

Improved Pyrrolysyl-tRNA Synthetase Derived Orthogonal Translation Systems

vorgelegt von

M. Sc.

Nikolaj Georg Koch

ORCID: 0000-0003-2745-0020

an der Fakultät III – Prozesswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

Dr. rer. nat.

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Peter Neubauer

Gutachter: Prof. Dr. Nediljko Budisa

Gutachter: Prof. Dr. Juri Rappsilber

Gutachterin: Prof. Dr. Zoya Ignatova

Tag der wissenschaftlichen Aussprache: 2. September 2022

Berlin 2023

Danksagung

Folgend möchte mich bei allen herzlich bedanken, die mich im Laufe meiner Doktorarbeit begleitet und unterstützt haben.

Ned danke ich ganz besonders für die Möglichkeit des Verfassens dieser Arbeit und den unermesslich kreativen Freiraum, den er mir bei der Ausrichtung meiner Forschungsthemen gegeben hat. Seine Art von Support ohne groß Druck aufzubauen habe ich sehr genossen und ich denke die zahlreichen Verbesserungen und gewonnenen Erkenntnisse über das Pyrrolsyl-System sprechen für sich.

Bei Prof. Dr. Juri Rappsilber möchte ich mich bedanken, dass er mir trotz Neds Umzugs nach Kanada ermöglicht hat meine Doktorarbeit ohne großartige Komplikationen abzuschließen und durch die Aufnahme in seine Arbeitsgruppe extrem spannende Einblicke in den Proteomics Forschungsbereich zu erlangen.

Prof Dr. Zoya Ignatova danke ich für die Übernahme des Gutachtens und Prof. Dr. Peter Neubauer danke ich für die Übernahme des Prüfungsvorsitz.

Ganz besonders bedanken möchte ich mich bei El Cheffe, Dr. Tobias Baumann, der mich seit meiner Masterarbeit unterstützt hat und selbst seitdem er nicht mehr in der Arbeitsgruppe war immer noch ein offenes Ohr und Ratschläge parat hatte, falls benötigt. Tausend Dank Chef.

Ein Riesen Dank zur Realisierung meiner Forschung geht auch an Dr. Philipp Ochtrup und Christian Stieger (beide AK Hackenberger) für die doch sehr zahlreichen ESI-MS Messungen.

Vielen Dank geht an alle Kollegen des AK Budisa, die immer sehr hilfsbereit waren und diese Gruppe mit einem entspannten Vibe ausgestattet haben (obwohl zum Schluss ja nicht mehr viele von uns übrig waren :D). Auch bei den Kollegen des AK Rappsilber möchte ich mich für die einfache und kollegiale Zusammenarbeit bedanken.

Weiterhin möchte ich Anette Poch (AK Süßmuth) für die immer sehr schnelle zur Verfügungstellung der Keio Stämme danken, wenn benötigt.

Schlussendlich möchte ich der Person danken, die den größten Anteil daran hatte, dass ich diese Arbeit verfassen konnte. Dieser unendliche Dank geht an meine Mutter, ohne deren Unterstützung ich diese Arbeit nie hätte verfassen können. Ihr Support und ihre positive Aussicht auf die Welt haben mich auch in den vielen nicht so einfachen Zeiten (weit vor dieser Doktorarbeit) meines Lebens gerettet. Ohne ihren Beistand wäre das alles nicht möglich gewesen.

Parts of this work were published as listed below:

Koch, N. G.; Goettig, P.; Rappsilber, J.; Budisa, N. Engineering Pyrrolysyl-tRNA Synthetase for the Incorporation of Non-Canonical Amino Acids with Smaller Side Chains. *Int. J. Mol. Sci.* **2021**, 22 (20), 11194. <https://doi.org/10.3390/ijms222011194>.

Koch, N. G.; Baumann, T.; Budisa, N. Efficient Unnatural Protein Production by Pyrrolysyl-tRNA Synthetase with Genetically Fused Solubility Tags. *Front. Bioeng. Biotechnol.* **2021**, 9, 1–14. <https://doi.org/10.3389/fbioe.2021.807438>.

Koch, N. G.; Budisa, N. Pyrrolysyl-tRNA-Synthetase: Methanogenese Und Gencode-Erweiterung. *BIOspektrum* **2021**, 27 (6), 616–619. <https://doi.org/10.1007/s12268-021-1653-x>.

TABLE OF CONTENTS

Table of Contents

I. SUMMARY.....	X
II. ZUSAMMENFASSUNG	XI
III. ABBREVIATIONS	XIII
IV. LIST OF FIGURES	XVI
V. LIST OF TABLES.....	XXII
1. INTRODUCTION.....	1
1.1. PROTEIN BIOSYNTHESIS AND THE LINK TO GENETICS: HISTORIC CONTEXT	1
1.2. PROTEIN BIOSYNTHESIS AND THE LINK TO GENETICS: THE PRESENT AGE	4
1.2.1. STRUCTURE OF THE GENETIC CODE	4
1.2.2. EXCURSUS: ORIGIN OF LIFE AND EVOLUTION OF THE GENETIC CODE: A NOTE ON COMPLEXITY, PREDICTABILITY AND DEDUCIBILITY	6
1.2.3. THE PROTEIN BIOSYNTHESIS.....	9
1.2.3.1. THE TRANSFER RNA	10
1.2.3.2. AMINOACYL TRNA SYNTHETASES	12
1.3. SYNTHETIC BIOLOGY	13
1.3.1. GENETIC CODE EXPANSION	13
1.3.1.1. STOP CODON SUPPRESSION	13
1.3.1.2. PYRROLYSYL TRNA SYNTHETASE.....	15
1.3.2. ENCODING NEW ENZYME FUNCTIONS: PERSPECTIVES ON PYLRS VS. <i>MJTYRRS</i> ENGINEERING	18
1.3.2.1. SOLUBILITY TAGS.....	19
1.4. USE OF EXTREMOPHILIC ENZYMES IN SYNTHETIC BIOLOGY	20
1.4.1. HEAT-ADAPTED ENZYMES	20
1.4.2. COLD-ADAPTED ENZYMES	20
1.5. AIM OF THIS STUDY	23
2. RESULTS AND DISCUSSION.....	24
2.1. GENERAL NCAA INCORPORATION READOUT AND PYLRS OTS SETUPS	24
2.2. INCREASING THE PYLRS OTS EFFICIENCY USING GENETICALLY FUSED SOLUBILITY TAGS.....	25
2.2.1. CHOICE OF USED SOLUBILITY TAGS.....	26
2.2.2. COMPARISON OF DIFFERENT N-TERMINAL FUSED SOLUBILITY TAGS	27
2.2.3. ANALYTICS OF SAC INCORPORATION.....	29
2.2.4. ANALYSIS OF PYLRS SOLUBILITY	32
2.2.5. TRANSFERABILITY OF TAG EFFECTS TO OTHER PYLRS VARIANTS	33
2.2.6. IMPACT OF ACTIVE-SITE MUTATIONS	35
2.2.7. TESTING THE IMPROVED <i>MBSACRS</i> FOR <i>IN VIVO</i> SAC SYNTHESIS AND INCORPORATION	37
2.3. ENGINEERING PYRROLYSYL-TRNA SYNTHETASE FOR THE INCORPORATION OF NON-CANONICAL AMINO ACIDS WITH SMALLER SIDE CHAINS	40
2.3.1. GENERAL FEATURES AND PERSPECTIVES OF USED NCAAS	42
2.3.2. GENERAL <i>MBPYLRS</i> SETUP	43

2.3.3.	TESTING <i>MBSACRS</i> FOR ALIPHATIC SUBSTRATES	44
2.3.4.	ELUCIDATING THE SAR OF <i>MBSACRS</i> VIA RATIONAL MUTATION STUDIES	44
2.3.5.	RATIONALIZING SAC INCORPORATION DATA AND CREATING ALIPHATIC SUBSTRATE INCORPORATING <i>MBPYLRS</i> VARIANTS	47
2.3.6.	SEMI-RATIONAL ENGINEERING OF <i>PYLRS</i> CONSTRUCTS FOR SMALL ALIPHATIC SUBSTRATE INCORPORATION	49
2.3.7.	EVALUATING THE INCORPORATION OF BIOCHEMICALLY USEFUL ALIPHATIC NCAA ANALOGS	50
2.3.8.	ANALYTICS OF CANONICAL/NON-CANONICAL AMINO ACIDS INCORPORATION	51
2.4.	CREATION OF NOVEL <i>PYLRS</i> OTSS BASED ON PSYCHRO- AND THERMOPHILIC HOMOLOGS	54
2.4.1.	<i>IN SILICO</i> ANALYSIS OF ARCHAEAL <i>PYLRS</i> VARIANTS	54
2.4.1.1.	PHYLOGENETIC <i>PYLRS</i> ANALYSIS.....	54
2.4.1.2.	ANALYSIS OF METHANOSARCINA <i>PYLRS</i> AND VARIABLE REGION	55
2.4.1.3.	ANALYSIS OF METHANOSARCINALES <i>PYLRS</i> AND VARIABLE REGION	57
2.4.2.	CHOICE AND ELABORATION OF USED <i>PYLRS</i> CONSTRUCTS	59
2.4.3.	COMPARISON OF OTS EFFICIENCIES OF PSYCHRO-, MESO-, AND THERMOPHILIC <i>PYLRS</i>	60
2.4.3.1.	MULTI-SITE NCAA INCORPORATION IN BL21(DE3) AND B95.ΔA E. COLI STRAINS	63
2.4.4.	ESTIMATION OF THE <i>IN VIVO</i> <i>PYLRS</i> SUBSTRATE PROMISCUITIES.....	66
2.4.5.	COMPARING OTS EFFICIENCIES OF PSYCHRO-, MESO-, AND THERMOPHILIC <i>PYLRS</i> MUTANTS.....	67
2.4.5.1.	ELUCIDATING THE GENERAL USEFULNESS OF THE SMBP-TAG FOR OTHER <i>PYLRS</i> MUTANTS WITH DESTABILIZING MUTATIONS	68
2.4.5.2.	COMPARING <i>PYLRS</i> DOUBLE ALANINE/GLYCINE MUTANTS FOR INCORPORATION OF TYROSINE AND PHENYLALANINE DERIVATIVES	68
2.4.5.3.	DETAILED <i>IN VIVO</i> ACTIVITY ANALYSES OF THE DOUBLE ALANINE/GLYCINE MUTANTS.....	70
2.4.5.4.	ELUCIDATING INCORPORATION EFFICIENCIES FOR ALIPHATIC NCAA INCORPORATING <i>PYLRS</i>	71
2.4.5.5.	COMPARING OTS EFFICIENCIES OF <i>SACRS</i> AND IMPROVED <i>SACRS</i> VARIANTS	73
2.4.5.6.	MULTI-SITE INCORPORATION OF <i>SAC</i> (1) AND <i>SPROC</i> (27)	75
2.4.5.7.	SCREENING <i>MBURSACRS</i> AND THE S379T MUTANT FOR <i>SAC</i> ANALOG INCORPORATION ACTIVITY.....	77
2.4.5.8.	IMPACT OF DIFFERENT <i>TRNA^{PYL}</i> FOR OTSS	77
2.4.5.8.1.	COMPARING <i>TRNA</i> FROM <i>M. MAZEI</i> AND <i>M. BURTONII</i>	78
2.4.6.	COMPARING <i>MJTYRRS</i> AND <i>MBURPYLRS</i> PERFORMANCE.....	80
2.4.7.	ANALYTICS OF NCAA INCORPORATION.....	82
2.5.	SIDE PROJECT: PRODUCTION OF GENETICALLY ENCODED BIOINSPIRED SMART WET ADHESIVE MATERIALS.....	83
2.5.1.	CHOICE, DESIGN AND CREATION OF PROTEIN SCAFFOLDS.....	84
2.5.2.	OPTIMIZING THE <i>MJONB-DOPARS</i>	86
2.5.3.	MULTISITE INCORPORATION OF (<i>M-ONB-DOPA</i>) INTO ELASTIN-LIKE-POLYPEPTIDE-SFGFP FUSION PROTEINS	87
2.5.4.	ANALYTICS OF ELP(5X <i>ONB-DOPA</i>)-SFGFP FUSION PROTEIN	89
2.5.5.	ENGINEERING A BACTERIAL STRAIN WITH REDUCED NITROREDUCTASE ACTIVITY	91
2.5.6.	EXPERIMENTS TO VERIFY THE DECREASED NK53 NITROREDUCTASE ACTIVITY	93
2.5.7.	EXPANDING THE NCAA INCORPORATION REPERTOIRE FOR WET TISSUE ADHESION	96

3.	SUMMARY AND OUTLOOK.....	99
4.	MATERIALS AND METHOD	103
4.1.	MATERIALS	103
4.1.1.	CHEMICALS.....	103
4.1.2.	MEDIA AND SUPPLEMENTS	104
4.1.3.	STRAINS.....	105
4.1.4.	TECHNICAL EQUIPMENT	106
4.1.5.	OLIGONUCLEOTIDES	107
4.1.6.	BIOMOLECULAR REAGENTS, ENZYMES AND KIT REAGENTS:.....	107
4.1.7.	BUFFERS AND SOLUTIONS	108
4.2.	MOLECULAR BIOLOGICAL METHODS	109
4.2.1.	TRANSFORMATION OF <i>E. COLI</i>	109
4.2.1.1.	TRANSFORMATION OF CHEMICALLY COMPETENT CELLS	109
4.2.1.2.	TRANSFORMATION OF ELECTROCOMPETENT CELLS	110
4.2.2.	ISOLATION OF PLASMID DNA	110
4.2.3.	POLYMERASE CHAIN REACTION.....	111
4.2.3.1.	COLONY PCR	111
4.2.3.2.	STANDARD PCR	111
4.2.3.2.1.	SITE-DIRECTED AND SITE-SATURATION MUTAGENESIS	112
4.2.4.	LIBRARY SCREENING	112
4.2.5.	<i>IN VITRO</i> MODIFICATION OF DNA.....	112
4.2.5.1.	PLASMID VECTOR CONSTRUCTION	112
4.2.5.2.	RESTRICTION DIGEST OF DNA	112
4.2.5.3.	LIGATION OF DNA FRAGMENTS.....	113
4.2.5.4.	RECURSIVE DIRECTIONAL LIGATION	113
4.2.6.	CHROMOSOMAL GENE DELETION BY HOMOLOGOUS RECOMBINATION	113
4.3.	PROTEIN EXPRESSION AND PURIFICATION	114
4.3.1.	RECOMBINANT PROTEIN EXPRESSION.....	114
4.3.2.	PROTEIN PURIFICATION.....	114
4.4.	BIOCHEMICAL METHODS	115
4.4.1.	PLATE READER ASSAYS	115
4.4.1.1.	ANALYSIS OF SUMO-SFGFP EXPRESSION BY INTACT CELL FLUORESCENCE	115
4.4.1.2.	TIME-RESOLVED CHROMOPHORE MATURATION OF AMILCP(Y63ONBY)	115
4.4.2.	ANTIMICROBIAL ACTIVITY ASSAY	115
4.4.3.	DECAGING ONBY	116
4.5.	ANALYTICAL METHODS.....	116
4.5.1.	AGAROSE GEL ELECTROPHORESIS FOR DNA ANALYSIS	116
4.5.2.	POLYACRYLAMIDE GEL ELECTROPHORESIS FOR PROTEIN ANALYSIS	116
4.5.3.	ANALYSIS OF SOLUBLE AND INSOLUBLE PROTEIN FRACTIONS.....	117

4.5.4.	DETERMINATION OF PROTEIN CONCENTRATION	117
4.6.	BIOPHYSICAL METHODS	118
4.6.1.	ESI-MS	118
4.6.1.1.	AGILENT 6530 QTOF	118
4.6.1.2.	WATERS XEVO G2-XS QTOF	118
5.	REFERENCES.....	119
6.	APPENDIX	143
6.1.	ADDITIONAL RESULTS	143
6.1.1.	FLUORESCENCE ASSAYS	143
6.1.2.	SDS-PAGE	179
6.1.3.	AGAROSE GELS	179
6.1.4.	MISCELLANEOUS	180
6.1.5.	ESI-MS SPECTRA	181
6.1.6.	PLASMIDS VECTOR MAPS	210
6.1.7.	NUCLEOTIDE SEQUENCES	213
6.1.7.1.	TRNAS	213
6.1.7.2.	SOLUBILITY TAGS.....	214
6.1.7.3.	PYLRS USED IN THIS STUDY.....	216
6.1.7.4.	REPORTER CONSTRUCTS.....	221
6.1.7.5.	GENOMES USED FOR IN SILICO ANALYSIS	226
6.1.8.	DETERMINATION OF THE PYLRS VARIABLE REGIONS (LINKER)	227
6.1.9.	SIMILARITY MATRIX OF METHANOSARCINA PYLRS	228
6.1.10.	DESCRIPTIVE STATISTICAL DATA OF PYLRSS.....	228
6.1.11.	EXAMPLE OF GENE DELETION WITH NFSB	232
7.	EIDESSTÄTTLICHE ERKLÄRUNG.....	233

I. Summary

Site-specific incorporation of non-canonical amino acids (ncAAs) into proteins has emerged as a universal tool for systems bioengineering at the interface of chemistry, biology, and technology. Introducing ncAAs by engineered orthogonal pairs of aminoacyl-tRNA synthetases and tRNAs has proven to be a highly useful for the expansion of the genetic code. Pyrrolysyl-tRNA synthetase (PylRS) from methanogenic archaeal species is particularly attractive due to its natural orthogonal reactivity in bacterial and eukaryotic cells. However, the scope of such a reprogrammed translation is often limited, due to low yields of chemically modified target protein. This can be the result of substrate specificity engineering, which decreases the aminoacyl-tRNA synthetase stability and reduces the abundance of active enzyme. Additionally, the substrate scope of the PylRS system is limited to amino acids with long and/or bulky side chains. These two main drawbacks of the PylRS system were successfully tackled in this work to increase the usefulness of this already widely used system. To increase the protein production yield, the ncAA incorporation efficiency was increased with two strategies. In the first strategy it could be shown that the solubility and folding of engineered PylRS enzymes can become a bottleneck for the production of ncAA-containing proteins *in vivo*. Solubility tags derived from various species provided the means to remedy this issue and enhanced the production of site-specifically labelled proteins for a variety of engineered PylRS variants by 200–540%, even the wild-type enzyme gained up to 245% efficiency. The second strategy involved PylRS enzymes from extremophilic organisms. Here, it could be shown that the PylRS from the psychrophilic organism *Methanococcoides burtonii* (*MburPylRS*) was a superior system in regard to ncAA incorporation efficiency. Especially impressive was the ability to incorporate multiple ncAAs with good yield into a target protein in comparison to the today's widely used PylRS systems. This was not only true for the wild-type *MburPylRS*, even an engineered variant to incorporate S-allyl-L-cysteine experienced almost wild-type activity. An engineered PylRS with wild-type target protein production capabilities has never been described before. Fortunately, I could also show that the known substrate promiscuity of psychrophilic enzymes is also true for *MburPylRS*.

This is very fortunate since the combination of the very high catalytic activity of *MburPylRS* and the greater substrate promiscuity could potentially give rise to a plethora of applications that were previously not possible due to the limitations of the PylRS system mentioned above. Considering, that there are only two other orthogonal translation systems (OTSs) known to come close to wild-type recombinant protein production levels (*Methanocaldococcus janaschii*^{1–3} and *Archaeoglobus fulgidus*, both TyrRS systems), the addition of a third OTS to this group is highly desired. This improves the probability that the number of different ncAAs that can be simultaneously incorporated with good efficiency can be increased. In this context, the most promising improvements can be expected with the *MburPylRS* system in conjunction

with semi-synthetic organisms possessing liberated codons, since the tRNA^{Pyl} anticodon can be freely chosen. This could be a crucial step closer to creating xenobiotic organisms.

To address the limited substrate range of PylRS systems I engineered PylRS variants capable of incorporating an entire library of aliphatic “small-tag” ncAAs. In particular, mutational studies of a specific PylRS, designed to incorporate the shortest non-bulky ncAA (S-allyl-L-cysteine) possible to date were performed to gain insights of the structure activity relationship. Based on this knowledge aliphatic ncAA derivatives were incorporated and also key residues responsible for maintaining orthogonality, while engineering the PylRS for these interesting substrates were determined. Based on the known plasticity of PylRS toward different substrates, this approach further expands the reassignment capacities of this enzyme toward aliphatic amino acids with smaller side chains endowed with valuable functionalities.

II. Zusammenfassung

Der ortsspezifische Einbau von nicht-kanonischen Aminosäuren (ncAAs) hat sich als Technologie etabliert, die es erlaubt maßgeschneidert biologische Systeme zu manipulieren und findet vielfältige Anwendungen in Bereichen der Lebenswissenschaften (Biochemie, synthetische Biologie, Biophysik). Um den genetischen Code mit ncAAs zu erweitern, haben sich orthogonale Paare von Aminoacyl-tRNA-Synthetasen und tRNAs als besonders nützlich erwiesen. Speziell die Pyrrolysyl-tRNA-Synthetase (PylRS), die von methanogenen Archaeen stammt, ist besonders wertvoll, da sie sowohl in Bakterien- als auch in Säugetierzellen eingesetzt werden kann. Leider ist der Anwendungsbereich mittels PylRS umprogrammierten Zellen zurzeit noch relativ limitiert, da nur geringe Ausbeuten von gewünschtem, mit ncAAs modifizierten, Zielprotein erhalten wird. Ein Grund für diese geringen Ausbeuten kann sein, dass die verwendeten *engineerten* Enzyme weniger stabil sind und somit weniger aktives Enzym in den Zellen vorliegt, um das gewünschte Zielprotein herzustellen. Eine weitere Limitierung, ist das zurzeit nur ncAAs die sehr voluminös und/oder lang sind, mittels des PylRS-Systems eingebaut werden können. Diese beiden Nachteile wurden in dieser Arbeit gezielt in Angriff genommen und erfolgreich minimiert. Diese deutlichen Verbesserungen in katalytischer Effizienz und Erhöhung der Substraterkennung macht das ohnehin schon sehr weit verbreitete PylRS-System noch nützlicher und breiter anwendbar. Die Ausbeutenerhöhung wurde mittels zweier Strategien verfolgt. Die erste Strategie bestand darin mehr lösliches und damit auch mehr funktionales Enzym in der Zelle zu erhalten, indem Löslichkeits-Tags aus verschiedenen Organismen an die PylRS fusioniert wurden. Dies hat zu beträchtlichen Zielproteinausbeutenerhöhungen geführt (zwischen 200-540%). Selbst das nicht *engineerte* PylRS-Enzym konnte an Einbaueffizienz gewinnen. In der zweiten Strategie wurde die Nutzung von PylRS-Systemen aus extremophilen Organismen verfolgt. Es konnte

gezeigt werden, dass das PyIRS-System aus dem psychrophilen Organismus *Methanococcoides burtonii* allen anderen PyIRS-Systemen (*MburPyIRS*) weit überlegen ist. Speziell der Einbau von mehreren ncAAs in ein Zielprotein war bemerkenswert effizienter als mit allen bisher verwendeten PyIRS-Systemen. Dies galt nicht nur für das Wildtyp-Enzym, selbst ein *engineertes MburPyIRS*-Enzym, welches S-Allyl-L-cystein (Sac) einbaut, erreichte ähnliche Effizienzen. So etwas konnte noch nie über ein PyIRS-System berichtet werden. Glücklicherweise stellte sich auch die für psychrophile Enzyme bekannte, erhöhte Substratpromiskuität als zutreffend für das *MburPyIRS*-Enzym heraus.

Die Kombination aus erhöhter katalytischer Effizienz und erhöhter Substratpromiskuität ist eine wirklich glückliche Fügung, welche dazu führen könnte, dass eine Reihe von weiteren Anwendungen basierend auf diesem System entwickelt werden könnten, die bis jetzt nicht möglich waren, wegen der oben beschriebenen Limitierungen. Es gibt bis heute nur zwei orthogonale Translationssysteme (OTSs) die mit ncAA modifizierte Zielproteine in ähnlicher Ausbeute herstellen können wie die Wildtyp-Proteine-Äquivalente. Diese stammen von *Methanocaldococcus janaschii*¹⁻³ und *Archaeoglobus fulgidus* (beides TyrRS Systeme). Ein drittes OTS in dieser Riege wäre deshalb äußerst wertvoll, weil somit die Wahrscheinlichkeit erhöht wird, dass mehrere verschiedene ncAAs gleichzeitig und bei guten Ausbeuten eingebaut werden können. In diesem Zusammenhang besteht das größte Entwicklungspotential in der Kombination aus diesen OTS in Verbindung mit semi-synthetischen Organismen, welche Codons besitzen die von ihrer eigentlichen genetischen Information befreit wurden. Diese könnten mit dem PyIRS OTS neu belegt werden, da das Anticodon der tRNA^{PyI} frei gewählt werden kann. Das könnte ein wichtiger Schritt zur Erstellung von xenobiotischen Organismen sein.

Um die Substratlimitierung des PyIRS-Systems zu beheben, wurden Varianten *engineert*, die ncAAs mit kleinen aliphatischen Seitenketten einbauen können. Dies war die folgt möglich. Zuerst wurden Mutationsstudien mit einer PyIRS Variante durchgeführt, die die bis heute die kleinste ncAA einbauen konnte, Sac. Basierend auf den Erkenntnissen dieser Struktur-Wirkungsbeziehungen wurden PyIRS Mutanten entwickelt, die die kleinen aliphatischen ncAAs einbauen konnten. Außerdem konnten Schlüsselpositionen in der PyIRS festgestellt werden die das zukünftige *engineeren* dieses Systems deutlich vereinfachen wird. Es ist äußerst wahrscheinlich, dass sich aufgrund der Plastizität der PyIRS und den hier gewonnenen Erkenntnissen noch viele weitere biotechnologisch nützliche kleinere ncAAs in Zielproteine einbauen lassen werden können.

III. Abbreviations

A	adenine
A₂₈₀	absorption at $\lambda = 280$ nm
AA	amino acid
aaRS	aminoacyl-tRNA synthetase
aa-tRNA	aminoacyl-tRNA
AllocK	<i>N</i> ^ε -Allyloxycarbonyl-L-lysine
Amp	Ampicillin
AMP	adenosine-5'-monophosphate
APS	ammonium persulfate
a.u.	arbitrary units
ATP	adenosine-5'-triphosphate
AzidoK	<i>N</i> ^ε -((2-Azidoethoxy)carbonyl)-L-lysine
BenzK	<i>N</i> ^ε -benzyloxycarbonyl-L-lysine
BockK	<i>N</i> ^ε - <i>tert</i> -butyloxycarbonyl-L-lysine
bp	base pair
Bpa	<i>p</i> -benzoylphenylalanine
C	cytosine
c	concentration
cAA	canonical amino acid
CAT	chloramphenicol acetyltransferase (Cm resistance gene)
Cm	chloramphenicol
CV	column volume
°C	degree Celsius
Da	dalton (1.66018×10^{-24} g)
dH₂O	distilled water
ddH₂O	double distilled water
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
Dopa	3,4-dihydroxyphenylalanine
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EF-Tu	elongation factor Tu
ELP	elastin-like polypeptide
ESI-MS	electrospray ionization mass spectrometry
EtBr	ethidiumbromide
EtOH	ethanol
ϵ_M	molar extinction coefficient

g	gram
G	guanine
GFP	green fluorescent protein
GTP	guanosine-5'-triphosphate
h	hour
His₆	hexahistidine tag
HPLC	high performance liquid chromatography
IMAC	immobilized metal ion affinity chromatography
IPTG	isopropyl-β-D-1-thiogalactopyranoside
Kan	kanamycin
LB	lysogeny broth
M	molar
MBP	maltose binding protein
M_w	molecular weight
min	minutes
mRNA	messenger RNA
ncAA	non-canonical amino acid
Ni-NTA	Ni ²⁺ -nitrilo-triacetic acid
NMM	new minimal medium
oNB	<i>ortho</i> -2-Nitrobenzyl
ONB-Dopa	<i>meta</i> -(<i>ortho</i> -(2-nitrobenzyl))-3,4-dihydroxyphenylalanine
ONBY	<i>ortho</i> -(2-Nitrobenzyl)-L-tyrosine
ori	origin of replication
OTS	orthogonal translation system
PCR	polymerase chain reaction
PhotoK	3'-Azibutyl- <i>N</i> ^ε -carbamoyl-L-lysine
pI	isoelectric point
PMSF	phenylmethylsulfonyl fluoride
PP_i	pyrophosphate
ProK	<i>N</i> ^ε -Propargyloxycarbonyl-L-lysine
PTM	post-translational modification
Pyl	pyrrolysine
RF1	release factor 1
RNA	ribonucleic acid
rpm	revolutions per minute
SCS	stop codon suppression
s.d.	standard deviation
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS polyacrylamide gel electrophoresis

s	seconds
sfGFP	superfolder GFP
Spec	spectinomycin
SmbP	Small metal binding protein
SPI	Selective Pressure Incorporation
SSM	site-saturation mutagenesis
SUMO	small ubiquitin-related modifier
T	thymine
T_m	melting temperature
TRIS	2-amino-2-hydroxymethyl-propane-1,3-diol
tRNA	transfer RNA
U	uracil
wt	wild-type

IV. List of Figures

FIGURE 1: CENTRAL DOGMA OF MOLECULAR BIOLOGY IN 1958 (GRAPHIC FROM 1970).	2
FIGURE 2: THE FIVE NUCLEOBASES WHICH ARE PART OF DNA (A,T,G,C) AND RNA (U INSTEAD T).....	3
FIGURE 3: COLLINEARITY OF THE CODING NUCLEOTIDE SEQUENCES OF DNA AND MRNA AND THE AMINO ACID SEQUENCE OF A POLYPEPTIDE CHAIN, ADAPTED.	4
FIGURE 4: A) CODON WHEEL OF THE GENETIC CODE IN RNA CODONS WITH CAA ABBREVIATED IN THE THREE LETTER CODE.	5
FIGURE 5: GENETIC CODE TABLE WITH TOPOGRAPHIC MAPPING.....	6
FIGURE 6: PROPERTIES OF COMPLEX SYSTEMS.....	7
FIGURE 7: SCHEMATIC REPRESENTATION OF THE PROTEIN BIOSYNTHESIS MACHINERY	9
FIGURE 8: A) TWO-DIMENSIONAL DEPICTION OF A GENERIC tRNA IN THE TYPICAL CLOVERLEAF STRUCTURE.	10
FIGURE 9: SCHEMATIC REPRESENTATION OF A tRNA WITH ITS MOST CHARACTERISTIC RECOGNITION SITES	11
FIGURE 10: REACTION OF tRNA AMINOACYLATION.....	12
FIGURE 11: SCS MECHANISM FOR AN IN-FRAME AMBER STOP-CODON READTHROUGH AND COMPETING REACTIONS.	14
FIGURE 12: FLOWCHART SHOWING THE TIMETABLE OF CHARACTERISTIC NCAAs THAT CAN BE INCORPORATED WITH THE PylRS OTS. THE THIRD WAVE (RED) WAS ACHIEVED IN THIS WORK.....	16
FIGURE 13: STRUCTURAL COMPARISON BETWEEN BACTERIAL PHERS AND ARCHAEAL PylRS.....	17
FIGURE 14: SUPERPOSITION OF THE <i>MmPylRS</i> NTD-tRNA ^{Pyl} COMPLEX (CYAN) AND THE <i>DESULFITOBACTERIUM HAFNIENSE</i> PylRS CTD-tRNA ^{Pyl} COMPLEX (GRAY; PDB ID 2ZNI).	17
FIGURE 15: AN ENERGETIC OVERVIEW OF THE TWO MAIN STRATEGIES FOR ENZYMATIC ADAPTION TO COLD ENVIRONMENTS.	22
FIGURE 16: PROPOSED MECHANISM OF HOW THE SMALL METAL BINDING PROTEIN (SMBP) TAG CONTRIBUTES TO <i>MbPylRS</i> BEING MORE CATALYTICALLY ACTIVE <i>IN VIVO</i> AND INCREASES THE YIELD OF NCAa-MODIFIED TARGET PROTEIN.	25
FIGURE 17: COMPARISON OF OTS EFFICIENCY OF 10 DIFFERENT <i>MbSacsRS</i> CONSTRUCTS MEASURED BY FLUORESCENCE OF INTACT CELLS OF <i>E. COLI</i> BL21(DE3) EXPRESSING THE sfGFP(R2 AMBER) CONSTRUCT.....	28
FIGURE 18: COMPARISON OF OTS EFFICIENCY FOR <i>MbSacsRS</i> FUSION CONSTRUCTS MEASURED BY FLUORESCENCE OF INTACT <i>E. COLI</i> BL21(DE3) CELLS.	29
FIGURE 19: ANALYTICS AND GEL PROFILES OF THE INCORPORATION EXPERIMENTS.....	31
FIGURE 20: SDS-PAGE ANALYSIS OF SAC INCORPORATION INTO AMILCP (26.7 kDa, A) AND PDZ (11.7 kDa, B).	32
FIGURE 21: SDS-PAGE ANALYSIS OF THE SOLUBLE (SF) AND INSOLUBLE (ISF) PROTEIN FRACTIONS OF CELL LYSATES.	33
FIGURE 22: CONCENTRATION-DEPENDENT PROTEIN PRODUCTION FOR DIFFERENT <i>MbPylRS</i> /NCAa COMBINATIONS.	34
FIGURE 23: CONCENTRATION-DEPENDENT UNNATURAL PROTEIN PRODUCTION FOR DIFFERENT <i>MbPylRS</i> /NCAa COMBINATIONS. ...	36
FIGURE 24: CONCENTRATION-DEPENDENT SAC INCORPORATION COMPARISON.	38
FIGURE 25: A PORTION OF THE DIAGRAM SHOWN IN FIGURE 24 (“ZOOM IN ON THE ALLYL MERCAPTAN PART”).	39
FIGURE 26: SURVEY OF AMINO ACIDS USED IN THIS STUDY.	42
FIGURE 27: MICROENVIRONMENT OF ACTIVE SITES DERIVED FROM THE CRYSTAL STRUCTURES OF <i>MmPylRS</i> AND THE <i>MmOmERS</i> MUTANT.	45
FIGURE 28: COMPARISON OF SAC INCORPORATION EFFICIENCY FOR <i>MbPylRS</i> CONSTRUCTS	46
FIGURE 29: COMPARISON OF ALIPHATIC NCAa INCORPORATION EFFICIENCY FOR <i>MbPylRS</i> (C313W) CONSTRUCTS MUTATED AT POSITION W382.	47
FIGURE 30: COMPARISON OF ALIPHATIC NCAa INCORPORATION EFFICIENCY FOR <i>MbPylRS</i> (C313W) CONSTRUCTS MUTATED AT POSITION N311.	48
FIGURE 31: COMPARISON OF ALIPHATIC NCAa INCORPORATION EFFICIENCY FOR A) <i>MbPylRS</i> (N311M:C313W) AND B) <i>MbPylRS</i> (N311Q:C313W) CONSTRUCTS BOTH MUTATED AT POSITION V366.	50
FIGURE 32: APPROXIMATE-MAXIMUM-LIKELIHOOD PHYLOGENETIC TREE BASED ON ALL <i>METHANOSARCINALES</i> AND <i>METHANOMASSILIICOCALES</i> PylRS SEQUENCES	55
FIGURE 33: SIMILARITY MATRIX OF THE C-TERMINUS OF KNOWN <i>METHANOSARCINA</i> PylRSs.....	56
FIGURE 34: AVERAGE AMINO ACID DISTRIBUTION IN PERCENT AND DESCRIPTIVE INDICATORS OF <i>METHANOSARCINAE</i> PylRS VARIANTS.	58
FIGURE 35: AVERAGE AMINO ACID DISTRIBUTION IN PERCENT AND DESCRIPTIVE INDICATORS OF <i>METHANOSARCINAE</i> PylRS VARIABLE REGIONS (LINKERS).	59
FIGURE 36: PYRROLYSINE (1A) AND DERIVATIVES USED IN THIS CHAPTER.	61
FIGURE 37: HEATMAPS OF THE INCORPORATION EFFICIENCIES OF THE SEVEN SELECTED PylRSs WITH DIFFERENT NCAa CONCENTRATIONS.....	62
FIGURE 38: CONCENTRATION-DEPENDENT UNNATURAL PROTEIN PRODUCTION WITH <i>MmPylRS</i> (A, C) AND <i>MbURPylRS</i> (B, D) FOR DIFFERENT NCAa/REPORTER CONSTRUCT COMBINATIONS.	64
FIGURE 39: CONCENTRATION-DEPENDENT UNNATURAL PROTEIN PRODUCTION WITH <i>MmPylRS</i> (A, C) AND <i>MbURPylRS</i> (B, D) FOR DIFFERENT NCAa/REPORTER CONSTRUCT COMBINATIONS.	65
FIGURE 40: PHE/TYR DERIVATIVES USED IN THIS STUDY.	68

FIGURE 41: HEATMAPS OF INCORPORATION EFFICIENCIES OF SELECTED PYLRS MUTANTS WITH 10 mM ADDED NCAA.....	69
FIGURE 42: HEATMAPS OF INCORPORATION EFFICIENCIES OF THE SELECTED PYLRSs WITH DIFFERENT NCAA CONCENTRATIONS.	70
FIGURE 43: HEATMAPS OF INCORPORATION EFFICIENCIES OF THE SELECTED PYLRSs WITH DIFFERENT NCAA CONCENTRATIONS.	72
FIGURE 44: CONCENTRATION-DEPENDENT UNNATURAL PROTEIN PRODUCTION WITH A) SmbP- <i>MbSACRS</i> , <i>MMSACRS</i> , <i>MtSACRS</i> (TM-1) AND <i>MBURsACRS</i> . B) <i>MtPYLRS</i> (TM-1) WITH AND WITHOUT SmbP-TAG.....	73
FIGURE 45: CONCENTRATION-DEPENDENT UNNATURAL PROTEIN PRODUCTION WITH A) <i>MBURPYLRS</i> MUTANTS AND SUPPLEMENTED SAC (1). B) <i>MBURPYLRS</i> MUTANTS AND SUPPLEMENTED SPROC (27).	74
FIGURE 46: CONCENTRATION-DEPENDENT UNNATURAL PROTEIN PRODUCTION FOR SACRS MUTANTS AND SUPPLEMENTED WITH A) SAC (1). B) SPROC (27).	74
FIGURE 47: CONCENTRATION-DEPENDENT UNNATURAL PROTEIN PRODUCTION WITH SmbP- <i>MbPYLRS</i> (A , C) AND <i>MBURPYLRS</i> (B , D) FOR DIFFERENT NCAA/REPORTER CONSTRUCT COMBINATIONS.	75
FIGURE 48: CONCENTRATION-DEPENDENT UNNATURAL PROTEIN PRODUCTION WITH SmbP- <i>MbPYLRS</i> (A , C) AND <i>MBURPYLRS</i> (B , D) FOR DIFFERENT NCAA/REPORTER CONSTRUCT COMBINATIONS.	76
FIGURE 49: APPROXIMATELY-MAXIMUM-LIKELIHOOD PHYLOGENETIC TREE BASED ON ALL <i>METHANOSARCINALES</i> AND <i>METHANOMASSILIICOCCALES</i> UNIQUE TRNA ^{PYL} SEQUENCES.....	77
FIGURE 50: SELECTION OF TRNA ^{PYL} (FROM FIGURE 202 AND FIGURE 203).	78
FIGURE 51: CONCENTRATION-DEPENDENT UNNATURAL PROTEIN PRODUCTION WITH <i>MBURPYLRS</i> FOR DIFFERENT TRNA(ORGANISM ABBREVIATION)/NCAA COMBINATIONS.....	79
FIGURE 52: CONCENTRATION-DEPENDENT UNNATURAL PROTEIN PRODUCTION WITH A) <i>MJONB-DOPARS</i> , B) <i>MJONBYRS</i> , C) <i>MBURPYLRS</i> AND D) <i>MBURsACRS</i>	81
FIGURE 53: NON-CANONICAL AMINO ACIDS USED IN THIS CHAPTER.	83
FIGURE 54: THE TWO ELP MONOMERS USED IN THIS STUDY.....	84
FIGURE 55: A) ALTERNATING AMBER STOP-CODON SEQUENCE OF ELP(8X AMBER). B) SEQUENCE OF THE ELP(5X AMBER) CONSTRUCT.	85
FIGURE 56: ELP CONSTRUCT DESIGN STRATEGY VIA THE RECURSIVE DIRECTIONAL LIGATION (RDL) APPROACH.	85
FIGURE 57: CREATING AN ELP-SFGFP REPORTER CONSTRUCT USING ELP(10X AMBER) AS AN EXAMPLE.....	86
FIGURE 58: EXAMINATION OF ONB-DOPA OTS EFFICIENCY BY DIFFERENT MUTATIONS IN THE BINDING POCKET OF MJTYRRS.	87
FIGURE 59: COMPARISON OF OTS EFFICIENCY OF THE CONSTRUCTS INDICATED IN THE TEXT.	88
FIGURE 60: COMPARISON OF OTS EFFICIENCY OF THE INDICATED CONSTRUCTS.....	89
FIGURE 61: ESI-MS OF ELP(5x 51)-SFGFP.....	90
FIGURE 62: WORKFLOW FOR GENE DELETIONS OF THE B-95.ΔA STRAIN.	92
FIGURE 63: VERIFICATION OF DELETED GENES VIA AGAROSE GEL ELECTROPHORESIS.....	93
FIGURE 64: MASS SPECTROMETRIC PROFILES OF PROTEIN EXPRESSED IN THE <i>E. COLI</i> B-95.ΔA NK53 STRAIN WITH ATTENUATED NITROREDUCTASE ACTIVITY.....	94
FIGURE 65: TIME-RESOLVED CHROMOPHORE MATURATION OF AMILCP(Y63ONBY) DETECTED BY AN ABSORPTION ASSAY AT 588 NM.	95
FIGURE 66: <i>L. LACTIS</i> INDICATOR ASSAY USING CELL LYSATES CONTAINING NISIN PRO-PEPTIDE MODIFIED WITH ONBY (41).....	96
FIGURE 67: HYPOTHETICAL LIGHT-INDUCIBLE CHEMICAL REACTION INVOLVING THE RELEASE OF H ₂ O FROM THE SCAFFOLD OF 54 LEADING TO A REACTIVE ALDEHYDE (56).	97
FIGURE 68: COMPARISON OF OTS EFFICIENCY OF THE INDICATED CONSTRUCTS.....	97
FIGURE 69: COMPARISON OF OTS EFFICIENCY OF THE CONSTRUCTS INDICATED IN THE TEXT.	98
FIGURE 70: OTS EFFICIENCY COMPARISON OF THREE DIFFERENT <i>MbSACRS</i> CONSTRUCTS.	143
FIGURE 71: OTS EFFICIENCY COMPARISON OF <i>MbPYLRS</i> (N311G:C313G) CO-EXPRESSION WITH AND WITHOUT SmbP-TAG.	144
FIGURE 72: OTS EFFICIENCY COMPARISON OF <i>MbPYLRS</i> (N311G:C313G) CO-EXPRESSION WITH AND WITHOUT SmbP-TAG.	144
FIGURE 73: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR <i>MbPYLRS</i> WITH AND WITHOUT SmbP-TAG.	145
FIGURE 74: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR DIFFERENT <i>MbPYLRS</i> /NCAA COMBINATIONS.....	145
FIGURE 75: COMPARISON OF RIBOSOMAL INCORPORATION EFFICIENCY OF <i>MbSACRS</i> WITH FOUR DIFFERENT NCAAAs (2 , 3 , 5 , 7 , 10) (10 mM) AND CONTROLS WITHOUT NCAA SUPPLEMENTATION.	146
FIGURE 76: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR <i>MbPYLRS</i> VARIANTS.	146
FIGURE 77: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR <i>MmPYLRS</i> VARIANTS.	147
FIGURE 78: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR <i>MbPYLRS</i> VARIANTS.	147
FIGURE 79: COMPARISON OF ALIPHATIC NCAA (2 , 3 , 5 , 7 , 10) INCORPORATION EFFICIENCY FOR <i>MbPYLRS</i> (N311M:C313W:W382H).	148
FIGURE 80: COMPARISON OF ALIPHATIC NCAA (4 , 6 , 8 , 11 , 12 , 13) INCORPORATION EFFICIENCY FOR <i>MbPYLRS</i> VARIANTS.	148
FIGURE 81: COMPARISON OF ALIPHATIC NCAA (4 , 6 , 8 , 11 , 12 , 13) INCORPORATION EFFICIENCY FOR <i>MbPYLRS</i> VARIANTS.	149
FIGURE 82: COMPARISON OF (S)-2-AMINO-5,5'-AZI-HEXANOIC ACID (20), (S)-2-AMINO-4-METHYLPENT-4-ENOIC ACID (21) AND (S)-2-AMINO-3-CYCLOPROPYLPROPANOIC ACID (9) INCORPORATION EFFICIENCY FOR <i>MbPYLRS</i> VARIANTS.	149

FIGURE 83: COMPARISON OF <i>S</i> -ALLYL-L-CYSTEINE (SAC, 1) INCORPORATION EFFICIENCY FOR <i>MbPYLRS</i> (N311M:C313W) VARIANTS MUTATE AT POSITION V366.	150
FIGURE 84: COMPARISON OF <i>S</i> -ALLYL-L-CYSTEINE (SAC, 1) INCORPORATION EFFICIENCY FOR <i>MbPYLRS</i> (N311Q:C313W) VARIANTS MUTATE AT POSITION V366.	150
FIGURE 85: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR <i>MbPYLRS</i> (N311M:C313W:V366A).	151
FIGURE 86: COMPARISON OF NCAA (11, 12, 13, 14, 15, 16, 17) INCORPORATION EFFICIENCY FOR <i>MbPYLRS</i> (N311M:C313W) IN DIFFERENT <i>E. COLI</i> HOST STRAINS.	151
FIGURE 87: COMPARISON OF NCAA (18, 19, 22, 23, 24, 25) INCORPORATION EFFICIENCY FOR <i>MbPYLRS</i> CONSTRUCTS IN DIFFERENT <i>E. COLI</i> HOST STRAINS.	152
FIGURE 88: COMPARISON OF NCAA (18, 19, 22, 23, 24, 25) INCORPORATION EFFICIENCY FOR <i>MbPYLRS</i> CONSTRUCTS IN DIFFERENT <i>E. COLI</i> HOST STRAINS.	152
FIGURE 89: COMPARISON OF NCAA (18, 19, 22, 23, 24, 25) INCORPORATION EFFICIENCY FOR <i>MbPYLRS</i> CONSTRUCTS IN DIFFERENT <i>E. COLI</i> HOST STRAINS.	153
FIGURE 90: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS VARIANTS.	153
FIGURE 91: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS VARIANTS.	154
FIGURE 92: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR SIX DIFFERENT PYLRS VARIANTS.	154
FIGURE 93: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS VARIANTS.	155
FIGURE 94: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS VARIANTS.	156
FIGURE 95: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR SIX DIFFERENT PYLRS VARIANTS.	156
FIGURE 96: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS VARIANTS.	157
FIGURE 97: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS VARIANTS.	157
FIGURE 98: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR SIX DIFFERENT PYLRS VARIANTS.	158
FIGURE 99: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS VARIANTS.	158
FIGURE 100: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS VARIANTS.	159
FIGURE 101: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR SIX DIFFERENT PYLRS VARIANTS.	159
FIGURE 102: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS VARIANTS.	160
FIGURE 103: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS VARIANTS.	160
FIGURE 104: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR SIX DIFFERENT PYLRS VARIANTS.	161
FIGURE 105: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS VARIANTS.	161
FIGURE 106: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS VARIANTS.	162
FIGURE 107: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR SIX DIFFERENT PYLRS VARIANTS.	162
FIGURE 108: COMPARISON OF NCAA (39, 40, 44, 45, 46, 49, 50) INCORPORATION EFFICIENCY FOR PYLRS DOUBLE ALA AND DOUBLE GLY MUTANTS.	163
FIGURE 109: COMPARISON OF NCAA (38, 47, 48, 51, 52) INCORPORATION EFFICIENCY FOR PYLRS DOUBLE ALA AND DOUBLE GLY MUTANTS.	163
FIGURE 110: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS DOUBLE ALA MUTANTS.	164
FIGURE 111: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS DOUBLE ALA MUTANTS.	164
FIGURE 112: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS DOUBLE ALA MUTANTS.	165
FIGURE 113: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS DOUBLE ALA MUTANTS.	165
FIGURE 114: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT PYLRS DOUBLE ALA MUTANTS.	166
FIGURE 115: COMPARISON OF NCAA (10, 7, 5, 2, 3, 8, 6) INCORPORATION EFFICIENCY FOR DIFFERENT PYLRS CONSTRUCTS.	166
FIGURE 116: COMPARISON OF NCAA (14, 15, 16, 13, 12, 11, 21) INCORPORATION EFFICIENCY FOR DIFFERENT PYLRS CONSTRUCTS.	167
FIGURE 117: COMPARISON OF NCAA (18, 19, 1, 27, 25, 23, 9) INCORPORATION EFFICIENCY FOR DIFFERENT PYLRS CONSTRUCTS.	167
FIGURE 118: COMPARISON OF NCAA (10, 7, 5, 2, 3, 8, 6) INCORPORATION EFFICIENCY FOR DIFFERENT PYLRS CONSTRUCTS.	168
FIGURE 119: COMPARISON OF NCAA (14, 15, 16, 13, 12, 11, 21) INCORPORATION EFFICIENCY FOR DIFFERENT PYLRS CONSTRUCTS.	168
FIGURE 120: COMPARISON OF NCAA (18, 19, 1, 27, 25, 23, 9) INCORPORATION EFFICIENCY FOR DIFFERENT PYLRS CONSTRUCTS.	169
FIGURE 121: COMPARISON OF NCAA (10, 7, 5, 2, 3, 8, 6) INCORPORATION EFFICIENCY FOR DIFFERENT PYLRS CONSTRUCTS.	169
FIGURE 122: COMPARISON OF NCAA (14, 15, 16, 13, 12, 11, 21) INCORPORATION EFFICIENCY FOR DIFFERENT PYLRS CONSTRUCTS.	170
FIGURE 123: COMPARISON OF NCAA (18, 19, 1, 27, 25, 23, 9) INCORPORATION EFFICIENCY FOR DIFFERENT PYLRS CONSTRUCTS.	170
FIGURE 124: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR TWO DIFFERENT PYLRS DOUBLE GLY MUTANTS.	171
FIGURE 125: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR TWO DIFFERENT PYLRS DOUBLE GLY MUTANTS.	171
FIGURE 126: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR TWO DIFFERENT PYLRS DOUBLE GLY MUTANTS.	172

FIGURE 127: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR TWO DIFFERENT PylRS DOUBLE GLY MUTANTS.....	172
FIGURE 128: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR TWO DIFFERENT PylRS DOUBLE GLY MUTANTS.....	173
FIGURE 129: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR TWO DIFFERENT PylRS DOUBLE GLY MUTANTS.....	173
FIGURE 130: COMPARISON OF NCAA INCORPORATION EFFICIENCY FOR DIFFERENT <i>MburPylRS</i> CONSTRUCTS.....	174
FIGURE 131: COMPARISON OF NCAA INCORPORATION EFFICIENCY FOR DIFFERENT <i>MburPylRS</i> CONSTRUCTS.....	174
FIGURE 132: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR FOUR DIFFERENT <i>MjTyrrs</i> /NCAA COMBINATIONS.	175
FIGURE 133: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR THREE DIFFERENT <i>MburPylRSRS</i> /NCAA COMBINATIONS.	176
FIGURE 134: PRESCREENING OF SEVERAL NCAAAs WITH <i>MjOnbyrs</i> AND <i>MjPcnfrs</i>	177
FIGURE 135: CONCENTRATION DEPENDENT PROTEIN PRODUCTION FOR TWO DIFFERENT <i>MjTyrrs</i> /NCAA COMBINATIONS.	177
FIGURE 136: SCREENING OF THE AZU-ALANINE LIBRARY. A) <i>P</i> -ONB-ALANINE (54) B) OR <i>M</i> -ONB-ALANINE (55).....	178
FIGURE 137: SCREENING OF THE DOPA LIBRARY. A) <i>M</i> -ONB-DOPA (51). B) OR <i>P</i> -ONB-DOPA (53).....	178
FIGURE 138: SDS-PAGE ANALYSIS OF PURIFIED His ₆ -SUMO-sfGFP(R2SAC:N39SAC:K101SAC) REPORTER EXPRESSED IN <i>E. COLI</i> BL21(DE3) WITH SMBP- <i>MbSACRS</i> CO-EXPRESSION.....	179
FIGURE 139: EXAMPLE OF AN ANALYTICAL AGAROSE GEL (1%) STAINED WITH EtBr. DIGESTION OF PET28A_ELP(5-60X AMBER-SFGFP) REPORTER CONSTRUCTS WITH XbaI AND BglI.	179
FIGURE 140: HARVESTED <i>E. COLI</i> BL21(DE3) CELLS AFTER SHAKE FLASK CULTIVATION FOR PRODUCTION OF SAC-MODIFIED SFGFP REPORTER PROTEIN.	180
FIGURE 141: OD ₆₀₀ -NORMALIZED, PELLETED <i>E. COLI</i> BL21(DE3) CELLS FROM AMILCP TARGET PROTEIN PRODUCTION WITH SITE-SPECIFIC NCAA INSTALLATION.	180
FIGURE 142: PHOTO-INDUCED CHROMOPHORE MATURATION OF AMILCP(Y63ONBY).	180
FIGURE 143: : DECONVOLUTED ESI-MS SPECTRA OF SUMO-sfGFP(R2SAC:N39SAC:K101SAC)-His ₆ PRODUCED IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i>	181
FIGURE 144: DECONVOLUTED ESI-MS SPECTRUM OF His ₆ -SUMO-sfGFP(R2(1))-STREP PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W:V366A).	181
FIGURE 145: DECONVOLUTED ESI-MS SPECTRUM OF His ₆ -SUMO-sfGFP(R2(2))-STREP PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311Q:C313W).	182
FIGURE 146: DECONVOLUTED ESI-MS SPECTRUM OF His ₆ -SUMO-sfGFP(R2(3))-STREP PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W:V366A).	182
FIGURE 147: DECONVOLUTED ESI-MS SPECTRUM OF His ₆ -SUMO-sfGFP(R2(4))-STREP PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W:V366A).	183
FIGURE 148: DECONVOLUTED ESI-MS SPECTRUM OF His ₆ -SUMO-sfGFP(R2(5))-STREP PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W).....	183
FIGURE 149: DECONVOLUTED ESI-MS SPECTRUM OF His ₆ -SUMO-sfGFP(R2(6))-STREP PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W:V366K).	184
FIGURE 150: DECONVOLUTED ESI-MS SPECTRUM OF His ₆ -SUMO-sfGFP(R2(7))-STREP PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W).....	184
FIGURE 151: DECONVOLUTED ESI-MS SPECTRUM OF His ₆ -SUMO-sfGFP(R2(8))-STREP PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W).....	185
FIGURE 152: DECONVOLUTED ESI-MS SPECTRUM OF His ₆ -SUMO-sfGFP(R2(9))-STREP PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W).....	185
FIGURE 153: DECONVOLUTED ESI-MS SPECTRUM OF His ₆ -SUMO-sfGFP(R2(10))-STREP PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W:V366A).	186
FIGURE 154: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(R2(11))-His ₆ PRODUCTION IN <i>E. COLI</i> C321.ΔA.EXP(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W).....	186
FIGURE 155: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(R2(12))-His ₆ PRODUCTION IN <i>E. COLI</i> C321.ΔA.EXP(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W).....	187
FIGURE 156: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(R2(13))-His ₆ PRODUCTION IN <i>E. COLI</i> C321.ΔA.EXP(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W).....	187
FIGURE 157: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(R2(14))-His ₆ PRODUCTION IN <i>E. COLI</i> JX33(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W).....	188
FIGURE 158: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(R2(15))-His ₆ PRODUCTION IN <i>E. COLI</i> C321.ΔA.EXP(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W).....	188
FIGURE 159: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(R2(16))-His ₆ PRODUCTION IN <i>E. COLI</i> C321.ΔA.EXP(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W).....	189
FIGURE 160: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(R2(17))-His ₆ PRODUCTION IN <i>E. COLI</i> C321.ΔA.EXP(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311M:C313W).....	189
FIGURE 161: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(R2(18))-His ₆ PRODUCTION IN <i>E. COLI</i> C321.ΔA.EXP(DE3) WITH CO-EXPRESSION OF SMBP- <i>MbSACRS</i> (N311Q:C313W:V366K).	190

FIGURE 162: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(ncAA))-His ₆ PRODUCTION IN <i>E. COLI</i> C321.ΔA.EXP(DE3) WITH CO-EXPRESSION OF SMBP-MbSACRS(N311Q:C313W:V366K).	190
FIGURE 163: DECONVOLUTED ESI-MS SPECTRUM OF His ₆ -SUMO-SFGFP(R2(20))-STREP PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP-MbSACRS(N311M:C313W:V366A).	191
FIGURE 164: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(21))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP-MbSACRS(N311M:C313W).	191
FIGURE 165: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(22))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP-MbSACRS(N311M:C313W).	192
FIGURE 166: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(23))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP-MbSACRS(N311Q:C313W).	192
FIGURE 167: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(24))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP-MbSACRS(N311M:C313W:V366K).	193
FIGURE 168: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(25))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP-MbSACRS(N311Q:C313W).	193
FIGURE 169: DECONVOLUTED ESI-MS SPECTRUM OF His ₆ -SUMO-SFGFP(R2(26))-STREP PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP-MbSACRS(N311M:C313W:V366A).	194
FIGURE 170: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(28))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF SMBP-MbSACRS(N311M:C313W:V366A:W382N).	194
FIGURE 171: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(1))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MBURSAACRS(S379T).	195
FIGURE 172: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(3x 1)-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MBURSAACRS(S379T).	195
FIGURE 173: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(5x 1)-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MBURSAACRS(S379T).	196
FIGURE 174: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(31))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MMPYLRS.	196
FIGURE 175: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(31))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MBURPYLRS.	197
FIGURE 176: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(34))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MMPYLRS.	197
FIGURE 177: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(34))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MBURPYLRS.	198
FIGURE 178: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(40))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MMPYLRS.	198
FIGURE 179: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(40))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MBURPYLRS.	199
FIGURE 180: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(45))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MMPYLRS.	199
FIGURE 181: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(45))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MBURPYLRS.	200
FIGURE 182: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(47))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MMPYLRS.	200
FIGURE 183: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(47))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MBURPYLRS.	201
FIGURE 184: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(39))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MMPYLRS.	201
FIGURE 185: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(39))-His ₆ PRODUCTION IN <i>E. COLI</i> BL21(DE3) WITH CO-EXPRESSION OF MBURPYLRS.	202
FIGURE 186: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(27))-His ₆ PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF SMBP-MbSACRS(S382T).	202
FIGURE 187: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(3x(27))-His ₆ PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF SMBP-MbSACRS(S382T).	203
FIGURE 188: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(5x(27))-His ₆ PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF SMBP-MbSACRS(S382T).	203
FIGURE 189: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(R2(27))-His ₆ PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF MBURSAACRS(S379T).	204
FIGURE 190: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-SFGFP(3x(27))-His ₆ PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF MBURSAACRS(S379T).	204

FIGURE 191: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(5x(27))-His ₆ PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF <i>MBUR</i> SACRS(S379T).	205
FIGURE 192: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP-His ₆ PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF <i>SMBP-MBS</i> SACRS(S382T).....	205
FIGURE 193: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(R2(1))-His ₆ PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF <i>SMBP-MBS</i> SACRS(S382T).....	206
FIGURE 194: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(3x(1))-His ₆ PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF <i>SMBP-MBS</i> SACRS(S382T).....	206
FIGURE 195: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(5x(1))-His ₆ PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF <i>SMBP-MBS</i> SACRS(S382T).....	207
FIGURE 196: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP-His ₆ PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF <i>MBUR</i> SACRS(S379T).	207
FIGURE 197: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(R2(1))-His ₆ PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF <i>MBUR</i> SACRS(S379T).	208
FIGURE 198: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(3x(1))-His ₆ PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF <i>MBUR</i> SACRS(S379T).	208
FIGURE 199: DECONVOLUTED ESI-MS SPECTRUM OF SUMO-sfGFP(5x(1))-His ₆ PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF <i>MBUR</i> SACRS(S379T).	209
FIGURE 200: DECONVOLUTED ESI-MS SPECTRUM OF AMILCP(Y66(ONBY))-His ₆ . PROTEIN PRODUCTION IN <i>E. COLI</i> B95.ΔA WITH CO-EXPRESSION OF <i>MJONB-DOPARS</i>	209
FIGURE 201: DECONVOLUTED ESI-MS SPECTRUM OF AMILCP(Y66(ONBY))-His ₆ . PROTEIN PRODUCTION IN <i>E. COLI</i> B95.ΔA NK53 WITH CO-EXPRESSION OF <i>MJONB-DOPARS</i>	210
FIGURE 202: <i>PSIJ8</i> GENE DELETION VECTOR.....	211
FIGURE 203: PET-28A TARGET PROTEIN EXPRESSION VECTOR.....	211
FIGURE 204: PTECH ORTHOGONAL TRANSLATION SYSTEM SETUP.	212
FIGURE 205: PULTRA ORTHOGONAL TRANSLATION SYSTEM SETUP.....	212
FIGURE 206: UNIQUE tRNA ^{Pyl} SEQUENCES.	213
FIGURE 207: UNIQUE tRNA ^{Pyl} SEQUENCES.	214
FIGURE 208: SIMILARITY MATRIX OF THE COMPLETE PYLRS N- AND C-TERMINUS, WITHOUT THE LINKER.	228
FIGURE 209: AMINO ACID DISTRIBUTION IN PERCENT AND DESCRIPTIVE INDICATORS OF <i>METHANOSARCINA</i> PYLRS VARIANTS.	228
FIGURE 210: AMINO ACID DISTRIBUTION IN PERCENT AND DESCRIPTIVE INDICATORS OF <i>METHANOSARCINA</i> PYLRS VARIABLE REGIONS (LINKER).	229
FIGURE 211: AMINO ACID DISTRIBUTION IN PERCENT AND DESCRIPTIVE INDICATORS OF <i>METHANOCOCCOIDES</i> PYLRS VARIANTS. ..	229
FIGURE 212: AMINO ACID DISTRIBUTION IN PERCENT AND DESCRIPTIVE INDICATORS OF <i>METHANOCOCCOIDES</i> PYLRS VARIABLE REGIONS (LINKER).....	230
FIGURE 213: AMINO ACID DISTRIBUTION IN PERCENT AND DESCRIPTIVE INDICATORS OF <i>METHANOHALOPHILUS</i> PYLRS VARIANTS. .	230
FIGURE 214: AMINO ACID DISTRIBUTION IN PERCENT AND DESCRIPTIVE INDICATORS OF <i>METHANOHALOPHILUS</i> PYLRS VARIABLE REGIONS (LINKER).....	231
FIGURE 215: AMINO ACID DISTRIBUTION IN PERCENT AND DESCRIPTIVE INDICATORS OF <i>METHANOLOBUS</i> PYLRS VARIANTS.....	231
FIGURE 216: AMINO ACID DISTRIBUTION IN PERCENT AND DESCRIPTIVE INDICATORS OF <i>METHANOLOBUS</i> PYLRS VARIABLE REGIONS (LINKER).	232
FIGURE 217: ALIGNMENT OF <i>NFSB</i> CONTAINING TARGET LOCUS.....	232

V. List of Tables

TABLE 1: CLASSIFICATION OF AARSS.	12
TABLE 2: ACTIVATION PARAMETERS OF PSYCHRO-, MESO- AND THERMOPHILIC α -AMYLASES AT 10°C.....	21
TABLE 3: LIST OF FUSION TAGS USED FOR SOLUBILITY ENHANCEMENT OF <i>MbSACRS*</i> WITH BRIEF DESCRIPTIONS, MOLECULAR WEIGHT (MW) AND REFERENCES.....	27
TABLE 4: YIELDS OF NCAA-MODIFIED TARGET PROTEIN AS MEASURE FOR OTS EFFICIENCY WITH 2 MM SAC PROVIDED IN THE GROWTH MEDIUM.....	30
TABLE 5: OPTIMAL REPORTER PROTEIN PRODUCTION SETUP, CALCULATED AND OBSERVED MOLECULAR WEIGHTS OF THE REPORTER PROTEINS HIS ₆ -SUMO-SFGFP(R2AA)-STREP (A) / SUMO-SFGFP(R2AA)-HIS ₆ (B) AND PROTEIN PRODUCTION YIELDS PER LITER OF CULTURE. THE MASSES WERE DETERMINED BY ESI-MS OF INTACT PROTEINS.....	52
TABLE 6: LIST OF ALL KNOWN METHANOSARCINA PYLRS AND THE ADDITIONAL PSYCHROPHILIC PYLRS USED IN THIS STUDY	60
TABLE 7: PYLRS PROMISCUITY ESTIMATIONS WITH 9 MM NCAA CONCENTRATION.	67
TABLE 8: SETUP OF PROTEIN PRODUCTION PLATFORM WITH CALCULATED AND OBSERVED MOLECULAR WEIGHTS OF REPORTER PROTEINS.....	82
TABLE 9: ESI-MS ANALYTICS OF ELP(5xONB-DOPA)-SFGFP-HIS ₆	90
TABLE 10: THE SIX GENES FROM THE <i>E. COLI</i> GENOME WITH KNOWN NITROREDUCTASE ACTIVITY DELETED FROM THE GENOME OF THE EXPRESSION HOST USED IN THIS STUDY.....	91
TABLE 11. AMINO ACIDS USED IN THIS WORK.	103
TABLE 12: GENOMES USED FOR EXTRACTION OF PYLRS AND TRNA ^{Pyl} SEQUENCES	226

1. Introduction

1.1. Protein Biosynthesis and the Link to Genetics: Historic Context

From about 1750 onwards, the study of proteins/enzymes was one of the main targets of physiological chemists.⁴ SUMNER performed one of the most crucial experiments for highlighting the central role of proteins in living organisms. He showed that the enzyme Urease was a protein.⁵ In addition, the pepsin experiments of NORTHROP supported the hypothesis that catalytic activity is a property of enzymes.⁴ Until then, it was assumed that enzymes don't necessarily have to be proteins. Further knowledge accumulation of proteins emphasized their role as the key players of life.

In the early 19th century, it was recognized that proteins were composed of L- α -amino acids (AA).⁶ Even though by the end of the century most of the canonical AAs (cAAs) were known, it was not clear how these building blocks were linked.⁷ There was generally disagreement as to whether a protein is a chemically defined structure (macromolecule) or a colloid (aggregate of lower molecular weight substances).⁷ FISCHER and HOFMEISTER both independently proposed the "peptide theory" in 1902, stating that proteins are polymers composed of cAAs covalently linked by a peptide bond.^{8,9} The work of SVEDBERG in 1929, particularly his ultracentrifugation experiments¹⁰ and the work of SANGER in 1951/53 with the elucidation of the insulin protein sequence^{11,12}, manifested the conclusion that proteins are chemically distinct structures with a defined cAA sequence.

In 1928, GRIFFITH showed that there is a certain substance within dead bacteria which contains hereditary information when he performed the first bacterial transformation.^{13,14} In 1944, AVERY confirmed the previous experiments and revealed that the unknown substrate was deoxyribonucleic acid (DNA) which was later also confirmed by HERSHEY and CHASE.¹⁴⁻¹⁶ This led CRICK to predict that protein biosynthesis and genetic information have to be linked, specifically the DNA sequence with the protein sequence.^{17,18} By including ribonucleic acid (RNA) in the previous prediction, the central dogma of molecular biology was derived (**Figure 1**). This dogma was further refined in 1970 to circumvent critique which was often based on misconceptions.¹⁹ CRICK later stated that he did not want to use the word dogma in its semantically correct form, but rather to emphasize the importance of his central hypothesis.

"I just didn't know what dogma meant. And I could just as well have called it the 'Central Hypothesis', or-you know. Which is what I meant to say. Dogma was just a catch phrase"²⁰.

The theory postulated by CRICK is still valid today and makes up the foundation of molecular biology and genetics. However, some extensions have been made to cover additional layers of information encoded in the DNA, not just the AA sequence.

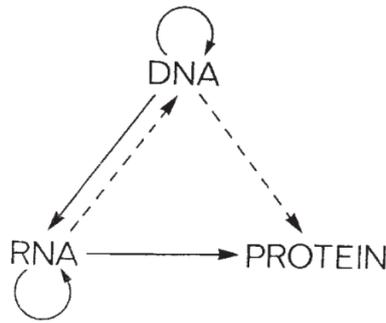


Figure 1: Central dogma of molecular biology in 1958 (graphic from 1970). Solid arrows show the predicted flow of genetic information. Dashed lines show the theoretically possible genetic information flow. The possible genetic information flows not shown with an arrow were postulated as non-existent.¹⁹

One example is the additional information in synonymous codons which contain information for protein folding or regulatory purposes (**see chapter 1.2.1**).^{21–24} Other extensions involve the flow of information from RNA to DNA in nature via reverse transcription.²⁵ The transduction of conformational information from certain proteins (prions) to proteins in very limited but nevertheless important circumstances (e.g., Creutzfeldt–Jakob disease) has also been observed.^{26,27}

It is noteworthy to mention that the journey of DNA elucidation and heredity began decades before CRICK'S first theory of how proteins are made.^{28–30} In 1866, the influential German biologist HEACKEL proposed that the inner part of the cell, the nucleus, is responsible for the inheritance of traits of organisms.³¹ This popularized research on the cell nucleus. In 1869, MIESCHER isolated a substance from pus cells which was distinct from proteins.³² He assumed that the substance originated from the cell nucleus and named it *nuclein*, now known as DNA. He is recognized as the first person to have isolated DNA³³ although there are strong indicators that other researchers before him isolated DNA (e.g., BRACONNOT (1831), QUEVENNE (1838) and SCHLOSSBERGER (1844)) but did not continue to analyze their isolates further.³⁴ At MIESCHER'S request PICCARD analyzed DNA and published in 1874 that the nuclein contains guanine.³⁵ In 1878, KOSSEL continued the work of MIESCHER and determined three more nucleobases, namely adenine^{36,37}, thymine³⁸ and cytosine^{39,40}. He also identified a carbohydrate as part of the DNA. This was possible in part because ALTMANN improved the nuclein extraction protocol and was the first to isolate nuclein in its acidic form, hence the name *nucleic acid*.⁴¹ Then in 1901, ASCOLI published the elucidation of the last nucleobase, Uracil (Figure 2).⁴² Eight years later, another component of DNA, D-Ribose, was published by LEVEN.⁴³ He also proposed the tetranucleotide theory as structure of the DNA, stating that all nucleotides occur in equimolar ratios in a string, with periodic repetitions.^{44–46} Even though LEVEN was one of the first to propose correctly that DNA is a biopolymer, his tetranucleotide postulation was unfortunate. His great reputation led to the false assumption that the question of DNA composition had been settled. Researchers turned away from DNA research, thus prolonging the quest of identifying the true nature of DNA.

Regrettably, information theory was introduced by SHANNON 40 years later, which could have pointed out the flaw with one repeating tetranucleotide.⁴⁷ The core concept, in a nutshell, is that entropy greater than zero is required to encode information.⁴⁸ Identical repeats of a tetranucleotide would have meant that the entropy is zero and no information could be encoded in this hypothetical DNA string.

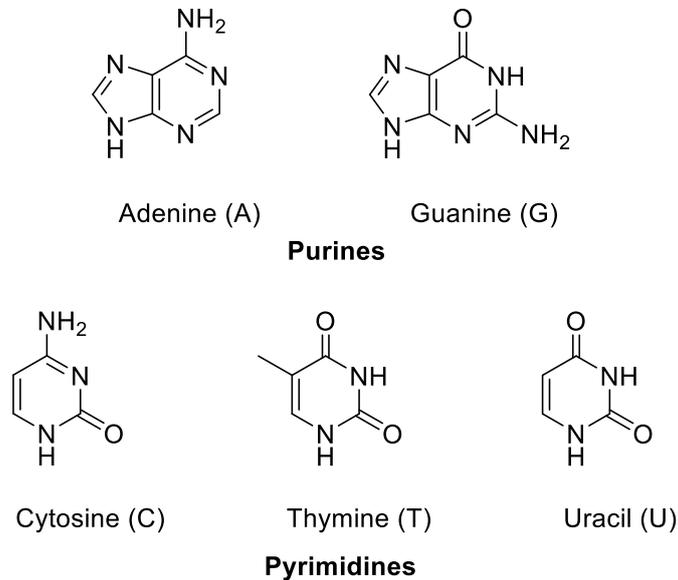


Figure 2: The five nucleobases which are part of DNA (A,T,G,C) and RNA (U instead T)

It took several decades for the tetranucleotide theory to be refuted by work of CHARGAFF and colleagues, who were able to show that the nucleobases always occur in a certain ratio to each other, but not in an equimolar ratio.⁴⁹ This evidence together with additional information from the X-Ray structure of DNA (Photo 51) made by Raymond Gosling, a PhD student of Rosalind Franklin^{50,51}, prompted WATSON and CRICK to propose the correct model of the structure of DNA in their 1953 paper.⁵² This model establishes that the DNA backbone consists of a deoxyribose and a phosphate backbone, and the nucleobases pair in a defined manner (**Figure 3**) resulting in a double helix structure.⁵² The final unanswered question was how the nucleobase sequence encodes the cAAs sequence. In 1961, Crick and colleagues found that the cAA information is encoded by a base pair triplet (codon).⁵³ In the same year, NIRENBERG and MATTHAEI decoded the first codon (UUU for phenylalanine) in their poly-U experiments with cell-free *in vitro* systems of *Escherichia coli* (*E. coli*).⁵⁴ By early 1966, every cAA was assigned at least one codon and by the end of 1966 the entire genetic code was elucidated.^{55,56} **Figure 3** illustrates the exact relationship between DNA, RNA and polypeptide synthesis. The entire genetic code is shown in **Figure 4** and **5**.

consequence of this clustering is that even if there is a mutation at the first and last position of the codon, the encoded cAA is the same or has similar physico-chemical properties and/or side chain volumes. This minimizes the probability of deleterious folding events and therefore loss of activity.

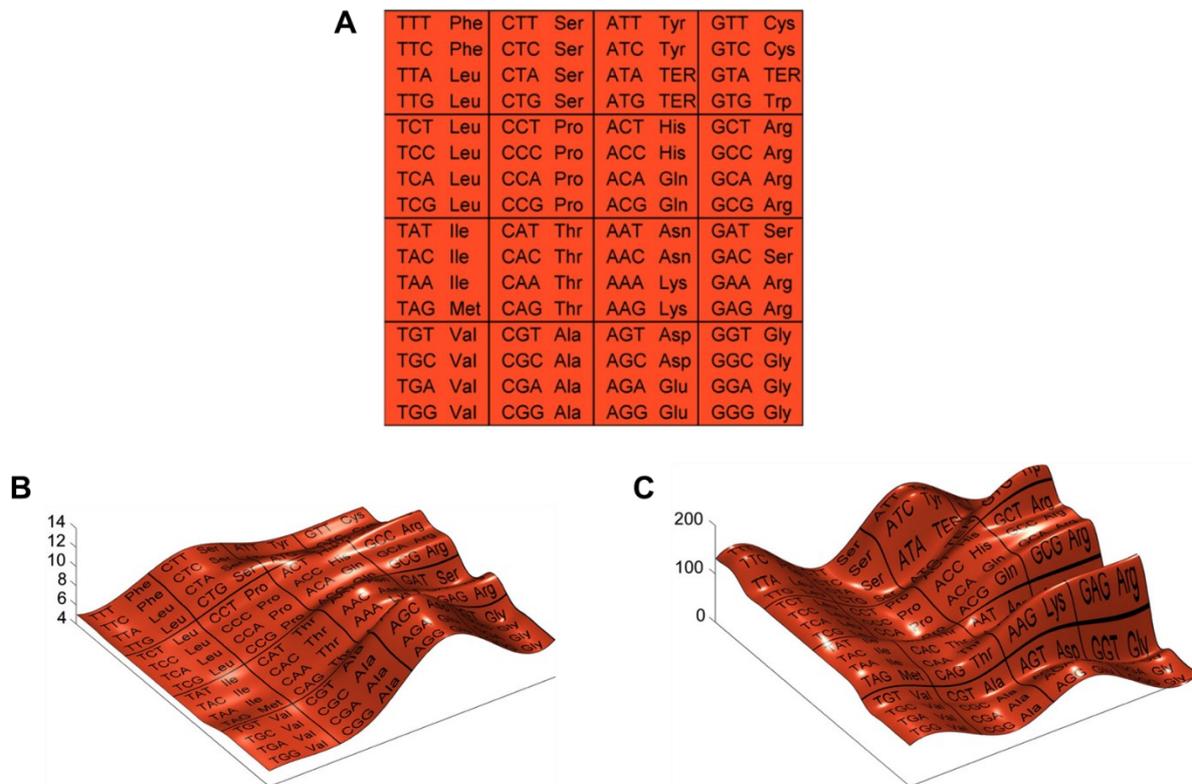


Figure 5: Genetic code table with topographic mapping. **A)** The 64 base-triplet codons are listed with the amino acids or stop signal (TER) they encode. **B)** The polarity of the amino-acids and **C)** the side-chain volume plotted as the height of a topographic map on the code table (values are the amino-acid polarity and molecular volume).⁷¹

1.2.2. Excursus: Origin of Life and Evolution of the Genetic Code: A Note on Complexity, Predictability and Deducibility

Chapter 1.2.1 discussed the structure and function of the genetic code. The question that remains to be answered is why the code has its current structure and whether information can be derived from it. To fully understand the structure of the code, one must look at the origin and evolution of life, which are intertwined with those of the genetic code.

A reasonable assumption for the origin of life is that it arose in a “primordial soup” (liquid phase)^{72,73}, possibly in hydrothermal vents^{74,75}, in which chemical reaction networks formed far from equilibrium states.^{76,77} These networks came into existence, either driven through the constant dissipation of energy absorbed from a strong external source,^{78,79} or spontaneously, provided that a sufficient number of molecules were present.^{80–82} It is also noteworthy that the catalytic efficiency need not be high for a network to emerge.⁸⁰ Such self-organization was the necessary starting point for the emergence of life. These so-called collective autocatalytic sets^{77,83} then evolved from modest to highly complex systems with ever increasing complexity,

until at some point they reached the stage of a cellular replicator (progenote).^{84–87} Clearly, these circumstances indicate that this process was the dynamic evolution of a complex system. Such systems exhibit certain characteristics that normally make them difficult to study. These properties are listed in **Figure 6**:

- i. Complex systems are composed of many elements.
- ii. These elements interact with each other through one or more interaction types. Interactions are often specific between elements.
- iii. Interactions are not static but change over time.
- iv. Elements are characterized by states. State changes can be deterministic or stochastic. They can be the result of an endogenous dynamics or of external driving
- v. Complex systems are characterized by the fact that states and interactions are often not independent but evolve together by mutually influencing each other; states and interactions co-evolve.
- vi. The dynamics of co-evolving multilayer networks is usually highly non-linear.
- vii. Complex systems are context-dependent (boundary conditions).
- viii. Complex systems are algorithmic. Their algorithmic nature is a direct consequence of the discrete interactions between interaction networks and states.
- ix. Complex systems are path-dependent and consequently often non-ergodic.
- x. Complex systems often have memory.

Figure 6: Properties of complex systems.⁸⁸ See below for the non-formal definition of point nine.¹

In order to extract information of a particular system with the help of statistical methods, point 2-10 must not apply. An example where statistical inference is routinely used is statistical mechanics, where a large number of non-interacting particles are analyzed. This is not applicable if the examined system consists of many particles and the interactions are of multiple types, non-linear, and/or vary over time.⁸⁹ Since this is true for all biological systems, Stephen J. Gould put these points together in a very convincing analogy:

“If the “tape of life”— the long evolutionary trajectory that has led to present life on earth—were rewound and played again, the outcomes would be very different”⁹⁰

The aforementioned properties of complex systems make it practically impossible to predict the exact behavior of such a system, although some of them could be predicted theoretically,

¹ In probability theory, an ergodic system is a system that exhibits the same behavior averaged over time as averaged over the space of all the system's states in its phase space.³⁷⁴ Or to put it another way, an ergodic system is a system that visits all its possible states over a reasonable time period.^{99,375} The final state of an ergodic system after a certain time is thus nearly independent of its initial state. This is not the case for a non-ergodic system. This property is also referred to as path dependency or history of the system.¹⁰⁰

if the detailed dynamic interactions of a system and the initial and boundary conditions are known precisely.⁹¹

Therefore, is it impossible to obtain any information about the origin and structure of the genetic code, given non-ergodicity and the lack of known initial and boundary conditions? No, there is information which can be deduced by studying genetic codes! But any knowledge gained must be viewed in the light of these processes. The properties of complex systems limit the possibilities for answering the question of how the genetic code and life itself originated. Because of computational irreducibility², it is very unlikely that there will ever be a true³ theory which can explain how the genetic code and life evolved to their current states.^{92,93} Nevertheless, for some complex systems there are methods available which can reveal certain aspects. One method is to describe the system in a more coarse-grained manner. This trades off the detailed description of the system for more general insights.⁹⁴ Intuitively, this can be conceived as HEISENBERG's uncertainty principle for the description of complex systems.

In that respect, CRICK's first attempt to explain the structure of the genetic code with the "frozen accident theory" can be viewed as a coarse-grained approach.⁹⁵ Astonishingly, he was able to include many features of complex systems even before the theoretical foundations were well established. Unfortunately, many arguments against his theory revolve around the fact that it is not precise enough and therefore applicable to a wide set of possibilities. This is precisely the case when some properties of a complex system are captured with a coarse-grained model at the expense of details. For example, a common misinterpretation of CRICK's theory is that all codon assignments were chance-based, although he said that only some of them were random, ignoring the detailed further explanation in his paper.^{96-98,4} This part of his argument just acknowledges the path dependency (non-ergodicity) of the evolution of the genetic code without explaining how it came to a particular starting point.

Non-ergodicity is a central feature of complex systems^{88,99} and of the evolution of the genetic code. It is difficult to imagine how an ergodic evolution of the genetic code could have been realized, since visiting all possible states would have meant that there would have been a point in time at which all possible genetic codes would have been present for a selection system to act upon. This would certainly have been impossible to realize considering the astronomical number of possible genetic codes. Moreover, it is not clear what the selection criteria would

² testing such a theory would most probably entail the simulation from prebiotic conditions to the current time point with every possible conditions, which is impossible due to computational constraints.

³ true in the Popperian sense of empirical falsification

⁴ These references are publications where CRICK's 1968 paper is misinterpreted in the introduction, even though they often recognize the insights of CRICK's analyses in the end.

have been at this stage of evolution (before cellular replicators existed). Darwinian selection applies only to the living biological world, other levels of complexity can have diverging selection methods.¹⁰⁰ For complex biological systems, these constraints are not merely physical but rather contain functionalities and therefore do not strictly adhere to physical laws.¹⁰¹ All of the above raised points should be considered when discussing the structure and evolution of genetic code.

1.2.3. The Protein Biosynthesis

Genes are the essential part of the genetic information. They code for the workhorses and the central part of the cell, proteins. Their distribution, function, and interplay within a cell constitute life. Protein biosynthesis starting from DNA in a cell is shown schematically in **Figure 1** and will be briefly explained here (**Figure 7**). For the translation of mRNA into a correct amino acid sequence, an adapter molecule is necessary, the transfer RNA (tRNA). One of the most important features of the tRNA is the anticodon (for more details see **chapter 1.2.3.1**).

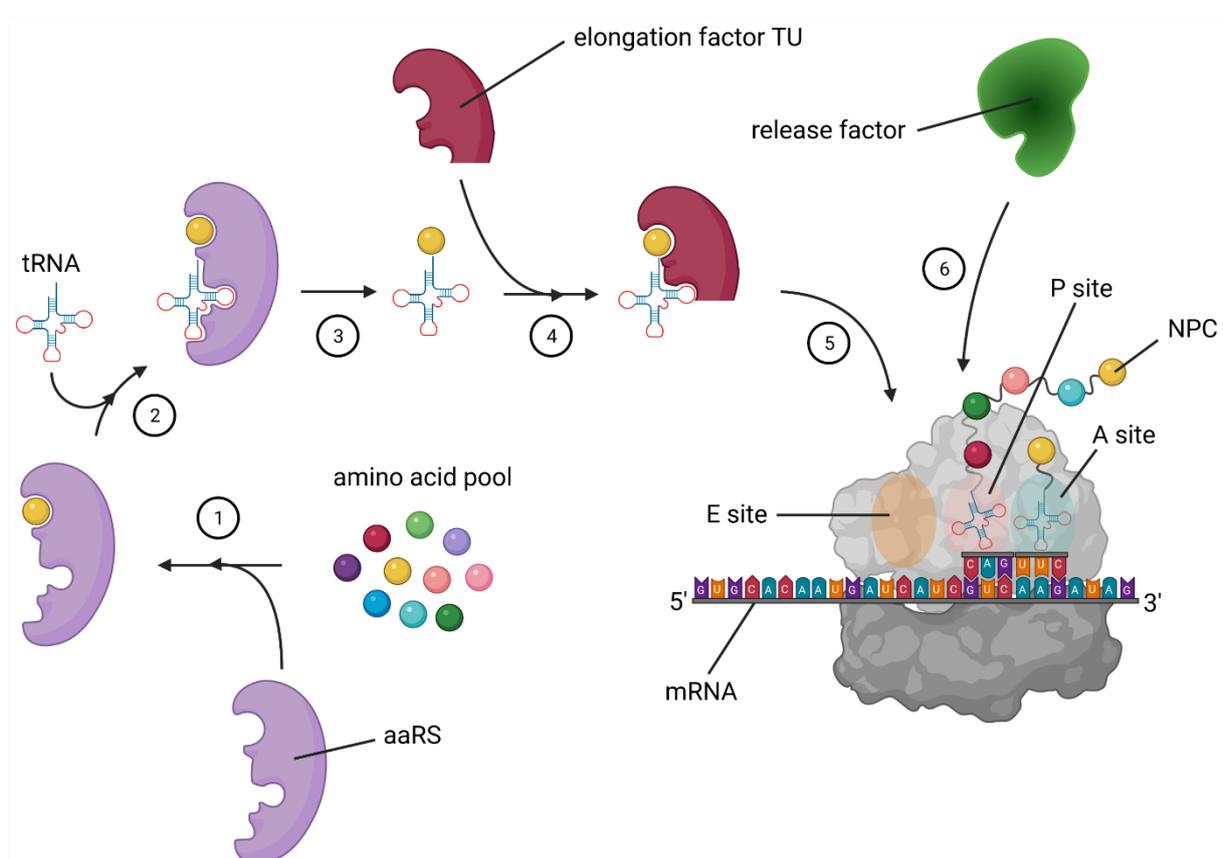


Figure 7: Schematic representation of the protein biosynthesis machinery: 1) Activation of the aaRS with cAA from the amino acid pool. 2) Binding of tRNA to aaRS ("loading" of the tRNA with cAA). 3) Aminoacylated tRNA (AA-tRNA) leaves the aaRS. 4) Binding of the elongation factor TU (EF-TU). 5) Binding to the aminoacyl site of the ribosome (A site). There, the anticodon of the tRNA forms hydrogen bridge bonds to the corresponding mRNA codon. Then at the peptidyl site (P site), the tRNA transfers the nascent polypeptide chain (NPC) to the AA-tRNA to elongate the NPC by the additional cAA. After transfer, the tRNA at the P site moves to the exit site (E site) and leaves the ribosome. 6) Translation ends when a stop-codon arrives at the A site. At that point, the release factor (RF) binds to the ribosome and initiates the release of the polypeptide chain.

With the codon-anticodon recognition, the tRNA assigns a specific AA to each codon in a mRNA sequence (**Figure 7**).¹⁰² For protein translation to occur with high fidelity, the tRNA must be loaded with the correct AA. This function is executed by aminoacyl tRNA synthetases (aaRSs). After the loaded tRNA is released from the aaRS, the charged tRNA binds to the elongation factor TU (EF-TU) which guides it to the ribosome. At the same time, the ribosome binds to the mRNA and the codon-anticodon recognition mechanism can take place. The ribosome consists of three subunits which is composed of over 50 proteins and ribosomal RNA (rRNA). One of the greatest achievements of biochemistry was the function elucidation of this complex. It was shown that the catalytic function of this complex is carried out by rRNA, demonstrating that the ribosome is in fact a ribozyme, an rRNA-based catalyst.

1.2.3.1. The Transfer RNA

The tRNA molecules serve as adapters to translate a codon into a given AA specified in an mRNA sequence. All known tRNAs consist of 73 to 93 ribonucleotides.¹⁰² They are composed of several nonstandard chemically altered nucleobases, with methylation as the simplest modification. These modifications modulate recognition interactions with aaRSs and the ribosome.¹⁰³ In canonical two-dimensional representation, tRNAs resemble a cloverleaf (**Figure 8A**). Approximately half of the nucleotides form base pairs with their corresponding double helical structures, leading to the formation of L-shapes in their three-dimensional representations (**Figure 8B**).¹⁰²

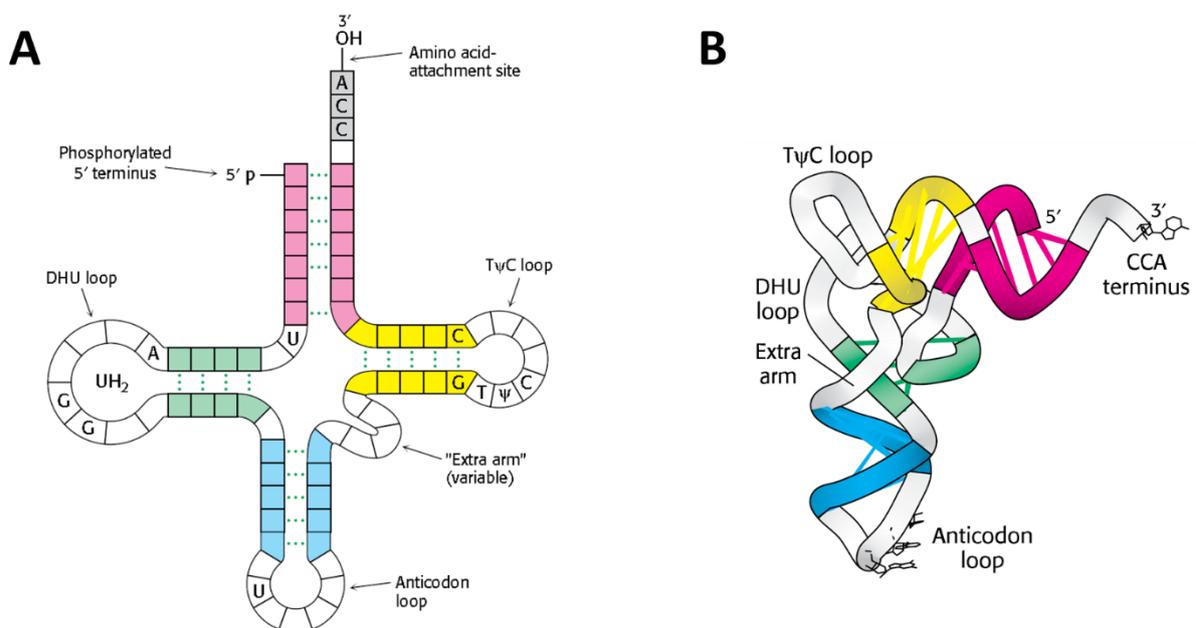


Figure 8: **A**) Two-dimensional depiction of a generic tRNA in the typical cloverleaf structure. The tRNA generally contains four characteristic loops. The dihydrouracil loop (DHU loop), the anticodon loop, the extra arm (variable loop, which contains a variable number of bases) and the TψC loop (containing ribothymidine (T), pseudouridine (ψ) and cytosine, hence the name). **B**) Three-dimensional structure of tRNA. The four double-stranded regions of tRNA stack to form an L-shaped structure.¹⁰⁴

The four helical segments form an apparently continuous double helix. Depending on the species, 30 to 40 different tRNAs have been identified.¹⁰² Since there are 61 codons, a tRNA must be able to decode more than one codon.¹⁰⁵ This exact mechanism was postulated by CRICK as the “Wobble Hypothesis”¹⁰⁶ which states, that only the first two bases form a strict Watson-Crick base pairing. Such wobble decoding is often enabled by posttranscriptional modifications of the first anticodon base.¹⁰⁷ Since these modifications vary from species to species, there are also differences in the abundance of codons, aaRS, and tRNAs in different organisms.¹⁰⁸ The tRNAs that can decode codons without strict Watson-Crick pairing are called isoacceptor tRNAs.¹⁰⁹ Interestingly, many species possess more than one copy of these tRNAs in their genome. For example, strain *E. coli* K-12 has 85 to 86 copies in its genome.^{110,111} TULLER and colleagues have shown that the tRNA copy number directly correlates with the intracellular concentration.¹¹² Many years earlier, DONG and colleagues had already found that tRNA and codon abundance correlate within an organism.¹¹³ It was also found that proteins highly overexpressed within an organism are coded for by codons which have the highest abundance in the respective species.¹¹⁴

The molecular recognition mechanism between the aaRSs and the tRNAs occurs in different regions called identity elements.¹¹⁵ These elements are shown in **Figure 9** where it is clearly shown that the anticodon and the acceptor stem are the most frequently used regions for recognition.

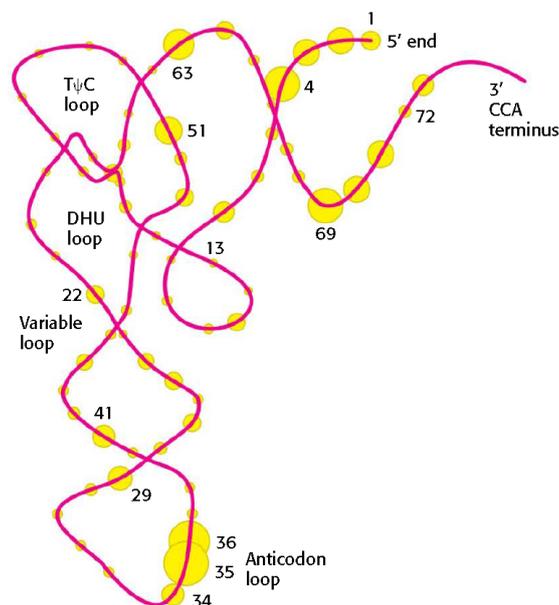


Figure 9: Schematic representation of a tRNA with its most characteristic recognition sites: The circles represent nucleotides, and the size of the circles is proportional to the frequency with which they are used as recognition sites by aminoacyl-tRNA synthetases. The numbers indicate the positions of the nucleotides in the base sequence, starting with the 5' end of the tRNA molecule.¹⁰⁴

1.2.3.2. Aminoacyl tRNA Synthetases

For the protein biosynthesis to occur the tRNA must be charged with an AA (**Figure 7**, step 1-3). This charging process describes the ester bond formation of an AA with its cognate tRNA. The crucial part of this reaction cascade is the recognition of AAs and their activation via formation of aminoacyl-adenylate (mixed anhydride) using ATP (**Figure 10**, step 1). Then, the nucleophilic AA transfer to the tRNA, forming an ester bond between AA and tRNA (aminoacyl-tRNA). These reactions catalyzed by aaRSs are shown in **Figure 10**.

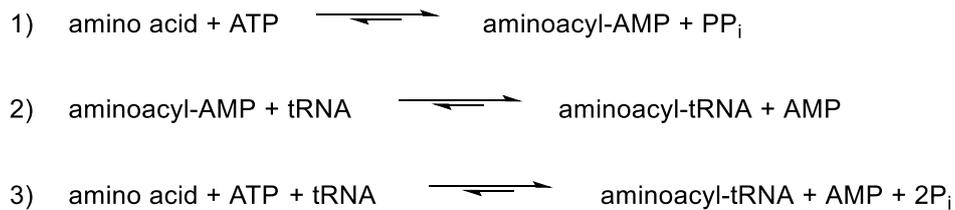


Figure 10: Reaction of tRNA aminoacylation. This reaction cascade is catalyzed by an aaRS. Driving force of the reaction is the pyrophosphate (PPi) hydrolysis. Hydrolysis of PPi is not depicted. 1) Activation of the amino acid. 2) Transfer of the amino acid to the tRNA. 3) Overall chemical reaction. Nucleophilic attack at the α position of ATP displaces pyrophosphate (PPi) and transfers adenylate (5'-AMP) to AA forming aminoacyl-AMP (the reaction is called adenylation in non-ribosomal peptide synthesis). The PPi is further hydrolyzed by inorganic pyrophosphatase to two Pi (-20 kJ/mol). This is the driving force of the reaction.

The fidelity of aaRS in AA recognition and tRNA loading is of paramount importance for the correct translation of mRNA. A robust decoding mechanism is only possible when this process takes place with a very low error rate. The error rate is estimated to be $1 - 6 \cdot 10^{-4}$ per codon translated.¹¹⁶ Most species possess at least 20 aaRS to load all tRNA isoacceptors.¹⁸ There are some known cases in which an organism possesses fewer than 20 (for which there are various compensatory strategies).¹¹⁷ Although the catalyzed reaction is always the same, aaRSs differ in size, structure and subunit composition.¹⁰² They are clustered into two classes based on their structure, function and evolutionary dependence (**Table 1**).¹¹⁸

Table 1: Classification of aaRSs.¹¹⁹

	Class I	Class II	Characteristic Feature	Class I	Class II
A	MetRS* LeuRS* IleRS* ValRS*	SerRS* ProRS* ThrRS* GlyRS HisRS	catalytic domain	Rossmann fold	antiparallel beta sheet
B	CysRS GlnRS GluRS	AspRS AsnRS LysRS-II*	aminoacylation site	2'-OH	3'-OH
C	TyrRS TrpRS	PheRS* GlyRS AlaRS* SepRS PylRS	tRNA-binding	minor groove	major groove
D	ArgRS		anticodon recognition	the majority	the minority
E	LysRS-I		quaternary structure	monomers	di- or oligomers

(*) possess editing function

1.3. Synthetic Biology

There is no clear definition of synthetic biology. It is an interdisciplinary field which overlaps with chemistry, biology, biotechnology and engineering (e.g., protein engineering). Synthetic biology differs from other scientific disciplines in that it applies specific knowledge to manipulate systems (in this case biological) in a desired way. The ‘forward engineering’ approach is predominant to create new molecules or functions within biological systems in contrast to the ‘reverse engineering’ approach, which is used to elucidate certain aspects of biological phenomena in molecular biology or biophysics/chemistry.¹²⁰ The synthetic biology approach can be used at any hierarchical level of living systems. It can be used to create DNA information storage devices¹²¹, biological computers¹²², and entire organisms that can degrade toxic chemicals¹²³ to just name a few examples.

1.3.1. Genetic Code Expansion

Expanding the genetic code has proven to be an important tool for adding new chemistries to the biological world and expand the chemical space of proteins beyond the standard 20 amino acids.^{124–126} In this regard, amino acids with non-proteinogenic functional groups can be used to manipulate, design, and elucidate protein structure, dynamics, function, allostereism, interactions, catalysis, folding, synthesis, trafficking, degradation, and aggregation^{127–133}. In nature, there are also methods to expand the chemical space beyond the standard 20 amino acids. This is done by enzymatic and non-enzymatic post-translational modifications (PTM) of protein side chains and backbones of which over 200 are known.¹³⁴ The most frequent posttranslational modifications are acylation, alkylation, glycosylation and oxidation.¹³⁵ To incorporate ncAAs ribosomally at defined positions specified by an mRNA, the existing code or its interpretation has to be manipulated. Today, there are four common site-specific approaches which are used *in vitro* (i and ii) or *in vivo* (iii and iv) methods. (i) The use of chemically acylated tRNAs with ncAAs in a cell free system. HECHT and colleagues developed the general method of chemical tRNA acylation^{136,137} and Schultz and colleagues used this strategy to first suppress an amber stop codon in yeast.¹³⁸ (ii) Solid phase peptide synthesis.¹³⁹ (iii) Modification of the translational machinery to recognize quadruplet codons.¹⁴⁰ (iv) The reassignment of a nonsense codon (stop codon suppression, SCS).^{124–126}

1.3.1.1. Stop Codon Suppression

The first SCS system for *in vivo* use was based on an orthogonal phenylalanine aminoacyl tRNA synthetase (PheRS) isolated from yeast which was developed from FURTER in 1998.¹⁴¹ This system had low incorporation efficiency and therefore delivered heterogeneous target protein mixtures. In 2001, SCHULTZ and colleagues developed an improved system derived from an archaeal organism.¹⁴² Coincidentally, with the development of FURTER’s and

SCHULTZ's system, a natural SCS system was serendipitously discovered, the pyrrolysyl-tRNA synthetase (PylRS)(see **chapter 1.3.1.2**).

These SCS systems are a particular straightforward strategy to incorporate ncAAs into proteins. They consist of an aaRS/tRNA pair that enables the ribosomal incorporation of ncAAs in response to a reprogrammed codon. The aaRS/tRNA pairs form the essential parts of an orthogonal translation system (OTS). The most commonly used approach for this site-specific peptide and protein modification is stop codon suppression which targets the amber stop codon. The aaRS/tRNA pairs form the essential parts of an orthogonal translation system (OTS). The most commonly used approach for this site-specific peptide and protein modification is stop codon suppression which targets the amber stop codon. Here, the ncAA is incorporated in response to an in-frame stop codon placed at a predefined position in the protein coding sequence which is ribosomally expressed either *in vivo* or *in vitro*.^{124–126} The majority of all engineered aaRS variants to date are derived from *Methanosarcina mazei/barkeri* PylRS (*MmPylRS*/*MbPylRS*) or *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase (*MjTyrRS*).^{124–126} The archaeal origin and therefore distant phylogeny are the reason for their orthogonality in bacterial and, for the former, even in eukaryotic cells.^{124–126} It is this dual orthogonality that allows easy engineering, screening and selection in bacterial cells, e.g. *E. coli*, and subsequent transfer of the machinery to more complex eukaryotic hosts (for more information see **chapter 1.3.1.2.1**).

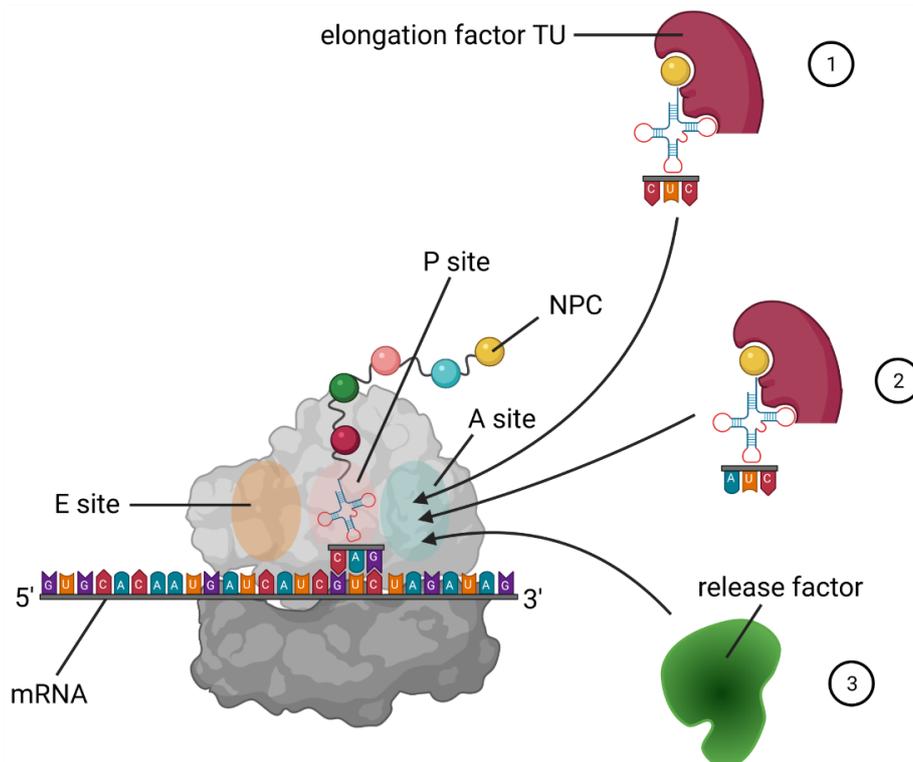


Figure 11: SCS mechanism for an in-frame amber stop-codon readthrough and competing reactions. 1) Near cognate suppression with Gln-tRNA (anticodon:5'-CUC-3'). 2) Suppression of amber stop codon with suppressor tRNA. 3) Competing process to an in-frame stop-codon readthrough: termination of protein translation triggered by RF.

Even though there is a broad spectrum of applications, a common obstacle remains the low efficiency of ribosomal ncAA incorporation resulting in low yields of target protein.^{143–145} There

are two competitive mechanisms that reduce these yields. The premature termination triggered by the release factor and near cognate suppression in which Gln is incorporated at the amber position due to wobble decoding (**Figure 11**).

Both effects are more pronounced when the orthogonal aaRS is inefficient. Low efficiency leads to low levels of ncAA loaded tRNAs and therefore low decoding efficiency. While some of the *Mj*Tyr systems provide target protein yields at wild-type levels, the PylRS system never achieved similar results. To some extent, the low efficiency could be compensated by supplying higher concentrations of ncAA to the growth media. The standard is 1 mM, but higher concentrations of up to 10 mM ncAA are often used, which dramatically increases the cost of unnatural protein production.^{145–147} The overall efficiency of production generally depends on a variety of factors but generally decreases as the number of in-frame stop codons in the target gene sequence increases.¹⁴⁸ Many efforts have already been made to incrementally increase protein production. These include optimization of OTS plasmid copy numbers and promoter strength of aaRS and/or tRNA genes¹⁴⁹, engineering of tRNA¹⁵⁰, directed evolution of aaRS¹⁴³, optimization of sequence context surrounding the target codon¹⁵¹, host cell engineering¹⁵² and engineering of parts of the translational machinery (e.g., elongation factor TU or RF1).^{153,154}

1.3.1.2. *Pyrrolysyl tRNA Synthetase*

The PylRS was discovered in *Methanosarcina* species when the function of enzymes important for methane metabolism was elucidated. But recent discoveries of additional archaeal organisms show that the PylRS system is not limited to the order of *Methanosarcinales* which was thought for a long time.¹⁵⁵ It was initially discovered that *M. barkeri* uses an amber stop codon in monomethylamine methyltransferase (MtmB) to encode for Pyl (**1a**, **Figure 12**) which was later designated the 22nd amino acid.^{156–158} Further investigations of the PylRS system revealed that there are only two essential components needed to encode Pyl (**1a**) in response to the amber codon: The PylRS and the tRNA^{Pyl}.^{159,160} Since its discovery, the PylRS system has been engineered to incorporate over 100 ncAAs via SCS.^{161,162} It should be noted, that the majority of ncAAs incorporated using the PylRS system are flexible, long-chained and bulky Pyl (**1a**) analogs^{163,164} or shorter but still bulky aromatic substrates, especially phenylalanine^{165–169}, tryptophan^{170,171} and histidine¹⁷² analogs (**Figure 12**).¹⁶¹ Additionally, long and bulky azobenzene-alanine-derivates have been incorporated.^{173,174} The third wave (red) was established in this work, see **chapter 2.3**. The PylRS from all known *Methanosarcina* consists of two domains, the N-terminal domain important for *in vivo* tRNA^{Pyl} recognition, the C-terminal domain for catalytic aminoacylation and a linker between these two domains that varies in size depending on the origin.¹⁷⁵ Structural phylogeny analysis assigned PylRS to class II aaRS.¹⁷⁶ From a protein engineering perspective, an interesting structural feature of these class II aaRSs is that new AA recognition has historically evolved by changing the AA side chains

within the binding pocket rather than by changes that might affect the position of the protein backbone or secondary structure. Thus, novel amino acids should be easily encoded, either evolutionary or rationally. From an evolutionary perspective, it seems plausible that the PylRS evolved from an ancient PheRS before the last common universal ancestor (LUCA).¹⁷⁶ To support this hypothesis, a comparison with a PheRS from *Thermus thermophilus* displays that the two structures are highly superimposable and have similarly organized hydrophobic pockets for amino acid binding (**Figure 13**).¹⁶¹ This is remarkable considering that the sequence similarity is less than 30%.

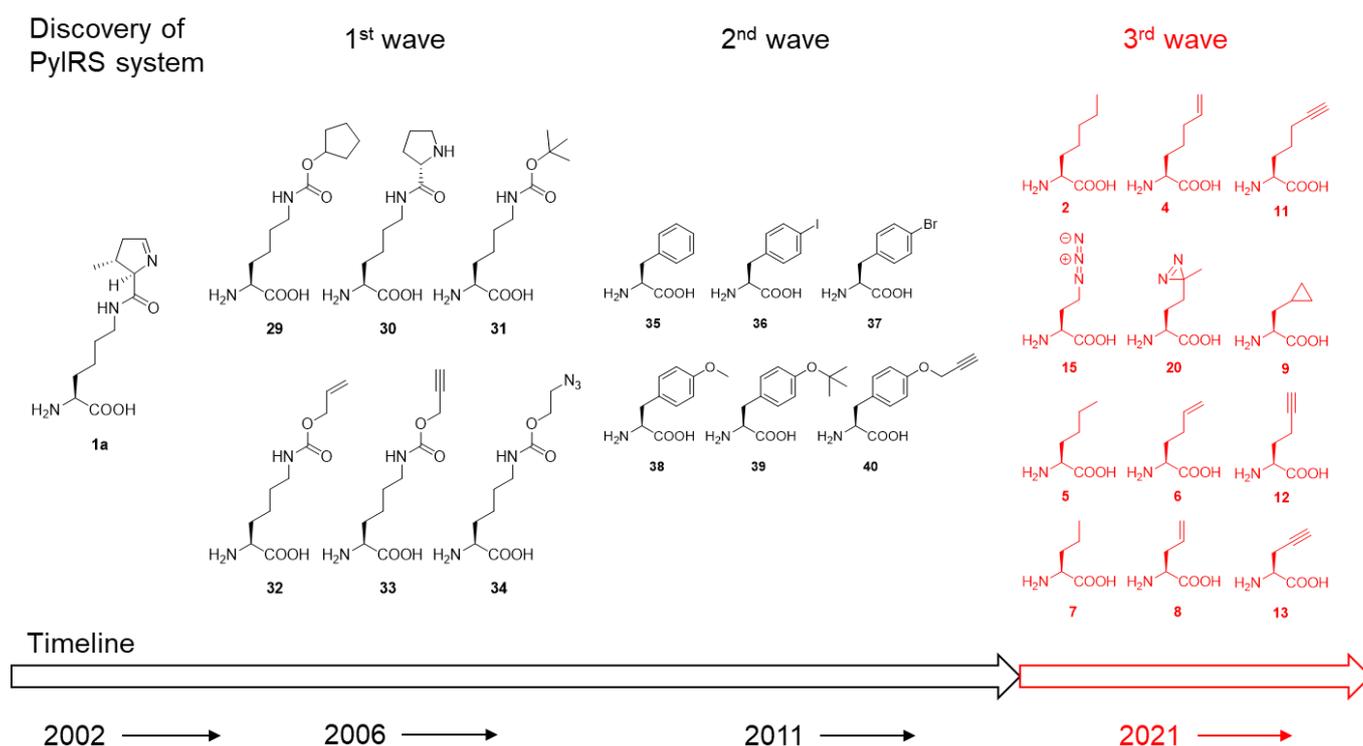


Figure 12: Flowchart showing the timetable of characteristic ncAAs that can be incorporated with the PylRS OTS. The third wave (red) was achieved in this work.¹⁷⁷

Methane-producing microorganisms are thought to be among the earliest cellular life forms that colonized various anaerobic habitats on our planet.¹⁷⁸ The discovery of a PylRS system that is distinct to the ones found in bacteria and archaea suggest that the most likely origin is in pre-LUCA bacterial or archaeal organism and later spread by horizontal gene transfer (HGT).¹⁵⁵ This scenario implies, that also the methanogenic genus *Methanosarcinae* acquired the Pyl-coding genes by HGT. The metabolic needs of methanogenesis were most likely the driving force for the natural expansion of the genetic code (i.e., reassignment of the Amber codon to Pyl) in methanogenic archaea living in extreme habitats.¹⁷⁹ The addition of Pyl could be a recent evolutionary event or a "fossil" derived from a pre-LUCA lineage representing a hypothetical extinct fourth domain of life, as speculated by Fournier and associates.^{179,180} In addition to an interesting evolutionary history, PylRS enzymes are remarkably tolerant of activating substrates with various amino acid side chains and even alpha-hydroxy acids.

