen, ON. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol. Cell Proteomics* **2005**, *4*, 310-327.

8 Figure list

9 Table list

10 Danksagung

Mein besonderer Dank gilt meinem Betreuer Herrn Prof. Dr. Nediljko Budisa, der mich mit seinem Enthusiasmus und seiner Begeisterungsfähigkeit stets motiviert und inspiriert hat. Seine stetige Unterstützung, Hilfs- und Diskussionsbereitschaft haben diese Arbeit erst möglich gemacht. Ich wünsche Dir, dass Du all Deine Visionen Realität werden lässt!

Herrn Prof. Dr. Johannes Buchner und Herrn Prof. Dr. Thomas Kiefhaber danke ich für die Bereitschaft zur Übernahme des Korreferats.

Bei Herrn Prof. Dr. Dieter Oesterhelt bedanke ich mich für die Möglichkeit, in seiner Abteilung zu arbeiten und die hervorragende Infrastruktur zu nutzen.

Herrn Prof. Dr. Garobed Antranikian, Herrn Dr. Ralf Grote and Frau Dr. Marina Royter will ich auf diesem Weg für die gute Zusammenarbeit in unserem „Lipaseprojekt“ danken. Herrn Dr. Martin Grininger, Herrn Prof. Dr. Josef Wachtveitl, Frau Heike Staudt und Herrn Prof. Dr. Andreas Dreuw danke ich herzlich für die hervorragende Zusammenarbeit im „Dodecinprojekt“.

Mein Dank geht auch an Herrn Prof. Dr. Peter Schultz und Herrn Prof. Dr. David Tirrell für die Bereitstellung der orthogonalen Paare, ohne die diese Arbeit nicht möglich gewesen wäre. Weiterhin bedanke ich mich bei Herrn Prof. Dr. Jay Keasling für die Bereitstellung des Propionat-Promoters.

Frau Dr. Sandra Lepthien, Frau Dr. Birgit Wiltschi, Herrn Dr. Lars Merkel und Herrn Dr. Carlos Acevedo Rocha danke ich für die Mühe bei der kritischen Durchsicht der Arbeit.

Mein Dank gilt weiterhin Frau Elisabeth „Lissy“ Weyher-Stingl, die mit ihrer Hilfsbereitschaft und ihrem großen Wissen im Bereich der Analytik einen bedeutenden Teil zum Gelingen dieser Doktorarbeit beigetragen hat. Bei Herrn Dr. Cyril Boulegue bedanke ich mich für die Orbitrap Analysen, bei Herrn Dr. Jürgen Musiol für die Synthese des Azulenylalanins, und bei Frau Petra Birle und Frau Tatjana Krywcun für die Bereitstellung von Plasmiden.

Sehr herzlich will ich mich bei allen meinen Kollegen aus der Budisa Gruppe bedanken. Ihr habt die Zeit der Doktorarbeit wie im Fluge vergehen lassen und unvergesslich gemacht. Traudl Wenger möchte ich vor allem für die tolle Unterstützung im Labor, ihre wertvollen Ratschläge und ihr stets offenes Ohr danken. Birgit Wiltschi danke ich für die zahlreichen Diskussionen und für ihre große Hilfestellung bei wissenschaftlichen Problemen und im Gewächshaus. Ein ganz großer Dank geht auch an Lars Merkel, der mit seiner Unterstützung, seinen Ratschlägen und seinem Fussballsachverstand meine Doktorandenzeit innerhalb wie außerhalb des Labors sehr bereichert hat. Sandra Lepthien und Tina Wolschner danke ich für die Unterstützung im Labor, die unvergesslichen Kaffeepausen und die lustigen Stunden innerhalb und außerhalb des Labors. Maud Larregola danke ich für ihre Unterstützung im Labor und das französische Flair, das sie in die Gruppe gebracht hat. Nicht zuletzt will ich Carlos Acevedo Rocha danken, der mir in vielen Lebens- und Laborlagen ein guter Freund und für jeden Schmarrn zu haben war.

Bedanken möchte ich mich auch bei allen meinen Studenten/innen — insbesondere bei Haissi Cui, Marco Perna, Sascha Serdjukov und Oliver Baker —, die im Rahmen ihrer Praktika, Bachelor- und Masterarbeiten ihren Beitrag zum Fortgang meiner Projekte geleistet haben.

Anne Frohn, Amy Schönegge, Roland Knispel und Carolin Fleischer danke ich für lustige Stunden und vor allem für mittwochs, 17 Uhr.

Ein ganz besonderer Dank geht an meine Freunde Hannes con gas, Stefan, Honki, Anderl, Andreas, Hannes sin gas, Rainer, Julia, Zicka, Luise, Iris, Agathe, Alina, Heike und nicht zuletzt Flo. Ihre Freundschaft über die Jahre hat mir viele tolle Tage, Nächte und Erlebnisse bereitet und mir immer Rückhalt gegeben.

Der größte Dank gebührt aber meiner Familie und vor allem meiner Freundin Ari, die immer für mich da waren, mich immer unterstützt und ertragen haben.