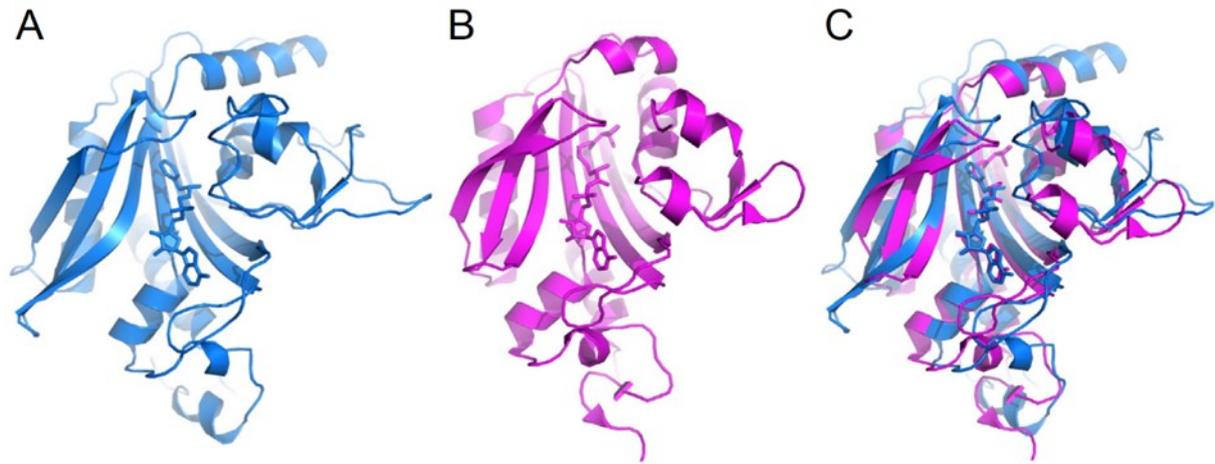


Figure 13: Structural comparison between bacterial PheRS and archaeal PylRS. A) Structure of *Thermus thermophilus* PheRS with bound phenyl AMP. (B) Structure of *M. mazei* PylRS with bound pyrrolysyl AMP. (C) Superimposed structures of A and B.¹⁶¹

Together with the lack of an anticodon recognition domain in PylRS, the Pyl incorporation machinery is probably the best accessible tool for experimental expansion of the genetic code to date. However, the poor solubility of the protein is a disadvantage for both, robust orthogonal translation and is the reason why only recently it has been possible to obtain a model of the full-length enzyme crystal structure (without the linker). The high-resolution 3D structure enabled to elucidate the PylRS/tRNA^{Pyl} binding recognition mechanism (**Figure 14**).¹⁸¹

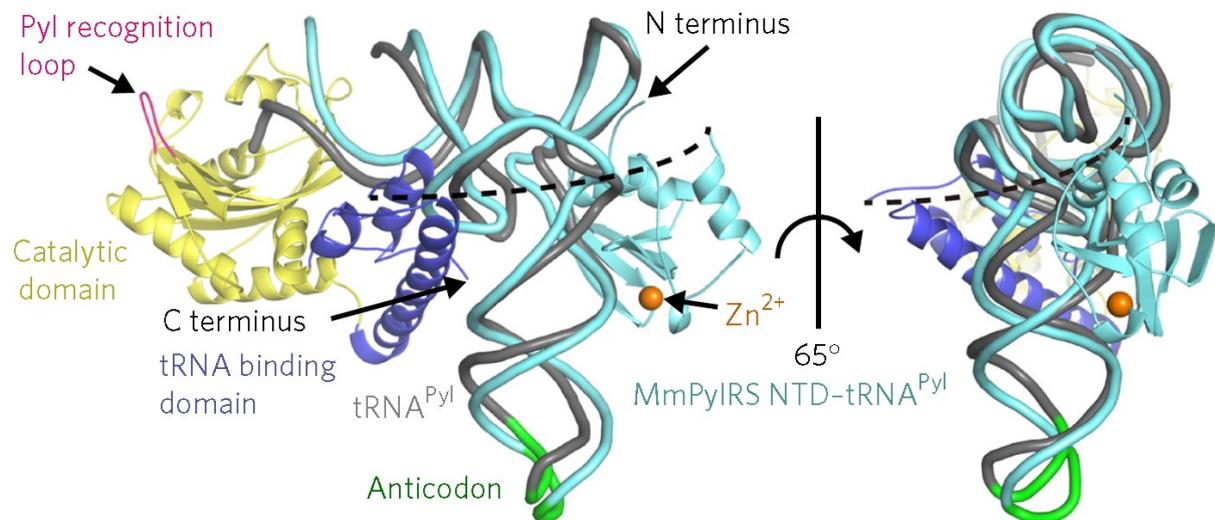


Figure 14: Superposition of the *MmPylRS* NTD-tRNA^{Pyl} complex (cyan) and the *Desulfitobacterium hafniense* PylRS CTD-tRNA^{Pyl} complex (gray; PDB ID 2ZNI). The cation Zn²⁺ is shown as an orange sphere. The PylRS CTD consists of catalytic (yellow) and tRNA binding (blue) domains. The Pyl recognition loop is indicated in pink. The possible path of the linker connecting NTD to CTD is shown by a black dashed line; the linker varies in AA-length depending on the source of PylRS (see **chapter 2.4** for more discussion about the linker).¹⁸¹

The reason for the orthogonality of the PylRS system is the unique binding mode of the PylRS to the tRNA^{Pyl}. As discussed in **chapter 1.2.3.1**, most aaRSs recognize their cognate tRNAs at the anticodon or acceptor stem. This is different for the tRNA^{Pyl} as it has almost no variable loop, the anticodon arm and T-loop possess tight interactions and act as recognition

elements.¹⁸¹ Moreover, this also explains the anticodon "blindness" of tRNA^{Pyl}¹⁸², which has also been exploited for reassignment of sense codons and assignments of newly created codons in a semisynthetic organism.^{183,184} Because the PylRS OTS can be used to reassign sense and nonsense codons and is orthogonal in bacterial and mammalian cells, its superiority over all other OTS is undeniable. Unfortunately, its low efficiency and limited substrate scope are still impeding its usefulness.¹⁸⁵

1.3.2. Encoding New Enzyme Functions: Perspectives on PylRS vs. *Mj*TyrRS Engineering

As mentioned in **chapter 1.3.1.1**, the PylRS- and *Mj*TyrRS-based OTS are the most commonly used systems for site-specific incorporation of ncAAs into a protein sequence. Their pros and cons will be discussed here in the context of implementing a new function of an aaRS and the likelihood of successfully producing such an enzyme.

In general, there are two key features that an enzyme should possess to ensure successful recognition of new substrates. First, in the best case, the target enzyme should have low levels of the desired or similar new activity. For aaRSs, enzymatic activity for ncAAs which strongly deviate from their native substrate would fulfill this prerequisite.^{186,187} Even if this is not the case, the PylRS has great promiscuity towards ncAA analogs based on its unique Pyl (**1a**) recognition mode, which relies on relatively nonspecific hydrophobic interactions in the large binding pocket. This makes it a superior candidate of OTS engineering than the *Mj*TyrRS which has very specific substrate interactions. As mentioned in **chapter 1.3.1.2**, the PylRS is a class II aaRS with structural features additionally increasing the probability to encode new substrate recognition.¹⁷⁶ In addition, there is empirical evidence that these PylRS-based OTSs are amenable to encode a variety of substrates.¹⁶¹ Compared to *Mj*TyrRS, a class I aaRS, this is a major advantage. Class I aaRSs require many more mutations to encode new substrates (up to 10 mutations^{128,188}) compared with class II aaRSs, making the engineering of these enzymes overly complicated (often requiring computational models) and reducing the likelihood of finding a desired enzyme. Since the incorporation efficiency of the *Mj*TyrRS system was always higher than that of the PylRS-based OTS, these drawbacks were tolerated.

Second, sufficient stability is required to buffer destabilizing mutations necessary for active site remodeling^{189–191}. Lower stability may result in a lower abundance of correctly folded and active enzymes.^{191,192} Unfortunately, PylRS is marginally stable under standard cultivation conditions in *E. coli*¹⁹³ which is also reflected in the low *in vitro* solubility of the enzyme.^{176,194} This is even more pronounced for the *Mb*PylRS variant than for *Mm*PylRS^{169,181}. Given the above changes affecting enzyme stability, it has been suggested that aaRS scaffolds with high thermostability provide better starting points for enzyme engineering.¹⁹¹ This is true for the *Mj*TyrRS system, which is derived from a hyperthermophilic organism. Since this aaRS

requires numerous mutations to encode for new substrates, the hyperthermophilic origin provides the necessary stability to tolerate these mutations. While most thermophilic enzymes are very inefficient at room temperature (compared to their meso- or psychrophilic counterparts), the *Mj*TyrRS OTS is extremely efficient at standard cultivation conditions (30-37°C) which is extremely unusual for a hyperthermophilic enzyme (it is conceivable that there are not many hyperthermophilic enzymes with that efficiency that could be engineered as OTS).³ On the other hand, the results of the thermophilic aaRS approach for the PylRS OTS are inconclusive. To date, only one study has been conducted, and it has two major shortcomings.¹⁹³ First, the reference system used in this study (*Mb*PylRS) is not the most efficient known to date (which is the *Mm*PylRS) resulting in a lack of information to make a clear conclusion. Second, although a phylogenetic tree is presented in the publication, the high sequence similarity between the *Mt*PylRS used and the most psychrophilic PylRS is not recognized (*Ml*PylRS) (see **chapter 2.4** for details). Finally, following the logic of the phylogenetic tree, they compare a mildly thermophilic PylRS (*Mt*PylRS) with the not-best-performing mesophilic variant (*Mb*PylRS) and a psychrophilic variant (*Ml*PylRS). Moreover, the whole thermophilic PylRS approach seems implausible, as PylRS has a relatively unspecific substrate recognition, which means that fewer mutations (usually 2-4) are required to code for new substrate recognitions. Summing up this discussion, the search for more efficient PylRS OTSs should focus on psychrophilic PylRS variants rather than thermophilic ones (see **chapter 1.4.2**). For more information on extremophilic enzymes used in synthetic biology see **chapter 1.4**

1.3.2.1. Solubility Tags

Soluble expression of enzymes depends, among other factors, on thermodynamic stability, the aggregation propensity of the enzyme, and the folding rate.¹⁹⁵ One problem in protein engineering is that the enzyme's active site must be mutated. Since all mutations are likely to be destabilizing than stabilizing, this destabilization can lead to protein unfolding (and thus inactive enzymes) and concomitant protein aggregation.^{196,197} The solution here may be the classical approach of increasing the production of soluble and active recombinant enzymes by genetically fusing a well-folding, highly soluble protein domain to the N- or C-terminus of the target protein.^{198,199} These tags are thought to prevent the target protein from entering detrimental folding states by providing more time for correct folding, either spontaneously or with the help of chaperones.²⁰⁰ Since these tags are fused to one of the enzyme's ends, the catalytic center of the enzyme is not altered, thereby solving the problem of beneficial solubility mutations near the center.

1.4. Use of Extremophilic Enzymes in Synthetic Biology

Life on Earth can exist in the strangest of spaces, from hyperthermophiles that can survive in environments with temperatures as high as 120 °C to psychrophiles that live in temperatures of -20 °C.^{201–203} Some organisms have developed strategies to adapt parts of the molecular machinery to function under these extreme conditions. In addition to adapting parts of the cell to maintain structural integrity, most of these adaptations involve the enzymes of these organisms. The successful colonization of almost all niches on Earth, from the permanently frozen polar regions or the Siberian permafrost, to the moderately and extremely hot deep-sea hydrothermal vents, is partly due to these special enzyme properties.

1.4.1. Heat-adapted Enzymes

Most thermophilic enzymes exhibit robust activity under harsh industrial production conditions (e.g., exposure to denaturing chemicals, surfactants and oxidants).²⁰⁴ Since they are thermostable, they can be easily purified by heat treatment, as long as they are produced in mesophilic organisms.²⁰⁵ The high working temperature brings other advantages under industrial conditions: a lower risk of contamination, easier mixing due to lower viscosity and higher solubility of the substrates, and a shift in the reaction equilibria in endothermic reactions, leading to higher product yields.²⁰⁴ The robustness also makes them perfect parent enzymes for the engineering of new substrate recognition systems, as discussed in **chapter 1.3.2**.²⁰⁶ From a structural biology perspective, the robustness is due to increased hydrophobic packing density in the inner core region, shortened surface loops, increased number of hydrogen bonds per residue, and increased number of charged residues. All these features lead to the stabilization and rigidification of the folded state.²⁰⁷ A major drawback resulting from the rigidification of the scaffolds is that almost all thermostable enzymes do not exhibit catalytic activity far from their optimum catalytic working conditions e.g., at room temperature.²⁰⁸ This also means that the use of these enzymes in biotechnological applications requires high energy input to achieve optimal working conditions, which is expensive.²⁰⁴

1.4.2. Cold-Adapted Enzymes

Cold-adapted enzymes are very attractive research targets, because their high enzymatic activity at low temperatures, can make them a valuable resource for various biotechnological applications.²⁰⁹ The disadvantages of thermostable enzymes mentioned above are the advantages of the cold-adapted ones and vice-versa. In addition, their higher surface hydrophobicity and hydrophilicity lead to a dense hydration shell even under low-water conditions, favoring catalysis in organic solvents.²¹⁰ Interestingly, psychrophilic enzymes exhibit a high degree of structural and sequence conservation with their meso- and thermophilic homologs.²⁰⁷ In particular, the residues crucial for the catalytic mechanism are always conserved in all homologs.

Even though the mechanisms of cold adaption of enzymes are diverse and often enzyme-specific, some general aspects have been elucidated. Compared to their meso- and thermophilic homologs, psychrophilic enzymes have a higher flexibility which is achieved by a lower hydrophobic packing density in the interior core region, longer surface loops, a lower number of hydrogen bonds per residue, a higher surface hydrophilicity and a decreased number of charged residues and thus fewer ionic interactions.^{207,209} From an energetic point of view this strategy manifests in a reaction enthalpy-entropy tradeoff.^{207,211} Eq. 1 describes the activation energy ΔG^* of a reaction, where ΔH^* is the enthalpic part, ΔS^* the entropic part, and T the temperature.

$$\Delta G^* = \Delta H^* - T\Delta S^* \quad (1)$$

In contrast to uncatalyzed reactions eq. 1 describes not only the pure substrate-product formation, it also includes the energetic values for the substrate and product interactions with the enzyme. In a nutshell, ΔH^* describes the stability of the enzyme which is determined by numbers and strength of stabilizing interactions (rigidity). The ΔS^* value includes the “penalty” an enzyme imposes on the reaction, depending on how coordinated (“ordered” or rigid) the enzyme is and becomes during that reaction. The more flexible the enzyme, the higher the entropic penalty during the reaction, since binding to a substrate increases the “order” (decreases the entropy) of the enzyme. However, this penalty decreases with decreasing temperatures. And because there is a ΔH^* - ΔS^* -tradeoff, it is plausible that psychrophilic enzymes have adapted with this strategy to speed up reactions at low temperatures, see **Table 2** with data for α -amylase homologs as an example. The increased entropy is reflected in increased flexibility of the enzyme. **Figure 15** shows the two main energetic effects of increased flexibility. The increased flexibility can occur at different locations in the enzyme.

Table 2: Activation parameters of psychro-, meso- and thermophilic α -amylases at 10°C²¹²

α -amylase	k_{cat} [s^{-1}]	ΔG^* [kcal mol ⁻¹]	ΔH^* [kcal mol ⁻¹]	$T\Delta S^*$ [kcal mol ⁻¹]
psychrophilic	294	13.8	8.3	-5.5
mesophilic	97	14	11.1	-2.9
thermophilic	14	15	16.8	1.8

First, when the increased flexibility is at or near the active site, substrate specificity decreases, which is reflected in increased K_m values (Michaelis constant).^{211,213,214} This means that the psychrophilic enzyme substrate complex (ES) is less stabilized compared with the mesophilic ES complex ($ES_P > ES_M$). This reduces the energy required to reach the transition state (ES^*) which accelerates the reaction.²¹⁵ Second, it increases flexibility in parts of the enzyme involved in subtle conformational changes during the reaction. This means that fewer bonds

need to be broken and reformed to reach the ES conformation associated with ES^* . This reduces the amount of energy required to reach ES^* ($ES_M^* > ES_P^*$). These two effects can occur in isolation or in combination in an enzyme. From a protein engineering perspective, these properties of psychrophilic enzymes are very promising. The increased flexibility can lead to higher promiscuity of the enzyme.²¹⁶ Combined with the higher enzymatic activity, they are an excellent starting point for specificity engineering to create new substrate recognition.²⁰⁹

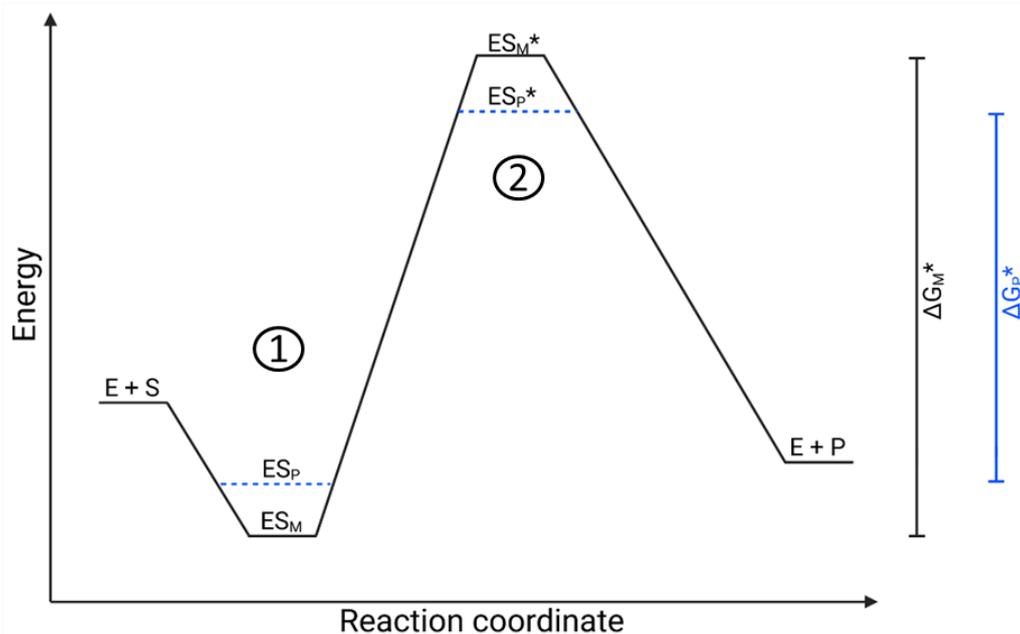


Figure 15: An energetic overview of the two main strategies for enzymatic adaptation to cold environments. The two main free energy changes (ΔG^*) that can occur in psychrophilic (p) enzymes compared to mesophilic (m) enzymes. Enzyme (E), substrate (S), enzyme substrate complex (ES), product (P). 1) Increased flexibility at or near the active site, increasing K_m and thus decreasing ΔG^* . 2) Increased flexibility in parts of the enzyme involved in subtle conformational changes during the reaction (e.g., surface loops). Fewer bonds must be broken and re-formed to achieve ES^* resulting in lower ΔG^* .

Although psychrophilic enzymes possess a number of promising features, most of them are unfortunately not stable enough to be used under standard biotechnological or normal laboratory conditions, at least without engineering them, due to their increased flexibility.

To sum up the requirements for high stability of a thermophilic enzyme, the trade-offs between activity and stability must be understood. The key assumption is that protein dynamics must be preserved at different ambient temperatures to maintain the structural integrity and functional state of the enzyme. These dynamics are inversely related to structural stability: the lower the mobility of the protein structure, the higher its stability, and vice versa. For psychrophilic amylase, there is a strong selection pressure for high activity at low and intermediate temperatures. In nature, this is achieved through the evolution of local structural changes of the protein/enzyme by adjusting the conformational energy landscapes through mutations, especially in the transition states of the activated substrate. Directed evolution of psychrophilic enzymes in the laboratory will follow exactly these principles.

1.5. Aim of this Study

The aim of this study was to increase the efficiency and substrate scope of the PyIRS-based orthogonal translation systems. As mentioned in the introduction, all PyIRS OTS which could be characterized in laboratory settings are rather catalytically inefficient. Approaches were sought to increase the efficiency of PyIRS enzymes and naturally find more efficient variants.

Two main strategies were pursued to increase efficiency: First, I have investigated whether the probability of mutation-induced protein aggregation, which in turn reduces OTS efficiency, is higher in engineered PyIRS OTS. In this context, an approach to eliminate this aggregation was developed. Second, PyIRS OTSs from extremophilic organisms were investigated with respect to possible efficiency enhancements. In addition, the mesophilic and extremophilic enzymes were examined for substrate promiscuity, since it is known in the literature that some psychrophilic enzymes exhibit higher promiscuity. If this is also true for the psychrophilic PyIRS OTSs, this would open up a multitude of possibilities to encode an even larger number of chemically diverse substrates.

2. Results and Discussion

2.1. General ncAA Incorporation Readout and PyIRS OTS Setups

To test the efficiency of ncAA incorporation, a superfolder GFP (sfGFP²¹⁷) with an N-terminally fused small ubiquitin-related modifier (SUMO)²¹⁸ or a construct with an N-terminally fused elastin-like polypeptide (ELP, see **chapter 2.5.1**) was used as a reporter protein. The SUMO-tag serves two purposes. First, by combining it with an N-terminal His₆-tag, the ratio of truncated to full-length reporter protein can be obtained in an IMAC purification step and relative changes in OTS efficiency can be observed by SDS-PAGE analysis (see **chapter 2.2.3**). Second, the sfGFP reporter proteins always have an amber stop codon at position R2. The residues at the N-terminus may play a crucial role in determining the cytosolic half-life of proteins *in vivo*. This phenomenon is referred to as N-end rule.²¹⁹ Generally destabilizing and therefore half-life-shortening residues (to the timescale of minutes²²⁰) typically include Tyr, Phe, Trp, Leu, Lys and Arg (and most likely derivatives thereof).²²¹ When containing a C-terminal His₆-tag, the SUMO-tag ensures that incorporated ncAAs that are potentially half-life-shortening can still be observed with a robust readout. Reporter proteins generally contain a C- or N-terminal His₆-tag. The sfGFP-based fluorescence readout is the simplest approach for this purpose because the fluorescence intensity of intact cells is directly correlated with the amount of protein produced. The sfGFP-based reporter constructs contain between one to five in-frame stop codons, while the ELP variants contain up to 60 and are indicated in the figure captions. The reporter protein sequences can be found in **chapter 6.1.7**. The ncAA concentrations are also given in the figure captions or mentioned in the corresponding text.

All PyIRS OTS in this study are in identical vector systems (pTECH vektor¹⁴³) with an *lpp* promoter driving PyIRS expression and a *proK* promoter for the tRNA^{PyI}. All PyIRS sequences were codon-optimized for *E. coli* and, based on a homology model, possessed the mutation corresponding to *MbPyIRS*(Y349F), as it is known to enhance aminoacylation and OTS efficiency in general.^{163,222} Since all tRNA^{PyI} from *Methanosarcina* vary only at a maximum of two positions (see **chapter 2.4.5.8.1**) and are generally known to perform comparably, PyIRS OTS from the genus *Methanosarcina* always possessed the tRNA^{PyI} from *M. mazei* (which was the original tRNA on the obtained pTECH vector, Addgene #104073). For the other *Methanosarcinales* constructs the organisms corresponding tRNA^{PyI} was used unless otherwise indicated.

2.2. Increasing the PyIRS OTS Efficiency Using Genetically Fused Solubility Tags

Enabling the incorporation of new ncAAs commonly requires mutations in the aaRS active site. Since most mutations are destabilizing, it has been recognized that engineering protein properties, including aaRS substrate specificity, can have substantially detrimental effects on the overall protein stability.^{196,197} This can lead to a lower abundance of correctly folded enzyme, which in turn can further induce protein aggregation.^{191,192} Surprisingly, there are no studies in which the physico-chemical properties of PyIRS are rationally modified, to determine if poor protein folding and low solubility are major bottlenecks for OTS efficiency *in vivo*. Understandably, the majority of application-driven aaRS engineering efforts have been directed towards substrate specificity - to allow incorporation of new ncAAs and genetic encoding of chemical functionalities. A classical approach to increase the production of soluble and active recombinant proteins is to genetically fuse a well-folding, highly soluble protein domain to the N- or C-terminus of the target (see **chapter 1.3.2.1**).¹⁹⁸ This approach has never been tested for improving the PyIRS-based OTS performance. The strategy presented here targets the N-terminus for alteration and the proposed working mechanism is shown in **Figure 16**. The active site and surrounding shell residues of the aaRS enzyme remain unaltered, facilitating transfer of successful findings to other enzymes. A set of solubility-tags (**Table 3**) was fused to the N-terminus of a *MbPyIRS* enzyme variant previously engineered for the incorporation of *S*-allyl-L-cysteine (Sac, **1**)¹⁴⁴ and subsequently screened for improved efficiency. Afterwards, the generality of the findings was assessed to gain knowledge about the applicability to other PyIRS systems.

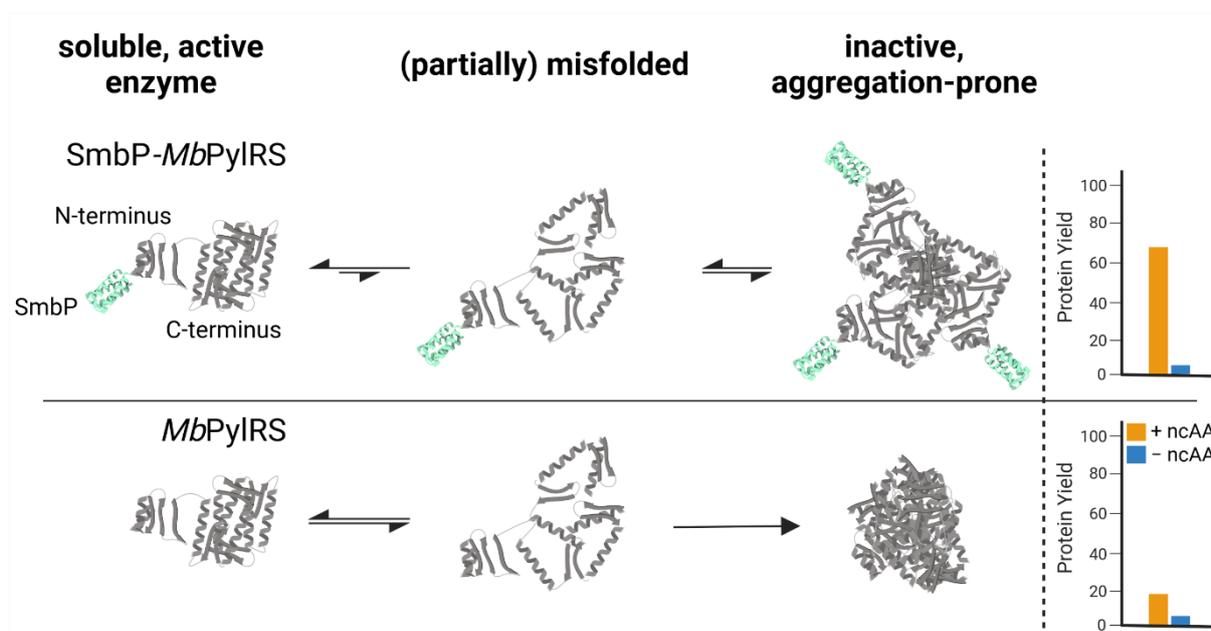


Figure 16: Proposed mechanism of how the small metal binding protein (SmbP) tag contributes to *MbPyIRS* being more catalytically active *in vivo* and increases the yield of ncAA-modified target protein.

Besides serving as the model substrate, the attractiveness for Sac incorporation into peptides and proteins stems from the introduction of a bioorthogonal functional group (i.e. an alkene tag) and its small size compared to e.g. Pyl (**1a**) derivatives equipped with the same functional group. To date, Sac represents one of the smallest non-aromatic ncAAs which can be genetically encoded via the PylRS system. Moreover, it is a low-cost substrate with high pH-, temperature- and aqueous-stability.^{144,223} Installed into peptides and proteins, Sac enables a variety of straightforward bioorthogonal protein conjugation reactions.^{224,225}

Early genetic code expansion studies have revealed low *in vitro* solubility of the PylRS enzyme.¹⁷⁶ Subsequent works revealed improvements stemming from N-terminal mutations and/or exchange of the N-terminal domain of *MbPylRS* by the more soluble counterpart of *MmPylRS*, creating a chimeric aaRS enzyme.¹⁴³ These studies clearly indicate that there is potential for OTS improvement besides optimizing the enzymatic recognition and activation of substrates. Given that other components such as intracellular ncAA and orthogonal tRNA abundance are not limiting, poor recombinant aaRS solubility would translate into a lower fraction of active enzyme and concomitant lower OTS efficiency. The low solubility is predominantly caused by the hydrophobic N-terminal domain which is essential for recognition, binding and charging of tRNA^{Pyl} *in vivo*.¹⁷⁵ This chapter shows a straightforward genetic strategy to improve orthogonal translation, increasing the production of site-specifically modified target proteins per amount of non-natural amino acid supplied.

2.2.1. Choice of Used Solubility Tags

To address the intrinsic solubility problem of the widely used PylRS scaffold, the focus was placed on the *Methanosarcina barkeri* enzyme (*MbPylRS*), whose solubility is even lower compared to *MmPylRS*.^{176,181} This engineered enzyme was intentionally chosen and not the wild-type under the assumption, that if aaRS solubility/folding are major bottlenecks for *in vivo* function, the beneficial effects should be more pronounced for relatively inefficient enzyme variants. The chosen variant is a double mutant reported to yield 0.6 mg enhanced green fluorescent protein (EGFP) modified with Sac at a single site (yield per liter of bacterial culture).¹⁴⁴ Its two active site mutations C313W:W382S crucial for activation of the ncAA and further two beneficial N-terminal mutations T13I:I36V (identified previously, cf. **Figure 70**) were introduced into the codon-optimized *MbPylRS* sequence (leading to the *MbSacRS* variant). Mutation Y349F mutation was also included by default, as it is known to generally enhance aminoacylation.¹⁶³ Nine common protein fusion partners with diversity in size and physico-chemical properties were chosen (**Table 3**). The assumption was, that by keeping the sequence of the aaRS enzyme unaltered, the ncAA specificity and activation kinetics are maintained.

Table 3: List of fusion tags used for solubility enhancement of *MbSacRS with brief descriptions, molecular weight (MW) and references.**

abbreviation	full name and description	MW [kDa]	reference
InfB(1-21)	First 21 nucleotides of the <i>E. coli</i> InfB gene encoding translation initiation factor IF-2. The mRNA secondary structure in the translation initiation region is weak, promoting ribosomal binding and high translation efficiencies.	0.8	J. G. Hansted <i>et al.</i> , 2011 ²²⁶
10xD	Ten aspartate residues: A polyanionic tag. The repulsive electrostatic interactions caused by the negative charge of the peptide tag are expected to enhance solubility and to facilitate correct protein folding by delaying protein aggregation.	1.2	T. Rathnayaka <i>et al.</i> , 2011 ²²⁷
10xR	Ten arginine residues: The improvement in protein solubility is attributed to the repulsive electrostatic interactions between similarly charged tags or protein stretches, which prevents aggregation and allows sufficient time for correct folding.	1.5	J. C. Smith <i>et al.</i> , 1984 ²²⁸
GB1	Immunoglobulin-binding domain B1 of protein G from group G <i>Streptococcus</i> : This domain possesses high thermal stability with a melting temperature (T_m) of 87 °C <i>in vitro</i> and completely reversible thermal denaturation, indicating excellent folding and solubility of the tag.	6.2	A. M. Gronenborn <i>et al.</i> , 1991 ²²⁹
Fh8	<i>Fasciola hepatica</i> antigen: 8 kDa calcium binding protein from the parasite <i>Fasciola hepatica</i> . Known to improve solubility of difficult-to-express recombinant target proteins upon fusion.	8	S. J. Costa <i>et al.</i> , 2013 ^{230,231}
SmbP	Small metal-binding protein from <i>Nitrosomonas europaea</i> : A monomeric protein characterized by a series of 10 repeats of a seven amino acid motif with an unusually high number of histidine residues. Its unique sequence without similarity to other proteins in current databases is considered to be a metal scavenging motif with an important role in cellular copper management.	9.9	T. Vargas-Cortez <i>et al.</i> , 2016 ²³²
SUMO	Small ubiquitin-related modifier: 100 AA residue protein which modulates protein structure and function by covalent modification of target proteins in eukaryotes. Well documented enhancer of recombinant protein expression and solubility.	11.2	M. P. Malakhov <i>et al.</i> , 2004 ²¹⁸
Trx	<i>E. coli</i> thioredoxin: A small, ubiquitous protein with a dithiol-disulfide in an exposed active center. Thioredoxin facilitates reduction of various proteins through the reversible oxidation via cysteine thiol-disulfide exchange.	11.8	A. Holmgren, 1985 ²³³
NusA	<i>E. coli</i> N-utilization substance A: Predicted and found to enhance cytoplasmic solubility of target proteins in <i>E. coli</i> using a statistical solubility model.	55	G. D. Davis <i>et al.</i> , 1999 ²³⁴

***The *MbSacRS* fusion partner is 419 AA (47.5 kDa) in size.**

2.2.2. Comparison of Different N-terminal Fused Solubility Tags

Comparing *MbSacRS* as reference to nine fusion proteins with solubility tags shows that all constructs except for Fh8-*MbSacRS* are functional *in vivo* (Figure 17). This is evident from increased fluorescence intensities in presence of ncAA supplementation. The best-performing construct has an N-terminal SmbP-tag (9.9 kDa in size), followed by the InfB(1-21) and 10xD-tag. Notably, the InfB(1-21)-tag is the smallest fusion partner (0.8 kDa) tested herein, known in literature to increase expression levels rather than the solubility of proteins.

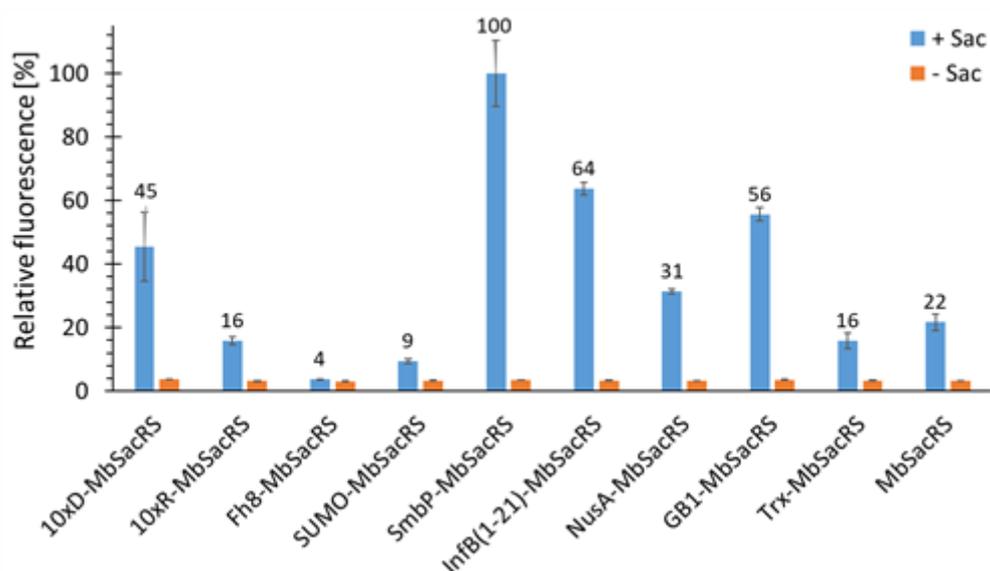


Figure 17: Comparison of OTS efficiency of 10 different *MbSacRS* constructs measured by fluorescence of intact cells of *E. coli* BL21(DE3) expressing the sfGFP(R2 amber) construct. Reporter signal is from stop codon suppression in the presence of (+ Sac = 2 mM) or without ncAA (-Sac) supplementation.

Strikingly, even small modifications (e.g. His₆-tags for aaRS purification) were previously shown to reduce the *in vivo* activity of engineered *MbPylRS* variants.²³⁵ Given the increased efficiency, it was next aimed to check if even higher improvements can be achieved by using multiple tags. It was chosen to combine tags with a different mode of action, being InfB(1-21) on the one hand and SmbP, GB1 and 10xD on the other. Unfortunately, no synergistic effects could be observed (**Figure 18A**). The collected data on single and combined tags do not reveal a clear pattern to rationalize the observed effects. Improvements in ribosomal ncAA incorporation could arise from enhanced aaRS expression levels, folding, solubility and combinations thereof. Three of the initially chosen tags reduced the efficiency compared to unmodified *MbSacRS*, once again highlighting that the aaRS scaffold is sensitive to certain terminal modifications. Albeit the multi-faceted nature of protein folding and solubility, previous studies have shown benefits of fusion partners which are highly robust towards thermal and chemical unfolding (like GB1). Moreover, it should be noted that both on nucleotide and amino acid level, the N-terminal sequence plays an important role for the initiation and efficiency of protein synthesis, which could promote changes in aaRS expression levels.²³⁶ After screening for the effect of different tags, the performance of the three best constructs were characterized in detail. Wild-type sfGFP served as reference for the maximum recombinant protein production. Based on fluorescence intensities, the overall efficiency for the suppression of one in-frame amber stop codon reaches around 56 % (**Figure 18B**).

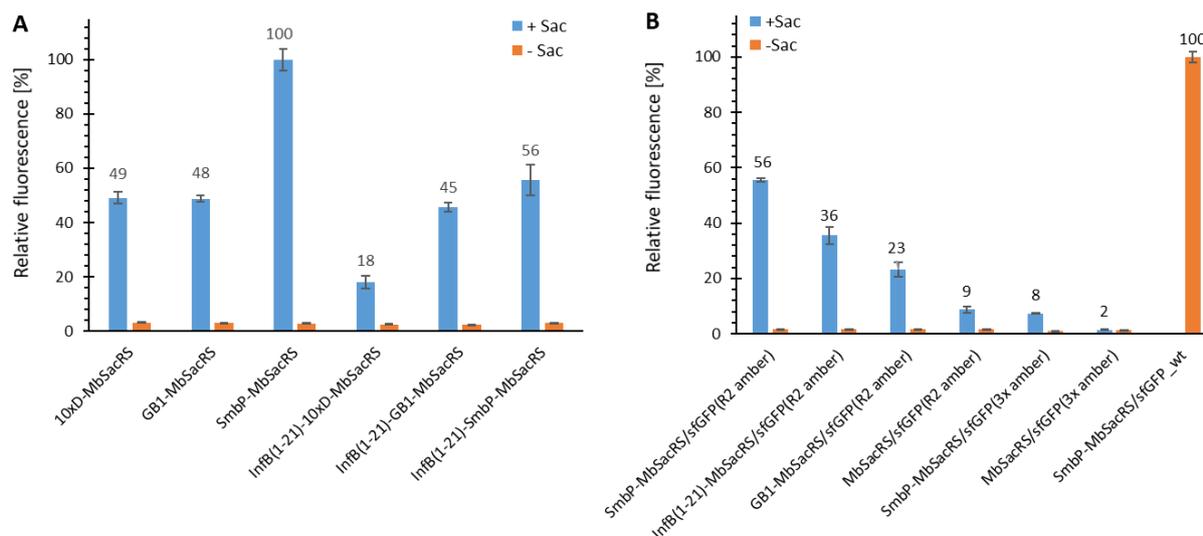


Figure 18: Comparison of OTS efficiency for *MbSacRS* fusion constructs measured by fluorescence of intact *E. coli* BL21(DE3) cells. **A**) Best-performing solubility tags were tested alone and in combination. **B**) Comparison of fusion constructs to the untagged starting enzyme. Reporter constructs for the suppression of one or three stop codons were sfGFP(R2 amber) and sfGFP(3x amber), respectively. Wild-type sfGFP without an in-frame stop codon was included as benchmark. Endpoint measurements the presence (+ Sac = 2 mM) or absence of ncAA supplementation (- Sac).

This level of performance is very high especially for a PylRS system, far exceeding commonly reached levels for stop codon suppression at this position.^{143,144} This fact led to the evaluation if even more than one in-frame stop codon could be efficiently suppressed with the improved setup. The best aaRS construct (SmbP-*MbSacRS*) was co-expressed with a sfGFP gene containing three in-frame stop codons. Protein production reached the same level as the untagged starting enzyme (*MbSacRS*) achieved for suppression of a single stop codon (**Figure 18B**). Co-expression of untagged *MbSacRS* did not lead to a detectable suppression of three stop codons.

2.2.3. Analytics of Sac Incorporation

To confirm the microtiter-scale intact cell fluorescence assays, the reporter constructs were expressed in bacterial cells in shake flasks and subsequently purified to determine target protein yields. This is the key metric for OTS performance and reflected by corresponding ratios between full-length target protein and truncation products. The reporter construct with an N-terminal His₆-SUMO-tag allows simultaneous purification of both protein species. The positive control (wild-type sfGFP as a metric for maximum recombinant protein production) was likewise expressed in cells carrying the improved SmbP-*MbSacRS* OTS to reach a similar metabolic burden for *E. coli* cells transformed with two plasmids. Addition of the ncAA to these cells had a negligible effect on the fluorescence signal (data not shown). The presence of Sac in the target protein was confirmed via electrospray ionization mass spectrometry (ESI-MS)

(**Figure 19A** and **B**). Obtained purified protein yields for the different setups are in good agreement with data from the fluorescence assays (**Table 1**).

Table 4: Yields of ncAA-modified target protein as measure for OTS efficiency with 2 mM Sac provided in the growth medium.

reporter construct	aaRS construct	protein yield [mg L ⁻¹]*
sfGFP wild-type	SmbP- <i>MbSacRS</i>	40.2
sfGFP(R2 amber)	SmbP- <i>MbSacRS</i>	15.2
sfGFP(R2 amber)	<i>MbSacRS</i>	1.2
sfGFP(3x amber)	SmbP- <i>MbSacRS</i>	0.8

*The amount of purified and dialyzed protein is given in mg per liter of bacterial culture.

The difference in reporter protein production is clearly visible by the naked eye (**Figure 140**). Suppression of a single amber codon yielded purified protein amounts equivalent to 50 % of wild-type, highlighting the improved performance of the SmbP-tagged PylRS enzyme. Using ESI-MS, almost no truncation products were detected for purified samples obtained with this co-expression setup (**Figure 19A**). This also strongly emphasizes the higher efficiency of this PylRS construct. Increased amounts of full-length target protein are further evident from SDS-PAGE analysis when compared to the starting aaRS (**Figure 19C** and **D**). For the challenging construct with three in-frame amber stop codons, SDS-PAGE analysis of purified target protein indicates incorporation of three Sac moieties, but also reveals a predominant formation of truncation products (**Figure 138**). This was expected, as it is well known that the amount of truncation product increases with the number of suppressed stop codons.¹⁸⁸ The relatively low amount of full-length, triple-modified target protein made a switch to a C-terminally His-tagged reporter for selective purification necessary. MS analysis of this construct confirmed the incorporation of Sac at three positions (**Figure 143**). To evaluate the robustness the findings, Sac was incorporated into two other sequence contexts and protein structures. The first construct is the blue chromoprotein from the coral *Acropora millepora* (*amilCP*, also a beta barrel in structure as sfGFP) where Sac was incorporated into the sequence context of an N-terminal tag (**6-R11-1**) evolved for highly efficient amber suppression.¹⁵¹ With flanking glycine residues, this small region would be expected to be unstructured and solvent-exposed. To evaluate Sac incorporation in a different context of protein structure, the PDZ3 domain of postsynaptic density protein-95 (PSD-95) was chosen. PDZ domains are of special interest as they represent highly abundant protein-protein interaction modules. These structural domains are found in the proteins of a variety of signal transduction complexes in multiple organisms (bacteria, yeast, plants and viruses). Taking the mouse genome as an example, they can be found 928 times.^{237,238} Sac was successfully incorporated at position F325, a location previously used for ncAA installation and part of the peptide ligand binding pocket.¹²⁸ The results of Sac incorporation into *amilCP* and PDZ are consistent with the previous sfGFP

results (**Figure 20**). There is clearly more target protein production with the SmbP-tagged aaRS, in the case of the chromoprotein amilCP even visible in the harvested bacterial cells by the naked eye (**Figure 141**).

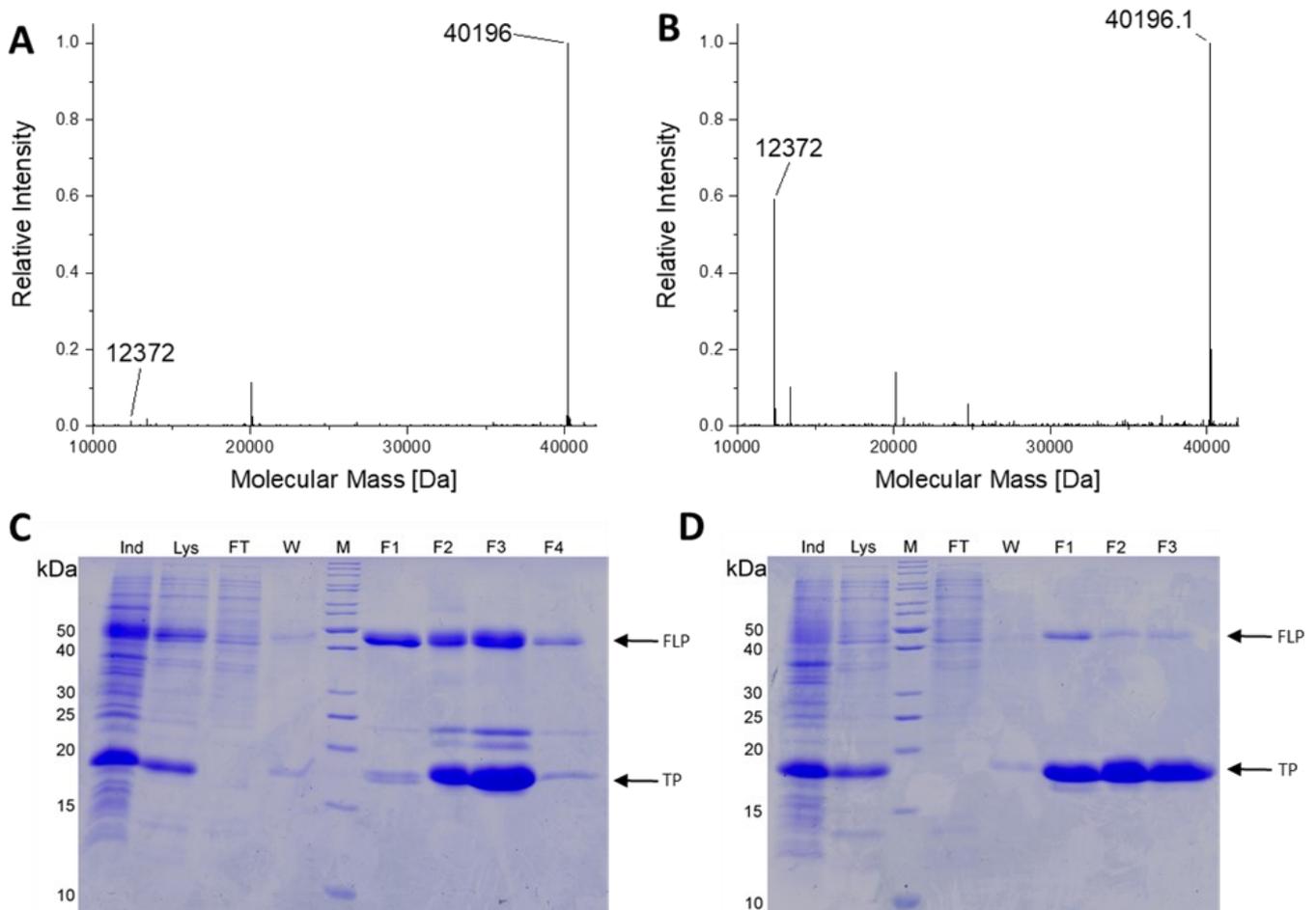


Figure 19: Analytics and gel profiles of the incorporation experiments. **A**) and **B**) Deconvoluted ESI-MS spectra of SUMO-sfGFP(R2Sac) production in *E. coli* BL21(DE3). **A**) Reporter co-expression with SmbP-*MbSacRS*. Expected protein mass of SUMO-sfGFP(R2Sac): 40194.9 Da. Observed mass: 40196 Da. **B**) Co-expression with untagged *MbSacRS*. Expected mass of SUMO-sfGFP(R2Sac): 40194.9 Da. Observed mass: 40196.1 Da. Expected mass of SUMO truncation product: 12372.8 Da. Observed mass: 12372 Da (**A** and **B**). **C**) and **D**) SDS-PAGE analysis of purified SUMO-sfGFP(R2Sac) reporter protein produced via co-expression of **C**) SmbP-*MbSacRS* and **D**) unmodified *MbSacRS*. Sample abbreviations: Whole cell extract of induced culture (Ind), soluble cellular lysate (Lys), liquid chromatography flow-through (FT), column wash of bound protein (W), collected fractions of eluate (F1-3/4), protein ladder (M), full-length product (FLP), truncation product (TP).

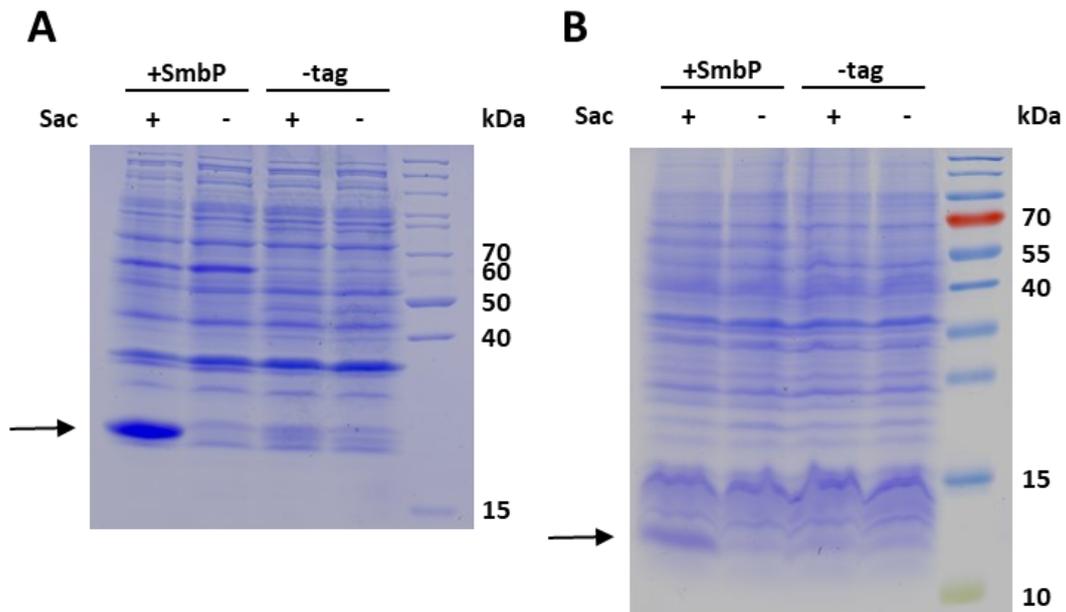


Figure 20: SDS-PAGE analysis of Sac incorporation into amilCP (26.7 kDa, **A**) and PDZ (11.7 kDa, **B**). The gels show samples of the soluble whole cell lysates of induced cultures produced with 2 mM Sac supplementation (+) and without (-), respectively.

2.2.4. Analysis of PyIRS Solubility

To confirm that the SmbP-tag increases the abundance of soluble enzyme in the cytoplasm of *E. coli*, the soluble and insoluble cell fractions for SmbP-*MbSacRS* and *MbSacRS* expression were analyzed via SDS-PAGE (**Figure 21**). For analytical reasons, this included an aaRS production setup with strong overexpression to facilitate detection. Accordingly, the aaRS gene was transferred into a pET plasmid vector with a strong T7 promoter. SDS-PAGE shows that most of the engineered aaRS with and without tag resides in the insoluble fraction. It is thus likely inactive *in vivo*, for instance present in bacterial inclusion bodies. The corresponding soluble fractions, however, display distinct differences in aaRSs abundance. SmbP-*MbSacRS* is clearly overproduced in soluble form, whereas untagged *MbSacRS* is hardly detectable in the corresponding fraction. In the same fashion, the aaRS co-expression setup was analyzed. For the *lpp* promoter driving aaRS production, clear overproduction of SmbP-*MbSacRS* in the soluble cell extract fraction is visible; just a small band is detectable in the corresponding insoluble fraction. In contrast, *MbSacRS* co-expression generates only a small detectable band in both fractions, indicating a relatively low aaRS production level. Comparing the findings of both aaRS promoter systems, it is noteworthy to mention that the abundance of soluble SmbP-*MbSacRS* appears comparable. This indicates that aaRS production reaches a similar level of soluble SmbP-*MbSacRS*, despite differences in promoter strength.

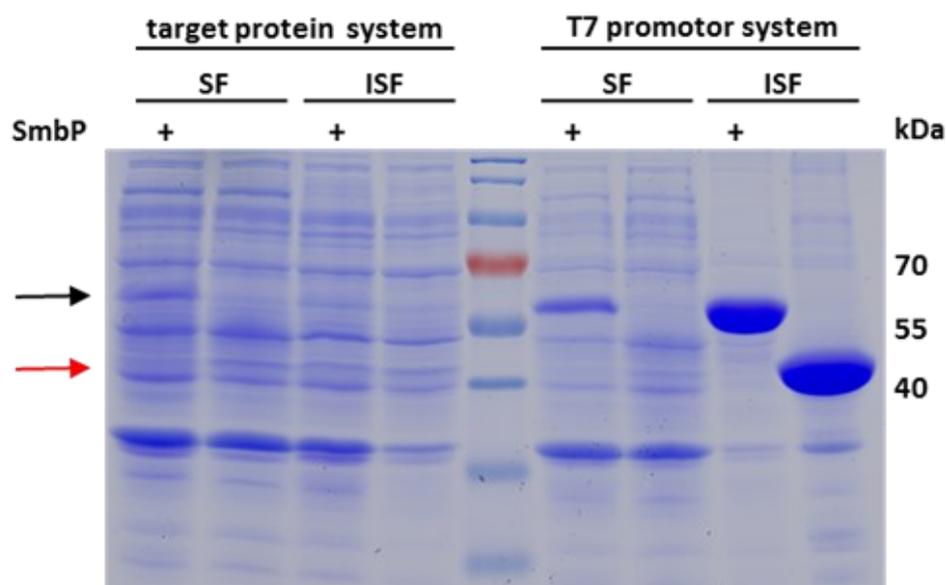


Figure 21: SDS-PAGE analysis of the soluble (SF) and insoluble (ISF) protein fractions of cell lysates. Left half: aaRS co-expression setup as used for ncAA incorporation (target protein production system, driven by *lpp* promoter). Right half: aaRS overproduction setup driven by the strong T7 promoter. Calculated molecular weights: SmbP-*MbSacRS* = 57.5 kDa (black arrow), *MbSacRS* = 47.6 kDa (red arrow).

2.2.5. Transferability of Tag Effects to Other PyIRS Variants

To investigate the generality of the N-terminal tagging strategy, different *MbPyIRS* variants were equipped with the SmbP-tag. Paired with the corresponding engineered enzyme, the following ncAA substrates were tested in a concentration dependent manner to gain detailed information about the OTS performance *in vivo*. First candidate was a *MbPyIRS* variant engineered for *o*-(2-nitrobenzyl)-L-tyrosine (ONBY (**41**)) incorporation.²³⁹ ONBY (**41**) is a versatile photocaged tyrosine derivative for example used to elucidate mechanisms of mammalian cell signal transduction and for light-controlled protein activation.²⁴⁰ The aaRS enzyme has two glycines among a total of four catalytic domain mutations (L270F:L274M:N311G:C313G) and also allows genetic encoding of bulky, reversibly photoswitchable azobenzene amino acids.²⁴¹ It is conceivable that the two side chain truncations are key for opening up the ncAA binding pocket towards accommodation of the bulky photocaged moiety. With the remaining mutations not in direct contact to the ncAA substrate, just the glycine double mutant was chosen. As in the case of *O*-*tert*-butyl-L-tyrosine (**39**) (Figure 23), the SmbP-tag drastically improved protein production for ONBY (320 %, Figure 22A). Considering that the two second-shell mutations (L270F:L274M) were omitted, the impact of the N-terminal tag on aaRS efficiency and the obtained protein production (even at low supplied ncAA concentrations) were astonishing. Mutations of aaRS active sites can result in an increased catalytic promiscuity.¹⁶⁵ This is especially true for the hydrophobic amino acid binding pocket of PyIRS, which is scarce in specific substrate interactions. Due to the enlarged binding pocket, it was considered and confirmed that the double Gly mutations allow the accommodation of other ncAA substrates (Figure 71 and Figure 72) which are even

bulkier than the substrates reported to be incorporatable with the double alanine mutant. This is plausible since the double Gly mutant's active site would be bigger. The activity of SmbP-*Mb*PyIRS(N311G:C313G) towards several ncAAs emphasizes the role of reducing the size of side chain moieties within the PyIRS pocket. Similar observations were made upon rationally engineering the enzyme's substrate tolerance via a double alanine mutation.¹⁶⁵

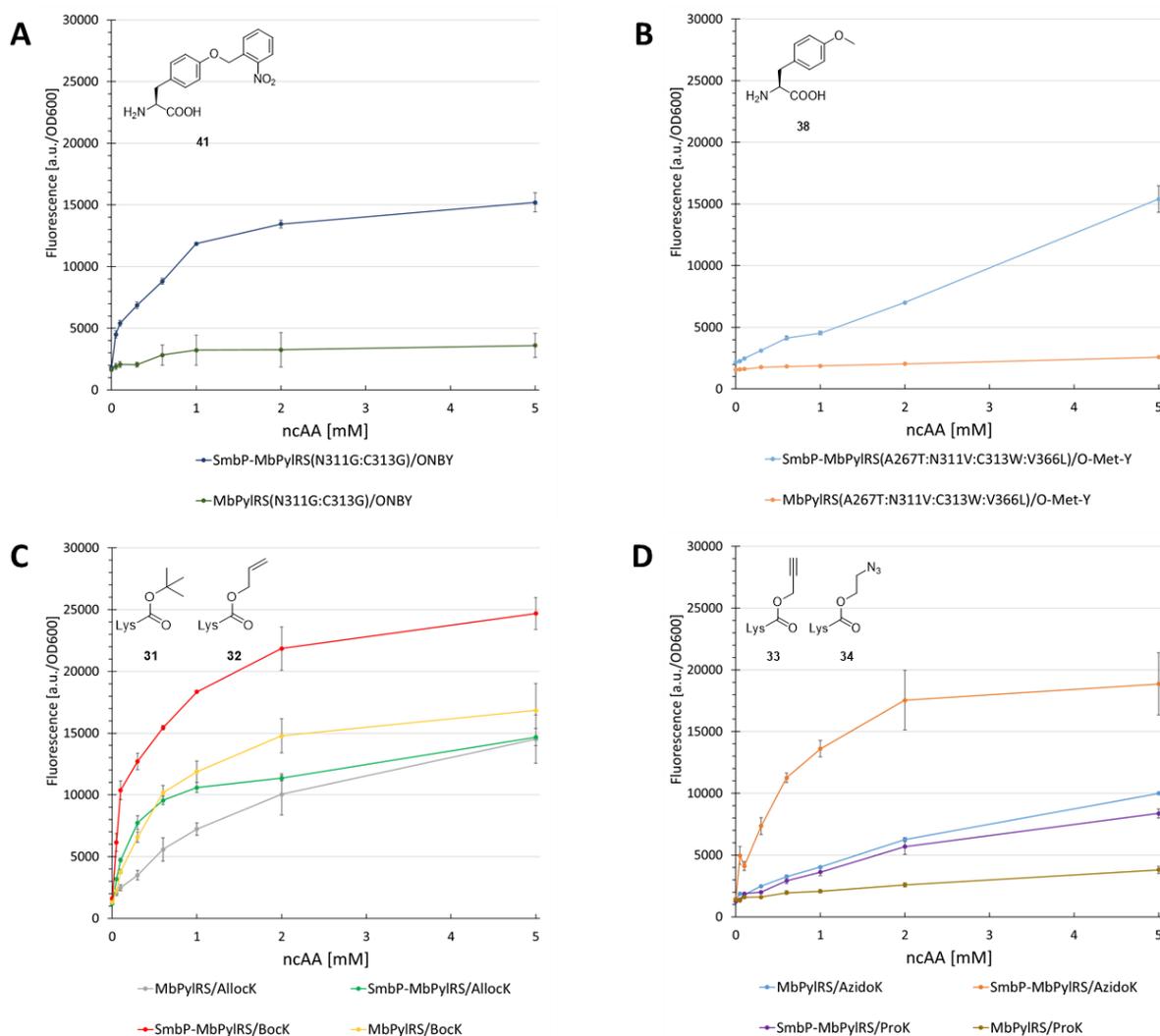


Figure 22: Concentration-dependent protein production for different *Mb*PyIRS/ncAA combinations. Endpoint measurements with different ncAA concentrations (0.05, 0.1, 0.3, 0.6, 1, 2, and 5 mM). The ncAAs are: ONBY = *ortho*-Nitro-benzyl-L-tyrosine (**41**), O-Met-Y = O-Methyl-L-tyrosine (**38**), AlloCk = *N*^ε-Allyloxycarbonyl-L-lysine (**32**), BocK = *N*^ε-*tert*-Butoxycarbonyl-L-lysine (**31**), AzidoK = *N*^ε-((2-Azidoethoxy)carbonyl)-L-lysine (**34**), ProK = *N*^ε-Propargyloxycarbonyl-L-lysine (**33**).

To further evaluate the transferability of the results, investigations of PyIRS variants evolved for activation of a small substrate (O-Methyl-L-tyrosine (**38**), **Figure 22B**) and *S*-propargyl-L-cysteine (SproC, **27**) (used as handle for the site-specific azide-alkyne Huisgen cycloaddition reaction, **Figure 73**) were conducted.^{146,242} Gratifyingly, increased protein yields were reached (200-490 %) upon aaRS tagging. For O-Methyl-L-tyrosine (**38**) installation, a linear response throughout the supplied ncAA concentration range was observed. No activity can be detected without the PyIRS tag, again highlighting detrimental effects of active site mutations. The

revealed transferability of the approach to other engineered *MbPylRS* variants led to the question if even the wild-type enzyme could profit from the tagging strategy. The ncAAs *N*^ε-Allyloxycarbonyl-L-lysine (AlloK (32)), *N*^ε-*tert*-butoxycarbonyl-L-lysine (BocK (31)), *N*^ε-((2-Azidoethoxy)carbonyl)-L-lysine (AzidoK (34)) and *N*^ε-propargyloxycarbonyl-L-lysine (ProK (33)) were screened, again in a concentration-dependent manner. Like *S*-propargyl-L-cysteine (27), AzidoK (34) and ProK (33) are commonly used as handles for the site-specific and bioorthogonal azide-alkyne HUISGEN cycloaddition reaction but exhibit a longer side chain length. All four ncAAs are established *PylRS* substrates, with BocK (31) more efficiently activated and charged to tRNA^{Pyl} compared to the other three. Protein production reached 170 % (0.1 mM BocK (31)) and 220 % (0.6 mM AlloK (32)) (Figure 22C), 245 % (0.6 mM AzidoK (34)) and 120 % (2 mM ProK (33)) (Figure 22D), highlighting that even the wild-type enzyme profits from the SmbP-tag. The efficiency observed for the tagged *MbPylRS* with BocK (31) supplied at very low concentrations is noteworthy. Almost half of the maximum protein production is already reached at 0.1 mM BocK (31). This is one order of magnitude below the commonly used ncAA concentration and far from quantities used in the majority of studies, where high ncAA concentrations (of up to 10 mM) are required to reach satisfactory protein production yields.^{145–147} Comparing the genetic encoding of BocK (31) and Sac (1), it is remarkable that the engineered SmbP-*MbSacRS* system can achieve even higher target protein yields. Substrate titration reveals that the very low efficiency of the *MbSacRS* starting enzyme can be boosted to even surpass the wild-type enzyme performance for an excellent, established substrate. In summary, all data underline the robust finding that the SmbP-tag increases the abundance of active *MbPylRS* *in vivo*, which leads to higher unnatural protein production yields. Especially for drastic changes within the engineered enzyme active site, the addition of the terminal tag compensates for the destabilizing effects and boosts the production of unnatural proteins.

2.2.6. Impact of Active-Site Mutations

Most natural enzymes are marginally stable at physiological temperatures. Recently, this has been specifically shown for *MbPylRS*.¹⁹³ Enzyme active site residues are inherently unfavorable for overall protein stability and mutations which drastically change the size and/or physico-chemical properties will likely aggravate this destabilization.¹⁹⁷ To investigate the hypothesis that this holds true for the mutations enabling Sac incorporation (via *MbSacRS* and its impactful C313W:W382S mutations), two control variants of the enzyme were created (Figure 23A and B).

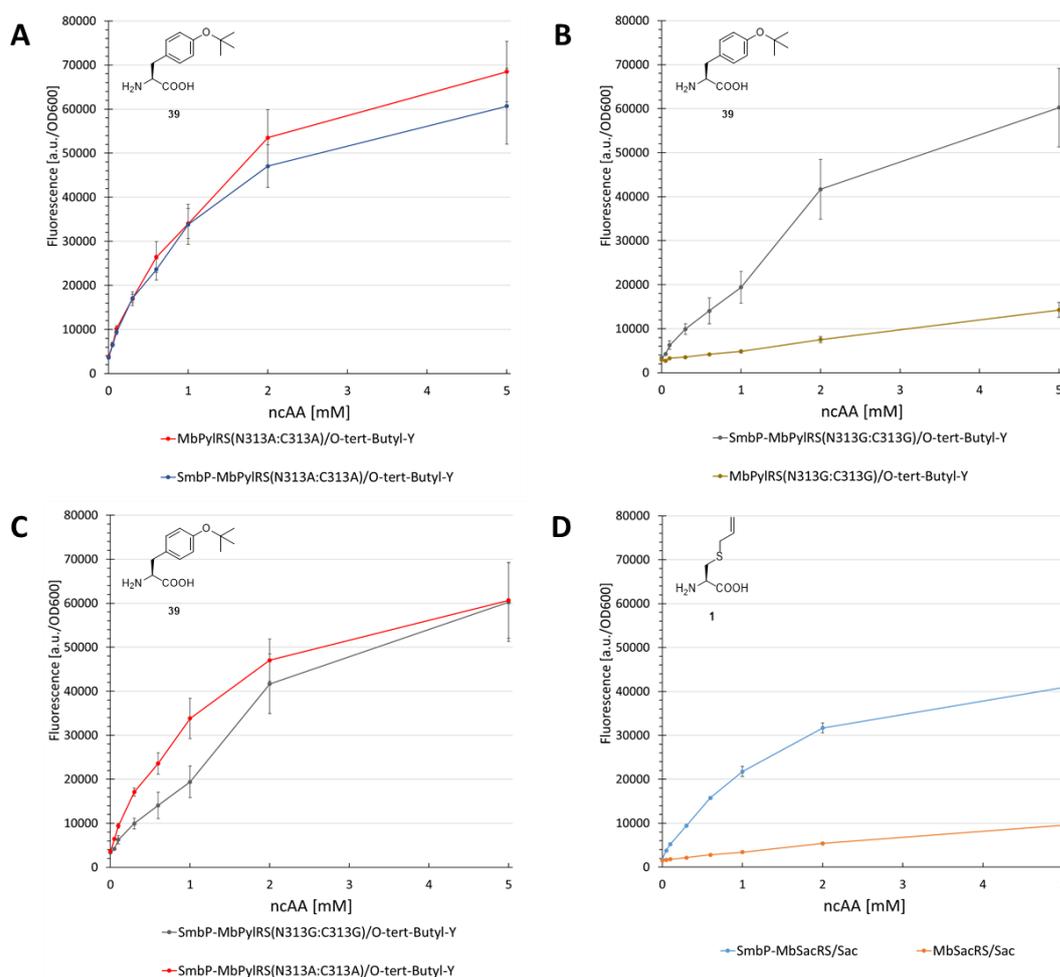


Figure 23: Concentration-dependent unnatural protein production for different *MbPylRS*/ncAA combinations. Endpoint measurements for ncAA concentrations of 0.05, 0.1, 0.3, 0.6, 1, 2 and 5 mM. The supplied ncAAs are O-tert-Butyl-Y = O-tert-butyl-L-tyrosine (**39**) and Sac = S-allyl-L-cysteine (**1**).

The mutations for these constructs targeted the polar active site residues N311 and C313. These positions are most frequently chosen for PylRS engineering as they are responsible for the recognition of the native Pyl substrate and have also proven to create a range of enzymes with new ncAA substrate recognition.¹⁶¹ Fortunately, mutating these residues does not impair the catalytic activity of PylRS as long as a sufficient level of overall enzyme stability is preserved. Both sites are distant to the catalytic center, and substrate recognition is encoded at the far end of the Pyl or ncAA side chain, respectively. The first control construct (where the SmbP-tag should have little to no effect) possesses the PylRS mutations N311A and C313A. Changing polar/charged residues in the active site to alanine is known to enhance the stability of enzymes but in most cases leads to a decrease in catalytic activity.²⁴³ The second construct (where mutations should severely destabilize the enzyme and therefore render the tag most effective) is the corresponding glycine double mutant used for ONBY (**41**) incorporation (see above). In most soluble proteins, the introduction of glycine disrupts secondary structure elements due to its high backbone flexibility.²⁴⁴ Both PylRS sites are part of the same β -sheet and come in close contact with the ncAA substrate. Taken the *MmPylRS* crystal structure as

homology model and the transferability of active site mutations between *Mm*PyIRS and *Mb*PyIRS as basis, both mutation sites N311 and C313 are part of the same β -sheet. They come in close contact with the ncAA substrate (<5 Å in PDB ID 3QTC; note that *Mb*PyIRS could not be crystallized thus far). Placing two glycines into this defined secondary structure element should perturb the folding and stability of the enzyme given the conformational flexibility of the smallest amino acid residue and the different backbone torsion angle preferences.²⁴⁴ Removal of the N311 and C313 side chains should profoundly destabilize the local secondary and subsequently overall aaRS structure. The initial assumptions were confirmed since no improvements could be detected for the SmbP-tagged double alanine mutant (**Figure 23A** and **Figure 74**). In contrast, protein production with the double glycine construct increased markedly, by 450 % (2 mM ncAA). This gain is comparable to the *Mb*SacRS improvement (540%, 1 mM Sac (**1**), cf. **Figure 23B** and **D**). Accepted by both enzyme variants, *O*-*tert*-butyl-L-tyrosine (**39**) provides a suitable substrate for this comparison. *O*-*tert*-Butyl-L-tyrosine (**39**) has been shown to work well as a structural reporter in NMR spectroscopy, but due to low protein yields of the PyIRS system initially reported, only the use of a *Mj*TyrRS-based system in bacterial cells proved applicable.²⁴⁵ The protein yield reported therein for a similar reporter construct paired with the *M. mazei* double alanine mutant was 2 mg/L.¹⁶⁵ On the basis of the used reporter, protein yields exceed 15 mg/L (note that this is estimated for a supply of 1 mM ncAA in contrast to 5 mM in the original publication).

2.2.7. Testing the Improved *Mb*SacRS for *In Vivo* Sac Synthesis and Incorporation

In the original Sac incorporation publication it was claimed that this can happen *in vivo* by just adding allyl mercaptan to the *E. coli* cultivation medium.¹⁴⁴ From a biotechnological point of view this has limited usefulness at best, since the ncAA (Sac (**1**)) which should be synthesized *in vivo* is very cheap. Allyl mercaptan is ten times cheaper (based on the data from TCI Germany in August 2021) than Sac (**1**) but is practically insoluble in water (this raises the question if allyl mercaptan can reach the cytoplasm of the cell in sufficient amounts), hazardous to water organisms and toxic (UN-number 1228) which creates additional costs for the wastewater disposal in comparison to the use of Sac (**1**). Additionally, allyl mercaptan functions as a histone deacetylase inhibitor which results in cytostatic cell effects and should therefore be handled with the utmost care.^{246,247} Histone acetylation and deacetylation play an important role in the epigenetic modification of chromatin with corresponding up- and down-regulation of gene transcription. All these aspects make the work with allyl mercaptan very difficult (cultivation and handling under the fume hood) and the disposal of the culture wastewater very complicated and expensive. Based on these considerations it is highly questionable if allyl mercaptan possesses any economical advantage in comparison to the use of Sac (**1**). Nevertheless, since a PyIRS with a highly increased incorporation efficiency of Sac (**1**) was developed it was sensible to test how this would change the incorporation

efficiency with *in vivo* generated Sac (1). Astonishingly, no Sac (1) incorporation could be detected when allyl mercaptan was added to the medium (**Figure 24**). In **Figure 24** clear incorporation is detectable with addition of Sac in the cultivation medium, even with 0.1 mM.

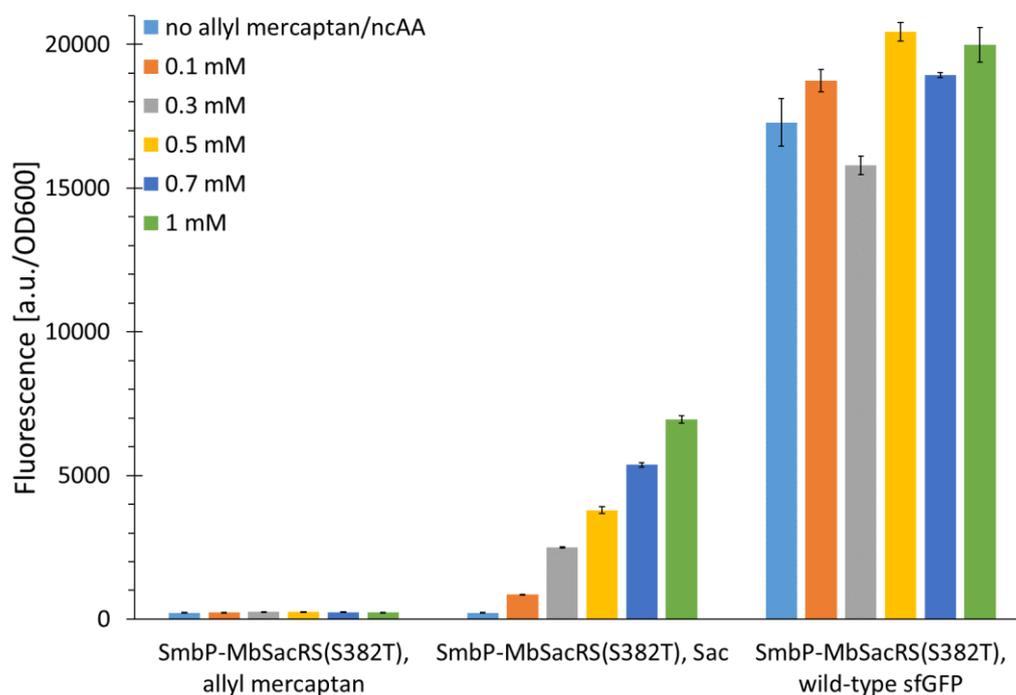


Figure 24: Concentration-dependent Sac incorporation comparison. Intact cell fluorescence of *E. coli* BL21(DE3) cells expressing the sfGFP(R2 amber) reporter with co-expression of *MbSacRS*(S382T). Endpoint measurements after 24 h of incubation

Zooming in on the allyl mercaptan part of **Figure 24** (**Figure 25**), the figure indicates that there could be Sac incorporation but the signal with 0.3 mM allyl mercaptan is just 10% higher than the background signal. For a robust ncAA incorporation a signal of at least 50-100% higher than the background signal is needed. It is conceivable that allyl mercaptan, which is extremely hydrophobic, could diffuse into the polystyrene of the 96-well plate and therefore give a false negative result. Diffusion of very hydrophobic precursors (azulene) to ncAAs have been observed in our group. To circumvent that problem all the allyl mercaptan cultivations were conducted in shaking flasks and just transferred to 96-well plates for the endpoint measurement.

After studying the original publication carefully, it is obvious that just background incorporation was measured when supplying allyl mercaptan, since the western blot method was used to “quantify” target protein yields. Generally, the question arises why western blots were used if enough protein could be obtained.

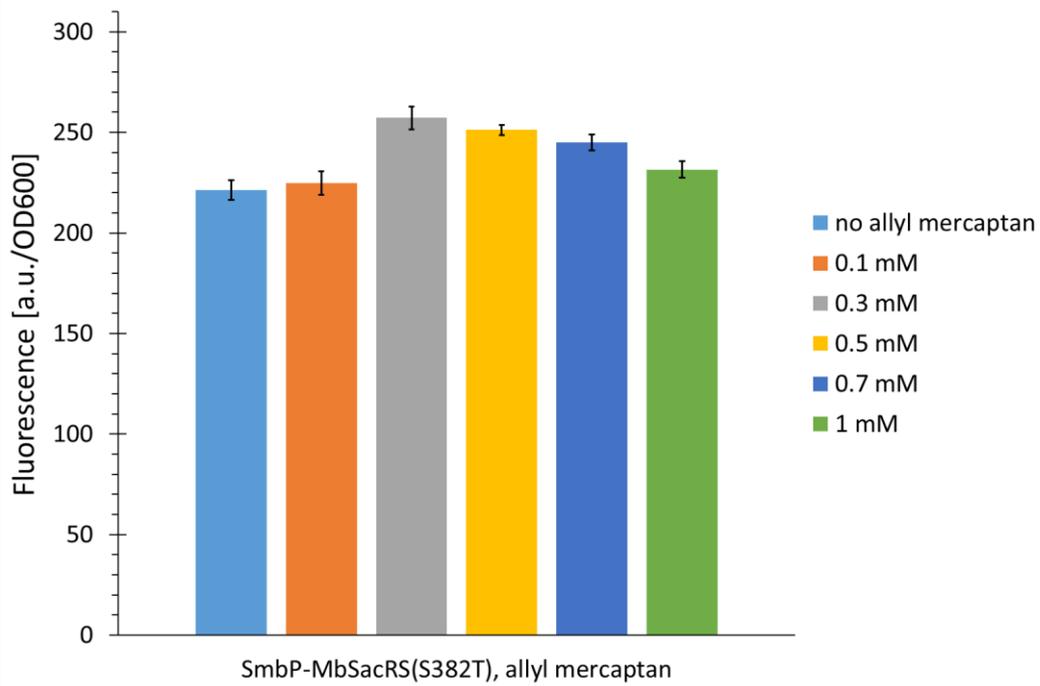


Figure 25: A portion of the diagram shown in **Figure 24** (“zoom in on the allyl mercaptan part”). Concentration dependent Sac incorporation comparison. Intact cell fluorescence of *E. coli* BL21(DE3) cells expressing the sfGFP(R2 amber) reporter with co-expression of *MbSacRS*(S382T). Endpoint measurements after 24 h of incubation

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

Research in the field of reprogrammed protein translation has now reached experimental and intellectual maturity: More than 200 non-canonical amino acids (ncAAs, i.e., a diversity that is an order of magnitude higher than that of the canonical amino acid repertoire) was introduced into proteins via various genetic code expansion routes: Selective pressure incorporation, stop codon suppression (SCS), fragment condensation, protein semisynthesis, and peptidomimetics²⁴⁸. It has been shown that AAs with non-proteinogenic functional groups can be used to manipulate, design, and elucidate protein structure, dynamics, function, allostereism, interactions, catalysis, folding, synthesis, trafficking, degradation, and aggregation^{127–133}.

Engineering aminoacyl-tRNA synthetase (aaRS)/tRNA pairs capable of recognizing, activating and loading ncAAs onto their cognate tRNAs is now a well-established strategy. It enables the site-specific ribosomal incorporation of ncAAs in response to a reprogrammed codon. The most commonly used approach for this purpose is in-frame stop codon suppression, targeting the amber stop codon^{124–126}. Hereby, the ncAA is incorporated in response to an in-frame stop codon placed at a predefined position in the protein coding sequence, ribosomally expressed either *in vivo* or *in vitro*^{124–126}. Most aaRS variants used for SCS so far are derived from *Methanosarcina mazei/barkeri* pyrrolysyl-tRNA synthetases (*MmPylRS/MbPylRS*) or *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase (*MjTyrRS*)^{124–126,161}. The archaeal origin and therefore distant phylogeny is responsible for their orthogonality in bacterial and eukaryotic cells¹⁶¹.

The native substrate of the PylRS is the rare proteinogenic amino acid Pyl (**1a**), a lysine analog with a 4-ethyl-pyrroline-5-carboxylate ring attached to the terminal amino function of the side chain (**Figure 26**). The wild type enzyme can activate several Pyl variants resembling ncAAs¹⁶¹. Moreover, catalytic promiscuity is widely exploited in both native and genetically engineered classes of PylRS enzymes to enable recognition, activation, and tRNA loading of the majority of all translationally active ncAAs.^{163–165,169} It should be noted, however, that the majority of incorporable ncAAs with the PylRS system are characterized with flexible, long-chained and bulky pyrrolysine analogs^{163,164} or shorter but still bulky aromatic substrates, especially phenylalanine¹⁶⁵, tryptophan¹⁷⁰ and histidine¹⁷² analogs. Therefore, a new class of PylRS enzymes capable of recognizing, activating and tRNA loading with shorter chain ncAAs endowed with useful functional groups is of great interest. Small ncAAs with shorter side chains containing azido, thioene, fluoro, cyano, and nitroso groups can be particularly useful, e.g., for FTIR, NMR, crosslinking, and spin labeling, because longer side chains are too flexible which usually results in a loss of spectral information or the necessary proximity for specific bioorthogonal reactions²⁴⁹. Moreover, still no efficient non-canonical counterparts are available

for Glu and Asp, which often form structurally important salt bridges or hydrogen-bond networks. It would be useful to modify these acidic residues, e.g., by removing their hydrogen bond donors or acceptors. Mimicking post-translational modifications of canonical amino acids (cAAs) with their genetically encoded ncAA counterparts are also attractive applications, to elucidate their functions.

The main reason for the substrate promiscuity of PyIRS is most likely the unique substrate binding mode with relatively nonspecific hydrophobic interactions in the large binding pocket of this enzyme. Therefore, it is not surprising that a multitude of ncAAs can be recognized and activated with very few mutations (the majority has just 2-4) mainly in the binding pocket¹⁶¹. For this reason, the PyIRS is predestined for the implementation of new functions¹⁷⁶. As mentioned in **chapter 1.3.2**, an enzyme should possess two key features to ensure successful recognition of new substrates. First, the target enzyme should have low levels of the desired new activity, which in case of PyIRS means enzymatic activity toward ncAAs that are highly divergent from the native substrate^{186,187}. Second, sufficient stability is required to buffer destabilizing mutations necessary for active site remodeling¹⁸⁹⁻¹⁹¹. Unfortunately, PyIRS is marginally stable under standard cultivation conditions in *E. coli*¹⁹³, which is also reflected by the low *in vitro* solubility of the enzyme^{176,194}. It was demonstrated that this drawback can be partly remedied with a solubility tag fused to the N-terminus of *MbPyIRS* (manuscript in preparation) which made this the enzyme of choice.

In this study, mutational analyses to elucidate the structure activity relationship of SacRS was performed. Based on this knowledge, several new *MbPyIRS* variants were engineered, in order to incorporate a variety of small side chain ncAAs. An entire library of small side-chain-containing ncAAs was used, which can be structurally and functionally categorized into five classes (**Figure 26**, detailed discussion in **chapter 2.3.1**). Briefly, they include (i) aliphatic analogs of Sac (**1**); (ii) bioorthogonal tags; (iii) small ncAAs with useful spectroscopic probes; (iv) methionine analogs; and (v) substrates with a terminal alkene as site-specific chemical cleavage site (being also bioorthogonal tags).

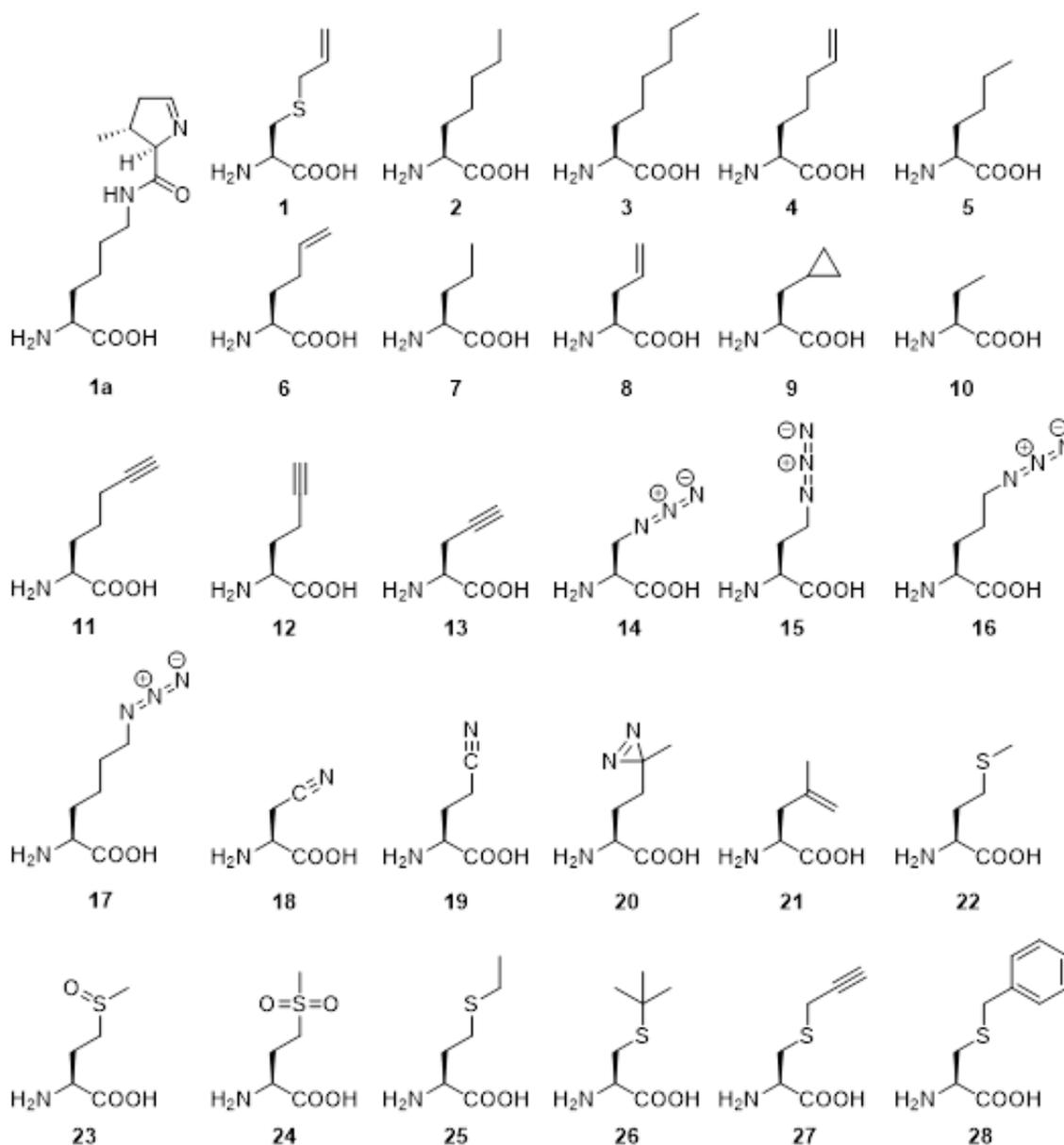


Figure 26: Survey of amino acids used in this study. Chemical structure of pyrrolysine (**1a**), S-allyl-L-cystein (**1**), (S)-2-aminoheptanoic acid (**2**), (S)-2-amino-octanoic acid (**3**), (S)-2-aminohept-6-enoic acid (**4**), (S)-2-amino-hexanoic acid (**5**), (S)-2-amino-hex-5-enoic acid (**6**),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

2.3.1. General Features and Perspectives of Used ncAAs

As discussed above, ncAAs with a side chain functionality closer to the protein backbone would be highly advantageous for several applications. An entire library of small side-chain-containing ncAAs was used. They can be structurally and functionally categorized into five classes, as shown in **Figure 26**. (i) Aliphatic Sac (**1**) analogs for structure activity elucidation (**2, 3, 5, 7, 10**). (ii) Site-specific bioorthogonal reaction handles which can be used for a variety

of bioconjugation reactions (**4**, **6**, **8**, **9**, **11**, **12**, **13**, **14**, **15**, **16**, **17**, **27**). These reactions include metal-free (e.g., click-chemistry, Staudinger ligations and strain-promoted cycloadditions) and transition metal-mediated (e.g., ruthenium-based olefin cross-metathesis or palladium based oxidative Heck and Sonogashira cross-coupling reactions) approaches.²²⁴ (iii) ncAAs used/could be used as biophysical probes in e.g., vibrational Stark, IR and NMR spectroscopy (**18**, **19**, **26**) and genetically encoded photo-crosslinker (**20**).^{245,249,250} (iv) Methionine analogs (**23**, **24**, **25**) as tools for elucidating the role of methionine oxidation in proteins, enzymes and cells.²⁵¹ (v) Substrates with a terminal alkene as site-specific chemical cleavage sites (**4**, **6**, **8**, **21**).^{252–255} For example, the cleavage reaction with substrate **8** proceeds presumably via iodolactonization which suggests that this reaction could also take place with substrates **6** and **8**. The transition state would change from a five-membered iodolactone to a 6/7 membered one.^{256–258} In contrast to the classical site-specific peptide cleavage method with cyanogen bromide at a methionine position, these ncAAs could be cleaved with non-toxic iodine under mild conditions.²⁵⁹

In yeast, substrates **3** and **4** have already been incorporated with an orthogonal *E. coli* leucyl-tRNA synthetase.²⁶⁰ Although the five aliphatic ncAAs (**2**, **3**, **5**, **7**, **10**) were used to estimate the size of the narrowed active site, some of these ncAAs could potentially help to address certain questions regarding posttranslational lipidation of proteins.^{261–264} Lipidation is also a common strategy to improve the pharmacokinetic properties of biopharmaceuticals, specifically to extend the systemic half-lives in patients with the corresponding increase in bioavailability.²⁶⁵ This opens a potentially interesting biomedical application area for these ncAAs. Notably, the ncAA substrates **5**, **9**, **12**, **15**, **17**, **20**, **25** have been incorporated in a residue specific manner (using auxotrophy based methods) but never in a site-specific mode, besides **17**.^{250,266–270} Substrate **9** contains a cyclopropane ring with properties closely resembling that of an olefinic double bond, which could be exploited for a whole range of site-specific bioorthogonal protein conjugation reactions.²⁷¹ These include the enzymatic halogenation with a haloperoxidase²⁷², enzymatic oxidation with a mono-oxygenase²⁷³, nucleophilic substitutions, electrophilic ring opening reactions and a plethora of other reactions.²⁷¹ Lastly, allylglycine (**8**) and propargylglycine (**13**) have recently been synthesized *in vivo* in *E. coli* and would therefore open up the possibility of coupling metabolic engineering and ncAA incorporation.²⁷⁴ This could eliminate the need to add ncAAs to the cultivation medium, which would drastically decrease costs and simplify associated applications.

2.3.2. General MbPylRS Setup

All PylRS variants used in this subsection possess a genetically fused N-terminal SmbP fusion solubility tag²³². The tag restores activity by drastically increasing the abundance of soluble and active enzyme in comparison to the untagged aaRS (see **chapter 2.2**). Most likely, this

phenomenon is due to an increase in kinetic stability and builds an improved and solid starting point for the presented enzyme engineering efforts. For clarity reasons the SmbP prefix is omitted in the notations for this chapter (2.3).

2.3.3. Testing *MbSacRS* for Aliphatic Substrates

In this section all *MbPylRS* constructs contain the N-terminal SmbP-tag. For clarity reasons this is omitted in the notations. The reporter protein is solely the SUMO-sfGFP(R2 amber) with an N-terminal or C-terminal His₆-tag. Also, assuming that the engineered constructs did not yield the most efficient *PylRS*s, high concentrations of ncAAs (10 mM, if not otherwise stated) were used throughout this study to prevent in-cell ncAA shortage and to enable the detection of even very low incorporation activity.

The starting hypothesis was, that due to the before mentioned relaxed substrate recognition of the *PylRS* system, the previously reported *SacRS* could incorporate close structural aliphatic analogs of Sac (1). If that would be true, that would be a good starting point to engineer the *SacRS* towards similar small-tag substrates. To scrutinize this hypothesis a *SacRS* variant was created based on a codon optimized *MbPylRS* sequence by introducing the two crucial active site mutations C313W:W382S (further referred to as *MbSacRS*). Also, the two previously identified beneficial N-terminal mutations T13I:I36V (²⁷⁵, cf. **Figure 70**) were introduced as well, which are also kept for all other constructs. Only very low incorporation of the aliphatic substrates for the *MbSacRS* could be detected (**Figure 75**). This specificity for very close structural analogs was surprising and is relatively uncommon for both native and mutant *PylRS* enzymes. As mentioned in **chapter 1.3.2** this low level of activity is still enough to engineer this enzyme for substrate recognition of aliphatic substrates. Since no well-functioning screening system was at hand to perform a directed evolution approach, it was decided to elucidate the *MbSacRS* structure-activity relationships (SARs) to rationally encode these ncAAs.

2.3.4. Elucidating the SAR of *MbSacRS* via Rational Mutation Studies

To date, there is no high-resolution crystal structure of *SacRS* available. Since there are only two *PylRS* mutations responsible for altering the substrate specificity to Sac (1), it was the best choice to perform rational mutation studies to elucidate the role of each residue regarding *Sac*-incorporation activity. Crystal structures of two *MmPylRS*, which guided the rational mutation approach, are displayed in **Figure 27**. Starting with residue S382, this position was reverted to wild-type Trp and less bulky amino acids. Since a serine was at position 382 in the original *SacRS*, it was also investigated whether polar functional groups are necessary for *Sac* (1) incorporation.

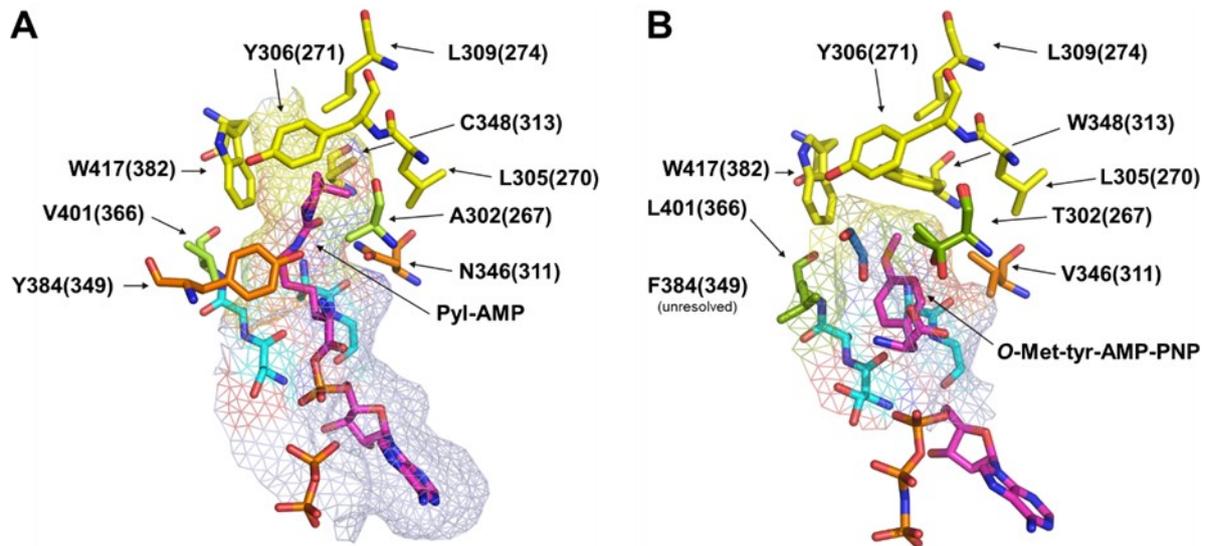


Figure 27: Microenvironment of active sites derived from the crystal structures of *MmPylRS* and the *MmOmeRS* mutant. These structures guided the rational mutation approach. Displayed are critical residues which shape the active site. Since there are just structures of *M. mazei* available these were used within a homology model for *M. barkeri*. Residue number in brackets reflect the number of the *M. barkeri* PylRS and the others indicate *M. mazei* PylRS numbering. **A**) Wild-type *MmPylRS* (PDB ID 2Q7H)¹⁷⁶ with bound Pyl-AMP. **B**) Mutant *MmOmeRS* (PDB ID 3QTC)²⁴² with bound O-Methyl-tyrosine-AMP-PNP.

As shown in **Figure 28A**, all constructed mutants led to comparable Sac incorporation. This was quite surprising since the authors in the original SacRS report found only three variants for Sac (1) incorporation and just one variant in the presence of an additional C313W mutation.¹⁴⁴ This highlights the enormous importance of quality control when constructing libraries and the need for better control when analyzing newly found variants after screening. **Figure 28A** does not provide a clear picture whether a hydroxyl group is an advantage, since the C313W:W382A variant performs at comparable levels. In contrast, the small size clearly plays an important role, with Phe and Trp mutations being the two most inefficient variants in the tested set so far. Interestingly, for the Trp mutant a strong cAA background incorporation is detected. This is perfectly in line with literature reports. It is known that PylRS enzymes with C313W mutations and a small residues at position N311 incorporate Phe¹⁶⁶. The original finding of the SacRS was based on positive and negative selection rounds. The inclusion of the negative step in the selection against variants with high background incorporations clearly shows why variant W382 was overlooked. All data included, the W382S mutation is most likely important for the reconstitution of *MbSacRS*'s orthogonality. The four best variants from **Figure 28A** were tested in a concentration dependent manner to gain detailed information about the OTS performance *in vivo*. Indeed, the best construct (C313W:W382T) displayed similar activity having double the OTS efficiency, at lower Sac (1) concentrations (1 mM), over the original SacRS (**Figure 76**).

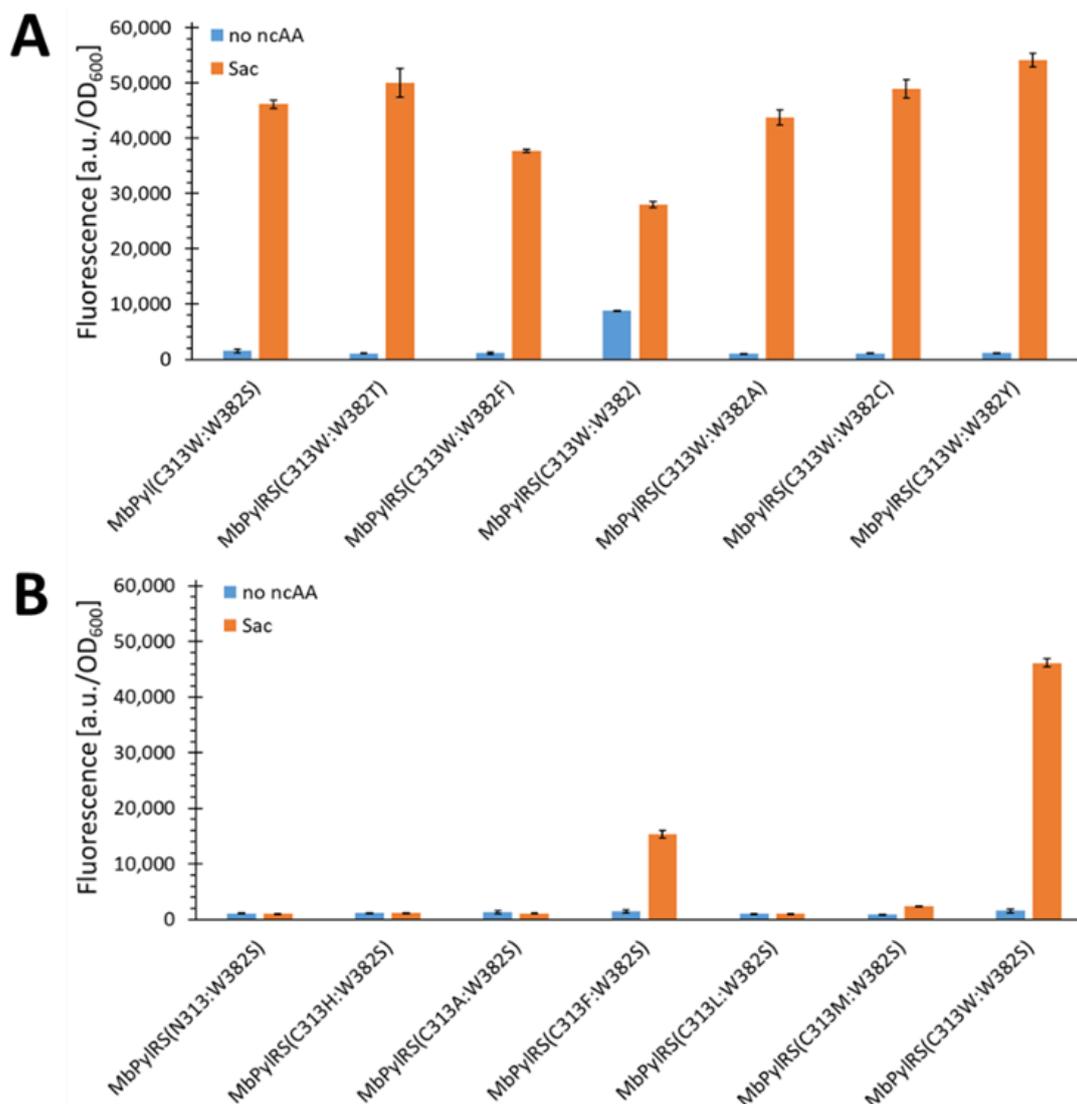


Figure 28: Comparison of Sac incorporation efficiency for *MbPylRS* constructs **A**) *MbPylRS*(C313W) and variants mutated at position W382 and **B**) *MbPylRS*(W382S) with variants mutated at position C313. Measured by the fluorescence of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2amber) reporter protein. Relative fluorescence is normalized to the highest value. 10 mM ncAA supplied.

To determine if this was just a specific property of the *MbPylRS* the mutations were also transferred to *MmPylRS*, the variant where the SacRS was first discovered. For this variant, the equivalent of the (C313W:W382T) mutations was observed as best performing as well, which indicates a robust result (**Figure 77**). Interestingly the second-best variant of the *Mb* and *MmSacRS* were different. While for the *M. barkeri* construct the W382Y was second best, for the *M. mazei* the corresponding W382S mutation was second. This highlights the fact that even though mutations are transplantable between the different PylRS systems of different species, the best mutation found in one species, does not have to be the best one in another. The C313W:W382T variant was also reported to incorporate *S*-propargyl-L-cysteine (**27**).¹⁴⁶ Therefore, the four best Sac (**1**) incorporating constructs were tested (also with substrate **27**). It was found that the (C313W:W382T/Y) constructs performed comparable (**Figure 78**). Investigating the effect of the C313W mutation, **Figure 28B** clearly indicates that the size of

the residue at this position is the most prominent factor (though not the only one). No variant with a small amino acid at this position could incorporate Sac. In contrast the bulkier Phe enables Sac incorporation, though at a lower level compared to all C313W mutants (seen in the ratio of incorporation and background signals). Even with Met at this position, there is low level incorporation detectable. The non-active C313H variant, comparable in size to C313F, indicates that a non-polar residue is necessary to be present in this microenvironment. In sum, the gathered data suggests that a reduced size of the binding pocket with the C313W mutation is essential for Sac incorporation. The W382 mutation has just to be a residue which is smaller than Trp to restore orthogonality. However, there are some Sac incorporation efficiency differences with different residues at this position. Especially variants with a polar OH- or SH-group at W382 position perform best.

2.3.5. Rationalizing Sac Incorporation Data and Creating Aliphatic Substrate incorporating MbPyIRS Variants

It was previously proposed that the C313W mutation is critical for activation of smaller substrates.^{144,242} This was also fully recapitulated in this work. Having established in this way that the C313W mutation is critical to incorporate Sac (**1**) and therefore probably also smaller aliphatic Sac (**1**) analogs, all these mutants were tested for incorporation of substrates **2**, **3**, **5**, **7** and **10** (Figure 29).

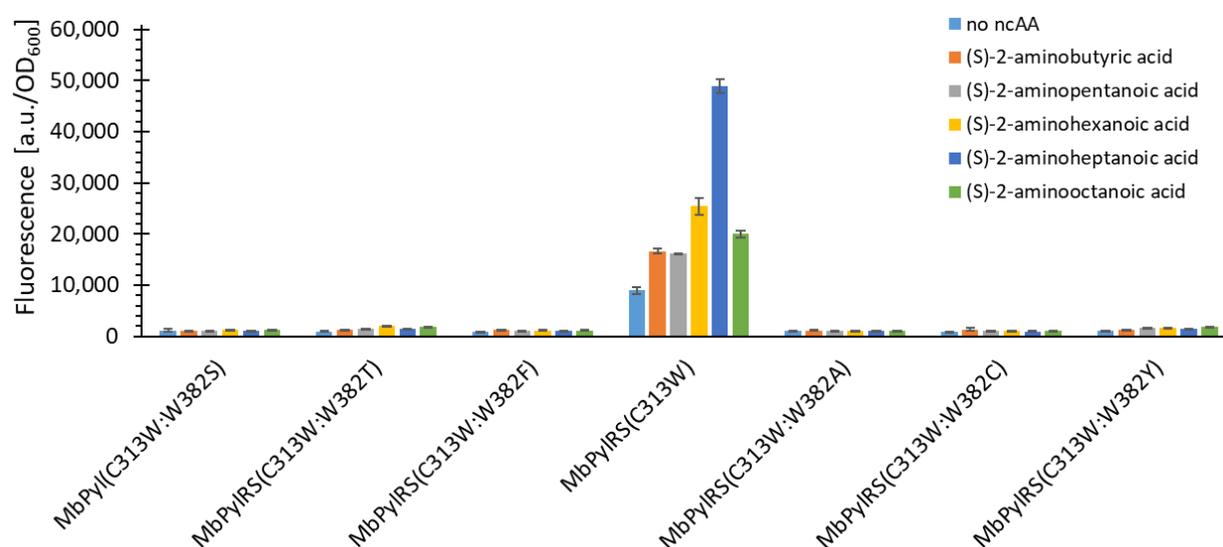


Figure 29: Comparison of aliphatic ncAA incorporation efficiency for MbPyIRS(C313W) constructs mutated at position W382. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Relative fluorescence is normalized to the highest value. 10 mM ncAAs supplied.

This fluorescence data revealed that all variants, besides the C313W mutant, did not incorporate the tested substrates. Unfortunately, the one construct which incorporated the aliphatic ncAAs exhibited a considerable level of background incorporation activity as well. As stated before, it is well known that bulky C313 mutations of MbPyIRS lead to background Phe incorporation^{166,176}. The hypothesis was that the key to recover the lost orthogonality (which

accompanied the C313W mutation) would be to mutate the *MbPylRS* at position N311 which is also known as one of the two gatekeeper residues for Pyl activation. The main working hypothesis was to restore the orthogonality by increasing the size of the amino acid at this position to interfere with the Phe substrate accommodation and simultaneously create a catalytic pocket suitable for the short-chain aliphatic ncAAs. As shown in **Figure 30A** by increasing the AA size at position N311, the orthogonality also increases until its full restoration with the N311L:C313W variant. An incorporation pattern of the tested aliphatic substrates was revealed, most likely corresponding to the size of the newly created catalytic pocket. The signal intensity for the N311L:C313W variant suggest that substrate **2** has the optimal size for this pocket. This substrate is the aliphatic equivalent to Sac (**1**). The data supported the hypothesis that it was possible to restore orthogonality with altering the residue at position N311 to bulkier AAs. This encouraged to test all amino acids bigger than valine at this position. In **Figure 30** all found active variants are displayed. Two of them (N311M/Q:C313W) can incorporate the substrates even better than the N311L:C313W construct. Interestingly, these two additional variants possess a different ncAA incorporation profile. The N311M construct favors the substrate with carbon chain length C6, while the N311Q favors C7. The incorporation profile of the N311Q variant is similar to the N311L.

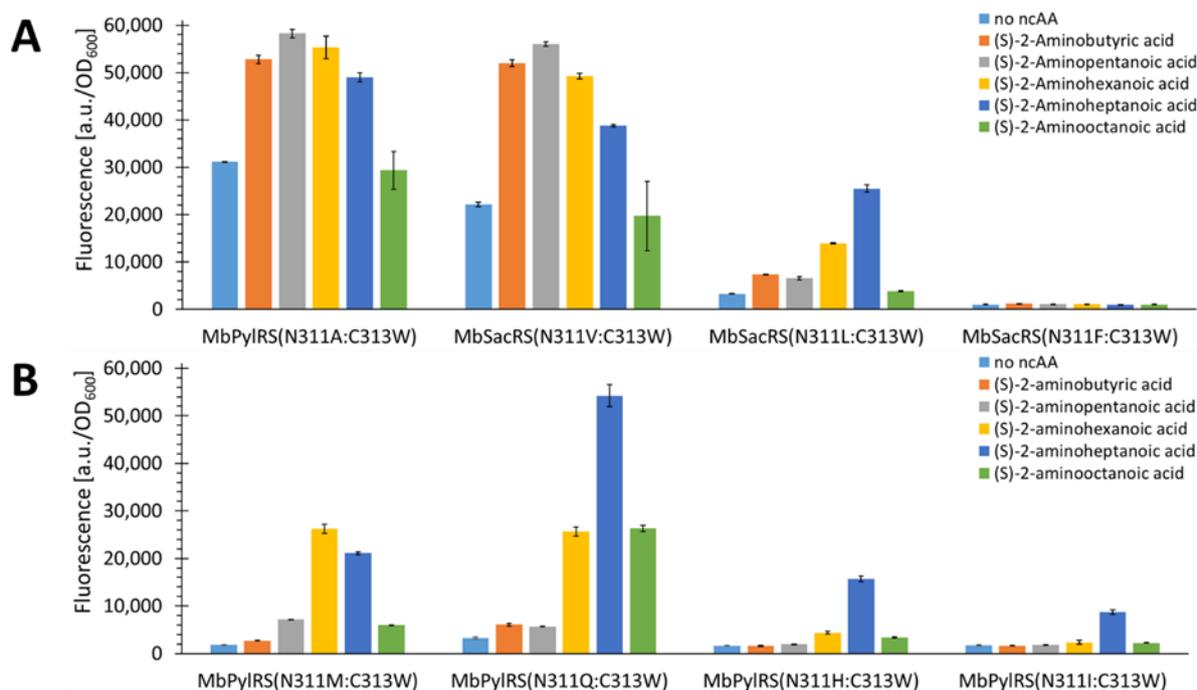


Figure 30: Comparison of aliphatic ncAA incorporation efficiency for *MbPylRS*(C313W) constructs mutated at position N311. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Relative fluorescence is normalized to the highest value. 10 mM ncAAs supplied.

The key residues important for maintaining the orthogonality of the PylRS system found here will facilitate future enzyme engineering efforts for the incorporation of structural analogs. The information will enable the creation of more intelligent PylRS-libraries based on reduced library size and extensively accumulated empirical know-how. In particular, the library size reduction

will enhance the probability of finding the desired enzymes and facilitate the establishment of ready-made enzyme batteries to test a vast variety of natural and synthetic ncAAs for translational activity. In addition, for close structural analogues, the number and type of selection cycles (e.g., negative selection) most likely can be reduced, making the whole selection procedure simpler and more feasible.

2.3.6. Semi-Rational Engineering of PyIRS Constructs for Small Aliphatic Substrate Incorporation

The previously described effort resulted in five constructs with efficient incorporation of aliphatic ncAAs. With them in hand, the two best performing ones (N311M:C313W and N311Q:C313W) were chosen for further engineering. The goal was to improve the incorporation efficiency for the aliphatic substrates. Therefore, residues potentially in close proximity to substrates were targeted. After inspection of the crystal structure of an *Mb*PyIRS analog, which can incorporate *O*-Methyl-L-tyrosine (*Mm*PyIRS, PDB: 2Q7H cf. ¹⁷⁶), the plan was to randomize residues A267, V366, Y349 and W382 in the active site, via site-saturation mutagenesis (SSM) with NNK primers (N=A/T/C/G; K=G/T). Since screening systems have their drawbacks as mentioned above, which can lead to missing active variants, a selection strategy based only on positive selection was chosen. With no available fluorescence-activated cell sorting (FACS) a strategy based on single site randomization and 96-well plate fluorescence screen was developed. Calculations with the Toplib tool estimated the probability of finding the best performing variant to be more than 95.3% and finding at least one of the two best performing variants with a probability of 99.8% (using a library yield of 85%, which is the lower limit of primer purity (for HPLC purified primers) and therefore also the lower yield limit of functioning created DNA constructs).²⁷⁶

The randomization was started with position V366 because it is directly located opposite to the N311 residue, which was increased in size. Therefore it was hypothesized that altering V366 could have the biggest tuning effect in regarding to the new ncAAs recognition. The randomization of this position revealed two new variants for each parent construct, with mutations V366A/K (**Figure 31**). **Figure 31A** shows the OTS performance of these constructs in comparison to the N311M:C313W starting construct. The V366K variant reveals no strong difference in the incorporation profile regarding the starting enzyme. Interestingly, the V366A mutation leads to a specificity shift towards long-chained aliphatic ncAAs in the incorporation profile. This seems plausible based on the *Mm*PyIRS crystal structure (PDB: 3QTC, cf.²⁴²), since this mutation potentially increases the space of the binding pocket which should facilitate the incorporation of longer substrates. **Figure 31B** shows that the two found variants based on the N311Q:C313W construct performed similarly to the parent enzyme. Similarly to the N311M:C313W:V366A construct, there is also a slight shift in the incorporation profile towards

longer ncAAs observable for the V366A mutant. Since this variant already favors ncAAs with C7 chain length over C6, the shift is smaller than for the N311M:C313W construct. Unfortunately, screening of the A267, Y349 and W382 randomizations revealed no better performing variants. However, one variant with a markedly different incorporation profile was found (N311M:C313W:W382H). This mutant preferably incorporated the longer chained substrate **3** (Figure 79).

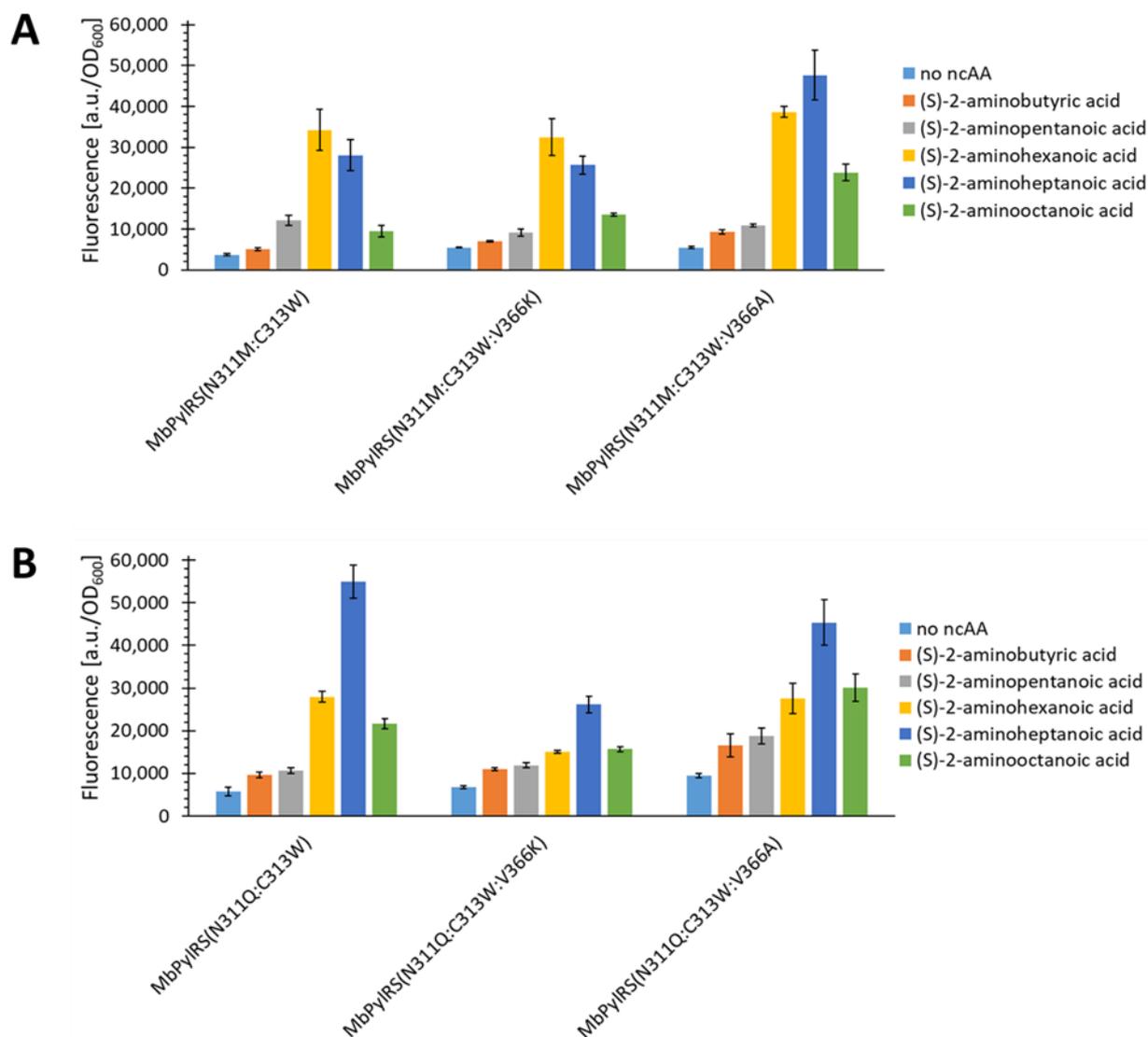


Figure 31: Comparison of aliphatic ncAA incorporation efficiency for **A**) *MbPyIRS*(N311M:C313W) and **B**) *MbPyIRS*(N311Q:C313W) constructs both mutated at position V366. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Relative fluorescence is normalized to the highest value. 10 mM ncAAs supplied.

2.3.7. Evaluating the Incorporation of Biochemically Useful Aliphatic ncAA Analogs

All created *MbPyIRS* constructs were screened for the incorporation of the useful aliphatic ncAA/AA analogs listed in Figure 26 (besides **2**, **3**, **5**, **7** and **10**). Substrates **1-9**, **20-26** and **28** were strongly incorporated (Figure 31, Figure 80, Figure 81 and Figure 82) and could be expressed in a standard *E. coli* BL21(DE) protein production strain to higher levels (up to

21 mg/L protein per culture, **Table 5**). Interestingly, several constructs also incorporated Sac (1) and substrate 27, with the best Sac (1) incorporating construct being N311M:C313W:V366A (**Figure 83****Figure 84**). This clearly illustrates that there can be multiple substrate-recognizing enzyme topologies based on the same scaffold. Since this variant could also incorporate 26 (**Figure 85**), another round of randomization was performed to find out if its performance could be enhanced or other interesting Cys-based amino acid derivatives could be incorporated. This led to the identification of N311M:C313W:V366A:W382N/T/Y mutants that were able to successfully incorporate substrate 28.

Substrates 11-19 which led to very low incorporation signals were additionally screened with release factor 1 (RF1) knock-out strains JX33, B-95.ΔA and C321.ΔA.exp (**Figure 86**-**Figure 89**)^{152,154,277}. Strains which lack RF1 are believed to produce higher amounts of full-length protein via amber suppression. All these RF1 knock-out strains were engineered in previous works of our group to chromosomally possess the lambda DE3 lysogen encoding the T7 polymerase compatible with the reporter protein setup. Even though the background incorporation increased for all RF1 knockout strains (this was observed before cf.¹⁴⁴), some setups indeed resulted in increased ratios of ncAA/AA incorporation, relative to background incorporation levels. The best performing strain and *MbPyIRS* combinations were selected for larger scale protein production (**Table 5**).

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

Larger scale protein production was performed to confirm the results from the small scale 96-well plate fluorescence assays. The protein yields are in good agreement with the trends observed in the fluorescence assays. Intact protein mass spectra of the reporter proteins via electrospray ionization mass spectrometry (ESI-MS) were acquired. To facilitate MS data evaluation for AAs with a low incorporation efficiency, a reporter protein with a C-terminal His₆-tag was used. For each substrate **Table 5** contains the optimal reporter protein production setup, the ESI-MS data, and the protein yields. The corresponding deconvoluted ESI-MS-spectra are found in **chapter 6.1.5 (Figure 144****Figure 170**). The incorporation of all substrates but five (**14**, **17**, **18**, **19** and **24**) could be confirmed with MS analytics. For the setup of substrate **14** the molecular weight of incorporated AA is 146.3 g·mol⁻¹ which corresponds to glutamine (146.2 g·mol⁻¹). It is known that near-cognate suppressor tRNAs, like tRNA^{Gln}_{CUG}, read amber codons to some extent²⁷⁸. This means Gln is incorporated at amber sites when the OTS is not working and then commonly observed in MS analytics. The high reporter protein yield for a non-functioning OTS is most likely the result of the general higher background suppression observed in RF1 knock-out strains (**Figure 86**-**Figure 89**). Gln incorporation is also observed for the setup of substrate **16**, **17**. For substrate setups **18**, **19** and **24** the molecular weight of the incorporated AA is 165.3/164.3 g·mol⁻¹. This corresponds to phenylalanine (165.2 g·mol⁻¹),

indicating that this PylRS leads to Phe incorporation when no/inefficient substrate is present. The observed high protein yield for the setup of substrates **18** and **19** is probably caused by a combination of the Phe incorporation activity and the use of a RF1 knock-out strain, although further analytics are required here. The same is true for substrate **24**, though on a lower level since no RF1 knock-out strain was used. Nonetheless, the fluorescence data with different *MbPylRS* constructs clearly indicate that incorporation is possible (**Figure 87**).

Table 5: Optimal reporter protein production setup, calculated and observed molecular weights of the reporter proteins His₆-SUMO-sfGFP(R2AA)-strep (a) / SUMO-sfGFP(R2AA)-His₆ (b) and protein production yields per liter of culture. The masses were determined by ESI-MS of intact proteins.

AA	<i>E. coli</i> strains ¹	<i>MbPylRS</i> construct	reporter construct	calculated mass [Da]	found mass [Da]	Δ mass [Da]	protein yield [mg·L ⁻¹] ²
1	BL21	N311M:C313W:V366A	a	40194.9	40196	1.1	10.8
2	BL21	N311Q:C313W	a	40178.8	40180	1.2	5.1
3	BL21	N311M:C313W:V366A	a	40192.9	40194	1.1	1.6
4	BL21	N311M:C313W:V366A	a	40176.8	40179	2.2	1.7
5	BL21	N311M:C313W	a	40164.8	40166	1.2	1.9
6	BL21	N311M:C313W:V366K	a	40162.8	40164	1.2	1.4
7	BL21	N311M:C313W	a	40150.8	40153	2.2	1.2
8	BL21	N311M:C313W	a	40148.8	40150	1.2	0.7
9	BL21	N311M:C313W	a	40162.8	40164	1.2	1.4
10	BL21	N311M:C313W:V366A	a	40136.8	40195	58.2	0.8
11	C321.ΔA.exp	N311M:C313W	b	38990.9	38992	1.1	5.1
12	C321.ΔA.exp	N311M:C313W	b	38976.9	38979	2.1	4.9
13	C321.ΔA.exp	N311M:C313W	b	38962.8	38965	2.2	11.3
14	JX33	N311M:C313W	b	38979.8	38996	16.2	4.3
15	C321.ΔA.exp	N311M:C313W	b	38993.8	38994	0.2	14.2
16	C321.ΔA.exp	N311M:C313W	b ⁴	39007.9	39007	0.9	5.3
17	C321.ΔA.exp	N311M:C313W	b	39021.9	38997	24.9	6.4
18	C321.ΔA.exp	N311Q:C313W:V366K	b	38963.8	39015	51.2	19
19	C321.ΔA.exp	N311Q:C313W:V366K	b	38977.8	39014	36.2	19.9
20	BL21	N311M:C313W:V366K	a	40190.8	40194	3.2	4.8
21	BL21	N311M:C313W	b	38978.9	38982	3.1	2.6
22	BL21	N311M:C313W	b	38998.9	38998	0.9	3.6
23	BL21	N311Q:C313W	b	39014.9	39012	2.9	4.6
24	BL21	N311Q:C313W:V366K	b	39030.9	39015	15.9	10.7
25	BL21	N311Q:C313W	b	39013	39014	1	9.8
26	BL21	N311M:C313W:V366A	a	40210.9	40211	0.1	21
27	BL21	N311M:C313W:V366A	a ³	-	-	-	-
28	BL21	N311M:C313W:V366A:W382N	b	39061	39063	2	1.1

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

In summary, several *Mb*PyIRS variants have been created that can incorporate 23 ncAAs and one cAA. To the best of my knowledge, 17 of these ncAAs (besides **1**, **3**, **4**, **17**, **27** and **28**) were not ribosomally incorporated by amber suppression and 20 of them were not incorporated with the PyIRS system before ^{144,146,260,279}.

2.4. Creation of Novel PyIRS OTSs Based on Psychro- and Thermophilic Homologs

As discussed before (**chapter 1.4.2**), psychrophilic enzymes can have remarkable properties from a protein engineering perspective (catalytic efficiency, increased promiscuity). Combined with the amenability of PyIRS to recognize novel substrates, this could lead to an extremely powerful tool for genetic code expansion. Therefore, *in silico* analyses were performed to identify potential candidates, followed by *in vivo* experiments.

2.4.1. *In Silico* Analysis of Archaeal PyIRS Variants

When *Bacteria* or *Archaea* are able to encode Pyl, their genes involved in the Pyl metabolism (PylBCD) and incorporation (PylST) are generally well conserved.¹⁷⁹ To reduce the scope of *in silico* analysis, only the PyIRS (pyIS) and tRNA^{Pyl} (pyIT) genes of archaea, specifically *Methanosarcinales* were analyzed. Even though there have been recent discoveries of PyIRS systems in a range of new archaeal organisms other than *Methanosarcinales* the scope was limited to the latter, since these systems have been proven functional in *E. coli*.¹⁵⁵ Also, bacterial PyIRS were omitted because attempts to create functional OTS based on them resulted in systems with low efficiency.²⁸⁰ This low efficiency is also supported by *in vitro* data.¹⁷⁵ Moreover, it is not clear how to construct bacterial-based PyIRS OTS that function efficiently. One hurdle certainly is the unknown *in vivo* function/interaction of one of the two genes (PylSn) that make up the PyIRS OTS in bacteria.^{280–282} In contrast, pylSc is known to be responsible for expression of the catalytic domain.

2.4.1.1. Phylogenetic PyIRS Analysis

Phylogenetic analysis was performed for all *Methanosarcinales* and *Methanomassiliicoccales* PyIRS sequences (**Figure 32**). These sequences were extracted from genome data from NCBI and, for *M. alaskense* from the JGI database. Also included was the sequence of *M. thermophila* which was determined with degenerative primer amplification from genomic DNA before the entire genome was available.¹⁷⁵ Unexpectedly, the sequence of this PyIRS did not match the PyIRS sequence that was extracted from the later published genome of *M. thermophila* (strain TM-1). Since there are two other sequences that were also extracted from different strains (CHTI-55 and MT-1) which match the TM-1 sequence, this is most likely the correct one and is referred to as *Mt*PyIRS(TM-1) in this work. The other *Mt*PyIRS appears to be more closely related to the only known psychrophilic *Methanosarcinae* PyIRS from *M. lacustris*. Further discussion on this issue and possible implications is provided below.

As expected, the four clusters of *Methanococcoides*, *Methanohalophilus*, *Methanolobus* and *Methanosarcina* are clearly recognizable. The *Methanococcoides* cluster can be divided into two subgroups. The other three cluster show relationships that are consistent with their

taxonomy. A fifth cluster is displayed consisting of the 7th order of methanogens (*Ca. M. alvus*, *Ca. M. intestinalis* and *M. luminyensis*)¹⁷⁹ and is closely related to *M. shengliensis* the first known highly thermophilic methylotrophic methanogen.²⁸³ This was unexpected because the first three belong to the order of *Methanomassiliicoccales* and the last was categorized as *Methanosarcinales* based on 16S rRNA phylogenetic analysis.²⁸⁴ Further investigations are needed to clarify this relationship which is out of scope of for this study.

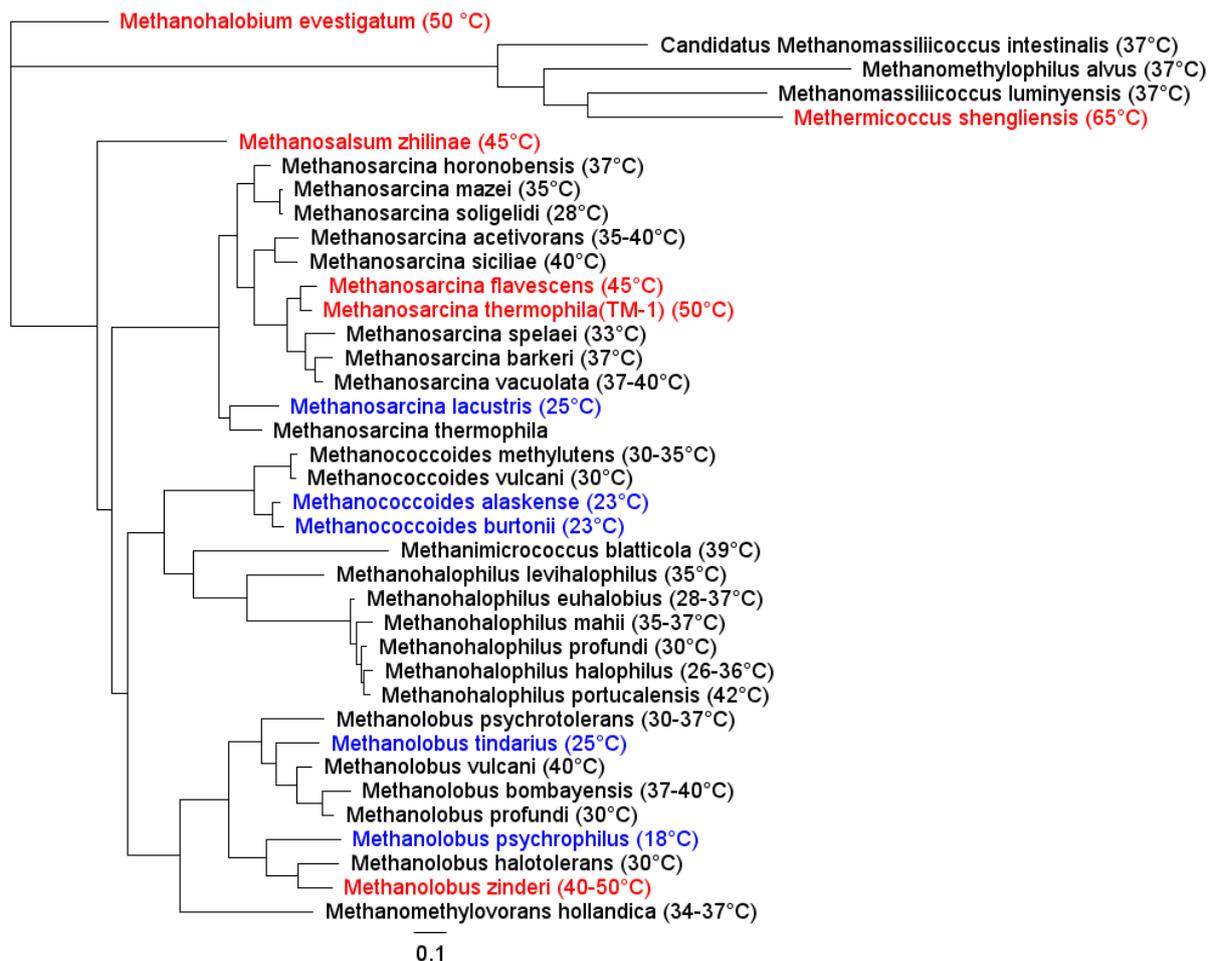


Figure 32: Approximate-maximum-likelihood phylogenetic tree based on all *Methanosarcinales* and *Methanomassiliicoccales* PylRS sequences available in the NCBI and *M. alaskense* from the JGI databases. The FastTree algorithm (version 2.1.11) was used for tree creation.²⁸⁵ The number(s) in parentheses represents the optimal growth temperature (OGT). Coloring indicates the origin of psychrophilic (blue, equal to and below 25°C) and thermophilic (red, equal to and above 45°C) PylRSs. The scale bar indicates 10% sequence divergence. Genome specifications of the organisms can be found in the Appendix.

2.4.1.2. Analysis of *Methanosarcina* PylRS and variable region

The phylogenetic tree (Figure 32) shows the relationships of the selected PylRSs. As expected, the PylRSs from the genus of *Methanosarcinae* form clusters based on their taxonomy. A notable feature of *Methanosarcina* PylRS, in contrast to all other PylRS of the same family, is that the lengths of the variable region are much more diverse. The shortest linker in the genus *Methanosarcina* is 30 AAs and the longest is 126 AAs long (Table 6, see chapter 6.1.8 for details on variable region determination). This span is unmatched compared

to all other genera in the same family (four genera with equal or more than 4 sequences were compared to obtain more robust results). For comparison, the range in the other genera is: *Methanococcoides* (27-29 AAs), *Methanohalophilus* (41-45 AAs), *Methanobolus* (70-80 AAs), so in general the differences in linker lengths range between 2 and 10 (**Figure 209-Figure 216**). This phenomenon prompted further investigation. It was hypothesized that the thermal PyIRS origin and linker length are correlated to some degree. To scrutinize this hypothesis, a similarity matrix was constructed for all C-terminal parts of the PyIRS of all *Methanosarcina* (**Figure 33**).

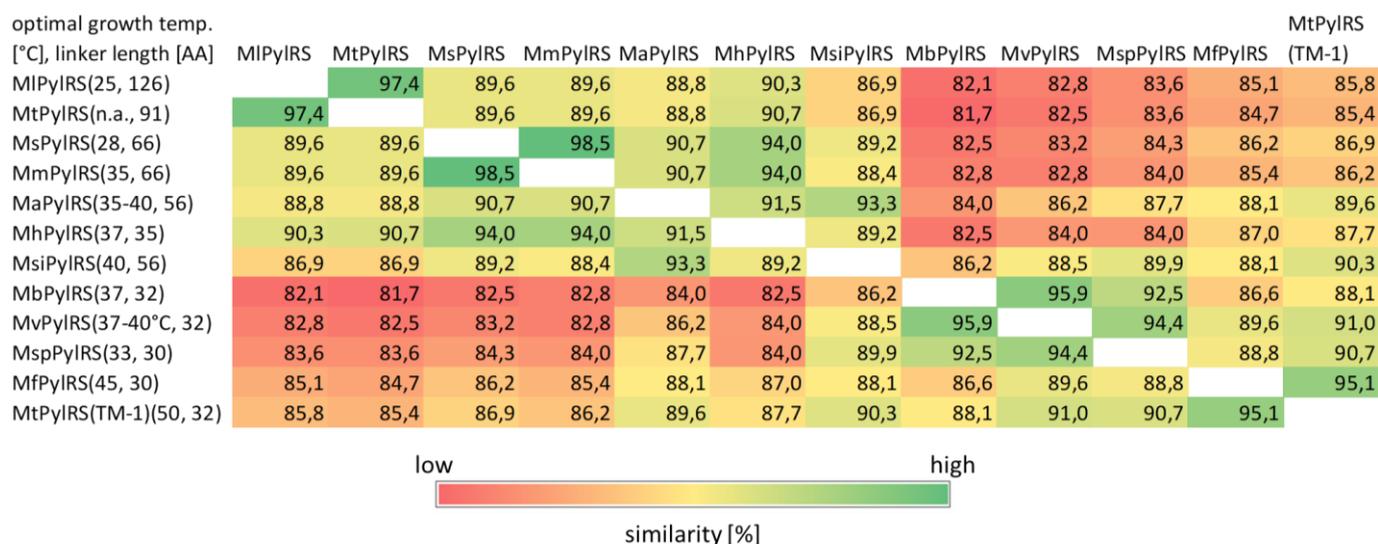


Figure 33: Similarity matrix of the C-terminus of known *Methanosarcina* PyIRSs. [AA] = number of amino acids as linker, n.a. = not available.

Only the C-termini were chosen because they were thought to provide more information about the relatedness in terms of catalytic activity. It is reasonable that evolutionary adaptations will most likely result in changes in the catalytic part of the enzyme, as is the case with the adaption of psychrophilic enzymes. In addition, a full-length enzyme alignment was performed without linker (to exclude alignment gap penalties due to linker length) resulting in the same trends (**Figure 208**). When sequences of different lengths are in the set, satisfactory alignment is difficult to achieve, resulting in low clustering accuracy.²⁸⁶ It is plausible that the linker plays a role in tuning the catalytic activity, therefore it is not the sequence itself (although Pro content certainly plays a role, e.g. for rigidity) but the length of the linker carries crucial information about enzyme properties. Not surprisingly, the similarity matrix shows the same correlations as the phylogenetic tree (**Figure 32**). Since the matrix also contains the information of linker lengths of the variants, connections can be drawn including this information. Focusing on the columns of *MIPyIRS* and *MtPyIRS*(TM-1), there appears to be an inverse correlation between optimal growth temperature and linker length. Since a longer linker would certainly lead to more flexibility, this is consistent with the rationale of psychrophilic enzyme adaption. While the amino acid composition of these PyIRS (with the exception of the variable region) is very

similar (**Figure 209**), the linker composition varies greatly. In addition to the aforementioned length, the major differences in the linkers concern the proportion of polar/unpolar amino acids and the proline content (**Figure 210**). There are certain trends which can be correlated with OGT. The variants with the highest OGT (thermophilic) have the highest content of polar and basic amino acids and the shortest linkers. In contrast, the single psychrophilic variant has by far the longest linker and one of the lowest levels of polar and basic amino acids. Interestingly, the Pro content follows the opposite trend with high percentage for long linkers and a very low percentage for short linkers. Psychrophilic homologs are known to exhibit increased enzyme flexibility due to a longer linker length between domains. For example, it has been observed that a psychrophilic cellulase has an unusually long linker (of 100 AA), five times longer than that of the mesophilic homolog.^{287,288} In the case of the PyIRS, the increase in Pro content together with the length increase appears to offset this increase in flexibility to a certain degree. Given all this information it is very plausible that the PyIRS, referred to as *MtPyIRS* in this work, does not have the AA composition and linker length (the C-terminus of the *MtPyIRS* has over 97% sequence similarity with *MIPyIRS*) to be classified as a thermophilic PyIRS.

It is also noteworthy that after removing the linker, all PyIRS (besides *MaPyIRS*) have almost the same sequence length and differ only in one AA (see **Table 6**). The divergence in the PyIRS length of *MaPyIRS* may indicate that the organism is genetically not as closely related to the genus as the phylogenetic tree would suggest.

2.4.1.3. Analysis of *Methanosarcinales* PyIRS and variable region

The similarity of the AA composition of the *Methanosarcina* PyIRS led to the question of how the enzymes of the other genera (*Methanococcoides*, *Methanohalophilus*, *Methanlobus*) are composed. Therefore, the averages of the PyIRS (**Figure 34**) and the variable region (**Figure 35**) were created based on the individual distributions (**Figure 209-Figure 216**). For the PyIRS enzymes, the averages are a good approximation of the individual PyIRS AA composition because the differences in amino acid content are small. In particular, the ratios of polar, nonpolar, basic and acidic AAs are similar. Only for the halophilic PyIRS there is a slight increase in polar AA which would be expected. For the variable region, it is more problematic to use the average as an approximation because there are large differences in AA usage. Thus, it is used only in a very descriptive way to emphasize a particular point and not to infer functional relationships. The most obvious differences in the linkers are the use of Lys, Ser and Ala. While the halophilic linkers contain a very high number of Lys residues, *Methanosarcina* linker contain very few. In comparison, *Methanosarcina* linker contain a high proportion of Ala and Ser although there is a high degree of heterogeneity between species (as mentioned before). It is difficult to explain these differences, but the most important result of the linker comparison is the high degree of heterogeneity. This suggests that the variable

region plays a crucial role in tuning enzymatic activity to the specific environments of these organisms and perhaps even tune the interaction with the tRNA^{Pyl}. There is one study that could clearly show that a change in linker length can have impact on the PylRS OTS of the wild-type and the engineered variants.²⁸⁹

PylRS averages	Methanosarcina	Methanococcus	Methanohalophilus	Methanobrevibacter	standard deviation	range	max	min
A	4.5	2.6	3.2	3.0	0.7	2.0	4.5	2.6
C	2.2	2.3	2.2	2.5	0.1	0.3	2.5	2.2
D	6.2	6.4	6.7	5.3	0.5	1.4	6.7	5.3
E	7.9	9.1	7.0	9.1	0.9	2.1	9.1	7.0
F	3.7	4.2	3.8	5.0	0.5	1.3	5.0	3.7
G	5.5	5.9	5.9	6.3	0.3	0.7	6.3	5.5
H	2.3	2.4	2.3	1.9	0.2	0.5	2.4	1.9
I	6.5	7.3	7.2	6.6	0.3	0.8	7.3	6.5
K	8.7	7.7	7.7	8.3	0.4	1.0	8.7	7.7
L	10.6	9.5	9.9	9.6	0.4	1.0	10.6	9.5
M	3.0	3.0	3.8	4.3	0.5	1.3	4.3	3.0
N	4.2	4.0	5.3	4.9	0.5	1.3	5.3	4.0
P	4.7	4.9	3.8	3.9	0.5	1.2	4.9	3.8
Q	1.2	2.3	2.2	1.8	0.5	1.2	2.3	1.2
R	7.9	9.1	8.6	8.1	0.5	1.2	9.1	7.9
S	7.4	6.2	6.1	7.9	0.8	1.9	7.9	6.1
T	4.2	3.8	5.4	3.4	0.8	2.0	5.4	3.4
V	5.5	5.4	4.7	5.0	0.3	0.8	5.5	4.7
W	0.8	0.8	0.8	0.8	0.0	0.0	0.8	0.8
Y	3.1	3.3	3.6	2.6	0.4	1.0	3.6	2.6
sum	100.0	100.0	100.1	100.0				
unpolar	39.3	37.6	37.0	38.1	0.8	2.3	39.3	37.0
polar	27.8	27.8	30.8	29.3	1.2	3.0	30.8	27.8
basic	18.8	19.1	18.7	18.2	0.3	0.9	19.1	18.2
acidic	14.1	15.5	13.7	14.4	0.6	1.7	15.5	13.7

Figure 34: Average amino acid distribution in percent and descriptive indicators of *Methanosarcinae* PylRS variants. The data describes amino acid distribution for the PylRS sequence excluding the variable region (linker). Coloring indicates green for high values and red for low values. Color scales are different for amino acids and range/standard deviation.

linker averages	Methanosa rcina	Methanoco ccoides	Methanohal ophilus	Methanolob us	standard deviation	range	max	min
A	19.9	9.7	12.7	8.1	4.5	11.8	19.9	8.1
C	0.2	0.0	0.0	0.0	0.1	0.2	0.2	0.0
D	0.0	5.2	0.0	2.0	2.1	5.2	5.2	0.0
E	1.9	6.0	8.3	5.4	2.3	6.4	8.3	1.9
F	0.4	0.0	0.0	1.2	0.5	1.2	1.2	0.0
G	2.3	1.8	0.4	1.7	0.7	2.0	2.3	0.4
H	0.8	0.0	0.0	3.4	1.4	3.4	3.4	0.0
I	1.9	5.0	1.2	4.1	1.6	3.8	5.0	1.2
K	3.5	10.4	25.4	12.2	7.9	21.9	25.4	3.5
L	1.6	0.0	0.0	1.2	0.7	1.6	1.6	0.0
M	1.2	0.0	2.0	1.0	0.7	2.0	2.0	0.0
N	4.0	7.0	3.9	9.0	2.1	5.1	9.0	3.9
P	14.5	17.3	13.8	12.9	1.7	4.4	17.3	12.9
Q	4.5	0.0	0.8	6.9	2.8	6.9	6.9	0.0
R	1.0	0.0	0.0	2.8	1.1	2.8	2.8	0.0
S	25.3	15.2	11.0	13.8	5.4	14.3	25.3	11.0
T	9.8	13.1	6.3	9.5	2.4	6.8	13.1	6.3
V	7.3	9.3	14.2	4.5	3.5	9.7	14.2	4.5
W	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Y	0.1	0.0	0.0	0.7	0.3	0.7	0.7	0.0
sum	100.1	100.0	100.0	100.1				
unpolar	46.7	41.3	43.8	32.9	5.2	13.8	46.7	32.9
polar	46.2	37.1	22.4	41.5	8.9	23.8	46.2	22.4
basic	5.3	10.4	25.4	18.4	7.7	20.1	25.4	5.3
acidic	1.9	11.3	8.3	7.4	3.4	9.4	11.3	1.9

Figure 35: Average amino acid distribution in percent and descriptive indicators of *Methanosarcinae* PyIRS variable regions (linkers). Coloring shows green for high values and red for low values. Color scales are different for amino acids and range/standard deviation.

2.4.2. Choice and Elaboration of Used PyIRS Constructs

The goal was to find a more efficient PyIRS based on a psychrophilic PyIRS. As the phylogenetic tree (**Figure 32**) shows, there are five potential candidates. At least one psychrophilic PyIRS from each genus was used (if existent). When more than one variant was considered for inclusion in the set, the psychrophilic set was reduced based on the following criteria to handle the workload (**Table 6**). For the genus *Methanolobus*, only the variant of the species with the lowest OGT was used. For the *Methanococcoides* genus, as the project was planned, just the *M. burtonii* genome was available and therefore used (the *M. alaskense* genome was available at 02.02.2021). In hindsight, the decision which construct to include would have been the same, since these two psychrophilic PyIRS variants (from *M. burtonii* and *M. alaskense*) have sequence similarity greater than 95%. But a closer look at the tRNAs revealed clear differences in free energy (**Figure 50**), with the tRNA^{PyI} of *M. burtonii* being almost 6 kcal/mol higher. This suggests that *M. burtonii* could even be more psychrophilically adapted, which is why this variant would have been chosen either way. To test the hypothesis that the *Mt*PyIRS was indeed derived from a psychrophilic organism, this variant was also

included. Finally, to have a reference for comparing performance, *Mm*PyIRS, *SmbP-Mb*PyIRS and *Mt*PyIRS(TM-1) were included. *Mt*PyIRS(TM-1) was included to find out whether the properties commonly attributed to thermophilic enzymes (low efficiency at mesophilic cultivation temperatures) also apply to the PyIRS group.

Table 6: List of all known Methanosarcina PyIRS and the additional psychrophilic PyIRS used in this study

abbreviation	organism name	optimal growth temp. [°C]	enzyme length [AA]	linker length [AA]	length without linker [AA]	linker proline content [%]	reference
<i>Mbur</i> PyIRS ¹	<i>Methanococcoides burtonii</i>	23	416	29	386	16.7	Allen et al. ²⁹⁰
<i>Mp</i> PyIRS ¹	<i>Methanlobus psychrophilus</i>	18	443	70	394-	14.3	Chen et al. ²⁹¹
<i>Ml</i> PyIRS ¹	<i>Methanosarcina lacustris</i>	25	513	126	387	23.8	Simankova et al. ²⁹²
<i>Mt</i> PyIRS ¹	<i>Methanosarcina thermophila</i>	n.a. ²	478	91	387	16.5	Herring et al. ¹⁷⁵
<i>Ms</i> PyIRS	<i>Methanosarcina soligelidi</i>	28	454	66	388	9.1	Wagner et al. ²⁹³
<i>Mm</i> PyIRS ¹	<i>Methanosarcina mazei</i>	35	454	66	388	9.1	Kavran et al. ¹⁷⁶
<i>Ma</i> PyIRS	<i>Methanosarcina acetivorans</i>	35-40	404	56	348	17.9	Galagan et al. ²⁹⁴
<i>Mh</i> PyIRS	<i>Methanosarcina horonobensis</i>	37	423	35	388	14.3	Shimizu et al. ²⁹⁵
<i>Msi</i> PyIRS	<i>Methanosarcina siciliae</i>	40	443	56	387	16.1	Elberson et al. ²⁹⁶
<i>Mb</i> PyIRS ¹	<i>Methanosarcina barkeri</i>	37	419	32	387	15.6	Srinivasan et al. ⁶²
<i>Mv</i> PyIRS	<i>Methanosarcina vacuolata</i>	37-40	419	32	387	21.9	Zhilina et al. ²⁹⁷
<i>Msp</i> PyIRS	<i>Methanosarcina spelaei</i>	33	418	30	387	16.1	Ganzert et al. ²⁹⁸
<i>Mf</i> PyIRS	<i>Methanosarcina flavescens</i>	45	417	30	387	6.7	Kern et al. ²⁹⁹
<i>Mt</i> PyIRS (TM-1) ¹	<i>Methanosarcina thermophila</i> (TM-1)	50	419	32	387	6.3	Zinder et al. ³⁰⁰

¹used in this work, ²not available

2.4.3. Comparison of OTS Efficiencies of Psychro-, Meso-, and Thermophilic PyIRS

Six Pyl-analogs (BocK (**31**), AllocK (**32**), ProK (**33**), AzidoK (**34**), PhotoK (**42**) and BenzK (**43**)) were selected to determine the OTS efficiency of PyIRS (**Figure 36**). The catalytic activities of PyIRSs for these substrates range from good (BocK (**31**), AllocK (**32**)) to moderate (AzidoK (**34**), ProK (**33**), PhotoK (**42**)³⁰¹) to poor (BenzK (**43**); almost no *in vitro* PyIRS tRNA aminoacylation activity with 1 mM but little activity with 3.5 mM¹⁶³). The different substrates were selected to provide a detailed picture of possible differences in catalytic efficiencies and promiscuity associated with the thermal origins of PyIRSs.

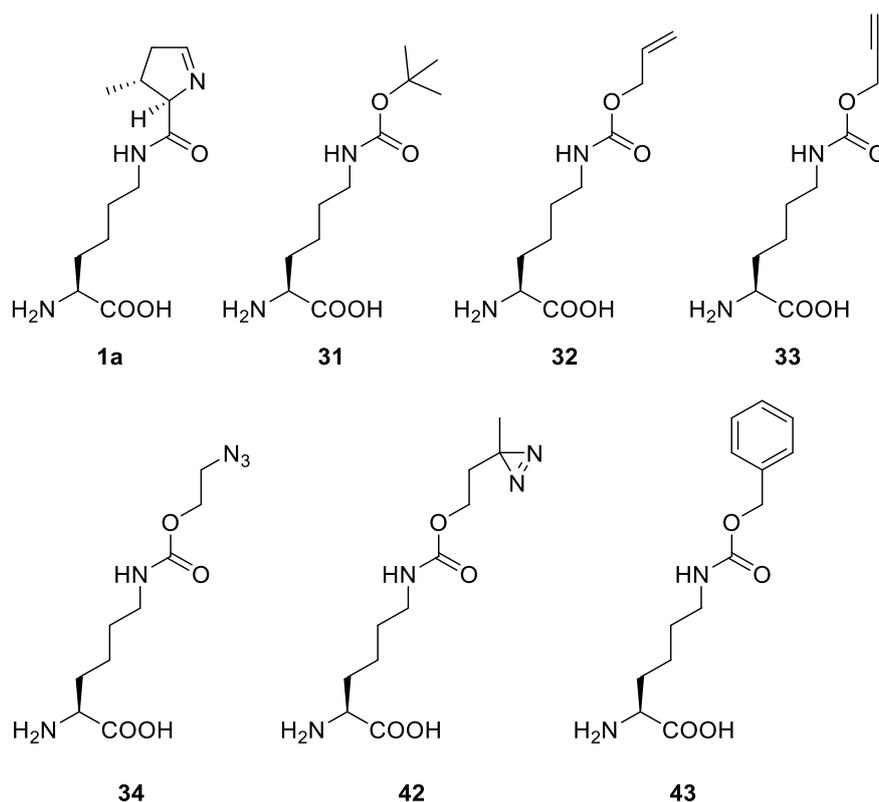


Figure 36: Pyrrolysine (**1a**) and derivatives used in this chapter. BocK = *N*^ε-*tert*-Butoxycarbonyl-L-lysine (**31**), AllocK = *N*^ε-Allyloxycarbonyl-L-lysine (**32**), ProK = *N*^ε-Propargyloxycarbonyl-L-lysine (**33**), AzidoK = *N*^ε-((2-Azidoethoxy)carbonyl)-L-lysine (**34**), PhotoK = 3'-Azibutyl-*N*^ε-carbamoyl-L-lysine (**42**), BenzK = *N*^ε-benzyloxycarbonyl-L-lysine (**43**).

Since it was shown in **chapter 2.2** that even the wild-type *Mb*PyIRS could achieve significant improvements with an SmbP-tag, the selected PyIRS set was also tested with and without this tag (**Figure 90-Figure 107**). Unfortunately, no other wild-type PyIRS showed improved OTS efficiency with the SmbP-tag. This could mean that at least all tested PyIRSs (besides *Mb*PyIRS) are sufficiently stable at 37°C or the solubility tag is dependent on the enzyme sequence, which would also be plausible. Further experiments on the SmbP-tag with PyIRS mutants can be found in **chapter 2.4.5.1**. Therefore, only the data from *Mb*PyIRS with an SmbP-tag are included in **Figure 37**.

Figure 37 shows six heatmaps depicting the incorporation performance of the selected PyIRSs variants at very low (0.05 mM, 0.1 mM), low (0.3 mM), medium (1 mM), medium-high (3 mM) and high (9 mM) ncAA concentrations. These maps show characteristically selected concentration data from the fluorescence assays (**Figure 90-Figure 107**). The fluorescence assays cover a wider concentration range (0.025, 0.05, 0.1, 0.3, 1, 3 and 9 mM). The ncAA gradient was chosen to provide more information about the efficiency of the OTSs. The background fluorescence signal at the right end of each box subtracted from the output signals serves as a reference. As a rough guide, robust ncAA incorporation is detectable via ESI-MS when the signal with ncAA added is twice that without ncAA. This threshold was defined as one of two cutoffs, i.e., all construct/ncAA combinations with a signal below this threshold were

colored red which means they are most likely non-functional. For convenience reasons the different cutoffs were defined based on the highest no ncAA signal from *MtPylRS*(TM-1). Since the no ncAA signals are not that different this just introduces a negligible error in coloring.

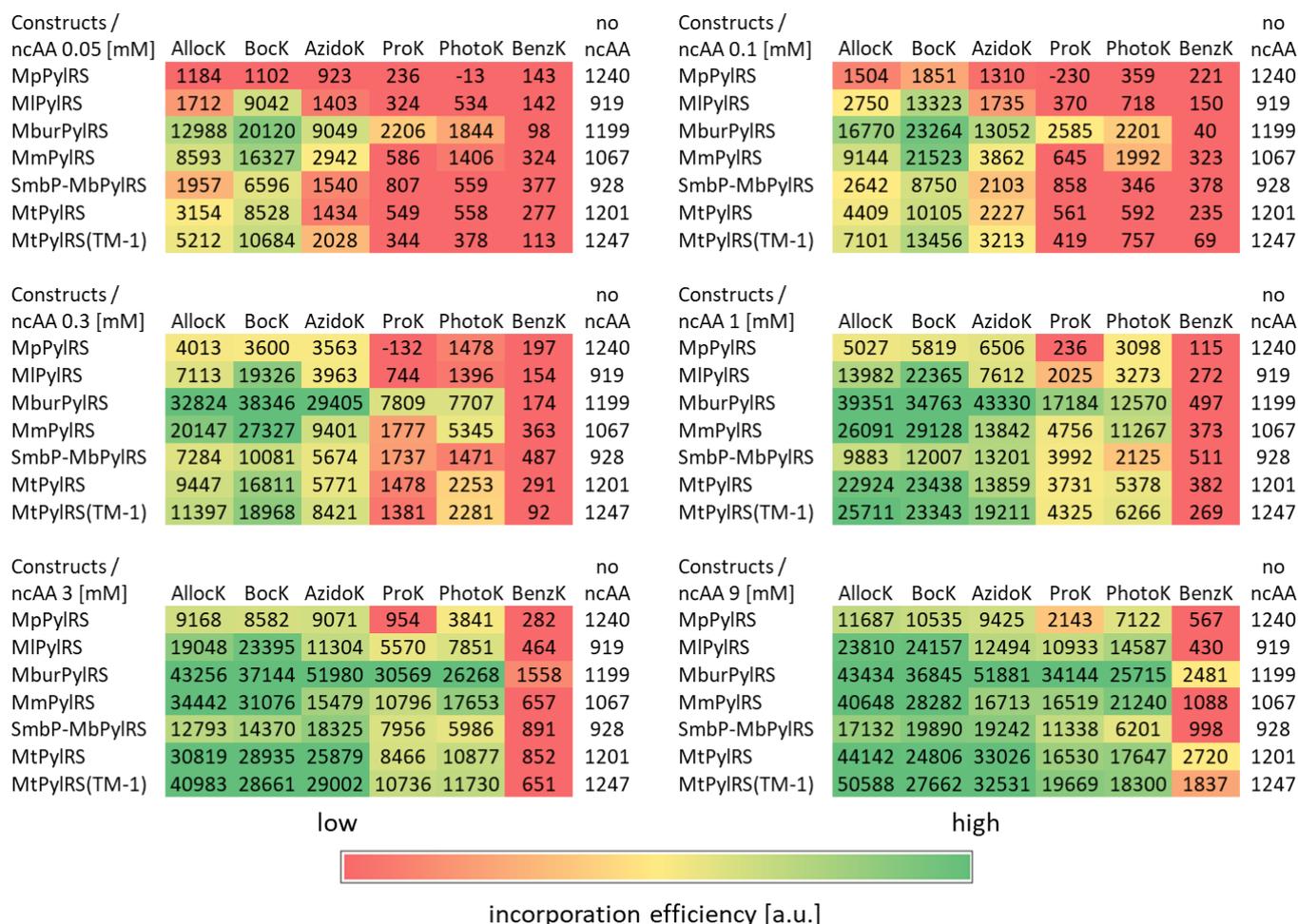


Figure 37: Heatmaps of the incorporation efficiencies of the seven selected PylRSs with different ncAA concentrations. The heatmaps are based on the fluorescence data (Figure 90-Figure 107) from sfGFP(R2amber) reporter protein assays. The background signal (no ncAA supplied) is subtracted from the initial signal values and also shown as reference. Cutoffs are defined on the bases of the no ncAA signal of *MtPylRS*(TM-1): below (no ncAA value) = red, 2*(no ncAA value) = yellow, 10*(no ncAA value) = green.

Figure 37 shows that all of the the created PylRS OTS are functional and recognize the substrates with varying efficiencies. Encouragingly, no significantly higher background incorporation is detectable for any of the psychrophilic PylRSs, indicating that they can discriminate against cAAs at the same level as the other PylRSs.

For the psychrophilic PylRS (*MIPylRS*, *MpPylRS*, *MburPylRS* and potential *MtPylRS*) there is no clear general trend in efficiency. Surprisingly, *MburPylRS* shows higher OTS efficiency for each substrate compared to all other PylRSs at almost every concentration. This enzyme shows significant activity even at very low (0.05 mM, 0.1 mM) substrate concentrations. Remarkably, the *MpPylRS* derived from the organism with the lowest OGT has the lowest incorporation efficiency. Here, further analyses would be interesting to find out whether the low

efficiency is an intrinsic property of the enzyme or just a byproduct of the low stability at cultivation conditions (37 °C)

In terms of efficiency, the two mesophilic PylRS present a mixed picture. The *MmPylRS* has a slightly higher efficiency at low ncAA concentrations in comparison with the *SmbP-MbPylRS* which is consistent with the literature.¹⁴³ At high AzidoK (**34**) concentrations (9 mM) the *SmbP-MbPylRS* can even outperform the *MmPylRS*.

Unexpectedly, the thermophilic *MtPylRS*(TM-1) operates on the same level as the mesophilic enzymes and even exhibits higher OTS efficiencies than the mesophilic *MmPylRS* for some substrates (AlloK (**32**), ProK (**33**), AzidoK (**34**) and BenzK (**43**)) at medium high/high ncAA concentrations. In contrast to hyperthermophilic enzymes, which always show no or very low activity at 37 °C, this seems plausible since the *MtPylRS*(TM-1) originates from an organism with an OGT of 50 °C which is only moderately thermophilic.

It is surprising that a standard ncAA concentration of 1 mM has been used so far even for good substrates (BocK (**31**) and AlloK (**32**)).^{163,302,303} For engineered PylRS, this concentration is often far exceeded (up to 10 mM) to reach satisfactory incorporation efficiency.^{146,147,304} At a concentration of 0.05 mM BocK (**31**), the *MburPylRS* and *MmPylRS* show more than 50% of their maximum activity. At a concentration greater than 0.3 mM BocK (**31**), their efficiencies no longer increase. In general, *MburPylRS* shows 30% (BocK (**31**)) to 280% (ProK (**33**)) higher OTS efficiency, at a concentration of 0.05 mM ncAA, than the second-best variant. At 0.3 mM ProK (**33**), the difference is even 440 %. Even though the absolute advantage of *MburPylRS* decreases with increasing ncAA concentrations, at the highest ncAA concentration it is still by far the most efficient enzyme for almost every substrate (besides AlloK (**32**)). For PhotoK (**42**) and BenzK (**43**), the OTS efficiency for *MburPylRS* does not increase in the last step between 3 and 9 mM ncAA concentration, again highlighting its efficiency at lower concentrations. Since there is an *MbPylRS* engineered specifically for the incorporation of PhotoK (**42**), the transfer of these mutations should significantly increase the already good incorporation efficiency for *MburPylRS* at low concentrations.³⁰⁵

Since many of the commonly used ncAAs are very expensive and therefore even prohibit certain research projects, increasing the incorporation efficiency at low ncAA concentrations with the *MburPylRS* could drastically advance these endeavors. Experiments with mammalian cells will also now become even more feasible, since these cell lines are generally more sensitive to high concentrations of additives that are unnecessary for survival.

2.4.3.1. Multi-Site ncAA Incorporation in BL21(DE3) and B95.ΔA E. coli strains

To compare the efficiency of *MburPylRS* with the best known PylRS OTS (*MmPylRS*), two substrates (AlloK(**32**) and AzidoK(**34**)) were tested in conjunction with SUMO-sfGFP reporter

constructs containing between one and five stop-codons. Performance was assessed in a normal *E. coli* (BL21(DE3)) and a BL21(DE3) derived RF-1 knockout strain (B95. Δ A).¹⁵⁴ The latter was optimized for the incorporation of ncAAs at multiple sites. As a reference, the wild-type construct without stop-codons was produced under exactly the same conditions.

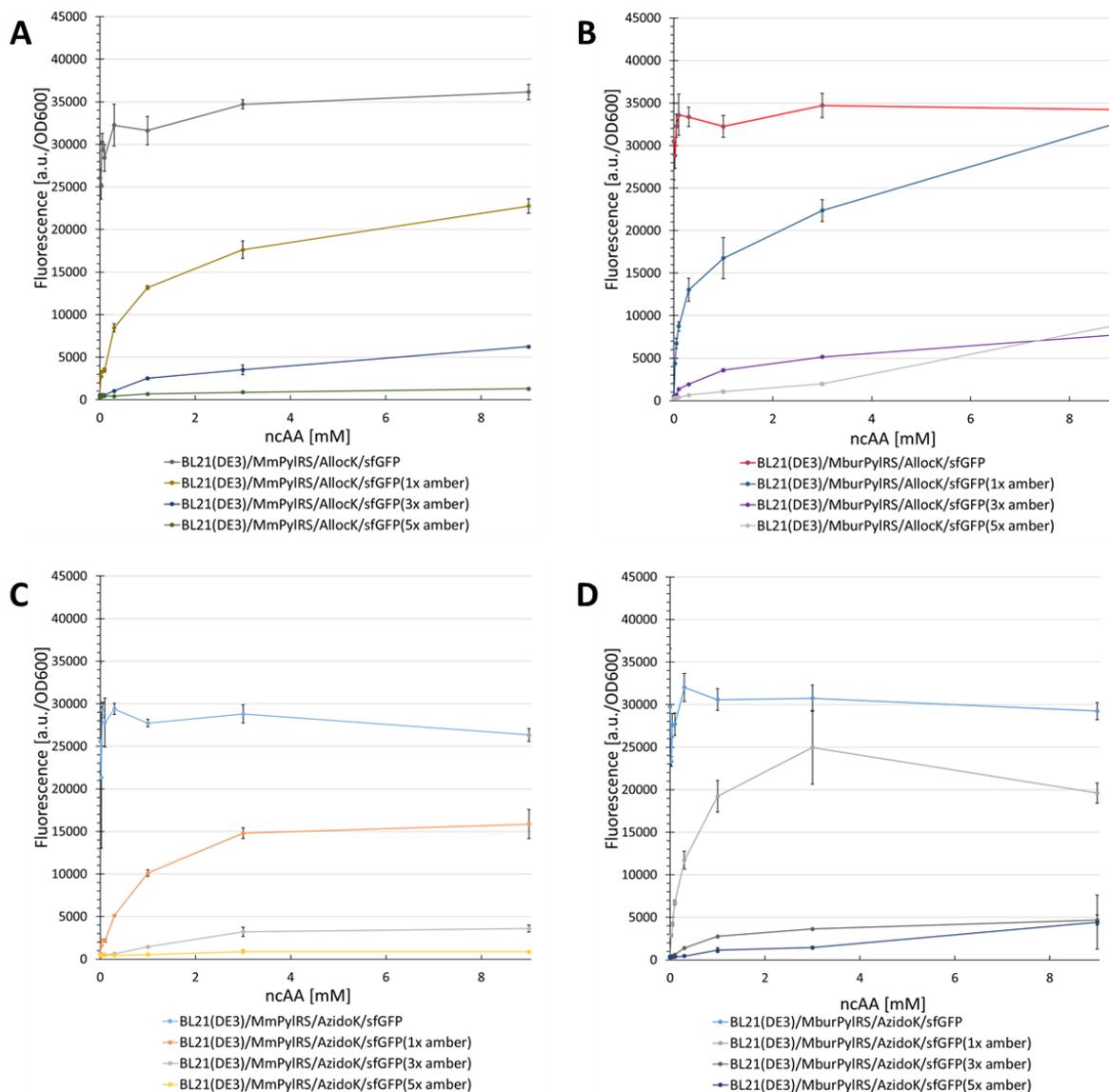


Figure 38: Concentration-dependent unnatural protein production with *MmPylRS* (A, C) and *MburPylRS* (B, D) for different ncAA/reporter construct combinations. The host for protein production was BL21(DE3). Endpoint measurements for ncAA concentrations of 0.025, 0.05, 0.1, 0.3, 1, 3 and 9 mM. Gain 80.

Figure 38: shows that that for BL21(DE3) expression of sfGFP(1x amber), the performances of both PylRS are comparable when supplied with AlloK(32). In the same setup but with AzidoK(34), the *MburPylRS* clearly outperforms *MmPylRS*, and is twice as efficient. These results are consistent with **Figure 37**. With the suppression of more than one stop-codon in the BL21(DE3) strain, the performance decreases significantly, but the *MburPylRS* shows at least twice the efficiency than the *MmPylRS* in the range between 1 and 3 mM ncAA, albeit at a low level. For protein production in the B95. Δ A strain, the picture for suppression of one stop-

codon is similar to the BL21(DE3) experiment, but with generally higher suppression efficiencies (**Figure 39**). Encouragingly, it is possible with *MburPylRS* to achieve wild-type level protein production for both ncAAs when 1 mM is supplied. A comparison of the OTS performances with *MburPylRS/AlloCk* (**32**) shows that the decrease in suppression efficiency of one to five stop-codons /when supplied with 1 mM) is 49%. For *MmPylRS*, the decrease is 77% which is over 50% higher.

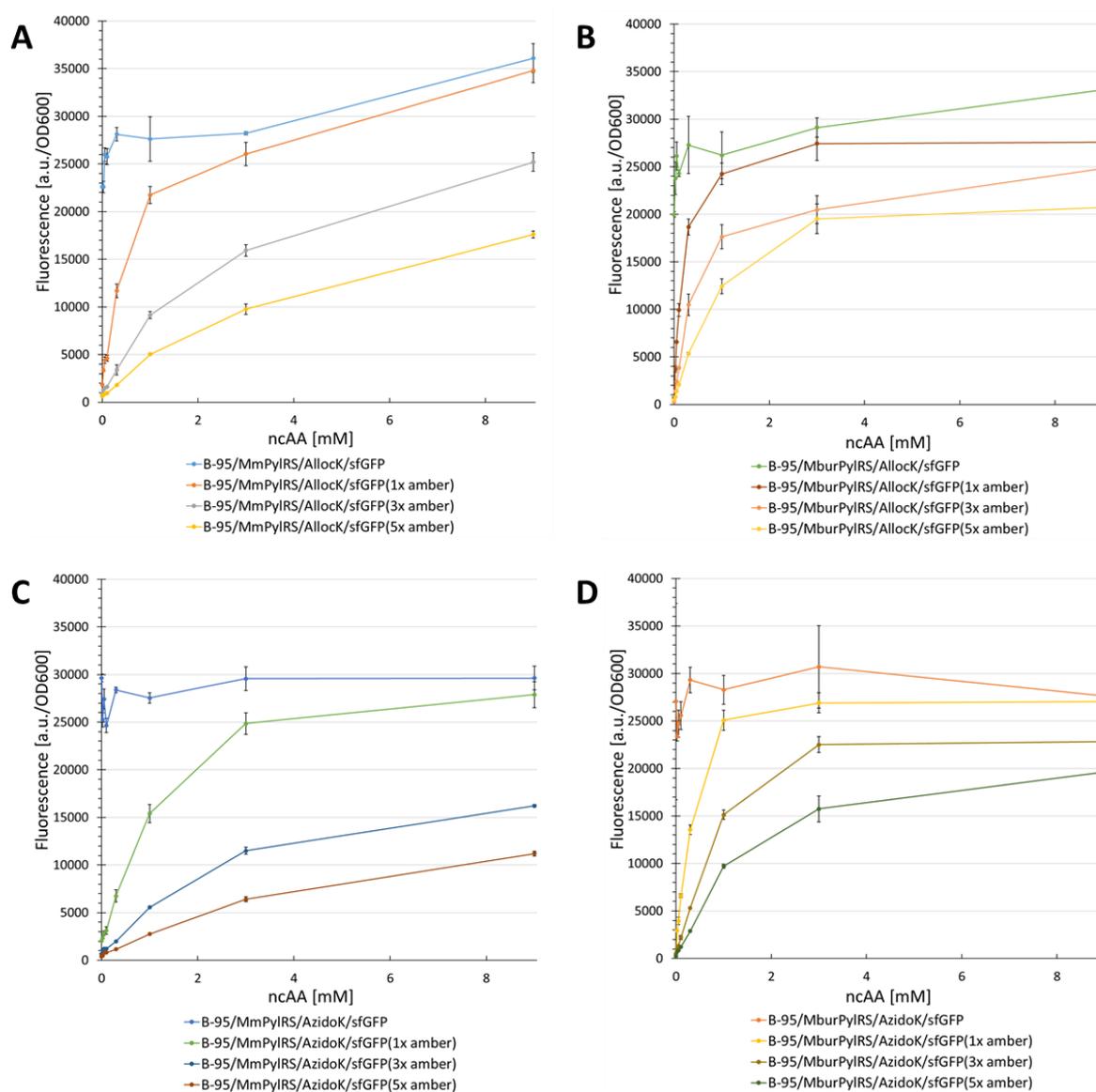


Figure 39: Concentration-dependent unnatural protein production with *MmPylRS* (**A, C**) and *MburPylRS* (**B, D**) for different ncAA/reporter construct combinations. The host for protein production was B-95.ΔA. Endpoint measurements for ncAA concentrations of 0.025, 0.05, 0.1, 0.3, 1, 3 and 9 mM. Gain 80.

Performance does not decrease as much when more than one stop codon is suppressed compared to BL21(DE3). This is where the advantage of *MburPylRS* becomes most apparent. At AzidoK(**34**) concentration of 1 mM and suppression of three and five stop-codons, the performance is three times better than that of *MmPylRS*. This suggests that *MburPylRS* is not only very efficient at low ncAA concentrations, but also more suitable for incorporation of ncAAs

at multiple sites. This advantage might be even more pronounced in the absence of RF-1 competition, for example, in organisms with liberated codons or with sense codon suppression.

2.4.4. Estimation of the *In Vivo* PylRSs Substrate Promiscuities

Interestingly, the concentration-dependent ncAA incorporation in **chapter 2.4.3** shows that at high ncAA concentrations (9 mM), all PylRSs can be classified as functioning (except for BenzK (**43**)). The substrate recognition gradient showed that the *Mbur*PylRS appears to be more efficient at low concentrations (0.05 mM) for all ncAAs and at 3 mM for BenzK (**43**). This could suggest a higher promiscuity of *Mbur*PylRS. Therefore, a method was developed to estimate the promiscuity of the PylRSs based on the *in vivo* data (**Table 7**). The magnitude of promiscuity is usually described by how the kinetic parameters for the promiscuous substrate compare with those for the native substrate.³⁰⁶ Generally, this is done with kinetic enzyme data acquired *in vitro*, mainly to exclude the multitude of effects that normally bias reaction rates in living cells. Ultimately, for recombinant protein production *in vivo* data are more important. An attempt was made to build a model based on these *in vivo* data. The assumption was that at a concentration of 9 mM ncAA in the medium, the concentration should be high enough to reach saturation of the enzyme in the cytosol. *Ceteris paribus*, this would imply that only the catalytic efficiency of PylRS is rate-limiting and therefore correlates with the fluorescence readout. This assumption is supported by cultivation data in similar setups with high AzidoK (**34**) and BockK (**31**) concentrations (between 1 and 10 mM) which showed that approximately 90% of the supplied ncAA is found inside the cells.¹⁴⁷ It was also assumed that this ncAA behavior is existent for the analogs.

The incorporation efficiencies of all substrates were correlated with the substrate that is best incorporated to give relative performance values (the value of the substrate was divided by the value of the best performing substrate). The numbers were averaged to obtain two coefficients (one omitting BenzK (**43**)) for the substrate promiscuity. For a perfectly promiscuous enzyme this value would be 100%. For a perfectly specific enzyme, the lowest number would depend on the number of substrates included. Here it would be $\frac{100\%}{6} = 17\%$ (or 20% without BenzK). Since BenzK (**43**) is not a good substrate, the coefficient values generally decrease when included in the average efficiency, but not equally for every construct. Unfortunately, the standard deviation for the BenzK (**43**) set is very high, therefore the coefficients without these values are the more robust ones. **Table 7** shows that the psychrophilic PylRSs are generally more promiscuous in comparison to the other constructs. An exception is the mesophilic SmbP-*Mb*PylRS which has the same level as the most promiscuous variants. The *in vitro* AA activation data of *Mb*PylRS and *Mm*PylRS also suggest that *Mb*PylRS is somewhat more promiscuous than *Mm*PylRS (indicated by the higher *Mb*PylRS K_m values for amino acid activation of Pyl).¹⁶⁹

Table 7: PyIRSs promiscuity estimations with 9 mM ncAA concentration.

PyIRS	AlloK [%]	Bock [%]	AzidoK [%]	ProK [%]	PhotoK [%]	BenzK [%]	average efficiency with BenzK [%]	average efficiency without BenzK [%]
<i>Mp</i> PyIRS	100	90	81	18	61	5	59	70
<i>MIP</i> PyIRS	99	100	52	45	60	2	60	71
<i>Mbur</i> PyIRS	84	71	100	66	50	5	63	74
<i>Mm</i> PyIRS	100	70	41	41	52	3	51	61
<i>SmbP-Mb</i> PyIRS	86	100	97	57	31	5	63	74
<i>Mt</i> PyIRS	100	56	75	37	40	6	52	62
<i>Mt</i> PyIRS(TM-1)	100	55	64	39	36	4	50	59

The lowest promiscuity is observed for *Mt*PyIRS(TM-1) which would be expected for a thermophilic enzyme. Interestingly, the promiscuity values of *MIP*PyIRS and *Mt*PyIRS differ by 15%, although their catalytic domain sequence is 97% identical (see **chapter 2.4.2**). This suggests that the differences in linker length and/or N-terminal domain are responsible for that difference. If that the difference stems from the catalytic domain this would certainly suggest unidentified key residue involvement. In general, the substrate promiscuity trends observed in **Table 7** are in line with what is expected from the literature knowledge for extremophilic enzymes, although the sample size is limited.^{209,212}

The promiscuity of an enzyme is, by definition, unpredictable.³⁰⁶ Therefore, the promiscuity coefficient could be useful as an indicator for assessing the engineering ability of PyIRS variants for novel substrate recognition. The practicality of these coefficients is underscored by the fact they agree well with experimental data for three distinct classes of mutant PyIRS in the following chapters (see **chapter 2.4.5**). The mutants of *Mbur*PyIRS and *SmbP-Mb*PyIRS always performed best.

2.4.5. Comparing OTS Efficiencies of Psychro-, Meso-, and Thermophilic PyIRS Mutants

To verify the promising results obtained in **chapter 2.4.3** and **2.4.4**, further experiments were performed with PyIRS sets containing variants known to enable the recognition of various substrates that are very different from Pyl. To obtain a complete picture of PyIRS promiscuity, the mutations for these variants were transferred to the most powerful psychro-, meso- and thermophilic enzymes. More specifically, *Mbur*PyIRS, *SmbP-Mb*PyIRS, *Mm*PyIRS and *Mt*PyIRS(TM-1)). A homology model was used to transfer the mutations. The main question was whether the differences in thermal enzyme origin were reflected in differences in substrate specificity, as shown in **Table 7**.

2.4.5.1. Elucidating the General Usefulness of the SmbP-Tag for Other PylRS Mutants with Destabilizing Mutations

Before obtaining information on substrate specificities, the general mutant applicability of the SmbP-tag was clarified. To this end, the performance of the PylRSs double Gly and double Ala variants (in analogy to **chapter 2.2.6**) were compared to see if these destabilizing double Gly mutations would have an impact on the *in vivo* OTS performance. The equivalent of the *MbPylRS*(N311G:C313G) mutations were transferred to the *MburPylRS*, *MmPylRS* and *MtPylRS*(TM-1) respectively with and without the SmbP-tag. The efficiency of incorporation was then tested with O-*tert*-butyl-Y (**39**) and ONBY (**41**) (**Figure 124****Figure 129**). No improvements were observed except for the *MbPylRS* variant. Therefore, the tag was omitted for all variants in the following experiments, except *MbPylRS*. This was also in agreement with the results from **chapter 2.4.3**.

2.4.5.2. Comparing PylRS Double Alanine/Glycine Mutants for Incorporation of Tyrosine and Phenylalanine Derivatives

The PylRS double alanine variant formed a new class of PylRS enzymes specific to a number of tyrosine and phenylalanine analogs (see **chapter 1.3.1.2**).¹⁶⁵ It was shown in **chapter 2.2.5** that the corresponding double glycine mutant can incorporate some of the same ncAAs and was therefore also investigated. **Figure 40** shows the ncAAs used in this chapter.

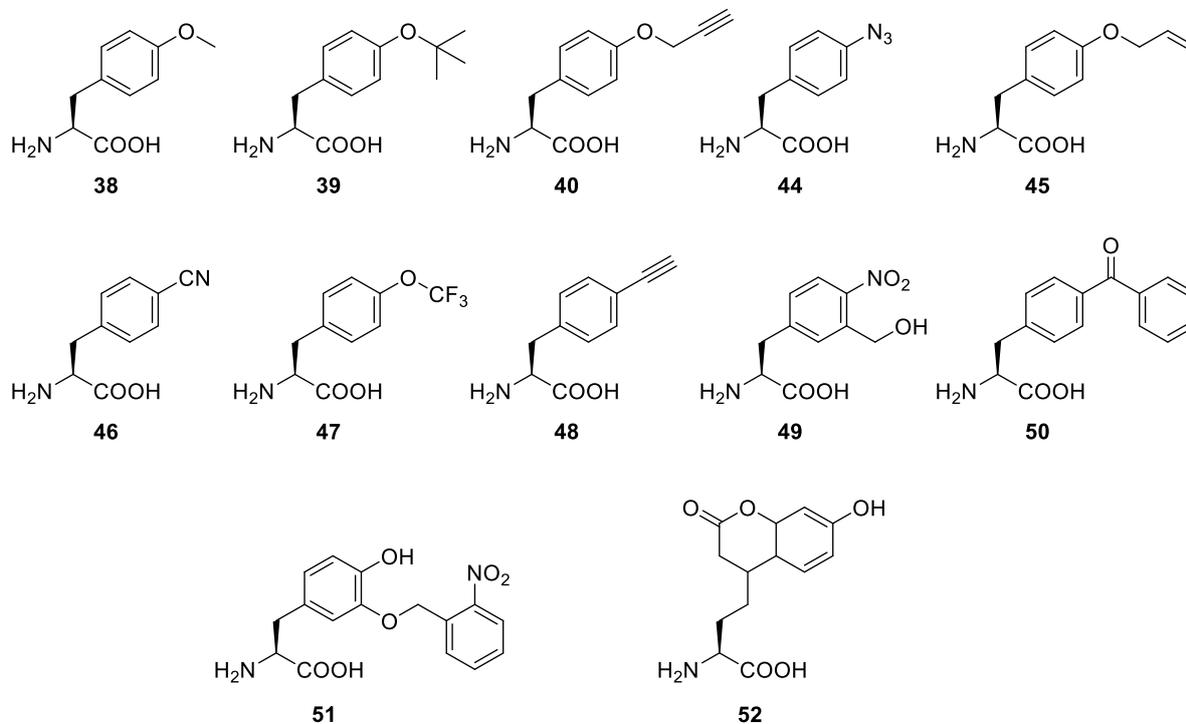


Figure 40: Phe/Tyr derivatives used in this study. O-methyl-Y = O-methyl-L-tyrosine (**38**), O-*tert*-butyl-L-tyrosine (**39**), O-prop-Y = O-propargyl-L-tyrosine (**40**), azido-F = 4-azido-L-phenylalanine (**44**), O-allyl-Y = O-allyl-L-tyrosine (**45**), cyano-F = 4-cyano-L-phenylalanine (**46**), O-CF₃-Y = O-CF₃-L-tyrosine (**47**), ethynyl-F = 4-ethynyl-L-phenylalanine (**48**), p-oNB-alanin (**49**), Bpa = 4-benzoyl-L-phenylalanin (**50**), *m*-oNB-Dopa = *o*-(2-Nitrobenzyl)-3,4-dihydroxyphenylalanine (**51**), coumarin = H-(7-Hydroxycoumarin-4-yl)-ethyl-Gly-OH (**52**).

The ncAAs include substrates known to be recognized by PyIRSs OTSs (**38**, **39**, **40**, **45**, **50**) and substrates that would be useful for incorporation with this OTS (**44**, **46**, **47**, **49**, **51**, **52**). The ncAAs contain a variety of useful biochemical functionalities, e.g. for NMR (**39**²⁴⁵, **47**³⁰⁷), click-chemistry (**40**, **44**, **48**)²²⁴, cross-metathesis (**45**)³⁰⁸ or as genetically encoded PTM (**38**)^{242,309}, spectroscopic probe (**46**³¹⁰, **48**³¹¹, **52**³¹²), photocaged ncAA (**51**¹⁸⁸) and photo cross-linker (**44**³¹³, **50**³¹⁴). For possible applications of ncAA **49** see **chapter 2.5**.

The experimental data for all substrates are shown in a heat map (**Figure 41**). SmbP-*Mb*PyIRS or *Mbur*PyIRS show the highest incorporation efficiencies for all substrates underlining the promiscuity results from **chapter 2.4.4**. Moiety **52** was not included in the heat map because the OD₆₀₀ values for the *Mt*PyIRS(TM-1) and *Mm*PyIRS constructs were unusually low (below 0.1). This suggests some toxicity for **52** and thus no robust results. Further analysis is needed to determine whether incorporation with this substrate is possible. In addition to the general efficiency trends, some subtle information can be extracted from the heat maps. It can be observed that the double Gly mutants prefer the slightly shorter but bulkier substrate **39** and that substrate **50** is exclusively incorporated with the double Gly mutant. This would be reasonable, since the double-G mutant creates even more space for the substrates than the double-A constructs. Interestingly, the general incorporation trend of all PyIRSs deviates only once.

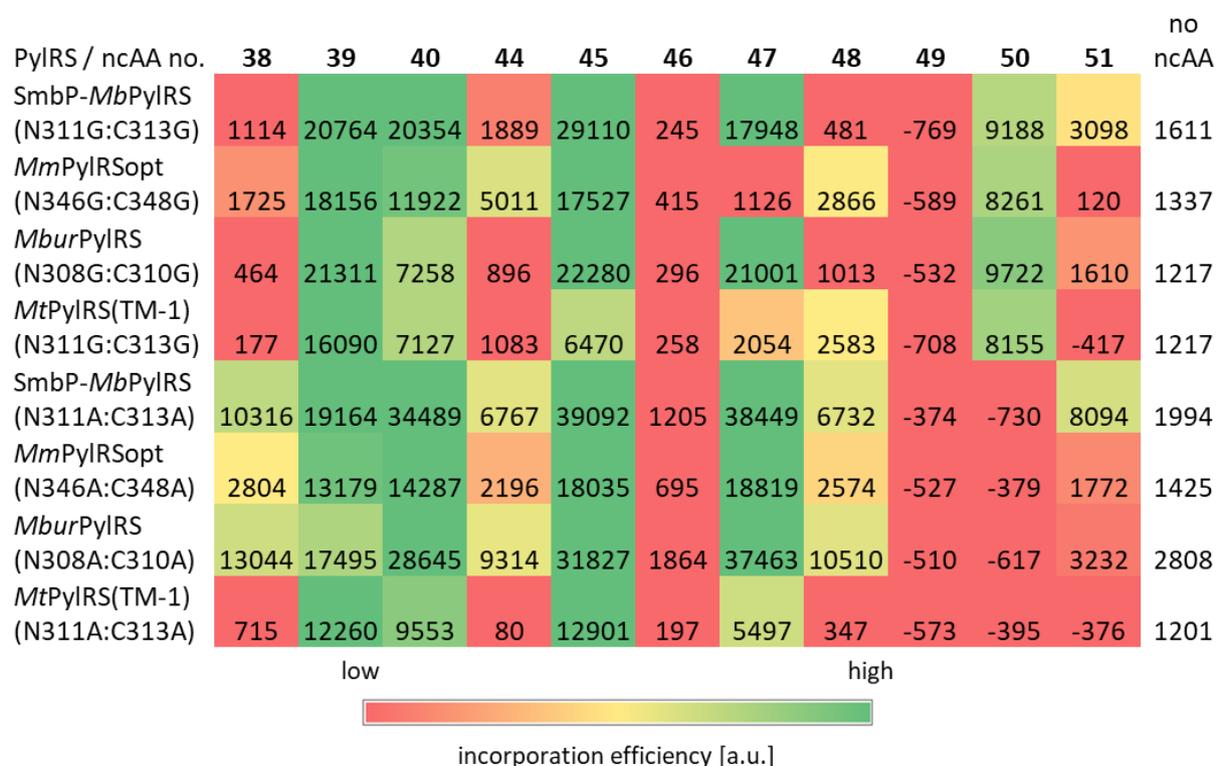


Figure 41: Heatmaps of incorporation efficiencies of selected PyIRS mutants with 10 mM added ncAA. Heatmaps are based on the fluorescence data (Figure 108 and Figure 109) from sfGFP(R2amber) reporter protein assays. Background signal (no ncAA supplied) is subtracted from the initial signal values and also shown as reference. Cutoffs: below (no ncAA value) = red, 2*(no ncAA value) = yellow, 10*(no ncAA value) = green.

For substrate **44**, the *Mm*PyIRS double Gly construct incorporates the substrate more efficiently than SmbP-*Mb*PyIRS and *Mbur*PyIRS, where the double Ala variant performs better. This indicates a slightly different active site architecture between these variants. It is also noteworthy that **51** is incorporated well only with SmbP-*Mb*PyIRS which indicates more space available in the active site for meta substituted Phe derivatives. 4-Azido-F (**44**), ethynyl-F (**48**) and *m*-oNB-Dopa (**51**) have never been incorporated with an archaeal PyIRS OTS. 4-Azido-F (**44**) incorporation was claimed with the *Mm*PyIRS(N346A:F347L:C348G) mutant, but the data were not published, only the data for the bacterial equivalent.²⁸⁰

2.4.5.3. Detailed In Vivo Activity Analyses of the Double Alanine/Glycine Mutants

To obtain a detailed picture of OTS performance, the ncAAs with the highest incorporation in **Figure 41** were measured in a concentration-dependent manner. Moieties **40**, **45** and **47** were selected as good substrates, **39** as medium substrate, and **44** as poor substrate. The selected concentrations are shown in the heat maps (**Figure 42**). These heat maps were derived from the fluorescence data in the Appendix (**Figure 110-Figure 114**). Compared to the wild-type PyIRS enzymes, all of these variants exhibit lower efficiencies. High fluorescence signals were not obtained before a concentration 5 mM ncAA, except for O-CF₃-Y (**47**), which is well incorporated at a concentration of 1 mM with the *Mbur*PyIRS variant. This efficiency is at least an order of magnitude lower than for the wild-type. At a concentration of 5 mM ncAA, the efficiencies of *Mbur*PyIRS's and SmbP-*Mb*PyIRS's are equal and exceed those of *Mm*PyIRS and *Mt*PyIRS(TM-1).

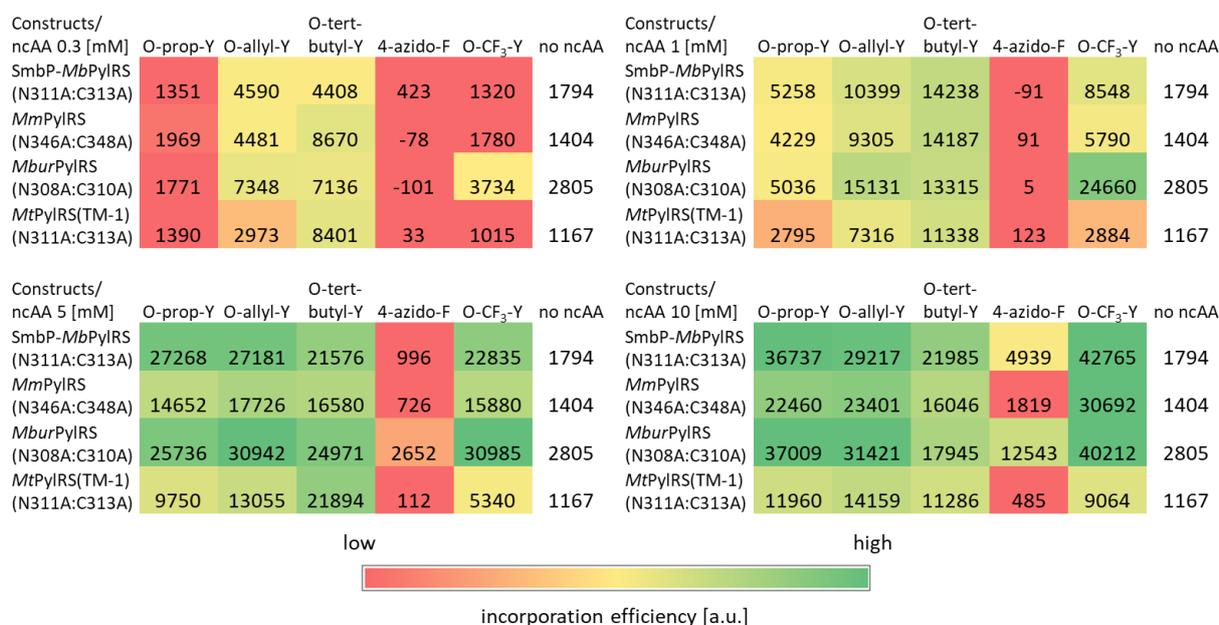


Figure 42: Heatmaps of incorporation efficiencies of the selected PyIRSs with different ncAA concentrations. The heatmaps are based on the fluorescence data (Figure 110-Figure 114) from sfGFP(R2amber) reporter protein assays. The background signal (no ncAA supplied) is subtracted from the initial signal values and also shown as reference. Cutoffs: below (no ncAA value) = red, 2*(no ncAA value) = yellow, 10*(no ncAA value) = green.

At a concentration of 1 mM, the psychrophilic and mesophilic PyIRS are at the same level, except for O-CF₃-Y (**38**). The major difference in incorporation activity is that the *Mbur*PyIRS can incorporate 4-Azido-F (**44**) well at a concentration of 10 mM, but also has a slightly higher background incorporation than the other constructs.

2.4.5.4. Elucidating Incorporation Efficiencies for Aliphatic ncAA Incorporating PyIRSs

In **chapter 2.3** a new class of PyIRS was developed that can incorporate aliphatic ncAAs with smaller side chains. This class was based on the SmbP-*Mb*PyIRS variant that was the result of the improvement experiments from **chapter 2.2**. The key mutations found were transferred to the PyIRS set from before to get an overview of the OTSs performances in conjunction with aliphatic ncAA incorporation. They correspond to the *Mb*PyIRS mutations N311M:C313W, N311Q:C313W and N311L:C313W. Most of the ncAAs tested in **chapter 2.3** were also used for these assays. **Figure 43** shows heat maps derived from the fluorescence data in the appendix (**Figure 115-Figure 123**). In general, it can be stated that not all mutations that are functional in the SmbP-*Mb*PyIRS variants lead to similar results in other PyIRSs. This should highlight the fact that not all PyIRS mutations are 1:1 transferrable, but there is a good overlap. For all ncAAs that can be well incorporated with the SmbP-*Mb*PyIRS constructs, the *Mbur*PyIRS variants are more efficient, except for the incorporation of (*S*)-2-aminoheptanoic acid (**2**), (*S*)-2-aminooctanoic acid (**3**) and L-ethionine (**25**). The *Mbur*PyIRS(N308M:C310W) exhibited even a stronger tendency to incorporate ncAAs with a chain length shorter than C7 in comparison to the *M. barkeri* construct. Compared with the SmbP-*Mb*PyIRS(N311Q:C313W) the *Mbur*PyIRS(N308Q:C310W) was almost completely inactive. Most surprising was the fact that the mesophilic SmbP-*Mb*PyIRS variants exhibited a wide range of incorporation activity, whereas the *Mm*PyIRS constructs, also mesophilic, were almost all inactive. This also underscores the promiscuity results in **chapter 2.4.4**. The *Mb*PyIRS variants found in **chapter 2.2** would have never been found if the *Mm*PyIRS were used instead of the SmbP-*Mb*PyIRS. Surprisingly, not a single reasonably active variant was found for the thermophilic *Mt*PyIRS(TM-1).

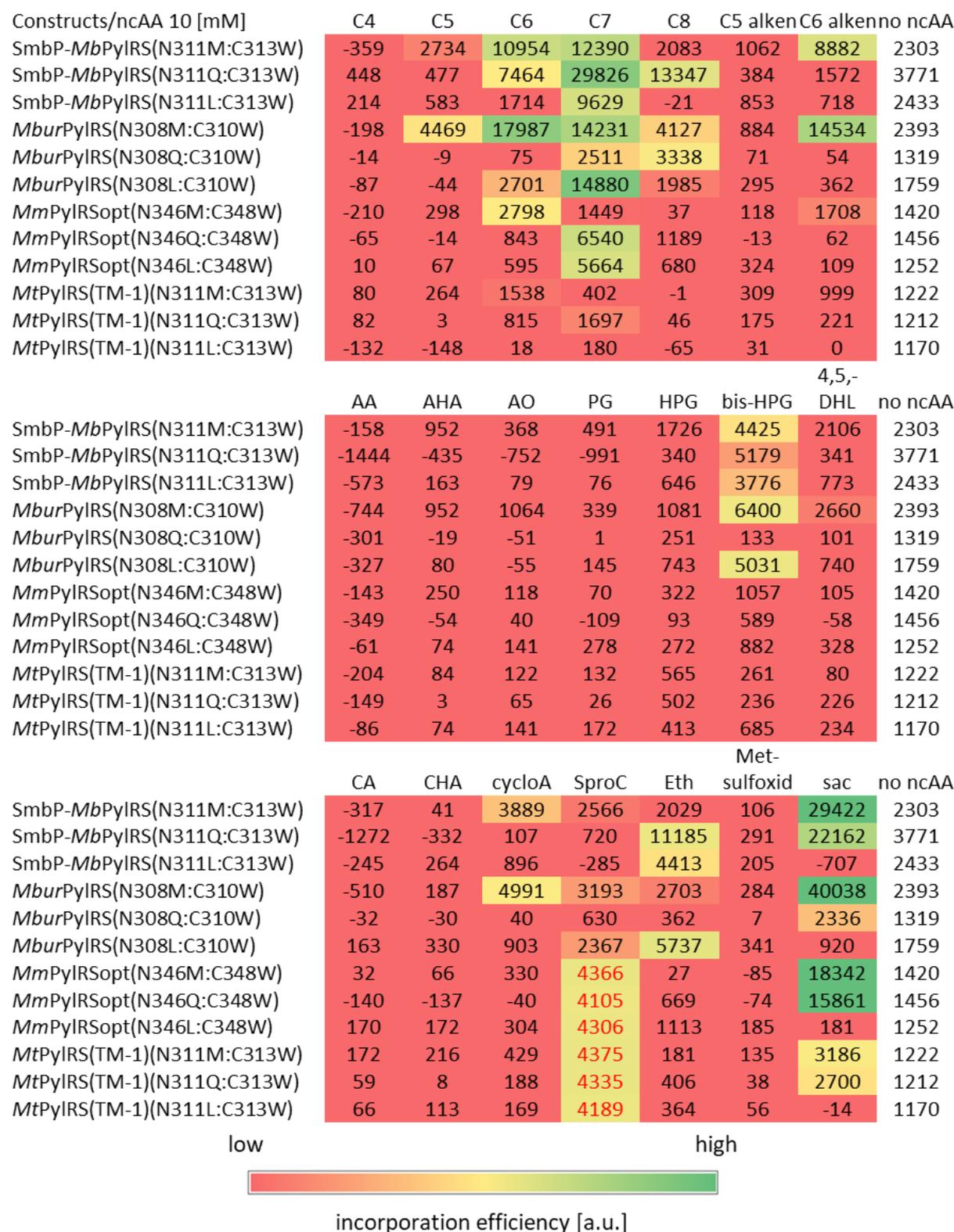


Figure 43: Heatmaps of incorporation efficiencies of the selected PyIRs with different ncAA concentrations. The heatmaps are based on the fluorescence data (Figure 115-Figure 123) from sfGFP(R2amber) reporter protein assays. The background signal (without ncAA) is subtracted from the initial signal values and also shown as reference. Cutoffs: below (no ncAA value) = red, 2*(no ncAA value) = yellow, 10*(no ncAA value) = green. Red numbers mean that the measured OD was very low (below 0.1). These values could therefore represent distorted activity and are therefore excluded in the interpretation. The ncAAs are: C4 = (S)-2-aminobutyric acid (**10**), C5 = (S)-2-aminopentanoic acid (**7**), C6 = (S)-2-aminohexanoic acid (**5**), C7 = (S)-2-aminoheptanoic acid (**2**), C8 = (S)-2-aminooctanoic acid (**3**), C5 alken = (S)-2-aminopent-4-enoic acid (**8**), C6 alken = (S)-2-aminohex-5-enoic

acid (**6**), AA = (S)-2-amino-3-azidopropanoic acid (**14**), AHA = (S)-2-amino-4-azidobutanoic acid (**15**), AO = (S)-2-amino-5-azidopentanoic acid (**16**), PG = (S)-2-aminopent-4-ynoic acid (**13**), HPG = (S)-2-aminohept-6-ynoic acid (**12**), bis-HPG = (S)-2-aminohept-6-ynoic acid (**11**), 4,5-DHL = amino-4-methylpent-4-enoic acid (**21**), CA = (S)-2-amino-3-cyanopropanoic acid (**18**), CHA = (S)-2-amino-4-cyanobutanoic acid (**19**), cycloA = (S)-2-amino-3-cyclopropylpropanoic acid (**9**), SproC = S-propargyl-L-cystein (**27**), Eth = L-ethionine (**25**), Met-sulfoxid = L-methionine sulfoxide (**23**), Sac = S-allyl-L-cystein (**1**). Gain 85.

2.4.5.5. Comparing OTS Efficiencies of SacRS and Improved SacRS Variants

The last PyIRS set tested is the more specific variant SacRS. The two mutations required by the PyIRSs to recognize Sac are the same as those of *Mb*PyIRS(C313W:W382S). The four PyIRS tested are shown in **Figure 44A**. There, the superiority of *Mbur*SacRS at low Sac concentrations can be clearly seen. At a Sac concentration of 0.6 mM, the improvement is approximately 120% compared to the two mesophilic PyIRSs. Suspiciously, no activity could be detected with *Mt*SacRS(TM-1). Therefore, this variant was tested again with the SmbP-tag, although its usefulness was ruled out earlier (see **chapter 2.4.5.1**). Fortunately, the tag showed some recovery of activity, highlighting the applicability of the solubility tag strategy to other PyIRSs (**chapter 2.2**).

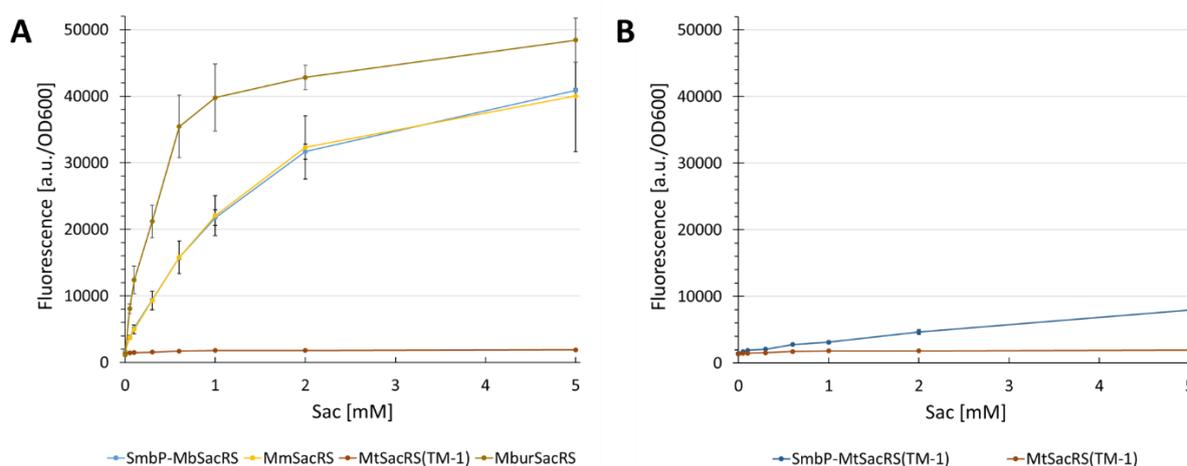


Figure 44: Concentration-dependent unnatural protein production with **A**) SmbP-*Mb*SacRS, *Mm*SacRS, *Mt*SacRS(TM-1) and *Mbur*SacRS. **B**) *Mt*PyIRS(TM-1) with and without SmbP-tag. Using BL21(DE3) cells and sfGFP(1x amber) as reporter. Endpoint measurements for Sac (**1**) concentrations of 0.05, 0.1, 0.3, 0.6, 1, 2 and 5 mM. Gain 85.

The mutations found (in **chapter 2.3.4**) to increase Sac (**1**) incorporation efficiency (**Figure 76** and **Figure 77**) were also tested for the *Mbur*SacRS, using Sac (**1**) and SproC (**27**) as substrates (**Figure 45**). Although *Mbur*SacRS was already very active, the S379T mutation increased efficiency by 80% with 0.6 mM Sac (**1**) and for SproC (**27**) by 120%. Interestingly, nearly 90% of the maximum suppression signal is achieved at 0.3 mM supplied Sac (**1**), indicating very high activity that rivals the best performing PyIRS wild-type substrate BockK (**31**). The observation of *Mbur*PyIRS exhibiting an even higher relative increase in performance on suboptimal substrates like SproC (**27**) compared with the other PyIRS constructs is consistent with the overall higher promiscuity of *Mbur*PyIRS as demonstrated in this study. All SacRS

variants with the S→T mutation were compared in an assay to obtain information on their relative performance (**Figure 46**). Of note is the generally high catalytic efficiency for Sac (**1**) incorporation, with the efficiency of *MburSacRS*(S379T) being extremely high at low concentrations. Interestingly, the two mesophilic PylRS variants outperform *MburSacRS*(S379T) at concentrations above 1 mM Sac (**1**), but at a very high level (the signal is between 16 and 18 times the background suppression signal). The incorporation performances for SproC (**27**) are consistent with the promiscuity coefficients and show that the *M. burtonii* variant performs best, while *M. barkeri* is second and *M. mazei* third. The low signal at 9 mM SproC (**27**) with *MburSacRS*(S379T) is due to inexplicably low growth ($OD_{600} < 0.1$). However, since the maximum signal is already reached at 3 mM, this is negligible.

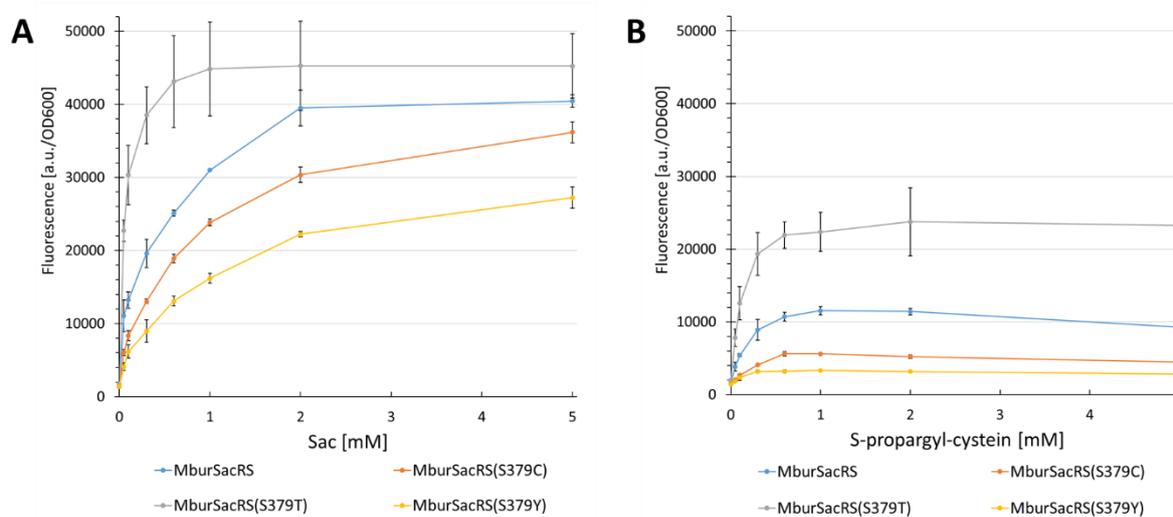


Figure 45: Concentration-dependent unnatural protein production with **A**) *MburPylRS* mutants and supplemented Sac (**1**). **B**) *MburPylRS* mutants and supplemented SproC (**27**). Endpoint measurements for ncAA concentrations of 0.05, 0.1, 0.3, 0.6, 1, 2 and 5 mM. Gain 85.

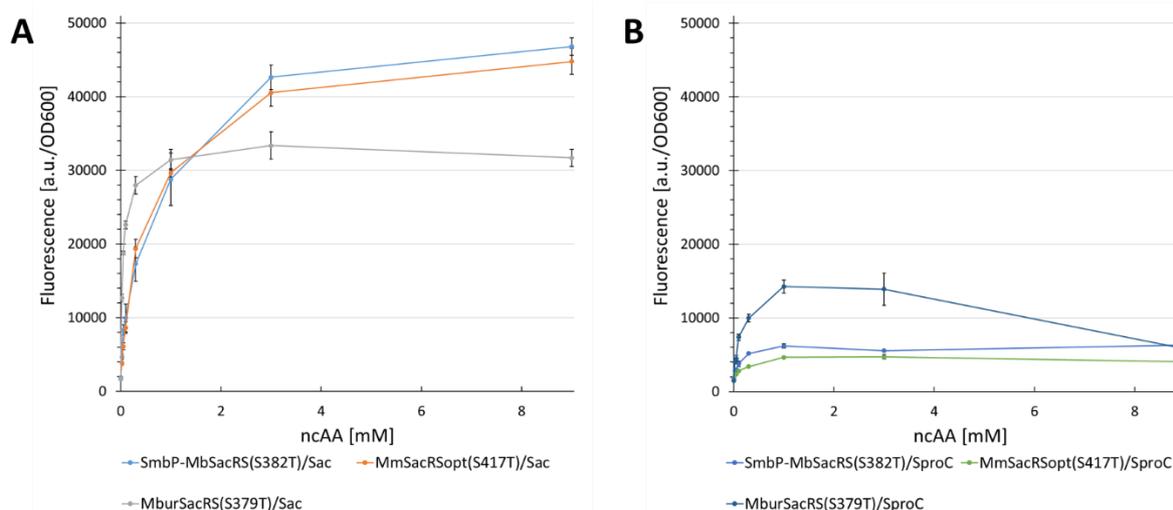


Figure 46: Concentration-dependent unnatural protein production for SacRS mutants and supplemented with **A**) Sac (**1**). **B**) SproC (**27**). Using BL21(DE3) cells and sfGFP(1x amber) as reporter. Endpoint measurements for ncAA concentrations of 0.025, 0.05, 0.1, 0.3, 1, 3 and 9 mM. Gain 80.

2.4.5.6. Multi-Site Incorporation of Sac (1) and SproC (27)

Based on the impressive efficiency of the SacRS variants with the S→T mutation, the *M. barkeri* and *M. burtonii* variants were investigated for the multi-site incorporation of ncAAs with two different *E. coli* strains, analogous to **chapter 2.4.3.1 (Figure 47)**. The *M. barkeri* variant rather than the *M. mazei* variant was chosen because of slightly better performance with SproC (27) (and equal Sac (1) activity). For reporter protein production in BL21(DE3) with Sac (1) added, the relative performance of *MburSacRS*(S379T) compared with *SmbP-MbSacRS*(S382T) increases with the number of in-frame stop-codons. With one stop codon, the performance is 230% higher (at 0.3 mM), with three stop codons it is 490% higher (also at 0.3 mM), and five stop codons result in no incorporation with *SmbP-MbSacRS*(S382T).

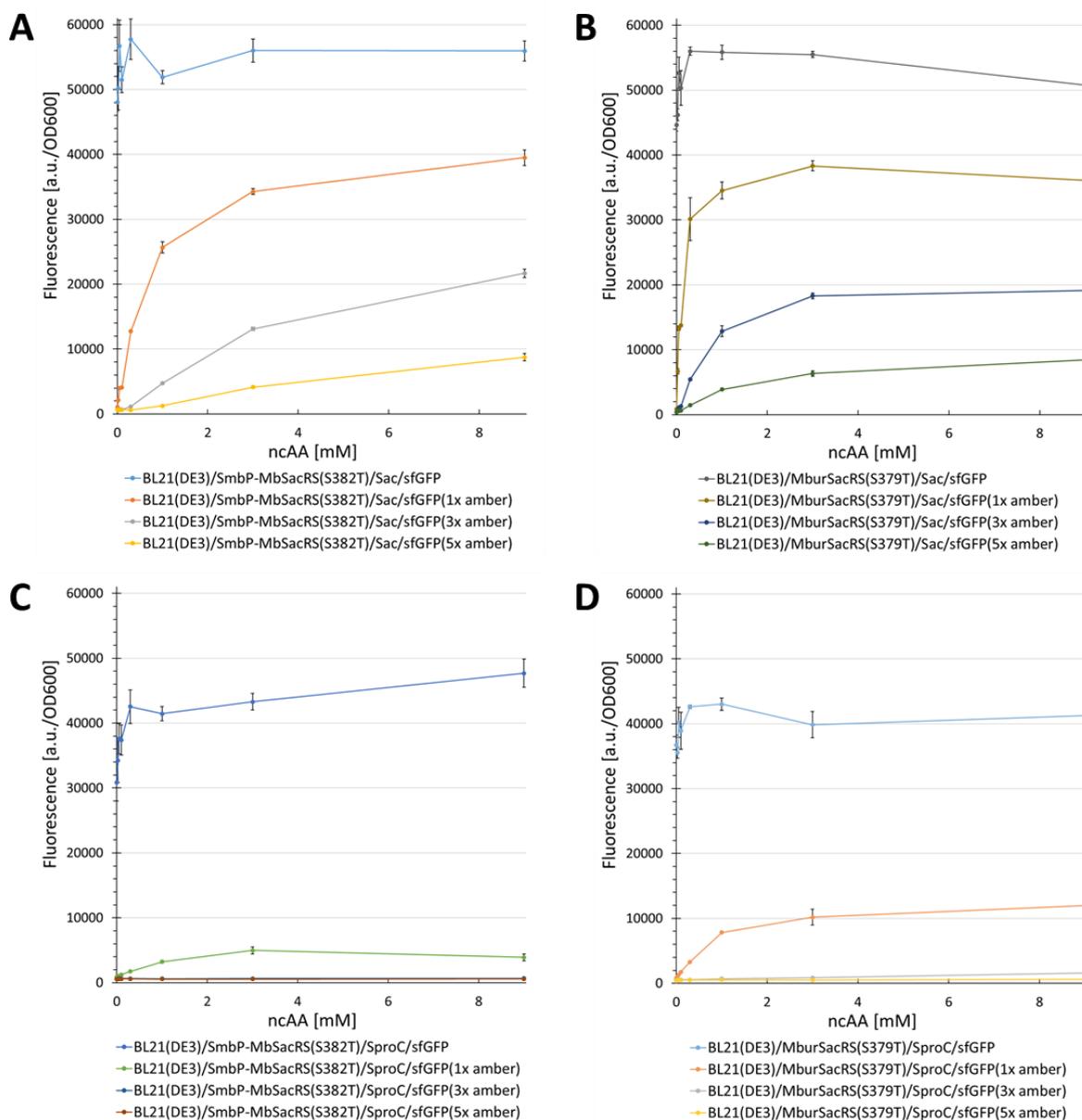


Figure 47: Concentration-dependent unnatural protein production with *SmbP-MbPylRS* (**A, C**) and *MburPylRS* (**B, D**) for different ncAA/reporter construct combinations. The host for protein production was BL21(DE3). Endpoint measurements for ncAA concentrations of 0.025, 0.05, 0.1, 0.3, 1, 3 and 9 mM. Gain 80.

The same trends are observed in the B-95.ΔA strain with the difference that both OTS show significantly higher incorporation efficiencies when more than one in-frame stop codon is suppressed (**Figure 48**). But as with other *MburPylRS* variants, the efficiency is much higher at low ncAA concentrations and surprisingly, the decrease in SCS efficiency with increasing number of in-frame stop codons is extremely low (for Sac (**1**)). A comparison of the OTSs performances with Sac (**1**) shows that the decrease for the *MburSacRS*(S379T) construct (fed with 1 mM) is 25% from the suppression of one to five in-frame stop codons. For *SmbP-MbSacRS*(S382T) the decrease is 72%. As mentioned above, the SCS efficiency of *MburSacRS*(S379T) is at the same level as the wild-type enzyme when suppressing one stop codon, but when incorporating ncAAs at multi-sites, this mutant even surpasses the wild-type performance (compare **Figure 39** for the good substrate AllockK (**32**)).

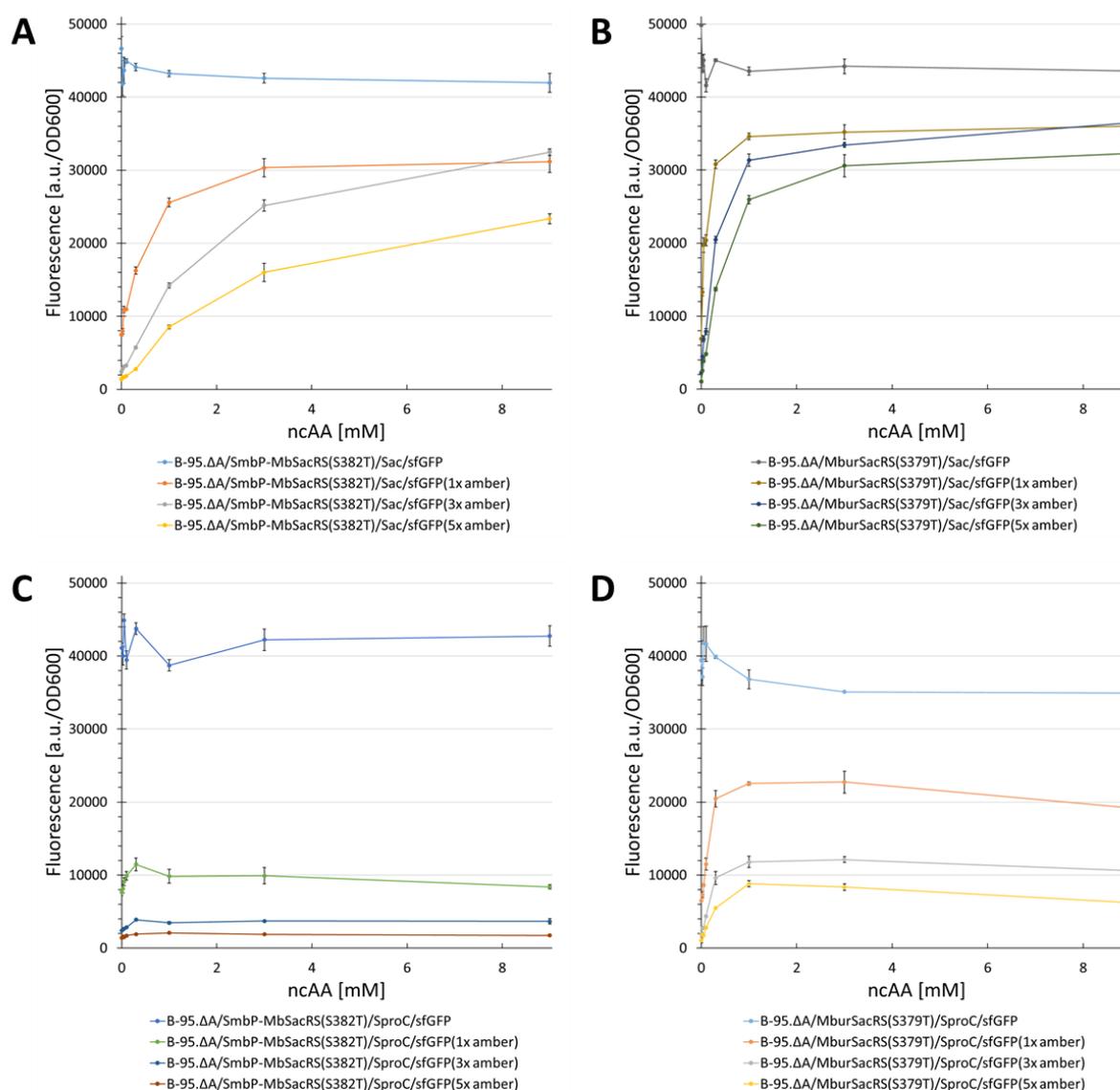


Figure 48: Concentration-dependent unnatural protein production with *SmbP-MbPylRS* (**A**, **C**) and *MburPylRS* (**B**, **D**) for different ncAA/reporter construct combinations. The protein production host was B-95.ΔA. Endpoint measurements for ncAA concentrations of 0.025, 0.05, 0.1, 0.3, 1, 3 and 9 mM. Gain 80.

A PylRS mutant with better catalytic efficiency than the wild-type has never been reported before, highlighting the extreme advantage of *M. burtonii* PylRS over all other PylRS tested here in encoding novel substrate recognitions.

2.4.5.7. Screening *MburSacRS* and the *S379T* mutant for *Sac* Analog Incorporation Activity

To evaluate the promiscuity of *Sac* (**1**) and *SproC* (**27**) incorporating *M. burtonii* variants, *MburSacRS* and *MburSacRS*(S379T) were screened for *Sac* (**1**) analogs, including the ncAAs from **chapter 2.3**. Very low incorporation activity was observed for several ncAAs. However, for amino-4-methylpent-4-enoic acid (**21**), good incorporation activity was detectable with twice the signal compared to activity with *MburPylRS*(N308M:C311W) (comparison of **Figure 43** and **Figure 130**).

2.4.5.8. Impact of different *tRNA^{Pyl}* for OTSs

The PylRS system occurs naturally in a variety of organisms.¹⁷⁹ Although all of these systems encode Pyl, there is an evolutionary divergence (**Figure 32**). The divergence is so great that further engineering led to the creation triply orthogonal PylRS OTS.³⁰⁴

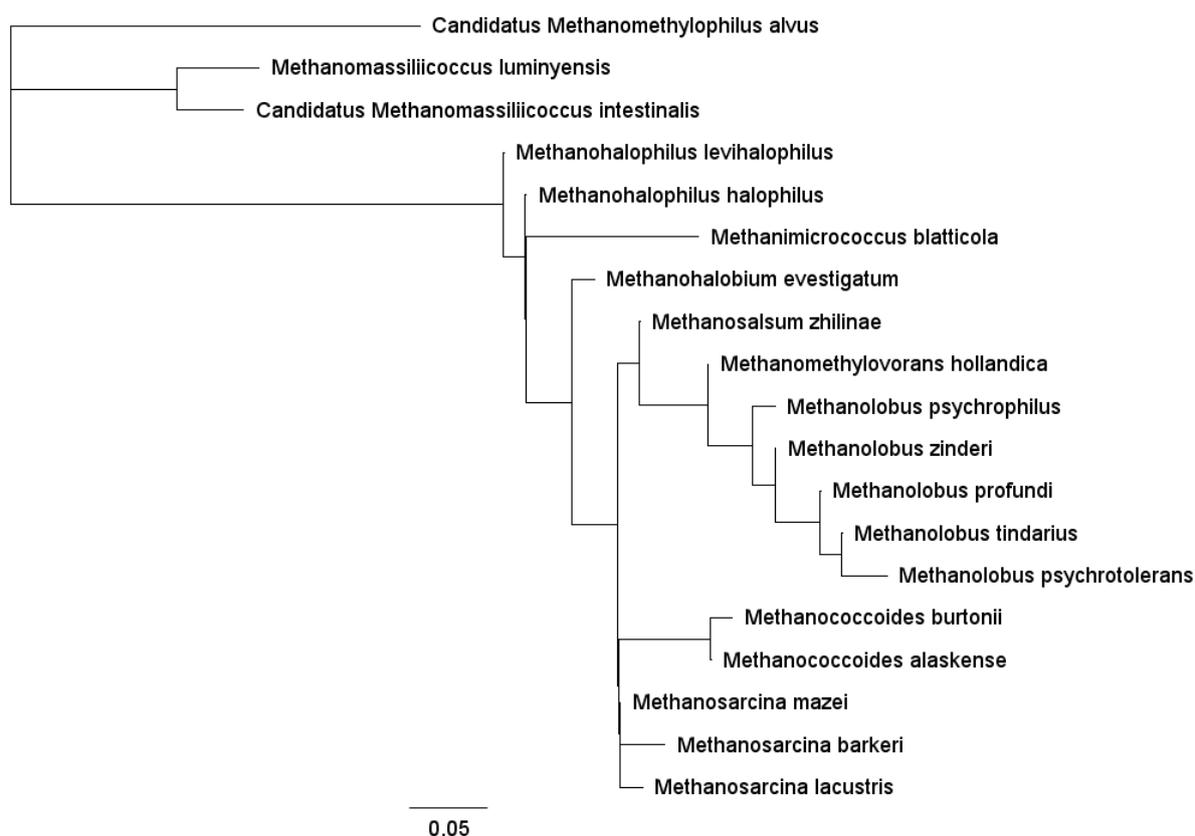


Figure 49: Approximately-maximum-likelihood phylogenetic tree based on all *Methanosarcinales* and *Methanomassiliicoccales* unique *tRNA^{Pyl}* sequences available in the NCBI and JGI database. The FastTree 2 algorithm was used for tree creation.²⁸⁵ Significance of branch support is inferred by resamples ($n = 1000$). The scale bar indicates 5% sequence divergence. Specific accession numbers can be found in the appendix.

Two of these orthogonal pairs are based on PylRS enzymes from the *Methanomassiliicoccales* genus that lack the N-terminal domain. However, given the phylogenetic analysis (**Figure 49**) and the structural information on tRNA folding prediction (**Figure 50**), it seems even possible to form mutually orthogonal PylRS OTS based on enzymes containing the N-terminal domain. For example, the structure and phylogenetic distance of the tRNA^{Pyl} of *Methanosarcina* seems divergent enough compared to *Methanohalophilus* or *Methanimicrococcus*. The most striking differences between tRNAs are found in the D-loop, the variable arm, and even in the anticodon stem (**Figure 50**).

2.4.5.8.1. Comparing tRNA from *M. mazei* and *M. burtonii*

Engineering of tRNA^{Pyl} can have a non-negligible effect on the performance of an OTS.¹⁵⁰ In **Figure 50** a selection of tRNA^{Pyl}s is shown to illustrate the diversity for this tRNA within the subgroup of N-terminal domain containing PylRS systems. Due to differences in folding, their thermal stability at 37°C is different, as indicated by the ΔG values. This is also shown graphically with a color code (red = high, green = low and blue = no binding strength).

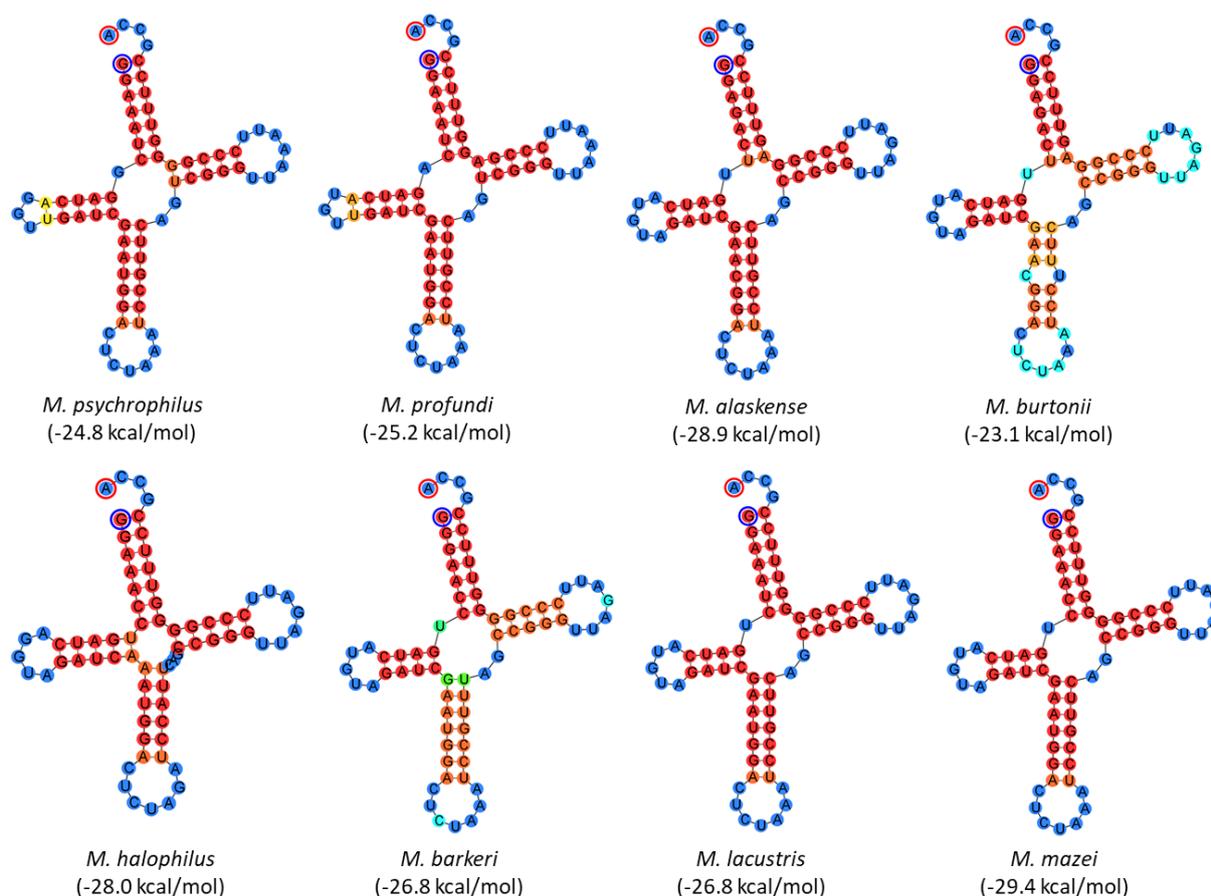


Figure 50: Selection of tRNA^{Pyl} (from **Figure 206** and **Figure 207**). ΔG values in brackets, calculated at 37 °C. The nucleotide binding probability is indicated by color; red = high, green = middle and blue = low. The fold and free energy prediction was performed with *Geneious* which uses the *ViennaRNA* Package.³¹⁵ The nucleotide with blue a circle is the 5' end and with a red circle the 3' end.

The tRNA^{Pyl}s of the genus *Methanosarcina* differ at a maximum of two sites (the *M. barkeri*) which can also be seen in **Figure 49** with close relationships. The tRNA^{Pyl} of *M. burtonii* has five changes compared to *M. mazei* and has about 30% less thermal stability. However, the overall folding is very similar, therefore no orthogonality would be expected. To determine if the stability differences affect *in vivo* OTS performance, the tRNA^{Pyl} of *M. burtonii* was replaced with the tRNA^{Pyl} from *M. mazei* to create an orthogonal hybrid pair. As can be seen in **Figure 51**, the differences in OTS performance are small, with a slightly higher efficiency of tRNA^{Pyl} from *M. burtonii* in connection with *MburPylRS*.

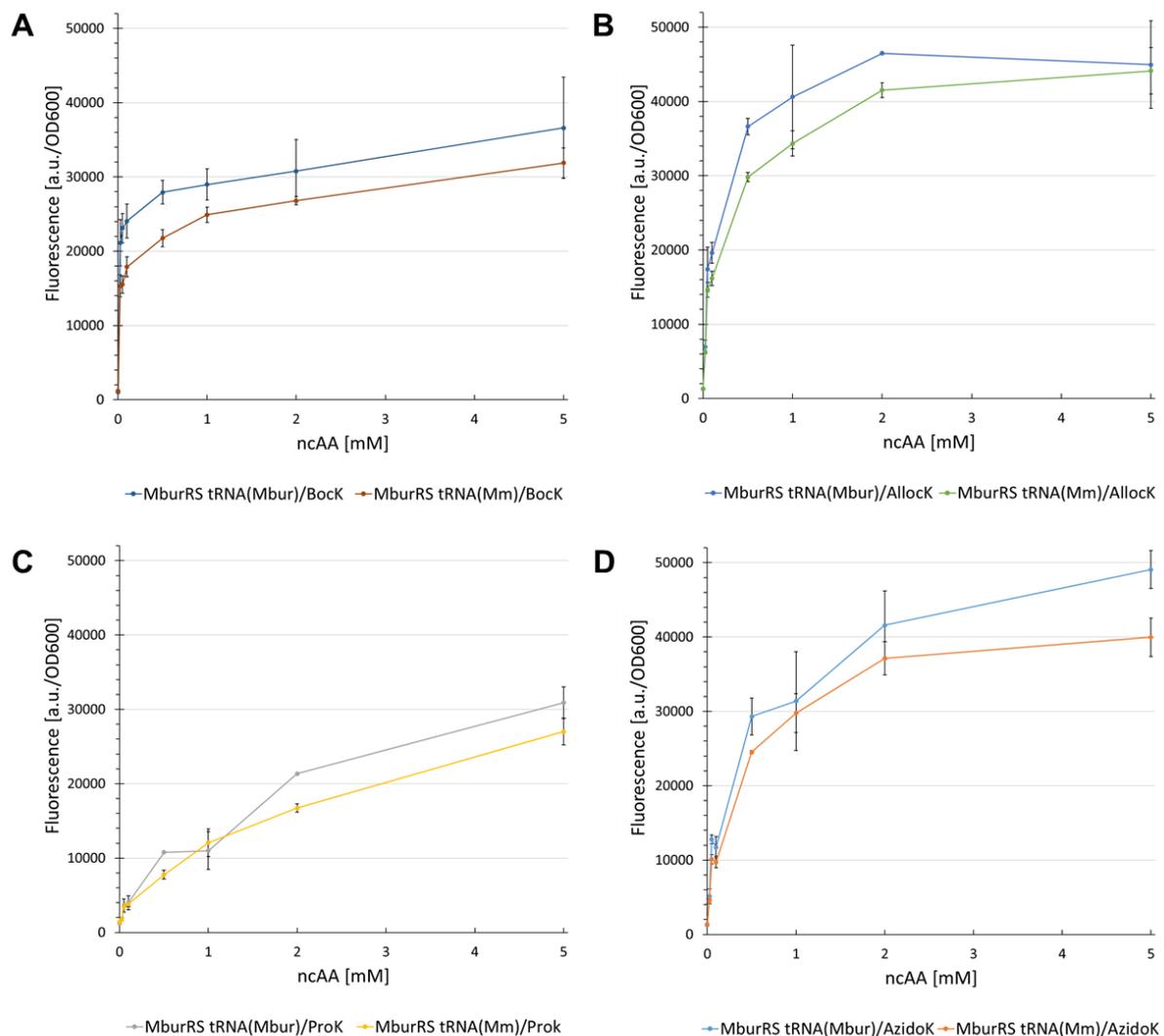


Figure 51: Concentration-dependent unnatural protein production with *MburPylRS* for different tRNA(organism abbreviation)/ncAA combinations. Using BL21(DE3) cells and sfGFP(1x amber) as reporter. Endpoint measurements for ncAA concentrations of 0.025, 0.05, 0.1, 0.3, 1, 2 and 5 mM.

2.4.6. Comparing *MjTyrRS* and *MburPylRS* performance

The *MjTyrRS* system, along with the TyrRS system from *Archaeoglobus fulgidus* (*AfTyrRS*), is one of the well-known OTS that can produce target proteins with in-frame stop codons on a wild-type level.^{1,3,188,316} To gain insights into the efficiency of *MburPylRS* compared to the efficient *MjTyrRS* OTS, two *MjTyrRS* variants were compared to two *MburPylRS* constructs in a side-by-side experiment. The *MjONB-Dopa* OTS was chosen because the high efficiency is well documented (see **chapter 2.5.3**).^{188,317} To decide on a second variant, a prescreen was performed using *MjONBYRS* and *MjPCNFRS*^{310,318} with several ncAAs (**Figure 134**). After prescreening, concentration-dependent screening was conducted that revealed O-propargyl- and O-allyl-L-tyrosine (**40** and **46**) as best performing (**Figure 135**). That showed that *MjONBYRS* was slightly more efficient than *MjPCNFRS*, but given the much lower background suppression of *MjONBYRS*, it was the preferred construct. Albeit it is known that *MjONBYRS* can produce target proteins with incorporated ONBY (**41**) up to wild-type levels, this was not the preferred combination for the assay.³ Unfortunately, ONBY (**41**) is very insoluble in water based solvents and is therefore hard to handle in small scale assays, as the ncAA usually have to be transferred directly into the final medium volume to ensure the desired amount. The assay was still performed with ONBY (**41**) from stock solutions, but resulted in inconsistent data, most likely due to high variance in ONBY (**41**) content in individual wells (**Figure 132D**). O-allyl-L-tyrosine (**46**) also seems to work better in multi-site incorporation modus. Taken together, O-allyl-L-tyrosine (**46**) was selected for direct comparison. The *MjTyrRS* variants are in a different system (pUltra) than the PylRS OTS (pTECH). The pUltra system revealed a slightly higher overall wild-type target protein production than that used for *MburPylRS*. There can be various reasons for this: among others, differences in the promotor strength of the aaRS, different copy number and/or origin of replication (pUltra has a higher copy number). All these variables represent different metabolic loads and stress responses on the cells, which translate into different yields of target proteins. This can be observed in the raw data (**Figure 132** and **Figure 133**). In order to still be able to estimate the differences in efficiency despite different OTS setups, the fluorescence signals were normalized to the corresponding wild-type target protein signal (sfGFP without an in-frame stop codon, **Figure 52**). Remarkably, *MburPylRS* performed best with 0.1 mM Bock (**31**). In addition, *MjONB-DopaRS* also has a very high efficiency with 0.1 ncAA supplied. In general, the efficiencies for suppressing a single in-frame stop codon are at the same level. Astonishingly, the *MburPylRS*-based OTS exceeded in suppressing multiple in-frame stop codons compared to the *MjTyrRS* OTS. This is observable for the fact that at 1 mM ncAA concentration the drop in incorporation efficiency for multiple ncAA incorporation is far lower for the *MburPylRS* systems than for the *MjTyrRS* systems. The *MburPylRS* efficiency for suppressing five stop codons is 2.5 to 3 times higher in comparison. The high efficiency for multi-site incorporation was observed in **chapter 2.4.5.6** but it was not

expected that the *MburPyIRS* system would even surpass the *MjTyrRS* efficiency. The higher performance of *MburPyIRS* in a RF1 deficient strain could also be an indication for a superior performance when used in connection with strains which have liberated codons.

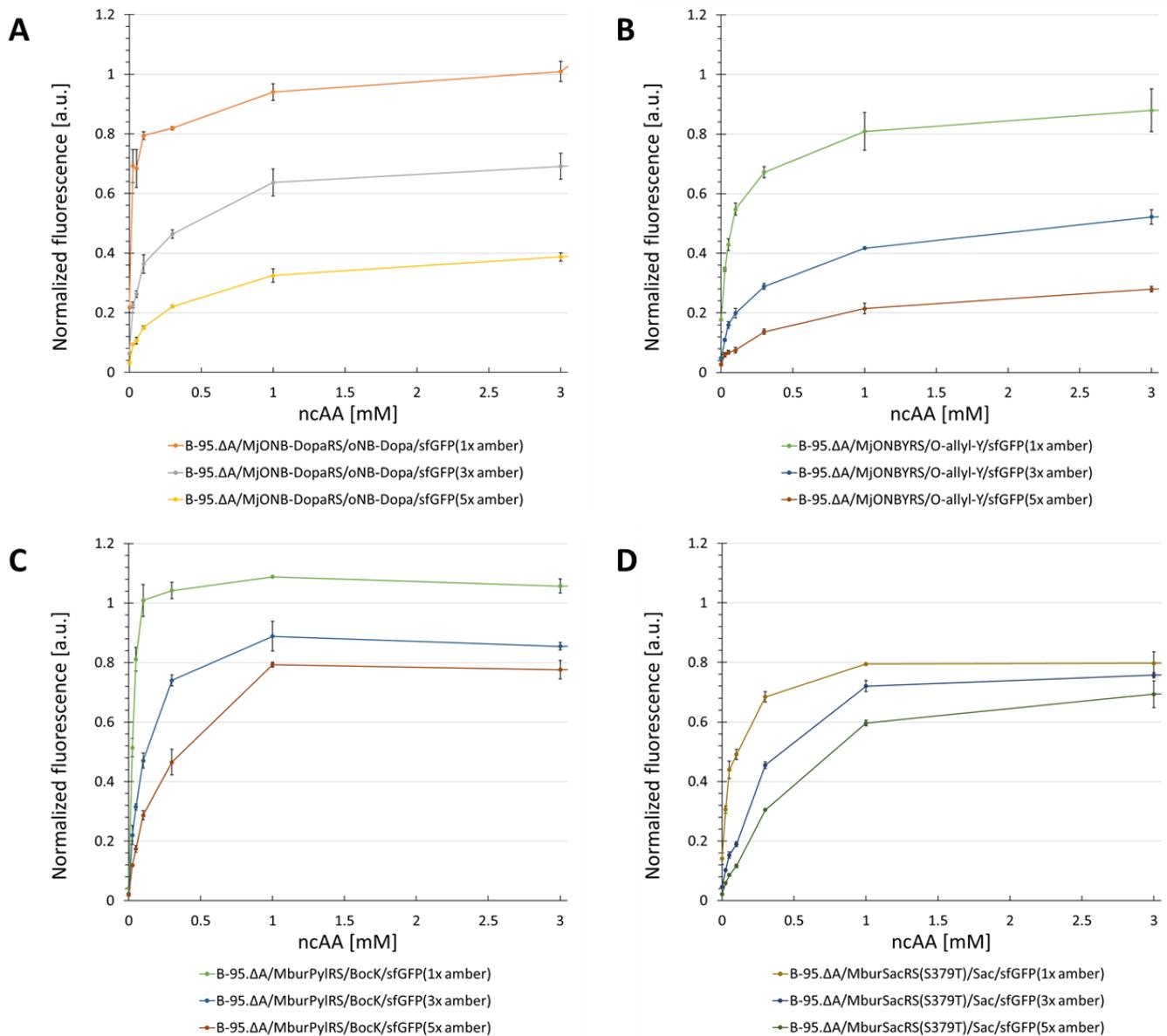


Figure 52: Concentration-dependent unnatural protein production with **A)** *MjONB-DopaRS*, **B)** *MjONBYRS*, **C)** *MburPyIRS* and **D)** *MburSacRS*. The protein production host was *Escherichia coli* B95.ΔA. Endpoint measurements for ncAA concentrations of 0.025, 0.05, 0.1, 0.3, 1 and 3 mM. The fluorescence values were normalized with the values for the wild-type sfGFP reporter constructs (without an in-frame stop codon). A value of 1 represents the same level of protein production as the wild-type.

2.4.7. Analytics of ncAA Incorporation

To test how the high efficiency of *MburPylRS* and its mutants affects the production of the target protein, protein production was scaled up to confirm the results of the small-scale fluorescence assays using 96-well plates. Target proteins were purified using Ni-NTA immobilized metal ion affinity chromatography. The protein yields are in good agreement with the trends observed in the fluorescence experiments, except for the SproC (27) expression with *MburSacRS*(S379T). Here the target protein yields were far above what was expected from **Figure 48**. The mass spectra of the intact reporter proteins were determined by ESI-MS (**Figure 171-Figure 199**). It can be observed that the higher amount of obtained target protein with the *MburPylRS* system results in smaller mass shifts. This is to be expected since more target protein will be enriched in the sample, resulting in smaller errors in MS data acquisition.

Table 8: Setup of protein production platform with calculated and observed molecular weights of reporter proteins corresponding to a) SUMO-sfGFP(1x amber), b) SUMO-sfGFP(3x amber), c) SUMO-sfGFP(5x amber), d) SUMO-sfGFP(wild-type). The masses were determined by ESI-MS of the intact proteins.

ncAA	[mM]	<i>E. coli</i> strains ¹	PylRS construct	reporter	calculated mass [Da]	found mass [Da]	Δ mass [Da]	protein yield [mg·L ⁻¹] ²
31	0.05	BL21	<i>MmPylRS</i>	a ³	39096.0	-	-	18.5
31	0.05	BL21	<i>MburPylRS</i>	a ³	39096.0	-	-	20.0
31	1	BL21	<i>MmPylRS</i>	a	39096.0	39094	2	97.2
31	1	BL21	<i>MburPylRS</i>	a	39096.0	39094	2	123.6
34	0.05	BL21	<i>MmPylRS</i>	a ³	39109.0	-	-	4.3
34	0.05	BL21	<i>MburPylRS</i>	a ³	39109.0	-	-	11.2
34	1	BL21	<i>MmPylRS</i>	a	39109.0	39105	4	33.0
34	1	BL21	<i>MburPylRS</i>	a	39109.0	39106	3	84.9
40	3	BL21	<i>MmPylRS</i> (N346A:C348A)	a	39069.0	39065	4	11.9
40	3	BL21	<i>MburPylRS</i> (N308A:C310A)	a	39069.0	39065	4	19.9
45	3	BL21	<i>MmPylRS</i> (N346A:C348A)	a	39071.0	39068	3	26.4
45	3	BL21	<i>MburPylRS</i> (N308A:C310A)	a	39071.0	39068	3	52.0
47	1	BL21	<i>MmPylRS</i> (N346A:C348A)	a	39098.9	39095	3.9	4.6
47	1	BL21	<i>MburPylRS</i> (N308A:C310A)	a	39098.9	39095	3.9	15.1
39	3	BL21	<i>MmPylRS</i> (N346A:C348A)	a	39087.0	39084	3	30.2
39	3	BL21	<i>MburPylRS</i> (N308A:C310A)	a	39087.0	39084	3	51.0
1	0.3	B-95.ΔA	<i>SmbP-MbSacRS</i> (S382T)	d	39023.9	39023	0.9	89.8
1	0.3	B-95.ΔA	<i>SmbP-MbSacRS</i> (S382T)	a	39010.9	39010	0.9	42.2
1	0.3	B-95.ΔA	<i>SmbP-MbSacRS</i> (S382T)	b	39055.2	39053	2.2	13.8
1	0.3	B-95.ΔA	<i>SmbP-MbSacRS</i> (S382T)	c	39097.5	39095	2.5	6.7
1	0.3	B-95.ΔA	<i>MburSacRS</i> (S379T)	d	39023.9	39023	0.9	113.2
1	0.3	B-95.ΔA	<i>MburSacRS</i> (S379T)	a	39010.9	39011	0.1	86.2
1	0.3	B-95.ΔA	<i>MburSacRS</i> (S379T)	b	39055.2	39054	1.2	69.1
1	0.3	B-95.ΔA	<i>MburSacRS</i> (S379T)	c	39097.5	39096	1.5	38.6
27	1	B-95.ΔA	<i>SmbP-MbSacRS</i> (S382T)	a	39008.9	39009	0.1	60.1
27	1	B-95.ΔA	<i>SmbP-MbSacRS</i> (S382T)	b	39049.0	39047	2.0	19.6
27	1	B-95.ΔA	<i>SmbP-MbSacRS</i> (S382T)	c	39087.2	39084	3.2	10.9
27	1	B-95.ΔA	<i>MburSacRS</i> (S379T)	a	39008.9	39008	0.9	104.9
27	1	B-95.ΔA	<i>MburSacRS</i> (S379T)	b	39049.0	39048	1.0	59.3
27	1	B-95.ΔA	<i>MburSacRS</i> (S379T)	c	39087.2	39086	1.2	41.0

¹all DE3, ²yield per liter of cell culture, ³was not analyzed by ESI-MS

2.5. Side Project: Production of Genetically Encoded Bioinspired Smart Wet Adhesive Materials

A long-term goal in the Budisa group is to create biomaterial-based wet-adhesion agents that can be used for *in vivo* tissue or bone wound healing, with the help of genetic code expansion tools. The strategy followed so far has been to try to mimic the underwater adhesion of mussels. Mussels can adhere on surfaces which are generally difficult targets (e.g., wet, corroded, slimy) for conventional adhesion agents (e.g., cyanoacrylates; they are esters of cyanoacrylic acid and are generally known as strong, fast-acting dry-adhesives). The underwater adhesion ability of mussels is a complex phenomenon in which a mussel places an insulated reaction chamber (byssus) onto the target surface, where a mixture of adhesion proteins is produced. These mixtures then interact with the surface and create the adhesion phenomenon.³¹⁹ The most striking difference in the composition of mussel adhesion proteins compared to normal proteins is the high content of 3,4-dihydroxyphenyl-L-alanine (Dopa).³¹⁹ Dopa has been previously shown to be incorporated in a residue specific manner into one of the main mussel adhesion proteins (mussel adhesion protein 3, mfp-3).³²⁰ Unfortunately, Dopa tends to convert to its quinone species under normal (oxidative) protein production and purification conditions. Therefore, Dopa derivatives with a photo-cleavable protecting group have been developed (**51**, **53**).³²¹ The photo-cleavable protecting group also provides timely control over a potential adhesion mechanism. Using a library-based protein engineering strategy, a *Mj*TyrRS OTS was designed to incorporate *meta*-(*ortho*-(2-Nitrobenzyl))-3,4-dihydroxyphenylalanine (*m*-oNB-Dopa, **51**) into proteins.¹⁸⁸ However, this strategy is not free of serious flaws. For example, the production of mussel proteins known to be most abundant in surface interaction (mfp-3, mfp-5) is not feasible in *E. coli* (very low yields), most likely due to their low solubility and/or SCS efficiency.

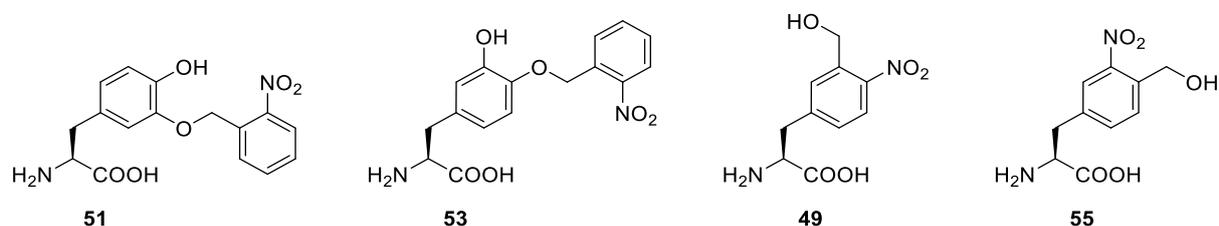


Figure 53: Non-canonical amino acids used in this chapter. *Meta*-(*ortho*-(2-nitrobenzyl))-3,4-dihydroxyphenylalanine (*m*-oNB-Dopa, **51**), *para*-(*ortho*-(2-nitrobenzyl))-3,4-dihydroxyphenylalanine (*p*-oNB-Dopa, **53**), *p*-oNB-alanine (**54**), *m*-oNB-alanine (**55**).

It makes sense that mussels produce adhesion proteins that are insoluble in water-based solutions, otherwise contradicting the purpose. Moreover, the dependence of SCS efficiency based on protein scaffolds is known to be a context effect.^{1,151,322} The low yield of protein production is an obstacle because the amounts of protein obtained are too small for milligram-scale adhesion tests, which are usually needed to infer whether the material can adhere to

skin or bone tissue. This usually requires several 100 mg. Therefore, an alternative adhesion protein scaffold was sought for that combines high production yield (also in terms of high SCS efficiency) with potential adhesion properties (see **chapter 2.5.1**).

2.5.1. Choice, Design and Creation of Protein scaffolds

The first choice as potential new scaffold that meet the necessary criteria for *in vivo* applications was Elastin-Like Polypeptides (ELPs). ELPs are a class of self-assembling peptides derived from the repetitive peptide sequences which are derived from the water-soluble part of the tropoelastin protein.^{323–325} Elastin is a protein of the extracellular matrix. In multicellular organisms it is part of the connective, skin, lung, and tendon tissue. Due to its high biological similarity to native tissue, it is very well tolerated when used *in vivo* and results in a very mild immune response.³²⁶ In addition, ELPs have a lower critical solution temperature (LCST) which is very useful for developing the adhesion agents. The LCST depends on their sequence length and ionic strength of the solution they are in. Since the length of the ELPs can be precisely genetically encoded, the LCST can be easily tuned to the desired temperature. The long-term goal is to set the right parameters to obtain soluble protein in a biotechnological production strain (e.g., *E. coli*) and final product, but after application to the desired body part it should reach the coacervation point. So, the coacervation temperature should be somewhat below human body temperature. The combined properties of adhesion and forming a cohesive matrix is essential for adhesion materials to form a robust adhesion agent and ELPs seem like a good bioinspired candidate, given all the constraints in genetic code expansion.

ELPs consist of a repeating pentapeptide sequence (VPGXG)_n (**Figure 54**). Surprisingly, when the X is selected as the stop codon, there appear to be excellent context effects compared to other protein scaffolds (see **chapter 2.4.6**), allowing for the suppression of up to 60 stop codons (30 with a reasonable protein yield), making it a prime candidate for the incorporation of ncAAs at multiple sites using the SCS method.

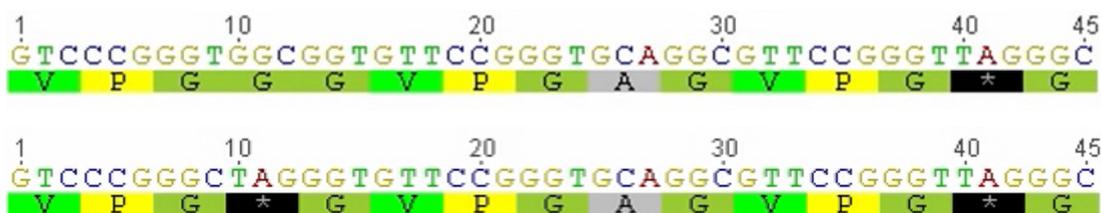


Figure 54: The two ELP monomers used in this study. Top: Three repeats of the VPGXG sequence, where X₃ being an amber stop-codon. Bottom: Three repeats of the VPGXG sequence, where X₁ and X₃ are an amber stop-codon, referred to as alternating construct.

The ELPs were designed to have a GGC/TAG as X₁, GCA as X₂ and a TAG as X₃ codons. These are the core motifs for all ELPs in this work and are adapted from the work of Amiram

and colleagues.³¹⁷ To create the ELP(5x amber)/ELP(8x amber) constructs, the core motifs were repeated five times and produced by gene synthesis (**Figure 55**).

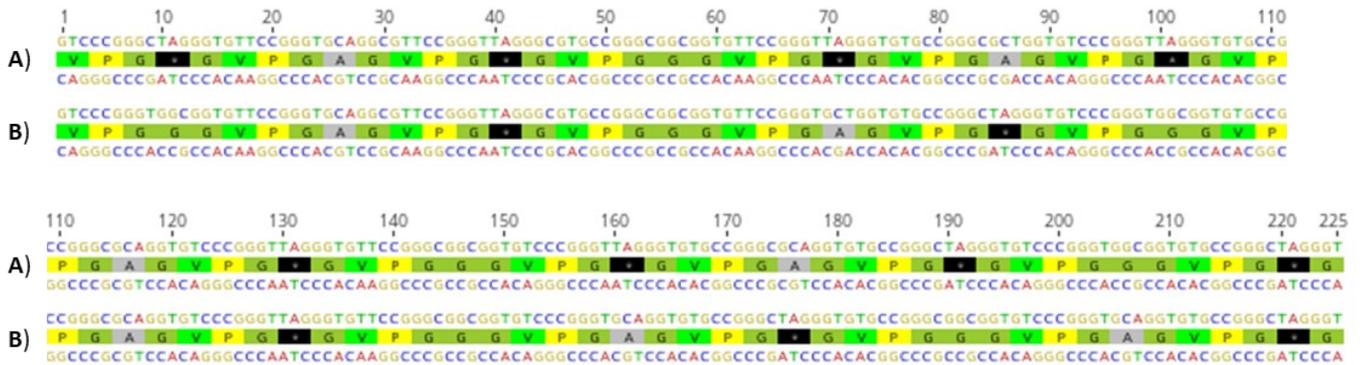


Figure 55: **A**) Alternating amber stop-codon sequence of ELP(8x amber). **B**) Sequence of the ELP(5x amber) construct.

All other constructs were made from these two constructs using the well-established method of recursive directional ligation (RDL, see **Figure 56**).³²⁷ The alternating stop-codon constructs were made in cooperation with Anna Dziegielewska (Technische Universität Berlin, Germany). The ELP scaffold with the alternating amber stop-codon sequence was designed to study the SCS efficiency as a function of ELP sequence length.

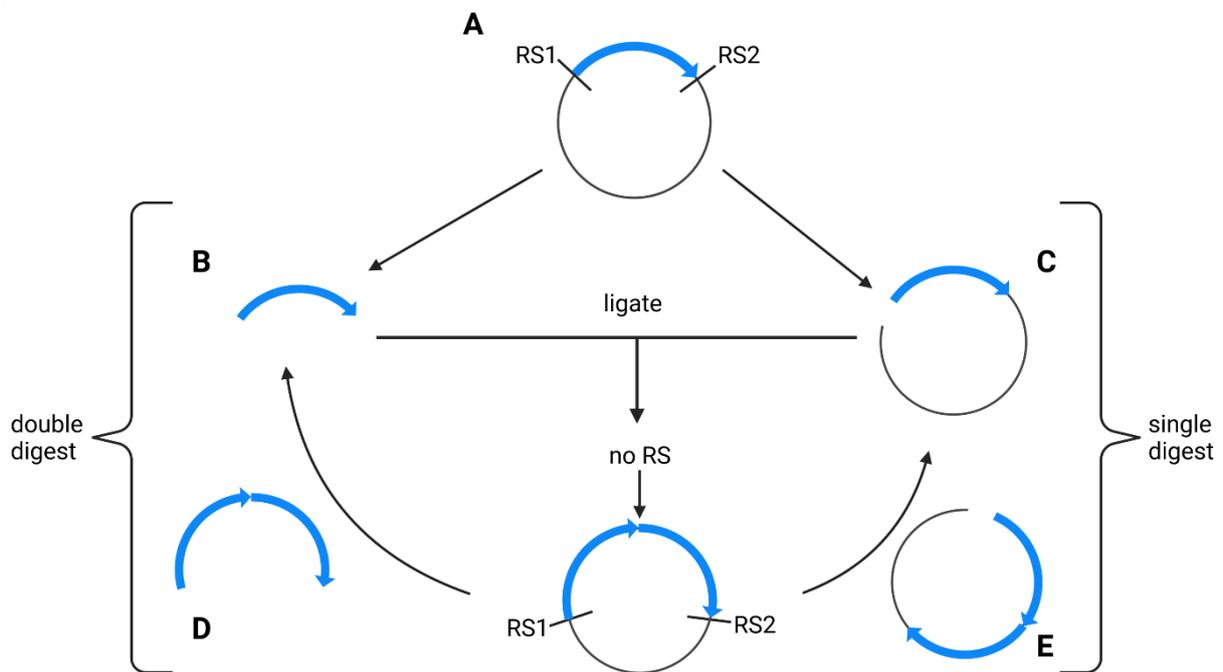


Figure 56: ELP construct design strategy via the recursive directional ligation (RDL) approach.³²⁷ Restriction sites (RS) are: RS1 = PflMI, RS2 = BglI. **A**) starting construct, **B**) ELP gene, **C**) construct digested with at one RS, **D**) elongated ELP gene, **E**) digested construct with elongated ELP gene.

After generation of the ELP variants with a cloning plasmid (pSB1C3), the ELP genes were inserted into an expression plasmid (pET28a) upstream of an sfGFP gene, resulting in the corresponding reporter constructs (**Figure 57**). Correct insertion of the ELP sequences was

evaluated by restriction digestions (**Figure 139**) and, for the shorter ELPs of up to 2000 bp length, with sanger sequencing. Protein production of these constructs was then analyzed via fluorescence assays (see **chapter 2.5.3**).

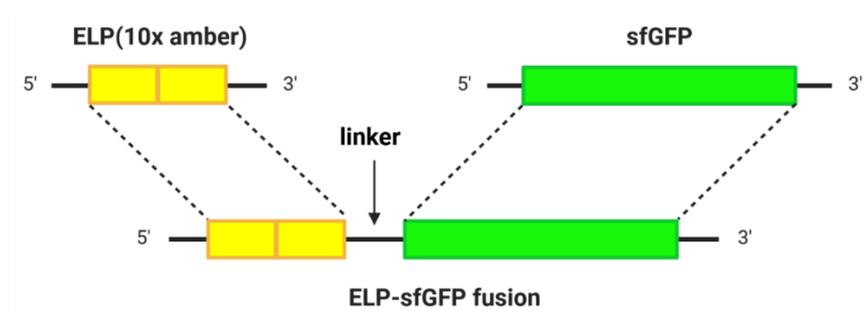


Figure 57: Creating an ELP-sfGFP reporter construct using ELP(10x amber) as an example.

2.5.2. Optimizing the *Mj*oNB-DopaRS

The *Mj*TyrRS recognizes the para OH group of Tyrosine with two critical residues Y32 and D158.³²⁸ Therefore, it has always been assumed that this recognition must be switched off. These two residues are usually mutated to smaller AAs. An example is the first aaRS that encoded ONBY (**41**).³²⁹ This was also done for the Budisa group oNB-Dopa library, whose original goal was to incorporate *p*-oNB-Dopa (**53**).¹⁸⁸ Curiously, the *Mj*TyrRS found recognized *m*-oNB-Dopa (**51**) and not *p*-oNB-Dopa (**53**). Thus, the question arose whether the Y32A and D158A mutations were necessary to encode *m*-oNB-Dopa (**51**), as there could be a potential interaction with the OH group in para position. This was not verified in the original publication.³³⁰ Reversal of these mutations, one at a time and in combination, shows that the A158D mutation is necessary to maintain orthogonality but the Y32A mutation is not (**Figure 58A**). Interestingly, background suppression appears to be 15-50% lower for the A32Y construct than for the original construct (**Figure 58B**).

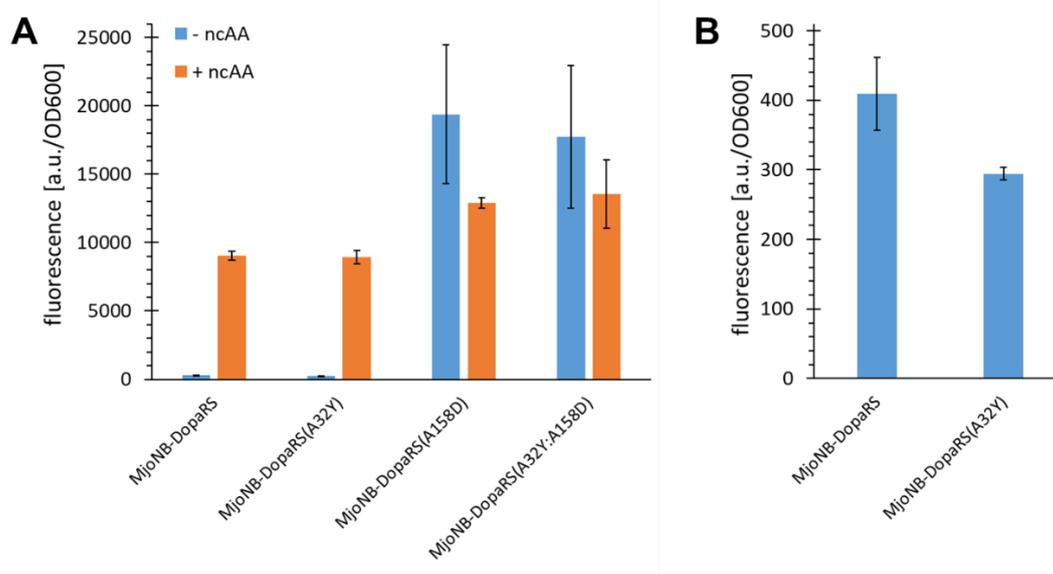


Figure 58: Examination of oNB-DOPA OTS efficiency by different mutations in the binding pocket of MjTyrRS. **A**) Comparison of OTS efficiency of the constructs indicated in the text. Ribosomal incorporation with ncAA (+ ncAA = 1 mM *m*-oNB-Dopa(**51**)) and controls without ncAA supplementation. **B**) Background suppression. Intact cell fluorescence of *E. coli* BL21(DE3), endpoint measurements after 24 h of incubation. Data (incl. standard deviation) represent the mean of three biological replicates.

2.5.3. Multisite Incorporation of (*m*-oNB-Dopa) into Elastin-Like-Polypeptide-sfGFP Fusion Proteins

After determining the most efficient *m*-oNB-Dopa (**51**) incorporating MjTyrRS variant, the preparation of ELP scaffolds with up to 60 amber stop-codons was attempted (**Figure 59**). To facilitate readout, the ELP variants were fused to the N-terminus of sfGFP as explained in **chapter 2.5.1**. **Figure 59** shows that high yields can be expected for ELP constructs containing between 5 and 30 stop-codons. Careful inspection of the constructs with lower fluorescence signals revealed that constructs containing up to 60 stop-codon should be producible in sufficient quantities for ESI-MS measurements (**Figure 59B**). The highest number of in-frame suppressed stop codons to date was 30.³¹⁷ As mentioned earlier (**chapter 2.4.4**) a fluorescence signal twice that of the background indicates robust incorporation. It is noteworthy to mention that it is not known why such a great number of stop codons can be suppressed in the ELP scaffold since for other scaffolds (sfGFP) the protein yield decreases dramatically already at five in-frame stop codons (see **chapter 2.4.6**). This should be investigated to eventually reveal superior context effects that can be used for SCS.

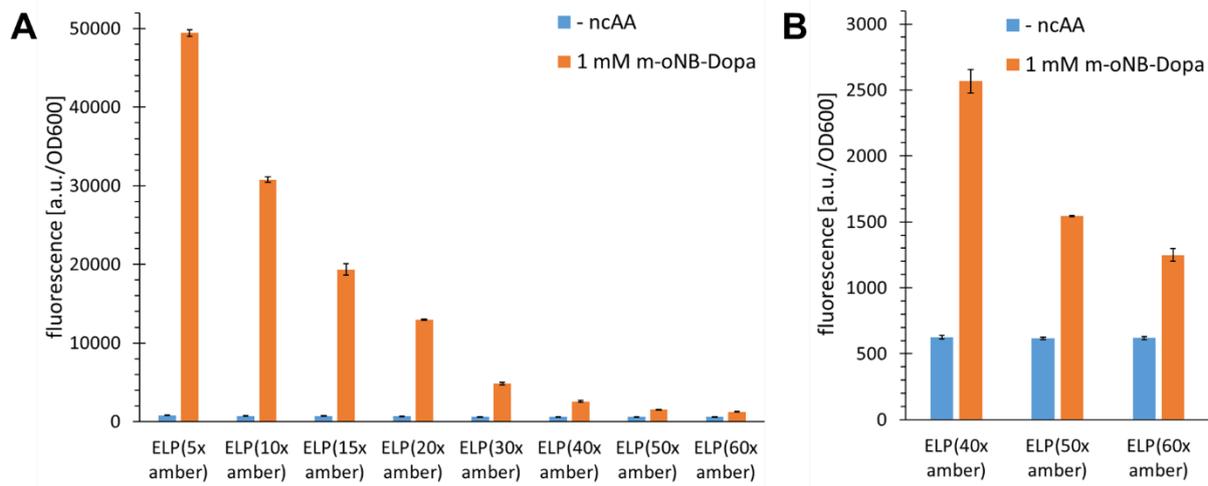


Figure 59: Comparison of OTS efficiency of the constructs indicated in the text. **A**) Ribosomal incorporation with ncAA (1 mM *m-o*NBDopa(**51**)) in all constructs with varying numbers (5x-60x) of in-frame amber stop codons including controls without ncAA supplementation (-ncAA). **B**) A portion of the diagram shown left with focus on 40x, 50x and 60x in-frame stop codon suppression. Intact cell fluorescence of B-95.ΔA, endpoint measurements after 24 h of incubation. Data (incl. standard deviation) represent the mean of three biological replicates.

To clarify the relationship between the number of in-frame stop codons and the length of the ELP sequence, the constructs with alternating stop-codons were also inserted upstream of a sfGFP gene and the production of these reporter proteins was analyzed using the fluorescence readout. This showed that the amount of produced protein was correlated only with the number of in-frame stop-codons in the ELP sequence and that the distance between these codons had little to no effect (**Figure 60**). This is best observed when comparing ELP(16x amber)_{alt} with ELP(15x amber)_{nor.}, ELP(24x amber)_{alt.} with ELP(20x amber)_{nor.} and ELP(32x amber)_{alt.} with ELP(30x amber)_{nor.} These pairwise comparisons clearly show that constructs with a similar number of stop codons have a similar fluorescence signal, indicating similar protein yields.

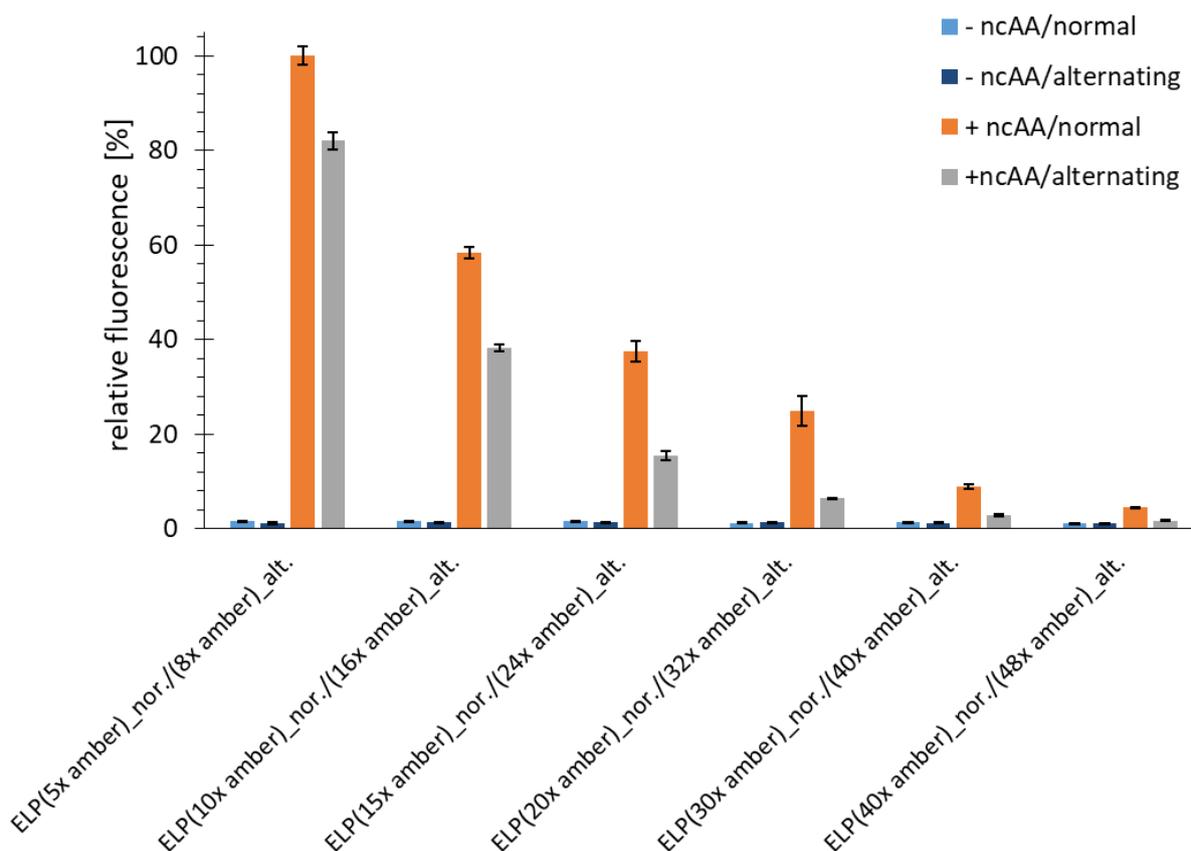


Figure 60: Comparison of OTS efficiency of the indicated constructs. Ribosomal incorporation of *m*-oNB-Dopa (**51**) (1 mM). Alt. = alternating amber stop-codons; normal = one amber stop-codon per three VPGXG repeats. Intact cell fluorescence of B-95.ΔA, endpoint measurements after 24 h of incubation. Data (incl. standard deviation) represent the mean of three biological replicates.

2.5.4. Analytics of ELP(5x oNB-Dopa)-sfGFP Fusion Protein

For ESI-MS analysis, the construct was prepared with five stop-codons in a flask cultivation (50 mL). **Figure 61** shows the successful incorporation of five *m*-oNB-Dopa (**51**). It is well known that, particularly in bacterial hosts, aromatic nitro groups can be enzymatically reduced to hydroxylamino and/or amino derivatives.^{3,331–333} This phenomenon was also observed with the *m*-oNB-Dopa (**51**) incorporating OTS as well.¹⁸⁸ Notably, a substantial fraction (circa 50% of the total protein, estimated from spectrum peaks) of the protein produced contains one to five reduced nitro groups. For clarity, the measured mass-shifts are shown in **Figure 61** and explained in **Table 9**.

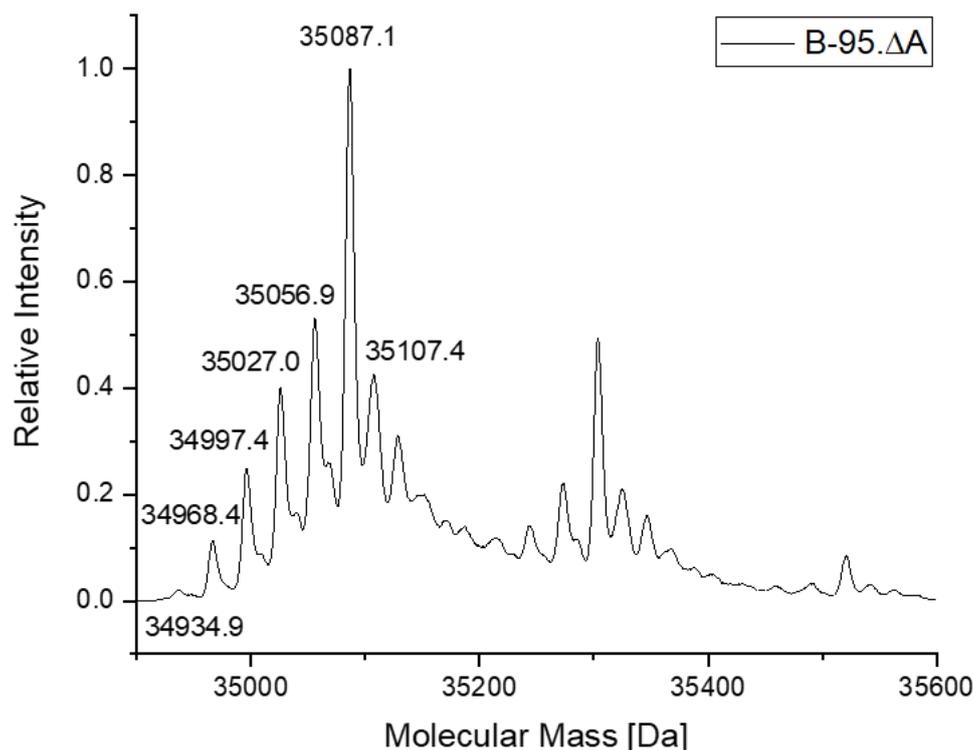


Figure 61: ESI-MS of ELP(5x51)-sfGFP. The mass spectrometric profiles of ELP with 5x *m*-oNB-dopa-sfGFP show a chemical change (the reduction of the nitro group). The parameters for the mass shifts (i.e. comparisons of expected/calculated and measured masses) are given in **Table 9** and the corresponding chemistry is explained in the text.

From **Table 9** it can be seen that the predicted mass peaks for possible nitroreduction agree very well with those found. As expected, the differences (Δ mass) become larger for the less abundant protein species. To address the observed problem of nitro group reduction, a strategy to delete ('knock-out') the genes associated with high nitroreductase activity in *E. coli* was developed for the B-95. Δ A strain (**chapter 2.5.5**).

Table 9: ESI-MS analytics of ELP(5x α NB-Dopa)-sfGFP-His₆

Species*	Calculated mass [Da]	Found mass [Da]	Δ mass [Da]
mat., unreduced	35087.1	35087.1	0
mat., 1x reduced	35057.1	35056.9	0.2
mat., 2x reduced	35027.1	35027.0	0.1
mat., 3x reduced	34997.1	34997.4	0.3
mat., 4x reduced	34967.1	34968.4	1.3
mat., 5x reduced	34937.1	34934.9	2.2
Non-maturated, unreduced	35107.1	35107.4	0.3

*all without the starting Met, mat. = sfGFP fluorophore is maturated

2.5.5. Engineering a bacterial strain with reduced nitroreductase activity

In a routine expression setup, the *m*-oNB-Dopa (**51**) containing proteins possess reduced nitro groups as shown in **chapter 2.5.4**. The nitro group of **51** is reduced to the corresponding amine, which removes the photocleavable ability of the protecting group. In general, at least eleven known genes are known to have nitroreductase activity.^{334,335} The *E. coli* NfsA and NfsB have been found to have the highest ability to reduce nitro groups.^{334,336,337} Unfortunately, deletion of *nfsA* and *nfsB* has been shown to be insufficient to prevent nitroreduction of ncAAs in *E. coli*³³³, although biotransformation of small molecules can be prevented with these deletions.³³⁷ Therefore, a strategy was developed to delete ('knock-out') six genes known to have adequate nitroreductase activity by the corresponding enzymes (see **Table 10**).

Table 10: The six genes from the *E. coli* genome with known nitroreductase activity deleted from the genome of the expression host used in this study.

Gene	Description	Nitroreductase activity ¹	Reference
<i>nfsA</i>	Major Oxygen-insensitive nitroreductase.	high	Zenno et al. ³³⁶ Copp et al. ³³⁴
<i>nfsB</i>	Minor Oxygen-insensitive nitroreductase.	middle	Zenno et al. ³³⁸ Copp et al. ³³⁴
<i>azoR</i>	NADH-azoreductase which can reduce azo dyes. Nitroreductase activity was observed with ortholog from <i>Rhodobacter sphaeroides</i> .	middle/low	Mercier et al. ³³⁵ ; Copp et al. ³³⁴
<i>ydja</i>	One of the smallest nitroreductases from <i>E. coli</i> .	low	Choi et al. ³³⁹ Copp et al. ³³⁴
<i>nema</i>	Flavin-dependent oxidoreductases related to the old yellow enzyme family.	low	Williams et al. ³⁴⁰ Copp et al. ³³⁴
<i>rutE</i>	Reduces malonic semialdehyde to 3-hydroxypropionic acid. RutG appears to be a uracil transporter.	low	Kim et al. ³⁴¹ Copp et al. ³³⁴

¹rough estimation based on the substrate activities in the references

Gene deletions were performed using an updated Lambda Red recombineering system (pSIJ8 plasmid).³⁴² This method is based on the well-established DATSENKO and WANNER method, with the difference that it is a single plasmid system.³⁴³ In contrast to the DATSENKO and WANNER method, the antibiotic resistance cassette was not amplified from a template plasmid but from genomic DNA of *E. coli* clones from the Keio Single Gene Knockout Collection.³⁴⁴ These single-knockout clones were generated containing the antibiotic resistance cassette with flanking FRT sites, making the PCR amplification product from these clones compatible with the DATSENKO and WANNER approach. The workflow is shown in **Figure 62** and has also been described in different variants before.^{345–348} Such a hybrid approach had the advantage that the length of homology region could be freely chosen and was not restricted by primer length constraints (synthesis and PCR efficiency limitations). This is advantageous because the efficiency of recombineering correlates with the length of the homology region (the longer the region, the higher the efficiency).^{349–351} For one or a few deletions, efficiency works well with the standard homology region of 35-50 bp. However, for each deleted gene, an FRT scar remains in the genome, which in turn increases the likelihood that the antibiotic cassette will

be inserted into this already existing FRT site, increasing the number of false positive deletions. The longer homology regions mitigate this effect by shifting the distribution of false-positive to true-positive deletions towards the true positives, as the efficiency of insertion of the antibiotic resistance cassette is increased. Another advantage is that the Keio knockouts were carefully designed to mitigate the negative effects that can occur when a deleted gene is part of an operon. The FRT-scar is in-frame and the last seven bps of the deleted gene are retained, decreasing the likelihood of adverse effects on downstream gene translation.

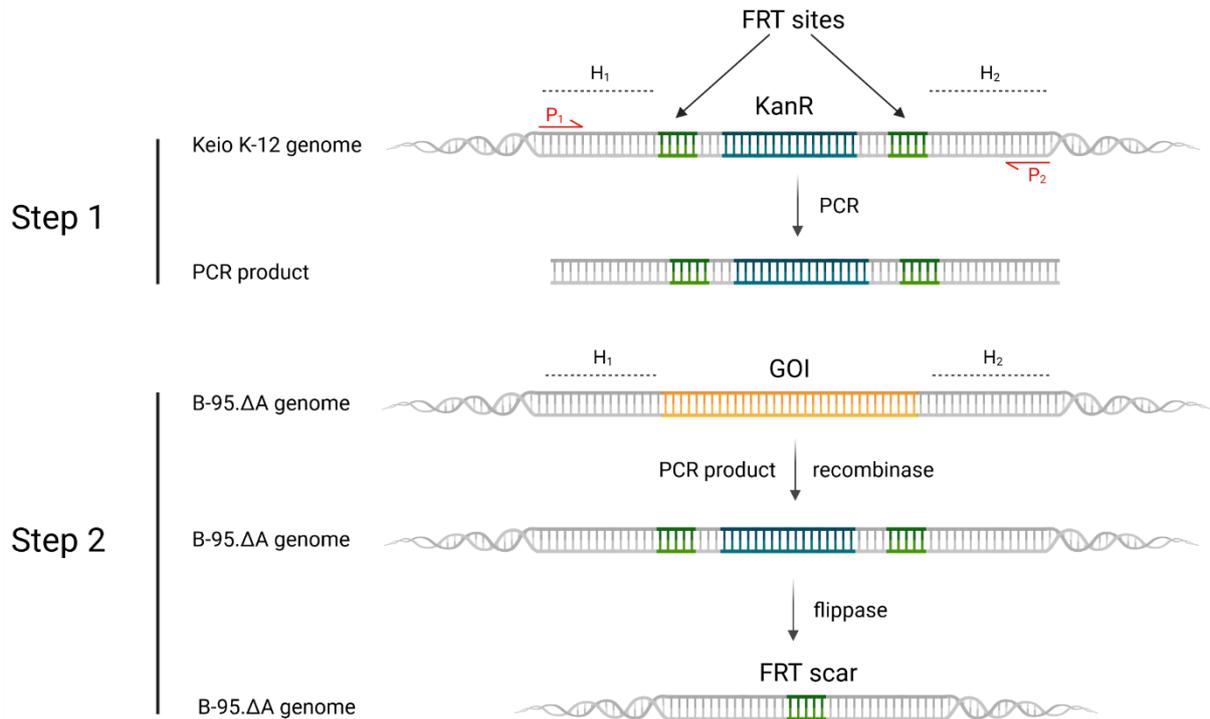


Figure 62: Workflow for gene deletions of the B-95.ΔA strain. H_{1,2} = homology region 1 or 2; P_{1,2} = Primer 1 or 2; KanR = Kanamycin resistance cassette; GOI = gene of interest.

Therefore, lengths of 150-200 bp were chosen, which is hardly achievable for amplification from a template plasmid because the length of the homology region affects the amplification efficiency of the antibiotic resistance cassette from the template plasmid. Moreover, amplification of the cassette from the genome of the Keio clone is possible with only desalted primers (as used in this work). In contrast, HPLC purified primers are essential for amplification of the cassette from the template. Combining the advantages of shorter and less purified primers, the method used here is about an order of magnitude cheaper per gene deletion than the DATSENKO and WANNER method, assuming the Keio collection is already available at no additional cost. If the creation of a longer homology region is necessary due to low deletion efficiency, the method presented here is at least one order of magnitude cheaper and less laborious.

After deletion of the six genes from **Table 10**, *E. coli* strain B-95.ΔA(ΔNfsA::FRT; ΔNfsB::FRT; ΔAzoR::FRT; ΔYdja::FRT; ΔNemA::FRT; ΔRutE::FRT), hereafter referred to as NK53 was obtained. After each round of deletion, the clones were probed with primers binding upstream and downstream of the homology region and also with the corresponding primers that bind within the antibiotic resistance cassette. Sequencing of the PCR products was also performed. This confirmed that the insertion was in the correct region of the genome. Correct removal of the antibiotic resistance cassette was verified only with the primers binding upstream and downstream. After deletion of all 6 genes, NK53 was rescreened for deletions to ensure that all deletions were present, using the initial strain as negative control (**Figure 63**). The same was also done with the primers for upstream and downstream. In general, the PCR products from every step were sequenced to verify the correctness of the sequence. This is shown by the example of the *nfsB* deletion in **Figure 217**. **Figure 63** clearly shows that the PCR product for all targeted genes is shorter than the wild-type PCR product, indicating successful deletion of the genes.

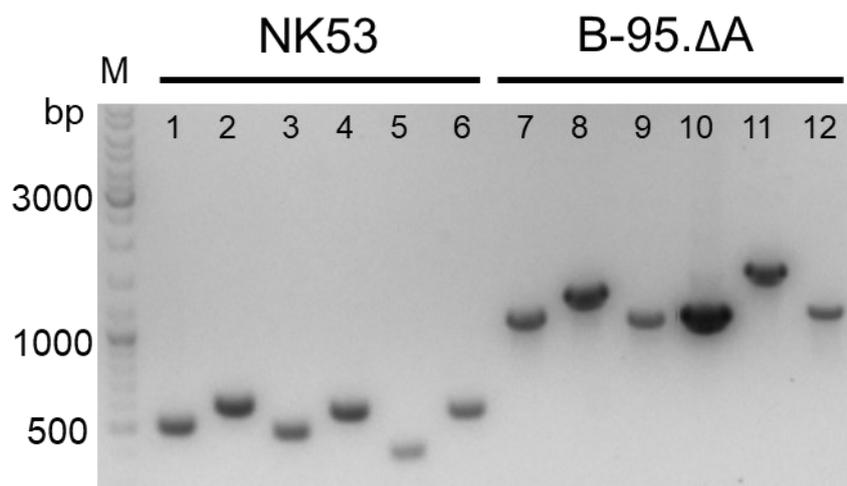


Figure 63: Verification of deleted genes via agarose gel electrophoresis. Comparison of *E. coli* NK53 and *E. coli* B-95.ΔA. 1) ΔNfsA::FRT; 2) ΔNfsB::FRT; 3) ΔAzoR::FRT; 4) ΔYdja::FRT; 5) ΔNemA::FRT; 6) ΔRutE::FRT; 7) NfsA; 8) NfsB; 9) AzoR; 10) Ydja; 11) NemA; 12) RutE.

The strategy of engineering host strains (*Escherichia coli* B-95.ΔA NK53) with attenuated nitroreductase activity by gene knockout proved to be very efficient in practice. It should also be mentioned that only some genes that were knocked out last were found to have a small percentage of clones that exhibited the insertion of antibiotic resistance cassette into the pre-existing FRT scar. However, this problem was easily circumvented because for each deletion round the selection of only five clones was always sufficient to find a correct strain.

2.5.6. Experiments to Verify the Decreased NK53 Nitroreductase Activity

To assess the effects of gene deletions, the same reporter protein construct as in **Figure 61** (ELP(5x *m*-oNB-Dopa)-sfGFP) was produced in knock-out strain B-95.ΔA NK53 and analyzed

by ESI-MS (**Figure 64A**). Overlay of **Figure 61** and **Figure 64A** yields **Figure 64B**, where it is clear that almost no target protein with nitro group reduction is present. Only a very small peak is seen, indicating that a target protein with only one reduced nitro group is present in the sample at a very low level, making the genome engineering efforts of the B-95.ΔA a success. Judging from the peaks in the spectrum, around 3% of the total protein contains one reduced nitro group, which corresponds to an overall reduction of 94%.

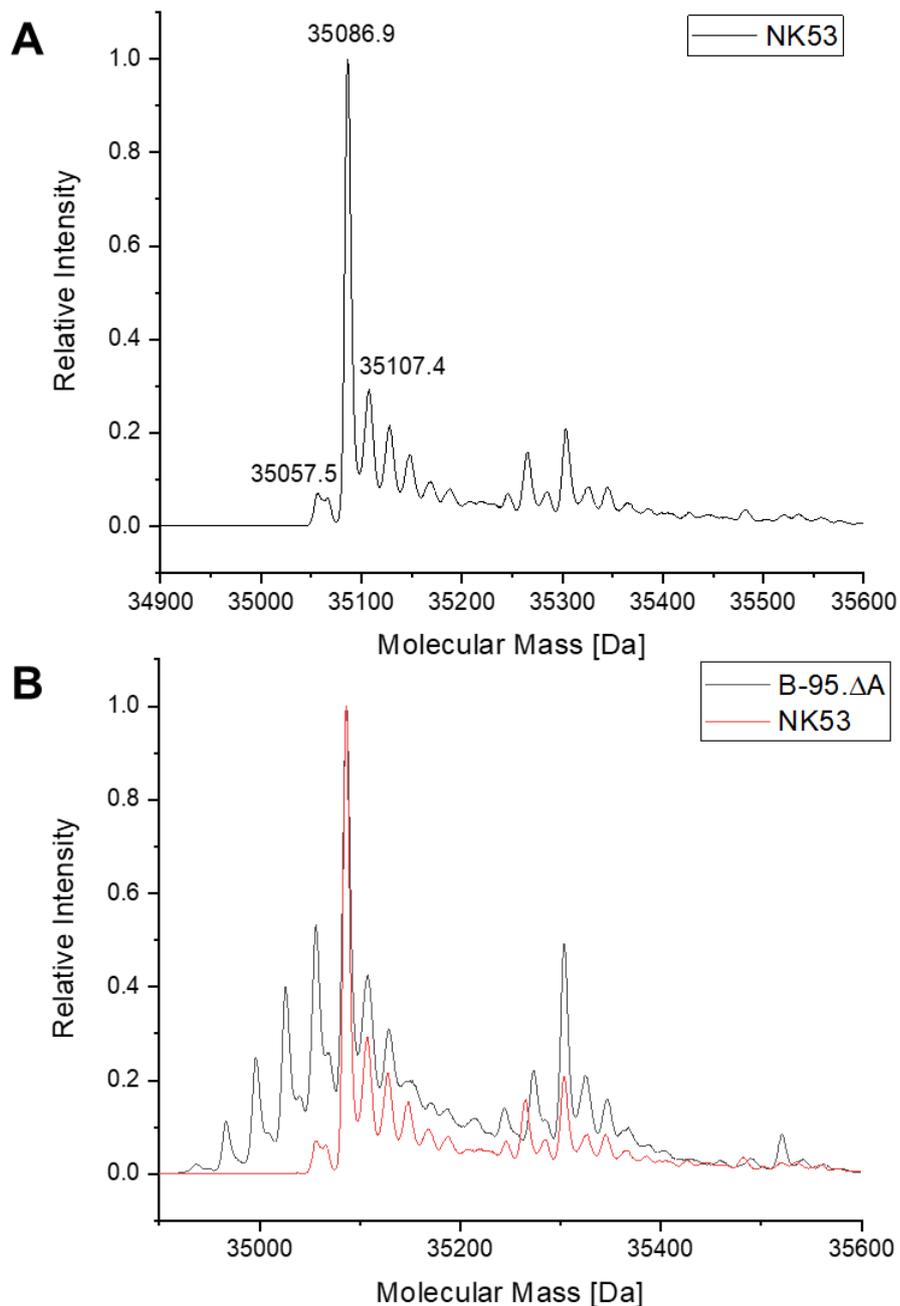


Figure 64: Mass spectrometric profiles of protein expressed in the *E. coli* B-95.ΔA NK53 strain with attenuated nitroreductase activity. **A**) ESI-MS spectrum of ELP(5x amber)-sfGFP produced in NK53. **B**) Overlay of ESI-MS spectra of ELP(5x amber)-sfGFP produced in B-95.ΔA NK53 and B-95.ΔA (for peak annotation see **chapter 2.5.4**). Expected protein mass for five incorporated *m*-oNB-Dopa (**51**): 35087.1 Da. Observed mass: 35086.9 Da. Expected protein mass for five incorporated *m*-oNB-Dopa (**51**) with one reduced ncAA: 35057.1 Da. Observed mass: 35057.5 Da.

To further validate the ESI-MS results and assess the practicality of the NK53 strain for the incorporation of other nitro group containing ncAAs, two additional experiments were performed. These experiments were designed to determine whether differences in ONBY (**41**) nitro-reduction could be detected depending on the production strain, in analogy to the *m*-ONB-Dopa (**51**) experiments. The system used for ONBY (**41**) incorporation has been described previously.³ In the first experiment, ONBY (**41**) was introduced at position Y63 of amilCP (equivalent of Y66 in GFP), a chromoprotein from the coral *Acropora millepora*.³⁵² A similar experiment was performed with a GFP derivative, because amilCP has a similar structure it was considered to also be also feasible with amilCP.³⁵³ The target protein amilCP(Y63ONBY) was then produced in B-95.ΔA and the engineered B-95.ΔA NK53 strain. The purified protein (100 μL) was then normalized to the same concentration (1 mg/mL) and irradiated with UV-light (**Figure 65**). The absorbance at 588 nm was then monitored as a function of time. Unfortunately, no significant differences in the absorption were detectable. If there were a difference in the reduction of the ONBY nitro group, one would expect the absorption to be lower for the amilCP produced in the B-95.ΔA strain without the gene deletions. It has been shown that ncAAs are reduced after being incorporated into the polypeptide chain it is plausible that the non-occurrence of the nitroreduction is due to protein context effects. Since amilCP, similar to the GFP protein family, has a β-barrel surrounding the chromophore, this could shield the ONBY moiety from enzymes performing the reduction. This could explain why also no reduction of ONBY (**41**) was detected in the GFP publication.³⁵³ The same results could also be observed by the naked eye (**Figure 142**). To confirm this hypothesis ESI-MS measurements of proteins produced in both strains were conducted that confirmed that no nitroreduction did occur in both strains (**Figure 200** and **Figure 201**).

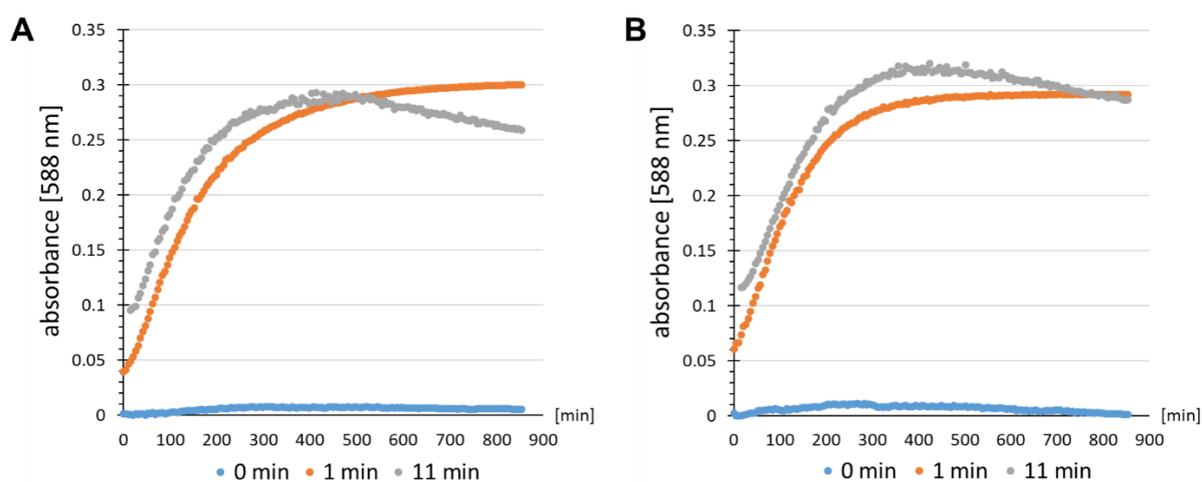


Figure 65: Time-resolved chromophore maturation of amilCP(Y63ONBY) detected by an absorption assay at 588 nm. Colored dots indicate the time amilCP(Y63ONBY) was irradiated with UV-light. Production of amilCP(Y63ONBY) with **A**) B-95.ΔA NK53 or **B**) B-95.ΔA.

The second experiment consisted of an established Nisin *Lactococcus lactis* (*L. lactis*) indicator strain assay and was performed by Jessica H. Nickling (Technische Universität Berlin, Germany).^{354,355} Here, ONBY (**41**) was introduced at permissive Nisin sites indicated in **Figure 66**. It can clearly be seen that only the cell lysates produced in the nitroreductase-deficient strain B-95.ΔA NK53 show a dramatic increase in antimicrobial activity after UV-irradiation associated with ONBY (**41**)-induced cleavage. The strongest increase is observed when ONBY is introduced at Nisin position 17. This experiment highlights that especially proteins lacking a tertiary structure, which can protect the oNB group from reduction, can benefit from protein production in a nitroreductase deficient *E. coli* strain. Also, the strategy described here serves as a template to drastically reduce the nitroreductase activity in any *E. coli* strain in a fast and unexpensive manner. If well prepared this sextuple knock out can be made within one month.

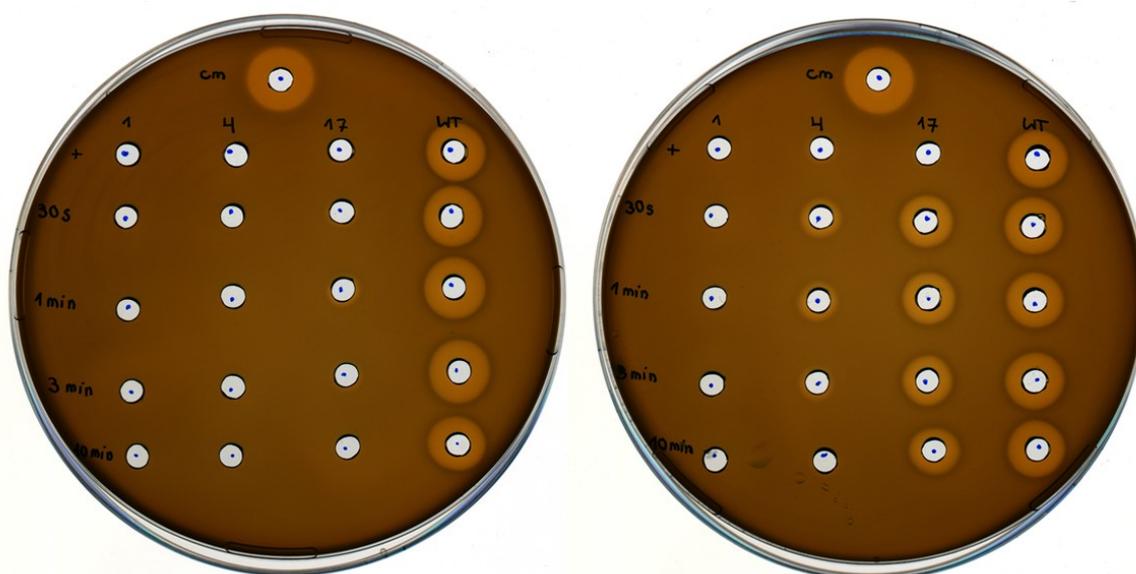


Figure 66: *L. lactis* indicator assay using cell lysates containing Nisin pro-peptide modified with ONBY (**41**) at position 1, 4 and 17, with additional wild-type Nisin pro-peptide (WT). Lysates were prepared with either B-95.ΔA (left) or B-95.ΔA NK53 (right). Cm = Chloramphenicol positive control, first row (+) = with ONBY (**41**) but without UV-light irradiation of the cell lysate. Numbers from top to bottom indicate UV-light irradiation time (30s, 1 min, 3 min, 10 min). The numbers above the column indicate the position number of Nisin where ONBY (**41**) is introduced.

2.5.7. Expanding the ncAA Incorporation Repertoire for Wet Tissue Adhesion

Despite the previously described efforts, mussel adhesion is most likely not suitable for wet tissue adhesion because mussel adhesion relies on relatively weak physical interactions with low adhesion energies in the range of 1 to 10 Jm⁻².³⁵⁶ A recent breakthrough by Hong and colleagues showed that wet tissue adhesion is possible with synthetically produced photo-inducible hydrogels with side chain moieties similar to **54**. They demonstrated the applicability of their polymer even to tissues under high pressure (stopping arterial and cardiac bleeding).³⁵⁷ Inspired by this approach, a strategy was developed to incorporate ncAAs into protein scaffolds that could mimic this mode of adhesion (**Figure 67**).

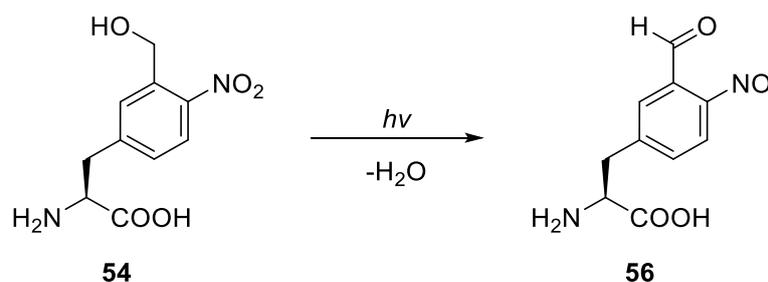


Figure 67: Hypothetical light-inducible chemical reaction involving the release of H_2O from the scaffold of **54** leading to a reactive aldehyde (**56**).

The adhesion mechanism would be based on a photo-activatable aldehyde (**56**) which should be able to crosslink with primary amines (e.g., lysine side chains) on tissue surfaces. Although a synthetic route to such adhesion materials now exists, it would still be sensible to produce such agents based on biomaterials, especially ELPs, as they have several advantages over synthetically produced materials for *in vivo* use (see 2.5.1). A fluorescence assay was used to determine whether *p*-oNB-alanine (**54**) or *m*-oNB-alanine (**55**) could be incorporated with the two *Mjo*NB-DopaRS variants (Figure 68). It was assumed that the recognition would not be very good, if it existed at all. Therefore, the ncAA concentration chosen for a *Mj*TyrRS-based system was high (5 mM) to detect even low levels of activity. It can be observed that some incorporation is possible, but with very low efficiency. Surprisingly, the efficiency of both variants is at the same level, but with lower background suppression, the A32Y mutant performs slightly better.

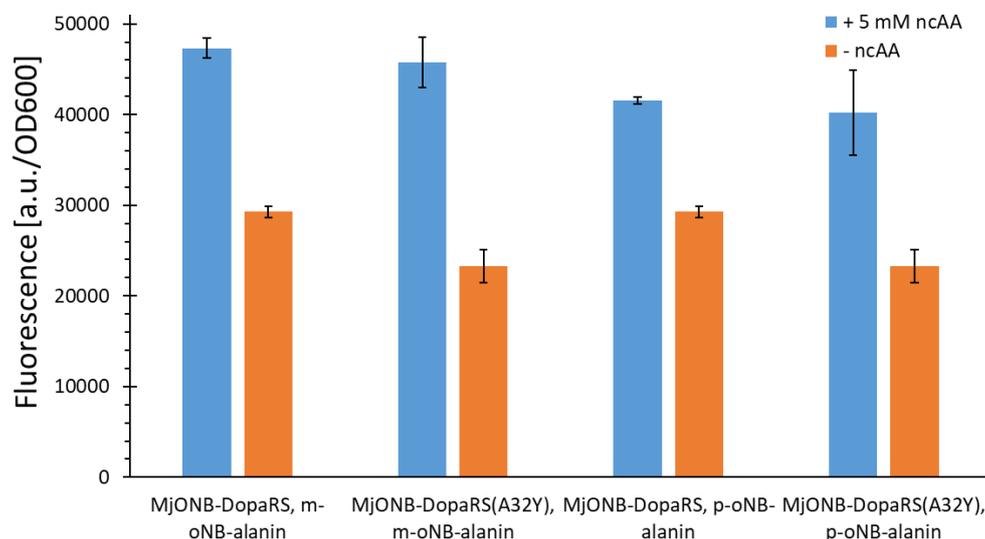


Figure 68: Comparison of OTS efficiency of the indicated constructs. Ribosomal incorporation with 5 mM ncAA (*p*-oNB-alanine (**54**) or *m*-oNB-alanine (**55**)). Intact cell fluorescence of B-95. Δ A endpoint measurements after 24 h of incubation. Data (incl. standard deviation) represent the mean of three biological replicates.

Since neither of the two *Mjo*NB-Dopa variants incorporated substrate **54** and **55** well, the existing libraries in the Budisa group were screened for suitable variants. Unfortunately, no variant could be found (Figure 136 and Figure 137). This was not a big surprise since these

libraries were designed for completely different substrates (**41** and **53**), and the *MjTyrRS* enzyme needs more specific mutations to recognize different substrates than e.g., the *PyIRS* (see **chapter 1.3.2**). However, the initially low activity of the two *MjoNB-DopaRS* variants indicates that engineering of an efficient enzyme for substrates *p*-*o*NB-alanine (**54**) or *m*-*o*NB-alanine (**55**) should be possible. Because there is an *MjTyrRS* variant with an appropriate crystal structure that can incorporate *m*-nitro-tyrosine, engineering efforts based on this scaffold might be advantageous for incorporation of *m*-*o*NB-alanine (**55**) because the difference between this substrate and *m*-nitro-tyrosine is only a CH₂ group.³⁵⁸ This variant and an additional one that showed general promiscuity towards *para*-substituted Phe derivatives (*MjpCNF*) were tested but exhibited no/low incorporation activity regarding substrates **54** and **55** (**Figure 69**).³¹⁸ In general, the *MjpCNF* variant showed considerable amount of background suppression, suggesting suboptimal orthogonality.

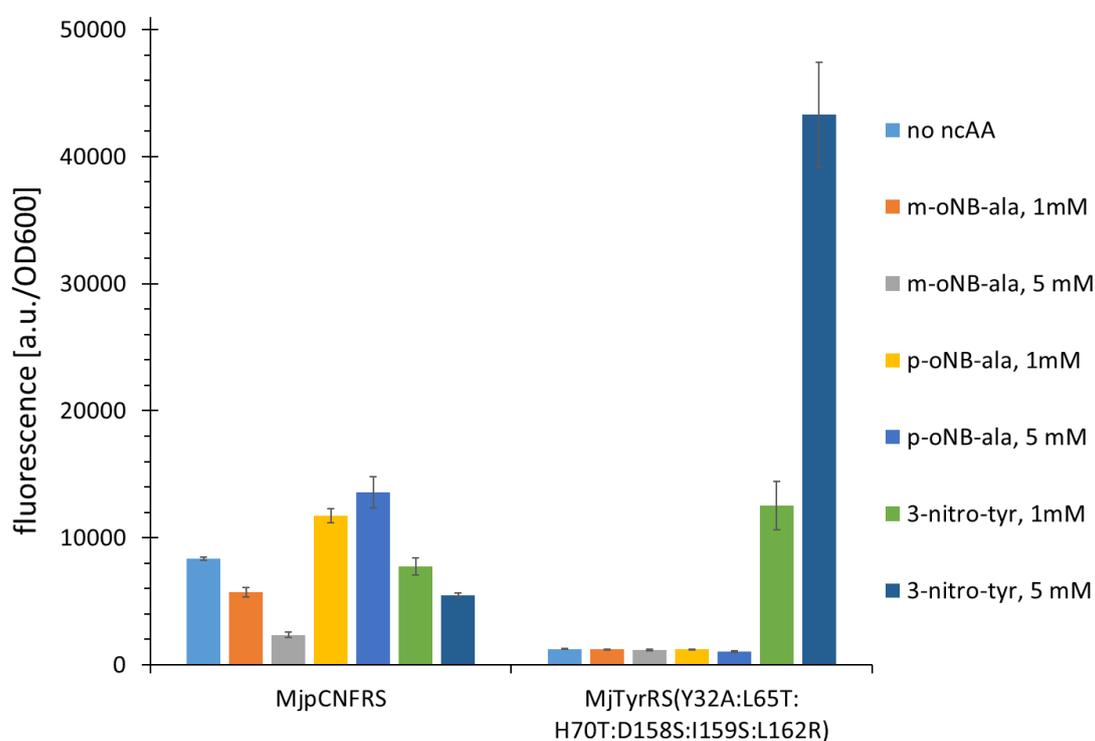


Figure 69: Comparison of OTS efficiency of the constructs indicated in the text. Ribosomal incorporation with ncAA (*p*-*o*NB-ala = *p*-*o*NB-alanine (**54**); *m*-*o*NB-ala = *m*-*o*NB-alanine (**55**); 3-nitro-tyr = 3-nitro-tyrosine). Intact cell fluorescence of B-95.ΔA endpoint measurements after 24 h of incubation. Data (incl. standard deviation) represent the mean of three biological replicates.

3. Summary and Outlook

In the first part of this study, the low intrinsic protein solubility of PylRS and its impact on the production of ncAA-modified recombinant proteins were investigated in a rational manner. A set of genetically encoded solubility-tags derived from diverse organisms were tested for improvements in *MbPylRS* OTS performance *in vivo*. The results showed that the N-terminal SmbP-tag boosts the performance of wild-type (120-245 %) and especially engineered *MbPylRS* systems (200-540 %) by delivering more protein for the amount of ncAA supplied. The most potent tag has been shown to increase soluble expression of the enzyme, which is most likely the cause of the observed increased efficiency. The straightforward genetic addition of the solubility tag to the otherwise inefficient enzyme resulted in unusually high performance for a PylRS-based system. The tag appears to remediate the destabilizing effects of active site mutations that were originally introduced to allow genetic encoding of new ncAAs. This explains why the observed improvements are higher for the engineered aaRS variants than for the wild-type enzyme. The SmbP-tag enables the enzyme engineered for Sac installation (*MbSacRS*) to outperform the the wild-type enzyme. The obtained comprehensive data for a number of ncAA and aaRS enzyme pairs strongly supports the working hypothesis that intracellular folding and solubility of engineered PylRS enzymes readily becomes a bottleneck for orthogonal translation - besides the kinetics of ncAA binding, activation and charging to the cognate tRNA. The results highlight that the expression level and solubility of engineered PylRS enzymes are important parameters that should not be neglected when optimizing cells for the production of ncAA-modified unnatural proteins. In the light of previous reports and the reaction pathway towards charged tRNA, it was remarkable to see that even the attachment of large fusion partners did not abolish the *in vivo* function of the orthogonal pair, with the N-terminal aaRS domain primarily involved in tRNA recognition and binding. Gratifyingly, it was also shown that the tag can help improve other PylRS (*MtSacRS*(TM-1)).

Overall, it was demonstrated that the classic method of obtaining more soluble and active enzymes by fusion partners can be transferred to OTS engineering. Remarkably, this strategy for improving aaRS performance *in vivo* does not require any changes in the catalytic domain or active site of the enzyme. The fidelity of the system is maintained, as no increased background stop codon suppression was detected. It is conceivable that transferring this approach to even more PylRS-based systems used *in vivo* and/or *in vitro* could substantially boost the efficiency of unnatural protein production. The findings suggest that engineered enzymes with drastic changes in the active site particularly benefit from rescue by solubility tags. In addition, the increased amount of soluble *MbPylRS* could eventually facilitate the crystal structure elucidation of this enzyme, and generally increased aaRS activity and robustness may allow the engineering of enzymes able to activate new classes of ncAA substrates. It is reasonable to assume that many aaRS variants with suitable active site

residues (generated by various methods) remain hidden due to insufficient aaRS stability/activity for detection by screening systems. The approach presented here could help to find these hidden variants.

The second part of this study was aimed at expanding the substrate scope of PyIRS OTS which was greatly facilitated by the previously improved *Mb*PyIRS variant (shown in the third part). Earlier works have shown that the PyIRS, which is naturally specialized for large bulky substrates, could also be repurposed for substrates with smaller side chains (Sac (**1**)). The *Mb*SacRS incorporates the shortest non-bulky ncAA (Sac (**1**)) known to date. In the absence of a crystal structure, the structure-activity relationship of *Mb*SacRS was elucidated by mutational studies. Based on these findings, several *Mb*PyIRS variants were engineered for the incorporation of aliphatic amino acids and variety biochemically/structurally useful derivatives thereof. Some of the incorporated substrates (allylglycine **8** and propargylglycine **13**) were recently synthesized *in vivo* in *E. coli* and would therefore open up the possibility of coupling metabolic engineering and incorporation of ncAAs.²⁷⁴ This could eliminate the need to add ncAAs to the culture medium, which would drastically reduce costs and simplify associated applications. In total, this study created several *Mb*PyIRS variants capable of incorporating 23 ncAAs and one cAA. A thorough research of the relevant literature revealed that 17 of these ncAAs (besides **1**, **3**, **4**, **17**, **27** and **28**) have not yet been ribosomally incorporated by amber suppression and 20 of them have not yet been inserted with the PyIRS system.^{144,146,260,279} The variants created in this study can selectively incorporate a new class of ncAAs, allowing the physicochemical diversity of amino acids that can be incorporated into recombinant proteins to be significantly increased. In addition, a number of important information has been collected on the role of specific residues responsible for the orthogonality of PyIRS. Key residues responsible for maintaining orthogonality while engineering the *Mb*PyIRS enzyme to accommodate smaller substrates were determined. This will facilitate future enzyme engineering efforts for the incorporation of structural analogs. To this end, PyIRS-libraries with a reduced sizes can now be created, which simultaneously improves the likelihood of finding the desired enzymes and reduces the amount of work required to do so because negative selection can be omitted.

In the third part, an attempt was made to address the low catalytic efficiency of existing PyIRS OTSs, as in the first part, but with a different approach. Based on considerations of extremophilic enzymes (see **chapter 1.4**) a comparative study of PyIRS variants of these origins was performed. In particular, the OTS efficiency and substrate promiscuities of a series (wild-type and mutants) of psychro-, meso-, and thermophilic PyIRS were investigated. Encouragingly, one of the psychrophilic PyIRS (*Mbur*PyIRS) was found to exhibit remarkable ncAA incorporation efficiency even at very low ncAA concentrations. Interestingly this high efficiency was always found when residue N308 (*M. burtonii* notation) was unchanged (for

Pyl (**1a**) derivative and Sac (**1**)). Including other studies into the considerations, this might indicate that a polar/charged residue at position N308 or C310 (*M. burtonii* notation) confers a catalytic advantage.³⁵⁹ If true this puts a critical constraint onto the creation of catalytically efficient PylRSs and should definitely be considered when creating genetic libraries. Given the higher promiscuity of *MburPylRS* the probability coping with this limitation would be increased, making it the enzyme of choice. However, even when N308 was mutated to change the substrate specificity for Phe derivatives and small aliphatic ncAAs, the *MburPylRS* showed the best performance with few exceptions. This behavior was found for a wide variety of substrates, including Pyl and Phe derivatives as well as small-chain aliphatic ncAAs, highlighting the robustness of the results. Experimental data were used to derive a promiscuity coefficient for the PylRS variants that agreed well with later experiments. The higher substrate promiscuities of the psychrophilic PylRSs would be expected regarding the literature and was confirmed, even though the SmbP-*MbPylRS* could perform at the same level for Phe- and for some small aliphatic ncAA derivatives. The most impressive result was the multi-site ncAA incorporation efficiency of *MburSacRS*(S379T) for Sac (**1**) which was on par with the performance of wild-type *MburPylRS*, for Bock (**31**). It must be noted that Bock (**31**) is an even better substrate than Pyl (**1a**) for the wild-type PylRSs. When compared with to one of the best performing OTS known (*MjTyrRS*) the *MburPylRS* and *MburSacRS*(S379T) proved superior in multi-site ncAA incorporation.

The combination of the very high catalytic activity of *MburPylRS*'s and the greater substrate promiscuity could potentially give rise to a plethora of applications that were previously not possible due to the limitations of the PylRS system mentioned above. Considering, that there are only two OTSs known to come close to wild-type recombinant protein production levels (*MjTyrRS*¹⁻³ and *AfTyrRS*³¹⁶), the addition of a third OTS to this group is highly desired. This improves the probability that the number of ncAAs, that can be simultaneously incorporated with good efficiency, can be increased. In particular, because the *MburPylRS* is a class II aaRS and possesses a variety of accompanying advantages in comparison to the class I aaRS (*Mj/AfTyrRS*, as discussed in the introduction). In this context, the most promising improvements can be expected with the *MburPylRS* system in conjunction with semi-synthetic organisms possessing liberated codons^{184,360}, since the tRNA^{Pyl} anticodon can be freely chosen. It has been shown that when the *MmPylRS* is used in a recoded *E. coli* strain, OTS efficiency is higher for some of these newly created codons, but only twofold.¹⁸⁴ This suggests that even when using free codons not competing with RF1, catalytic efficiency of the PylRS OTS is still a limiting factor. This limiting catalytic efficiency was successfully addressed in this study in several ways. If the higher efficiency of *MburPylRS* holds true in mammalian cells as well, this could translate into enhanced and expanded applicability within these hosts. Since in most cases mammalian cells are used to produce therapeutically relevant proteins this could

mean that the *MburPy*IRS system is even attractive for commercial use, e.g. for the production of Antibody-Drug-Conjugates (ADCs) or the like.^{361–363} ADCs are already produced by companies such as Ambrx and even preliminary approved for certain kinds of breast and gastric cancers.^{364,365}

4. Materials and Method

4.1. Materials

4.1.1. Chemicals

Unless indicated, all standard chemicals were purchase from Carl Roth GmbH (Karlsruhe, Germany), Merck (Waltham, MA, USA), VWR International GmbH (Waltham, MA, USA) or Sigma-Aldrich (Taufkirchen, Germany).

Table 11. Amino acids used in this work.

No.	Name	Abbreviation	Cas No.	Company
1	S-allyl-L-cystein	Sac	21593-77-1	TCI Deutschland
2	(S)-2-aminoheptanoic acid	C7	44902-02-5	Fluorochem
3	(S)-2-aminooctanoic acid	C8	116783-26-7	Fluorochem
4	(S)-2-aminohept-6-enoic acid	C7 alken	166734-64-1	Chempur
5	(S)-2-aminohexanoic acid	C6	327-57-1	TCI Deutschland
6	(S)-2-aminohex-5-enoic acid	C6 alken	90989-12-1	Fluorochem
7	(S)-2-aminopentanoic acid	C5	6600-40-4	TCI Deutschland
8	(S)-2-aminopent-4-enoic acid	C5 alken	16338-48-0	Fluorochem
9	(S)-2-amino-3-cyclopropylpropanoic acid	cyclo-ala	102735-53-5	Fluorochem
10	(S)-2-aminobutyric acid	C4	1492-24-6	TCI Deutschland
11	(S)-2-aminohept-6-ynoic acid	Bis-Hpg	835627-45-7	Chiralix
12	(S)-2-aminohex-5-ynoic acid	Hpg	98891-36-2	Toronto Research Chemicals
13	(S)-2-aminopent-4-ynoic acid	propG	23235-01-0	Fluorochem
14	(S)-2-amino-3-azidopropanoic acid hydrochloride	azido-ala	105661-40-3	Iris Biotech
15	(S)-2-amino-4-azidobutanoic acid hydrochloride	AHA	942518-29-8	Carl Roth
16	(S)-2-amino-5-azidopentanoic acid hydrochloride	azido-ornithin	1782935-10-7	Iris Biotech
17	(S)-2-amino-6-azidohexanoic acid hydrochloride	Azido- norleucine	159610-92-1	Fluorochem
18	(S)-2-amino-3-cyanopropanoic acid	CA	6232-19-5	Iris Biotech
19	(S)-2-amino-4-cyanobutanoic acid	CHA	6232-22-0	Iris Biotech
20	(S)-2-amino-5,5'-azi-hexanoic acid	pMet	851960-68-4	Thermo Fisher Scientific
21	(S)-2-amino-4-methylpent-4-enoic acid	4,5-DHL	87392-13-0	Fluorochem
22	L-methionine	Met	63-68-3	Carl Roth
23	L-methionine sulfoxide	Met-sulfoxide	3226-65-1	Sigma-Aldrich (Merck)
24	L-methionine sulfone	Met-sulfone	7314-32-1	Iris biotech
25	L-ethionine	Eth	13073-35-3	Sigma-Aldrich (Merck)
26	S-tert-butyl-L-cystein		2481-09-6	Fluorochem
27	S-propargyl-L-cystein	SproC	3262-64-4	Fluorochem
28	S-benzyl-L-cystein		3054-01-1	Fluorochem
31	N ^ε -tert-Butoxycarbonyl-L-lysine	BocK	2418-95-3	Budisa Group
32	N ^ε -Allyloxycarbonyl-L-lysine	AllocK	6298-03-9	Fluorochem
33	N ^ε -Propargyloxycarbonyl-L-lysine	ProK	1428330-91-9	Iris Biotech

34	<i>N</i> ^ε -((2-Azidoethoxy)carbonyl)-L-lysine	AzidoK	1994331-17-7	Iris Biotech
38	O-methyl-L-tyrosine	O-methyl-Y	6230-11-1	Fluorochem
39	O-tert-butyl-L-tyrosine	O-tert-Butyl-Y	18822-59-8	Fluorochem
40	O-propargyl-L-tyrosine	O-prop-Y	610794-20-2	Iris Biotech
41	<i>ortho</i> -nitro-benzyl-L-tyrosine	ONBY	207727-86-4	Sigma-Aldrich (Merck)
42	3'-azibutyl- <i>N</i> ^ε -carbamoyl-L-lysine	PhotoK	1253643-88-7	Iris Biotech
43	<i>N</i> ^ε -benzyloxycarbonyl-L-lysine	BenzK	1155-64-2	TCI Deutschland
44	4-azido-L-phenylalanine	azido-F	33173-53-4	Fluorochem
45	O-allyl-L-tyrosine	O-allyl-Y	107903-42-4	Iris Biotech
46	4-cyano-L-phenylalanine	cyano-F	167479-78-9	Alfa Aesar
47	O-CF ₃ -L-tyrosine	O-CF ₃ -Y	921609-34-9	Fluorochem
48	4-ethynyl-L-phenylalanine	ethynyl-F	278605-15-5	Sigma-Aldrich (Merck)
49	(S)-2-amino-3-(3-(hydroxymethyl)-4-nitrophenyl)propanoic acid	<i>p</i> -oNB-alanin		Budisa Group
50	4-benzoyl-L-phenylalanin	Bpa	104504-45-2	Bachem AG
51	<i>o</i> -(2-nitrobenzyl)-3,4-dihydroxyphenylalanine	<i>m</i> -oNB-Dopa		Budisa Group
52	H-(7-Hydroxycoumarin-4-yl)-ethyl-Gly-OH	Coumarin	905442-42-4	Bachem AG
53	<i>para</i> -(<i>ortho</i> -(2-Nitrobenzyl))-3,4-dihydroxyphenylalaninen	<i>p</i> -oNB-Dopa		Budisa Group
54	(S)-2-amino-3-(3-(hydroxymethyl)-3-nitrophenyl)propanoic acid	<i>m</i> -oNB-alanin		Budisa Group

4.1.2. Media and Supplements

Liquid media were autoclaved for 20 min at 121 °C and 1.5 bars; agar plates were poured with 1.5 % (w/w) agar, which was added before autoclaving.

LB medium:

Ingredient	Concentration
yeast extract	5 g/L
tryptone	10 g/L
NaCl	10 g/L

Auto-induction medium ZYP-5052 was used for protein production in shake flasks or in 96 well plates in order to monitor the fluorescence of sfGFP.³⁶⁶ After depletion of glucose, induction of lac promoters is achieved with residual lactose. Stock solutions of ZY were autoclaved, 20x P, 50x 5052, Mg₂SO₄ (1M) and 1000x trace elements were sterilized by filtration (pore size 0.22 μm). Concentration of Kan 100 μg/mL when using ZYP-5052 media.

ZYP 5052 medium:

Stock	Ingredient	Concentration	Volume per L
ZY	tryptone	1 %	928 mL
	yeast extract	0.5 %	
20x P	Na ₂ HPO ₄	50 mM	50 mL
	KH ₂ PO ₄	50 mM	
	(NH ₄) ₂ SO ₄	25 mM	
50x 5052	glycerol	20 mL	20 mL
	glucose	0.05 %	
	α-lactose	0.2 %	
1 M MgSO ₄	MgSO ₄	2 mM	2 mL
1000x trace elements	trace elements	0.2 x	0.2 mL

Supplements:

Supplement	Stock Concentration	Final concentration
Ampicillin	100 mg/mL	100 µg/mL
Chloramphenicol	37 mg/mL	37 µg/mL
Kanamycin	50 mg/mL	50 µg/mL
Spectinomycin	100 mg/mL	100 µg/mL
Streptomycin	50 mg/mL	50 µg/mL
Zeocin	100 mg/mL	100 µg/mL
IPTG	1 M	0.5 mM – 1 mM
Arabinose	20 %	0.2 % - 0.002 % (w/v)
Glucose	1 M	20 mM
Ampicillin	100 mg/mL	100 µg/mL
Chloramphenicol	37 mg/mL	37 µg/mL

4.1.3. Strains

Strain name	Genotype	Source
<i>E. coli</i> BL21(DE3)	<i>E. coli</i> B F ⁻ <i>ompT gal dcm lon hsdS_B(r_B⁻m_B⁻)</i> λ(DE3 [<i>lacI lacUV5-T7p07 ind1 sam7 nin5</i>]) [<i>malB</i> ⁺] _{K-12} (λ ^S)	Budisa group Studier et al. ³⁶⁷
<i>E. coli</i> TOP10	F- <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74 nupG recA1 araD139</i> Δ(<i>ara-leu</i>)7697 <i>galE15 galK16 rpsL(Str^R) endA1 λ⁻</i>	Invitrogen, now ThermoFisher Scientific (Waltham, MA, USA)
<i>E. coli</i> B-95.ΔA	<i>E. coli</i> BL21(DE3) Δ <i>prfA</i> ; 95 UAG codons changed to UAA or UGA	From Kensaku Sakamoto ¹⁵⁴
<i>E. coli</i> B-95.ΔA NK53	<i>E. coli</i> BL21(DE3) Δ <i>prfA</i> ; 95 UAG codons changed to UAA or UGA, Δ <i>NfsA</i> ::FRT; Δ <i>NfsB</i> ::FRT; Δ <i>AzoR</i> ::FRT; Δ <i>Ydja</i> ::FRT; Δ <i>NemA</i> ::FRT; Δ <i>RutE</i> ::FRT	This work

<i>E. coli</i> C321.ΔA.exp(DE3)	<i>E. coli</i> MG1655 Δ(<i>ybhB-bioAB</i>):: <i>zeoR ΔprfA</i> λ(DE3); all 321 UAG codons changed to UAA	Lajoie et al. ¹⁵² Budisa group
<i>E. coli</i> JX33(DE3)	<i>E. coli</i> MDS42 Δ <i>prfA prfB</i> ^f	Budisa group, Lei Wang ²⁷⁷

4.1.4. Technical Equipment

Balances:

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

Centrifuges:

Centrifuge 5810 R	Eppendorf AG (Hamburg, Germany)
Centrifuge 5418 R	Eppendorf AG (Hamburg, Germany)
MiniSpin plus	Eppendorf AG (Hamburg, Germany)

Gel electrophoresis:

Horizontal agarose gel system	Factory of the Max-Planck Institute for Biochemistry (Martinsried, Germany)
Vertical SDS-gel system	Factory of the Max-Planck Institute for Biochemistry (Martinsried, Germany)

Incubators, mixers & shakers:

Ecotron	Infors HT (Einsbach, Germany)
Multitron	Infors HT (Einsbach, Germany)
Incubator series B, KB	Binder (Tuttlingen, Germany)

Liquid Chromatography:

Peristaltic pump P1	Pharmacia Biotech (now: Cytiva, Marlborough, MA, USA)
---------------------	--

Mass spectrometry:

Agilent 6530 Accurate-Mass Q-TOF	Agilent (Santa Clara, CA, USA)
Waters Xevo G2-XS QTof	Waters Corporation (Milford, MA, USA)

Spectroscopy:

Ultrospec 6300 pro	Amersham Biosciences (now: GE Healthcare, München, Germany)
BioPhotometer plus	Eppendorf AG (Hamburg, Germany)
Microplate reader Infinite M200	Tecan (Männedorf, Switzerland)

Thermocyclers:

Peqstar 2x Gradient	Peqlab (Erlangen, Germany)
---------------------	----------------------------

Thermomixer:

Mixing Block MB-102	Bioer Technology (Binjiang, China)
---------------------	------------------------------------

Other devices:

Sonopuls HD 3200	Bandelin (Berlin, Germany)
Sonotrodes MS72, KE76	Bandelin (Berlin, Germany)
Power supply Power Pack P25 T	Biometra (Jena, Germany)
Power supply Consort EV 261 und E 143	Sigma-Aldrich (Taufkirchen, Germany)
Power supply Consort E 143	Sigma-Aldrich (Taufkirchen, Germany)
Orbital shaker Rotamax 120	Heidolph (Schwabach, Germany)
IKA Combimag RET	IKA (Staufen, Germany)
Vortex Genie™	Bender & Hobein AG (Zürich, Switzerland)
Ice machine Scotsman AF 80	Scotsman (Vernon Hills, IL, USA)
pH-Meter S20-SevenEasy™	Mettler Toledo (Gießen, Germany)
Gel-documentation system Felix 2050	Biostep (Jahnsdorf, Germany)
Scanner ViewPix 700	Biostep (Jahnsdorf, Germany)
MicroPulser™	Bio-Rad Laboratories GmbH (München, Germany)

4.1.5. Oligonucleotides

All oligonucleotides were purchased from Sigma-Aldrich (Taufkirchen, Germany), resuspended in ddH₂O to a final concentration of 100 µM and stored at -20 °C. Working concentration of primers were adjusted to 10 µM. Primers shorter than 50 bp were generally purchased in desalted form. Primers between 50-80 bp were ordered in cartridge-purified form, while longer ones were obtained in HPLC-purified grade. Primers are not listed because far over five hundred were used and the utility of this information is limited at best.

4.1.6. Biomolecular Reagents, Enzymes and Kit Reagents:

Name	Supplier
GeneRuler Mix Ladder	ThermoFisher Scientific (Waltham, MA, USA)
Prestained Protein Marker	ThermoFisher Scientific (Waltham, MA, USA)
Unstained Protein Marker	ThermoFisher Scientific (Waltham, MA, USA)
B PER® Bacterial Protein Extraction Reagent	ThermoFisher Scientific (Waltham, MA, USA)

Enzymes:

Name	Supplier
FastAP Alkaline Phosphatase	ThermoFisher Scientific (Waltham, MA, USA)
Phusion High-Fidelity DNA Polymerase	ThermoFisher Scientific (Waltham, MA, USA)
Phusion High-fidelity PCR Master Mix	ThermoFisher Scientific (Waltham, MA, USA)
Q5 High-Fidelity DNA Polymerase	New England Biolabs, Frankfurt am Main, Germany)
<i>Taq</i> DNA Polymerase	In-house and ThermoFisher Scientific (Waltham, MA, USA)
FastDigest restriction enzymes	ThermoFisher Scientific (Waltham, MA, USA)
Restriction enzymes	ThermoFisher Scientific (Waltham, MA, USA), New England Biolabs (Frankfurt am Main)
T4 DNA Ligase	ThermoFisher Scientific (Waltham, MA, USA)
Lysozym	Carl Roth (Karlsruhe, Germany)
DNase	Carl Roth (Karlsruhe, Germany)
RNase	Carl Roth (Karlsruhe, Germany)

Kits:

Name	Supplier
GeneJET™ Plasmid Mini-prep Kit	ThermoFisher Scientific (Waltham, MA, USA)
GeneJET™ PCR Purification Kit	ThermoFisher Scientific (Waltham, MA, USA)
GeneJET™ Gel Extraction Kit	ThermoFisher Scientific (Waltham, MA, USA)

4.1.7. Buffers and Solutions

All buffers were prepared with distilled water (dH₂O) or ddH₂O (Milli-Q, Merck-Millipore).

Buffers for Polyacrylamide gel electrophoresis gels:

Name	Composition
Coomassie staining solution	1 g Coomassie Brilliant Blue R-250 500 mL ethanol 100 mL glacial acetic acid ad 1 L dH ₂ O
5X SDS loading dye	80 mM TRIS pH 6.8 10 % SDS 12.5 % glycerol 4 % (v/v) mercaptoethanol 0.2 % (w/v) bromophenol blue
SDS running buffer	190 mM glycine 25 mM TRIS 3.5 mM SDS

Resolving gel	380 mM Tris-HCl pH 8.8 15 % Acrylamid / bis-acrylamide (37.5:1) 0.1 % SDS 0.05 % APS 0.05 % TEMED
Stacking gel	125 mM TRIS-HCl pH 6.8 5 % acrylamide / bis-actylamide (37.5:1) 0.1 % SDS 0.05 % APS 0.17 % TEMED

Buffers for Agarose gel electrophoresis:

Name	Composition
50X TAE buffer	2 M Tris 2 M acetic acid 10 % (v/v) 0.5 M EDTA pH 8.0
6X DNA loading dye	0.25 % bromphenol blue 0.25 % xylencyanole 30 % glycerol
TE buffer	10 mM Tris pH 7.4 1 mM EDTA pH 8.0

Buffers for Protein Purification:

Name	Composition
Lysis / washing buffer	50 mM NaH ₂ PO ₄ 300 mM NaCl 20 mM Imidazole pH 8.0
Elution buffer	50 mM NaH ₂ PO ₄ 300 mM NaCl 500 mM Imidazole pH 8

4.2. Molecular Biological Methods

4.2.1. Transformation of *E. coli*

4.2.1.1. Transformation of chemically competent cells

For the preparation of chemically competent *E. coli* cells, 300 mL LB media were inoculated with an overnight *E. coli* culture (1:100). Cells were grown (37 °C, 200 rpm) until an OD₆₀₀ of 0.4 – 0.5 was reached. The culture was chilled on ice, and cells were harvested by

centrifugation (4 °C, 5 min, 4000 x g) in 50 mL falcons. After discarding the supernatant, the pellet was resuspended in 50 mL ice cold buffer (100 mM MgCl₂) and incubated for 30 min on ice. After another round of centrifugation, the cell pellet was resuspended in appropriate amount of ice cold buffer (100 mM MgCl₂) and pooled into on falcon. After the final centrifugation step, the pellet was resuspended in ice cold storage buffer (CaCl₂/glycerol(15%)). The amount of storage buffer used was to increase the cell density 50-fold (e.g., 2 mL for 100 mL culture). The competent cells were either directly used or frozen in liquid nitrogen in 100 µL aliquots and stored at -80 C.

For transformation of chemically competent *E. coli* cells, 50 µL of cells were mixed with 0.2-2.5 µL of DNA (2.5 µL in case of a ligation product). After incubation on ice for 20 min, the suspension was heated to 42 °C for 45 s. Then, 750 µL of LB medium was added, followed by recovery of the cells for 1 h at 37 °C and 220 rpm. Finally, the cells were plated (using glass beads) on agar plates with appropriate supplements and incubated overnight at 37 °C.

4.2.1.2. Transformation of Electrocompetent Cells

For the preparation of electrocompetent *E. coli* cells, an overnight culture was grown (37 °C, 200 rpm) and used to inoculate 300 mL LB media (1:100). Cells were grown (37 °C, 200 rpm) until an OD₆₀₀ of 0.4 – 0.5 was reached. The culture was immediately chilled on ice and cells were harvested by centrifugation (4 °C, 5 min, 4000 x g) in 50 mL falcon tubes. Next, the supernatant was discarded, and the cells were washed three times (50 mL, 25 mL, 5 mL) with ice cold glycerol (10%), whereby the cells were pooled before the last washing step. Finally, cells were resuspended with ice-cold glycerol (10%). The amount of storage buffer used was to increase the cell density 50-fold (e.g., 2 mL glycerol for 100 mL culture). The cells were either directly used for electroporation or frozen in liquid nitrogen in 50 µL aliquots and stored at -80 °C until use.

For transformation of *E. coli* via electroporation, 50 µL of electrocompetent cells were mixed with 0.2 - 1 µL of DNA. The mixture was placed in a pre-chilled electroporation cuvette (1 mm gap width). Electroporation was performed with an electroporator (BioRad Gene Pulser Xcell) by applying an electrical pulse of 1.8 kV. For recovery, 750 mL LB medium were immediately added and the suspension was transferred to a sterile 1,5 mL microcentrifuge tube. After incubated at 37 °C (220 rpm) for 1 h, cells were plated (using glass beads) on agar plates with appropriate selection markers and incubated overnight at 37 °C.

4.2.2. Isolation of Plasmid DNA

Small-scale plasmid DNA isolation from *E. coli* cells was performed according to the the GeneJET Plasmid Mini-prep Kit protocol (ThermoFisher Scientific, Waltham, MA, USA).

Typically, 5 mL cell culture were used for small-scale plasmid extraction. The DNA was eluted with 30 μL elution buffer and stored at -20°C .

4.2.3. Polymerase Chain Reaction

4.2.3.1. Colony PCR

To verify insertion of DNA fragments into plasmids or to verify removal of genes from the chromosome, colony PCR was applied employing Taq polymerase (made in-house for plasmids or purchased from ThermoFisher Scientific, Waltham, MA, USA for chromosomal DNA). For this purpose, relevant colonies were picked and suspended each in 20 μL ddH₂O. In order to ensure cell lysis, an initial boiling step at 95°C for 5 min was used.

Component	Volume	Concentration
Colony	1 μL	-
10x Dream Taq Polymerase Buffer	2 μL	1x
dNTPs (10 mM)	0.4 μL	0.2 mM
Forward primer (10 μM)	1 μL	0.5 μM
Reverse primer (10 μM)	1 μL	0.5 μM
Taq polymerase (made in house)	0.4 μL	-
ddH ₂ O	adjust to 20 μL	-

PCR step	Temperature	Time	Cycles
Initial Denaturation	95°C	5 min	1
Denaturation	95°C	60 s	
Annealing	$55 - 65^{\circ}\text{C}$	30 s	30
Extension	72°C	60 s per 1 kb	
Final extension	72°C	5 min	1

4.2.3.2. Standard PCR

For DNA amplifications via standard PCR, the proofreading Phusion High-Fidelity DNA polymerase (ThermoFisher Scientific, Waltham, MA, USA) or Q5 High-Fidelity DNA Polymerase (New England Biolabs, Frankfurt am Main, Germany) was used.

Component	Volume	Concentration
Template	1 μL	2 ng μL^{-1}
5x Phusion HF or Q5 reaction buffer	10 μL	1x
dNTPs (10 mM)	1 μL	0.2 mM
Forward primer (10 μM)	2.5 μL	0.5 μM
Reverse primer (10 μM)	2.5 μL	0.5 μM
Phusion / Q5 polymerase	0.5 μL	-
ddH ₂ O	adjust to 50 μL	-

PCR step	Temperature	Time	Cycles
Initial Denaturation	98 °C	30 s	1
Denaturation	98 °C	12 s	
Annealing	60 - 72 °C	20 s	35
Extension	72 °C	20 s per 1 kb	
Final extension	72 °C	2 min	1

PCR products were analyzed for purity on agarose gels. PCR products were purified via GeneJET Gel Extraction Kit (ThermoFisher Scientific, Waltham, MA, USA) or via preparative gel electrophoresis and subsequent purification utilizing the same kit.

4.2.3.2.1. Site-Directed and Site-Saturation Mutagenesis

Point mutations were introduced by non-overlapping inverse PCR.³⁶⁸ Focused *MbPylRS* gene libraries were created also with non-overlapping inverse PCR, but randomization was performed using mutagenic primers (with NNK (N = A, T, G, or C; K = G or T)) at designated positions.

4.2.4. Library Screening

After library creation and transformation, 96 clones were picked and grown overnight in a 96-well plate in 100 μ L LB with 1% glucose and appropriate antibiotics. The next day a 96-well plate with 100 μ L ZYP-5052, appropriate antibiotics, and ncAAs was inoculated with 1 μ L culture, grown for 24 h and measured afterwards as stated above. The 96-well plate which was used for inoculation was sealed with aluminum foil and stored at 4 °C. From this plate, desired clones were analyzed via PCR gene amplification and sequencing of this PCR product afterwards. Calculations with the Toplib tool estimate the probability of finding the best performing variant to be 95.3% and finding at least one of the two best performing variants with a probability of 99.8% (using a yield of 85%, which is the lower limit of primer purity and therefore also assumed as the lower yield limit of created DNA constructs).²⁷⁶

4.2.5. In Vitro Modification of DNA

4.2.5.1. Plasmid Vector Construction

All plasmids were assembled by Golden Gate cloning and confirmed by DNA sequencing.³⁶⁹ Plasmids harboring the OTS (aaRS/tRNA^{Pyl}) were constructed by cloning the target aaRS gene into the pTECH vector (Addgene plasmid #104073).¹⁴³

4.2.5.2. Restriction Digest of DNA

For cloning procedures or in order to analyze plasmids, DNA was digested with restriction enzymes. Digests with the appropriate restriction enzymes were done according to the manufacturer's instructions at 37 °C. For an analytical digest, 200 – 500 ng of DNA were

incubated for 2 h in a total volume of 20 μ L. For analysis, 6x DNA loading dye was added to restriction digests, and the DNA was analyzed via gel electrophoresis. In standard preparative digests 1 μ g of DNA was digested for 2 h in a total volume of 50 μ L. To prevent re-ligation of vector backbone fragments FastAP Alkaline Phosphatase (ThermoFisher Scientific, Waltham, MA, USA) was used for dephosphorylation (for 15 min at 37 °C). The digested DNA was then either purified by GeneJET™ Gel Extraction Kit or by gel extraction after separation on an agarose gel also using the GeneJET™ Gel Extraction Kit (both ThermoFisher Scientific, Waltham, MA, USA).

4.2.5.3. Ligation of DNA fragments

For the ligation of digested DNA fragments, T4 DNA ligase was utilized. A typical 10 μ L ligation reaction was composed of 1 μ L T4 DNA-Ligase (5 Weiss U/ μ L, New England Biolabs, Frankfurt am Main, Germany), 1 μ L 10x ligation buffer as well as vector backbone, insert DNA and ddH₂O. Typically, 50 ng of vector DNA were ligated with the insert DNA in five- to ten-fold molar excess. Ligation was carried out at room temperature for 1 h. Afterwards, 2.5 μ L of the ligation mixture was used for transformation of chemically competent cells.

4.2.5.4. Recursive Directional Ligation

The monomers for the different ELP scaffolds were purchased by GeneArt (ThermoFisher Scientific, Waltham, MA, USA). The overall cloning strategy is described and depicted in **chapter 2.5.1**. For ELP construct creation the monomers were transferred to the pSB1C3 vector. The monomer elongation was performed by digesting the desired plasmid with one restriction enzyme (BglI) and subsequent dephosphorylation with alkaline phosphatase (FastAP, ThermoFisher Scientific, Waltham, MA, USA). Afterwards, the desired insert double digested (with BglI and PflMI) monomer was ligated into this construct to receive an elongated ELP scaffold. To check if the correct insert was introduced an analytical digest was performed with BglI and PstI. This was necessary because colony PCR with ELP construct bigger than the ELP(10x amber) constructed resulted in unsuccessful amplifications most probably due to the high GC content of the ELP gene. Several setups were tested (PCR with GC-enhancer with and without DMSO) but did not lead to a PCR-product. ELP scaffolds > 2000 bp (up to construct ELP(40x amber)) were also sequenced (from both sides). When the ELP scaffold reached the desired length, this construct was then double digested (with BglI and PflMI) and ligated into a pET-28a vector. This was also verified by double digest with BglI and XbaI. ELP scaffolds > 1000 bp (up to construct ELP(20x amber)) were also sequenced (5'→3').

4.2.6. Chromosomal Gene Deletion by Homologous Recombination

Chromosomal gene deletions were performed according to a modified method of Datsenko and Wanner³⁴². For this purpose, the Kan-resistance cassette including flippase recognition target (FRT) sites was amplified from genomic DNA of *E. coli* clones from the Keio single gene

knockout collection.³⁴⁴ The primers amplified the cassette and additionally of up to 200 bp homologous regions up- and downstream of the target genes. The PCR product was purified via gel extraction and 1 µg was used for transformation of electrocompetent B-95.ΔA cells carrying the pSIJ8 plasmid. The expression of the pSIJ8-encoded λ - Red recombination system was ensured by inducing protein expression with 0.2% arabinose (w/v) at an OD₆₀₀ of 0.3 for 30–45 min during preparation of these electrocompetent cells. After transformation with 250ng of the antibiotic resistance cassette PCR-product and 2 hours recovery (30 °C, 220 rpm), cells were plated on LB-Kan agar plates and incubated overnight at 30 °C. The homologous recombination of the FRT-*kan*-FRT cassette was verified by colony PCR. Removal of antibiotic markers was done at 30 °C in LB by inducing FLP with 50 mM L-rhamnose (final concentration) at an OD₆₀₀ between 0.1–0.4 for 4–6 h prior to plating. This was again checked via PCR. Verified clones were incubated at 42 °C o/n to remove of temperature-sensitive pSIJ8 plasmid which was also checked via PCR.

4.3. Protein Expression and Purification

4.3.1. Recombinant Protein Expression

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

4.3.2. Protein Purification

Harvested cell pellets were resuspended (50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 8.0) and lysed with B PER® Bacterial Protein Extraction Reagent (ThermoFisher Scientific, Waltham, MA, USA) according to their protocol, with addition of phenylmethanesulfonyl fluoride (PMSF, 1 mM final concentration), DNase and RNase. Cleared lysates were loaded onto a equilibrated Ni-NTA column and purified via a peristaltic pump (Pharmacia Biotech, Stockholm, SE). After washing with 10 column volumes of resuspension buffer, elution buffer (50 mM sodium phosphate, 300 mM NaCl, 500 mM imidazole, pH 8.0) was applied to elute the his-tagged target proteins. The first 2 mL (covering the dead volume) were discarded. Afterwards, the eluate (1 mL) was collected and dialyzed in cellulose film tubings against 1 L buffer (50 mM sodium phosphate, 300 mM NaCl, pH 8.0) for at least 2 h with three buffer changes. Concentrations of purified reporter proteins were determined by measuring the sfGFP chromophore absorption at 488 nm or absorption at 280 nm.

4.4. Biochemical Methods

4.4.1. Plate Reader Assays

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

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

4.4.1.2. Time-Resolved Chromophore Maturation of *ami*CP(Y63ONBY)

Proteins were expressed and purified according to **chapter 4.3**. After purification 100 μ L of protein solution with concentration of 1 mg/mL was transferred into glass MS vials (glass was preferred since the UV-light is strongly absorbed by plastic). Irritation with UV-light was conducted according to **chapter 4.4.3**. After irradiation the sample was transferred to a 96-well clear bottom plate, sealed with foil and monitored for 24 hours. Absorption was measured for 589 nm.

4.4.2. Antimicrobial activity assay

This assay was performed by Jessica H. Nickling (Technical University Berlin, Germany).

To determine the antimicrobial activity, an overnight culture of the nisin sensitive indicator strain *L. lactis* NZ9000 pNZ nisPT was incubated in M17 medium with 1% (w/v) glucose and 5 μ g/mL chloramphenicol at 30 °C without agitation.³⁷⁰ Fresh GM17 medium was inoculated and cells were grown at 30 °C until OD₆₀₀ of 0.4-0.6 was reached. 1 mL of the culture was added into 50 mL molten GM17-agar supplemented with chloramphenicol and poured into a

large petri dish. The wide end of a glass Pasteur pipette was used to create holes in the GM17-agar.

1 mL of the *E. coli* expression cultures was centrifuged (3 min, 7,000 x *g*) and resuspended in 500 μ L Na-P buffer (50 mM, pH 7,4). The Cells were lysated by sonification (Sonoplus HD3200, MS72 electrode) at 30% amplitude with puls of 1 s on and 5 s off for 3 min. Subsequently, cellular debris was removed by centrifugation (4 °C, 10 min, 13,000 x *g*), supernatants were diluted and normalized to 1 mL OD₆₀₀ = 0.6 or 1.0 relative to the harvested cell density. 40 μ L of each normalized sample were added into the holes of the indicator plate and incubated overnight at 30 °C. Lysate with Nisin containing non-decaged ONBY served as a negative control and chloramphenicol at 400 μ g/mL as positive control compound.

4.4.3. Decaging ONBY

For decaging proteins or whole cell lysates containing ONBY, samples were placed in a glass vial normally used for MS samples. A self-built 365 nm LED lamp (radiant flux ~720 mW) was used at a distance of ~ 3 mm for indicated periods of time.

4.5. Analytical Methods

4.5.1. Agarose Gel Electrophoresis for DNA Analysis

For the separation of double-stranded DNA fragments depending on their size, agarose gel electrophoresis in presence of ethidium bromide (EtBr) was performed. Analytical agarose gels were used to evaluate restriction digest of plasmids or PCR products. Preparative gels were used to purify desired DNA fragments by isolating them from the gel. Typically, 1 % agarose was dissolved in TAE buffer by heating. 0.5–1 μ L EtBr (10 mg mL⁻¹ in ethanol) was then added to 30 - 50 mL of dissolved agarose. For shorter nucleotide sequences of up to 100 bp, gels containing 2 % agarose were used. DNA samples were mixed with 6x DNA loading dye and applied on the gel. To separate DNA according to their size, a voltage of 80 V was applied and variable time, depending on the size of targeted DNA. The GeneRuler Mix Ladder or GeneRuler 1 kb DNA Ladder (ThermoFisher Scientific, Waltham, MA, USA) was used as a standard. Visualization of DNA fragments was achieved via illumination with UV light. In case of preparative gel electrophoresis, the desired DNA fragment was extracted from the gel and purified with GeneJET Gel Extraction Kit (ThermoFisher Scientific, Waltham, MA, USA).

4.5.2. Polyacrylamide Gel Electrophoresis for Protein Analysis

To analyze proteins according to their molecular mass, SDS polyacrylamide gel electrophoresis was performed. For separation, 12 % or 15 % acrylamide gels were used. Samples were mixed with 5x SDS loading dye, incubated (95 °C, 5 min) and loaded on the gel. As protein standard, PageRuler Prestained Protein Ladder (ThermoFisher Scientific,

Waltham, MA, USA) or PageRuler Unstained Protein Ladder (ThermoFisher Scientific, Waltham, MA, USA) were used. Electrophoresis was performed with SDS running buffer (190 mM glycine, 25 mM Tris, 3.5 mM SDS) at 80 V until the samples passed the collection gel. Immediately afterwards the voltage was increased to 120 V until the end of the run. Gels were stained with Coomassie staining solution overnight and destained with heated dH₂O where the water was changed after 30 min.

4.5.3. Analysis of Soluble and Insoluble Protein Fractions

E. coli strain BL21(DE) was used in 2 mL ZYP-5052 medium supplemented with 2 mM Sac (to match target protein expression conditions) in 14 mL round bottom tubes (Falcon, Corning Life Sciences) and appropriate antibiotics. The expression medium was inoculated with a starter culture (1:100) and incubated for 20 h at 37 °C, 200 rpm. 80 µL of culture were normalized to OD₆₀₀=6 via dilution and absorption measurements with an Ultrospec 6300 pro spectrophotometer (Amersham Biosciences). Cells were harvested by centrifugation and lysed with B-PER® Bacterial Protein Extraction Reagent (ThermoFisher Scientific Waltham, MA, USA) according to protocol, with addition of phenylmethanesulfonyl fluoride (1 mM final). Lysed cells were centrifuged (15,000 g, 5 min, RT) and the supernatant isolated (soluble protein fraction). The precipitate (insoluble protein fraction) was resolved (50 mM Tris-HCl, 7.5 M Urea, pH 8) with the same volume as the initial lysis buffer. Afterwards, soluble and insoluble protein fractions were combined with 5x SDS loading dye (80 mM TRIS pH 6.8, 10 % SDS, 12.5 % glycerol, 4 % (v/v) β-mercaptoethanol, 0.2 % (w/v) bromophenol blue) and 5 µL samples were used for 15 % acrylamide gels. PageRuler Prestained Protein Ladder (ThermoFisher Scientific, Waltham, MA, USA) or PageRuler Unstained Protein Ladder (ThermoFisher Scientific, Waltham, MA, USA) were used. Electrophoresis was performed at 80 V for 60 minutes and afterwards at 120 V with SDS running buffer (190 mM glycine, 25 mM Tris, 3.5 mM SDS). Gels were stained with Coomassie solution overnight and destained with dH₂O prior to documentation.

4.5.4. Determination of Protein Concentration

Protein concentrations were determined by measuring their absorbance at 280 nm or 488 nm for sfGFP variants. The molar extinction coefficient (ϵ_M) at 280 nm was calculated (Expasy ProtParam)³⁷¹ using the Edelhoch method,³⁷² with the extinction coefficients for Tryptophane, Tyrosine and Cystine determined by Pace *et al.*³⁷³. For calculations of concentrations sfGFP variants, the published extinction coefficient for the sfGFP chromophore absorption at 488 nm was used.²¹⁷

4.6. Biophysical Methods

4.6.1. ESI-MS

Intact protein mass measurements were performed by electrospray LC-MS on an Agilent 6530 QTOF instrument (Agilent, Santa Clara, CA, USA) after external calibration coupled with an Agilent 1260 HPLC system or Waters H-class QToF. Measurements with the latter were performed by Dr. Philipp Ochtrup or Christian Stieger (Leibniz-Forschungsinstitut für Molekulare Pharmakologie, Hackenberger Group).

4.6.1.1. *Agilent 6530 QToF*

This instrument was used for the measurements in **chapter 2.5**. Samples were infused at a flow rate of 0.3 - 0.5 mL min⁻¹ onto a gradient from 5 % Acetonitrile w/0.1 % formic acid in water to 80 % Acetonitrile w/0.1 % formic acid in water through a Discovery Bio Wide Pore C5 column, 2.1x100, 3 micron (Supelco analytical, Sigma-Aldrich, St. Lois, USA) over 20 minutes. Spectra deconvolution was performed with Agilent MassHunter Qualitative Analysis software version B.06.00 Bioconfirm Intact mass module, employing the maximum entropy deconvolution algorithm. Raw data were analyzed employing the maximum entropy deconvolution algorithm and plotted with Origin or QtiPlot.

4.6.1.2. *Waters Xevo G2-XS QTof*

This instrument was used for all other Intact protein mass measurements (SUMO-sfGFP variants). A Waters Acquity UPLC protein BEH C4 column (300 Å, 1.7 µm, 2.1 mm x 50 mm) was used and gradient at a flow rate of 0.3 mL/min: A: 0.01% formic acid in H₂O; B: 0.01% formic acid in MeCN. 5-95% B 0-6 min). Mass analysis was conducted with a Waters Xevo G2-XS QTof analyzer (positive mode, cone voltage=40 kV). Deconvoluted raw data were analyzed employing the maximum entropy deconvolution algorithm and plotted with QtiPlot.

5. References

- (1) Schwark, D. G.; Schmitt, M. A.; Fisk, J. D. Dissecting the Contribution of Release Factor Interactions to Amber Stop Codon Reassignment Efficiencies of the Methanocaldococcus Jannaschii Orthogonal Pair. *Genes (Basel)*. **2018**, *9* (11). <https://doi.org/10.3390/genes9110546>.
- (2) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. Coli. *J. Mol. Biol.* **2010**, *395* (2), 361–374. <https://doi.org/10.1016/j.jmb.2009.10.030>.
- (3) Baumann, T.; Hauf, M.; Richter, F.; Albers, S.; Möglich, A.; Ignatova, Z.; Budisa, N. Computational Aminoacyl-TRNA Synthetase Library Design for Photocaged Tyrosine. *Int. J. Mol. Sci.* **2019**, *20* (9). <https://doi.org/10.3390/ijms20092343>.
- (4) Fruton, J. S. *Proteins, Enzymes, Genes: The Interplay of Chemistry and Biology*; Yale University Press: New Haven and London, 1999.
- (5) Sumner, J. B. The Isolation and Crystallization of the Enzyme Urease. *J. Biol. Chem.* **1926**, *69*, 435–441.
- (6) Vickery, H. B.; Schmidt, C. L. A. The History of the Discovery of the Amino Acids. *Chem. Rev.* **1931**, *9* (2), 169–318. <https://doi.org/10.1021/cr60033a001>.
- (7) Fruton, J. S. Early Theories of Protein Structure. *Ann. N. Y. Acad. Sci.* **1979**, *325* (1), 1–20. <https://doi.org/10.1111/j.1749-6632.1979.tb14125.x>.
- (8) Fischer, E. Syntheses in the Purine and Sugar Group. *Nobel Lect.* **1902**.
- (9) Hofmeister, F. Über Bau Und Gruppierung Der Eiweisskörper. *Ergebnisse der Physiol.* **1902**, *1*, 759–802. <https://doi.org/10.1007/BF02323641>.
- (10) Svedberg, T. Mass and Size of Protein Molecules. *Nature* **1929**, *123* (3110), 871.
- (11) Sanger, F.; Tuppy, H. The Amino-Acid Sequence in the Phenylalanyl Chain of Insulin. I. The Identification of Lower Peptides from Partial Hydrolysates. *Biochem. J.* **1951**, *49*, 463–481. <https://doi.org/10.1042/bj0490463>.
- (12) Sanger, F.; Thompson, E. O. P. The Amino-Acid Sequence in the Glycyl Chain of Insulin. II. The Investigation of Peptides from Enzymic Hydrolysates. *Biochem. J.* **1953**, *53*, 366–374. <https://doi.org/10.1042/bj0530366>.
- (13) Griffith, F. The Significance of Pneumococcal Types. *J. Hyg. (Lond)*. **1928**, *27* (2), 113–159. <https://doi.org/10.1017/S0022172400031879>.
- (14) Downie, A. W. Pneumococcal Transformation-A Backward View Fourth Griffith Memorial Lecture. *J. Gen. Microbiol.* **1972**, *73* (1), 1–11. <https://doi.org/10.1099/00221287-73-1-1>.
- (15) Avery, O. T.; MacLeod, C. M.; 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* (2), 137–158. <https://doi.org/10.1084/jem.79.2.137>.
- (16) Hershey, A. D.; Chase, M. INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE. *J. Gen. Physiol.* **1952**, *36* (1), 39–56. <https://doi.org/10.1085/jgp.36.1.39>.
- (17) Crick, F. H. C. On Protein Synthesis. *Symp. Soc. Exp. Biol.* **1958**, *12*, 138–163.
- (18) Nelson, D. L.; Cox, M. M. Lehninger Principles of Biochemistry, 6th Ed. In *Lehninger Principles of Biochemistry, 6th ed.*; W. H. Freeman and Company: New York, 2012.
- (19) Crick, F. Central Dogma of Molecular Biology. *Nature* **1970**, *227* (5258), 561–563. https://doi.org/10.1007/978-1-4020-6754-9_2672.

- (20) Judson, H. F. The Eighth Day of Creation: Makers of the Revolution in Biology, Commemorative Edition. In *The Eighth Day of Creation: Makers of the Revolution in Biology, Commemorative Edition*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 1996; p 333.
- (21) Biro, J. C. Suggestion to Upgrade the Canonical Concept of Translation. *J. Proteomics Bioinforma.* **2014**, 7 (5), 112–120. <https://doi.org/10.4172/0974-276X.1000311>.
- (22) Kubyshkin, V.; Acevedo-Rocha, C. G.; Budisa, N. On Universal Coding Events in Protein Biogenesis. *BioSystems* **2018**, 164, 16–25. <https://doi.org/10.1016/j.biosystems.2017.10.004>.
- (23) Buhr, F.; Jha, S.; Thommen, M.; Mittelstaet, J.; Kutz, F.; Schwalbe, H.; Rodnina, M. V.; Komar, A. A. Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. *Mol. Cell* **2016**, 61 (3), 341–351. <https://doi.org/10.1016/j.molcel.2016.01.008>.
- (24) Maraia, R. J.; Iben, J. R. Different Types of Secondary Information in the Genetic Code. *RNA* **2014**, 20 (7), 977–984. <https://doi.org/10.1261/rna.044115.113>.
- (25) Baltimore, D. Viral RNA-Dependent DNA Polymerase: RNA-Dependent DNA Polymerase in Virions of RNA Tumour Viruses. *Nature* **1970**, 226 (5252), 1209–1211. <https://doi.org/10.1038/2261209a0>.
- (26) Sigurdson, C. J.; Bartz, J. C.; Glatzel, M. Cellular and Molecular Mechanisms of Prion Disease. *Annu. Rev. Pathol. Mech. Dis.* **2019**, 14 (1), 497–516. <https://doi.org/10.1146/annurev-pathmechdis-012418-013109>.
- (27) Terry, C.; Wadsworth, J. D. F. Recent Advances in Understanding Mammalian Prion Structure: A Mini Review. *Front. Mol. Neurosci.* **2019**, 12 (July), 1–7. <https://doi.org/10.3389/fnmol.2019.00169>.
- (28) Mendel, G. Versuche Über Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereines Brünn* **1866**, 4, 3–47.
- (29) Vries, H. de. *Intracellulare Pangenesis*; Gustav Fischer: Jena, 1889. <https://doi.org/10.5962/bhl.title.94284>.
- (30) Weismann, A. *Das Keimplasma. Eine Theorie Der Vererbung*, First.; Gustav Fischer: Jena, 1892.
- (31) Haeckel, E. *Generelle Morphologie Der Organismen*; Georg Reimer: Berlin, 1866. <https://doi.org/10.1515/9783110848281>.
- (32) Miescher, F. Ueber Die Chemische Zusammensetzung Der Eiterzellen. *Med. Untersuchungen* **1871**, 4, 441–462.
- (33) Dahm, R. Discovering DNA: Friedrich Miescher and the Early Years of Nucleic Acid Research. *Hum. Genet.* **2008**, 122 (6), 565–581. <https://doi.org/10.1007/s00439-007-0433-0>.
- (34) Kossel, A. Ueber Das Nuclein Der Hefe. *Zeitschrift für Physiol. Chemie* **1879**, 3, 284–291.
- (35) Piccard, J. Ueber Protamin, Guanin Und Sarkin, Als Bestandtheile Des Lachssperma. *Berichte der Dtsch. Chem. Gesellschaft* **1874**, 7 (2), 1714–1719. <https://doi.org/10.1002/cber.187400702234>.
- (36) Kossel, A. Ueber Eine Neue Base Aus Dem Thierkörper. *Berichte der Dtsch. Chem. Gesellschaft* **1885**, 18 (1), 79–81. <https://doi.org/10.1002/cber.18850180119>.
- (37) Kossel, A. Weitere Beiträge Zur Chemie Des Zellkerns. *Zeitschrift für Physiol. Chemie* **1886**, 10, 248–246.
- (38) Kossel, A.; Neumann, A. Ueber Das Thymin, Ein Spaltungsproduct Der Nucleänsäure.

- Berichte der Dtsch. Chem. Gesellschaft* **1893**, 26 (3), 2753–2756. <https://doi.org/10.1002/cber.18930260379>.
- (39) Kossel, A.; Neumann, A. Darstellung Und Spaltungsprodukte Der Nucleinsäure (Adenylsäure). *Berichte der Dtsch. Chem. Gesellschaft* **1894**, 27 (2), 2215–2222. <https://doi.org/10.1002/cber.189402702206>.
- (40) Kossel, A.; Neumann, A. Ueber Nucleinsäure Und Thyminsäure. **1896**, 22, 74–81.
- (41) Altmann, R. Ueber Nucleinsäuren. *Arch. für Anat. und Physiol. Physiol. Abteilung* **1889**, 524–536.
- (42) Ascoli, A. Ueber Ein Neues Spaltungsproduct Des Hefenucleins. *Zeitschrift für Physiol. Chemie* **1900**, 31, 161–164.
- (43) Levene, P. A.; Jacobs, W. A. Über Die Pentose in Den Nucleinsäuren. *Berichte der Dtsch. Chem. Gesellschaft* **1909**, 42 (3), 3247–3251. <https://doi.org/10.1002/cber.19090420351>.
- (44) Hargittai, I. The Tetranucleotide Hypothesis: A Centennial. *Struct. Chem.* **2009**, 20 (5), 753–756. <https://doi.org/10.1007/s11224-009-9497-x>.
- (45) Levene, P. A.; Mandel, J. A. Über Die Konstitution Der Thymo-Nucleinsäure. *Berichte der Dtsch. Chem. Gesellschaft* **1908**, 41 (2), 1905–1909. <https://doi.org/10.1002/cber.19080410266>.
- (46) Frixione, E.; Ruiz-Zamarripa, L. The “Scientific Catastrophe” in Nucleic Acids Research That Boosted Molecular Biology. *J. Biol. Chem.* **2019**, 294 (7), 2249–2255. <https://doi.org/10.1074/jbc.CL119.007397>.
- (47) Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, 27 (4), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>.
- (48) Aaronson, S.; Carroll, S. M.; Ouellette, L. *Quantifying the Rise and Fall of Complexity in Closed Systems: The Coffee Automaton*; 2014.
- (49) Chargaff, E.; Zamenhof, S.; Green, C. Human Desoxypentose Nucleic Acid: Composition of Human Desoxypentose Nucleic Acid. *Nature* **1950**, 165 (4202), 756–757. <https://doi.org/10.1038/165756b0>.
- (50) Wilkins, M. H. F.; Gosling, R. G.; Seeds, W. E. Physical Studies of Nucleic Acid: Nucleic Acid: An Extensible Molecule? *Nature* **1951**, 167 (4254), 759–760. <https://doi.org/10.1038/167759a0>.
- (51) Franklin, R. E.; Gosling, R. G. The Structure of Sodium Thymonucleate Fibres. I. The Influence of Water Content. *Acta Crystallogr.* **1953**, 6 (8), 673–677. <https://doi.org/10.1107/s0365110x53001939>.
- (52) Watson, J. D.; Crick, F. H. C. Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* **1953**, 171 (4361), 964–967. <https://doi.org/10.1038/171964b0>.
- (53) Crick, F. H. C.; Barnett, L.; Brenner, S.; Watts-Tobin, R. J. General Nature of the Genetic Code for Proteins. *Nature* **1961**, 192 (4809), 1227–1232. <https://doi.org/10.1038/1921227a0>.
- (54) Nirenberg, M. W.; Matthaei, J. H. The Dependence of Cell-Free Protein Synthesis in *E. Coli* upon Naturally Occurring or Synthetic Polyribonucleotides. *Proc. Natl. Acad. Sci.* **1961**, 47 (10), 1588–1602. <https://doi.org/10.1073/pnas.47.10.1588>.
- (55) Nirenberg, M. Historical Review: Deciphering the Genetic Code – a Personal Account. *Trends Biochem. Sci.* **2004**, 29 (1), 46–54. <https://doi.org/10.1016/j.tibs.2003.11.009>.
- (56) Judson, H. F. The Eighth Day of Creation: Makers of the Revolution in Biology, Commemorative Edition. In *No Title The Eighth Day of Creation: Makers of the Revolution in Biology, Commemorative Edition*; Cold Spring Harbor Laboratory Press:

- Cold Spring Harbor, New York, 1996; p 470.
- (57) Woese, C. R. *The Genetic Code the Molecular Basis for Genetic Expression*; Harper & Row: New York, NY, 1967.
- (58) Marshall, R. E.; Caskey, C. T.; Nirenberg, M. Fine Structure of RNA Codewords Recognized by Bacterial, Amphibian, and Mammalian Transfer RNA. *Science (80-)*. **1967**, *155* (3764), 820–826. <https://doi.org/10.1126/science.155.3764.820>.
- (59) Woese, C. R.; Hinegardner, R. T.; Engelberg, J. Universality in the Genetic Code. *Science (80-)*. **1964**, *144* (3621), 1030–1031. <https://doi.org/10.1126/science.144.3621.1030>.
- (60) Watson, J. D.; Baker, T. A.; Bell, S. P.; Gann, A.; Levine, M.; Losick, R. *Molecular Biology of the Gene*, 7th Ed.; Pearson: London, 2013; pp 587–590.
- (61) Böck, A.; Forchhammer, K.; Heider, J.; Leinfelder, W.; Sawers, G.; Veprek, B.; Zinoni, F. Selenocysteine: The 21st Amino Acid. *Mol. Microbiol.* **1991**, *5* (3), 515–520. <https://doi.org/10.1111/j.1365-2958.1991.tb00722.x>.
- (62) Srinivasan, G.; James, C. M.; Krzycki, J. A. Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA. *Science (80-)*. **2002**, *296* (5572), 1459–1462. <https://doi.org/10.1126/science.1069588>.
- (63) Haig, D.; Hurst, L. D. A Quantitative Measure of Error Minimization in the Genetic Code. *J. Mol. Evol.* **1991**, *33* (5), 412–417. <https://doi.org/10.1007/BF02103132>.
- (64) Saier, M. H. Understanding the Genetic Code. *J. Bacteriol.* **2019**, *201* (15), 1–12. <https://doi.org/10.1128/JB.00091-19>.
- (65) Freeland, S. J.; Hurst, L. D. The Genetic Code Is One in a Million. *J. Mol. Evol.* **1998**, *47* (3), 238–248. <https://doi.org/10.1007/PL00006381>.
- (66) Goodarzi, H.; Nejad, H. A.; Torabi, N. On the Optimality of the Genetic Code, with the Consideration of Termination Codons. *Biosystems* **2004**, *77* (1–3), 163–173. <https://doi.org/10.1016/j.biosystems.2004.05.031>.
- (67) Lagerkvist, U. “Two out of Three”: An Alternative Method for Codon Reading. *Proc. Natl. Acad. Sci.* **1978**, *75* (4), 1759–1762. <https://doi.org/10.1073/pnas.75.4.1759>.
- (68) Lehmann, J.; Libchaber, A. Degeneracy of the Genetic Code and Stability of the Base Pair at the Second Position of the Anticodon. *RNA* **2008**, *14* (7), 1264–1269. <https://doi.org/10.1261/rna.1029808>.
- (69) Lagerkvist, U. Unorthodox Codon Reading and the Evolution of the Genetic Code. *Cell* **1981**, *23* (2), 305–306. [https://doi.org/10.1016/0092-8674\(81\)90124-0](https://doi.org/10.1016/0092-8674(81)90124-0).
- (70) Trinquier, G.; Sanejouand, Y. H. Which Effective Property of Amino Acids Is Best Preserved by the Genetic Code? *Protein Eng. Des. Sel.* **1998**, *11* (3), 153–169. <https://doi.org/10.1093/protein/11.3.153>.
- (71) Tlusty, T. A Colorful Origin for the Genetic Code: Information Theory, Statistical Mechanics and the Emergence of Molecular Codes. *Phys. Life Rev.* **2010**, *7* (3), 362–376. <https://doi.org/10.1016/j.plrev.2010.06.002>.
- (72) Miller, S. L. A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science (80-)*. **1953**, *117* (3046), 528–529. <https://doi.org/10.1126/science.117.3046.528>.
- (73) Patel, B. H.; Percivalle, C.; Ritson, D. J.; Duffy, C. D.; Sutherland, J. D. Common Origins of RNA, Protein and Lipid Precursors in a Cyanosulfidic Protometabolism. *Nat. Chem.* **2015**, *7* (4), 301–307. <https://doi.org/10.1038/nchem.2202>.
- (74) Damer, B.; Deamer, D. Coupled Phases and Combinatorial Selection in Fluctuating Hydrothermal Pools: A Scenario to Guide Experimental Approaches to the Origin of

- Cellular Life. *Life* **2015**, *5* (1), 872–887. <https://doi.org/10.3390/life5010872>.
- (75) Kitadai, N.; Maruyama, S. Origins of Building Blocks of Life: A Review. *Geosci. Front.* **2018**, *9* (4), 1117–1153. <https://doi.org/10.1016/j.gsf.2017.07.007>.
- (76) Perunov, N.; Marsland, R. A.; England, J. L. Statistical Physics of Adaptation. *Phys. Rev. X* **2016**, *6* (2), 021036. <https://doi.org/10.1103/PhysRevX.6.021036>.
- (77) Vasas, V.; Fernando, C.; Santos, M.; Kauffman, S.; Szathmáry, E. Evolution before Genes. *Biol. Direct* **2012**, *7* (1), 1. <https://doi.org/10.1186/1745-6150-7-1>.
- (78) Horowitz, J. M.; England, J. L. Spontaneous Fine-Tuning to Environment in Many-Species Chemical Reaction Networks. *Proc. Natl. Acad. Sci.* **2017**, *114* (29), 7565–7570. <https://doi.org/10.1073/pnas.1700617114>.
- (79) England, J. L. Dissipative Adaptation in Driven Self-Assembly. *Nat. Nanotechnol.* **2015**, *10* (11), 919–923. <https://doi.org/10.1038/nnano.2015.250>.
- (80) Hordijk, W.; Kauffman, S. A.; Steel, M. Required Levels of Catalysis for Emergence of Autocatalytic Sets in Models of Chemical Reaction Systems. *Int. J. Mol. Sci.* **2011**, *12* (5), 3085–3101. <https://doi.org/10.3390/ijms12053085>.
- (81) Dyson, F. J. A Model for the Origin of Life. *J. Mol. Evol.* **1982**, *18* (5), 344–350. <https://doi.org/10.1007/BF01733901>.
- (82) Kauffman, S. A. Autocatalytic Sets of Proteins. *J. Theor. Biol.* **1986**, *119* (1), 1–24. [https://doi.org/10.1016/S0022-5193\(86\)80047-9](https://doi.org/10.1016/S0022-5193(86)80047-9).
- (83) Lincoln, T. A.; Joyce, G. F. Self-Sustained Replication of an RNA Enzyme. *Science* (80-.). **2009**, *323* (5918), 1229–1232. <https://doi.org/10.1126/science.1167856>.
- (84) Cronin, L.; Walker, S. I. Beyond Prebiotic Chemistry. *Science* (80-.). **2016**, *352* (6290), 1174–1175. <https://doi.org/10.1126/science.aaf6310>.
- (85) Hordijk, W.; Steel, M.; Kauffman, S. The Structure of Autocatalytic Sets: Evolvability, Enablement, and Emergence. *Acta Biotheor.* **2012**, *60* (4), 379–392. <https://doi.org/10.1007/s10441-012-9165-1>.
- (86) Hordijk, W.; Hein, J.; Steel, M. Autocatalytic Sets and the Origin of Life. *Entropy* **2010**, *12* (7), 1733–1742. <https://doi.org/10.3390/e12071733>.
- (87) Woese, C. R.; Fox, G. E. The Concept of Cellular Evolution. *J. Mol. Evol.* **1977**, *10* (1), 1–6. <https://doi.org/10.1007/BF01796132>.
- (88) Thurner, S.; Hanel, R.; Klimek, P. Introduction to the Theory of Complex Systems, 1st Ed. In *Introduction to the Theory of Complex Systems, 1st ed.*; Oxford University Press: Oxford, 2018; pp 22–23.
- (89) Thurner, S.; Hanel, R.; Klimek, P. Introduction to the Theory of Complex Systems, 1st Ed. In *Introduction to the Theory of Complex Systems, 1st ed.*; Oxford University Press: Oxford, 2018; p 5.
- (90) Gould, S. J. *Wonderful Life*, 1st ed.; W. W. Norton & Co.: New York, 1989.
- (91) Thurner, S.; Hanel, R.; Klimek, P. Introduction to the Theory of Complex Systems, 1st Ed. In *Introduction to the Theory of Complex Systems, 1st ed.*; Oxford University Press: Oxford, 2018; p 6.
- (92) Zwirn, H.; Delahaye, J.-P. Unpredictability and Computational Irreducibility. In *Irreducibility and Computational Equivalence*; Zenil, H., Ed.; Springer-Verlag: Berlin Heidelberg, 2013; pp 273–295. https://doi.org/10.1007/978-3-642-35482-3_19.
- (93) Beckage, B.; Kauffman, S.; Gross, L. J.; Zia, A.; Koliba, C. More Complex Complexity: Exploring the Nature of Computational Irreducibility across Physical, Biological, and Human Social Systems. In *Irreducibility and Computational Equivalence*; Zenil, H., Ed.; Springer-Verlag: Berlin Heidelberg, 2013; pp 79–88. <https://doi.org/10.1007/978-3-642->

- 35482-3_7.
- (94) Israeli, N.; Goldenfeld, N. Computational Irreducibility and the Predictability of Complex Physical Systems. *Phys. Rev. Lett.* **2004**, *92* (7), 074105. <https://doi.org/10.1103/PhysRevLett.92.074105>.
- (95) Crick, F. H. C. The Origin of the Genetic Code. *J. Mol. Biol.* **1968**, *38* (3), 367–379. [https://doi.org/10.1016/0022-2836\(68\)90392-6](https://doi.org/10.1016/0022-2836(68)90392-6).
- (96) Koonin, E. V.; Novozhilov, A. S. Origin and Evolution of the Universal Genetic Code. *Annu. Rev. Genet.* **2017**, *51* (1), 45–62. <https://doi.org/10.1146/annurev-genet-120116-024713>.
- (97) Koonin, E. Frozen Accident Pushing 50: Stereochemistry, Expansion, and Chance in the Evolution of the Genetic Code. *Life* **2017**, *7* (2), 22. <https://doi.org/10.3390/life7020022>.
- (98) Knight, R. D.; Freeland, S. J.; Landweber, L. F. Selection, History and Chemistry: The Three Faces of the Genetic Code. *Trends Biochem. Sci.* **1999**, *24* (6), 241–247. [https://doi.org/10.1016/S0968-0004\(99\)01392-4](https://doi.org/10.1016/S0968-0004(99)01392-4).
- (99) Kauffman, S. A. A World Beyond Physics. In *A World Beyond Physics*; Oxford University Press: Oxford, 2019; pp 2–4.
- (100) Jacob, F. Evolution and Tinkering. *Science (80-)*. **1977**, *196* (4295), 1161–1166. <https://doi.org/10.1126/science.860134>.
- (101) Kauffman, S. A. Prolegomenon to Patterns in Evolution. *Biosystems* **2014**, *123*, 3–8. <https://doi.org/10.1016/j.biosystems.2014.03.004>.
- (102) Berg, J. M.; Tymoczko, J. L.; Gatto, G. J. jr.; Stryer, L. *Biochemie*, 8th ed.; Springer-Verlag: Berlin Heidelberg, 2018.
- (103) Dirheimer, G.; Keith, G.; Dumas, P.; Westhof, E. *TRNA: Structure, Biosynthesis, and Function*; Söll, D., RajBhandary, U., Eds.; American Society of Microbiology: Washington, DC, 1995.
- (104) Berg, J. M.; Tymoczko, J. L.; Gatto, G. J. jr.; Stryer, L. *Biochemistry*, 8th ed.; W. H. Freeman and Company: New York, 2015.
- (105) Lodish, H.; Berk, A.; Kaiser, C. A.; Krieger, M.; Bretscher, A.; Ploegh, H.; Amon, A.; Martin, K. C. *Molecular Cell*, 8th ed.; W. H. Freeman and Company: New York, 2016.
- (106) Crick, F. H. C. Codon—Anticodon Pairing: The Wobble Hypothesis. *J. Mol. Biol.* **1966**, *19* (2), 548–555. [https://doi.org/10.1016/S0022-2836\(66\)80022-0](https://doi.org/10.1016/S0022-2836(66)80022-0).
- (107) Agris, P. F. Wobble Position Modified Nucleosides Evolved to Select Transfer RNA Codon Recognition: A Modified-Wobble Hypothesis. *Biochimie* **1991**, *73* (11), 1345–1349. [https://doi.org/10.1016/0300-9084\(91\)90163-U](https://doi.org/10.1016/0300-9084(91)90163-U).
- (108) Giege, R.; Sissler, M.; Florentz, C. Universal Rules and Idiosyncratic Features in TRNA Identity. *Nucleic Acids Res.* **1998**, *26* (22), 5017–5035. <https://doi.org/10.1093/nar/26.22.5017>.
- (109) Graw, J. *Genetik*, 6th ed.; Springer-Verlag: Berlin Heidelberg, 2015.
- (110) Blattner, F. R.; Plunkett, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y. The Complete Genome Sequence of Escherichia Coli K-12. *Science (80-)*. **1997**, *277* (5331), 1453–1462. <https://doi.org/10.1126/science.277.5331.1453>.
- (111) Chan, P. P.; Lowe, T. M. GtRNADB 2.0: An Expanded Database of Transfer RNA Genes Identified in Complete and Draft Genomes. *Nucleic Acids Res.* **2016**, *44* (D1), D184–D189. <https://doi.org/10.1093/nar/gkv1309>.

- (112) Tuller, T.; Carmi, A.; Vestsigian, K.; Navon, S.; Dorfan, Y.; Zaborske, J.; Pan, T.; Dahan, O.; Furman, I.; Pilpel, Y. An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell* **2010**, *141* (2), 344–354. <https://doi.org/10.1016/j.cell.2010.03.031>.
- (113) Dong, H.; Nilsson, L.; Kurland, C. G. Co-Variation of tRNA Abundance and Codon Usage In *Escherichia Coli* at Different Growth Rates. *J. Mol. Biol.* **1996**, *260* (5), 649–663. <https://doi.org/10.1006/jmbi.1996.0428>.
- (114) Berg, O. G.; Kurland, C. . Growth Rate-Optimised tRNA Abundance and Codon Usage. *J. Mol. Biol.* **1997**, *270* (4), 544–550. <https://doi.org/10.1006/jmbi.1997.1142>.
- (115) Pütz, J.; Puglisi, J. D.; Florentz, C.; Giegé, R. Identity Elements for Specific Aminoacylation of Yeast tRNA^{Asp} by Cognate Aspartyl-tRNA Synthetase. *Science (80-.)*. **1991**, *252* (5013), 1696–1699. <https://doi.org/10.1126/science.2047878>.
- (116) Lofffield, R. B.; Vanderjagt, D. The Frequency of Errors in Protein Biosynthesis. *Biochem. J.* **1972**, *128* (5), 1353–1356. <https://doi.org/10.1042/bj1281353>.
- (117) Dubios, D.; Lapointe, J.; Sekine, S. *The Aminoacyl-tRNA Synthetase*; Ibba, M., Francklyn, C., Cusack, S., Eds.; Landes Bioscience: Georgetown, 2005.
- (118) Ibba, M.; Söll, D. Aminoacyl-tRNA Synthesis. *Annu. Rev. Biochem.* **2000**, *69* (1), 617–650. <https://doi.org/10.1146/annurev.biochem.69.1.617>.
- (119) Perona, J. J.; Gruic-Sovulj, I. *Aminoacyl-tRNA Synthetases in Biology and Medicin*, 1st ed.; Kim, S., Ed.; Springer-Verlag: Berlin Heidelberg, 2014.
- (120) Cameron, D. E.; Bashor, C. J.; Collins, J. J. A Brief History of Synthetic Biology. *Nat. Rev. Microbiol.* **2014**, *12* (5), 381–390. <https://doi.org/10.1038/nrmicro3239>.
- (121) Erlich, Y.; Zielinski, D. DNA Fountain Enables a Robust and Efficient Storage Architecture. *Science (80-.)*. **2017**, *355* (6328), 950–954. <https://doi.org/10.1126/science.aaj2038>.
- (122) Benenson, Y. Biomolecular Computing Systems: Principles, Progress and Potential. *Nat. Rev. Genet.* **2012**, *13* (7), 455–468. <https://doi.org/10.1038/nrg3197>.
- (123) Dzionek, A.; Wojcieszynska, D.; Guzik, U. Natural Carriers in Bioremediation: A Review. *Electron. J. Biotechnol.* **2016**, *23*, 28–36. <https://doi.org/10.1016/j.ejbt.2016.07.003>.
- (124) Mukai, T.; Lajoie, M. J.; Englert, M.; Söll, D. Rewriting the Genetic Code. *Annu. Rev. Microbiol.* **2017**, *71* (1), 557–577. <https://doi.org/10.1146/annurev-micro-090816-093247>.
- (125) Chin, J. W. Expanding and Reprogramming the Genetic Code. *Nature* **2017**, *550* (7674), 53–60. <https://doi.org/10.1038/nature24031>.
- (126) Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, *79* (1), 413–444. <https://doi.org/10.1146/annurev.biochem.052308.105824>.
- (127) Groff, D.; Thielges, M. C.; Cellitti, S.; Schultz, P. G.; Romesberg, F. E. Efforts toward the Direct Experimental Characterization of Enzyme Microenvironments: Tyrosine 100 in Dihydrofolate Reductase. *Angew. Chemie - Int. Ed.* **2009**. <https://doi.org/10.1002/anie.200806239>.
- (128) Baumann, T.; Hauf, M.; Schildhauer, F.; Eberl, K. B.; Durkin, P. M.; Deniz, E.; Löffler, J. G.; Acevedo-Rocha, C. G.; Jaric, J.; Martins, B. M.; Dobbek, H.; Bredenbeck, J.; Budisa, N. Site-Resolved Observation of Vibrational Energy Transfer Using a Genetically Encoded Ultrafast Heater. *Angew. Chemie - Int. Ed.* **2019**. <https://doi.org/10.1002/anie.201812995>.
- (129) Minnihan, E. C.; Young, D. D.; Schultz, P. G.; Stubbe, J. Incorporation of Fluorotyrosines into Ribonucleotide Reductase Using an Evolved, Polyspecific Aminoacyl-tRNA

- Synthetase. *J. Am. Chem. Soc.* **2011**. <https://doi.org/10.1021/ja207719f>.
- (130) Li, J. C.; Nastertorabi, F.; Xuan, W.; Han, G. W.; Stevens, R. C.; Schultz, P. G. A Single Reactive Noncanonical Amino Acid Is Able to Dramatically Stabilize Protein Structure. *ACS Chem. Biol.* **2019**, *14* (6), 1150–1153. <https://doi.org/10.1021/acscchembio.9b00002>.
- (131) Agostini, F.; Völler, J. S.; Kokschi, B.; Acevedo-Rocha, C. G.; Kubyshkin, V.; Budisa, N. Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angew. Chemie - Int. Ed.* **2017**, *56* (33), 9680–9703. <https://doi.org/10.1002/anie.201610129>.
- (132) Drienovská, I.; Mayer, C.; Dulson, C.; Roelfes, G. SI_A Designer Enzyme for Hydrazone and Oxime Formation Featuring an Unnatural Catalytic Aniline Residue. *Nat. Chem.* **2018**. <https://doi.org/10.1038/s41557-018-0082-z>.
- (133) Burke, A. J.; Lovelock, S. L.; Frese, A.; Crawshaw, R.; Ortmayer, M.; Dunstan, M.; Levy, C.; Green, A. P. Design and Evolution of an Enzyme with a Non-Canonical Organocatalytic Mechanism. *Nature* **2019**, *570* (7760), 219–223. <https://doi.org/10.1038/s41586-019-1262-8>.
- (134) Walsh, C. T. *Postranslational Modification of Proteins: Expanding Nature's Inventory*, 1st ed.; Roberts & Company Publishers: Greenwood Village, 2005.
- (135) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angew. Chemie - Int. Ed.* **2005**, *44* (45), 7342–7372. <https://doi.org/10.1002/anie.200501023>.
- (136) Hecht, S. M.; Alford, B. L.; Kuroda, Y.; Kitano, S. “Chemical Aminoacylation” of tRNA's. *J. Biol. Chem.* **1978**, *253* (13), 4517–4520. [https://doi.org/10.1016/S0021-9258\(17\)30417-9](https://doi.org/10.1016/S0021-9258(17)30417-9).
- (137) Heckler, T. G.; Chang, L. H.; Zama, Y.; Naka, T.; Chorghade, M. S.; Hecht, S. M. T4 RNA Ligase Mediated Preparation of Novel “Chemically Misacylated” tRNAPhes. *Biochemistry* **1984**, *23* (7), 1468–1473. <https://doi.org/10.1021/bi00302a020>.
- (138) Noren, C.; Anthony-Cahill, S.; Griffith, M.; Schultz, P. G. A General Method for Site-Specific Incorporation of Unnatural Amino Acids into Proteins. *Science (80-.)*. **1989**, *244* (4901), 182–188. <https://doi.org/10.1126/science.2649980>.
- (139) Kimmerlin, T.; Seebach, D. ‘100 Years of Peptide Synthesis’: Ligation Methods for Peptide and Protein Synthesis with Applications to β -Peptide Assemblies*. *J. Pept. Res.* **2005**, *65* (2), 229–260. <https://doi.org/10.1111/j.1399-3011.2005.00214.x>.
- (140) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. Encoding Multiple Unnatural Amino Acids via Evolution of a Quadruplet-Decoding Ribosome. *Nature* **2010**, *464* (7287), 441–444. <https://doi.org/10.1038/nature08817>.
- (141) Furter, R. Expansion of the Genetic Code: Site-Directed p-Fluoro-Phenylalanine Incorporation in Escherichia Coli. *Protein Sci.* **1998**, *7* (2), 419–426. <https://doi.org/10.1002/pro.5560070223>.
- (142) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the Genetic Code of Escherichia Coli. *Science (80-.)*. **2001**, *292* (5516), 498–500. <https://doi.org/10.1126/science.1060077>.
- (143) Bryson, D. I.; Fan, C.; Guo, L. T.; Miller, C.; Söll, D.; Liu, D. R. Continuous Directed Evolution of Aminoacyl-tRNA Synthetases. *Nat. Chem. Biol.* **2017**, *13* (12), 1253–1260. <https://doi.org/10.1038/nchembio.2474>.
- (144) Exner, M. P.; Kuenzi, T.; To, T. M. T.; Ouyang, Z.; Schwagerus, S.; Hoesl, M. G.; Hackenberger, C. P. R.; Lensen, M. C.; Panke, S.; Budisa, N. Design of S-Allylcysteine in Situ Production and Incorporation Based on a Novel Pyrrolysyl-tRNA Synthetase Variant. *ChemBioChem* **2017**, *18* (1), 85–90. <https://doi.org/10.1002/cbic.201600537>.

- (145) Willis, J. C. W.; Chin, J. W. Mutually Orthogonal Pyrrolysyl-TRNA Synthetase/TRNA Pairs. *Nat. Chem.* **2018**, *10* (8), 831–837. <https://doi.org/10.1038/s41557-018-0052-5>.
- (146) Liu, J.; Cheng, R.; Wu, H.; Li, S.; Wang, P. G.; DeGrado, W. F.; Rozovsky, S.; Wang, L. Building and Breaking Bonds via a Compact S-Propargyl-Cysteine to Chemically Control Enzymes and Modify Proteins. *Angew. Chemie - Int. Ed.* **2018**, *57* (39), 12702–12706. <https://doi.org/10.1002/anie.201806197>.
- (147) Galindo Casas, M.; Stargardt, P.; Mairhofer, J.; Wiltschi, B. Decoupling Protein Production from Cell Growth Enhances the Site-Specific Incorporation of Noncanonical Amino Acids in *E. Coli*. *ACS Synth. Biol.* **2020**, *9* (11), 3052–3066. <https://doi.org/10.1021/acssynbio.0c00298>.
- (148) Zheng, Y.; Lajoie, M. J.; Italia, J. S.; Chin, M. A.; Church, G. M.; Chatterjee, A. Performance of Optimized Noncanonical Amino Acid Mutagenesis Systems in the Absence of Release Factor 1. *Mol. Biosyst.* **2016**, *12* (6), 1746–1749. <https://doi.org/10.1039/c6mb00070c>.
- (149) Chatterjee, A.; Sun, S. B.; Furman, J. L.; Xiao, H.; Schultz, P. G. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in *Escherichia Coli*. *Biochemistry* **2013**, *52* (10), 1828–1837. <https://doi.org/10.1021/bi4000244>.
- (150) Fan, C.; Xiong, H.; Reynolds, N. M.; Söll, D. Rationally Evolving TRNAPyl for Efficient Incorporation of Noncanonical Amino Acids. *Nucleic Acids Res.* **2015**, *43* (22), 1–10. <https://doi.org/10.1093/nar/gkv800>.
- (151) Pott, M.; Schmidt, M. J.; Summerer, D. Evolved Sequence Contexts for Highly Efficient Amber Suppression with Noncanonical Amino Acids. *ACS Chem. Biol.* **2014**, *9* (12), 2815–2822. <https://doi.org/10.1021/cb5006273>.
- (152) Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H. R.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. A.; Wang, H. H.; Carr, P. A.; Mosberg, J. A.; Rohland, N.; Schultz, P. G.; Jacobson, J. M.; Rinehart, J.; Church, G. M.; Isaacs, F. J. Genomically Recoded Organisms Expand Biological Functions. *Science (80-.)*. **2013**, *342* (6156), 357–360. <https://doi.org/10.1126/science.1241459>.
- (153) Park, H. S.; Hohn, M. J.; Umehara, T.; Guo, L. T.; Osborne, E. M.; Benner, J.; Noren, C. J.; Rinehart, J.; Söll, D. Expanding the Genetic Code of *Escherichia Coli* with Phosphoserine. *Science (80-.)*. **2011**, *333* (6046), 1151–1154. <https://doi.org/10.1126/science.1207203>.
- (154) Mukai, T.; Hoshi, H.; Ohtake, K.; Takahashi, M.; Yamaguchi, A.; Hayashi, A.; Yokoyama, S.; Sakamoto, K. Highly Reproductive *Escherichia Coli* Cells with No Specific Assignment to the UAG Codon. *Sci. Rep.* **2015**, *5*, 1–9. <https://doi.org/10.1038/srep09699>.
- (155) Brugère, J.-F.; Atkins, J. F.; O'Toole, P. W.; Borrel, G. Pyrrolysine in Archaea: A 22nd Amino Acid Encoded through a Genetic Code Expansion. *Emerg. Top. Life Sci.* **2018**, *2* (4), 607–618. <https://doi.org/10.1042/ETLS20180094>.
- (156) Burke, S. A.; Lo, S. L.; Krzycki, J. A. Clustered Genes Encoding the Methyltransferases of Methanogenesis from Monomethylamine. *J. Bacteriol.* **1998**, *180* (13), 3432–3440. <https://doi.org/10.1128/JB.180.13.3432-3440.1998>.
- (157) Paul, L.; Ferguson, D. J.; Krzycki, J. A. The Trimethylamine Methyltransferase Gene and Multiple Dimethylamine Methyltransferase Genes of *Methanosarcina Barkeri* Contain In-Frame and Read-Through Amber Codons. *J. Bacteriol.* **2000**, *182* (9), 2520–2529. <https://doi.org/10.1128/JB.182.9.2520-2529.2000>.
- (158) Hao, B.; Gong, W.; Ferguson, T. K.; James, C. M.; Krzycki, J. A.; Chan, M. K. A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase. *Science (80-.)*. **2002**, *296* (5572), 1462–1466. <https://doi.org/10.1126/science.1069556>.

- (159) Ambrogelly, A.; Gundllapalli, S.; Herring, S.; Polycarpo, C.; Frauer, C.; Soll, D. Pyrrolysine Is Not Hardwired for Cotranslational Insertion at UAG Codons. *Proc. Natl. Acad. Sci.* **2007**, *104* (9), 3141–3146. <https://doi.org/10.1073/pnas.0611634104>.
- (160) Namy, O.; Zhou, Y.; Gundllapalli, S.; Polycarpo, C. R.; Denise, A.; Rousset, J.-P.; Söll, D.; Ambrogelly, A. Adding Pyrrolysine to the Escherichia Coli Genetic Code. *FEBS Lett.* **2007**, *581* (27), 5282–5288. <https://doi.org/10.1016/j.febslet.2007.10.022>.
- (161) Wan, W.; Tharp, J. M.; Liu, W. R. Pyrrolysyl-TRNA Synthetase: An Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. *Biochim. Biophys. Acta - Proteins Proteomics* **2014**, *1844* (6), 1059–1070. <https://doi.org/10.1016/j.bbapap.2014.03.002>.
- (162) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. Designing Logical Codon Reassignment-Expanding the Chemistry in Biology. *Chem. Sci.* **2015**, *6* (1), 50–69. <https://doi.org/10.1039/c4sc01534g>.
- (163) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. Multistep Engineering of Pyrrolysyl-TRNA Synthetase to Genetically Encode N ϵ -(o-Azidobenzoyloxycarbonyl) Lysine for Site-Specific Protein Modification. *Chem. Biol.* **2008**, *15* (11), 1187–1197. <https://doi.org/10.1016/j.chembiol.2008.10.004>.
- (164) Yanagisawa, T.; Kuratani, M.; Seki, E.; Hino, N.; Sakamoto, K.; Yokoyama, S. Structural Basis for Genetic-Code Expansion with Bulky Lysine Derivatives by an Engineered Pyrrolysyl-TRNA Synthetase. *Cell Chem. Biol.* **2019**, *26* (7), 936–949.e13. <https://doi.org/10.1016/j.chembiol.2019.03.008>.
- (165) Wang, Y. S.; Fang, X.; Wallace, A. L.; Wu, B.; Liu, W. R. A Rationally Designed Pyrrolysyl-TRNA Synthetase Mutant with a Broad Substrate Spectrum. *J. Am. Chem. Soc.* **2012**, *134* (6), 2950–2953. <https://doi.org/10.1021/ja211972x>.
- (166) Wang, Y. S.; Russell, W. K.; Wang, Z.; Wan, W.; Dodd, L. E.; Pai, P. J.; Russell, D. H.; Liu, W. R. The de Novo Engineering of Pyrrolysyl-TRNA Synthetase for Genetic Incorporation of l-Phenylalanine and Its Derivatives. *Mol. Biosyst.* **2011**, *7* (3), 714–717. <https://doi.org/10.1039/c0mb00217h>.
- (167) Wang, Y.-S.; Fang, X.; Chen, H.-Y.; Wu, B.; Wang, Z. U.; Hilty, C.; Liu, W. R. Genetic Incorporation of Twelve Meta -Substituted Phenylalanine Derivatives Using a Single Pyrrolysyl-TRNA Synthetase Mutant. *ACS Chem. Biol.* **2013**, *8* (2), 405–415. <https://doi.org/10.1021/cb300512r>.
- (168) Tharp, J. M.; Wang, Y. S.; Lee, Y. J.; Yang, Y.; Liu, W. R. Genetic Incorporation of Seven Ortho-Substituted Phenylalanine Derivatives. *ACS Chem. Biol.* **2014**, *9* (4), 884–890. <https://doi.org/10.1021/cb400917a>.
- (169) Guo, L. T.; Wang, Y. S.; Nakamura, A.; Eiler, D.; Kavran, J. M.; Wong, M.; Kiessling, L. L.; Steitz, T. A.; O'Donoghue, P.; Söll, D. Polyspecific Pyrrolysyl-TRNA Synthetases from Directed Evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (47), 16724–16729. <https://doi.org/10.1073/pnas.1419737111>.
- (170) Tseng, H.; Baumann, T.; Sun, H.; Wang, Y.; Ignatova, Z.; Budisa, N. Expanding the Scope of Orthogonal Translation with Pyrrolysyl-TRNA Synthetases Dedicated to Aromatic Amino Acids. *Molecules* **2020**, *25* (19), 4418. <https://doi.org/10.3390/molecules25194418>.
- (171) Englert, M.; Nakamura, A.; Wang, Y.-S.; Eiler, D.; Söll, D.; Guo, L.-T. Probing the Active Site Tryptophan of Staphylococcus Aureus Thioredoxin with an Analog. *Nucleic Acids Res.* **2015**, *43* (22), 11061–11067. <https://doi.org/10.1093/nar/gkv1255>.
- (172) Xiao, H.; Peters, F. B.; Yang, P. Y.; Reed, S.; Chittuluru, J. R.; Schultz, P. G. Genetic Incorporation of Histidine Derivatives Using an Engineered Pyrrolysyl-TRNA Synthetase. *ACS Chem. Biol.* **2014**, *9* (5), 1092–1096. <https://doi.org/10.1021/cb500032c>.

- (173) Hoppmann, C.; Lacey, V. K.; Louie, G. V.; Wei, J.; Noel, J. P.; Wang, L. Genetically Encoding Photoswitchable Click Amino Acids in Escherichia Coli and Mammalian Cells. *Angew. Chemie - Int. Ed.* **2014**, *53* (15), 3932–3936. <https://doi.org/10.1002/anie.201400001>.
- (174) John, A. A.; Ramil, C. P.; Tian, Y.; Cheng, G.; Lin, Q. Synthesis and Site-Specific Incorporation of Red-Shifted Azobenzene Amino Acids into Proteins. *Org. Lett.* **2015**, *17* (24), 6258–6261. <https://doi.org/10.1021/acs.orglett.5b03268>.
- (175) Herring, S.; Ambrogelly, A.; Gundllapalli, S.; O'Donoghue, P.; Polycarpo, C. R.; Söll, D. The Amino-Terminal Domain of Pyrrolysyl-TRNA Synthetase Is Dispensable in Vitro but Required for in Vivo Activity. *FEBS Lett.* **2007**, *581* (17), 3197–3203. <https://doi.org/10.1016/j.febslet.2007.06.004>.
- (176) Kavran, J. M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Söll, D.; Steitz, T. A. Structure of Pyrrolysyl-TRNA Synthetase, an Archaeal Enzyme for Genetic Code Innovation. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (27), 11268–11273. <https://doi.org/10.1073/pnas.0704769104>.
- (177) Koch, N. G.; Goettig, P.; Rappsilber, J.; Budisa, N. Engineering Pyrrolysyl-TRNA Synthetase for the Incorporation of Non-Canonical Amino Acids with Smaller Side Chains. *Int. J. Mol. Sci.* **2021**, *22* (20), 11194. <https://doi.org/10.3390/ijms222011194>.
- (178) Berghuis, B. A.; Yu, F. B.; Schulz, F.; Blainey, P. C.; Woyke, T.; Quake, S. R. Hydrogenotrophic Methanogenesis in Archaeal Phylum Verstraetearchaeota Reveals the Shared Ancestry of All Methanogens. *Proc. Natl. Acad. Sci.* **2019**, *116* (11), 5037–5044. <https://doi.org/10.1073/pnas.1815631116>.
- (179) Borrel, G.; Gaci, N.; Peyret, P.; O'Toole, P. W.; Gribaldo, S.; Brugère, J.-F. Unique Characteristics of the Pyrrolysine System in the 7th Order of Methanogens: Implications for the Evolution of a Genetic Code Expansion Cassette. *Archaea* **2014**, *2014*, 1–11. <https://doi.org/10.1155/2014/374146>.
- (180) Fournier, G. P.; Huang, J.; Gogarten, J. P. Horizontal Gene Transfer from Extinct and Extant Lineages: Biological Innovation and the Coral of Life. *Philos. Trans. R. Soc. B Biol. Sci.* **2009**, *364* (1527), 2229–2239. <https://doi.org/10.1098/rstb.2009.0033>.
- (181) Suzuki, T.; Miller, C.; Guo, L. T.; Ho, J. M. L.; Bryson, D. I.; Wang, Y. S.; Liu, D. R.; Söll, D. Crystal Structures Reveal an Elusive Functional Domain of Pyrrolysyl-TRNA Synthetase. *Nat. Chem. Biol.* **2017**, *13* (12), 1261–1266. <https://doi.org/10.1038/nchembio.2497>.
- (182) Nozawa, K.; O'Donoghue, P.; Gundllapalli, S.; Arais, Y.; Ishitani, R.; Umehara, T.; Söll, D.; Nureki, O. Pyrrolysyl-TRNA Synthetase-TRNAPyl Structure Reveals the Molecular Basis of Orthogonality. *Nature* **2009**, *457* (7233), 1163–1167. <https://doi.org/10.1038/nature07611>.
- (183) Ho, J. M.; Reynolds, N. M.; Rivera, K.; Connolly, M.; Guo, L. T.; Ling, J.; Pappin, D. J.; Church, G. M.; Söll, D. Efficient Reassignment of a Frequent Serine Codon in Wild-Type Escherichia Coli. *ACS Synth. Biol.* **2016**, *5* (2), 163–171. <https://doi.org/10.1021/acssynbio.5b00197>.
- (184) Fischer, E. C.; Hashimoto, K.; Zhang, Y.; Feldman, A. W.; Dien, V. T.; Karadeema, R. J.; Adhikary, R.; Ledbetter, M. P.; Krishnamurthy, R.; Romesberg, F. E. New Codons for Efficient Production of Unnatural Proteins in a Semisynthetic Organism. *Nat. Chem. Biol.* **2020**, *16* (5), 570–576. <https://doi.org/10.1038/s41589-020-0507-z>.
- (185) O'Donoghue, P.; Ling, J.; Wang, Y.-S.; Söll, D. Upgrading Protein Synthesis for Synthetic Biology. *Nat. Chem. Biol.* **2013**, *9* (10), 594–598. <https://doi.org/10.1038/nchembio.1339>.
- (186) Trudeau, D. L.; Tawfik, D. S. Protein Engineers Turned Evolutionists—the Quest for the Optimal Starting Point. *Curr. Opin. Biotechnol.* **2019**, *60*, 46–52.

- <https://doi.org/10.1016/j.copbio.2018.12.002>.
- (187) Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angew. Chemie - Int. Ed.* **2018**, *57* (16), 4143–4148. <https://doi.org/10.1002/anie.201708408>.
- (188) Hauf, M.; Richter, F.; Schneider, T.; Faidt, T.; Martins, B. M.; Baumann, T.; Durkin, P.; Dobbek, H.; Jacobs, K.; Möglich, A.; Budisa, N. Photoactivatable Mussel-Based Underwater Adhesive Proteins by an Expanded Genetic Code. *ChemBioChem* **2017**, *18* (18), 1819–1823. <https://doi.org/10.1002/cbic.201700327>.
- (189) Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. Protein Stability Promotes Evolvability. *Proc. Natl. Acad. Sci.* **2006**, *103* (15), 5869–5874. <https://doi.org/10.1073/pnas.0510098103>.
- (190) Tokuriki, N.; Tawfik, D. S. Stability Effects of Mutations and Protein Evolvability. *Curr. Opin. Struct. Biol.* **2009**, *19* (5), 596–604. <https://doi.org/10.1016/j.sbi.2009.08.003>.
- (191) Grasso, K. T.; Yeo, M. J. R.; Hillenbrand, C. M.; Ficarella, E. D.; Italia, J. S.; Huang, R. L.; Chatterjee, A. Structural Robustness Affects the Engineerability of Aminoacyl-TRNA Synthetases for Genetic Code Expansion. *Biochemistry* **2021**, *60* (7), 489–493. <https://doi.org/10.1021/acs.biochem.1c00056>.
- (192) Bianco, V.; Franzese, G.; Coluzza, I. In Silico Evidence That Protein Unfolding Is a Precursor of Protein Aggregation. *ChemPhysChem* **2020**. <https://doi.org/10.1002/cphc.201900904>.
- (193) Hu, L.; Qin, X.; Huang, Y.; Cao, W.; Wang, C.; Wang, Y.; Ling, X.; Chen, H.; Wu, D.; Lin, Y.; Liu, T. Thermophilic Pyrrolysyl-TRNA Synthetase Mutants for Enhanced Mammalian Genetic Code Expansion. *ACS Synth. Biol.* **2020**, *acssynbio.0c00257*. <https://doi.org/10.1021/acssynbio.0c00257>.
- (194) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. Crystallographic Studies on Multiple Conformational States of Active-Site Loops in Pyrrolysyl-TRNA Synthetase. *J. Mol. Biol.* **2008**, *378* (3), 634–652. <https://doi.org/10.1016/j.jmb.2008.02.045>.
- (195) Klesmith, J. R.; Bacik, J.-P.; Wrenbeck, E. E.; Michalczyk, R.; Whitehead, T. A. Trade-Offs between Enzyme Fitness and Solubility Illuminated by Deep Mutational Scanning. *Proc. Natl. Acad. Sci.* **2017**, *114* (9), 2265–2270. <https://doi.org/10.1073/pnas.1614437114>.
- (196) Matthews, B. W. Structural and Genetic Analysis of Protein Stability. *Annu. Rev. Biochem.* **1993**, *62* (1), 139–160. <https://doi.org/10.1146/annurev.bi.62.070193.001035>.
- (197) Tokuriki, N.; Stricher, F.; Serrano, L.; Tawfik, D. S. How Protein Stability and New Functions Trade Off. *PLoS Comput. Biol.* **2008**, *4* (2), e1000002. <https://doi.org/10.1371/journal.pcbi.1000002>.
- (198) Paraskevopoulou, V.; Falcone, F. Polyionic Tags as Enhancers of Protein Solubility in Recombinant Protein Expression. *Microorganisms* **2018**, *6* (2), 47. <https://doi.org/10.3390/microorganisms6020047>.
- (199) Esposito, D.; Chatterjee, D. K. Enhancement of Soluble Protein Expression through the Use of Fusion Tags. *Curr. Opin. Biotechnol.* **2006**, *17* (4), 353–358. <https://doi.org/10.1016/j.copbio.2006.06.003>.
- (200) Raran-Kurussi, S.; Waugh, D. S. The Ability to Enhance the Solubility of Its Fusion Partners Is an Intrinsic Property of Maltose-Binding Protein but Their Folding Is Either Spontaneous or Chaperone-Mediated. *PLoS One* **2012**, *7* (11). <https://doi.org/10.1371/journal.pone.0049589>.
- (201) Takai, K.; Nakamura, K.; Toki, T.; Tsunogai, U.; Miyazaki, M.; Miyazaki, J.; Hirayama, H.; Nakagawa, S.; Nunoura, T.; Horikoshi, K. Cell Proliferation at 122 C and Isotopically Heavy CH₄ Production by a Hyperthermophilic Methanogen under High-Pressure

- Cultivation. *Proc. Natl. Acad. Sci.* **2008**, *105* (31), 10949–10954. <https://doi.org/10.1073/pnas.0712334105>.
- (202) Feller, G. Cryosphere and Psychrophiles: Insights into a Cold Origin of Life? *Life* **2017**, *7* (2). <https://doi.org/10.3390/life7020025>.
- (203) Mykytczuk, N. C. S.; Foote, S. J.; Omelon, C. R.; Southam, G.; Greer, C. W.; Whyte, L. G. Bacterial Growth at $-15\text{ }^{\circ}\text{C}$; Molecular Insights from the Permafrost Bacterium *Planococcus Halocryophilus* Or1. *ISME J.* **2013**, *7* (6), 1211–1226. <https://doi.org/10.1038/ismej.2013.8>.
- (204) Siddiqui, K. S. Some like It Hot, Some like It Cold: Temperature Dependent Biotechnological Applications and Improvements in Extremophilic Enzymes. *Biotechnol. Adv.* **2015**, *33* (8), 1912–1922. <https://doi.org/10.1016/j.biotechadv.2015.11.001>.
- (205) Huang, S.-Y.; Zhang, Y.-H. P.; Zhong, J.-J. A Thermostable Recombinant Transaldolase with High Activity over a Broad PH Range. *Appl. Microbiol. Biotechnol.* **2012**, *93* (6), 2403–2410. <https://doi.org/10.1007/s00253-011-3578-7>.
- (206) Watkins, E. J.; Almhjell, P. J.; Arnold, F. H. Direct Enzymatic Synthesis of a Deep-Blue Fluorescent Noncanonical Amino Acid from Azulene and Serine. *ChemBioChem* **2020**, *21* (1–2), 80–83. <https://doi.org/10.1002/cbic.201900497>.
- (207) Åqvist, J.; Isaksen, G. V.; Brandsdal, B. O. Computation of Enzyme Cold Adaptation. *Nat. Rev. Chem.* **2017**, *1* (7), 0051. <https://doi.org/10.1038/s41570-017-0051>.
- (208) Gerday, C.; Aittaleb, M.; Bentahir, M.; Chessa, J.-P.; Claverie, P.; Collins, T.; D'Amico, S.; Dumont, J.; Garsoux, G.; Georlette, D.; Hoyoux, A.; Lonhienne, T.; Meuwis, M.-A.; Feller, G. Cold-Adapted Enzymes: From Fundamentals to Biotechnology. *Trends Biotechnol.* **2000**, *18* (3), 103–107. [https://doi.org/10.1016/S0167-7799\(99\)01413-4](https://doi.org/10.1016/S0167-7799(99)01413-4).
- (209) Santiago, M.; Ramírez-Sarmiento, C. A.; Zamora, R. A.; Parra, L. P. Discovery, Molecular Mechanisms, and Industrial Applications of Cold-Active Enzymes. *Front. Microbiol.* **2016**, *7* (SEP). <https://doi.org/10.3389/fmicb.2016.01408>.
- (210) Karan, R.; Capes, M. D.; DasSarma, S. Function and Biotechnology of Extremophilic Enzymes in Low Water Activity. *Aquat. Biosyst.* **2012**, *8* (1), 4. <https://doi.org/10.1186/2046-9063-8-4>.
- (211) Feller, G. Psychrophilic Enzymes: From Folding to Function and Biotechnology. *Scientifica (Cairo)*. **2013**, *2013*, 1–28. <https://doi.org/10.1155/2013/512840>.
- (212) Feller, G.; Gerday, C. Psychrophilic Enzymes: Hot Topics in Cold Adaptation. *Nat. Rev. Microbiol.* **2003**, *1* (3), 200–208. <https://doi.org/10.1038/nrmicro773>.
- (213) Daniel, R. M.; Peterson, M. E.; Danson, M. J.; Price, N. C.; Kelly, S. M.; Monk, C. R.; Weinberg, C. S.; Oudshoorn, M. L.; Lee, C. K. The Molecular Basis of the Effect of Temperature on Enzyme Activity. *Biochem. J.* **2010**, *425* (2), 353–360. <https://doi.org/10.1042/BJ20091254>.
- (214) Daniel, R. M.; Danson, M. J. A New Understanding of How Temperature Affects the Catalytic Activity of Enzymes. *Trends Biochem. Sci.* **2010**, *35* (10), 584–591. <https://doi.org/10.1016/j.tibs.2010.05.001>.
- (215) Xia, Y. L.; Li, Y. P.; Fu, Y. X.; Liu, S. Q. The Energetic Origin of Different Catalytic Activities in Temperature-Adapted Trypsins. *ACS Omega* **2020**, *5* (39), 25077–25086. <https://doi.org/10.1021/acsomega.0c02401>.
- (216) Nobeli, I.; Favia, A. D.; Thornton, J. M. Protein Promiscuity and Its Implications for Biotechnology. *Nat. Biotechnol.* **2009**, *27* (2), 157–167. <https://doi.org/10.1038/nbt1519>.
- (217) Pédelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and Characterization of a Superfolder Green Fluorescent Protein. *Nat. Biotechnol.* **2006**, *24* (1), 79–88. <https://doi.org/10.1038/nbt1172>.

- (218) Malakhov, M. P.; Mattern, M. R.; Malakhova, O. A.; Drinker, M.; Weeks, S. D.; Butt, T. R. SUMO Fusions and SUMO-Specific Protease for Efficient Expression and Purification of Proteins. *J. Struct. Funct. Genomics* **2004**, No. 5, 75–86. <https://doi.org/10.1023/B:JSFG.0000029237.70316.52>.
- (219) Varshavsky, A. The N-End Rule Pathway and Regulation by Proteolysis. *Protein Sci.* **2011**, *20* (8), 1298–1345. <https://doi.org/10.1002/pro.666>.
- (220) Tobias, J. W.; Shrader, T. E.; Rocap, G.; Varshavsky, A. The N-End Rule in Bacteria. *Science* (80-.). **1991**, *254* (5036), 1374–1377. <https://doi.org/10.1126/science.1962196>.
- (221) Kunjapur, A. M.; Stork, D. A.; Kuru, E.; Vargas-Rodriguez, O.; Landon, M.; Söll, D.; Church, G. M. Engineering Posttranslational Proofreading to Discriminate Nonstandard Amino Acids. *Proc. Natl. Acad. Sci.* **2018**, *115* (3), 619–624. <https://doi.org/10.1073/pnas.1715137115>.
- (222) Ko, J. H.; Wang, Y. S.; Nakamura, A.; Guo, L. T.; Söll, D.; Umehara, T. Pyrrolysyl-TRNA Synthetase Variants Reveal Ancestral Aminoacylation Function. *FEBS Lett.* **2013**, *587* (19), 3243–3248. <https://doi.org/10.1016/j.febslet.2013.08.018>.
- (223) Kodera, Y.; Suzuki, A.; Imada, O.; Kasuga, S.; Sumioka, I.; Kanezawa, A.; Taru, N.; Fujikawa, M.; Nagae, S.; Masamoto, K.; Maeshige, K.; Ono, K. Physical, Chemical, and Biological Properties of S-Allylcysteine, an Amino Acid Derived from Garlic. *J. Agric. Food Chem.* **2002**, *50* (3), 622–632. <https://doi.org/10.1021/jf0106648>.
- (224) Boutureira, O.; Bernardes, G. J. L. Advances in Chemical Protein Modification. *Chem. Rev.* **2015**, *115* (5), 2174–2195. <https://doi.org/10.1021/cr500399p>.
- (225) Lang, K.; Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.* **2014**, *114* (9), 4764–4806. <https://doi.org/10.1021/cr400355w>.
- (226) Hansted, J. G.; Pietikäinen, L.; Hög, F.; Sperling-Petersen, H. U.; Mortensen, K. K. Expressivity Tag: A Novel Tool for Increased Expression in Escherichia Coli. *J. Biotechnol.* **2011**, *155* (3), 275–283. <https://doi.org/10.1016/j.jbiotec.2011.07.013>.
- (227) Rathnayaka, T.; Tawa, M.; Nakamura, T.; Sohya, S.; Kuwajima, K.; Yohda, M.; Kuroda, Y. Solubilization and Folding of a Fully Active Recombinant Gaussia Luciferase with Native Disulfide Bonds by Using a SEP-Tag. *Biochim. Biophys. Acta - Proteins Proteomics* **2011**, *1814* (12), 1775–1778. <https://doi.org/10.1016/j.bbapap.2011.09.001>.
- (228) Smith, J. C.; Derbyshire, R. B.; Cook, E.; Dunthorne, L.; Viney, J.; Brewer, S. J.; Sassenfeld, H. M.; Bell, L. D. Chemical Synthesis and Cloning of a Poly(Arginine)-Coding Gene Fragment Designed to Aid Polypeptide Purification. *Gene* **1984**, *32* (3), 321–327. [https://doi.org/10.1016/0378-1119\(84\)90007-6](https://doi.org/10.1016/0378-1119(84)90007-6).
- (229) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein G. *Science* (80-.). **1991**, *253* (5020), 657–661. <https://doi.org/10.1126/science.1871600>.
- (230) Costa, S. J.; Almeida, A.; Castro, A.; Domingues, L.; Besir, H. The Novel Fh8 and H Fusion Partners for Soluble Protein Expression in Escherichia Coli: A Comparison with the Traditional Gene Fusion Technology. *Appl. Microbiol. Biotechnol.* **2013**, *97* (15), 6779–6791. <https://doi.org/10.1007/s00253-012-4559-1>.
- (231) Costa, S. J.; Coelho, E.; Franco, L.; Almeida, A.; Castro, A.; Domingues, L. The Fh8 Tag: A Fusion Partner for Simple and Cost-Effective Protein Purification in Escherichia Coli. *Protein Expr. Purif.* **2013**, *92* (2), 163–170. <https://doi.org/10.1016/j.pep.2013.09.013>.
- (232) Vargas-Cortez, T.; Morones-Ramirez, J. R.; Balderas-Renteria, I.; Zarate, X. Expression

- and Purification of Recombinant Proteins in Escherichia Coli Tagged with a Small Metal-Binding Protein from Nitrosomonas Europaea. *Protein Expr. Purif.* **2016**, *118*, 49–54. <https://doi.org/10.1016/j.pep.2015.10.009>.
- (233) Holmgren, A. Thioredoxin. *Annu. Rev. Biochem.* **1985**, *54* (1), 237–271. <https://doi.org/10.1146/annurev.bi.54.070185.001321>.
- (234) Davis, G. D.; Elisee, C.; Newham, D. M.; Harrison, R. G. New Fusion Protein Systems Designed to Give Soluble Expression in Escherichia Coli. *Biotechnol. Bioeng.* **1999**, *65* (4), 382–388. [https://doi.org/10.1002/\(SICI\)1097-0290\(19991120\)65:4<382::AID-BIT2>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-0290(19991120)65:4<382::AID-BIT2>3.0.CO;2-I).
- (235) Lammers, C.; Hahn, L. E.; Neumann, H. Optimized Plasmid Systems for the Incorporation of Multiple Different Unnatural Amino Acids by Evolved Orthogonal Ribosomes. *ChemBioChem* **2014**, *15* (12), 1800–1804. <https://doi.org/10.1002/cbic.201402033>.
- (236) Verma, M.; Choi, J.; Cottrell, K. A.; Lavagnino, Z.; Thomas, E. N.; Pavlovic-Djuranovic, S.; Szczesny, P.; Piston, D. W.; Zaher, H. S.; Puglisi, J. D.; Djuranovic, S. A Short Translational Ramp Determines the Efficiency of Protein Synthesis. *Nat. Commun.* **2019**, No. 10, e5774. <https://doi.org/10.1038/s41467-019-13810-1>.
- (237) Lee, H.-J.; Zheng, J. J. PDZ Domains and Their Binding Partners: Structure, Specificity, and Modification. *Cell Commun. Signal.* **2010**, *8* (1), 8. <https://doi.org/10.1186/1478-811X-8-8>.
- (238) Doyle, D. A.; Lee, A.; Lewis, J.; Kim, E.; Sheng, M.; MacKinnon, R. Crystal Structures of a Complexed and Peptide-Free Membrane Protein–Binding Domain: Molecular Basis of Peptide Recognition by PDZ. *Cell* **1996**, *85* (7), 1067–1076. [https://doi.org/10.1016/S0092-8674\(00\)81307-0](https://doi.org/10.1016/S0092-8674(00)81307-0).
- (239) Arbely, E.; Torres-Kolbus, J.; Deiters, A.; Chin, J. W. Photocontrol of Tyrosine Phosphorylation in Mammalian Cells via Genetic Encoding of Photocaged Tyrosine. *J. Am. Chem. Soc.* **2012**, *134* (29), 11912–11915. <https://doi.org/10.1021/ja3046958>.
- (240) Wang, J.; Liu, Y.; Liu, Y.; Zheng, S.; Wang, X.; Zhao, J.; Yang, F.; Zhang, G.; Wang, C.; Chen, P. R. Time-Resolved Protein Activation by Proximal Decaging in Living Systems. *Nature* **2019**, *569* (7757), 509–513. <https://doi.org/10.1038/s41586-019-1188-1>.
- (241) Luo, J.; Samanta, S.; Convertino, M.; Dokholyan, N. V.; Deiters, A. Reversible and Tunable Photoswitching of Protein Function through Genetic Encoding of Azobenzene Amino Acids in Mammalian Cells. *ChemBioChem* **2018**, *19* (20), 2178–2185. <https://doi.org/10.1002/cbic.201800226>.
- (242) Takimoto, J. K.; Dellas, N.; Noel, J. P.; Wang, L. Stereochemical Basis for Engineered Pyrrolysyl-TRNA Synthetase and the Efficient in Vivo Incorporation of Structurally Divergent Non-Native Amino Acids. *ACS Chem. Biol.* **2011**, *6* (7), 733–743. <https://doi.org/10.1021/cb200057a>.
- (243) Beadle, B. M.; Shoichet, B. K. Structural Bases of Stability-Function Tradeoffs in Enzymes. *J. Mol. Biol.* **2002**, *321* (2), 285–296. [https://doi.org/10.1016/S0022-2836\(02\)00599-5](https://doi.org/10.1016/S0022-2836(02)00599-5).
- (244) Imai, K.; Mitaku, S. Mechanisms of Secondary Structure Breakers in Soluble Proteins. *Biophysics (Oxf)*. **2005**, *1*, 55–65. <https://doi.org/10.2142/biophysics.1.55>.
- (245) Chen, W.-N.; Kuppan, K. V.; Lee, M. D.; Jaudzems, K.; Huber, T.; Otting, G. O - Tert - Butyltyrosine, an NMR Tag for High-Molecular-Weight Systems and Measurements of Submicromolar Ligand Binding Affinities. *J. Am. Chem. Soc.* **2015**, *137* (13), 4581–4586. <https://doi.org/10.1021/jacs.5b01918>.
- (246) Rajendran, P.; Ho, E.; Williams, D. E.; Dashwood, R. H. Dietary Phytochemicals, HDAC Inhibition, and DNA Damage/Repair Defects in Cancer Cells. *Clin. Epigenetics* **2011**, *3*

- (1), 4. <https://doi.org/10.1186/1868-7083-3-4>.
- (247) Nian, H.; Delage, B.; Pinto, J. T.; Dashwood, R. H. Allyl Mercaptan, a Garlic-Derived Organosulfur Compound, Inhibits Histone Deacetylase and Enhances Sp3 Binding on the P21WAF1 Promoter. *Carcinogenesis* **2008**, *29* (9), 1816–1824. <https://doi.org/10.1093/carcin/bgn165>.
- (248) Pagar, A. D.; Patil, M. D.; Flood, D. T.; Yoo, T. H.; Dawson, P. E.; Yun, H. Recent Advances in Biocatalysis with Chemical Modification and Expanded Amino Acid Alphabet. *Chem. Rev.* **2021**, *121* (10), 6173–6245. <https://doi.org/10.1021/acs.chemrev.0c01201>.
- (249) Völler, J. S.; Biava, H.; Hildebrandt, P.; Budisa, N. An Expanded Genetic Code for Probing the Role of Electrostatics in Enzyme Catalysis by Vibrational Stark Spectroscopy. *Biochim. Biophys. Acta - Gen. Subj.* **2017**, *1861* (11), 3053–3059. <https://doi.org/10.1016/j.bbagen.2017.02.009>.
- (250) Mishra, P. K.; Yoo, C.; Hong, E.; Rhee, H. W. Photo-crosslinking: An Emerging Chemical Tool for Investigating Molecular Networks in Live Cells. *ChemBioChem* **2020**, *21* (7), 924–932. <https://doi.org/10.1002/cbic.201900600>.
- (251) Kim, G.; Weiss, S. J.; Levine, R. L. Methionine Oxidation and Reduction in Proteins. *Biochim. Biophys. Acta - Gen. Subj.* **2014**, *1840* (2), 901–905. <https://doi.org/10.1016/j.bbagen.2013.04.038>.
- (252) Wang, B.; Lodder, M.; Zhou, J.; Baird, T. T.; Brown, K. C.; Craik, C. S.; Hecht, S. M. Chemically Mediated Site-Specific Cleavage of Proteins. *J. Am. Chem. Soc.* **2000**, *122* (30), 7402–7403. <https://doi.org/10.1021/ja0002262>.
- (253) Baird, T.; Wang, B.; Lodder, M.; Hecht, S. M.; Craik, C. S. Generation of Active Trypsin by Chemical Cleavage. *Tetrahedron* **2000**, *56* (48), 9477–9485. [https://doi.org/10.1016/S0040-4020\(00\)00832-2](https://doi.org/10.1016/S0040-4020(00)00832-2).
- (254) Wang, B.; Brown, K. C.; Lodder, M.; Craik, C. S.; Hecht, S. M. Chemically Mediated Site-Specific Proteolysis. Alteration of Protein–Protein Interaction †. *Biochemistry* **2002**, *41* (8), 2805–2813. <https://doi.org/10.1021/bi011762p>.
- (255) Liutkus, M.; Fraser, S. A.; Caron, K.; Stigers, D. J.; Easton, C. J. Peptide Synthesis through Cell-Free Expression of Fusion Proteins Incorporating Modified Amino Acids as Latent Cleavage Sites for Peptide Release. *ChemBioChem* **2016**, *17* (10), 908–912. <https://doi.org/10.1002/cbic.201600091>.
- (256) Nolsøe, J. M. J.; Hansen, T. V. Asymmetric Iodolactonization: An Evolutionary Account. *European J. Org. Chem.* **2014**, *2014* (15), 3051–3065. <https://doi.org/10.1002/ejoc.201301400>.
- (257) Kristianslund, R.; Tungen, J. E.; Hansen, T. V. Catalytic Enantioselective Iodolactonization Reactions. *Org. Biomol. Chem.* **2019**, *17* (12), 3079–3092. <https://doi.org/10.1039/C8OB03160F>.
- (258) Lodder, M.; Golovine, S.; Laikhter, A. L.; Karginov, V. A.; Hecht, S. M. Misacylated Transfer RNAs Having a Chemically Removable Protecting Group. *J. Org. Chem.* **1998**, *63* (3), 794–803. <https://doi.org/10.1021/jo971692l>.
- (259) GROSS, E.; WITKOP, B. Nonenzymatic Cleavage of Peptide Bonds: The Methionine Residues in Bovine Pancreatic Ribonuclease. *J. Biol. Chem.* **1962**, *237* (6), 1856–1860.
- (260) Ai, H. W.; Shen, W.; Brustad, E.; Schultz, P. G. Genetically Encoded Alkenes in Yeast. *Angew. Chemie - Int. Ed.* **2010**, *49* (5), 935–937. <https://doi.org/10.1002/anie.200905590>.
- (261) Jiang, H.; Zhang, X.; Chen, X.; Aramsangtienchai, P.; Tong, Z.; Lin, H. Protein Lipidation: Occurrence, Mechanisms, Biological Functions, and Enabling Technologies. *Chem. Rev.* **2018**, *118* (3), 919–988. <https://doi.org/10.1021/acs.chemrev.6b00750>.

- (262) Chen, J. J.; Boehning, D. Protein Lipidation as a Regulator of Apoptotic Calcium Release: Relevance to Cancer. *Front. Oncol.* **2017**, *7* (JUN), 1–11. <https://doi.org/10.3389/fonc.2017.00138>.
- (263) Coleman, D. T.; Gray, A. L.; Kridel, S. J.; Cardelli, J. A. Palmitoylation Regulates the Intracellular Trafficking and Stability of C-Met. *Oncotarget* **2016**, *7* (22), 32664–32677. <https://doi.org/10.18632/oncotarget.8706>.
- (264) Hang, H. C.; Linder, M. E. Exploring Protein Lipidation with Chemical Biology. *Chem. Rev.* **2011**, *111* (10), 6341–6358. <https://doi.org/10.1021/cr2001977>.
- (265) Bech, E. M.; Pedersen, S. L.; Jensen, K. J. Chemical Strategies for Half-Life Extension of Biopharmaceuticals: Lipidation and Its Alternatives. *ACS Med. Chem. Lett.* **2018**, *9* (7), 577–580. <https://doi.org/10.1021/acsmchemlett.8b00226>.
- (266) Acevedo-Rocha, C. G.; Geiermann, A.-S.; Budisa, N.; Merkel, L. Design of Protein Congeners Containing β -Cyclopropylalanine. *Mol. Biosyst.* **2012**, *8* (10), 2719. <https://doi.org/10.1039/c2mb25193k>.
- (267) 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* (2), 788–796. <https://doi.org/10.1111/j.1432-1033.1995.0788h.x>.
- (268) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. Incorporation of Azides into Recombinant Proteins for Chemoselective Modification by the Staudinger Ligation. *Proc. Natl. Acad. Sci.* **2002**, *99* (1), 19–24. <https://doi.org/10.1073/pnas.012583299>.
- (269) Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A.; Tirrell, D. A. Discovery of Escherichia Coli Methionyl-TRNA Synthetase Mutants for Efficient Labeling of Proteins with Azidonorleucine in Vivo. *Proc. Natl. Acad. Sci.* **2009**, *106* (36), 15285–15290. <https://doi.org/10.1073/pnas.0905735106>.
- (270) Wang, Z. A.; Kurra, Y.; Wang, X.; Zeng, Y.; Lee, Y.; Sharma, V.; Lin, H.; Dai, S. Y.; Liu, W. R. A Versatile Approach for Site-Specific Lysine Acylation in Proteins. *Angew. Chemie Int. Ed.* **2017**, *56* (6), 1643–1647. <https://doi.org/10.1002/anie.201611415>.
- (271) Salaün, J. Cyclopropane Derivatives and Their Diverse Biological Activities; 2000; Vol. 207, pp 1–67. https://doi.org/10.1007/3-540-48255-5_1.
- (272) Geigert, J.; Neidleman, S. L.; Dalietos, D. J. Novel Haloperoxidase Substrates. Alkynes and Cyclopropanes. *J. Biol. Chem.* **1983**, *258* (4), 2273–2277.
- (273) Dalton, H.; Golding, B. T.; Waters, B. W.; Higgins, R.; Taylor, J. A. Oxidations of Cyclopropane, Methylcyclopropane, and Arenes with the Mono-Oxygenase System from Methylococcus Capsulatus. *J. Chem. Soc. Chem. Commun.* **1981**, *28* (10), 482. <https://doi.org/10.1039/c39810000482>.
- (274) Marchand, J. A.; Neugebauer, M. E.; Ing, M. C.; Lin, C.-I.; Pelton, J. G.; Chang, M. C. Y. Discovery of a Pathway for Terminal-Alkyne Amino Acid Biosynthesis. *Nature* **2019**, *567* (7748), 420–424. <https://doi.org/10.1038/s41586-019-1020-y>.
- (275) Owens, A. E.; Grasso, K. T.; Ziegler, C. A.; Fasan, R. Two-Tier Screening Platform for Directed Evolution of Aminoacyl-TRNA Synthetases with Enhanced Stop Codon Suppression Efficiency. *ChemBioChem* **2017**, *18* (12), 1109–1116. <https://doi.org/10.1002/cbic.201700039>.
- (276) Nov, Y. When Second Best Is Good Enough: Another Probabilistic Look at Saturation Mutagenesis. *Appl. Environ. Microbiol.* **2012**, *78* (1), 258–262. <https://doi.org/10.1128/AEM.06265-11>.
- (277) Johnson, D. B. F.; Xu, J.; Shen, Z.; Takimoto, J. K.; Schultz, M. D.; Schmitz, R. J.; Xiang, Z.; Ecker, J. R.; Briggs, S. P.; Wang, L. RF1 Knockout Allows Ribosomal Incorporation

- of Unnatural Amino Acids at Multiple Sites. *Nat. Chem. Biol.* **2011**, *7* (11), 779–786. <https://doi.org/10.1038/nchembio.657>.
- (278) O'Donoghue, P.; Prat, L.; Heinemann, I. U.; Ling, J.; Odoi, K.; Liu, W. R.; Söll, D. Near-Cognate Suppression of Amber, Opal and Quadruplet Codons Competes with Aminoacyl-TRNAPyl for Genetic Code Expansion. *FEBS Lett.* **2012**, *586* (21), 3931–3937. <https://doi.org/10.1016/j.febslet.2012.09.033>.
- (279) Wang, Y.; Chen, X.; Cai, W.; Tan, L.; Yu, Y.; Han, B.; Li, Y.; Xie, Y.; Su, Y.; Luo, X.; Liu, T. Expanding the Structural Diversity of Protein Building Blocks with Noncanonical Amino Acids Biosynthesized from Aromatic Thiols. *Angew. Chemie* **2021**, ange.202014540. <https://doi.org/10.1002/ange.202014540>.
- (280) Fladischer, P.; Weingartner, A.; Blamauer, J.; Darnhofer, B.; Birner-Gruenberger, R.; Kardashliev, T.; Ruff, A. J.; Schwaneberg, U.; Wiltschi, B. A Semi-Rationally Engineered Bacterial Pyrrolysyl-TRNA Synthetase Genetically Encodes Phenyl Azide Chemistry. *Biotechnol. J.* **2019**, *14* (3), 1–8. <https://doi.org/10.1002/biot.201800125>.
- (281) Jiang, R.; Krzycki, J. A. PylSn and the Homologous N-Terminal Domain of Pyrrolysyl-TRNA Synthetase Bind the TRNA That Is Essential for the Genetic Encoding of Pyrrolysine. *J. Biol. Chem.* **2012**, *287* (39), 32738–32746. <https://doi.org/10.1074/jbc.M112.396754>.
- (282) KATAYAMA, H.; NOZAWA, K.; NUREKI, O.; NAKAHARA, Y.; HOJO, H. Pyrrolysine Analogs as Substrates for Bacterial Pyrrolysyl-TRNA Synthetase in Vitro and in Vivo. *Biosci. Biotechnol. Biochem.* **2012**, *76* (1), 205–208. <https://doi.org/10.1271/bbb.110653>.
- (283) Albers, S. V.; Siebers, B. *The Prokaryotes*; Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E., Thompson, F., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; Vol. 9783642389. <https://doi.org/10.1007/978-3-642-38954-2>.
- (284) Cheng, L.; Qiu, T.-L.; Yin, X.-B.; Wu, X.-L.; Hu, G.-Q.; Deng, Y.; Zhang, H. *Methermicoccus Shengliensis* Gen. Nov., Sp. Nov., a Thermophilic, Methylophilic Methanogen Isolated from Oil-Production Water, and Proposal of Methermicoccaceae Fam. Nov. *Int. J. Syst. Evol. Microbiol.* **2007**, *57* (12), 2964–2969. <https://doi.org/10.1099/ijs.0.65049-0>.
- (285) Price, M. N.; Dehal, P. S.; Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **2010**, *5* (3), e9490. <https://doi.org/10.1371/journal.pone.0009490>.
- (286) Wei, D.; Jiang, Q.; Wei, Y.; Wang, S. A Novel Hierarchical Clustering Algorithm for Gene Sequences. *BMC Bioinformatics* **2012**, *13* (1), 174. <https://doi.org/10.1186/1471-2105-13-174>.
- (287) Violot, S.; Aghajari, N.; Czjzek, M.; Feller, G.; Sonan, G. K.; Gouet, P.; Gerday, C.; Haser, R.; Receveur-Bréchet, V. Structure of a Full Length Psychrophilic Cellulase from *Pseudoalteromonas Haloplanktis* Revealed by X-Ray Diffraction and Small Angle X-Ray Scattering. *J. Mol. Biol.* **2005**, *348* (5), 1211–1224. <https://doi.org/10.1016/j.jmb.2005.03.026>.
- (288) Garsoux, G.; Lamotte, J.; Gerday, C.; Feller, G. Kinetic and Structural Optimization to Catalysis at Low Temperatures in a Psychrophilic Cellulase from the Antarctic Bacterium *Pseudoalteromonas Haloplanktis*. *Biochem. J.* **2004**, *384* (2), 247–253. <https://doi.org/10.1042/BJ20040325>.
- (289) Jiang, H.-K.; Lee, M.-N.; Tsou, J.-C.; Chang, K.-W.; Tseng, H.-W.; Chen, K.-P.; Li, Y.-K.; Wang, Y.-S. Linker and N-Terminal Domain Engineering of Pyrrolysyl-TRNA Synthetase for Substrate Range Shifting and Activity Enhancement. *Front. Bioeng. Biotechnol.* **2020**, *8* (April), 1–14. <https://doi.org/10.3389/fbioe.2020.00235>.
- (290) Allen, M. A.; Lauro, F. M.; Williams, T. J.; Burg, D.; Siddiqui, K. S.; De Francisci, D.;

- Chong, K. W. Y.; Pilak, O.; Chew, H. H.; De Maere, M. Z.; Ting, L.; Katrib, M.; Ng, C.; Sowers, K. R.; Galperin, M. Y.; Anderson, I. J.; Ivanova, N.; Dalin, E.; Martinez, M.; Lapidus, A.; Hauser, L.; Land, M.; Thomas, T.; Cavicchioli, R. The Genome Sequence of the Psychrophilic Archaeon, *Methanococcoides Burtonii*: The Role of Genome Evolution in Cold Adaptation. *ISME J.* **2009**, *3* (9), 1012–1035. <https://doi.org/10.1038/ismej.2009.45>.
- (291) Chen, Z.; Yu, H.; Li, L.; Hu, S.; Dong, X. The Genome and Transcriptome of a Newly Described Psychrophilic Archaeon, *Methanobus Psychrophilus* R15, Reveal Its Cold Adaptive Characteristics. *Environ. Microbiol. Rep.* **2012**, *4* (6), n/a-n/a. <https://doi.org/10.1111/j.1758-2229.2012.00389.x>.
- (292) Simankova, M. V.; Parshina, S. N.; Tourova, T. P.; Kolganova, T. V.; Zehnder, A. J. B.; Nozhevnikova, A. N. *Methanosarcina Lacustris* Sp. Nov., a New Psychrotolerant Methanogenic Archaeon from Anoxic Lake Sediments. *Syst. Appl. Microbiol.* **2001**, *24* (3), 362–367. <https://doi.org/10.1078/0723-2020-00058>.
- (293) Wagner, D.; Schirmack, J.; Ganzert, L.; Morozova, D.; Mangelsdorf, K. *Methanosarcina Soligelidi* Sp. Nov., a Desiccation- and Freeze-Thaw-Resistant Methanogenic Archaeon from a Siberian Permafrost-Affected Soil. *Int. J. Syst. Evol. Microbiol.* **2013**, *63* (Pt_8), 2986–2991. <https://doi.org/10.1099/ijs.0.046565-0>.
- (294) Galagan, J. E.; Nusbaum, C.; Roy, A.; Endrizzi, M. G.; Macdonald, P.; FitzHugh, W.; Calvo, S.; Engels, R.; Smirnov, S.; Atnoor, D.; Brown, A.; Allen, N.; Naylor, J.; Stange-Thomann, N.; DeArellano, K.; Johnson, R.; Linton, L.; McEwan, P.; McKernan, K.; Talamas, J.; Tirrell, A.; Ye, W.; Zimmer, A.; Barber, R. D.; Cann, I.; Graham, D. E.; Grahame, D. A.; Guss, A. M.; Hedderich, R.; Ingram-Smith, C.; Kuettner, H. C.; Krzycki, J. A.; Leigh, J. A.; Li, W.; Liu, J.; Mukhopadhyay, B.; Reeve, J. N.; Smith, K.; Springer, T. A.; Umayam, L. A.; White, O.; White, R. H.; de Macario, E. C.; Ferry, J. G.; Jarrell, K. F.; Jing, H.; Macario, A. J. L.; Paulsen, I.; Pritchett, M.; Sowers, K. R.; Swanson, R. V.; Zinder, S. H.; Lander, E.; Metcalf, W. W.; Birren, B. The Genome of *M. Acetivorans* Reveals Extensive Metabolic and Physiological Diversity. *Genome Res.* **2002**, *12* (4), 532–542. <https://doi.org/10.1101/gr.223902>.
- (295) Shimizu, S.; Upadhye, R.; Ishijima, Y.; Naganuma, T. *Methanosarcina Horonobensis* Sp. Nov., a Methanogenic Archaeon Isolated from a Deep Subsurface Miocene Formation. *Int. J. Syst. Evol. Microbiol.* **2011**, *61* (10), 2503–2507. <https://doi.org/10.1099/ijs.0.028548-0>.
- (296) Elberson, M. A.; Sowers, K. R. Isolation of an Aceticlastic Strain of *Methanosarcina Siciliae* from Marine Canyon Sediments and Emendation of the Species Description for *Methanosarcina Siciliae*. *Int. J. Syst. Bacteriol.* **1997**, *47* (4), 1258–1261. <https://doi.org/10.1099/00207713-47-4-1258>.
- (297) Zhilina, T. N.; Zavarzin, G. A. NOTES: *Methanosarcina Vacuolata* Sp. Nov., a Vacuolated *Methanosarcina*. *Int. J. Syst. Bacteriol.* **1987**, *37* (3), 281–283. <https://doi.org/10.1099/00207713-37-3-281>.
- (298) Ganzert, L.; Schirmack, J.; Alawi, M.; Mangelsdorf, K.; Sand, W.; Hillebrand-Voiculescu, A.; Wagner, D. *Methanosarcina Spelaei* Sp. Nov., a Methanogenic Archaeon Isolated from a Floating Biofilm of a Subsurface Sulphurous Lake. *Int. J. Syst. Evol. Microbiol.* **2014**, *64* (Pt_10), 3478–3484. <https://doi.org/10.1099/ijs.0.064956-0>.
- (299) Kern, T.; Fischer, M. A.; Deppenmeier, U.; Schmitz, R. A.; Rother, M. *Methanosarcina Flavescens* Sp. Nov., a Methanogenic Archaeon Isolated from a Full-Scale Anaerobic Digester. *Int. J. Syst. Evol. Microbiol.* **2016**, *66* (3), 1533–1538. <https://doi.org/10.1099/ijsem.0.000894>.
- (300) ZINDER, S. H.; SOWERS, K. R.; FERRY, J. G. NOTES: *Methanosarcina Thermophila* Sp. Nov., a Thermophilic, Acetotrophic, Methane-Producing Bacterium. *Int. J. Syst. Bacteriol.* **1985**, *35* (4), 522–523. <https://doi.org/10.1099/00207713-35-4-522>.

- (301) Chou, C.; Uprety, R.; Davis, L.; Chin, J. W.; Deiters, A. Genetically Encoding an Aliphatic Diazirine for Protein Photocrosslinking. *Chem. Sci.* **2011**, *2* (3), 480–483. <https://doi.org/10.1039/C0SC00373E>.
- (302) Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W. Genetic Encoding and Labeling of Aliphatic Azides and Alkynes in Recombinant Proteins via a Pyrrolysyl-TRNA Synthetase/TRNA CUA Pair and Click Chemistry. *J. Am. Chem. Soc.* **2009**, *131* (25), 8720–8721. <https://doi.org/10.1021/ja900553w>.
- (303) 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* (4), 818–822. <https://doi.org/10.1016/j.bbrc.2008.04.164>.
- (304) Dunkelmann, D. L.; Willis, J. C. W.; Beattie, A. T.; Chin, J. W. Engineered Triply Orthogonal Pyrrolysyl-TRNA Synthetase/TRNA Pairs Enable the Genetic Encoding of Three Distinct Non-Canonical Amino Acids. *Nat. Chem.* **2020**, *12* (6), 535–544. <https://doi.org/10.1038/s41557-020-0472-x>.
- (305) Ai, H.; Shen, W.; Sagi, A.; Chen, P. R.; Schultz, P. G. Probing Protein-Protein Interactions with a Genetically Encoded Photo-Crosslinking Amino Acid. *ChemBioChem* **2011**, *12* (12), 1854–1857. <https://doi.org/10.1002/cbic.201100194>.
- (306) Khersonsky, O.; Tawfik, D. S. *Enzyme Promiscuity—Evolutionary and Mechanistic Aspects*, 8th ed.; Barton, D., Meth-Cohn, O., Eds.; Elsevier: Oxford, 2010.
- (307) Cellitti, S. E.; Jones, D. H.; Lagpacan, L.; Hao, X.; Zhang, Q.; Hu, H.; Brittain, S. M.; Brinker, A.; Caldwell, J.; Bursulaya, B.; Spraggon, G.; Brock, A.; Ryu, Y.; Uno, T.; Schultz, P. G.; Geierstanger, B. H. 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* (29), 9268–9281. <https://doi.org/10.1021/ja801602q>.
- (308) Lin, Y. A.; Boutureira, O.; Lercher, L.; Bhushan, B.; Paton, R. S.; Davis, B. G. Rapid Cross-Metathesis for Reversible Protein Modifications via Chemical Access to Se -Allyl-Selenocysteine in Proteins. *J. Am. Chem. Soc.* **2013**, *135* (33), 12156–12159. <https://doi.org/10.1021/ja403191g>.
- (309) Zhang, Y. A. N.; Wang, L. E. I.; Schultz, P. G.; Wilson, I. A. Tyrosyl-TRNA Synthetase (TyrRS) and an Engineered TyrRS Specific for O-Methyl-L-Tyrosine. *Protein Sci.* **2005**, *14*, 1340–1349. <https://doi.org/10.1110/ps.041239305.4>.
- (310) Schultz, K. C.; Supekova, L.; Ryu, Y.; Xie, J.; Perera, R.; Schultz, P. G. A Genetically Encoded Infrared Probe. *J. Am. Chem. Soc.* **2006**, *128* (43), 13984–13985. <https://doi.org/10.1021/ja0636690>.
- (311) Miyake-Stoner, S. J.; Miller, A. M.; Hammill, J. T.; Peeler, J. C.; Hess, K. R.; Mehl, R. A.; Brewer, S. H. Probing Protein Folding Using Site-Specifically Encoded Unnatural Amino Acids as FRET Donors with Tryptophan. *Biochemistry* **2009**, *48* (25), 5953–5962. <https://doi.org/10.1021/bi900426d>.
- (312) Wang, J.; Xie, J.; Schultz, P. G. A Genetically Encoded Fluorescent Amino Acid. *J. Am. Chem. Soc.* **2006**, *128* (27), 8738–8739. <https://doi.org/10.1021/ja062666k>.
- (313) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of p -Azido- l -Phenylalanine to the Genetic Code of Escherichia coli. *J. Am. Chem. Soc.* **2002**, *124* (31), 9026–9027. <https://doi.org/10.1021/ja027007w>.
- (314) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of a Photocrosslinking Amino Acid to the Genetic Code of Escherichia coli. *Proc. Natl. Acad. Sci.* **2002**, *99* (17), 11020–11024. <https://doi.org/10.1073/pnas.172226299>.
- (315) Lorenz, R.; Bernhart, S. H.; Höner zu Siederdissen, C.; Tafer, H.; Flamm, C.; Stadler,

- P. F.; Hofacker, I. L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **2011**, *6* (1), 26. <https://doi.org/10.1186/1748-7188-6-26>.
- (316) Cervettini, D.; Tang, S.; Fried, S. D.; Willis, J. C. W.; Funke, L. F. H.; Colwell, L. J.; Chin, J. W. Rapid Discovery and Evolution of Orthogonal Aminoacyl-TRNA Synthetase-TRNA Pairs. *Nat. Biotechnol.* **2020**, *38* (8), 989–999. <https://doi.org/10.1038/s41587-020-0479-2>.
- (317) Amiram, M.; Haimovich, A. D.; Fan, C.; Wang, Y. S.; Aerni, H. R.; Ntai, I.; Moonan, D. W.; Ma, N. J.; Rovner, A. J.; Hong, S. H.; Kelleher, N. L.; Goodman, A. L.; Jewett, M. C.; Söll, D.; Rinehart, J.; Isaacs, F. J. Evolution of Translation Machinery in Recoded Bacteria Enables Multi-Site Incorporation of Nonstandard Amino Acids. *Nat. Biotechnol.* **2015**, *33* (12), 1272–1279. <https://doi.org/10.1038/nbt.3372>.
- (318) Young, D. D.; Young, T. S.; Jahnz, M.; Ahmad, I.; Spraggon, G.; Schultz, P. G. An Evolved Aminoacyl-TRNA Synthetase with Atypical Polysubstrate Specificity. *Biochemistry* **2011**, *50* (11), 1894–1900. <https://doi.org/10.1021/bi101929e>.
- (319) Waite, J. H. Mussel Adhesion - Essential Footwork. *J. Exp. Biol.* **2017**, *220* (4), 517–530. <https://doi.org/10.1242/jeb.134056>.
- (320) Yang, B.; Ayyadurai, N.; Yun, H.; Choi, Y. S.; Hwang, B. H.; Huang, J.; Lu, Q.; Zeng, H.; Cha, H. J. In Vivo Residue-Specific Dopa-Incorporated Engineered Mussel Bioglue with Enhanced Adhesion and Water Resistance. *Angew. Chemie - Int. Ed.* **2014**, *53* (49), 13360–13364. <https://doi.org/10.1002/anie.201406099>.
- (321) Schneider, T.; Kubyschkin, V.; Budisa, N. Synthesis of a Photo-Caged DOPA Derivative by Selective Alkylation of 3,4-Dihydroxybenzaldehyde. *European J. Org. Chem.* **2018**, *2018* (18), 2053–2063. <https://doi.org/10.1002/ejoc.201701749>.
- (322) Chemla, Y.; Ozer, E.; Algov, I.; Alfonta, L. Context Effects of Genetic Code Expansion by Stop Codon Suppression. *Curr. Opin. Chem. Biol.* **2018**, *46* (Box 1), 146–155. <https://doi.org/10.1016/j.cbpa.2018.07.012>.
- (323) Urry, D. W. Physical Chemistry of Biological Free Energy Transduction as Demonstrated by Elastic Protein-Based Polymers. *J. Phys. Chem. B* **1997**, *101* (51), 11007–11028. <https://doi.org/10.1021/jp972167t>.
- (324) Vrhovski, B.; Weiss, A. S. Biochemistry of Tropoelastin. *Eur. J. Biochem.* **1998**, *258* (1), 1–18. <https://doi.org/10.1046/j.1432-1327.1998.2580001.x>.
- (325) Bellingham, C. M.; Lillie, M. A.; Gosline, J. M.; Wright, G. M.; Starcher, B. C.; Bailey, A. J.; Woodhouse, K. A.; Keeley, F. W. Recombinant Human Elastin Polypeptides Self-Assemble into Biomaterials with Elastin-like Properties. *Biopolymers* **2003**, *70* (4), 445–455. <https://doi.org/10.1002/bip.10512>.
- (326) Hale, M. K. M. C.; Setton, L. a; D, P.; Chilkoti, A. Cartilaginous Tissue Repair. *Tissue Eng.* **2005**, *11* (11), 1768–1779.
- (327) Meyer, D. E.; Chilkoti, A. Genetically Encoded Synthesis of Protein-Based Polymers with Precisely Specified Molecular Weight and Sequence by Recursive Directional Ligation: Examples from the the Elastin-like Polypeptide System. *Biomacromolecules* **2002**, *3* (2), 357–367. <https://doi.org/10.1021/bm015630n>.
- (328) Antonczak, A. K.; Simova, Z.; Yonemoto, I. T.; Bochtler, M.; Piasecka, A.; Czapińska, H.; Brancale, A.; Tippmann, E. M. Importance of Single Molecular Determinants in the Fidelity of Expanded Genetic Codes. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (4), 1320–1325. <https://doi.org/10.1073/pnas.1012276108>.
- (329) Deiters, A.; Groff, D.; Ryu, Y.; Xie, J.; Schultz, P. G. A Genetically Encoded Photocaged Tyrosine. *Angew. Chemie Int. Ed.* **2006**, *45* (17), 2728–2731. <https://doi.org/10.1002/anie.200600264>.
- (330) Hauf, M. Genetically Encoded Amino Acid-Based Photocages and Spectral Probes for

- the Design of Controllable Bioadhesives and Non-Invasive Labeling, Technical University of Berlin, 2017.
- (331) Nguyen, D. P.; Mahesh, M.; Elsässer, S. J.; Hancock, S. M.; Uttamapinant, C.; Chin, J. W. Genetic Encoding of Photocaged Cysteine Allows Photoactivation of TEV Protease in Live Mammalian Cells. *J. Am. Chem. Soc.* **2014**, *136* (6), 2240–2243. <https://doi.org/10.1021/ja412191m>.
- (332) Böcker, J. K.; Dörner, W.; Mootz, H. D. Light-Control of the Ultra-Fast Gp41-1 Split Intein with Preserved Stability of a Genetically Encoded Photo-Caged Amino Acid in Bacterial Cells. *Chem. Commun.* **2019**, *55* (9), 1287–1290. <https://doi.org/10.1039/c8cc09204d>.
- (333) Ren, W.; Ji, A.; Wang, M. X.; Ai, H. Expanding the Genetic Code for a Dinitrophenyl Hapten. *ChemBioChem* **2015**, *16* (14), 2007–2010. <https://doi.org/10.1002/cbic.201500204>.
- (334) Copp, J. N.; Mowday, A. M.; Williams, E. M.; Guise, C. P.; Ashoorzadeh, A.; Sharrock, A. V.; Flanagan, J. U.; Smail, J. B.; Patterson, A. V.; Ackerley, D. F. Engineering a Multifunctional Nitroreductase for Improved Activation of Prodrugs and PET Probes for Cancer Gene Therapy. *Cell Chem. Biol.* **2017**, *24* (3), 391–403. <https://doi.org/10.1016/j.chembiol.2017.02.005>.
- (335) Mercier, C.; Chalansonnet, V.; Orenge, S.; Gilbert, C. Characteristics of Major Escherichia Coli Reductases Involved in Aerobic Nitro and Azo Reduction. *J. Appl. Microbiol.* **2013**, *115* (4), 1012–1022. <https://doi.org/10.1111/jam.12294>.
- (336) Zenno, S.; Koike, H.; Kumar, A. N.; Jayaraman, R.; Tanokura, M.; Saigo, K. Biochemical Characterization of NfsA, the Escherichia Coli Major Nitroreductase Exhibiting a High Amino Acid Sequence Homology to Frp, a Vibrio Harveyi Flavin Oxidoreductase. *J. Bacteriol.* **1996**, *178* (15), 4508–4514. <https://doi.org/10.1128/jb.178.15.4508-4514.1996>.
- (337) Valle, A.; Le Borgne, S.; Bolívar, J.; Cabrera, G.; Cantero, D. Study of the Role Played by NfsA, NfsB Nitroreductase and NemA Flavin Reductase from Escherichia Coli in the Conversion of Ethyl 2-(2'- Nitrophenoxy)Acetate to 4-Hydroxy-(2H)-1,4-Benzoxazin-3(4H)-One (D-DIBOA), a Benzohydroxamic Acid with Interesting Biol. *Appl. Microbiol. Biotechnol.* **2012**, *94* (1), 163–171. <https://doi.org/10.1007/s00253-011-3787-0>.
- (338) Zenno, S.; Koike, H.; Tanokura, M.; Saigo, K. Gene Cloning, Purification, and Characterization of NfsB, a Minor Oxygen-Insensitive Nitroreductase from Escherichia Coli, Similar in Biochemical Properties to FRase I, the Major Flavin Reductase in Vibrio Fischeri. *J. Biochem.* **1996**, *120* (4), 736–744. <https://doi.org/10.1093/oxfordjournals.jbchem.a021473>.
- (339) Choi, J. W.; Lee, J.; Kosuke, N.; Jung, C. H.; Kim, J. S. Crystallization and Preliminary X-Ray Diffraction Analysis of YdjA, a Minimal Nitroreductase from Escherichia Coli K12. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2007**, *63* (12), 1064–1066. <https://doi.org/10.1107/S1744309107057636>.
- (340) Williams, R. E.; Rathbone, D. A.; Scrutton, N. S.; Bruce, N. C. Biotransformation of Explosives by the Old Yellow Enzyme Family of Flavoproteins. *Appl. Environ. Microbiol.* **2004**, *70* (6), 3566–3574. <https://doi.org/10.1128/AEM.70.6.3566-3574.2004>.
- (341) Kim, K. S.; Pelton, J. G.; Inwood, W. B.; Andersen, U.; Kustu, S.; Wemmer, D. E. The Rut Pathway for Pyrimidine Degradation: Novel Chemistry and Toxicity Problems. *J. Bacteriol.* **2010**, *192* (16), 4089–4102. <https://doi.org/10.1128/JB.00201-10>.
- (342) Jensen, S. I.; Lennen, R. M.; Herrgård, M. J.; Nielsen, A. T. Seven Gene Deletions in Seven Days: Fast Generation of Escherichia Coli Strains Tolerant to Acetate and Osmotic Stress. *Sci. Rep.* **2015**, *5* (June), 1–10. <https://doi.org/10.1038/srep17874>.
- (343) Datsenko, K. A.; Wanner, B. L. One-Step Inactivation of Chromosomal Genes in Escherichia Coli K-12 Using PCR Products. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*

- (12), 6640–6645. <https://doi.org/10.1073/pnas.120163297>.
- (344) Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H. Construction of Escherichia Coli K-12 In-frame, Single-gene Knockout Mutants: The Keio Collection. *Mol. Syst. Biol.* **2006**, *2* (1). <https://doi.org/10.1038/msb4100050>.
- (345) Rodríguez-Rojas, A.; Kim, J. J.; Johnston, P. R.; Makarova, O.; Eravci, M.; Weise, C.; Hengge, R.; Rolff, J. Non-Lethal Exposure to H₂O₂ Boosts Bacterial Survival and Evolvability against Oxidative Stress. *PLOS Genet.* **2020**, *16* (3), e1008649. <https://doi.org/10.1371/journal.pgen.1008649>.
- (346) Swings, T.; Marciano, D. C.; Atri, B.; Bosserman, R. E.; Wang, C.; Leysen, M.; Bonte, C.; Schalck, T.; Furey, I.; Van den Bergh, B.; Verstraeten, N.; Christie, P. J.; Herman, C.; Lichtarge, O.; Michiels, J. CRISPR-FRT Targets Shared Sites in a Knock-out Collection for off-the-Shelf Genome Editing. *Nat. Commun.* **2018**, *9* (1), 2231. <https://doi.org/10.1038/s41467-018-04651-5>.
- (347) Meuskens, I.; Michalik, M.; Chauhan, N.; Linke, D.; Leo, J. C. A New Strain Collection for Improved Expression of Outer Membrane Proteins. *Front. Cell. Infect. Microbiol.* **2017**, *7* (NOV), 1–13. <https://doi.org/10.3389/fcimb.2017.00464>.
- (348) Aedo, S. J.; Ma, H. R.; Brynildsen, M. P. Checks and Balances with Use of the Keio Collection for Phenotype Testing. In *Microbial Metabolic Engineering*; Santos C., A. P., Ed.; Humana Press: New York, NY, 2019; Vol. 1927, pp 125–138. https://doi.org/10.1007/978-1-4939-9142-6_9.
- (349) Sharan, S. K.; Thomason, L. C.; Kuznetsov, S. G.; Court, D. L. Recombineering: A Homologous Recombination-Based Method of Genetic Engineering. *Nat. Protoc.* **2009**, *4* (2), 206–223. <https://doi.org/10.1038/nprot.2008.227>.
- (350) Murphy, K. C. Use of Bacteriophage λ Recombination Functions To Promote Gene Replacement in Escherichia Coli. *J. Bacteriol.* **1998**, *180* (8), 2063–2071. <https://doi.org/10.1128/JB.180.8.2063-2071.1998>.
- (351) Liu, P.; Jenkins, N. A.; Copeland, N. G. A Highly Efficient Recombineering-Based Method for Generating Conditional Knockout Mutations. *Genome Res.* **2003**, *13* (3), 476–484. <https://doi.org/10.1101/gr.749203>.
- (352) Alieva, N. O.; Konzen, K. A.; Field, S. F.; Meleshkevitch, E. A.; Hunt, M. E.; Beltran-Ramirez, V.; Miller, D. J.; Wiedenmann, J.; Salih, A.; Matz, M. V. Diversity and Evolution of Coral Fluorescent Proteins. *PLoS One* **2008**, *3* (7), e2680. <https://doi.org/10.1371/journal.pone.0002680>.
- (353) Groff, D.; Wang, F.; Jockusch, S.; Turro, N. J.; Schultz, P. G. A New Strategy to Photoactivate Green Fluorescent Protein. *Angew. Chemie - Int. Ed.* **2010**, *49* (42), 7677–7679. <https://doi.org/10.1002/anie.201003797>.
- (354) Nickling, J. H.; Baumann, T.; Schmitt, F.-J.; Bartholomae, M.; Kuipers, O. P.; Friedrich, T.; Budisa, N. Antimicrobial Peptides Produced by Selective Pressure Incorporation of Non-Canonical Amino Acids. *J. Vis. Exp.* **2018**, No. 135, 1–12. <https://doi.org/10.3791/57551>.
- (355) Baumann, T.; Nickling, J. H.; Bartholomae, M.; Buivydas, A.; Kuipers, O. P.; Budisa, N. Prospects of In Vivo Incorporation of Non-Canonical Amino Acids for the Chemical Diversification of Antimicrobial Peptides. *Front. Microbiol.* **2017**, *8* (FEB), 1–9. <https://doi.org/10.3389/fmicb.2017.00124>.
- (356) Li, J.; Celiz, A. D.; Yang, J.; Yang, Q.; Wamala, I.; Whyte, W.; Seo, B. R.; Vasilyev, N. V.; Vlassak, J. J.; Suo, Z.; Mooney, D. J. Tough Adhesives for Diverse Wet Surfaces. *Science (80-)*. **2017**, *357* (6349), 378–381. <https://doi.org/10.1126/science.aah6362>.
- (357) Hong, Y.; Zhou, F.; Hua, Y.; Zhang, X.; Ni, C.; Pan, D.; Zhang, Y.; Jiang, D.; Yang, L.;

- Lin, Q.; Zou, Y.; Yu, D.; Arnot, D. E.; Zou, X.; Zhu, L.; Zhang, S.; Ouyang, H. A Strongly Adhesive Hemostatic Hydrogel for the Repair of Arterial and Heart Bleeds. *Nat. Commun.* **2019**, *10* (1), 1–11. <https://doi.org/10.1038/s41467-019-10004-7>.
- (358) Cooley, R. B.; Feldman, J. L.; Driggers, C. M.; Bundy, T. A.; Stokes, A. L.; Karplus, P. A.; Mehl, R. A. Structural Basis of Improved Second-Generation 3-Nitro-Tyrosine TRNA Synthetases. *Biochemistry* **2014**, *53* (12), 1916–1924. <https://doi.org/10.1021/bi5001239>.
- (359) Schwark, D. G.; Schmitt, M. A.; Fisk, J. D. Directed Evolution of the Methanosarcina Barkeri Pyrrolysyl TRNA/Aminoacyl TRNA Synthetase Pair for Rapid Evaluation of Sense Codon Reassignment Potential. *Int. J. Mol. Sci.* **2021**, *22* (2), 895. <https://doi.org/10.3390/ijms22020895>.
- (360) Robertson, W. E.; Funke, L. F. H.; de la Torre, D.; Fredens, J.; Elliott, T. S.; Spinck, M.; Christova, Y.; Cervettini, D.; Böge, F. L.; Liu, K. C.; Buse, S.; Maslen, S.; Salmond, G. P. C.; Chin, J. W. Sense Codon Reassignment Enables Viral Resistance and Encoded Polymer Synthesis. *Science* (80-.). **2021**, *372* (6546), 1057–1062. <https://doi.org/10.1126/science.abg3029>.
- (361) Tian, F.; Lu, Y.; Manibusan, A.; Sellers, A.; Tran, H.; Sun, Y.; Phuong, T.; Barnett, R.; Hehli, B.; Song, F.; DeGuzman, M. J.; Ensari, S.; Pinkstaff, J. K.; Sullivan, L. M.; Biroc, S. L.; Cho, H.; Schultz, P. G.; DiJoseph, J.; Dougher, M.; Ma, D.; Dushin, R.; Leal, M.; Tchistiakova, L.; Feyfant, E.; Gerber, H.-P.; Sapra, P. A General Approach to Site-Specific Antibody Drug Conjugates. *Proc. Natl. Acad. Sci.* **2014**, *111* (5), 1766–1771. <https://doi.org/10.1073/pnas.1321237111>.
- (362) Skidmore, L.; Sakamuri, S.; Knudsen, N. A.; Hewet, A. G.; Milutinovic, S.; Barkho, W.; Biroc, S. L.; Kirtley, J.; Marsden, R.; Storey, K.; Lopez, I.; Yu, W.; Fang, S.-Y.; Yao, S.; Gu, Y.; Tian, F. ARX788, a Site-Specific Anti-HER2 Antibody–Drug Conjugate, Demonstrates Potent and Selective Activity in HER2-Low and T-DM1–Resistant Breast and Gastric Cancers. *Mol. Cancer Ther.* **2020**, *19* (9), 1833–1843. <https://doi.org/10.1158/1535-7163.MCT-19-1004>.
- (363) Jin, X.; Park, O.-J.; Hong, S. H. Incorporation of Non-Standard Amino Acids into Proteins: Challenges, Recent Achievements, and Emerging Applications. *Appl. Microbiol. Biotechnol.* **2019**, *103* (7), 2947–2958. <https://doi.org/10.1007/s00253-019-09690-6>.
- (364) Conrad, A. FDA Grants ARX788 Fast Track Designation for HER2-positive Metastatic Breast Cancer <https://www.prnewswire.com/news-releases/fda-grants-arx788-fast-track-designation-for-her2-positive-metastatic-breast-cancer-301199951.html>.
- (365) Sternberg, A. FDA Action Moves ARX788 Forward in the Treatment of HER2-Positive Breast Cancer <https://www.cancernetwork.com/view/fda-action-moves-arx788-forward-in-the-treatment-of-her2-positive-breast-cancer>.
- (366) Studier, F. W. Protein Production by Auto-Induction in High Density Shaking Cultures. *Protein Expr. Purif.* **2005**, No. 41, 207–234. <https://doi.org/10.1016/j.pep.2005.01.016>.
- (367) Studier, F. W.; Moffatt, B. A. Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-Level Expression of Cloned Genes. *J. Mol. Biol.* **1986**, *189* (1), 113–130.
- (368) Dominy, C. N.; Andrews, D. W. Site-Directed Mutagenesis by Inverse PCR. In *E. coli Plasmid Vectors*; Humana Press: New Jersey; pp 209–224. <https://doi.org/10.1385/1-59259-409-3:209>.
- (369) Engler, C.; Kandzia, R.; Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS One* **2008**, *3* (11). <https://doi.org/10.1371/journal.pone.0003647>.
- (370) Khusainov, R.; Kuipers, O. P. The Presence of Modifiable Residues in the Core Peptide Part of Precursor Nisin Is Not Crucial for Precursor Nisin Interactions with NisB- and

NisC. *PLoS One* **2013**, *8* (9), e74890. <https://doi.org/10.1371/journal.pone.0074890>.

- (371) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Humana Press: Totowa, NJ, 2005; pp 571–607. <https://doi.org/10.1385/1-59259-890-0:571>.
- (372) Edelhoch, H. Spectroscopic Determination of Tryptophan and Tyrosine in Proteins. *Biochemistry* **1967**, *6* (7), 1948–1954.
- (373) Pace, C. N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T. How to Measure and Predict the Molar Absorption Coefficient of a Protein. *Protein Sci.* **1995**, *4* (11), 2411–2423. <https://doi.org/10.1002/pro.5560041120>.
- (374) Gray, R. M. *Probability, Random Processes, and Ergodic Properties, 2nd Ed.*; Springer US: New York, NY, 2009. <https://doi.org/10.1007/978-1-4419-1090-5>.
- (375) Boltzmann, L. Vorlesungen Über Gastheorie. In *Vorlesungen über Gastheorie*; Johann Ambrosius Barth: Leipzig, 1898; p 89.

6. Appendix

6.1. Additional Results

6.1.1. Fluorescence Assays

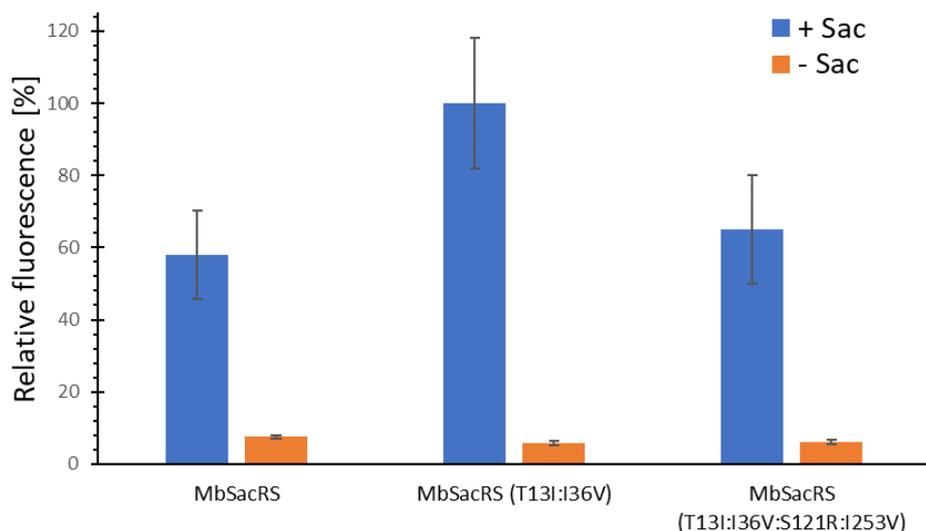


Figure 70: OTS efficiency comparison of three different *MbSacRS* constructs. Ribosomal incorporation of Sac (+ Sac = 2 mM) and controls without ncAA (- Sac) supplementation. Intact cell fluorescence of *E. coli* BL21(DE3), endpoint measurements after 24 h of incubation. Relative fluorescence is normalized to the highest value. The data (incl. standard deviation) represent the mean of three biological replicates.

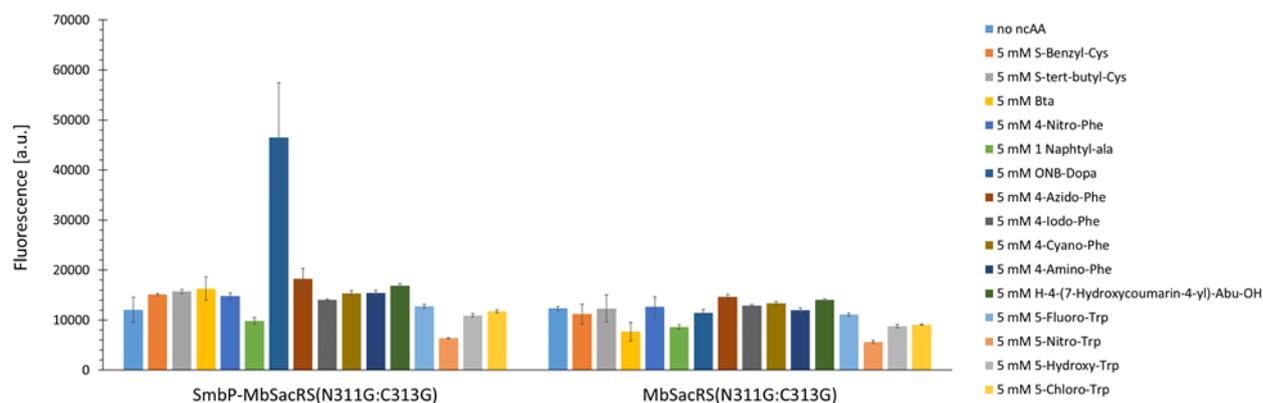


Figure 71: OTS efficiency comparison of *MbPylRS*(N311G:C313G) co-expression with and without SmbP-tag. Ribosomal incorporation of 15 different ncAAs. Intact cell fluorescence of *E. coli* BL21(DE3) cells expressing the sfGFP(R2 amber) reporter, endpoint measurements after 16 h of incubation for the presence or absence of ncAA supplementation. Relative fluorescence is normalized to the highest value. The data (incl. standard deviation) represent the mean of three biological replicates.

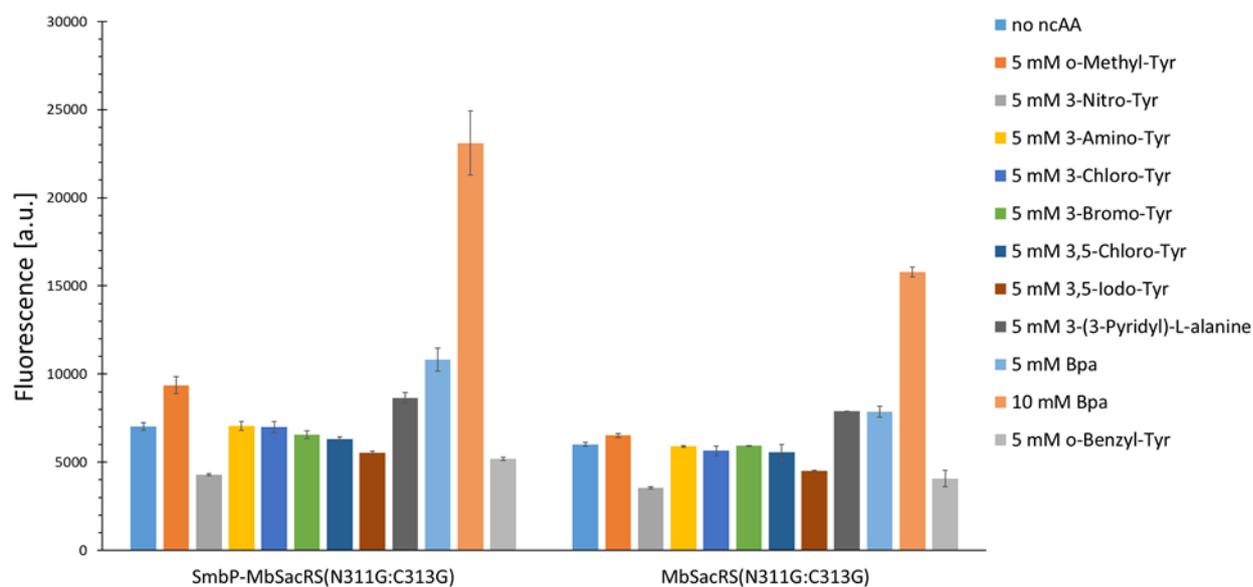


Figure 72: OTS efficiency comparison of *MbPylRS*(N311G:C313G) co-expression with and without SmbP-tag. Ribosomal incorporation of 11 different ncAAs. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the sfGFP(R2 amber) reporter, endpoint measurements after 24 h of incubation for the presence or absence of ncAA supplementation. Relative fluorescence is normalized to the highest value. The data (incl. standard deviation) represent the mean of three biological replicates.

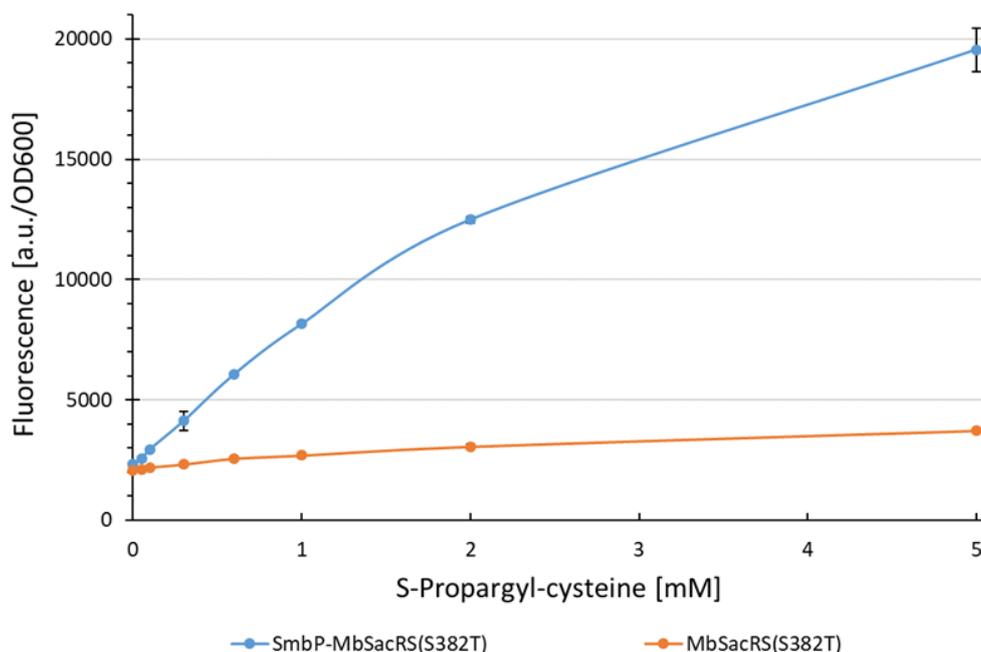


Figure 73: Concentration dependent protein production for *MbPylRS* with and without SmbP-tag. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the sfGFP(R2 amber) reporter, endpoint measurements after 24 h. Endpoint measurements with different ncAA concentrations (0.05, 0.1, 0.3, 0.6, 1, 2, and 5 mM) of S-Propargyl-L-cysteine (27).

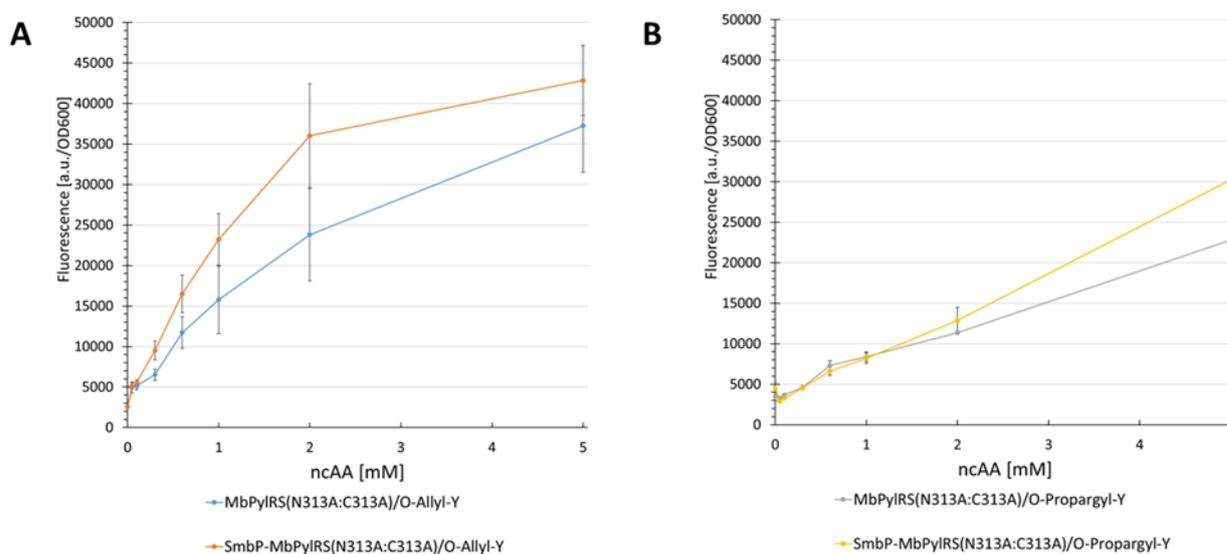


Figure 74: Concentration dependent protein production for different *MbPylRS*/ncAA combinations. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the sfGFP(R2 amber) reporter, endpoint measurements after 24 h. Endpoint measurements with different ncAA concentrations (0.05, 0.1, 0.3, 0.6, 1, 2, and 5 mM). The ncAAs are: **A**) O-Allyl-Y = O-Allyl-L-tyrosine **B**) O-Propargyl-Y = O-Propargyl-L-tyrosine.

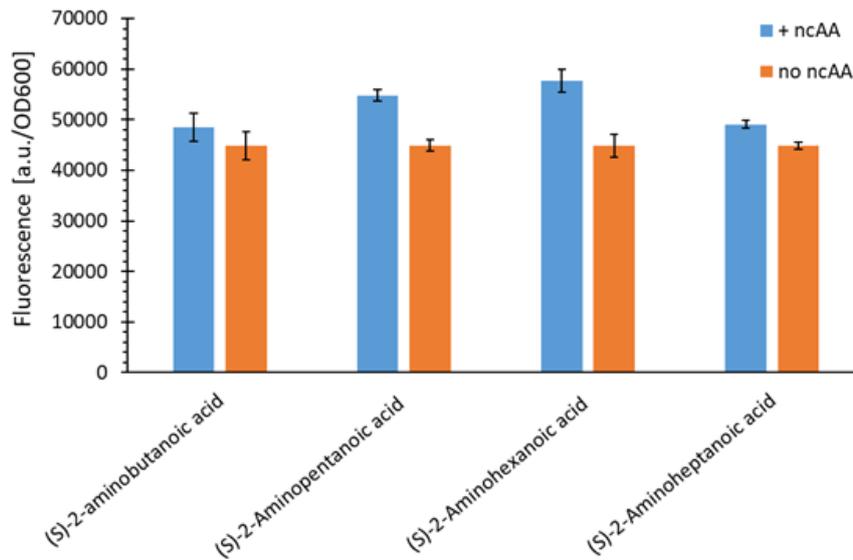


Figure 75: Comparison of ribosomal incorporation efficiency of MbSacRS with four different ncAAs (2, 3, 5, 7, 10) (10 mM) and controls without ncAA supplementation. Intact cell fluorescence of *E. coli* BL21(DE3), endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied.

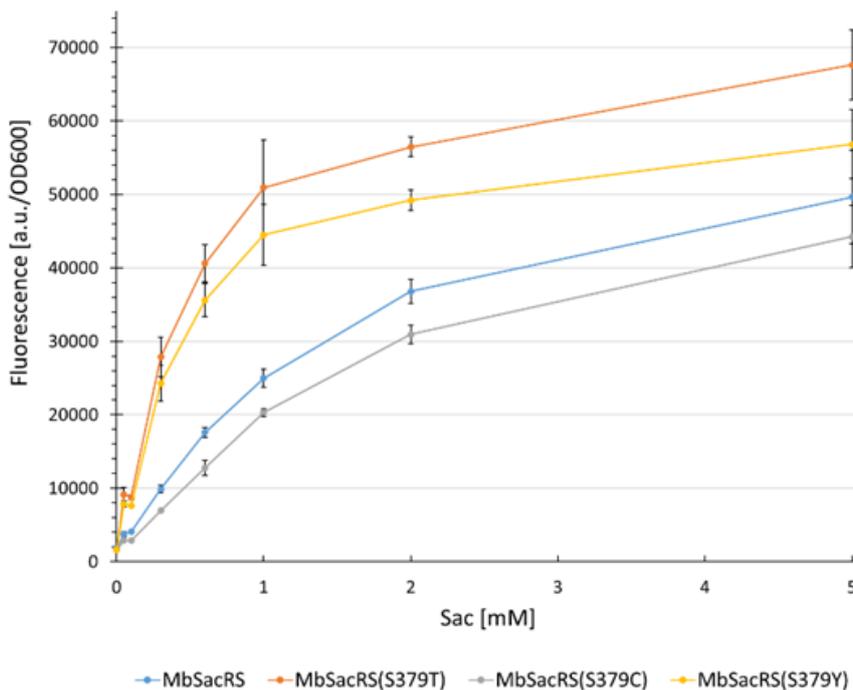


Figure 76: Concentration dependent protein production for MbPyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the sfGFP(R2 amber) reporter, endpoint measurements after 24 h. Endpoint measurements with different ncAA concentrations (0.05, 0.1, 0.3, 0.6, 1, 2, and 5 mM) of *S*-allyl-L-cysteine (Sac, 1). The data (incl. standard deviation) represent the mean of three biological replicates. Gain 90.

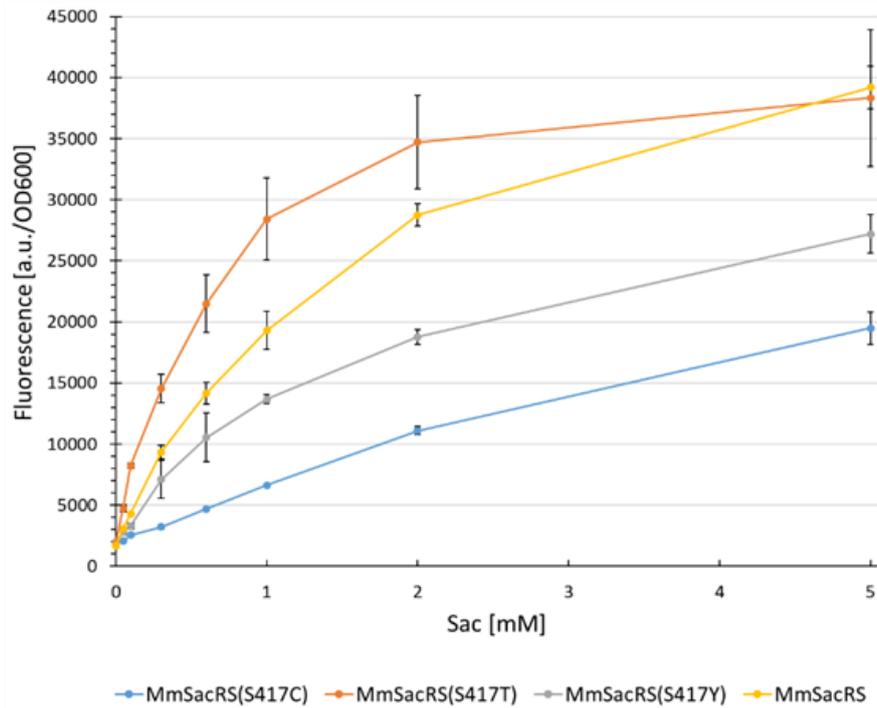


Figure 77: Concentration dependent protein production for *MmPyIRS* variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the sfGFP(R2 amber) reporter, endpoint measurements after 24 h. Endpoint measurements with different ncAA concentrations (0.05, 0.1, 0.3, 0.6, 1, 2, and 5 mM) of *S*-allyl-L-cysteine (Sac, **1**). The data (incl. standard deviation) represent the mean of three biological replicates. Gain 85.

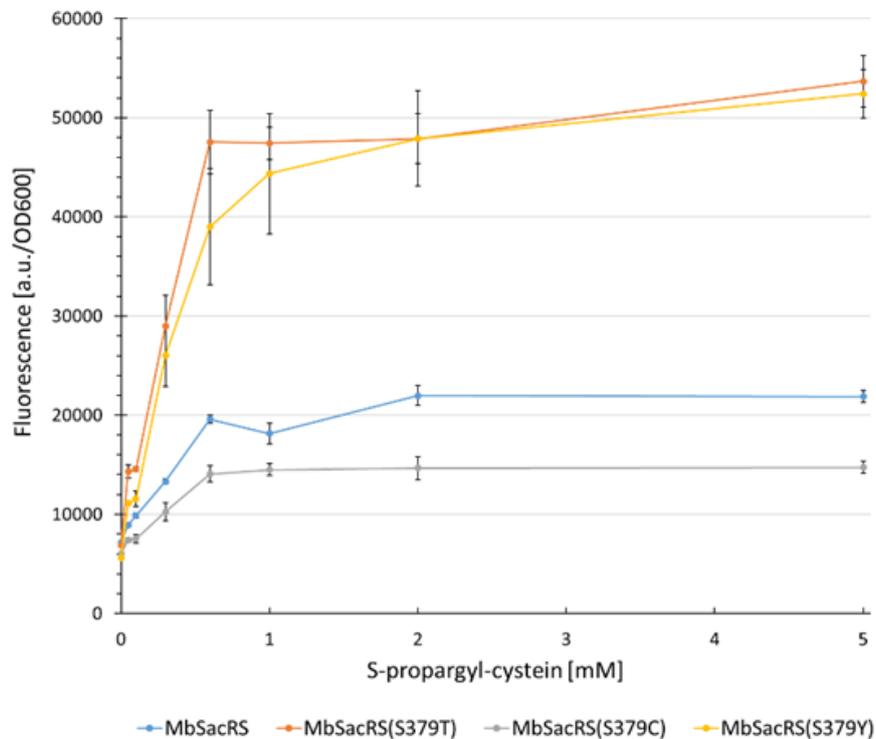


Figure 78: Concentration dependent protein production for *MbPyIRS* variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the sfGFP(R2 amber) reporter, endpoint measurements after 24 h. Endpoint measurements with different ncAA concentrations (0.05, 0.1, 0.3, 0.6, 1, 2, and 5 mM) of *S*-propargyl-L-cysteine (**27**). The data (incl. standard deviation) represent the mean of three biological replicates.

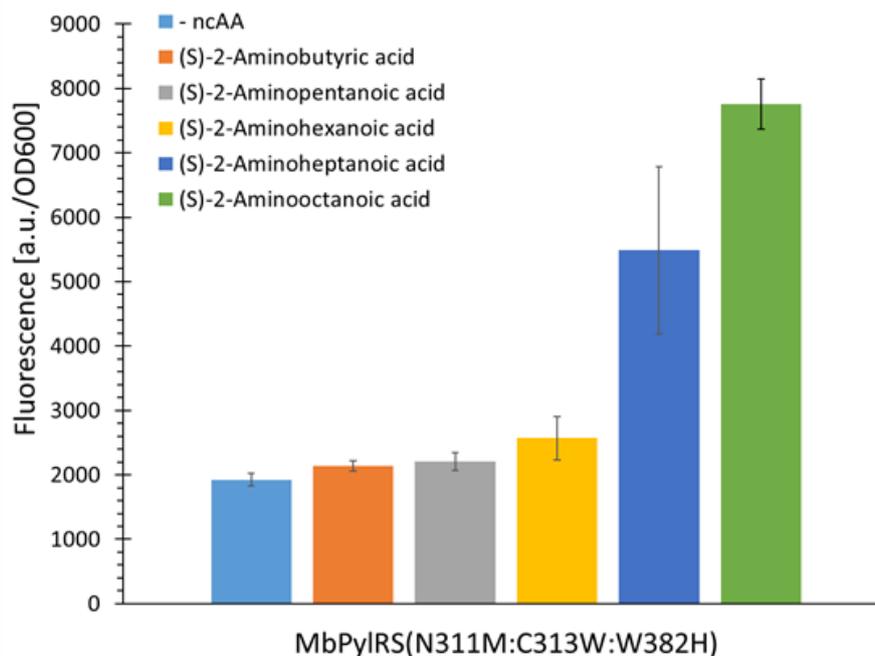


Figure 79: Comparison of aliphatic ncAA (**2**, **3**, **5**, **7**, **10**) incorporation efficiency for *MbPylRS(N311M:C313W:W382H)*. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Intact cell fluorescence of *E. coli* BL21(DE3), endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied.

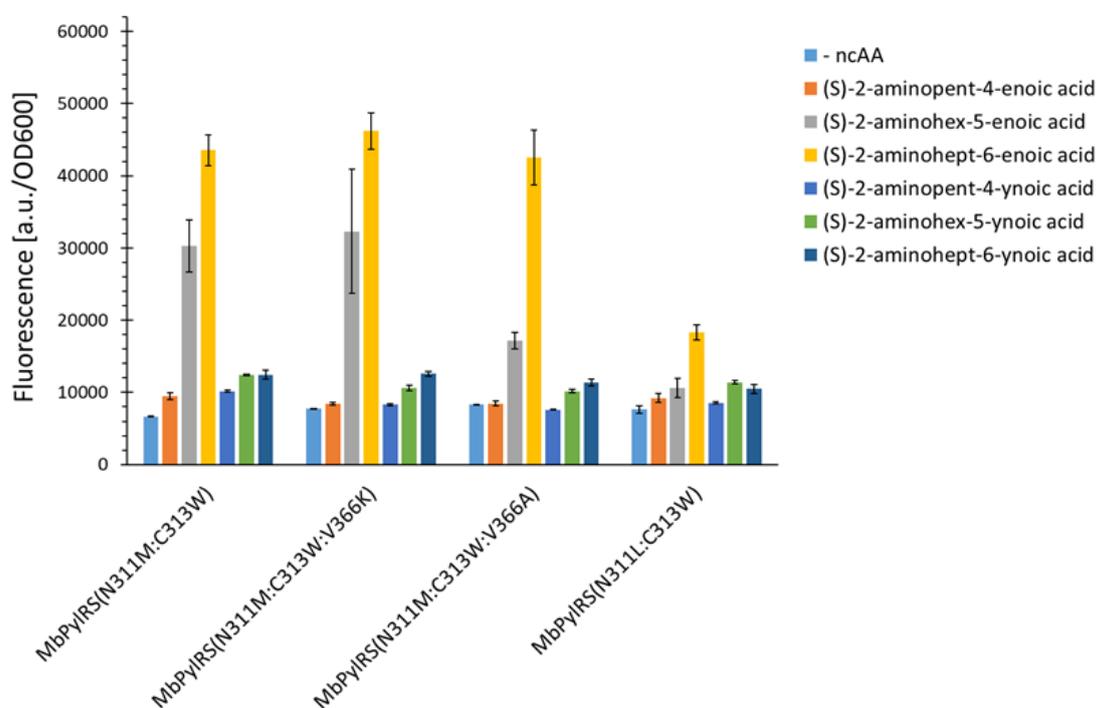


Figure 80: Comparison of aliphatic ncAA (**4**, **6**, **8**, **11**, **12**, **13**) incorporation efficiency for *MbPylRS* variants. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Intact cell fluorescence of *E. coli* BL21(DE3), endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied.

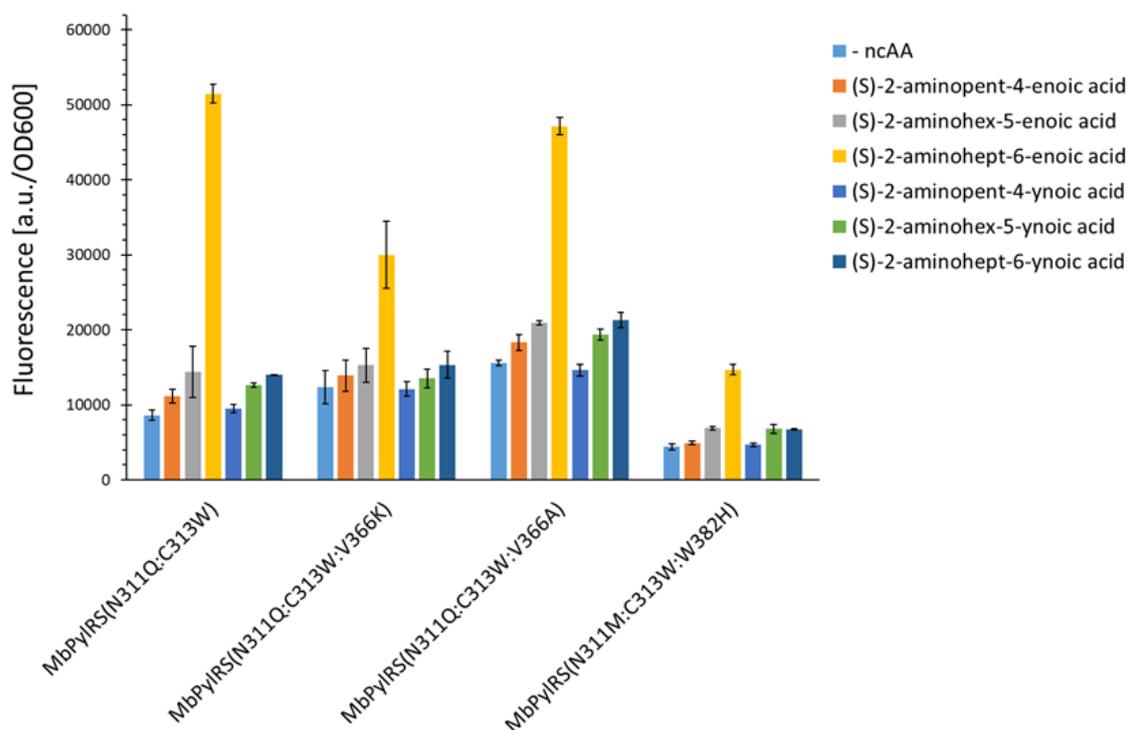


Figure 81: Comparison of aliphatic ncAA (**4**, **6**, **8**, **11**, **12**, **13**) incorporation efficiency for *MbPyIRS* variants. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Intact cell fluorescence of *E. coli* BL21(DE3), endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

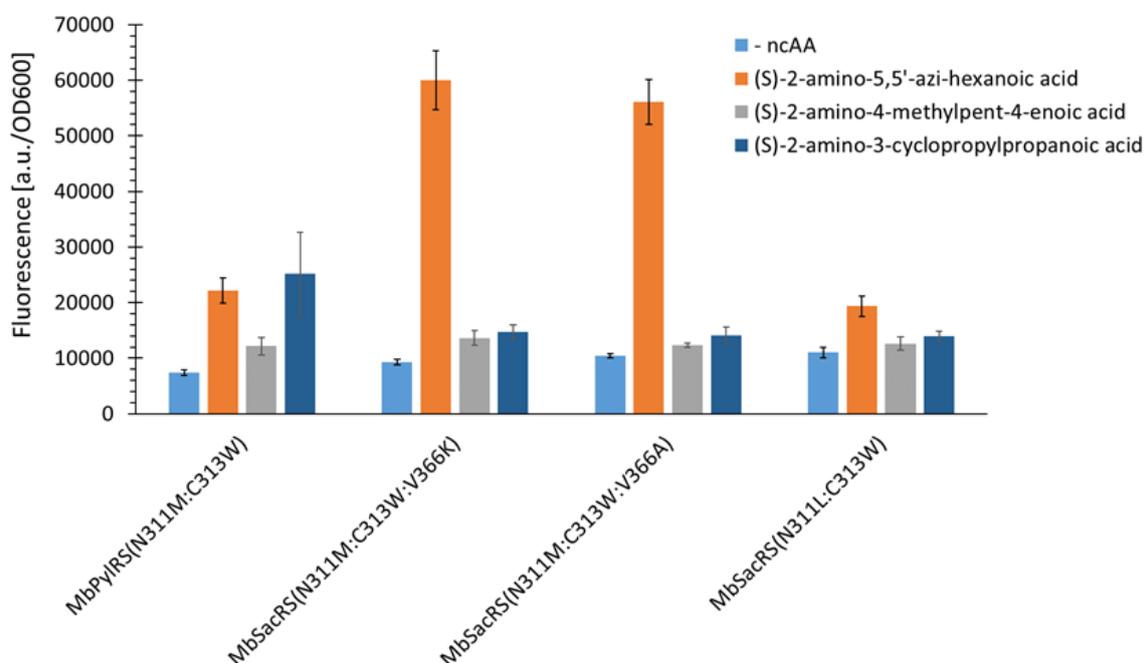


Figure 82: Comparison of (S)-2-amino-5,5'-azi-hexanoic acid (**20**), (S)-2-amino-4-methylpent-4-enoic acid (**21**) and (S)-2-amino-3-cyclopropylpropanoic acid (**9**) incorporation efficiency for *MbPyIRS* variants. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Intact cell fluorescence of *E. coli* BL21(DE3), endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

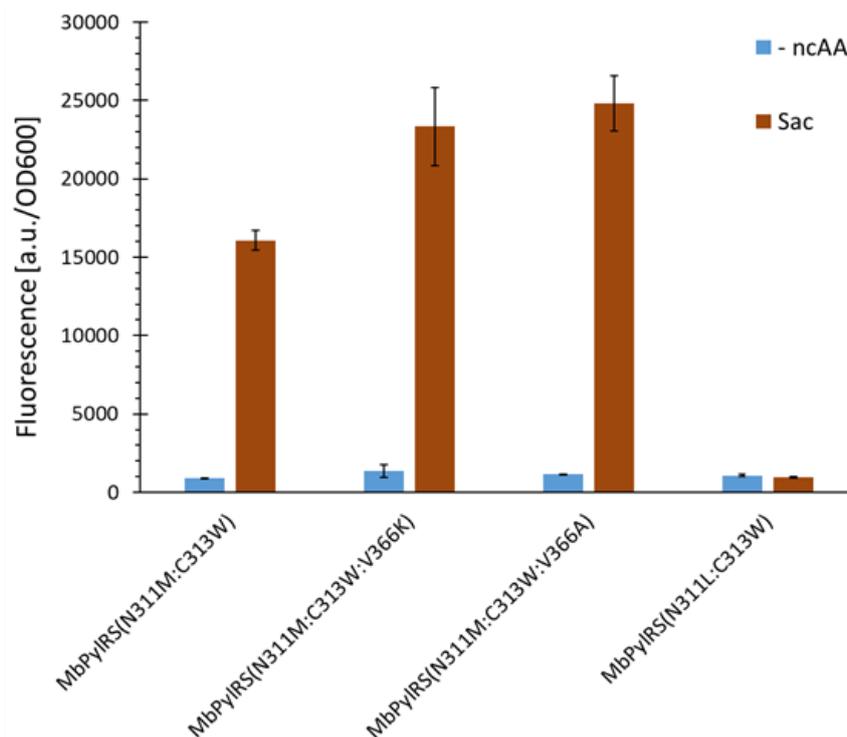


Figure 83: Comparison of S-allyl-L-cysteine (Sac, **1**) incorporation efficiency for *MbPyIRS*(N311M:C313W) variants mutate at position V366. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Relative fluorescence is normalized to the highest value. Intact cell fluorescence of *E. coli* BL21(DE3), endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

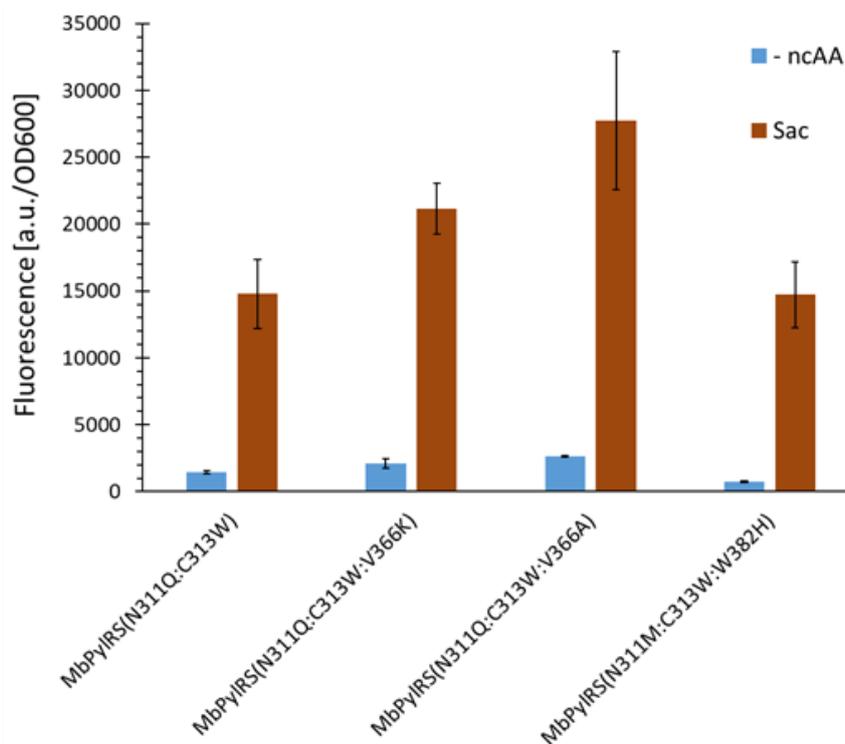


Figure 84: Comparison of S-allyl-L-cysteine (Sac, **1**) incorporation efficiency for *MbPyIRS*(N311Q:C313W) variants mutate at position V366. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Relative fluorescence is normalized to the highest value. Intact cell fluorescence of *E. coli* BL21(DE3), endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

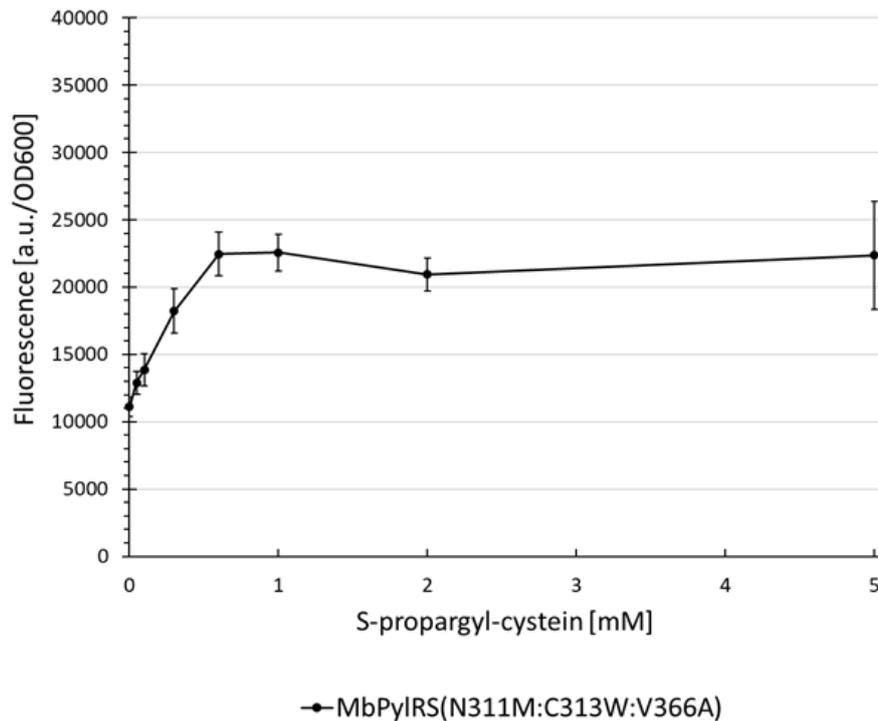


Figure 85: Concentration dependent protein production for *MbPyIRS(N311M:C313W:V366A)*. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the sfGFP(R2 amber) reporter, endpoint measurements after 24 h. Endpoint measurements with different ncAA concentrations (0.05, 0.1, 0.3, 0.6, 1, 2, and 5 mM) of *S*-propargyl-L-cysteine (27). The data (incl. standard deviation) represent the mean of three biological replicates.

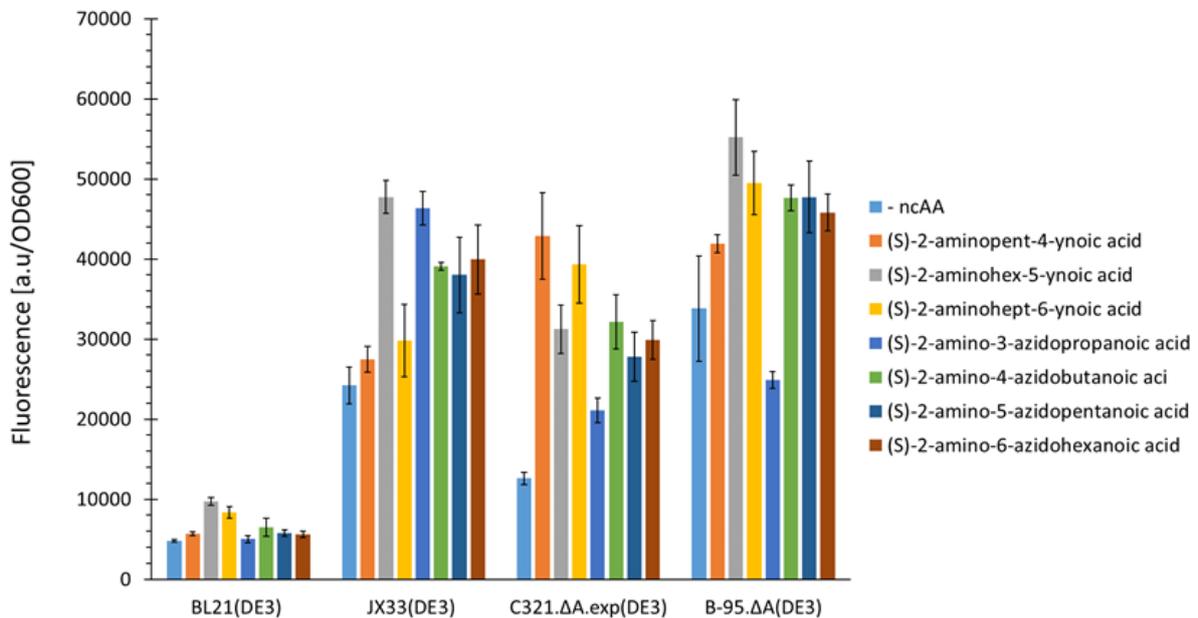


Figure 86: Comparison of ncAA (11, 12, 13, 14, 15, 16, 17) incorporation efficiency for *MbPyIRS(N311M:C313W)* in different *E. coli* host strains. Fluorescence measurement of intact *E. coli* cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

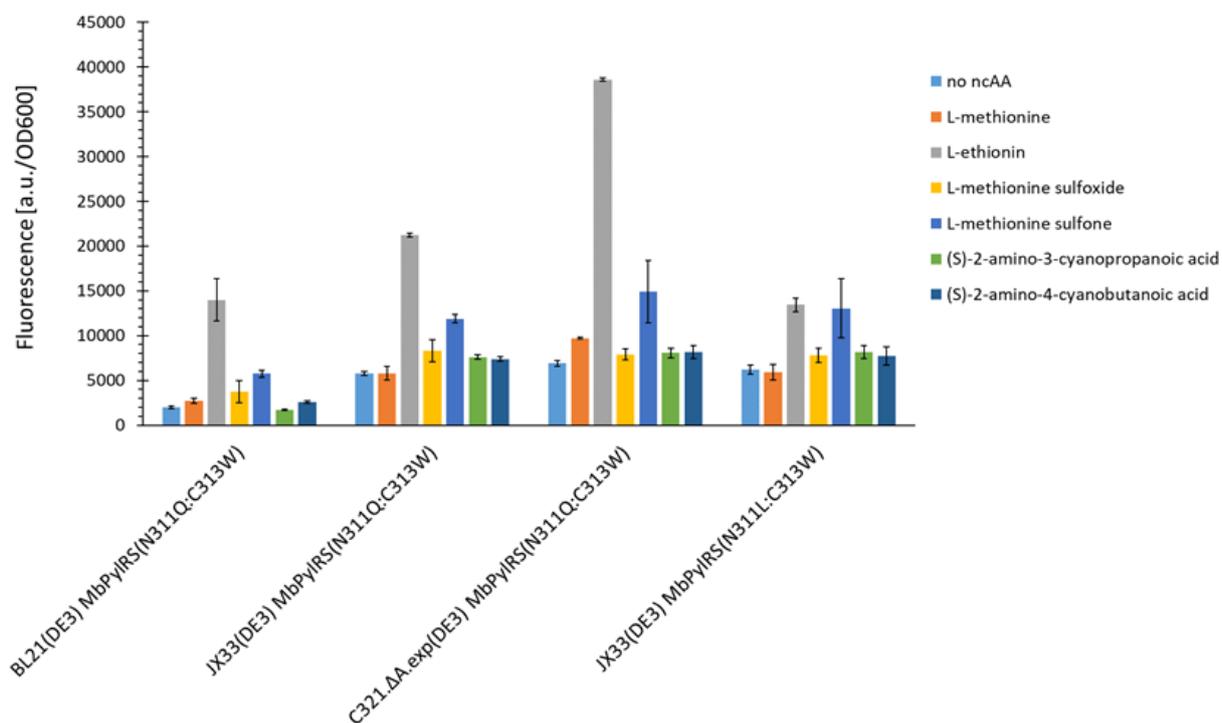


Figure 87: Comparison of ncAA (**18**, **19**, **22**, **23**, **24**, **25**) incorporation efficiency for *MbPyIRS* constructs in different *E. coli* host strains. Fluorescence measurement of intact *E. coli* cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

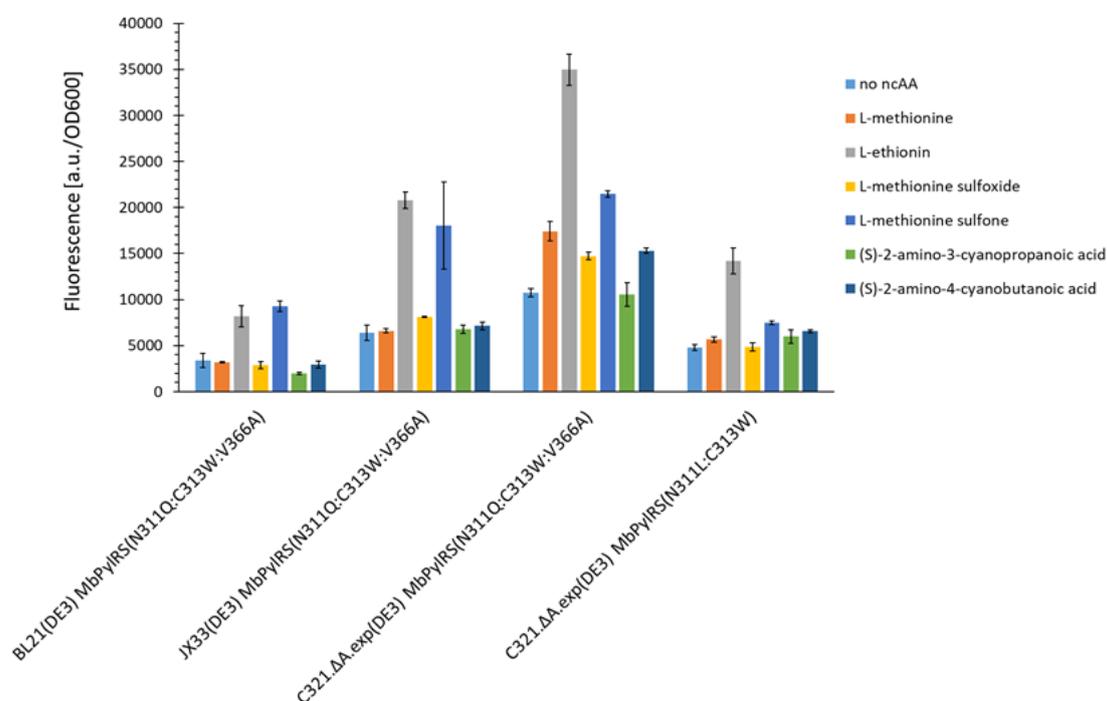


Figure 88: Comparison of ncAA (**18**, **19**, **22**, **23**, **24**, **25**) incorporation efficiency for *MbPyIRS* constructs in different *E. coli* host strains. Fluorescence measurement of intact *E. coli* cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied.

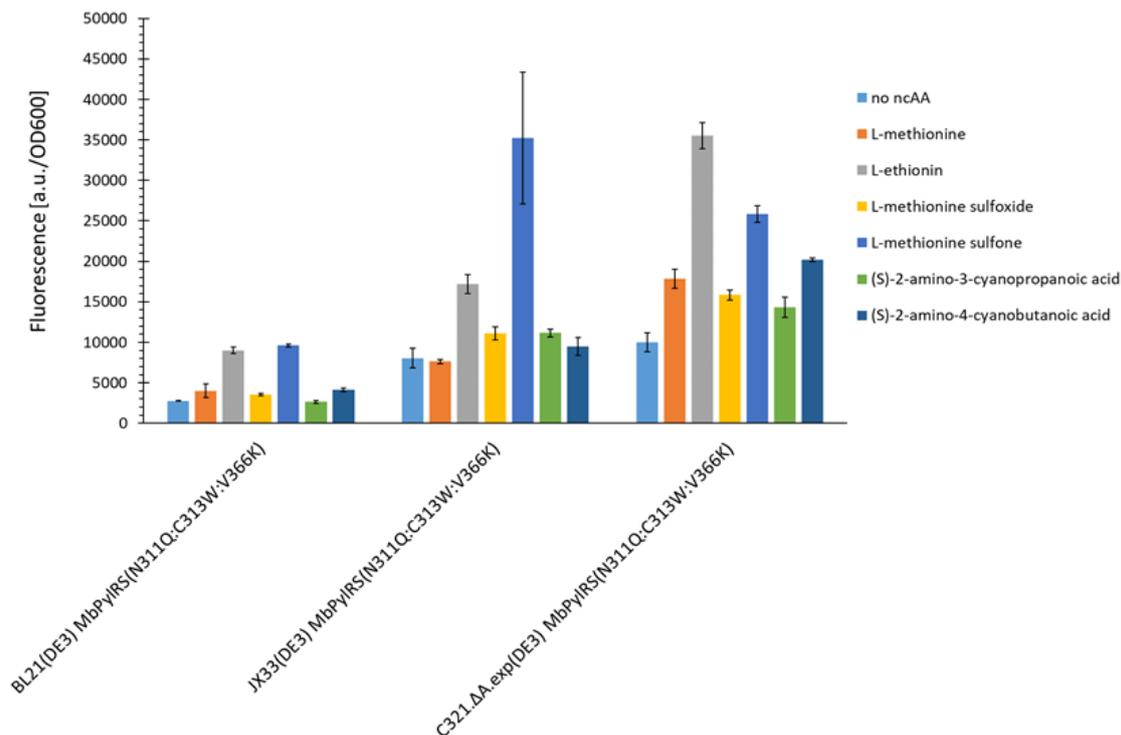


Figure 89: Comparison of ncAA (**18**, **19**, **22**, **23**, **24**, **25**) incorporation efficiency for MbPyIRS constructs in different *E. coli* host strains. Fluorescence measurement of intact *E. coli* cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

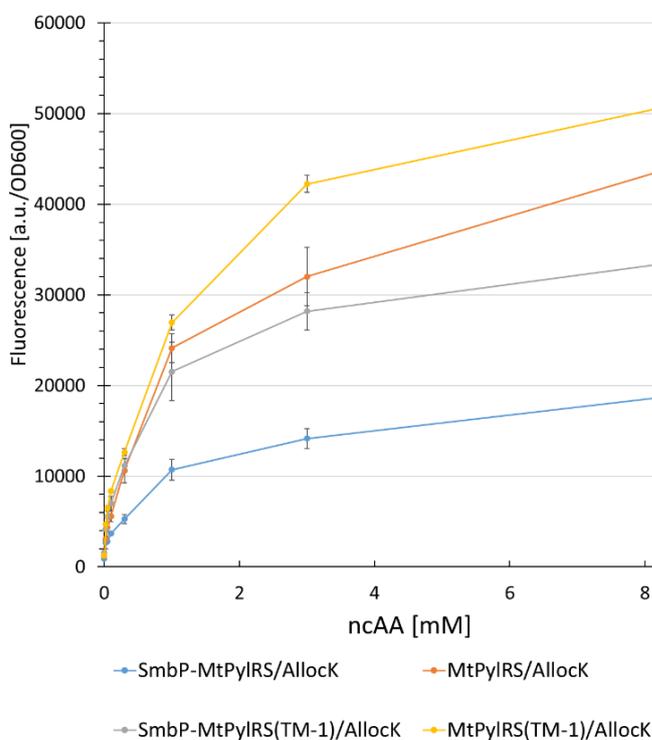


Figure 90: Concentration dependent protein production for four different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

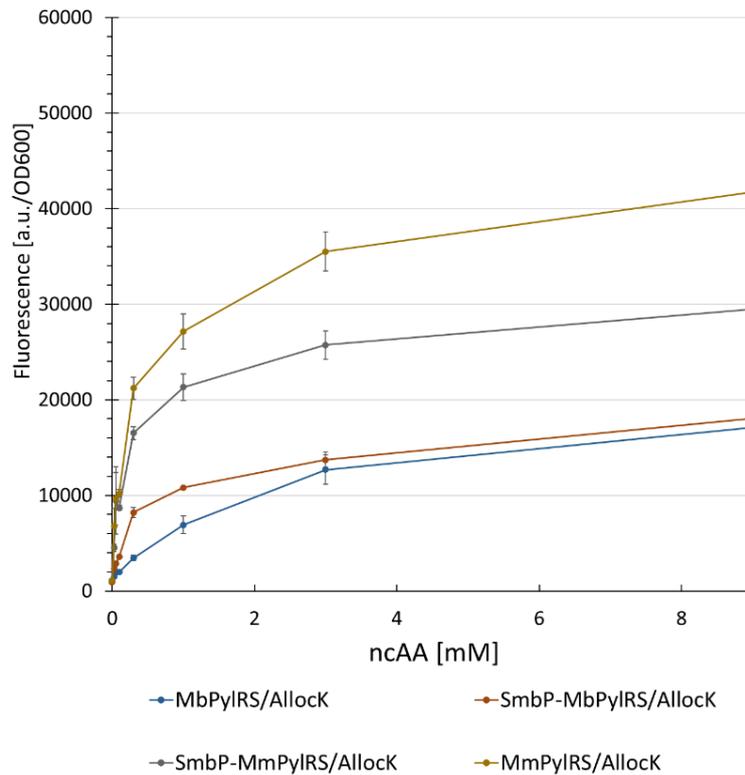


Figure 91: Concentration dependent protein production for four different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

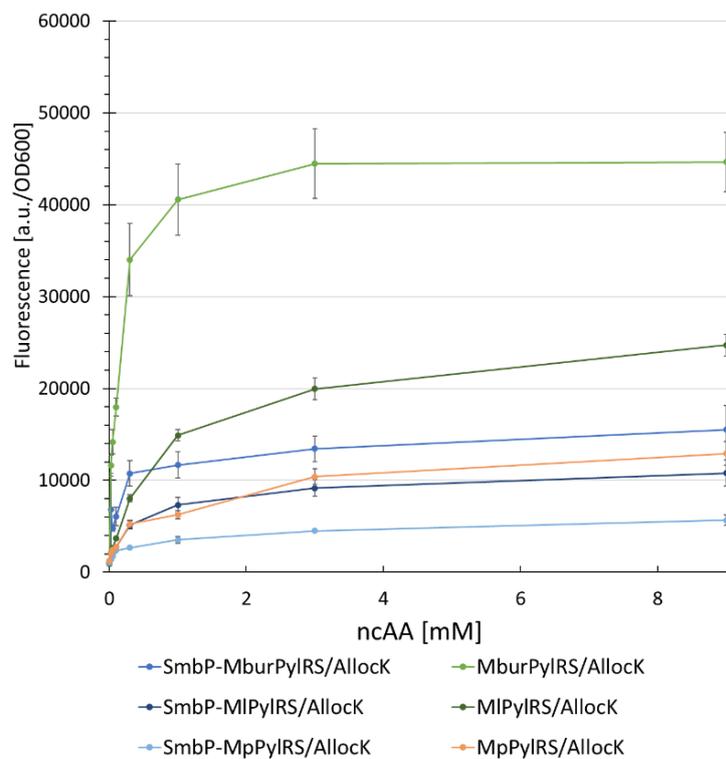


Figure 92: Concentration dependent protein production for six different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements

after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

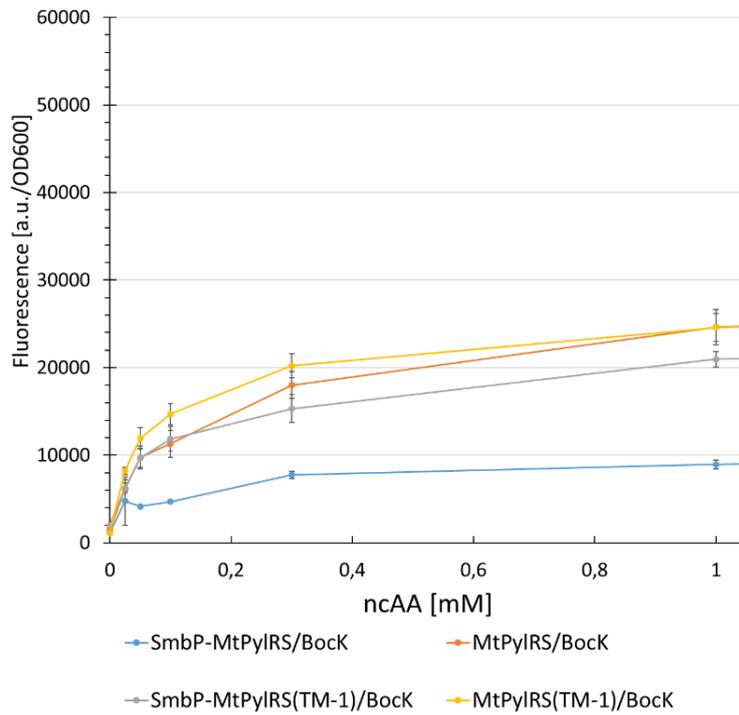


Figure 93: Concentration dependent protein production for four different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

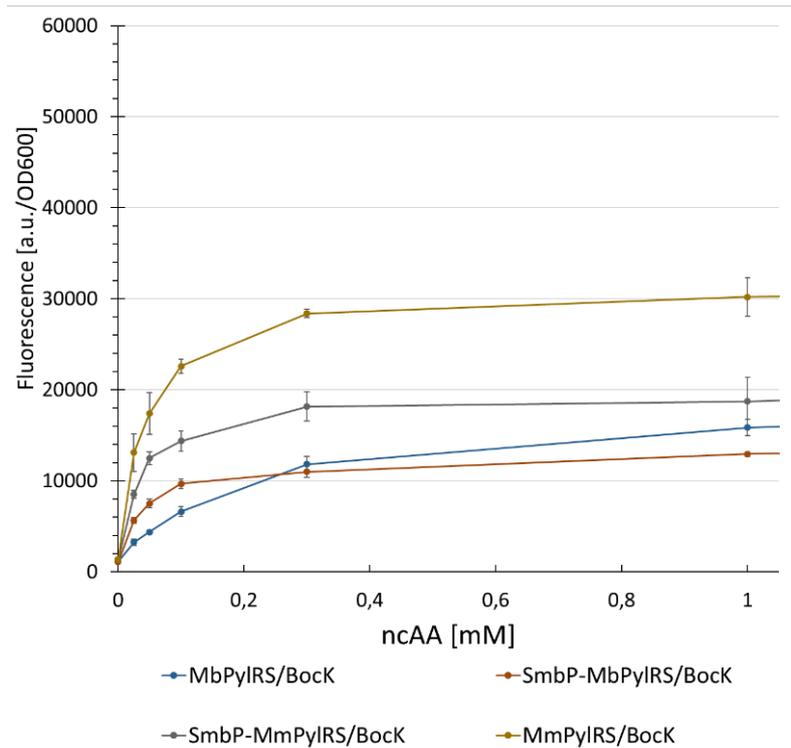


Figure 94: Concentration dependent protein production for four different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

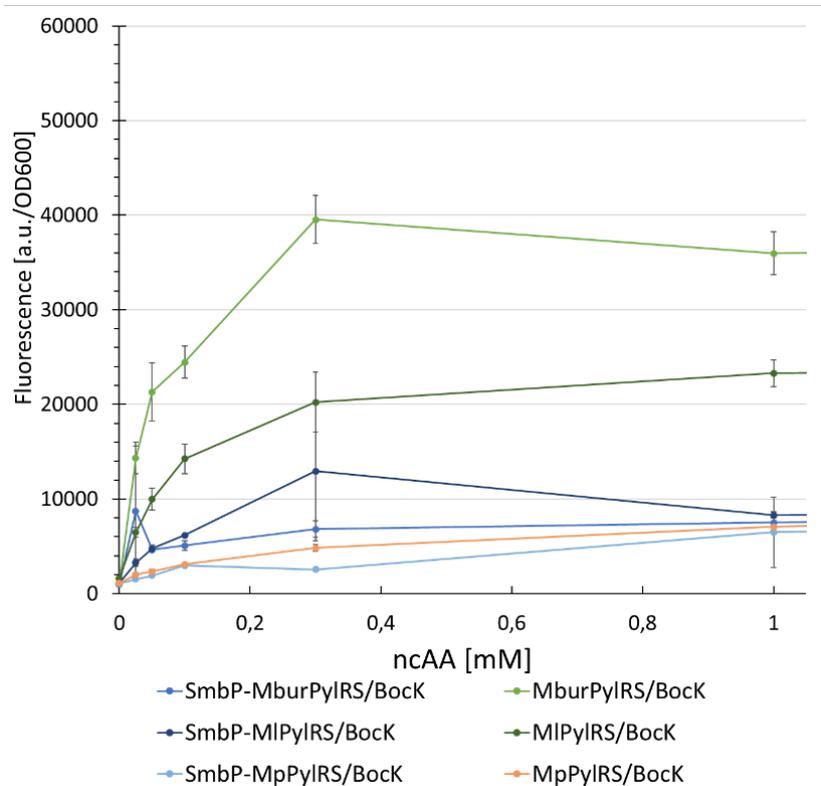


Figure 95: Concentration dependent protein production for six different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

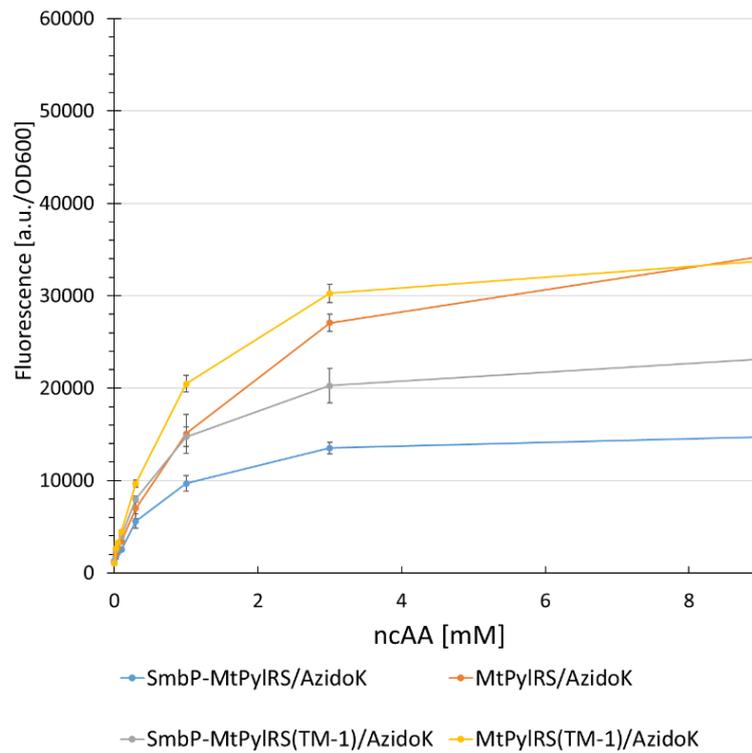


Figure 96: Concentration dependent protein production for four different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

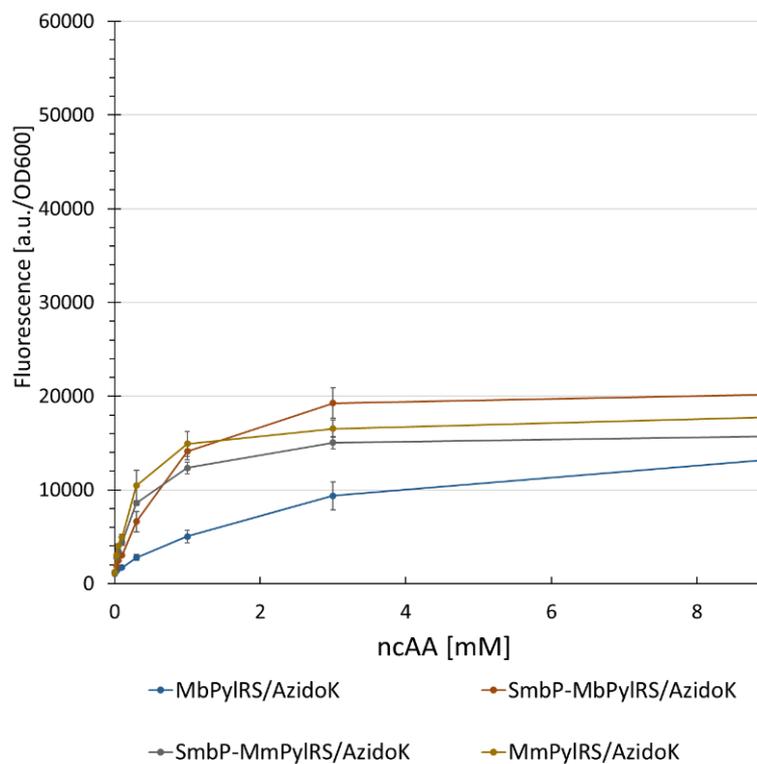


Figure 97: Concentration dependent protein production for four different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

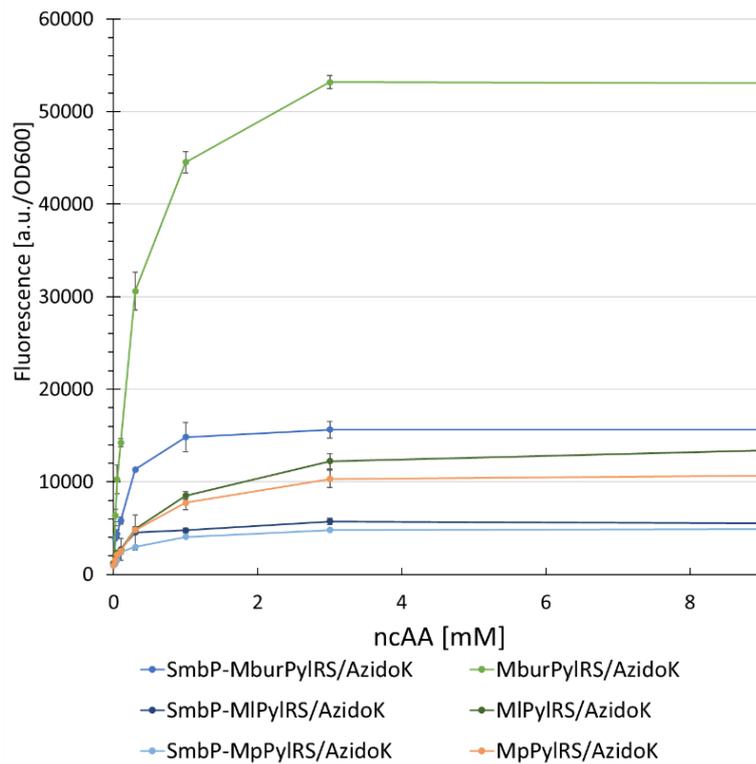


Figure 98: Concentration dependent protein production for six different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

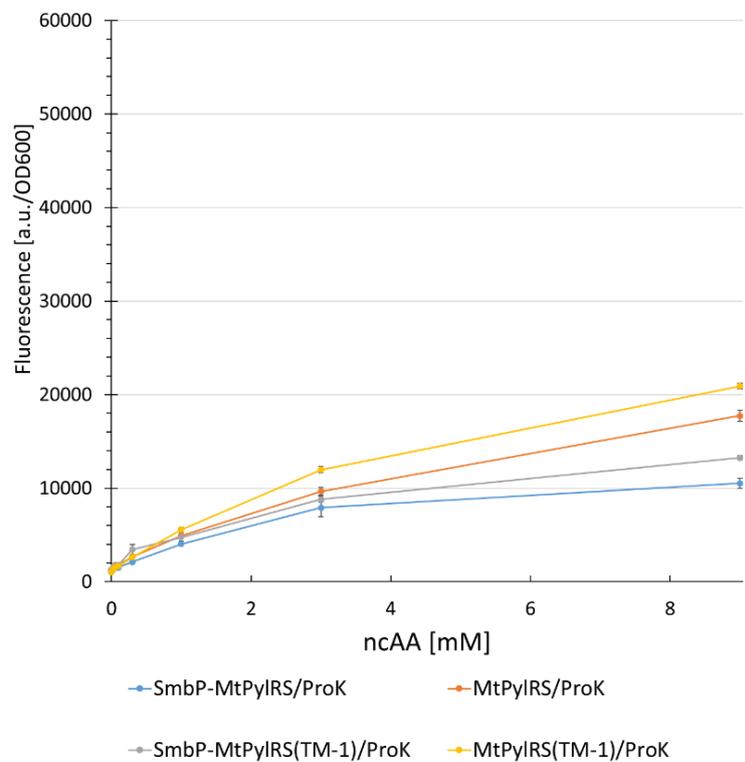


Figure 99: Concentration dependent protein production for four different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

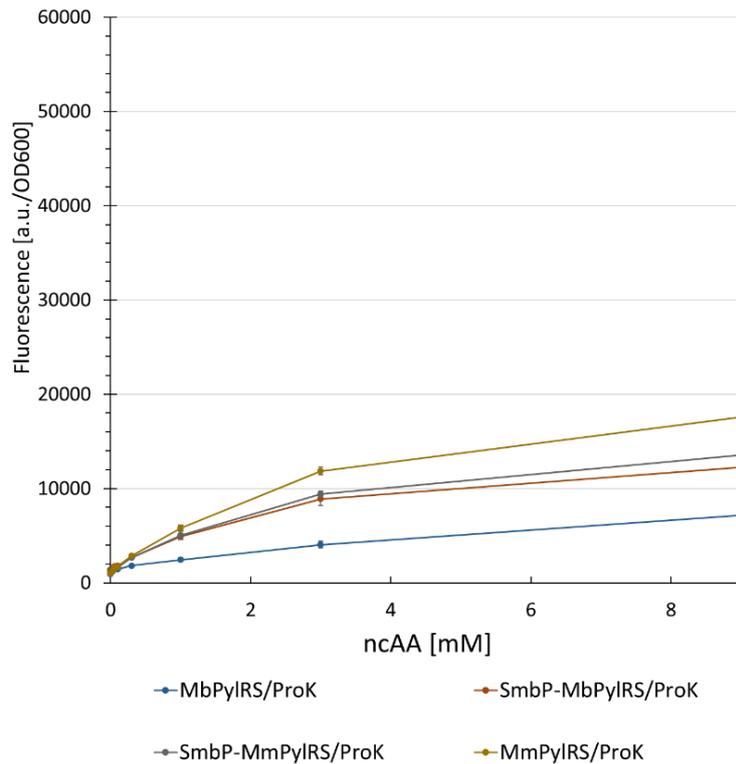


Figure 100: Concentration dependent protein production for four different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

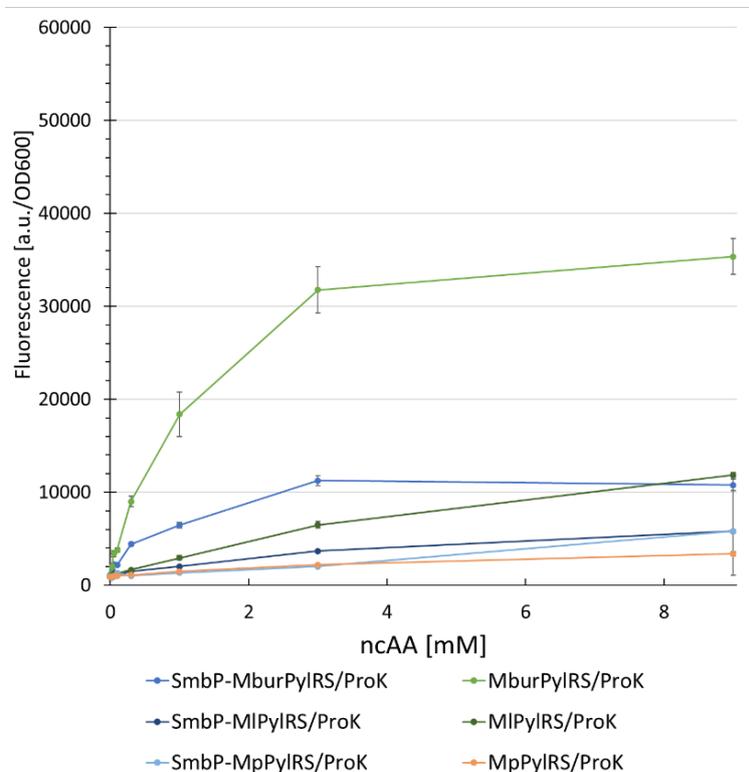


Figure 101: Concentration dependent protein production for six different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

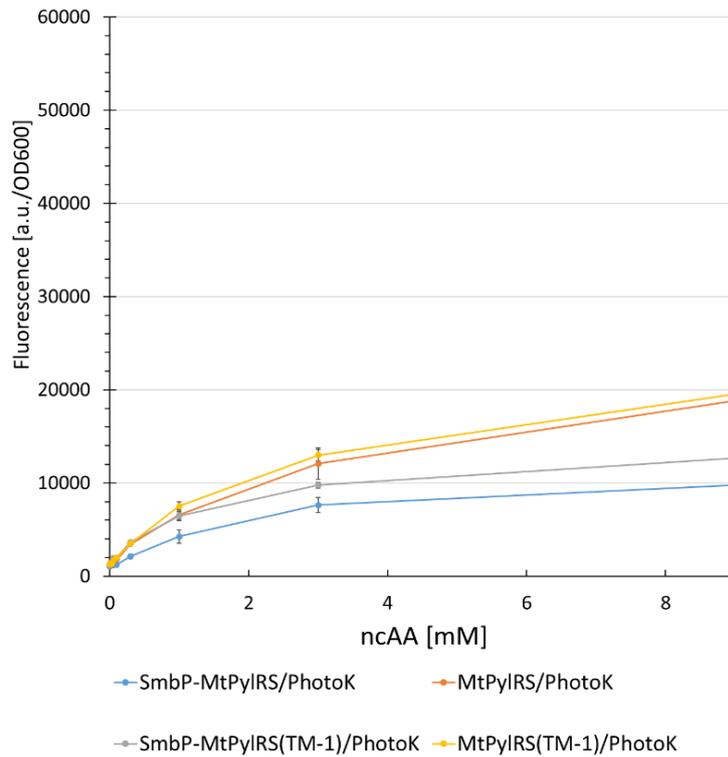


Figure 102: Concentration dependent protein production for four different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

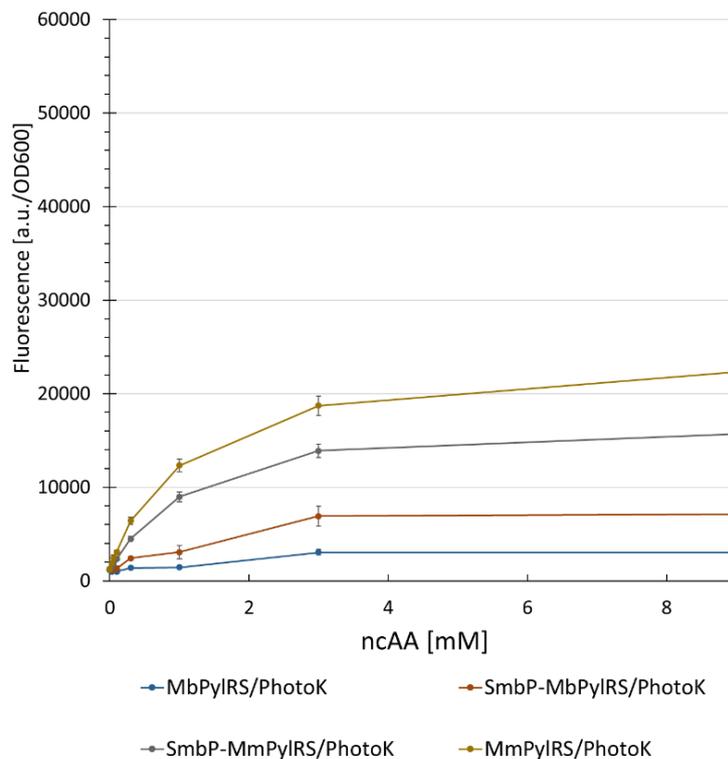


Figure 103: Concentration dependent protein production for four different MbPyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

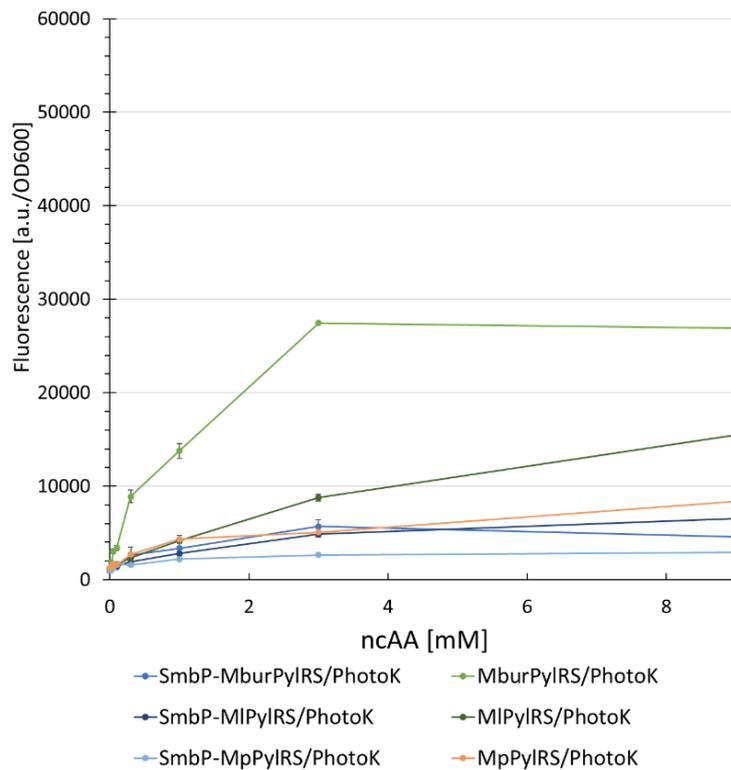


Figure 104: Concentration dependent protein production for six different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

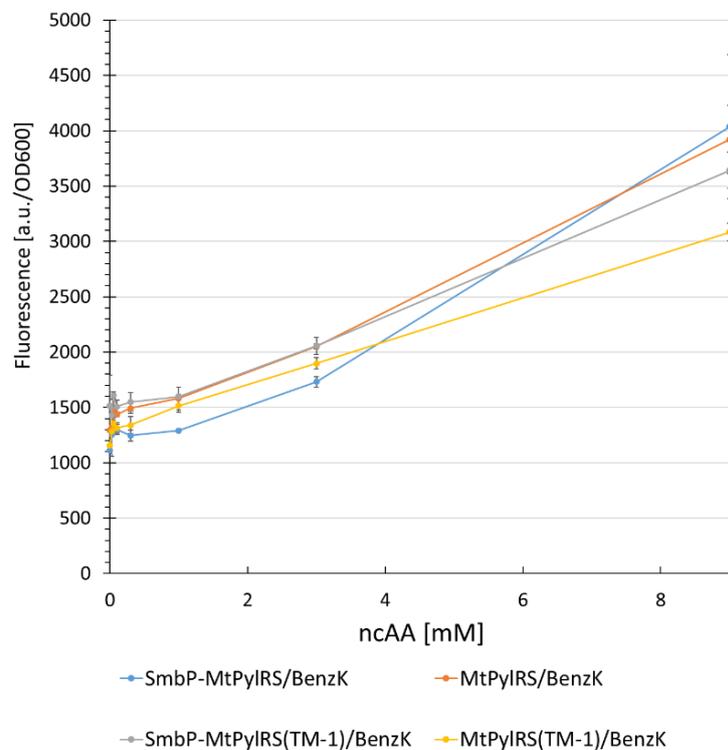


Figure 105: Concentration dependent protein production for four different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

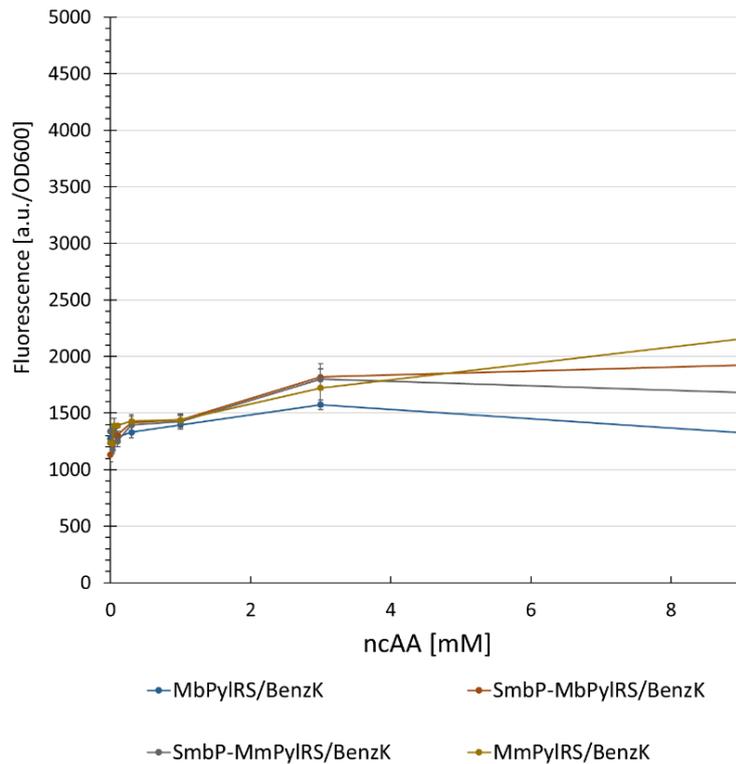


Figure 106: Concentration dependent protein production for four different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

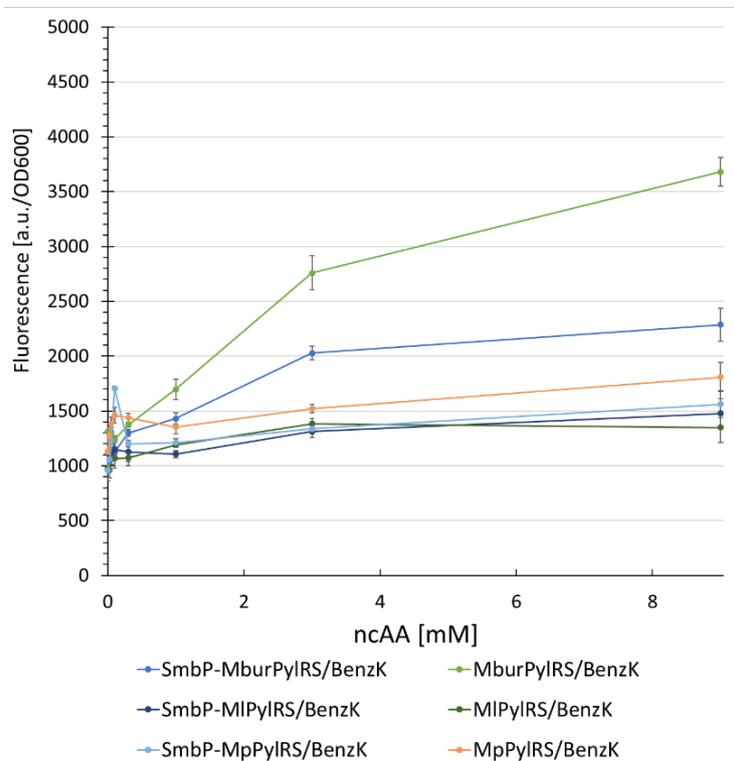


Figure 107: Concentration dependent protein production for six different PyIRS variants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter, endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

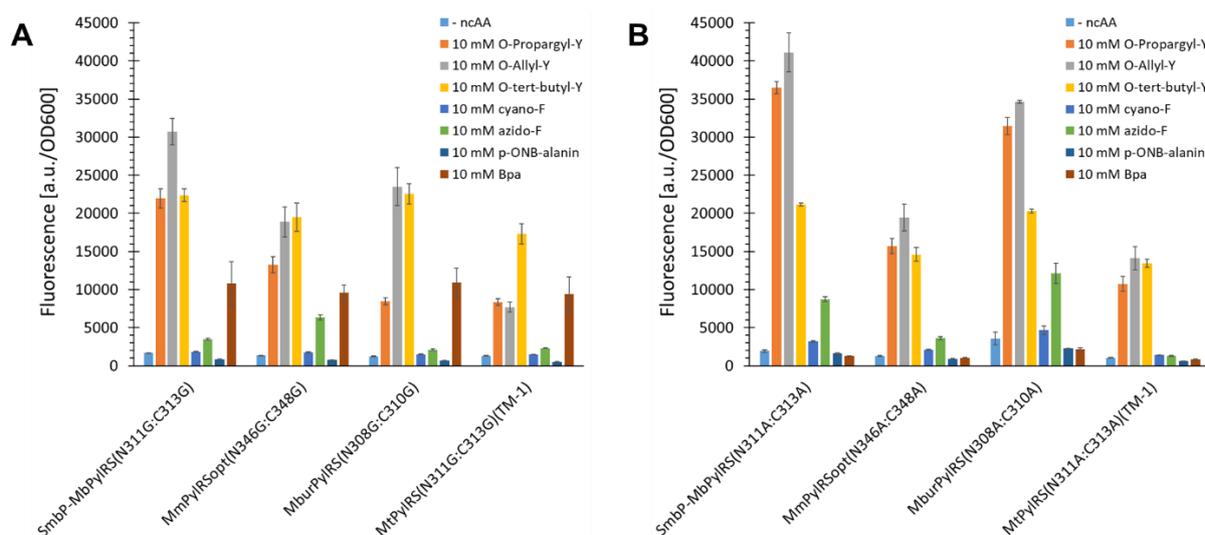


Figure 108: Comparison of ncAA (**39**, **40**, **44**, **45**, **46**, **49**, **50**) incorporation efficiency for PyIRS double Ala and double Gly mutants. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

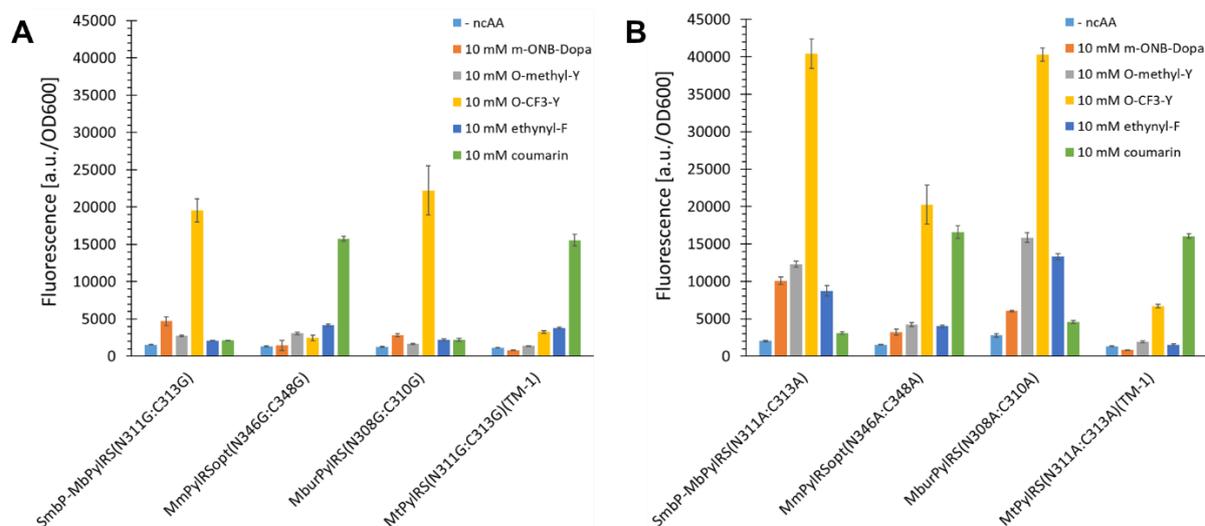


Figure 109: Comparison of ncAA (**38**, **47**, **48**, **51**, **52**) incorporation efficiency for PyIRS double Ala and double Gly mutants. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

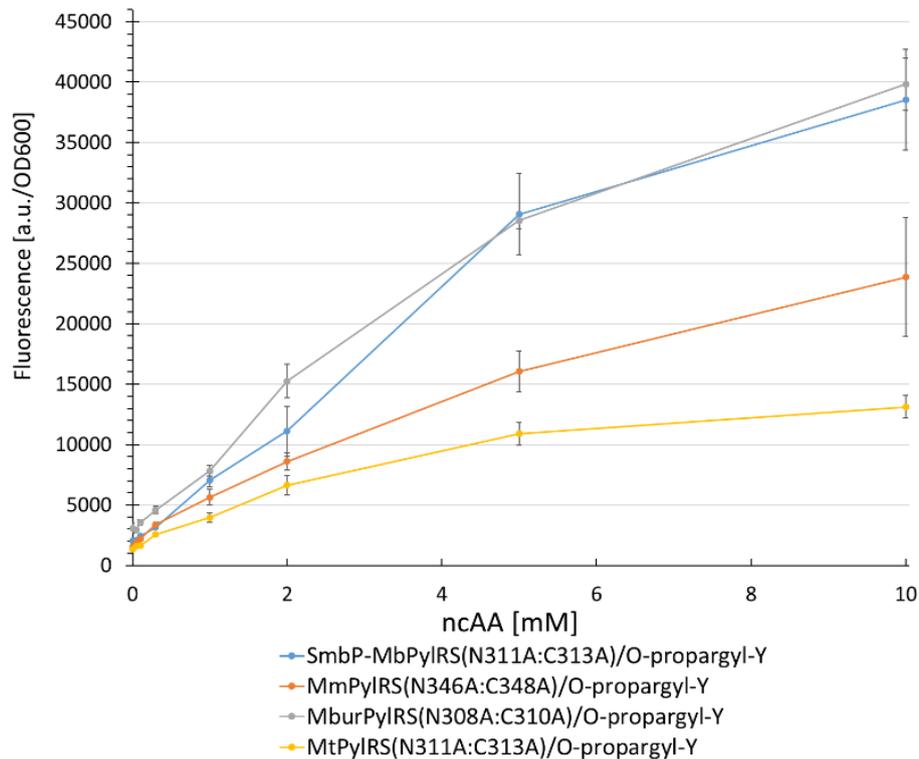


Figure 110: Concentration dependent protein production for four different PyIRS double Ala mutants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter. Endpoint measurements after 24 h with different ncAA concentrations (0.5, 0.1, 0.3, 1, 2, 3, and 10 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

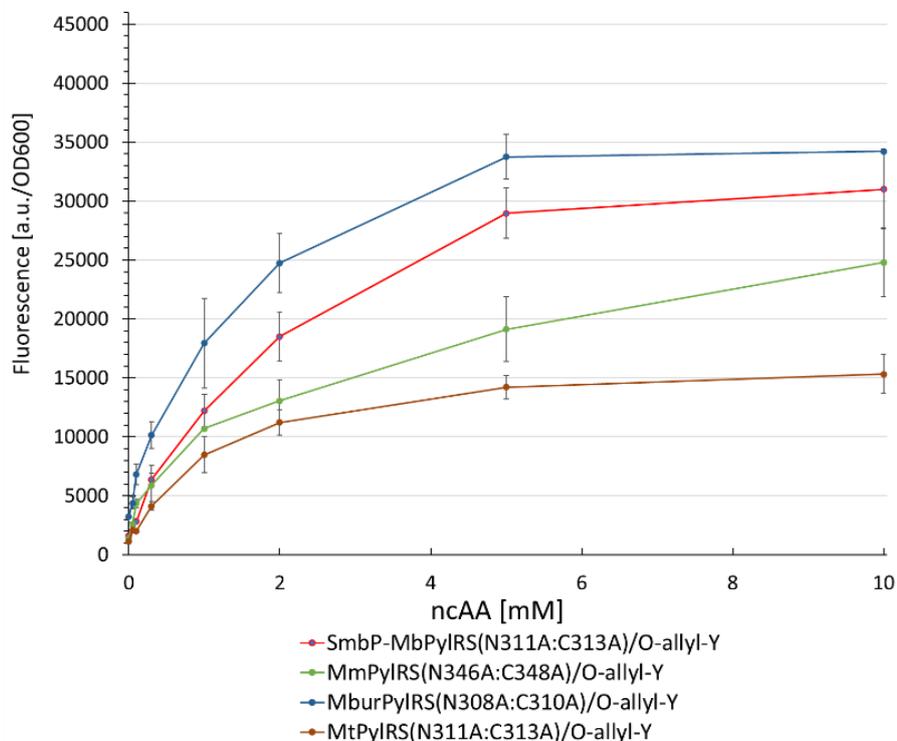


Figure 111: Concentration dependent protein production for four different PyIRS double Ala mutants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter. Endpoint measurements after 24 h with different ncAA concentrations (0.5, 0.1, 0.3, 1, 2, 3, and 10 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

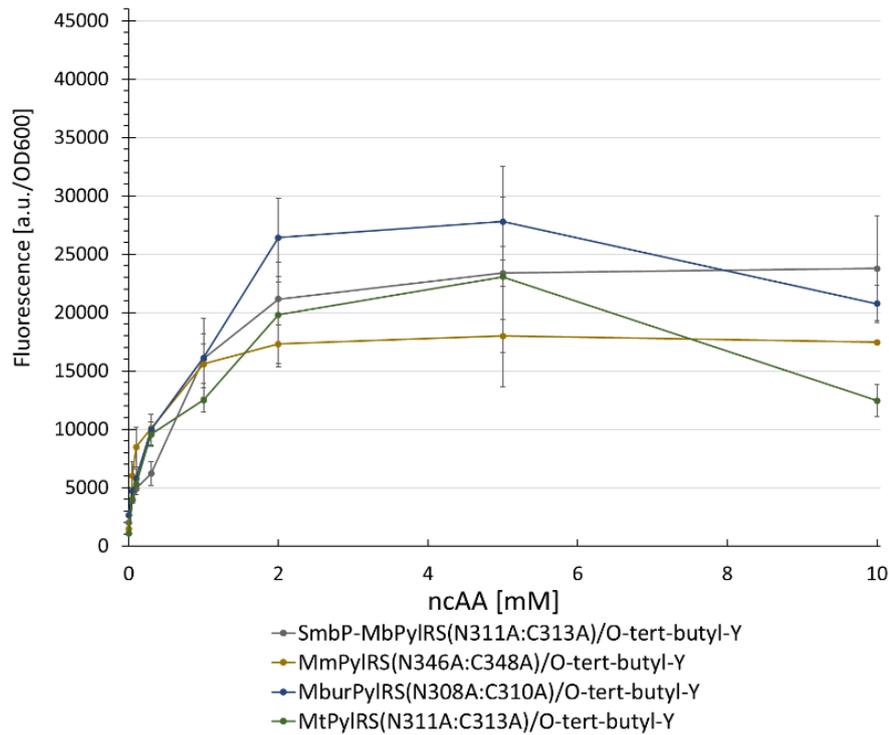


Figure 112: Concentration dependent protein production for four different PyIRS double Ala mutants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter. Endpoint measurements after 24 h with different ncAA concentrations (0.5, 0.1, 0.3, 1, 2, 3, and 10 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

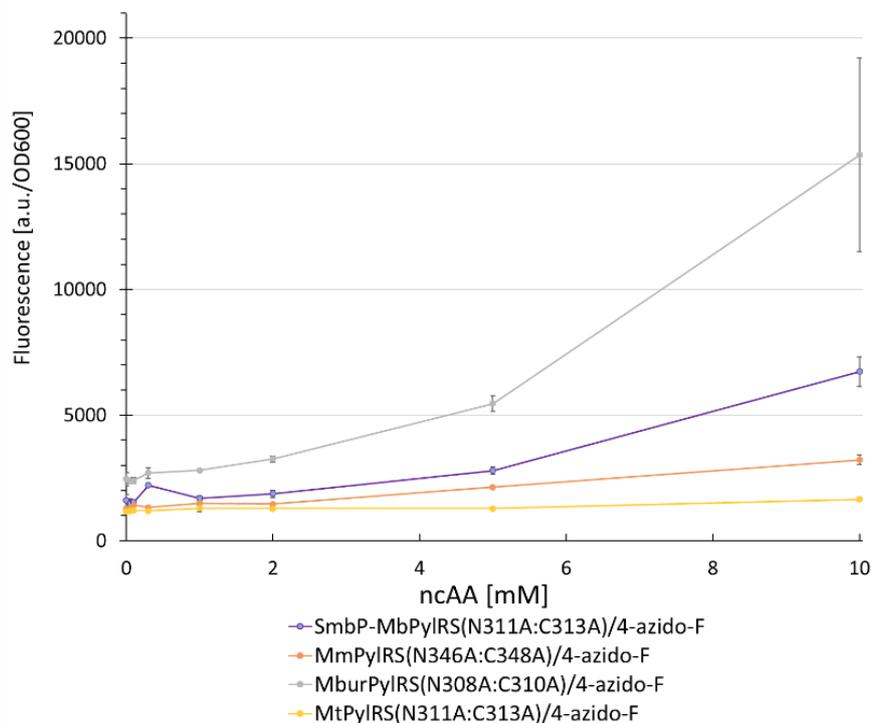


Figure 113: Concentration dependent protein production for four different PyIRS double Ala mutants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter. Endpoint measurements after 24 h with different ncAA concentrations (0.5, 0.1, 0.3, 1, 2, 3, and 10 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

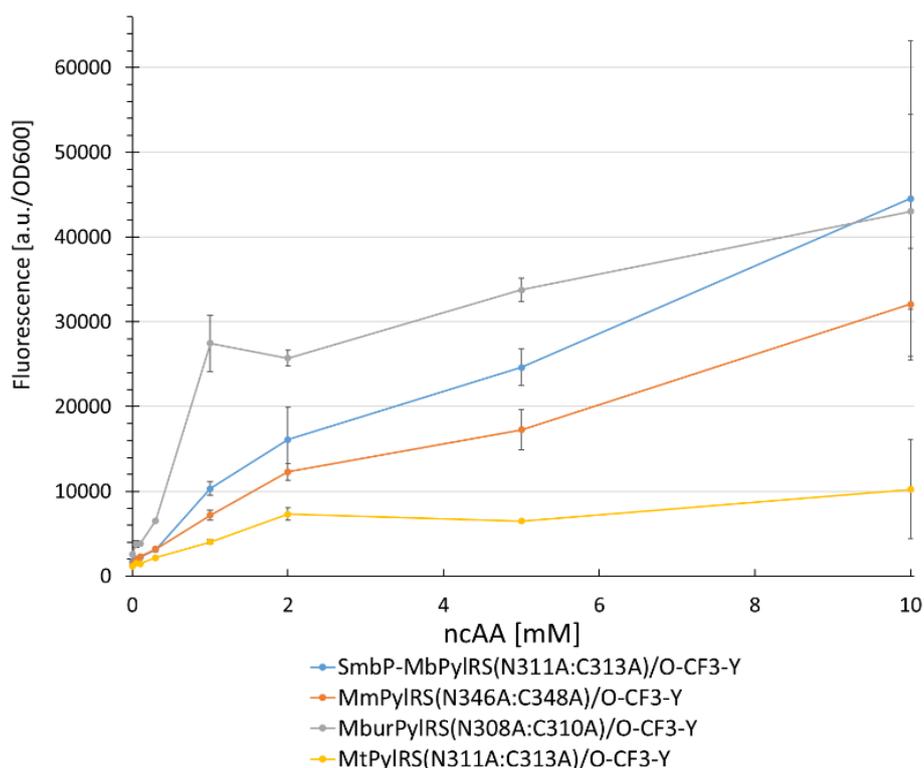


Figure 114: Concentration dependent protein production for four different PyIRS double Ala mutants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter. Endpoint measurements after 24 h with different ncAA concentrations (0.5, 0.1, 0.3, 1, 2, 3, and 10 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

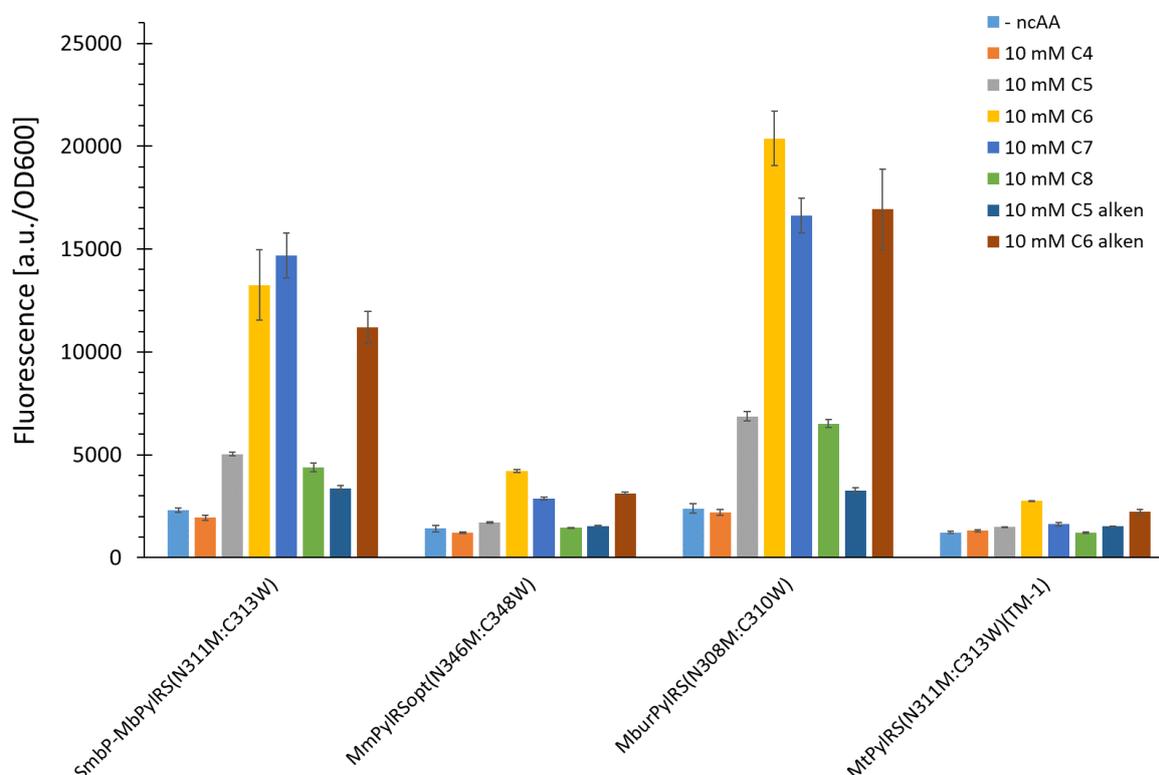


Figure 115: Comparison of ncAA (10, 7, 5, 2, 3, 8, 6) incorporation efficiency for different PyIRS constructs. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

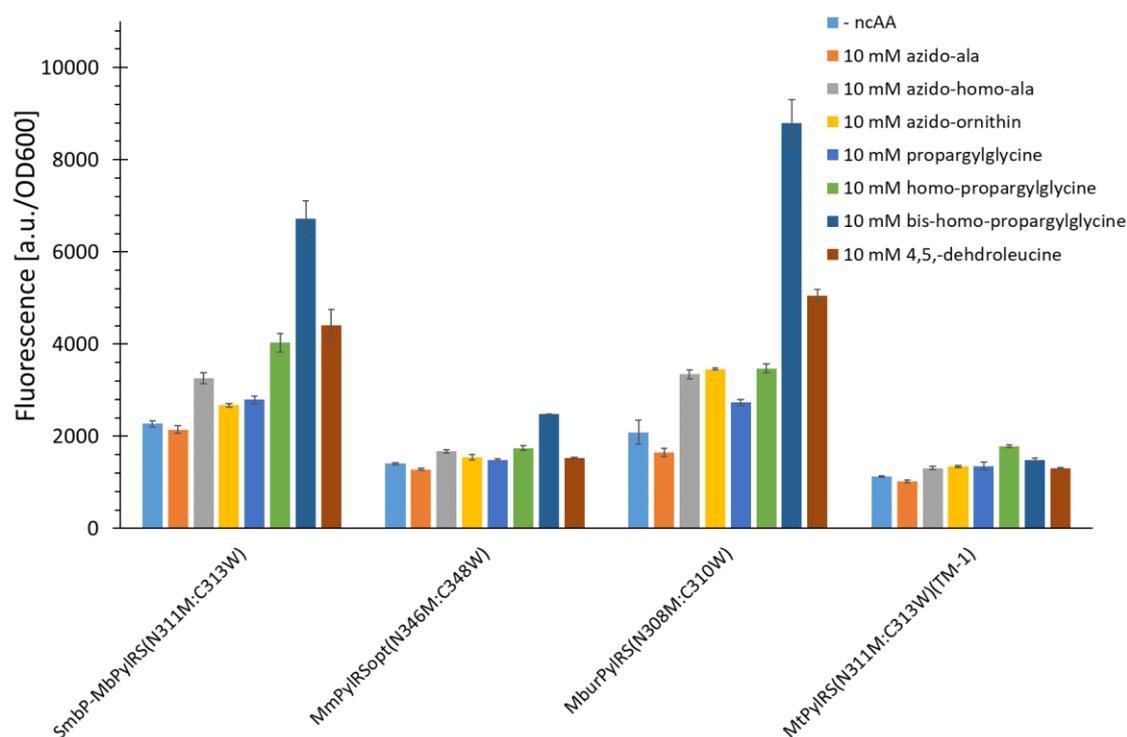


Figure 116: Comparison of ncAA (14, 15, 16, 13, 12, 11, 21) incorporation efficiency for different PyIRS constructs. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

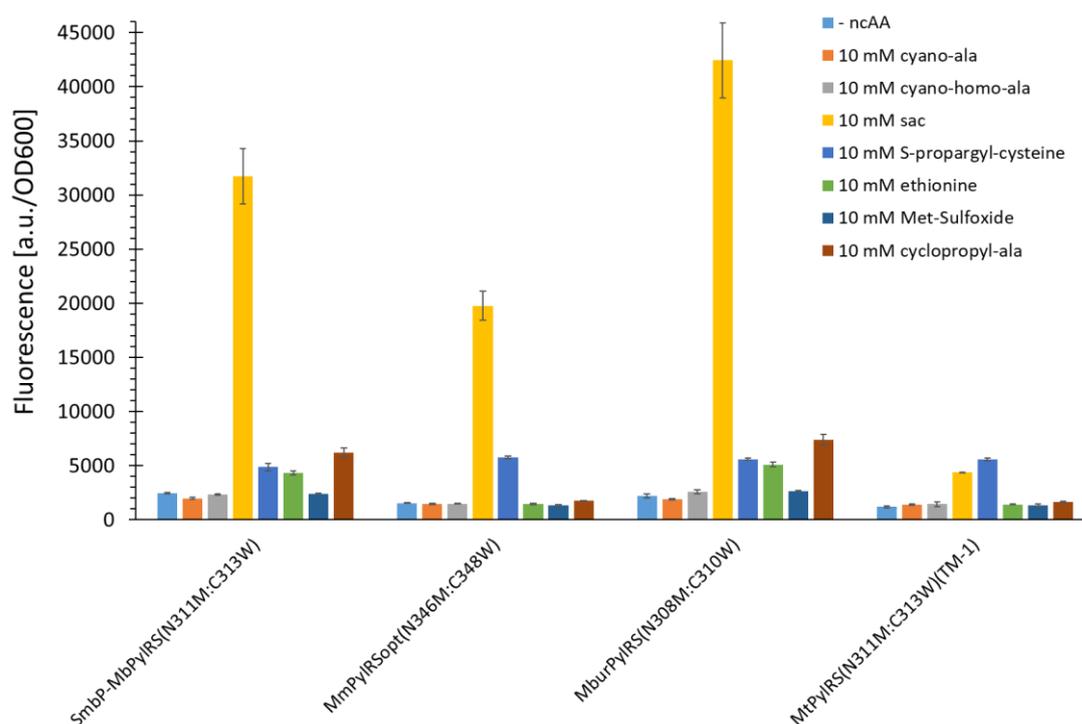


Figure 117: Comparison of ncAA (18, 19, 1, 27, 25, 23, 9) incorporation efficiency for different PyIRS constructs. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

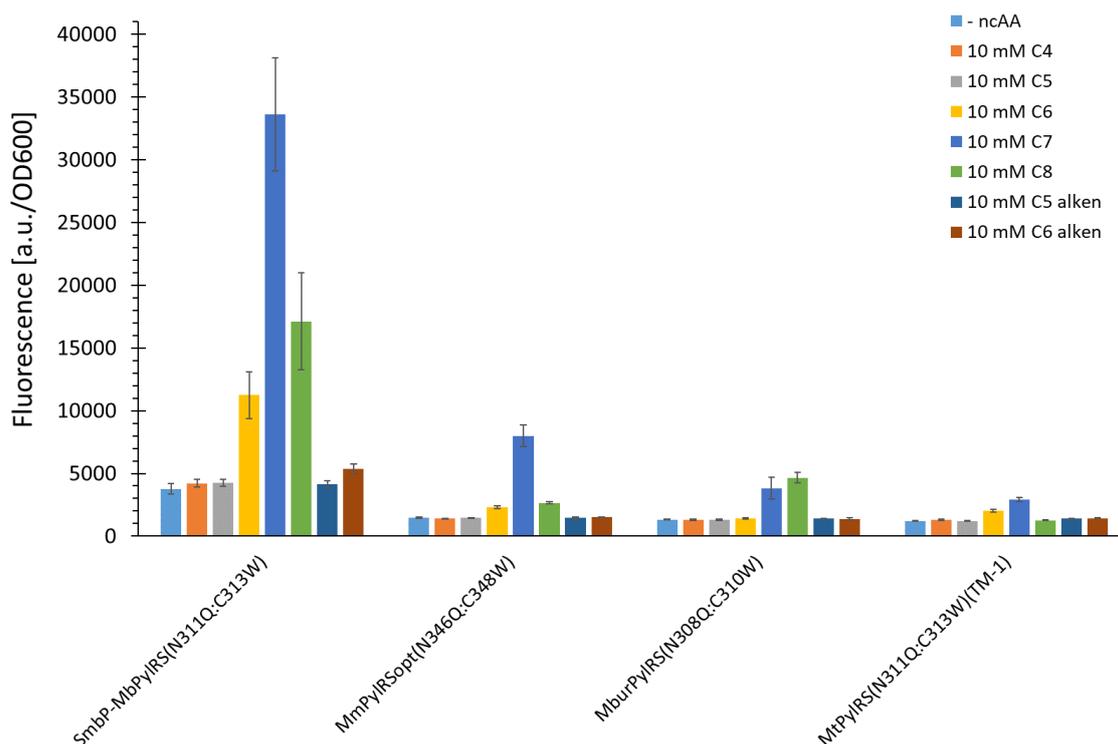


Figure 118: Comparison of ncAA (**10**, **7**, **5**, **2**, **3**, **8**, **6**) incorporation efficiency for different PyIRS constructs. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

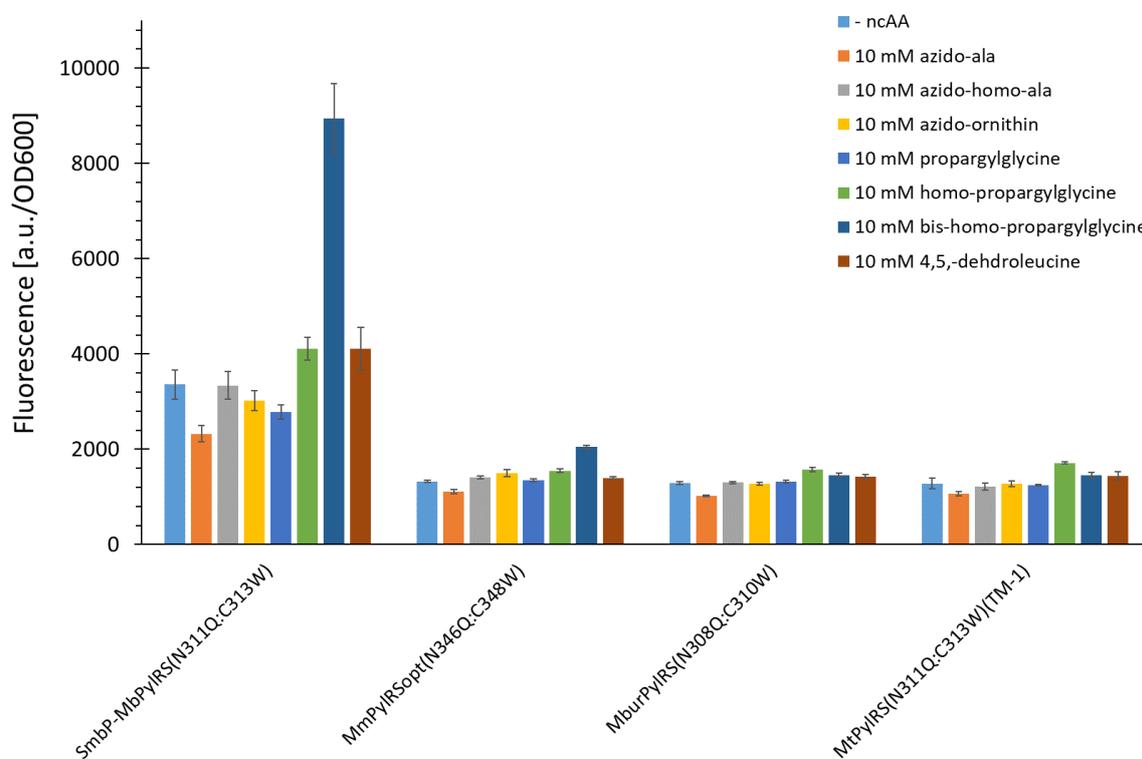


Figure 119: Comparison of ncAA (**14**, **15**, **16**, **13**, **12**, **11**, **21**) incorporation efficiency for different PyIRS constructs. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

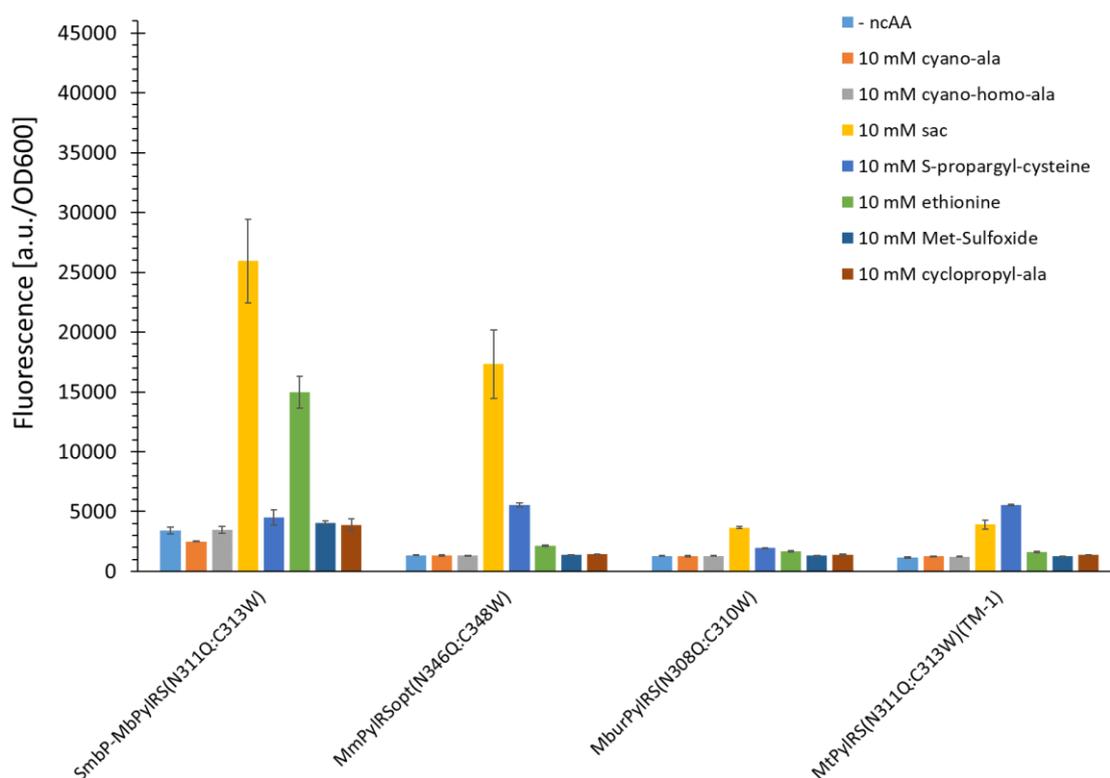


Figure 120: Comparison of ncAA (**18**, **19**, **1**, **27**, **25**, **23**, **9**) incorporation efficiency for different PyIRS constructs. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

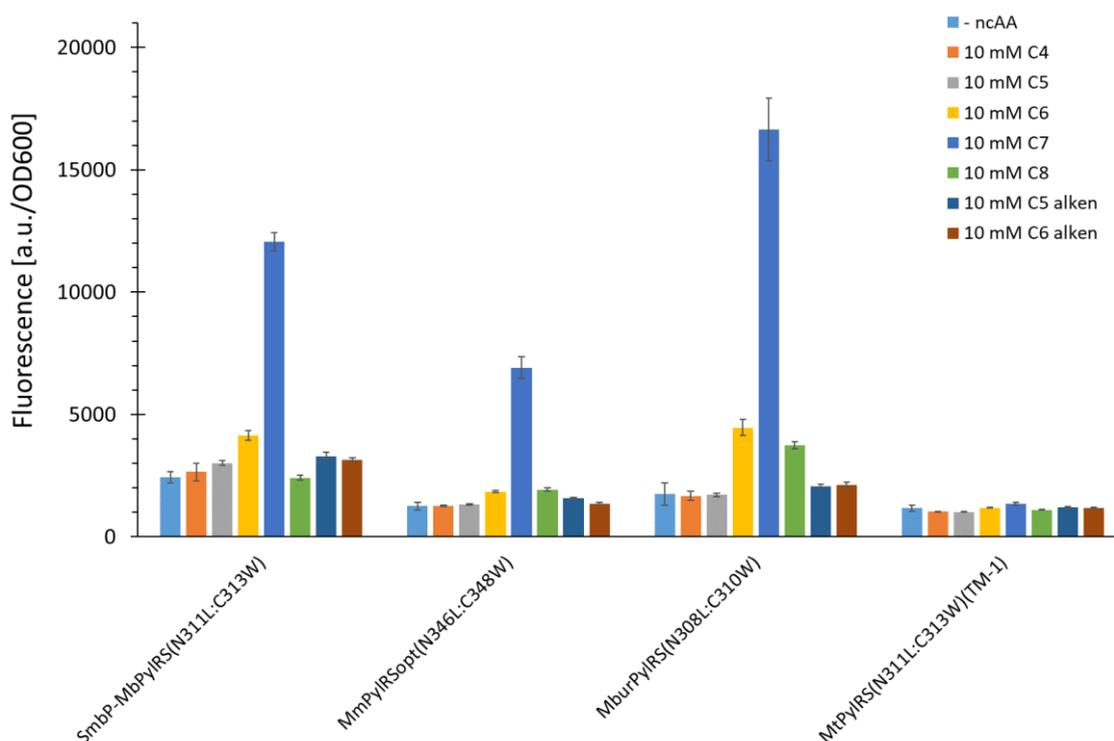


Figure 121: Comparison of ncAA (**10**, **7**, **5**, **2**, **3**, **8**, **6**) incorporation efficiency for different PyIRS constructs. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

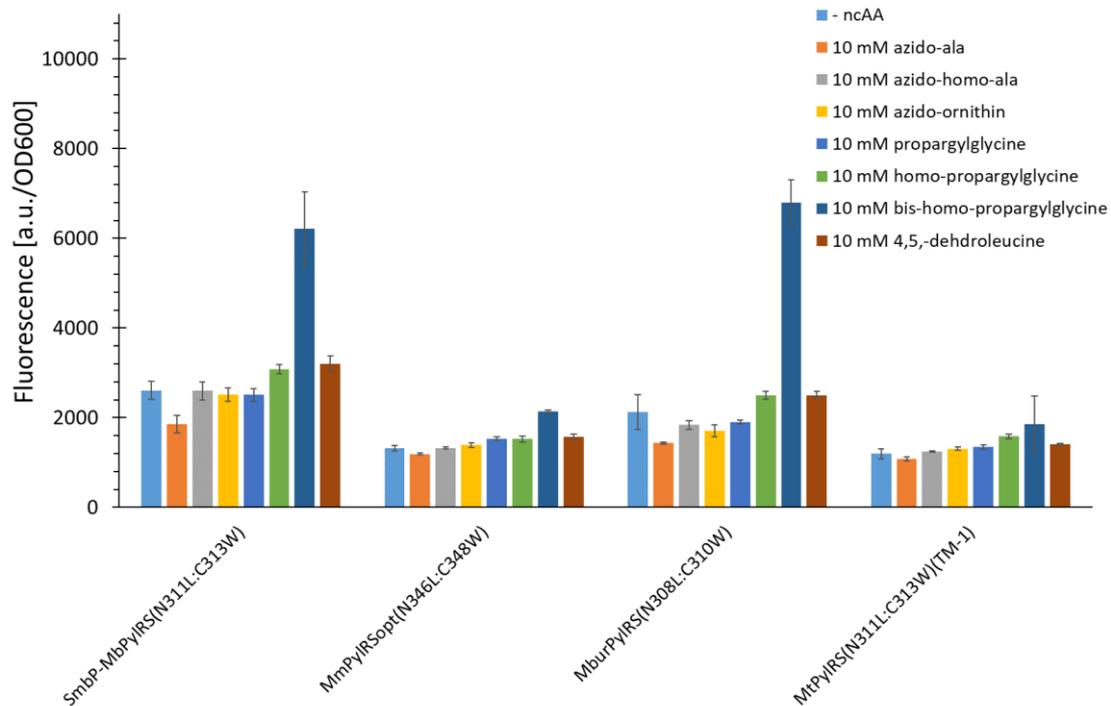


Figure 122: Comparison of ncAA (14, 15, 16, 13, 12, 11, 21) incorporation efficiency for different PyIRS constructs. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

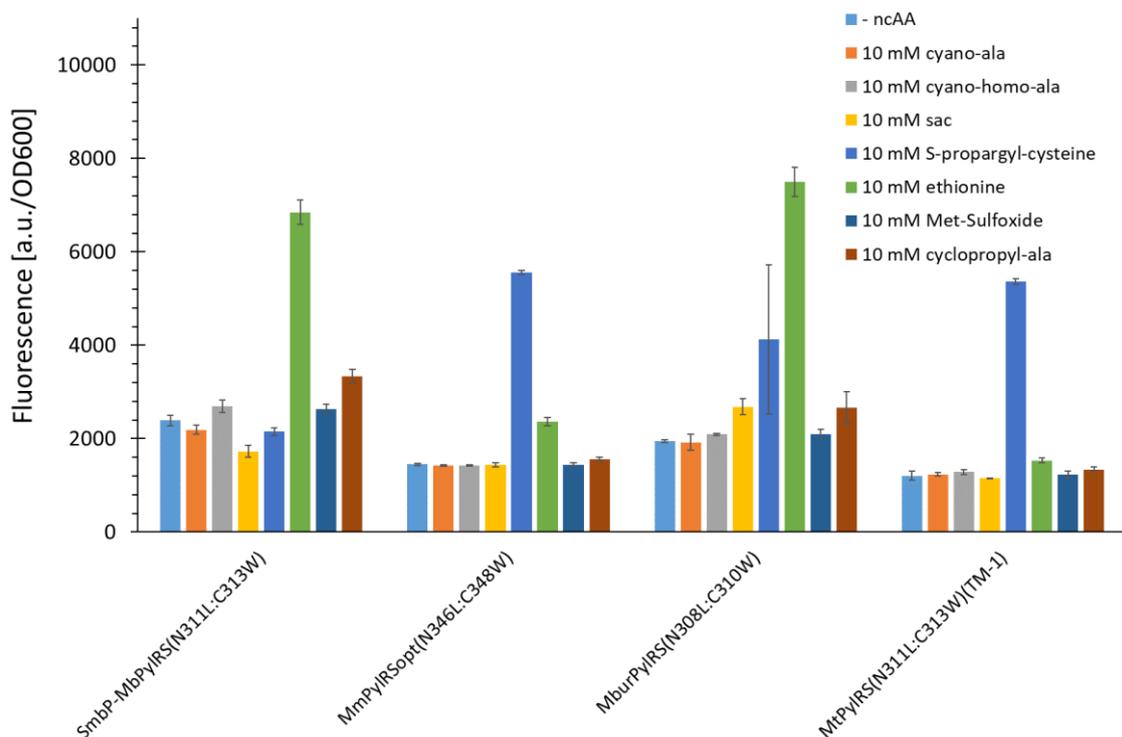


Figure 123: Comparison of ncAA (18, 19, 1, 27, 25, 23, 9) incorporation efficiency for different PyIRS constructs. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

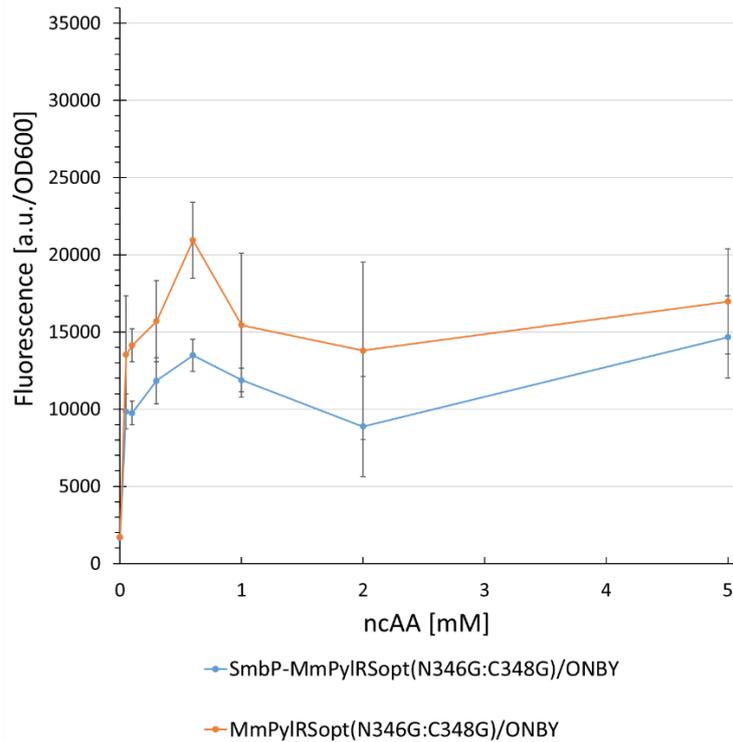


Figure 124: Concentration dependent protein production for two different PyIRS double Gly mutants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter. Endpoint measurements after 24 h with different ncAA concentrations (0.05, 0.1, 0.3, 0.6, 1, 2, and 5 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

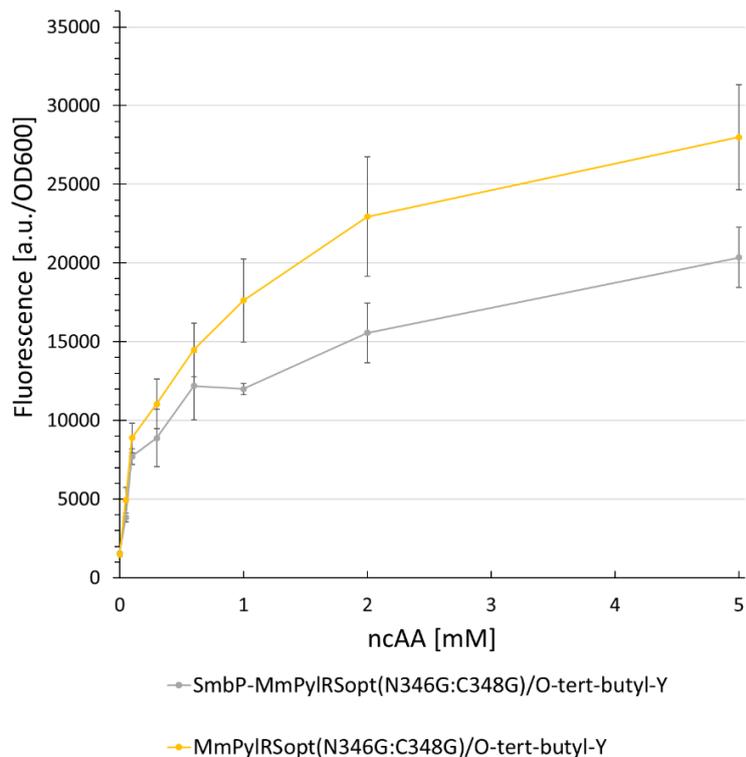


Figure 125: Concentration dependent protein production for two different PyIRS double Gly mutants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter. Endpoint measurements after 24 h with different ncAA concentrations (0.05, 0.1, 0.3, 0.6, 1, 2, and 5 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

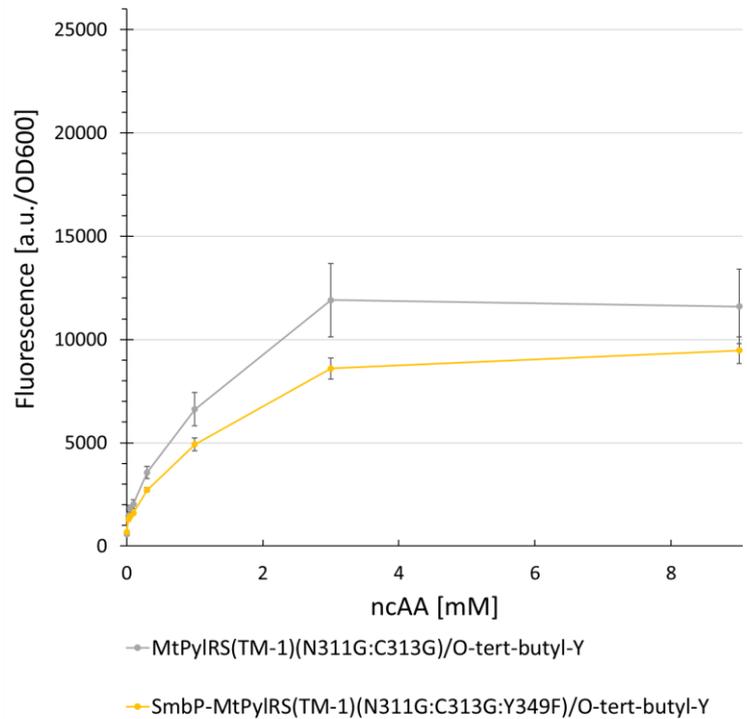


Figure 126: Concentration dependent protein production for two different PyIRS double Gly mutants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter. Endpoint measurements after 24 h with different ncAA concentrations (0.05, 0.1, 0.3, 0.6, 1, 2, and 5 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

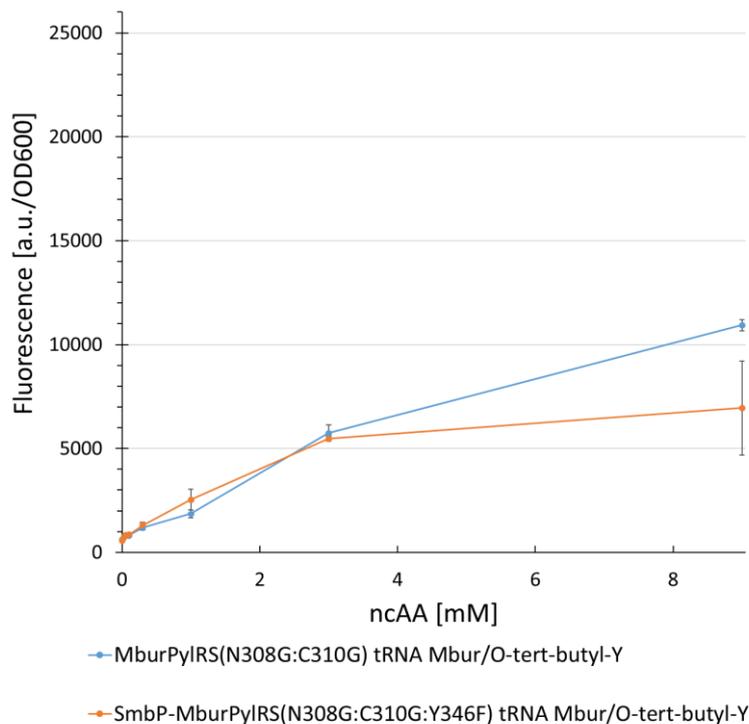


Figure 127: Concentration dependent protein production for two different PyIRS double Gly mutants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter. Endpoint measurements after 24 h with different ncAA concentrations (0.05, 0.1, 0.3, 0.6, 1, 2, and 5 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

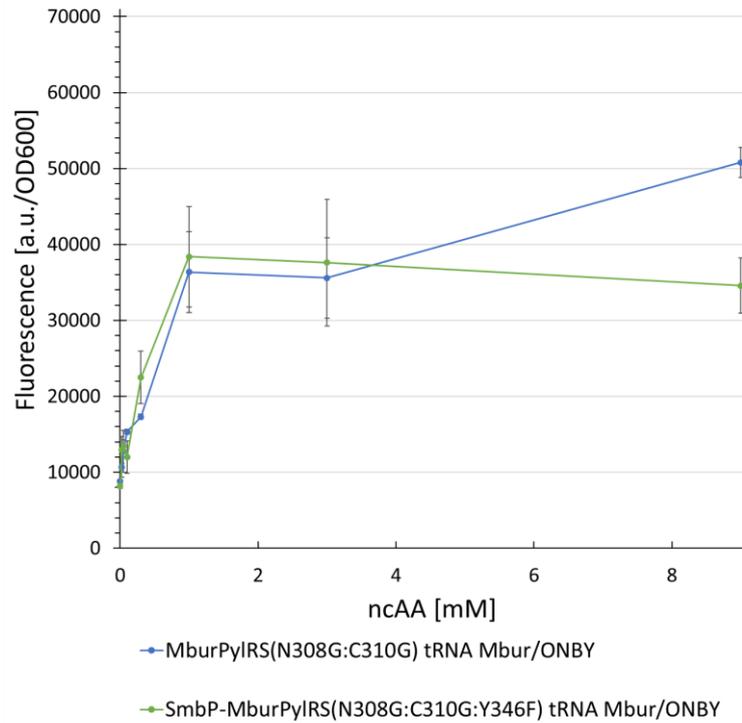


Figure 128: Concentration dependent protein production for two different PyIRS double Gly mutants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter. Endpoint measurements after 24 h with different ncAA concentrations (0.05, 0.1, 0.3, 0.6, 1, 2, and 5 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

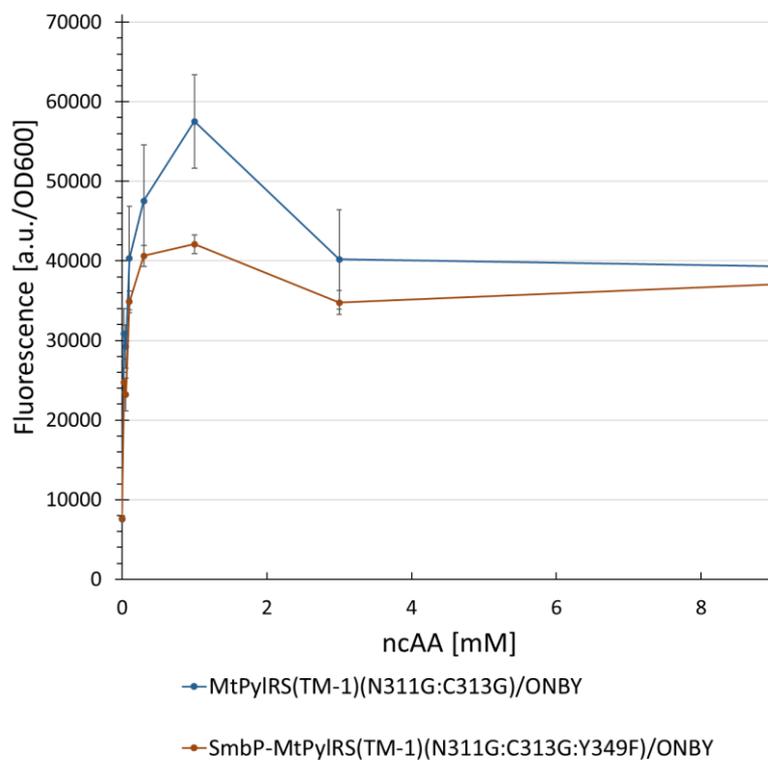


Figure 129: Concentration dependent protein production for two different PyIRS double Gly mutants. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter. Endpoint measurements after 24 h with different ncAA concentrations (0.05, 0.1, 0.3, 0.6, 1, 2, and 5 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

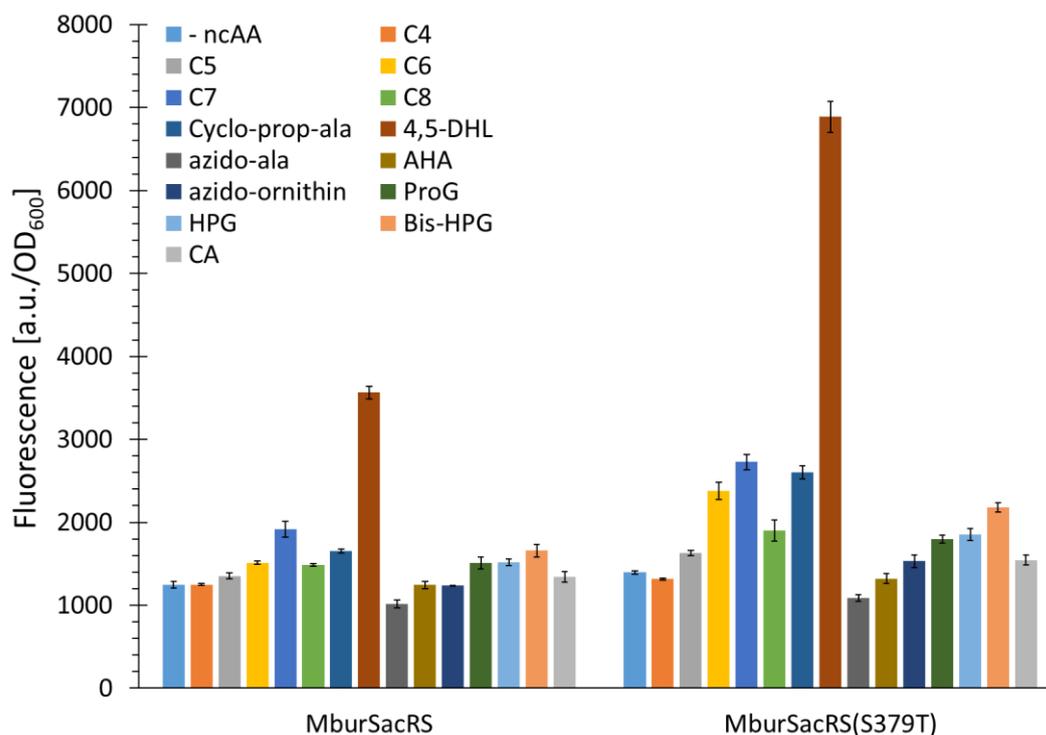


Figure 130: Comparison of ncAA incorporation efficiency for different *MburPylRS* constructs. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

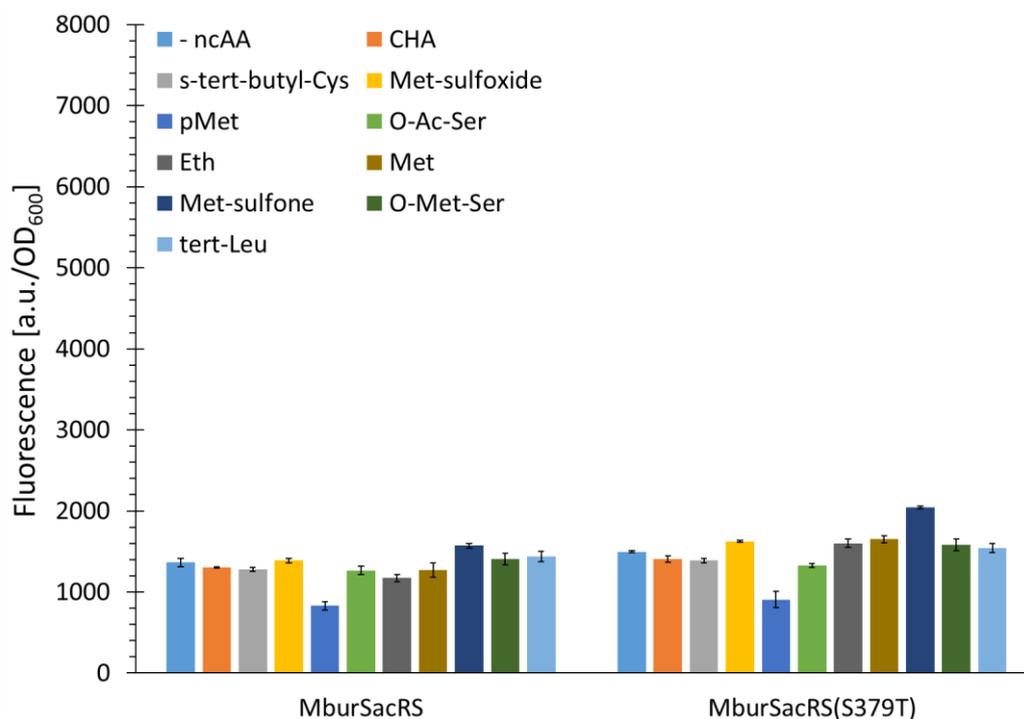


Figure 131: Comparison of ncAA incorporation efficiency for different *MburPylRS* constructs. Fluorescence measurement of intact *E. coli* BL21(DE3) cells producing the SUMO-sfGFP(R2amber) reporter protein. Endpoint measurements after 24 h of incubation. The data (incl. standard deviation) represent the mean of three biological replicates. 10 mM ncAAs supplied

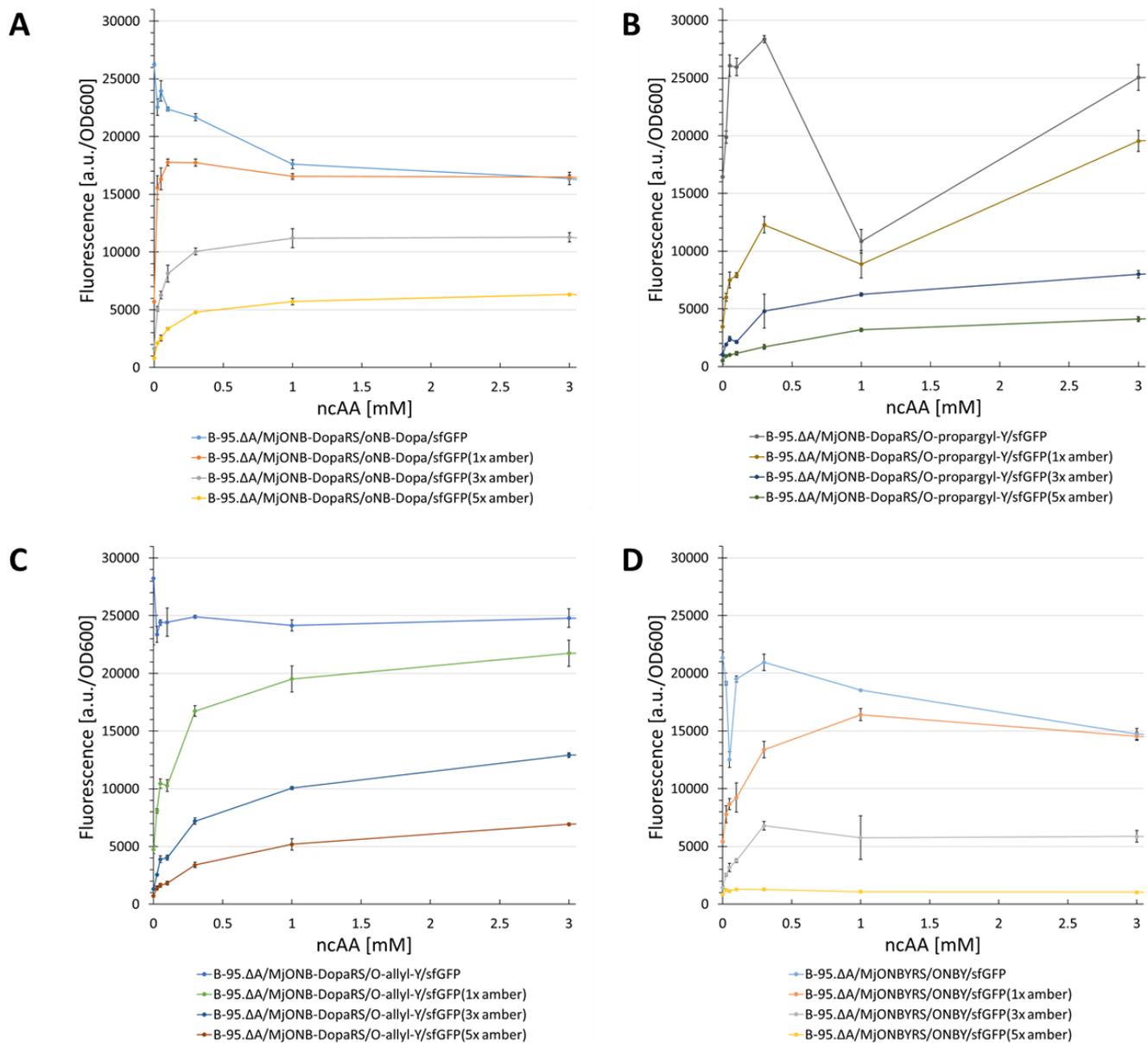


Figure 132: Concentration dependent protein production for four different *MjTyrRS*/ncAA combinations. Measured fluorescence intensity of intact B-95.Δ cells expressing the SUMO-sfGFP reporter containing indicated number of stop codons. Endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 (not shown) mM). The data (incl. standard deviation) represent the mean of three biological replicates.

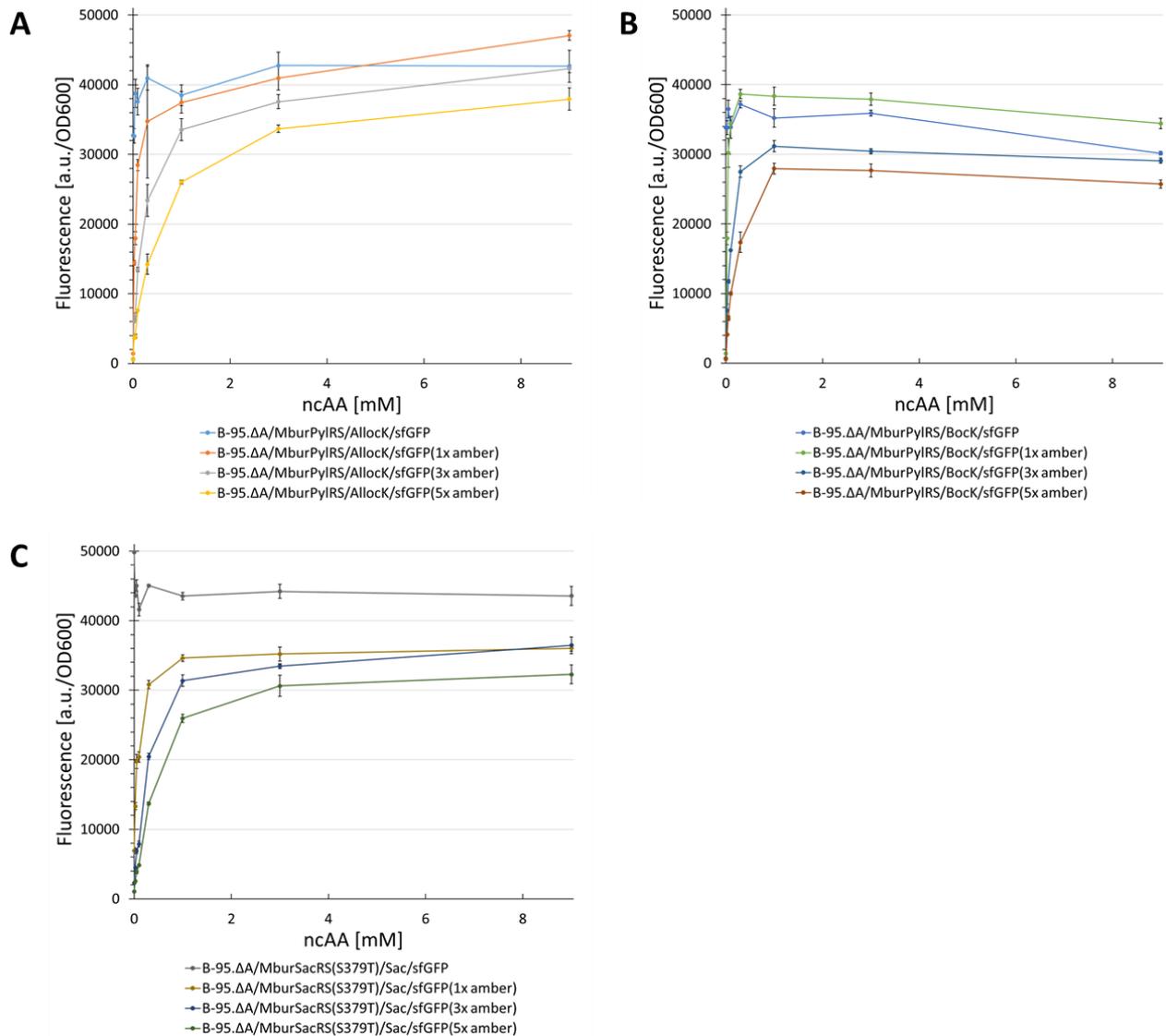


Figure 133: Concentration dependent protein production for three different *MburPylRSRS*/ncAA combinations. Measured fluorescence intensity of intact B-95.Δ cells expressing the SUMO-sfGFP reporter containing indicated number of stop codons. Endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 3, and 9 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

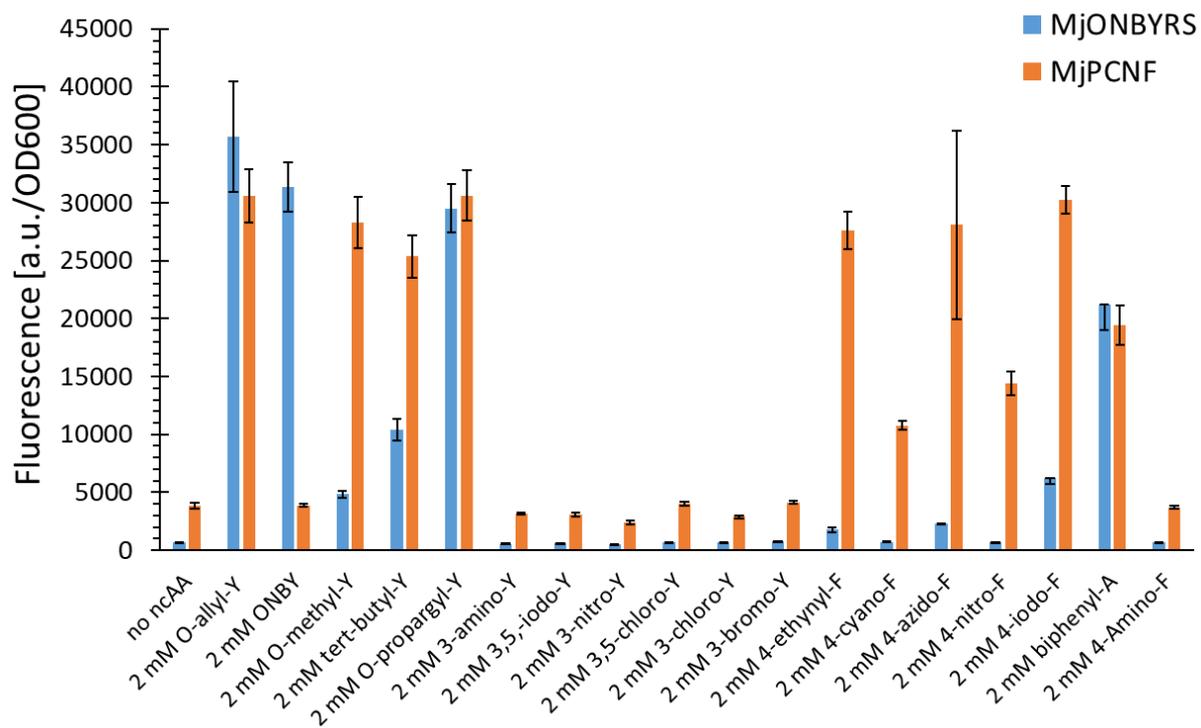


Figure 134: Prescreening of several ncAAs with MjONBYRS and MjPCNF. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter. Endpoint measurements after 24 h with 2 mM ncAAs supplied. The data (incl. standard deviation) represent the mean of three biological replicates.

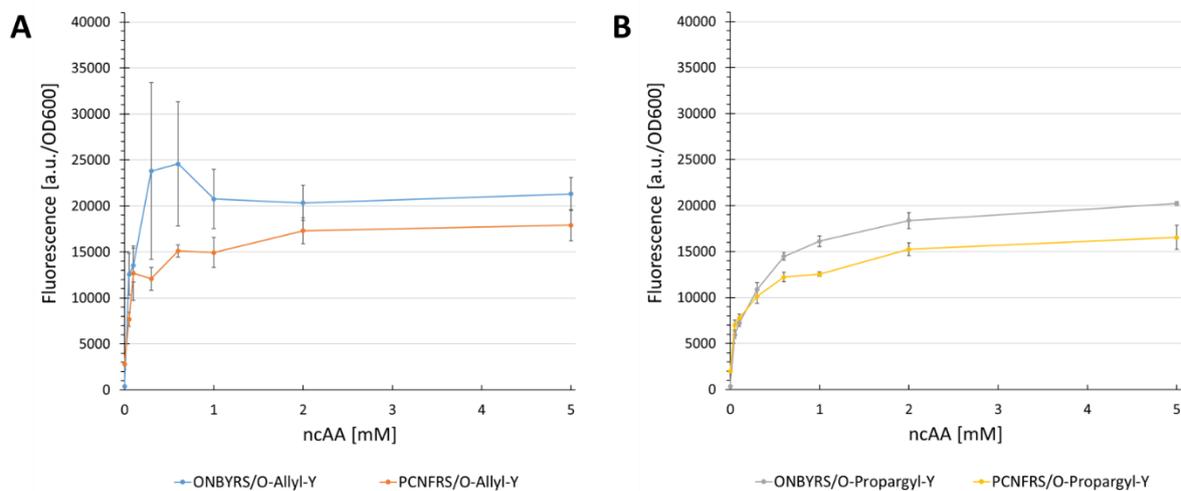


Figure 135: Concentration dependent protein production for two different MjTyrRS/ncAA combinations. Measured fluorescence intensity of intact *E. coli* BL21(DE3) cells expressing the SUMO-sfGFP(R2 amber) reporter. Endpoint measurements after 24 h with different ncAA concentrations (0.025, 0.05, 0.1, 0.3, 1, 2, and 5 mM). The data (incl. standard deviation) represent the mean of three biological replicates.

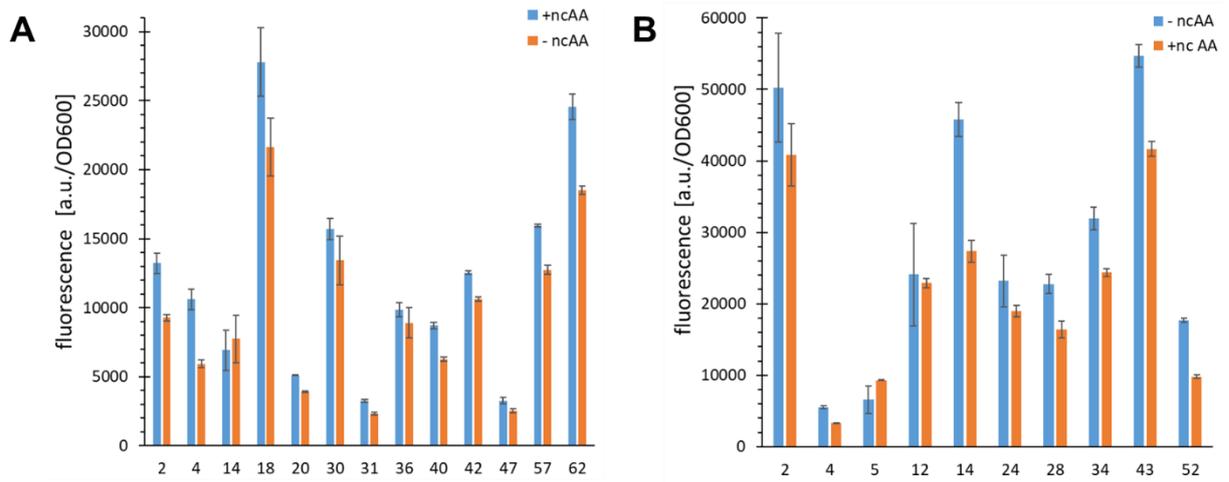


Figure 136: Screening of the Azu-alanine library. **A**) *p*-oNB-alanine (**54**) **B**) or *m*-oNB-alanine (**55**)

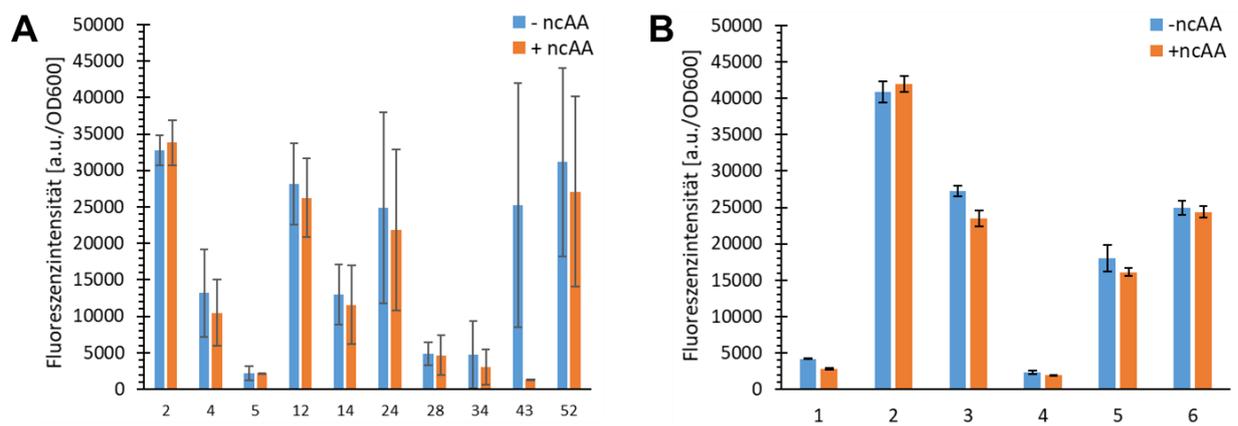


Figure 137: Screening of the Dopa library. **A**) *m*-oNB-Dopa (**51**). **B**) or *p*-oNB-Dopa (**53**)

6.1.2. SDS-PAGE

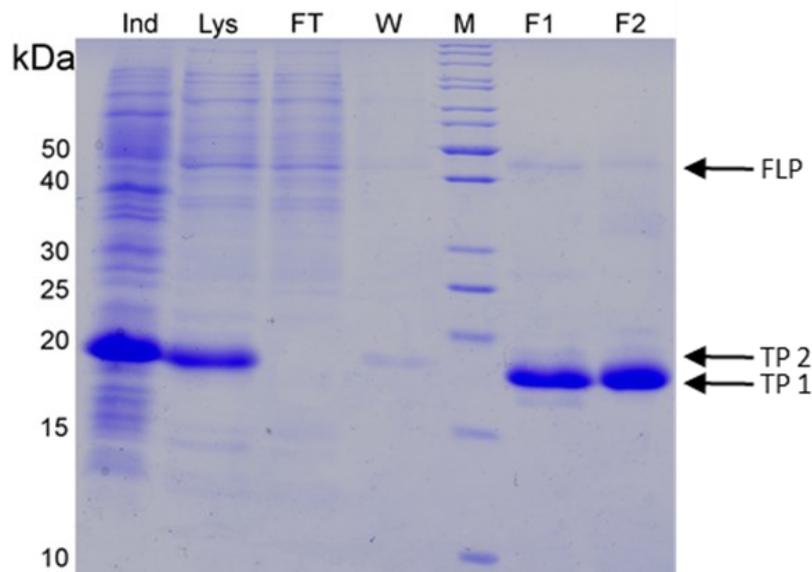


Figure 138: SDS-PAGE analysis of purified His₆-SUMO-sfGFP(R2Sac:N39Sac:K101Sac) reporter expressed in *E. coli* BL21(DE3) with SmbP-MbSacRS co-expression. Sample abbreviations: Whole cell extract of induced culture (Ind), soluble cellular lysate (Lys), liquid chromatography flow through (FT), column wash of bound protein (W), different fractions of eluate (F1-2), protein ladder (M), full length product (FLP), truncation product 1 (TP1), truncation product 2 (TP2).

6.1.3. Agarose Gels

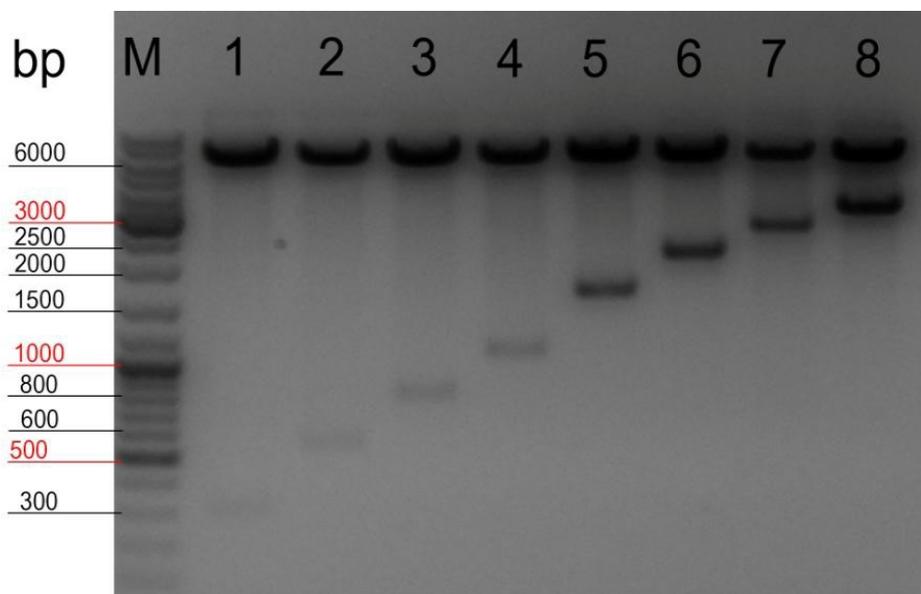


Figure 139: Example of an analytical agarose gel (1%) stained with EtBr. Digestion of pET28a_ELP(5-60x amber-sfGFP) reporter constructs with XbaI and BglI. Marker (M); backbone theoretical size = 5943; **1**) ELP(5x amber) theoretical size = 302 bp; **2**) ELP(10x amber) theoretical size = 542 bp; **3**) ELP(15x amber) theoretical size = 782 bp; **4**) ELP(20x amber) theoretical size = 1022 bp; **5**) ELP(30x amber) theoretical size = 1502 bp; **6**) ELP(40x amber) theoretical size = 1982 bp; **7**) ELP(50x amber) theoretical size = 2462 bp; **8**) ELP(60x amber) theoretical size = 2942 bp.

6.1.4. Miscellaneous



Figure 140: Harvested *E. coli* BL21(DE3) cells after shake flask cultivation for production of Sac-modified sfGFP reporter protein. Untagged (left) vs. tagged (right) *MbSacRS* OTS. Production of ncAA-modified target protein was conducted as described in the main text.

SmbP-tag	+	+	-	-
2 mM Sac	+	-	+	-



Figure 141: OD₆₀₀-normalized, pelleted *E. coli* BL21(DE3) cells from amilCP target protein production with site-specific ncAA installation. Co-expression of the SmbP-tagged aaRS leads to visibly higher production of the blue chromoprotein.

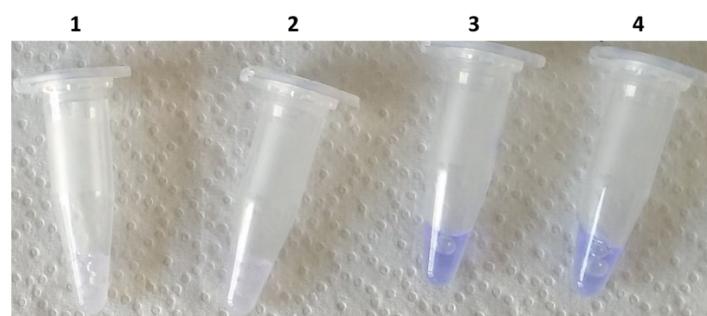


Figure 142: Photo-induced chromophore maturation of amilCP(Y63ONBY). **1)** Target protein produced with NK53, no UV irradiation; **2)** Target protein produced with B-95.ΔA, no UV irradiation; **3)** Target protein produced with NK53, with UV irradiation; **4)** Target protein produced with B-95.ΔA, with UV irradiation. All samples contain the target protein at 1 mg/L concentration.

6.1.5. ESI-MS Spectra

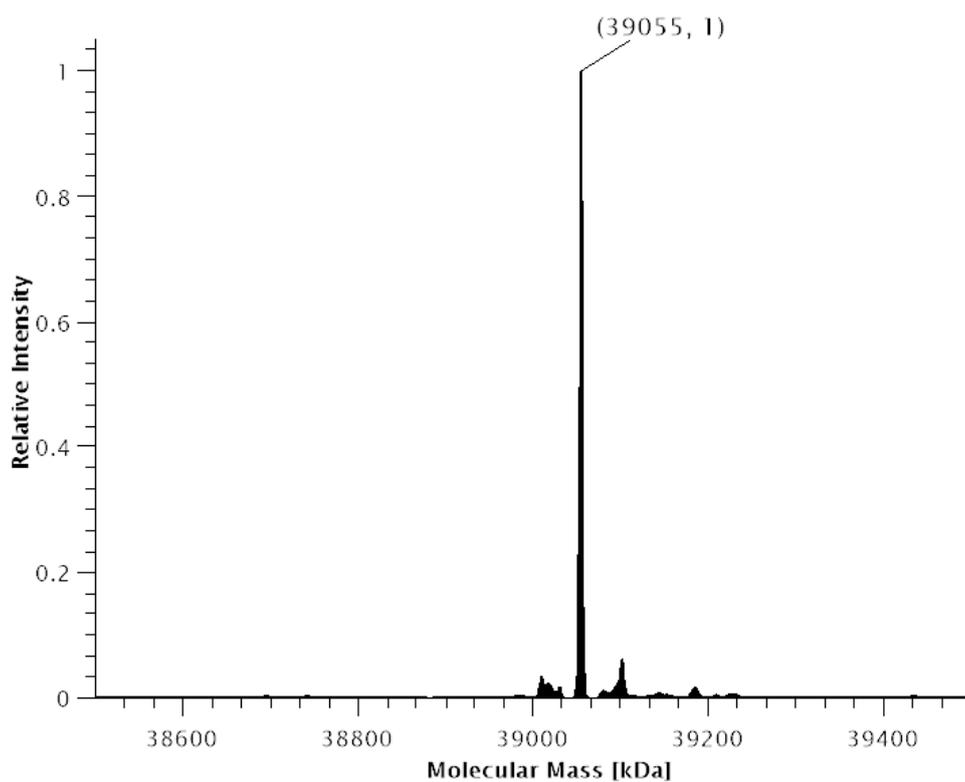


Figure 143: : Deconvoluted ESI-MS spectra of SUMO-sfGFP(R2Sac:N39Sac:K101Sac)-His₆ produced in *E. coli* BL21(DE3) with co-expression of SmbP-MbSacRS. Expected protein mass: 39055.2 Da. Observed mass: 39055 Da.

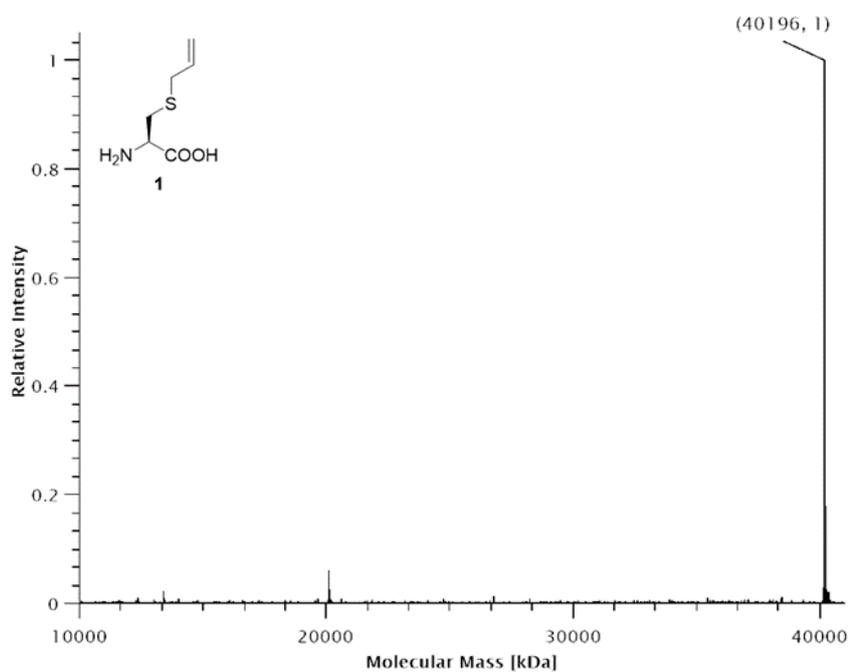


Figure 144: Deconvoluted ESI-MS spectrum of His₆-SUMO-sfGFP(R2(1))-Strep production in *E. coli* BL21(DE3) with co-expression of SmbP-MbSacRS(N311M:C313W:V366A). Expected protein mass: 40194.9 Da. Observed mass: 40196 Da.

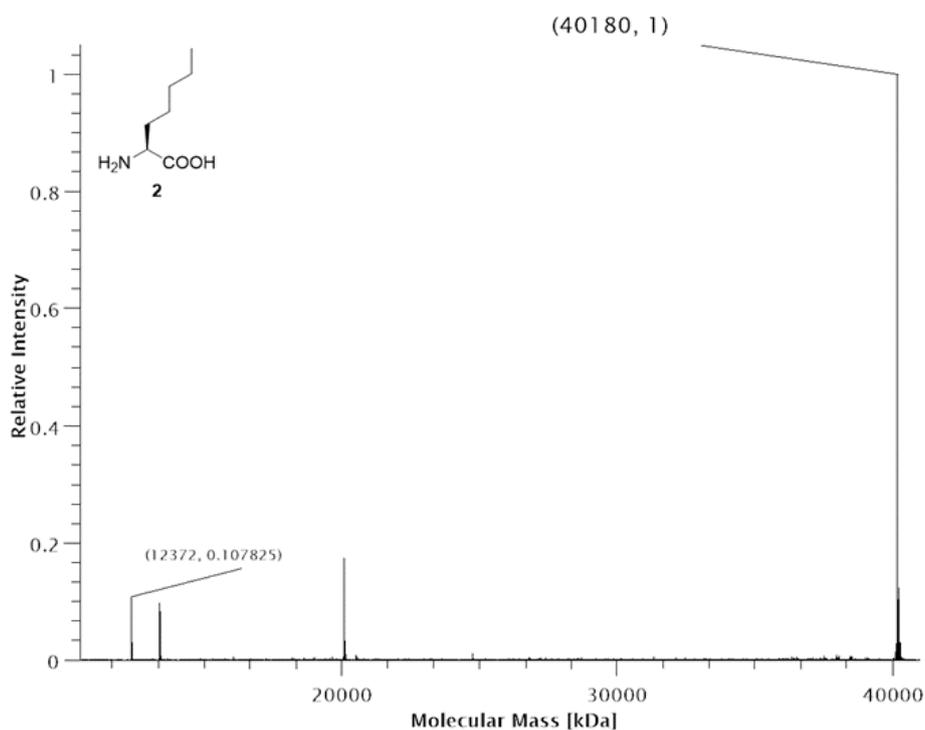


Figure 145: Deconvoluted ESI-MS spectrum of His₆-SUMO-sfGFP(R2(**2**))-Strep production in *E. coli* BL21(DE3) with co-expression of SmbP-*MbSacRS*(N311Q:C313W). Expected protein mass: 40178.8 Da. Observed mass: 40180 Da. Expected mass of His₆-SUMO truncation product: 12372.8 Da. Observed mass: 12372 Da.

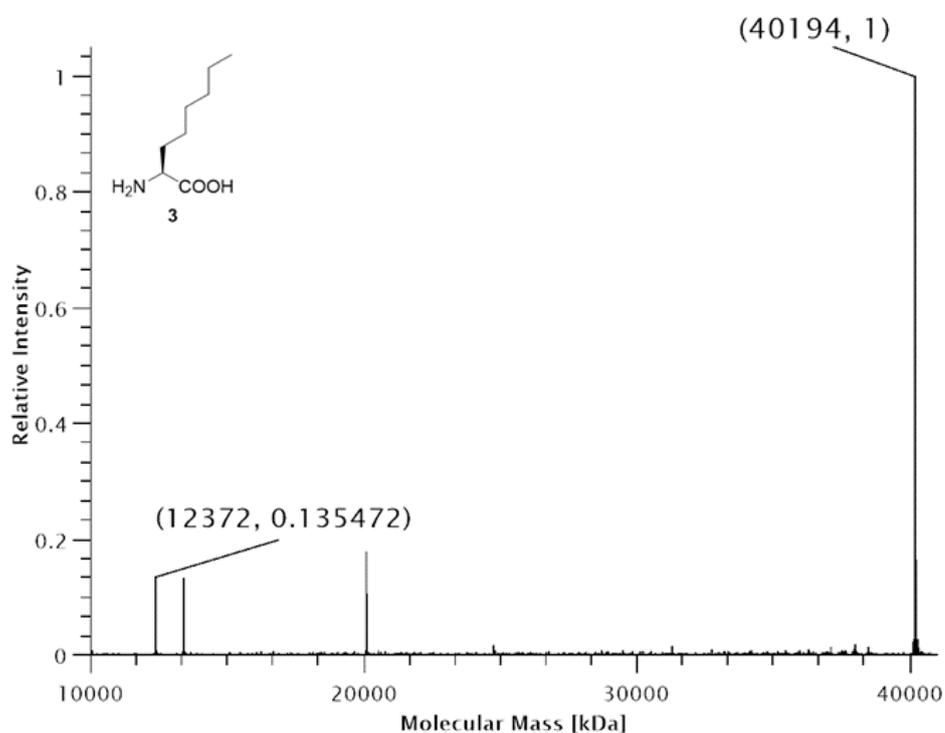


Figure 146: Deconvoluted ESI-MS spectrum of His₆-SUMO-sfGFP(R2(**3**))-Strep production in *E. coli* BL21(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W:V366A). Expected protein mass: 40192.9 Da. Observed mass: 40194 Da. Expected mass of His₆-SUMO truncation product: 12372.8 Da. Observed mass: 12372 Da.

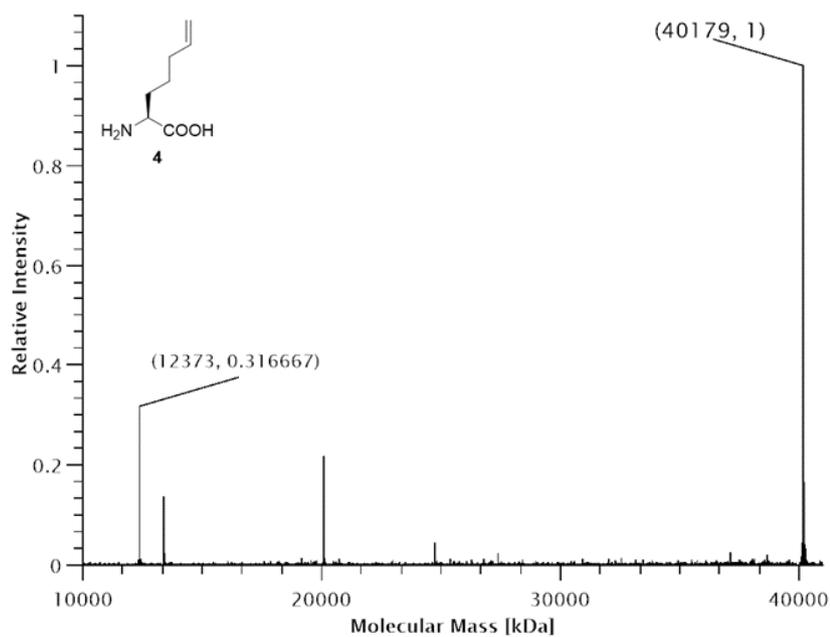


Figure 147: Deconvoluted ESI-MS spectrum of His₆-SUMO-sfGFP(R2(4))-Strep production in *E. coli* BL21(DE3) with co-expression of SmbP-MbSacRS(N311M:C313W:V366A). Expected protein mass: 40176.8 Da. Observed mass: 40179 Da. Expected mass of His₆-SUMO truncation product: 12372.8 Da. Observed mass: 12373 Da.

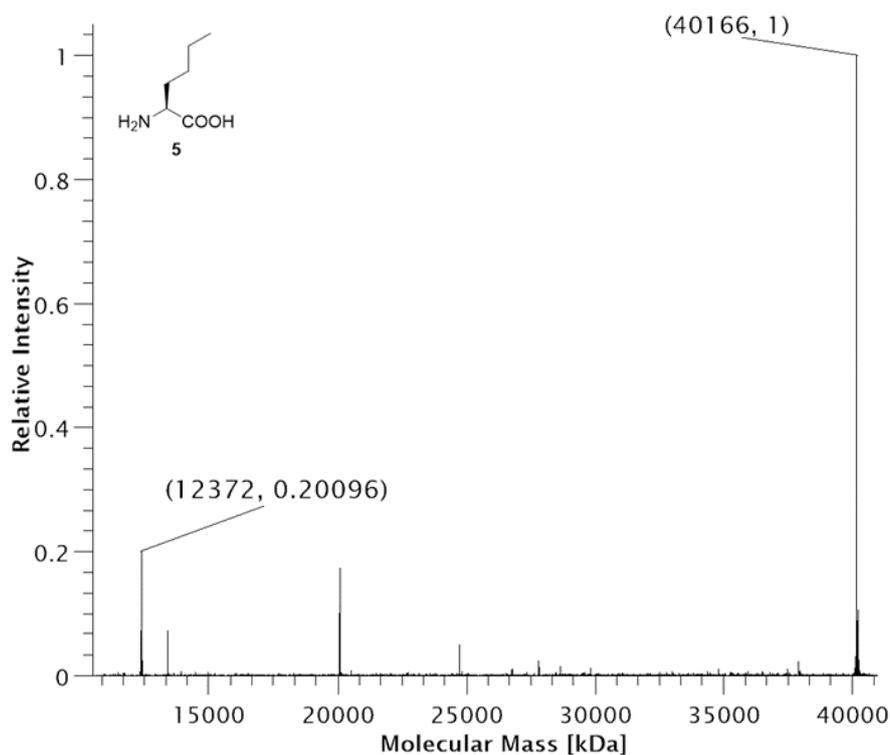


Figure 148: Deconvoluted ESI-MS spectrum of His₆-SUMO-sfGFP(R2(5))-Strep production in *E. coli* BL21(DE3) with co-expression of SmbP-MbSacRS(N311M:C313W). Expected protein mass: 40164.8 Da. Observed mass: 40166 Da. Expected mass of His₆-SUMO truncation product: 12372.8 Da. Observed mass: 12372 Da.

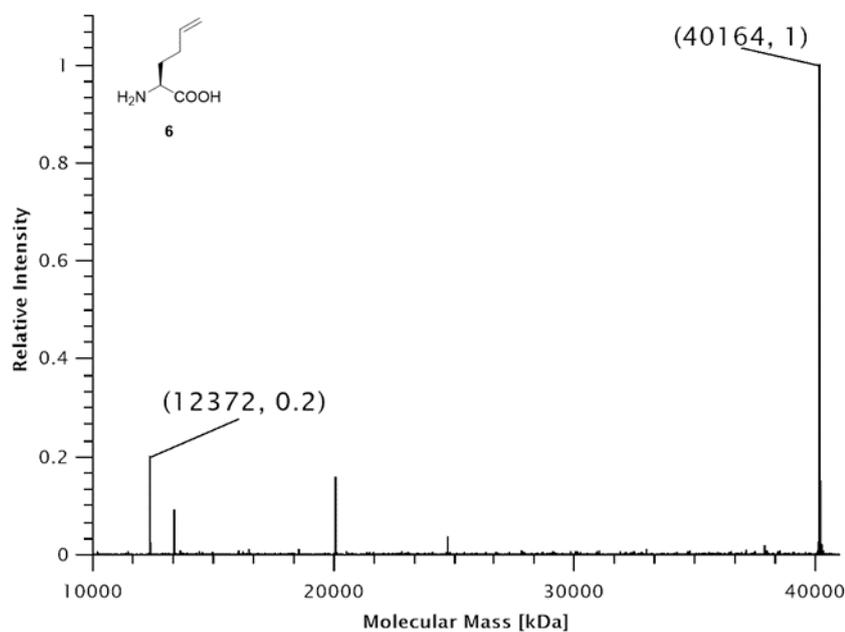


Figure 149: Deconvoluted ESI-MS spectrum of His₆-SUMO-sfGFP(R2(**6**))-Strep production in *E. coli* BL21(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W:V366K). Expected protein mass: 40162.8 Da. Observed mass: 40164 Da. Expected mass of His₆-SUMO truncation product: 12372.8 Da. Observed mass: 12372 Da.

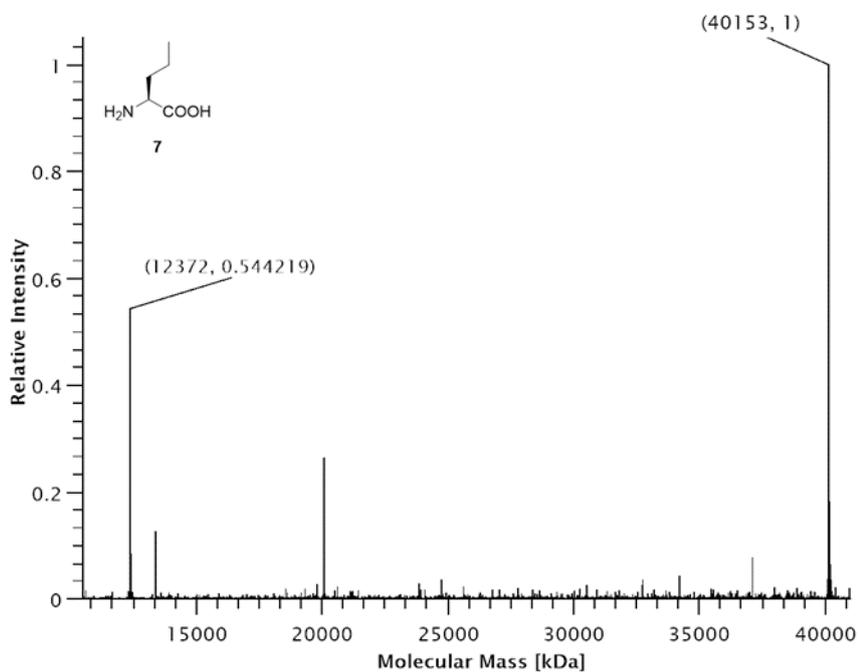


Figure 150: Deconvoluted ESI-MS spectrum of His₆-SUMO-sfGFP(R2(**7**))-Strep production in *E. coli* BL21(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W). Expected protein mass: 40150.8 Da. Observed mass: 40153 Da. Expected mass of His₆-SUMO truncation product: 12372.8 Da. Observed mass: 12372 Da.

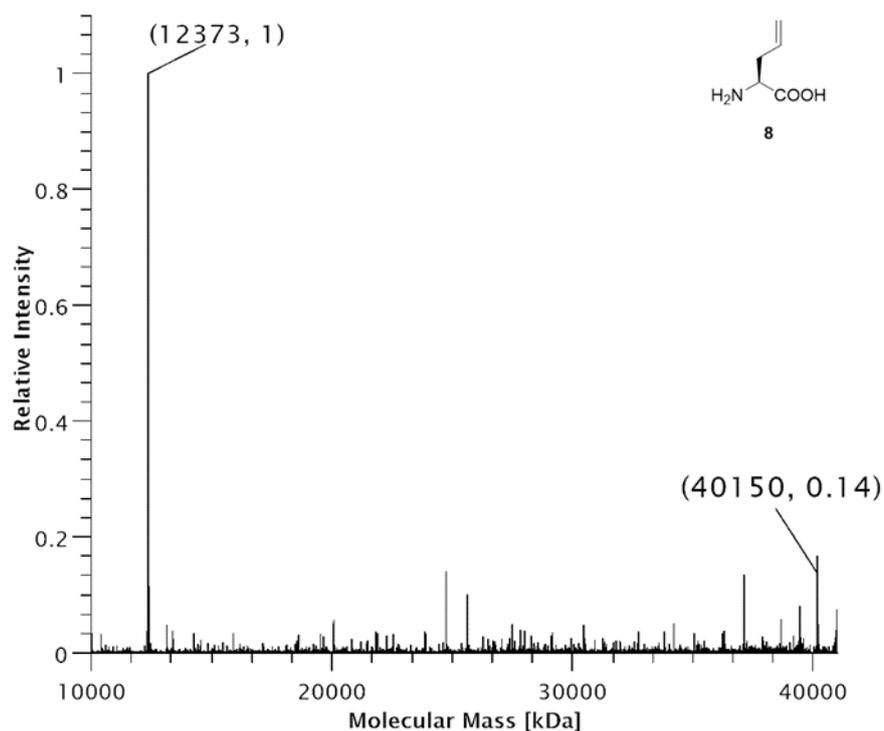


Figure 151: Deconvoluted ESI-MS spectrum of His₆-SUMO-sfGFP(R2(**8**))-Strep production in *E. coli* BL21(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W). Expected protein mass: 40148.8 Da. Observed mass: 40150 Da. Expected mass of His₆-SUMO truncation product: 12372.8 Da. Observed mass: 12373 Da.

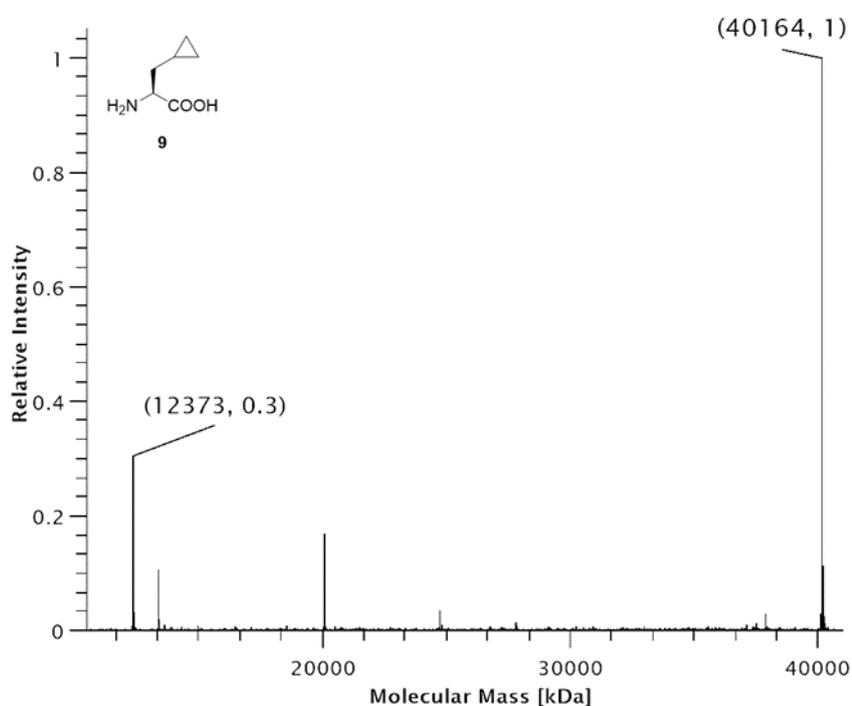


Figure 152: Deconvoluted ESI-MS spectrum of His₆-SUMO-sfGFP(R2(**9**))-Strep production in *E. coli* BL21(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W). Expected protein mass: 40162.8 Da. Observed mass: 40164 Da. Expected mass of His₆-SUMO truncation product: 12372.8 Da. Observed mass: 12373 Da.

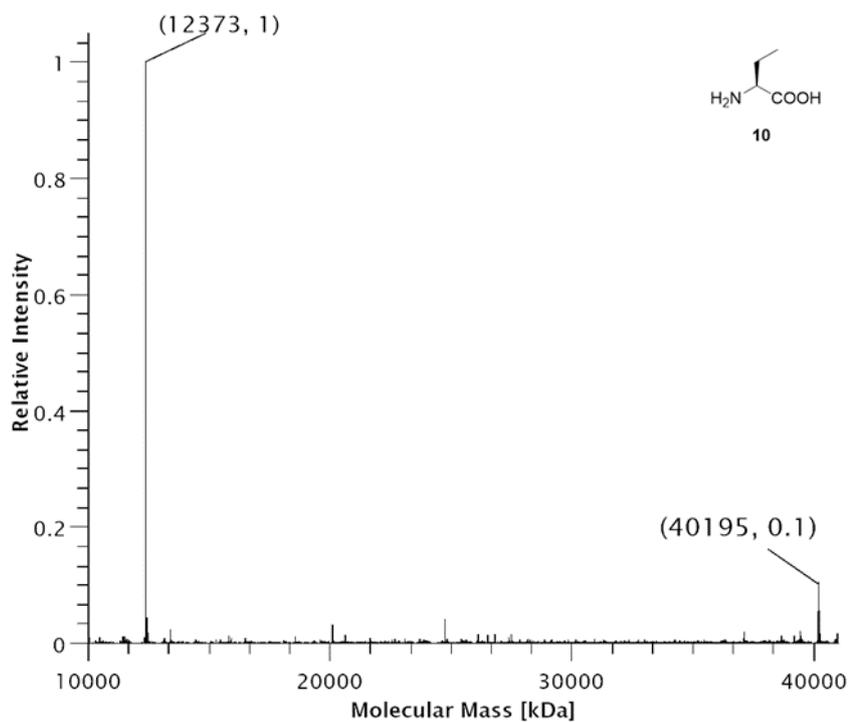


Figure 153: Deconvoluted ESI-MS spectrum of His₆-SUMO-sfGFP(R2(**10**))-Strep production in *E. coli* BL21(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W:V366A). Expected protein mass: 40136.8 Da. Observed mass: 40195 Da. Expected mass of His₆-SUMO truncation product: 12372.8 Da. Observed mass: 12373 Da.

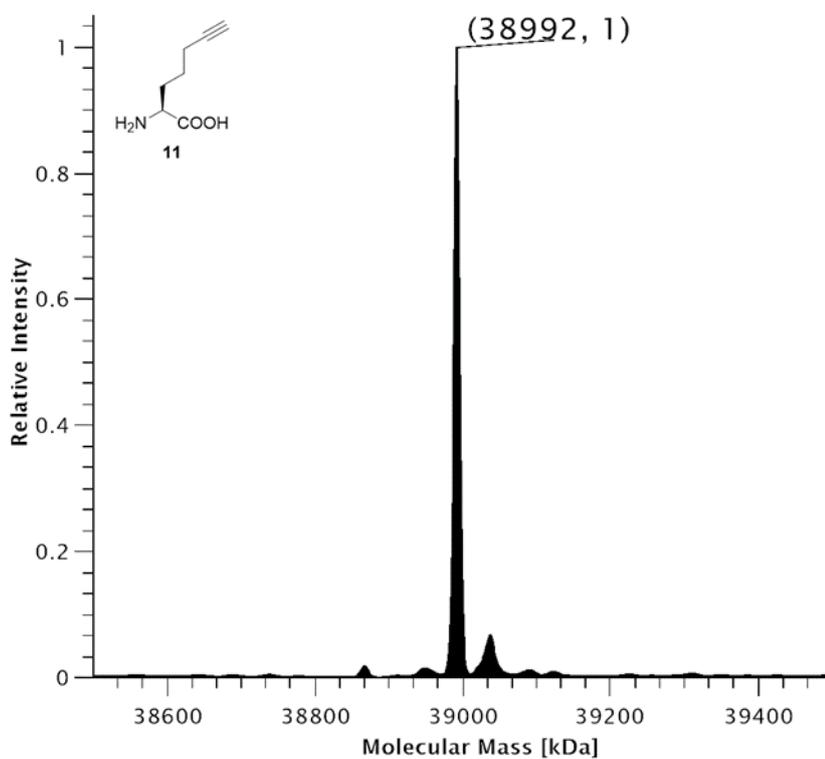


Figure 154: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(**11**))-His₆ production in *E. coli* C321.ΔA.exp(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W). Expected protein mass: 38990.9 Da. Observed mass: 38992 Da.

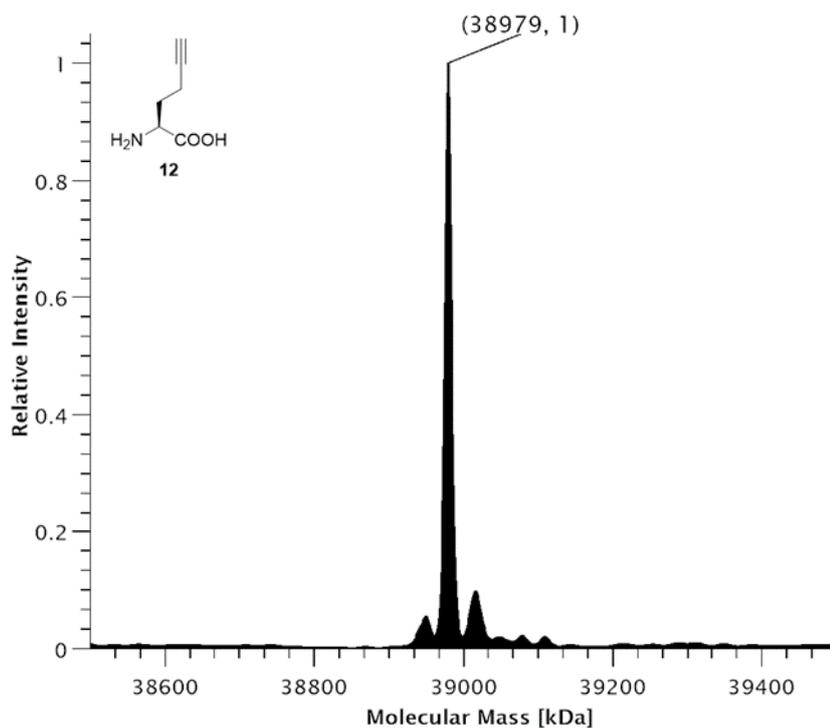


Figure 155: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(**12**))-His₆ production in *E. coli* C321.ΔA.exp(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W). Expected protein mass: 38976.9 Da. Observed mass: 38979 Da.

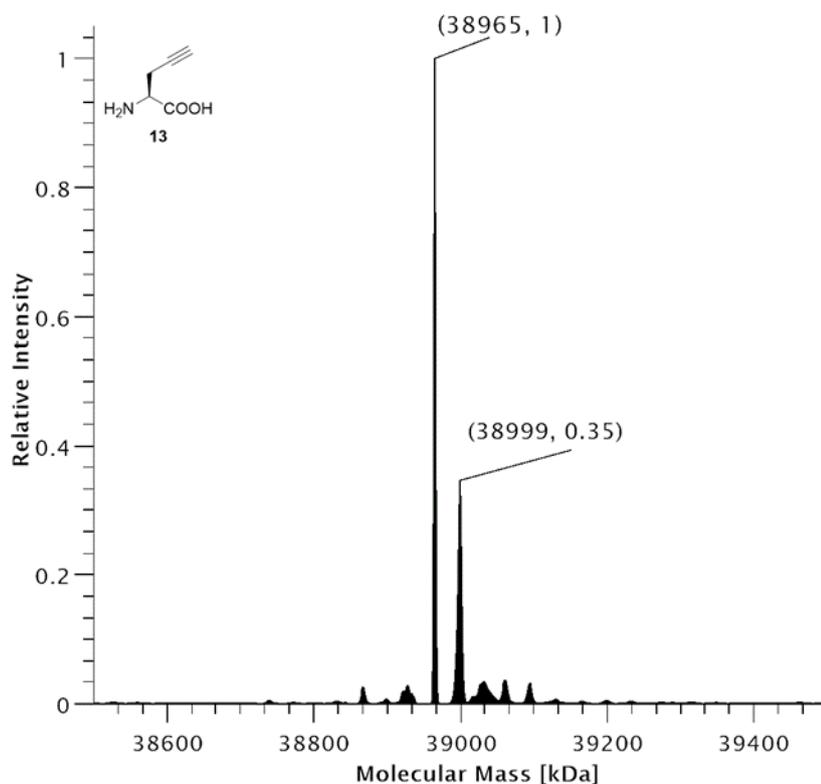


Figure 156: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(**13**))-His₆ production in *E. coli* C321.ΔA.exp(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W). Expected protein mass: 38962.8 Da. Observed mass: 38965 Da.

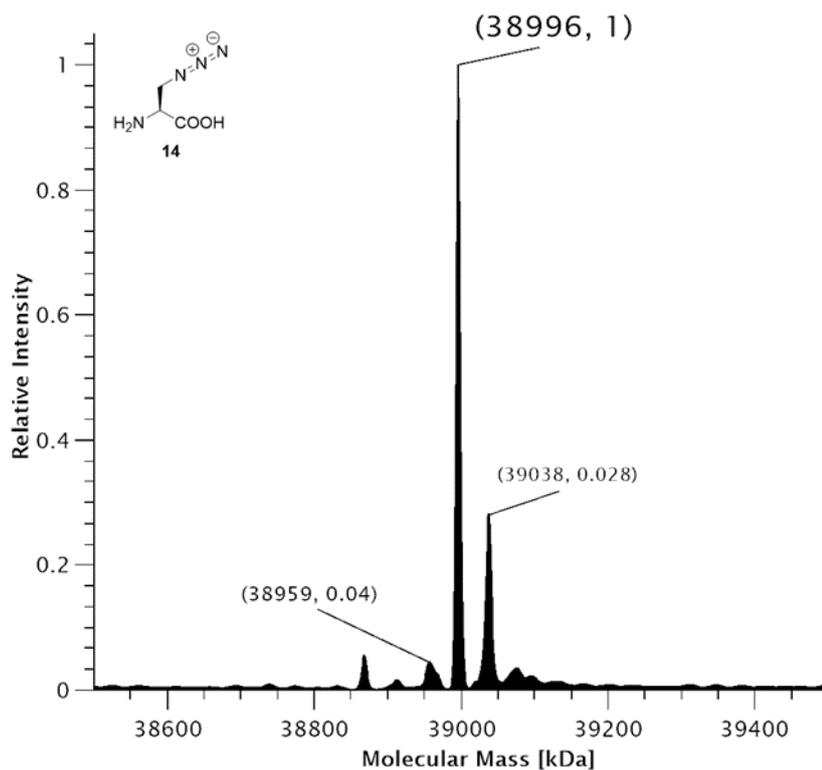


Figure 157: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(**14**))-His₆ production in *E. coli* JX33(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W). Expected protein mass: 38979.8 Da. Observed mass: 38996 Da.

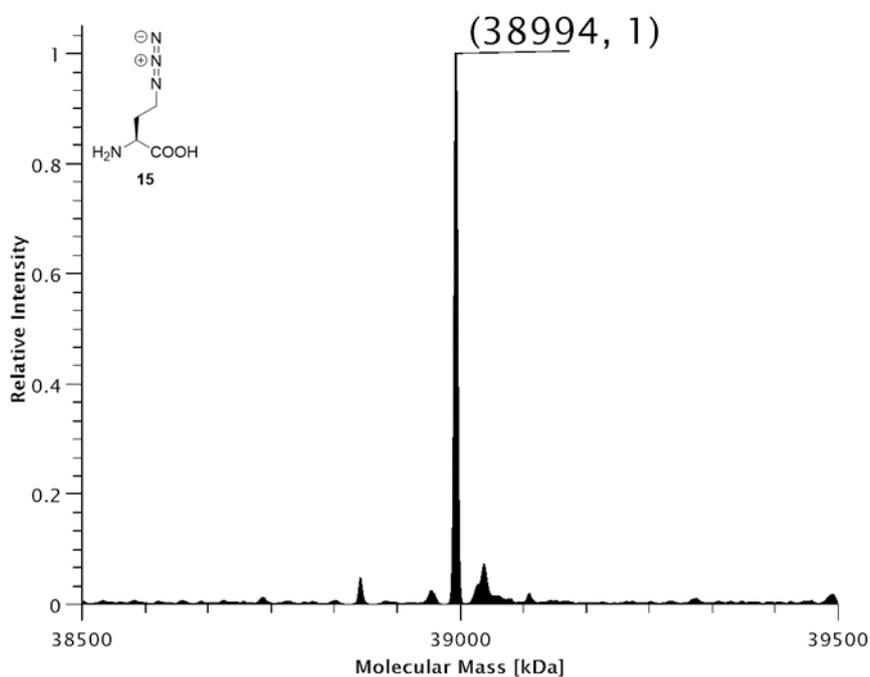


Figure 158: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(**15**))-His₆ production in *E. coli* C321.ΔA.exp(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W). Expected protein mass: 38993.8 Da. Observed mass: 38994 Da.

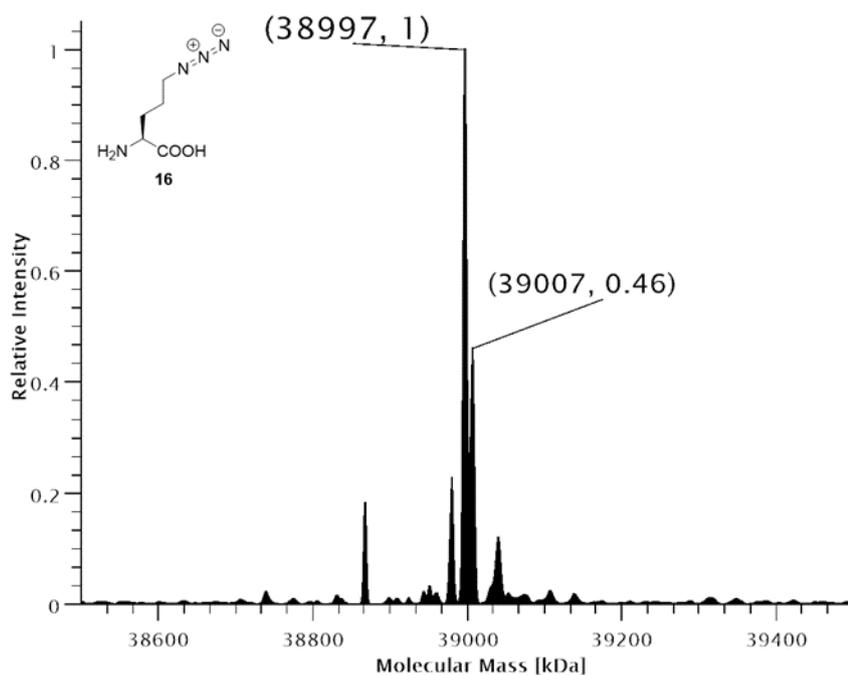


Figure 159: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(**16**))-His₆ production in *E. coli* C321.ΔA.exp(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W). Expected protein mass: 39007.9 Da. Observed mass: 39007 Da.

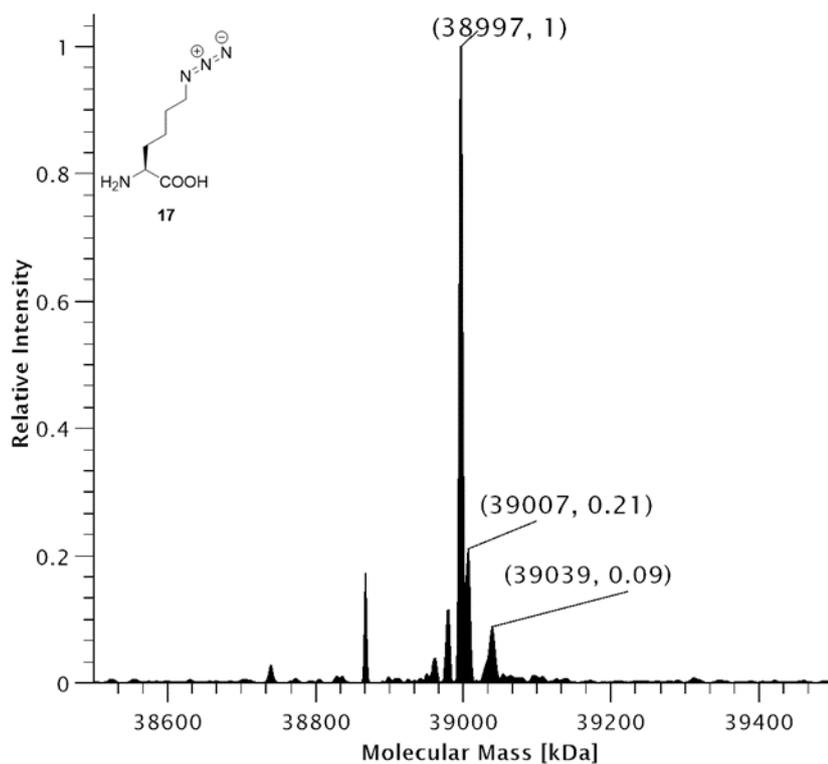
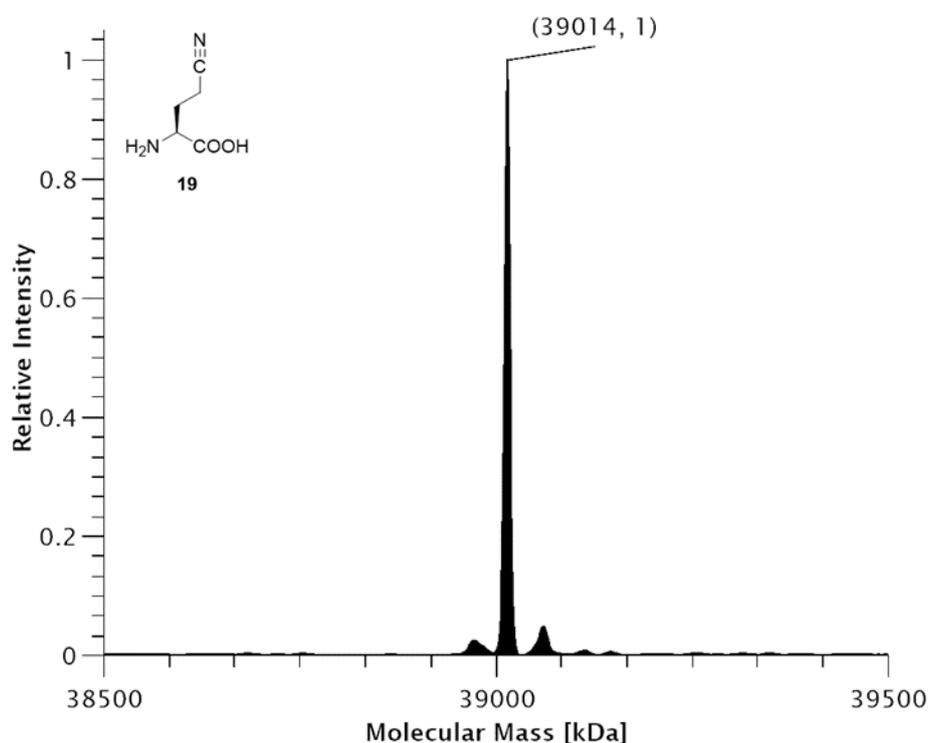
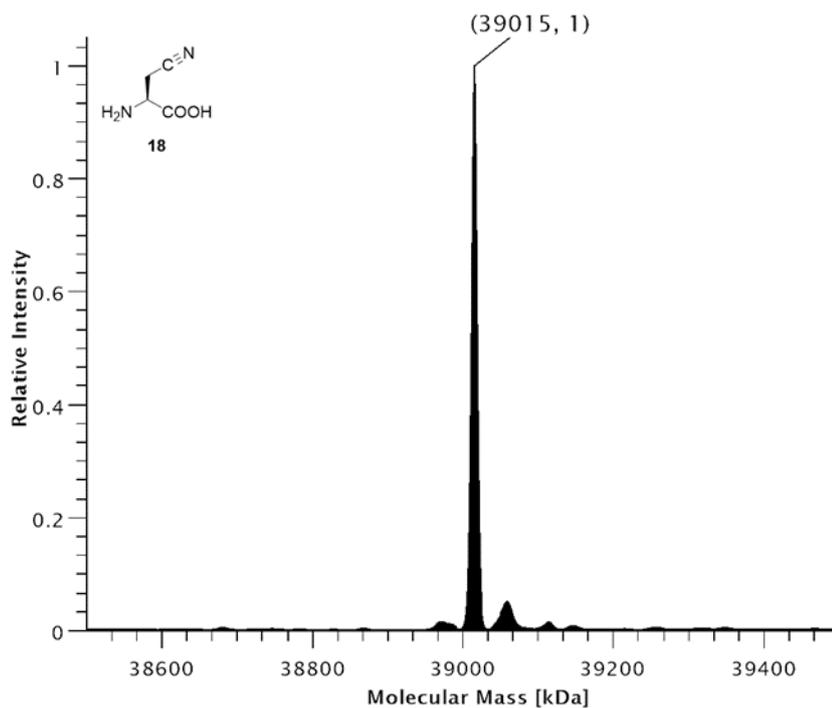


Figure 160: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(**17**))-His₆ production in *E. coli* C321.ΔA.exp(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W). Expected protein mass: 39021.9 Da. Observed mass: 38997 Da.



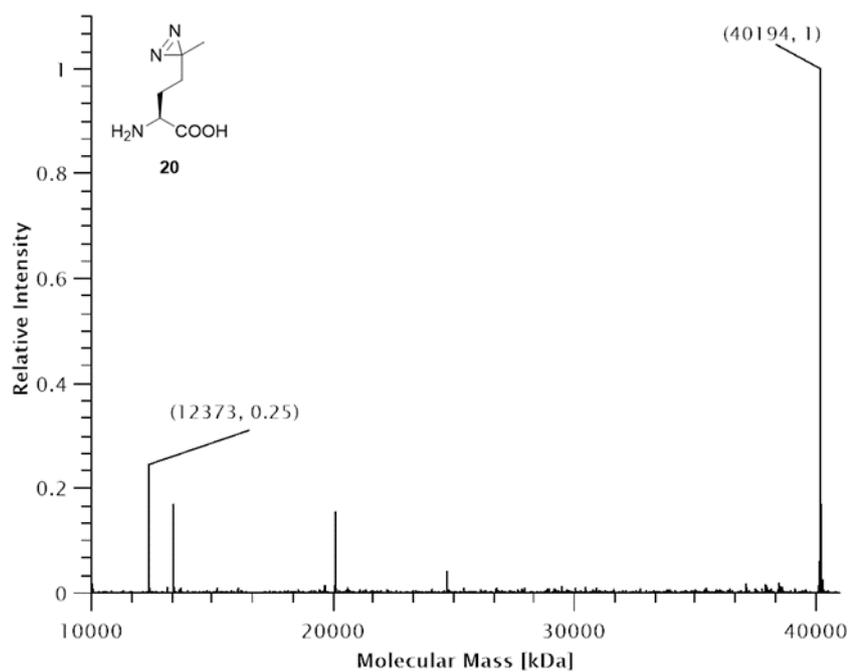


Figure 163: Deconvoluted ESI-MS spectrum of His₆-SUMO-sfGFP(R2(**20**))-Strep production in *E. coli* BL21(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W:V366A). Expected protein mass: 40190.8 Da. Observed mass: 40194 Da. Expected mass of His₆-SUMO truncation product: 12372.8 Da. Observed mass: 12373 Da.

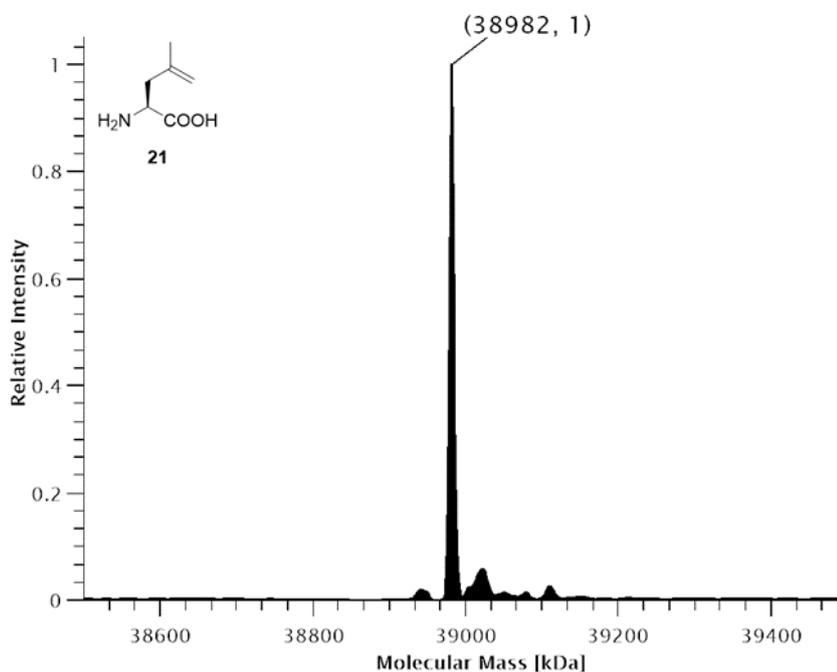


Figure 164: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(**21**))-His₆ production in *E. coli* BL21(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W). Expected protein mass: 38978.9 Da. Observed mass: 38982 Da.

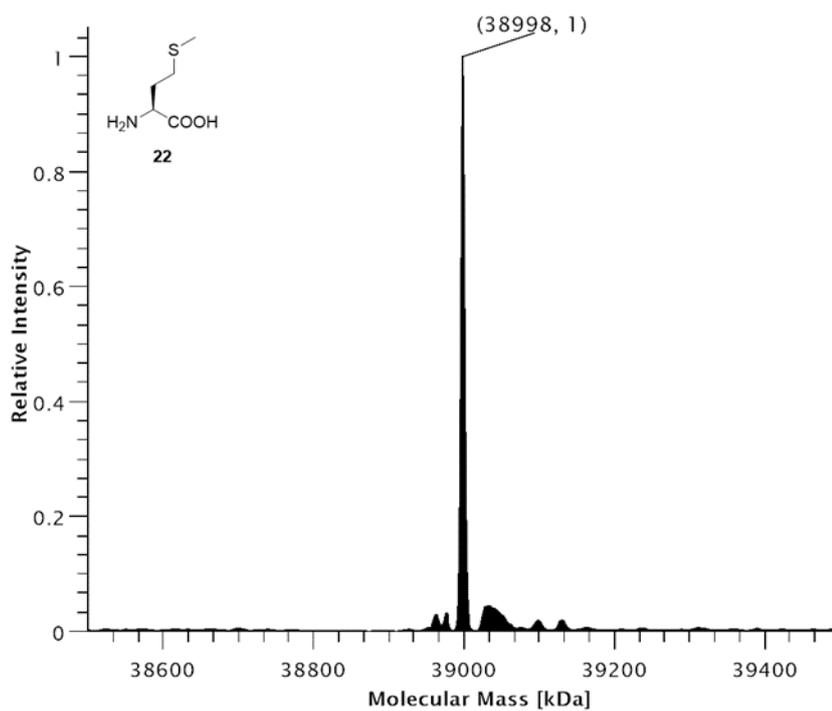


Figure 165: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(**22**))-His₆ production in *E. coli* BL21(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W). Expected protein mass: 38998.9 Da. Observed mass: 38998 Da.

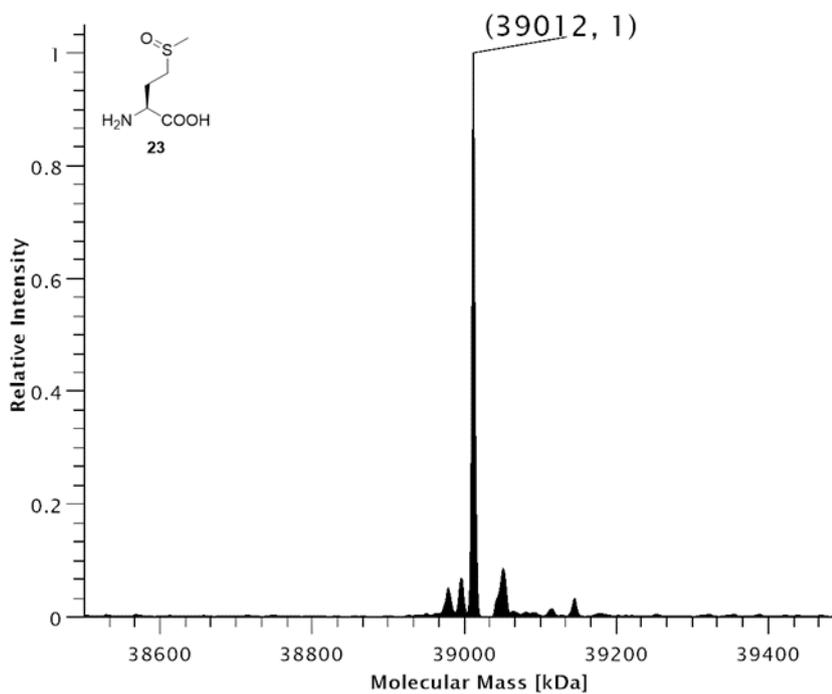
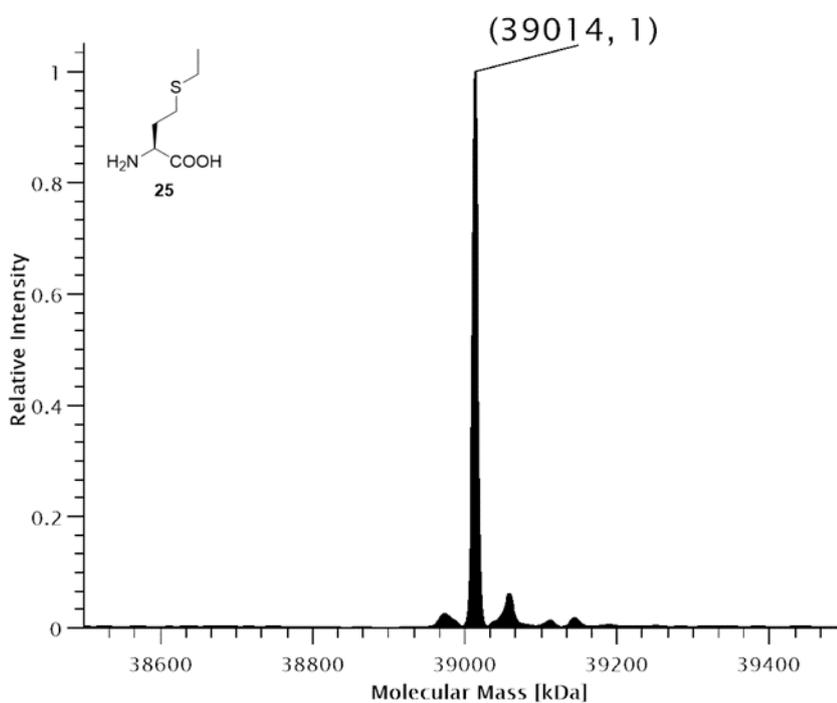
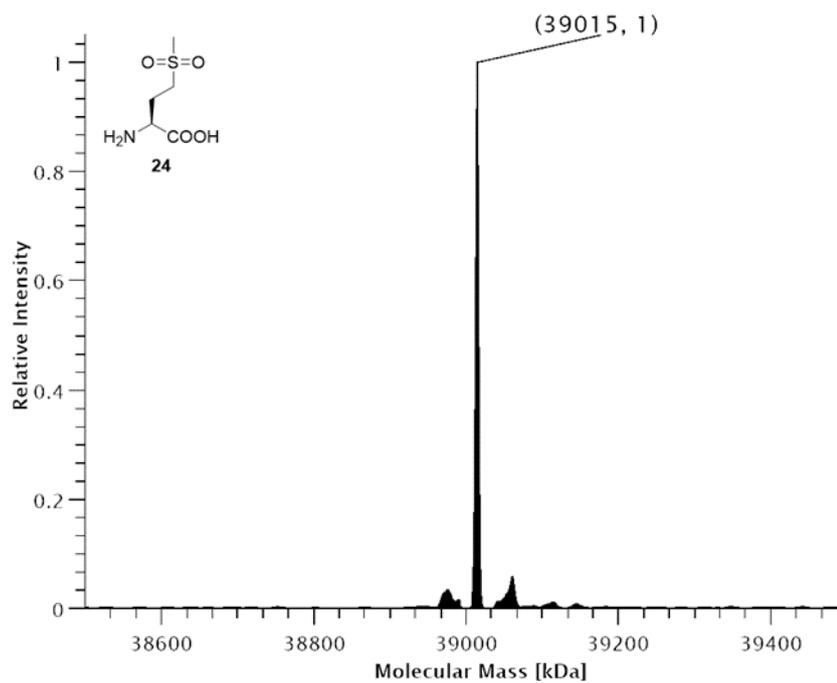


Figure 166: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(**23**))-His₆ production in *E. coli* BL21(DE3) with co-expression of SmbP-*MbSacRS*(N311Q:C313W). Expected protein mass: 39014.9 Da. Observed mass: 39012 Da.



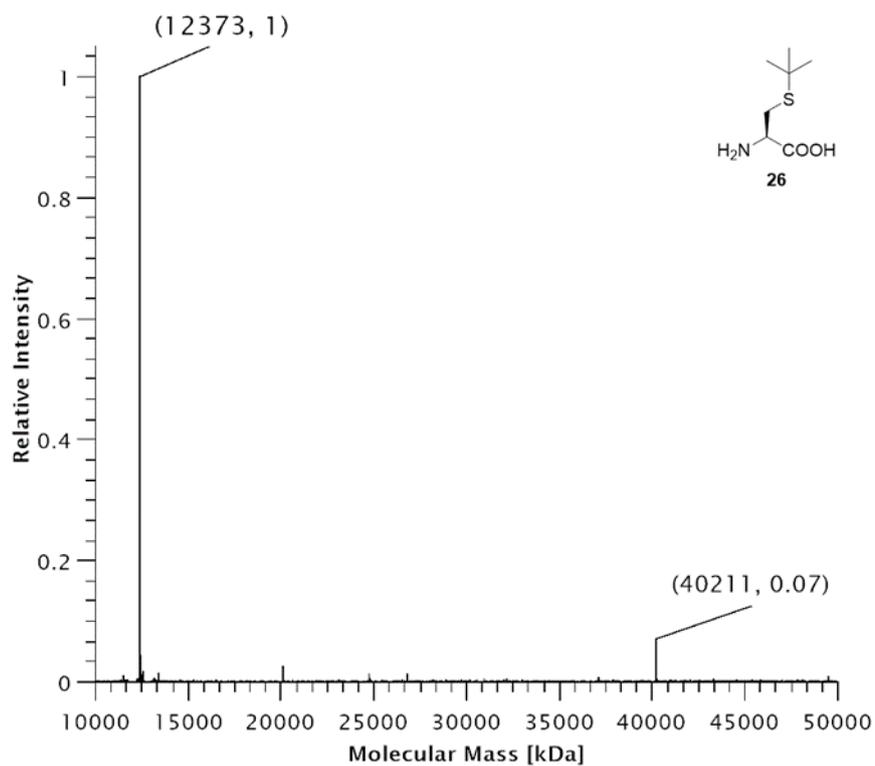


Figure 169: Deconvoluted ESI-MS spectrum of His₆-SUMO-sfGFP(R2(**26**))-Strep production in *E. coli* BL21(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W:V366A). Expected protein mass: 40190.8 Da. Observed mass: 40194 Da. Expected mass of His₆-SUMO truncation product: 12372.8 Da. Observed mass: 12373 Da.

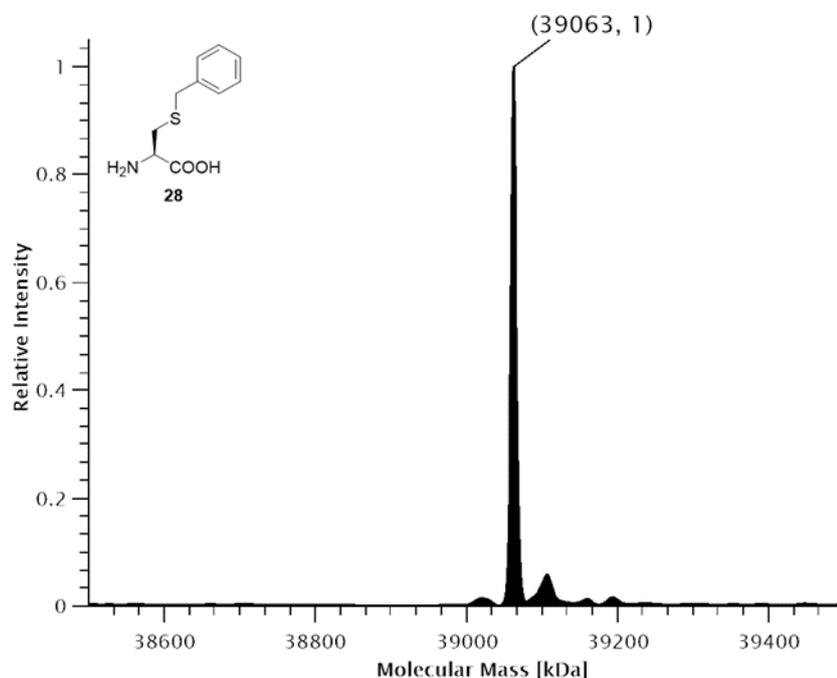


Figure 170: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(**28**))-His₆ production in *E. coli* BL21(DE3) with co-expression of SmbP-*MbSacRS*(N311M:C313W:V366A:W382N). Expected protein mass: 39061 Da. Observed mass: 39063 Da.

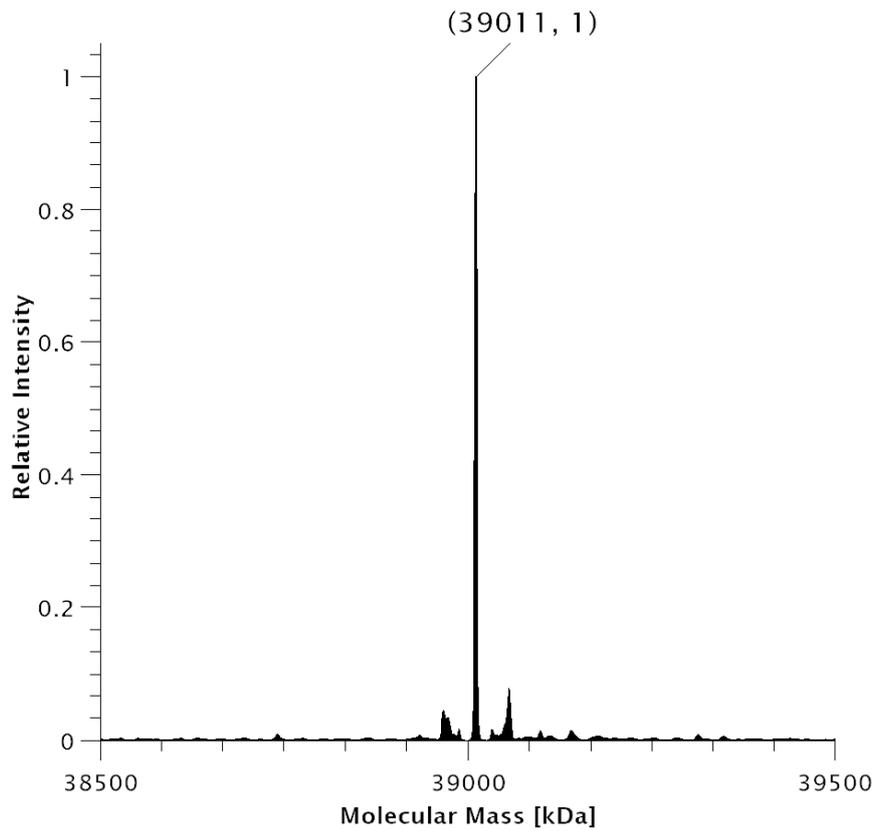


Figure 171: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(1))-His₆ production in *E. coli* BL21(DE3) with co-expression of *MburSacRS*(S379T). Expected protein mass: 39011 Da. Observed mass: 39011 Da.

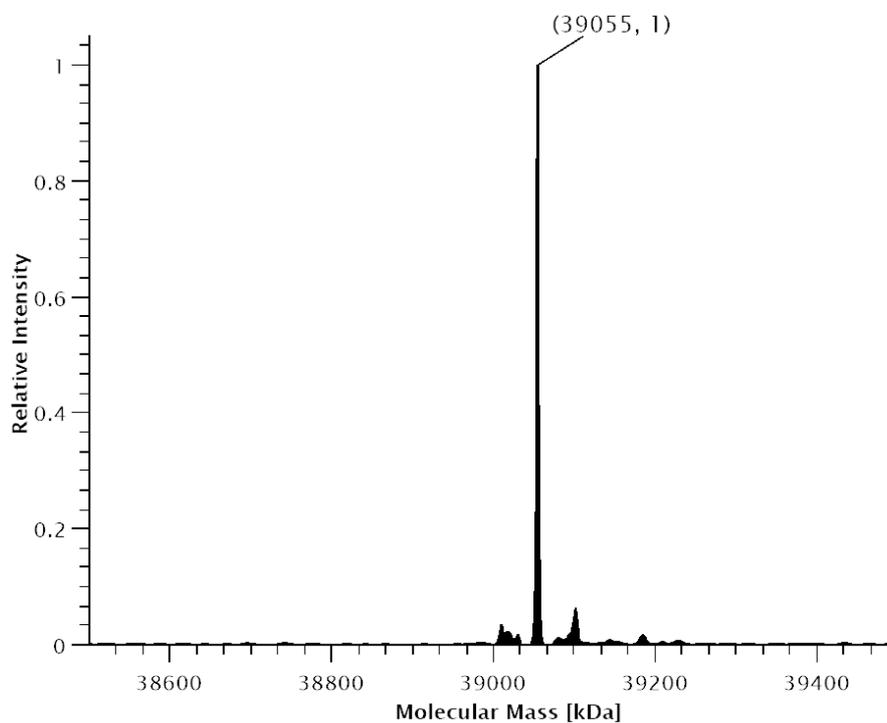


Figure 172: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(3x 1)-His₆ production in *E. coli* BL21(DE3) with co-expression of *MburSacRS*(S379T). Expected protein mass: 39055 Da. Observed mass: 39055 Da.

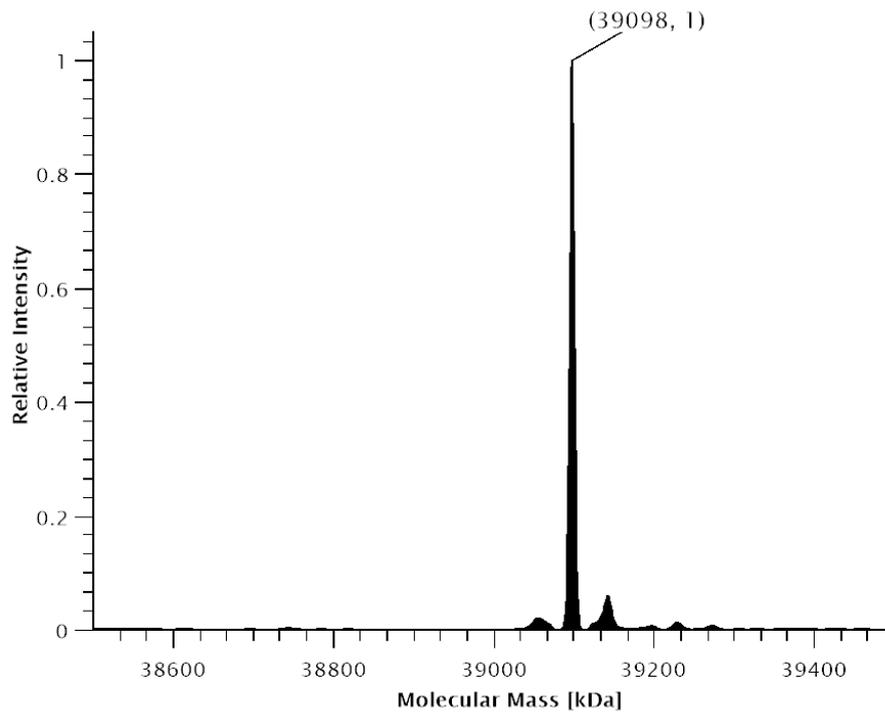


Figure 173: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(5x 1)-His₆ production in *E. coli* BL21(DE3) with co-expression of *MburSacRS*(S379T). Expected protein mass: 39098 Da. Observed mass: 39098 Da.

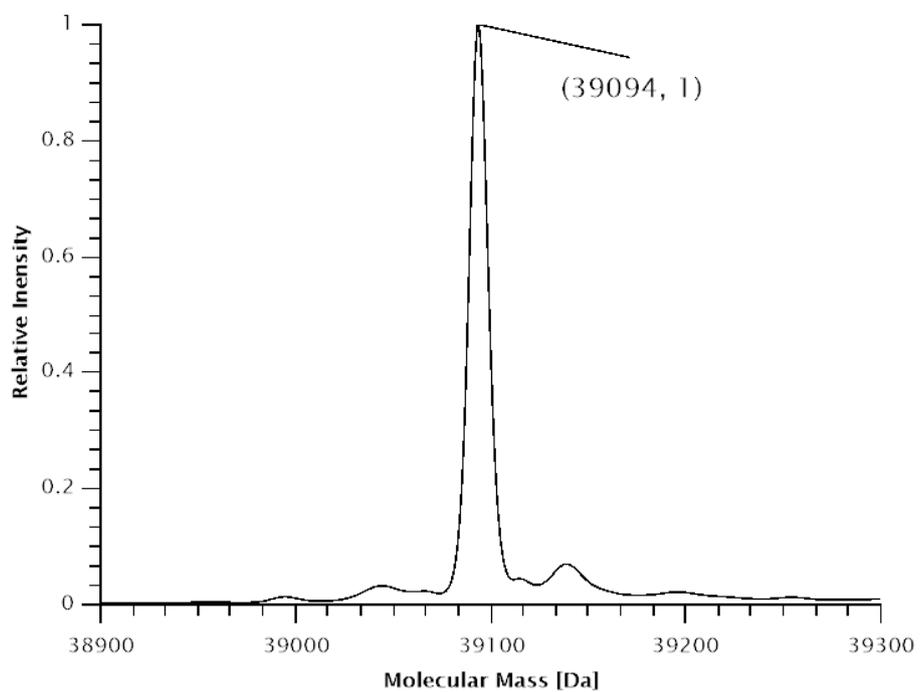


Figure 174: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(31))-His₆ production in *E. coli* BL21(DE3) with co-expression of *MmPylRS*. Expected protein mass: 39096 Da. Observed mass: 39094 Da.

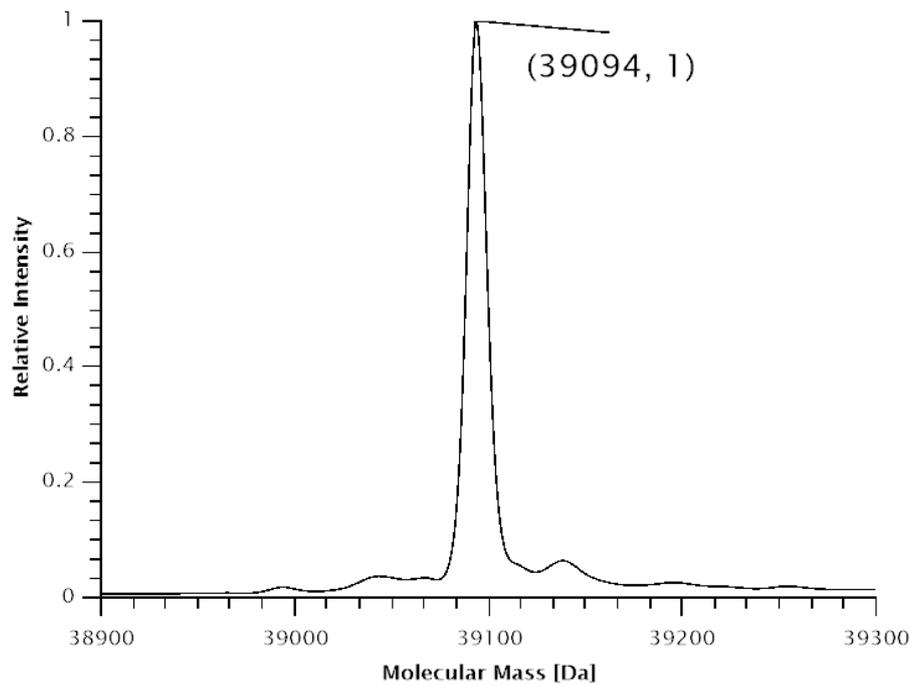


Figure 175: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(31))-His₆ production in *E. coli* BL21(DE3) with co-expression of *MburPyIRS*. Expected protein mass: 39096 Da. Observed mass: 39094 Da.

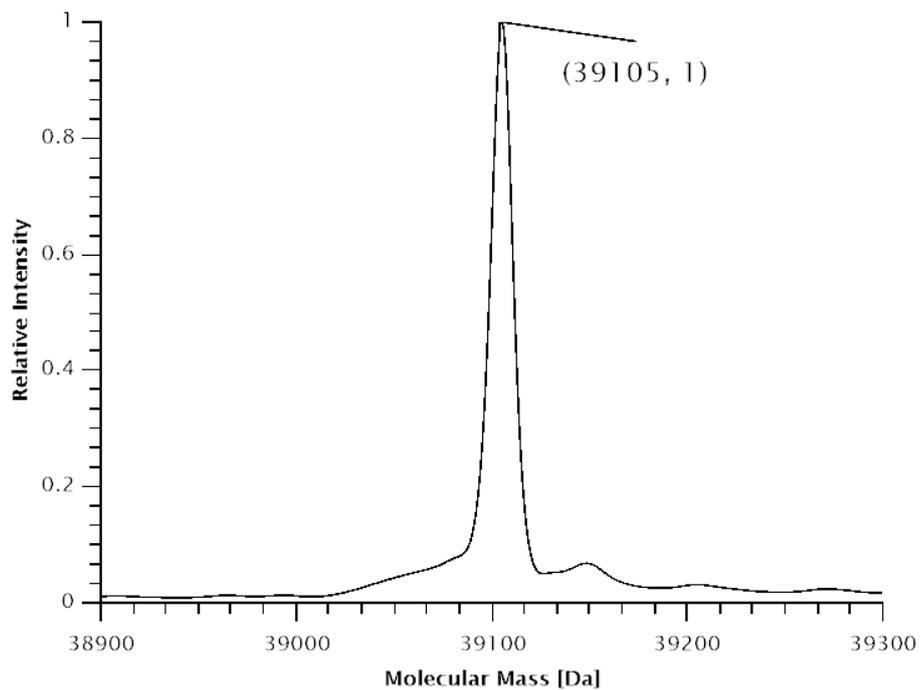


Figure 176: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(34))-His₆ production in *E. coli* BL21(DE3) with co-expression of *MmPyIRS*. Expected protein mass: 39099 Da. Observed mass: 39095 Da.

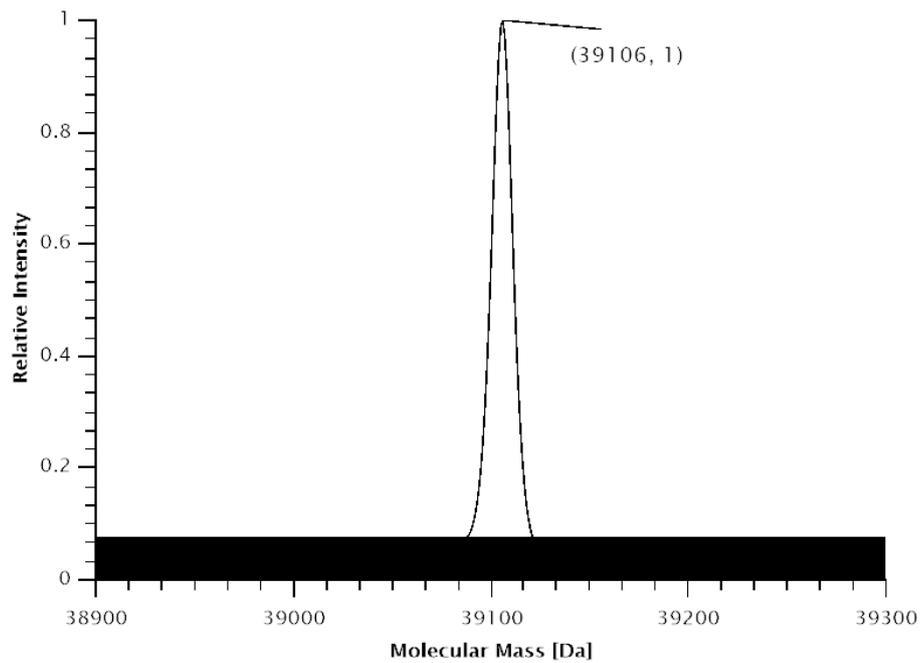


Figure 177: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(34))-His₆ production in *E. coli* BL21(DE3) with co-expression of *MburPyIRS*. Expected protein mass: 39099 Da. Observed mass: 39096 Da.

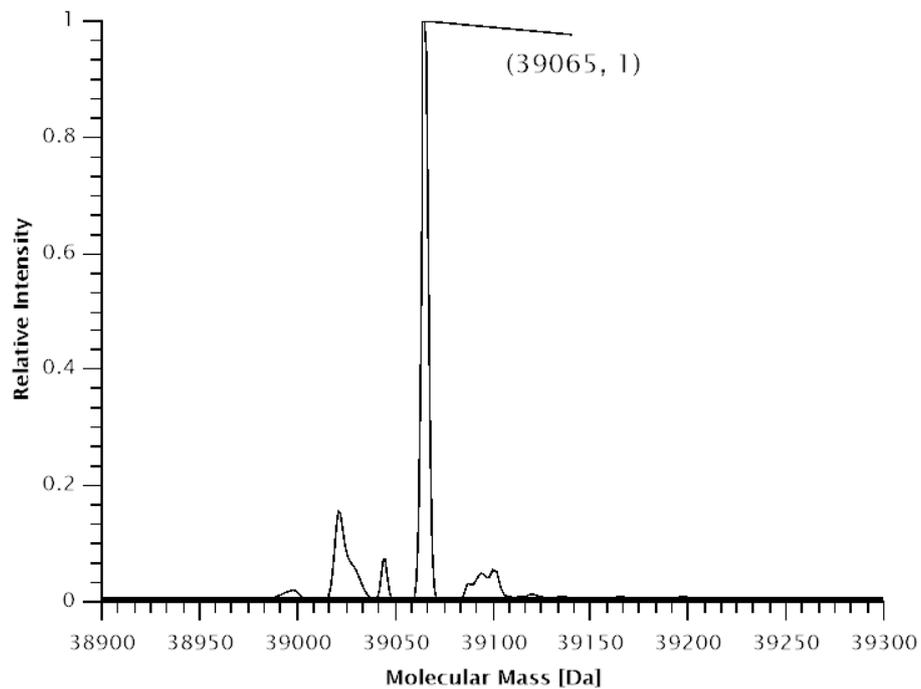


Figure 178: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(40))-His₆ production in *E. coli* BL21(DE3) with co-expression of *MmPyIRS*. Expected protein mass: 39069 Da. Observed mass: 39065 Da.

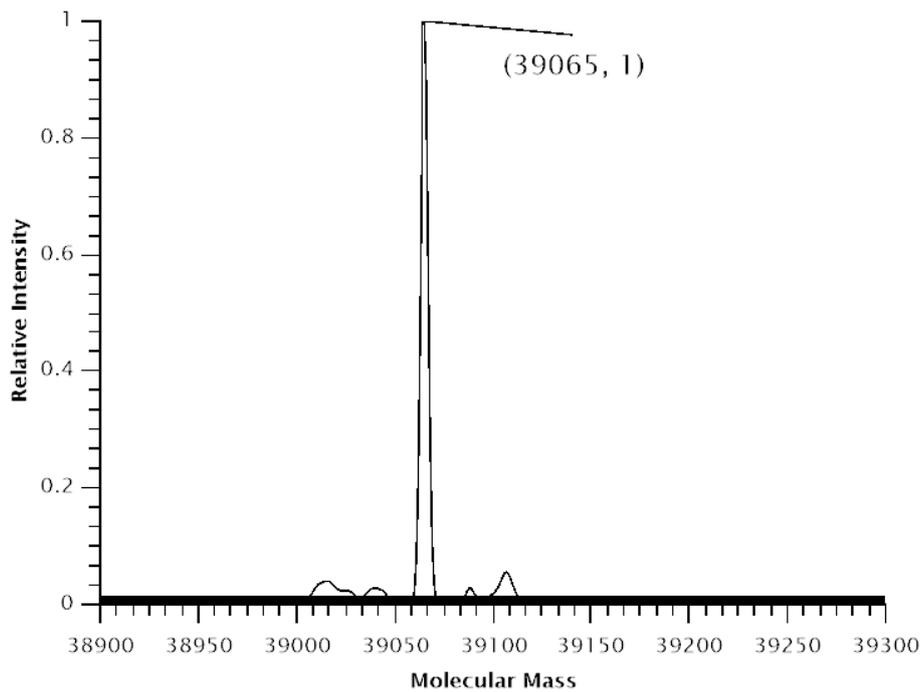


Figure 179: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(40))-His₆ production in *E. coli* BL21(DE3) with co-expression of *MburPyIRS*. Expected protein mass: 39069 Da. Observed mass: 39065 Da.

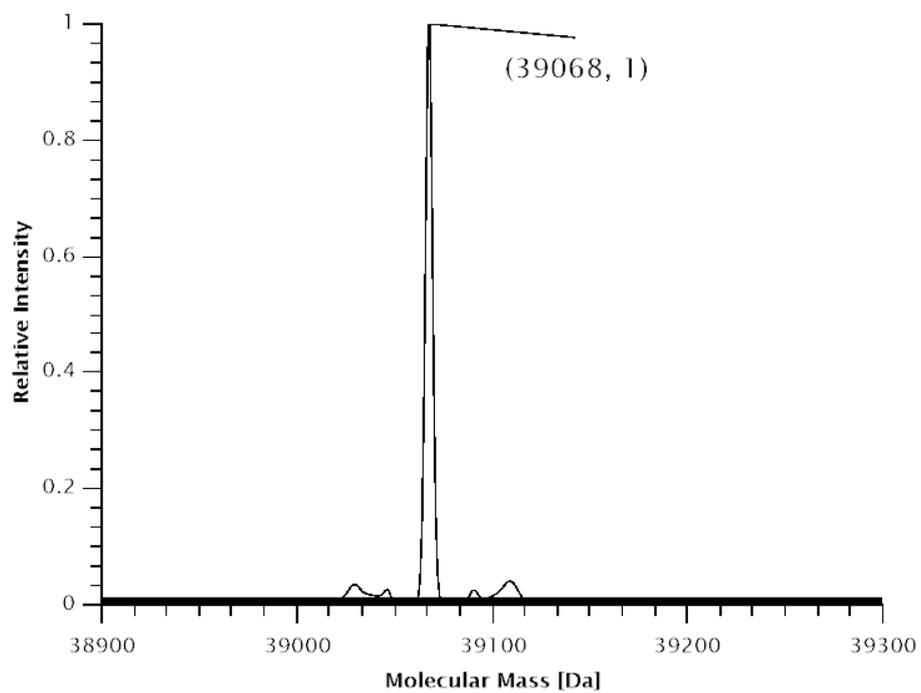


Figure 180: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(45))-His₆ production in *E. coli* BL21(DE3) with co-expression of *MmPyIRS*. Expected protein mass: 39071 Da. Observed mass: 39068 Da.

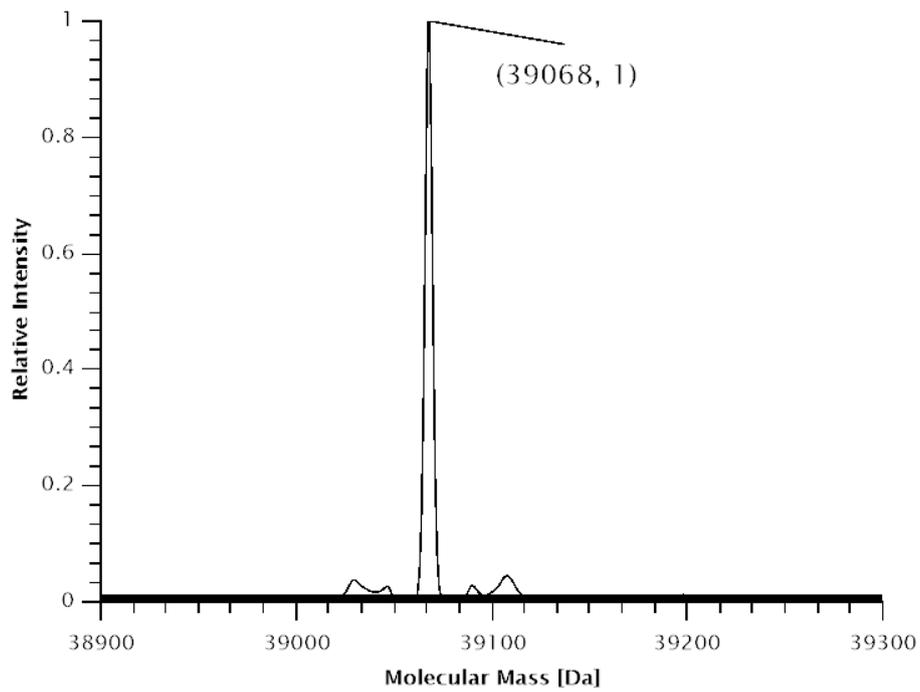


Figure 181: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(45))-His₆ production in *E. coli* BL21(DE3) with co-expression of *MburPyIRS*. Expected protein mass: 39071 Da. Observed mass: 39068 Da.

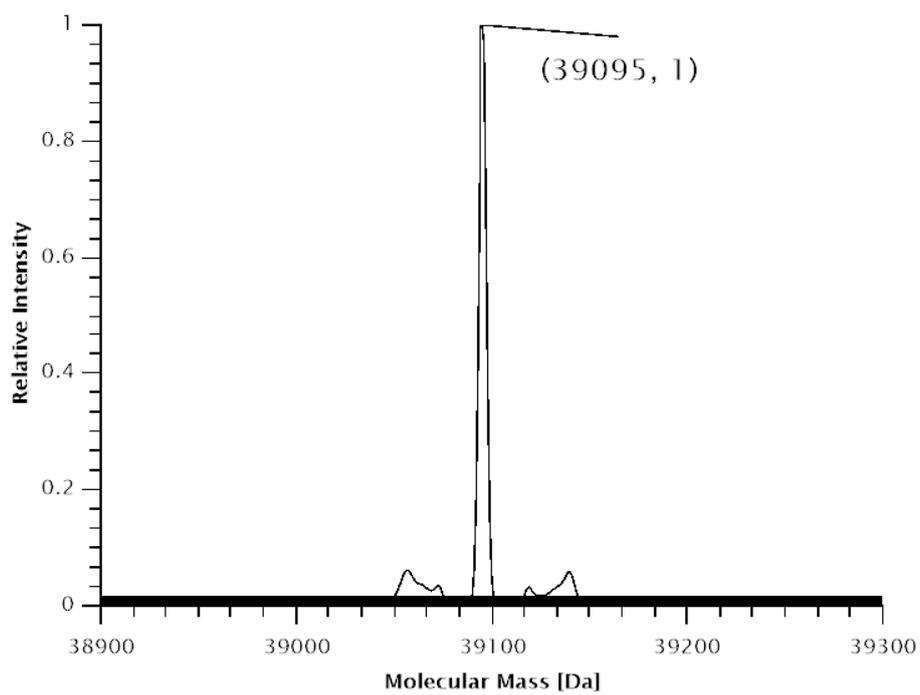


Figure 182: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(47))-His₆ production in *E. coli* BL21(DE3) with co-expression of *MmPyIRS*. Expected protein mass: 39099 Da. Observed mass: 39095 Da.

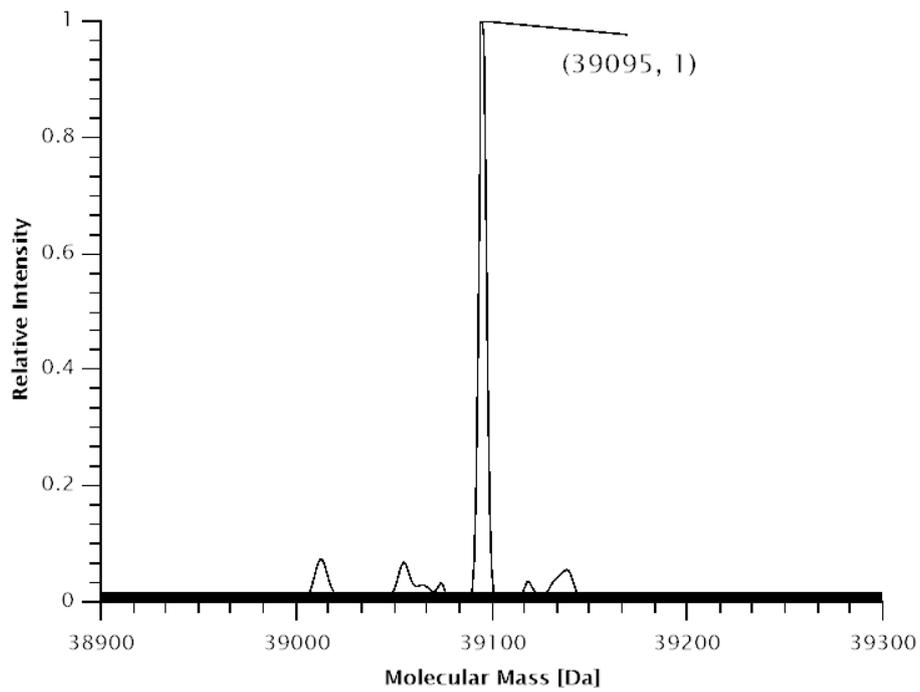


Figure 183: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(47))-His₆ production in *E. coli* BL21(DE3) with co-expression of *MburPyIRS*. Expected protein mass: 39099 Da. Observed mass: 39095 Da.

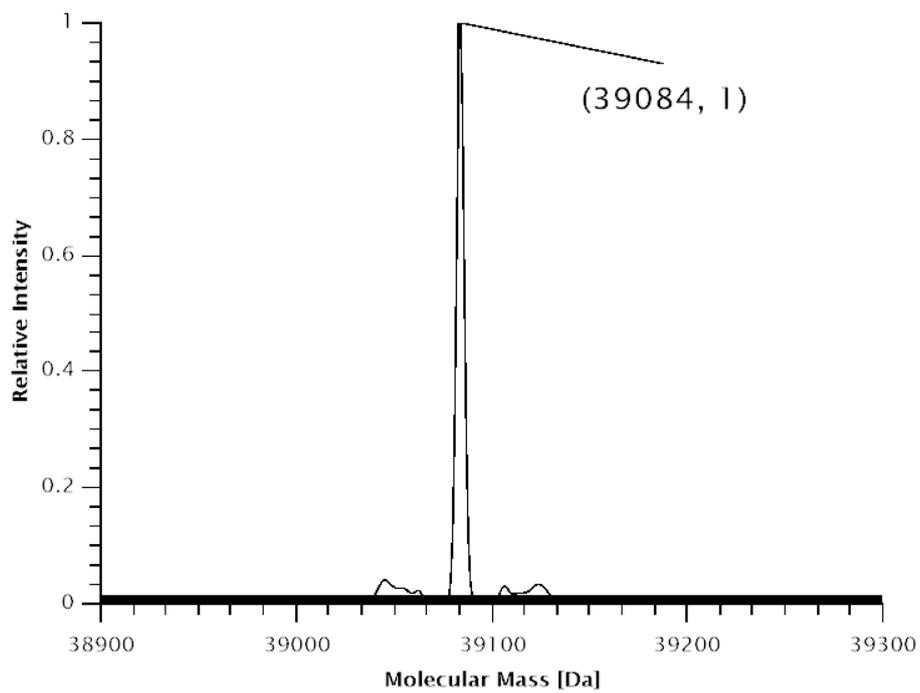


Figure 184: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(39))-His₆ production in *E. coli* BL21(DE3) with co-expression of *MmPyIRS*. Expected protein mass: 39087 Da. Observed mass: 39084 Da.

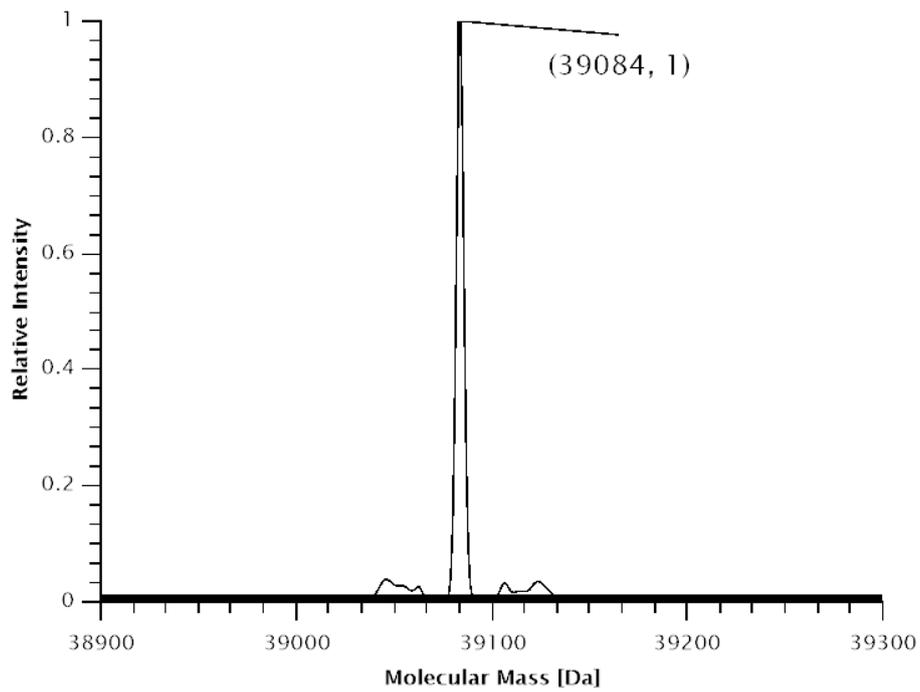


Figure 185: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(**39**))-His₆ production in *E. coli* BL21(DE3) with co-expression of *MburPyIRS*. Expected protein mass: 39087 Da. Observed mass: 39084 Da.

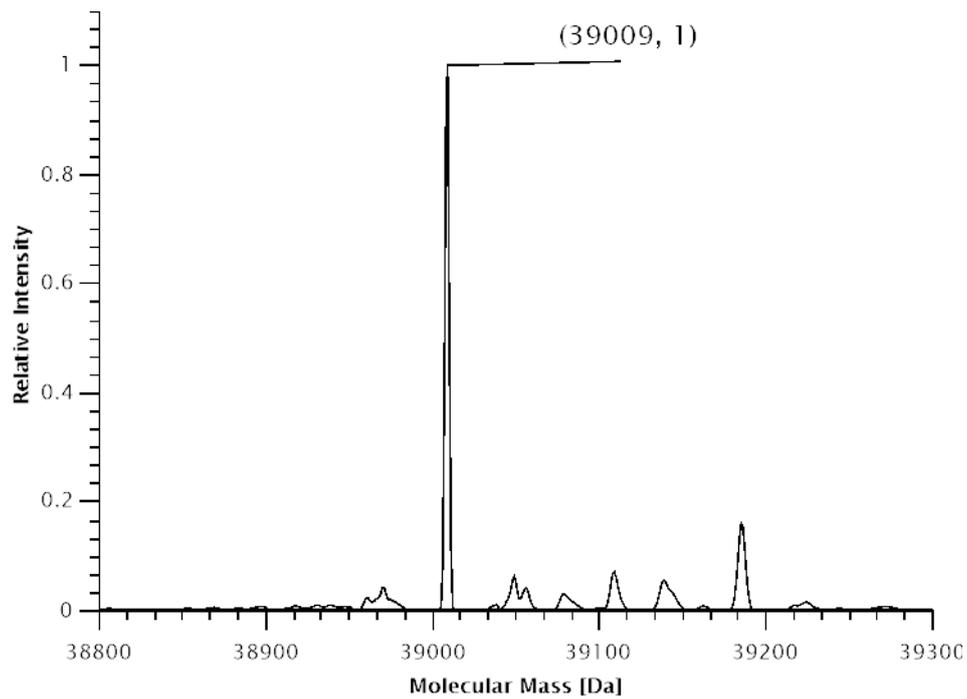


Figure 186: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(**27**))-His₆ production in *E. coli* B95.ΔA with co-expression of *SmbP-MbSacRS(S382T)*. Expected protein mass: 39008.9 Da. Observed mass: 39009 Da.

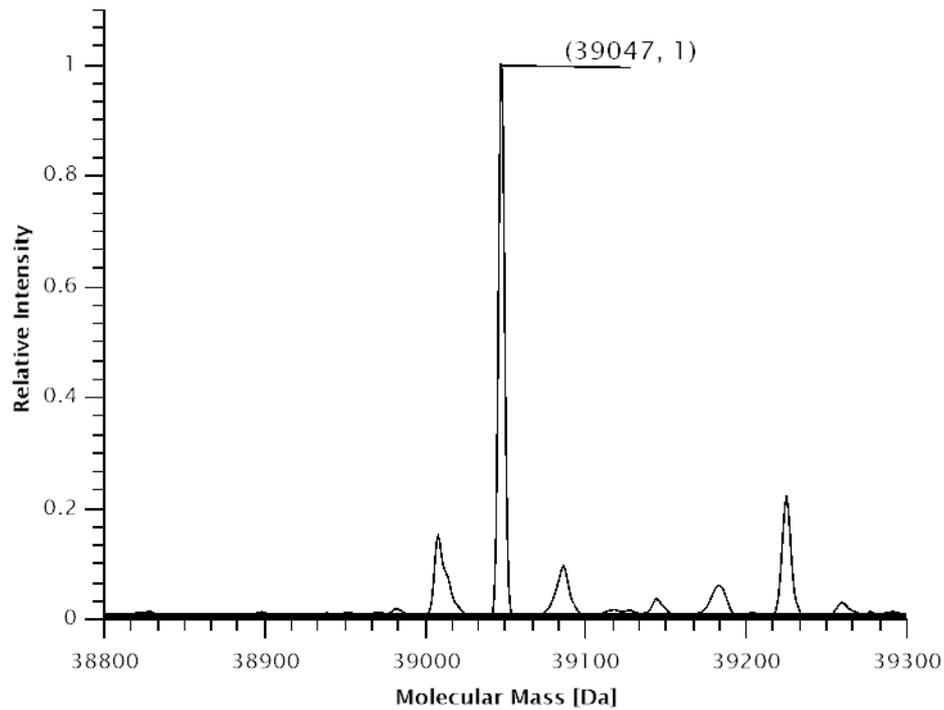


Figure 187: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(3x(27))-His₆ production in *E. coli* B95.ΔA with co-expression of SmbP-*MbSacRS*(S382T). Expected protein mass: 39049.0 Da. Observed mass: 39047 Da.

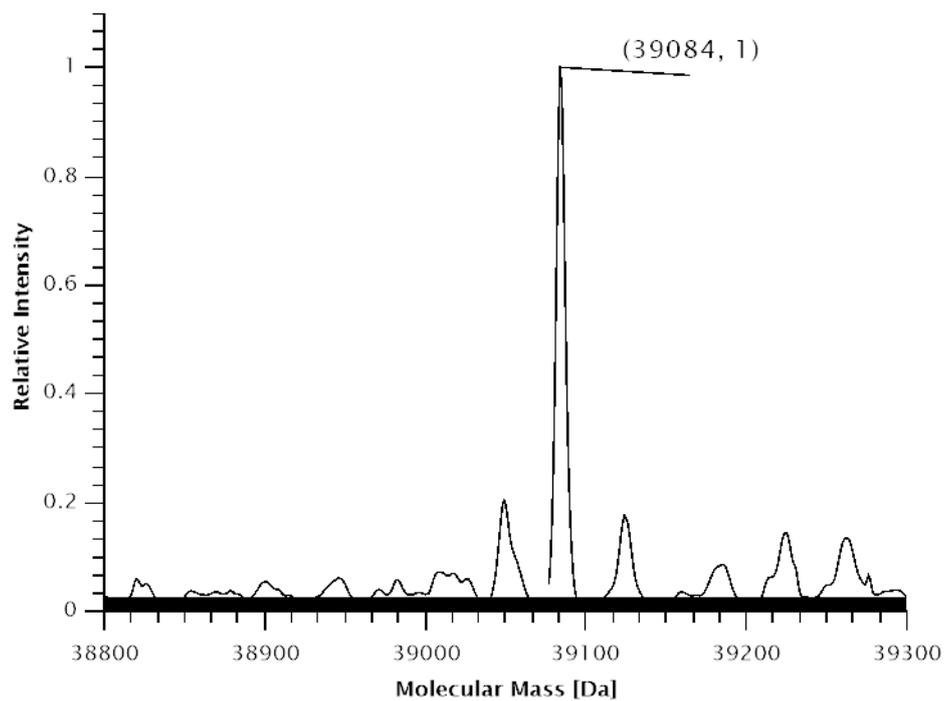


Figure 188: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(5x(27))-His₆ production in *E. coli* B95.ΔA with co-expression of SmbP-*MbSacRS*(S382T). Expected protein mass: 39087.2 Da. Observed mass: 39084 Da.

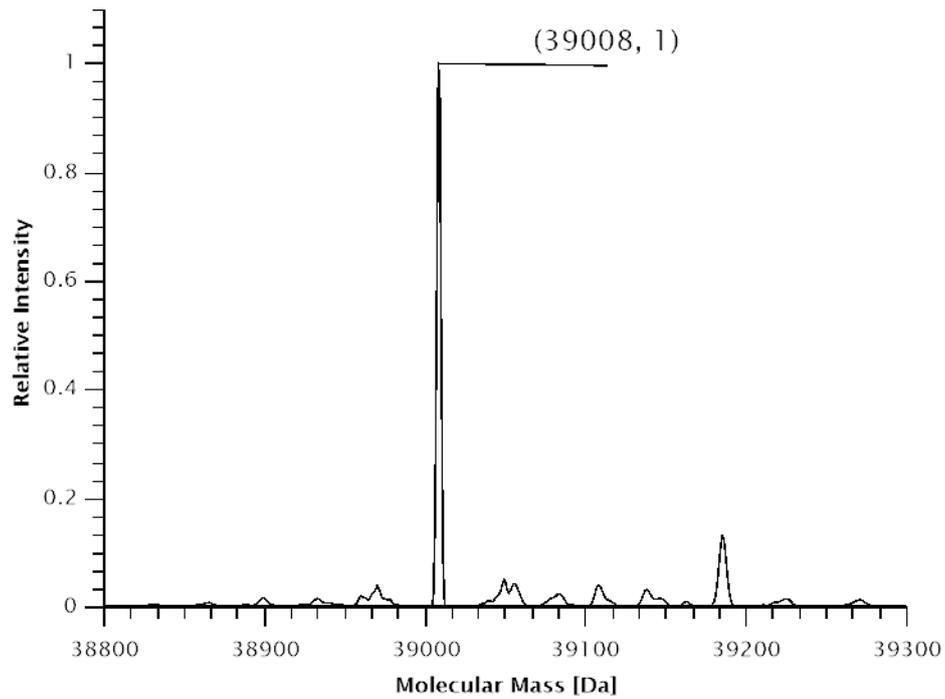


Figure 189: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(27))-His₆ production in *E. coli* B95.ΔA with co-expression of *MburSacRS*(S379T). Expected protein mass: 39008.9 Da. Observed mass: 39008 Da.

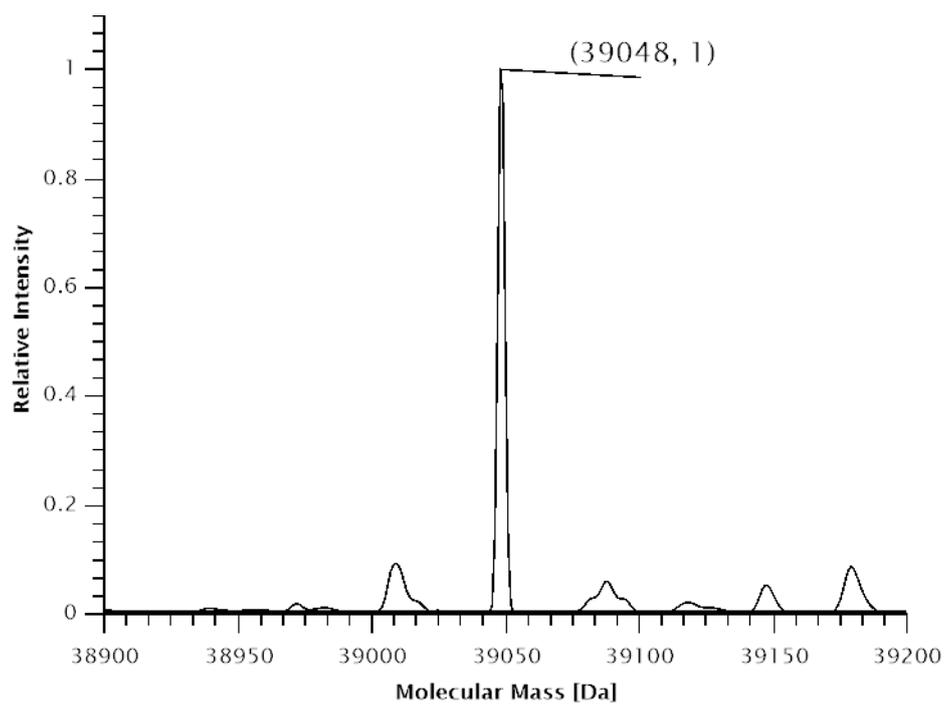


Figure 190: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(3x(27))-His₆ production in *E. coli* B95.ΔA with co-expression of *MburSacRS*(S379T). Expected protein mass: 39049.0 Da. Observed mass: 39048 Da.

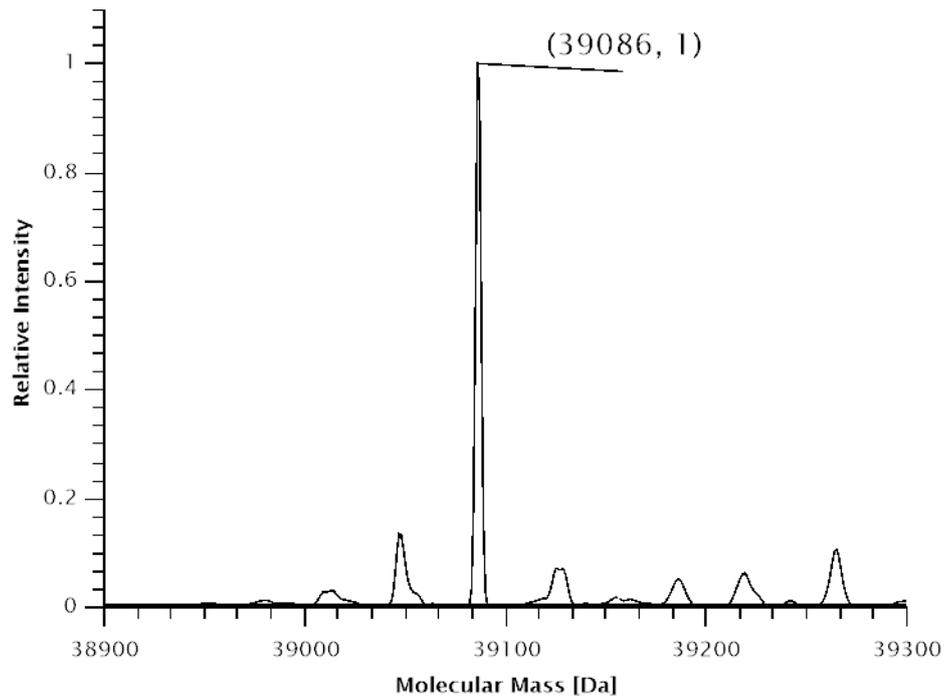


Figure 191: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(5x(27))-His₆ production in *E. coli* B95.ΔA with co-expression of *MburSacRS*(S379T). Expected protein mass: 39087.2 Da. Observed mass: 39086 Da.

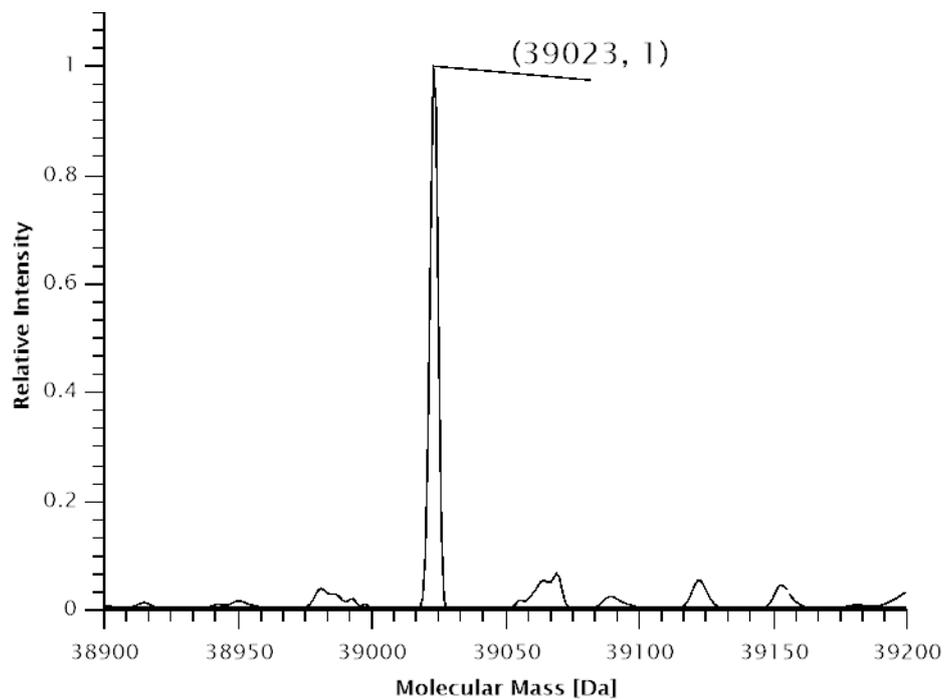


Figure 192: Deconvoluted ESI-MS spectrum of SUMO-sfGFP-His₆ production in *E. coli* B95.ΔA with co-expression of *SmbP-MbSacRS*(S382T). Expected protein mass: 39023.9 Da. Observed mass: 39023 Da.

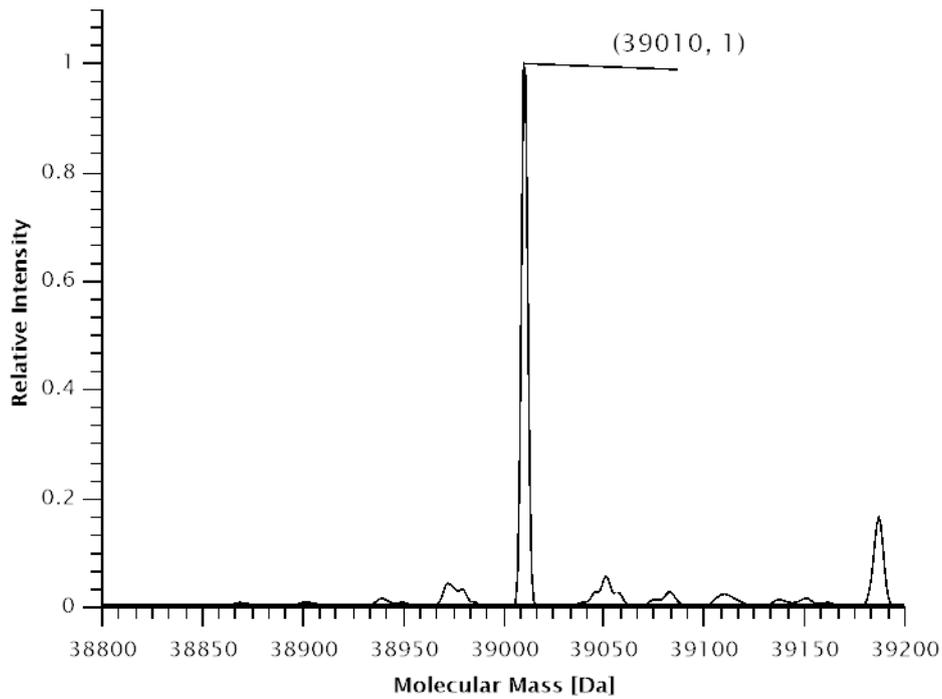


Figure 193: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(1))-His₆ production in *E. coli* B95.ΔA with co-expression of SmbP-*MbSacRS*(S382T). Expected protein mass: 39010.9 Da. Observed mass: 39010 Da.

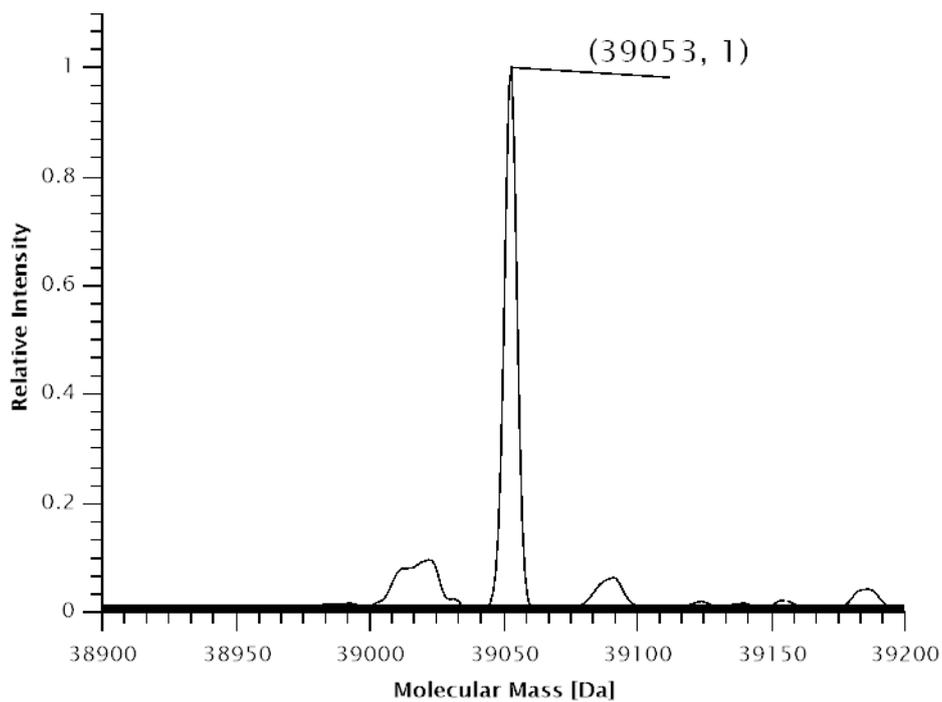


Figure 194: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(3x(1))-His₆ production in *E. coli* B95.ΔA with co-expression of SmbP-*MbSacRS*(S382T). Expected protein mass: 39055.2 Da. Observed mass: 39053 Da.

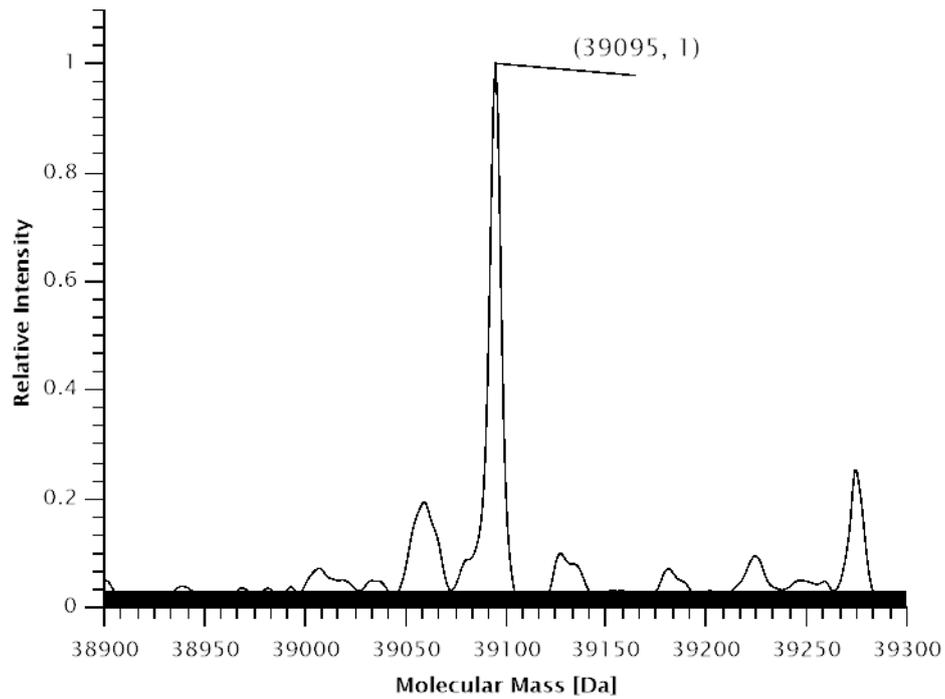


Figure 195: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(5x(1))-His₆ production in *E. coli* B95.ΔA with co-expression of SmbP-MbSacRS(S382T). Expected protein mass: 39097.5 Da. Observed mass: 39095 Da.

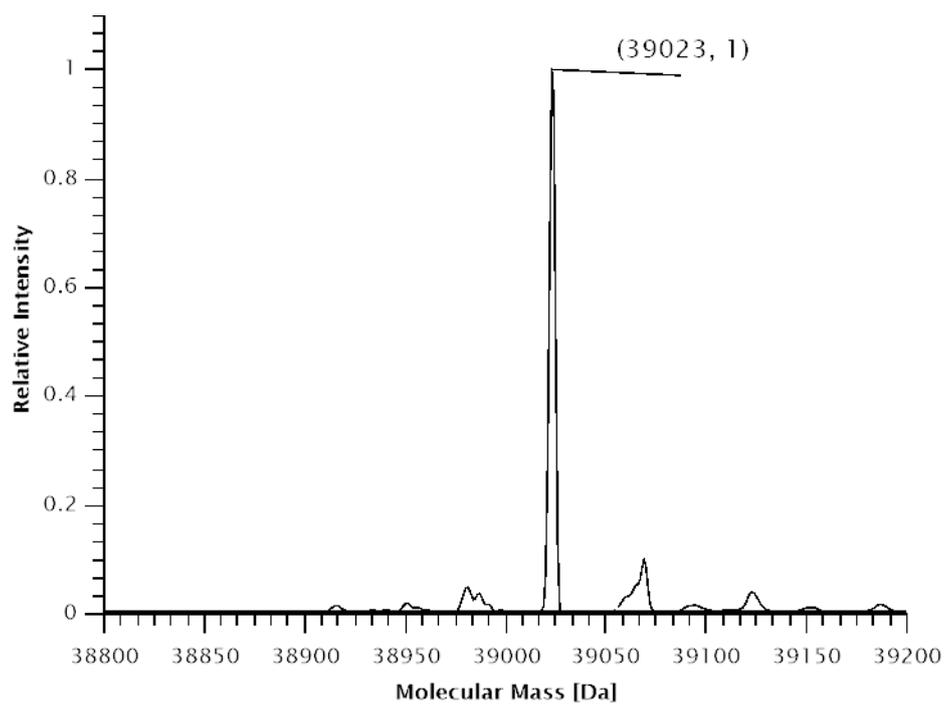


Figure 196: Deconvoluted ESI-MS spectrum of SUMO-sfGFP-His₆ production in *E. coli* B95.ΔA with co-expression of MburSacRS(S379T). Expected protein mass: 39023.9 Da. Observed mass: 39023 Da.

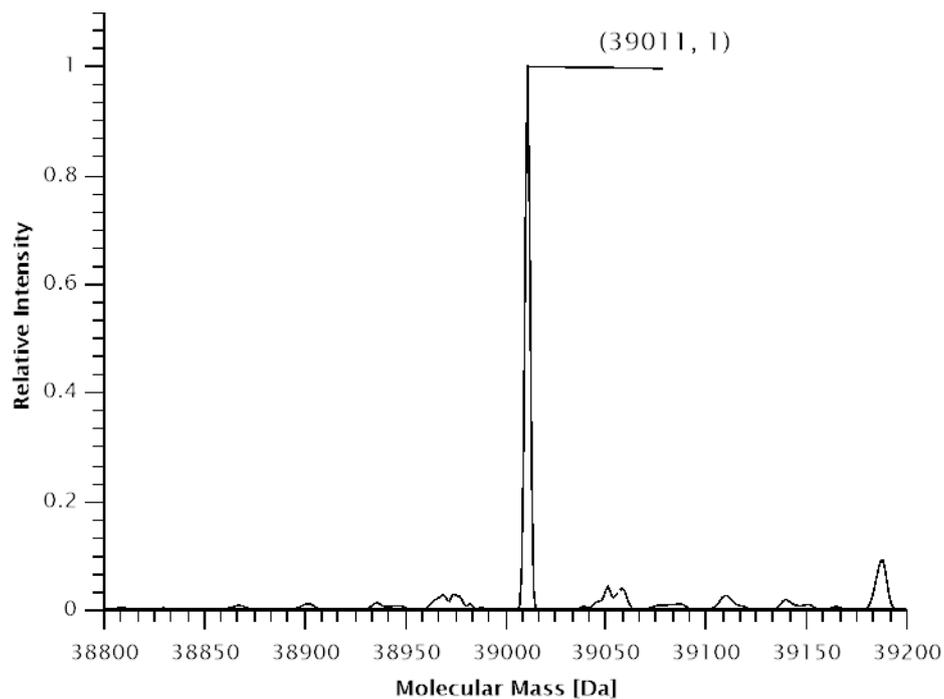


Figure 197: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(R2(1))-His₆ production in *E. coli* B95.ΔA with co-expression of *MburSacRS*(S379T). Expected protein mass: 39010.9 Da. Observed mass: 39011 Da.

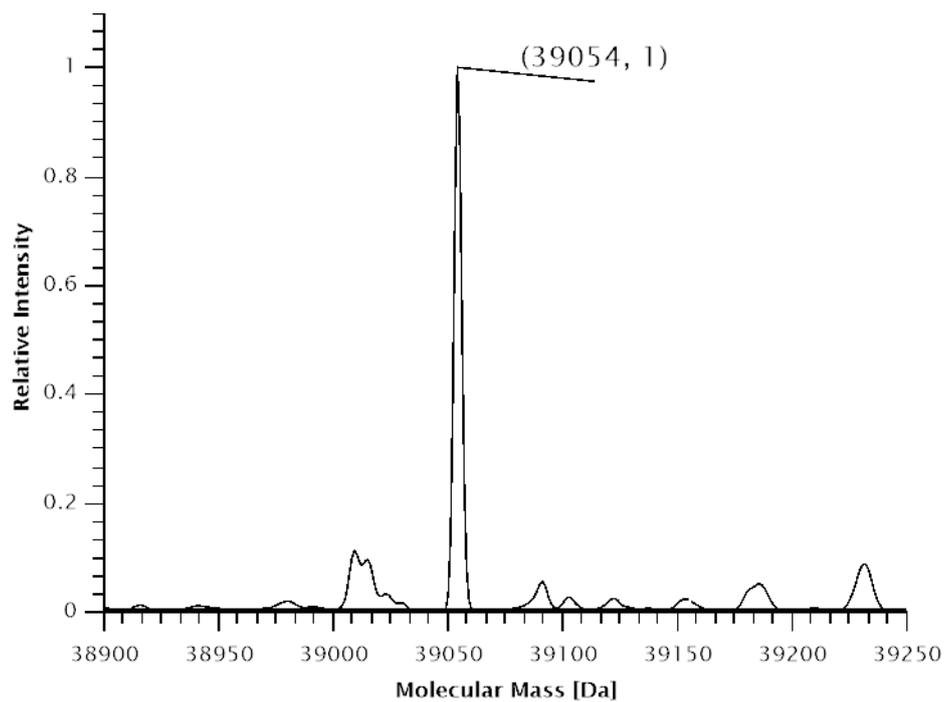


Figure 198: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(3x(1))-His₆ production in *E. coli* B95.ΔA with co-expression of *MburSacRS*(S379T). Expected protein mass: 39055.2 Da. Observed mass: 39054 Da.

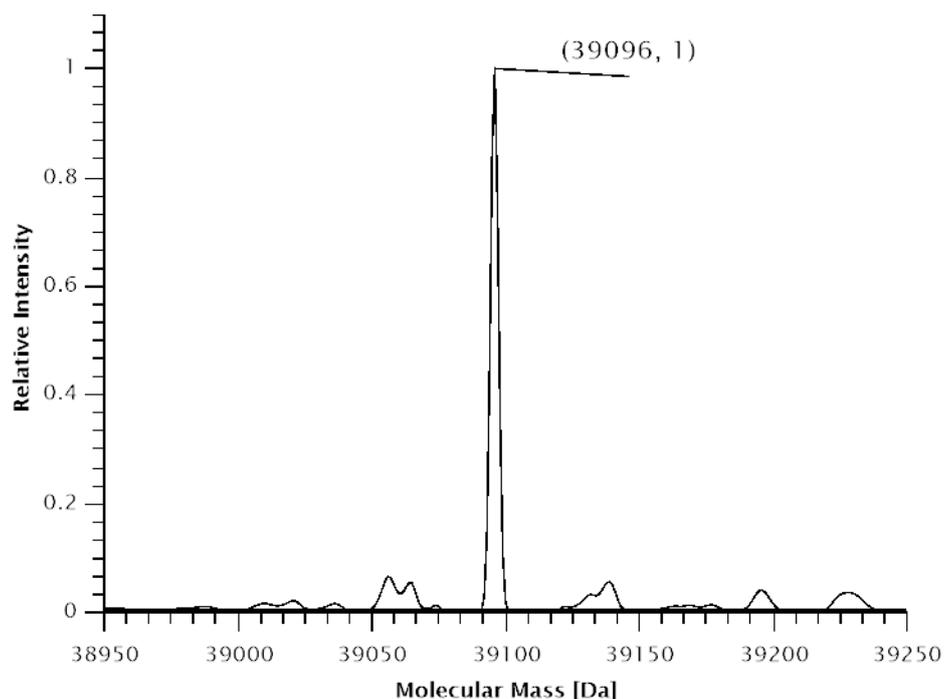


Figure 199: Deconvoluted ESI-MS spectrum of SUMO-sfGFP(5x(1))-His₆ production in *E. coli* B95.ΔA with co-expression of *MburSacRS*(S379T). Expected protein mass: 39097.5 Da. Observed mass: 39096 Da.

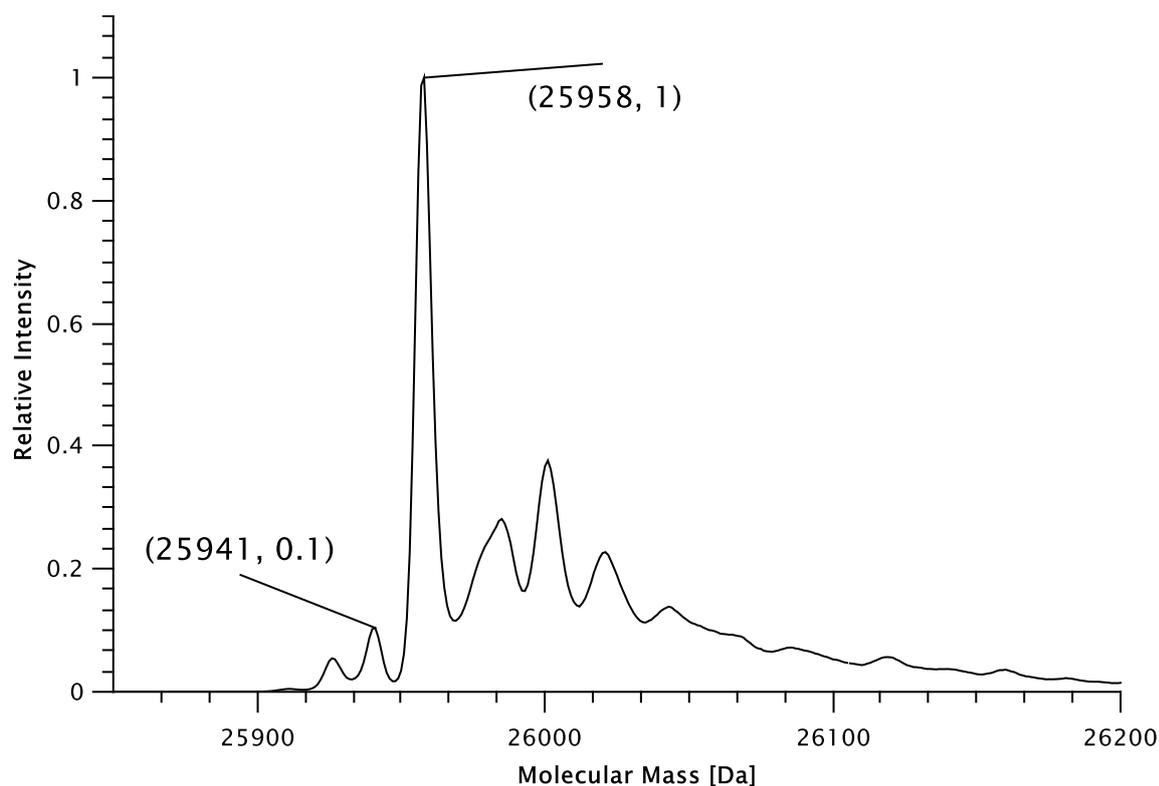


Figure 200: Deconvoluted ESI-MS spectrum of amilCP(Y66(ONBY))-His₆. Protein production in *E. coli* B95.ΔA with co-expression of *MjONB-DopaRS*. Expected protein mass with non-matured chromophore and unreduced nitro group : 25957.9 Da. Observed mass: 25958 Da. The 25941 Da peak is most probably matured amilCP because of background suppression with Tyr.

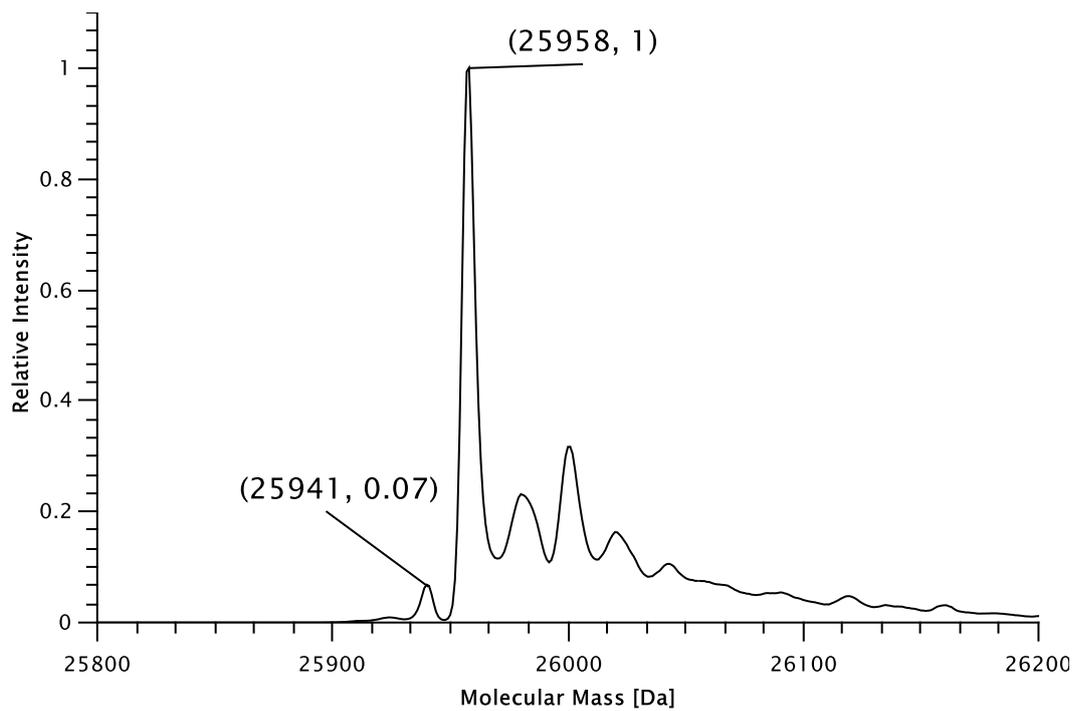


Figure 201: Deconvoluted ESI-MS spectrum of amilCP(Y66(ONBY)-His₆. Protein production in *E. coli* B95.ΔA NK53 with co-expression of *Mj*ONB-DopaRS. Expected protein mass with non-matured chromophore and unreduced nitro group : 25957.9 Da. Observed mass: 25958 Da. The 25941 Da peak is most probably matured amilCP because of background suppression with Tyr.

6.1.6. Plasmids Vector Maps

Color scheme of genetic elements:

Part color	Genetic element
light blue	origin of replication
green	Shine-Dalgarno sequence (RBS)
light green	promoter
red	terminator
magenta	tRNA
yellow	protein coding sequence

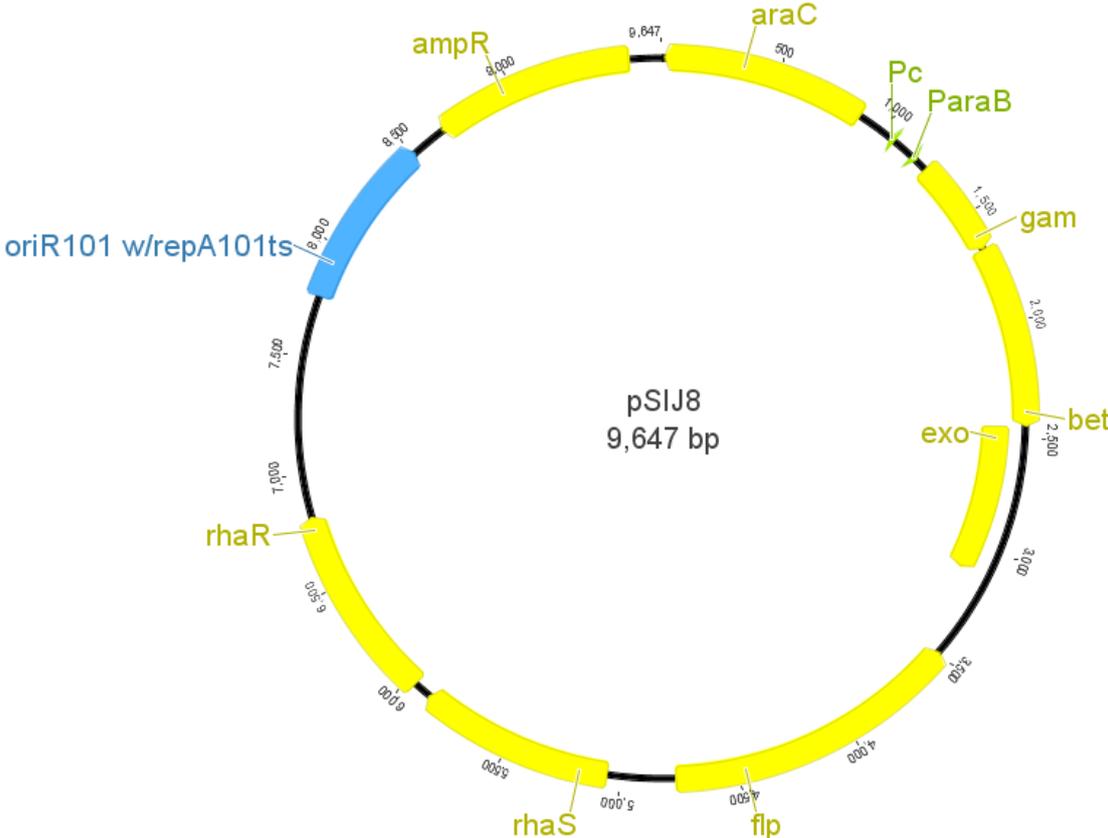


Figure 202: pSIJ8 gene deletion vector.

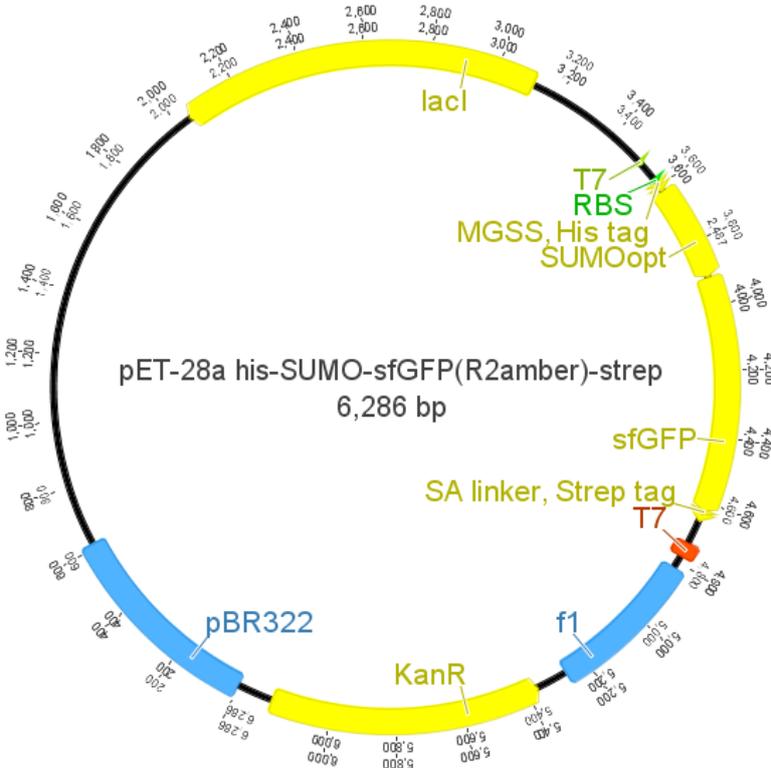


Figure 203: pET-28a target protein expression vector.

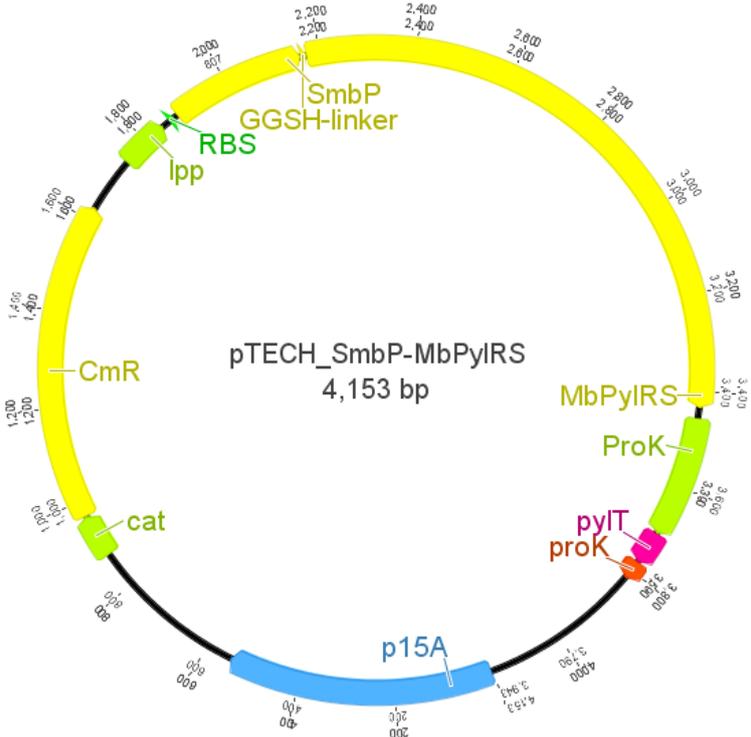


Figure 204: pTECH orthogonal translation system setup.

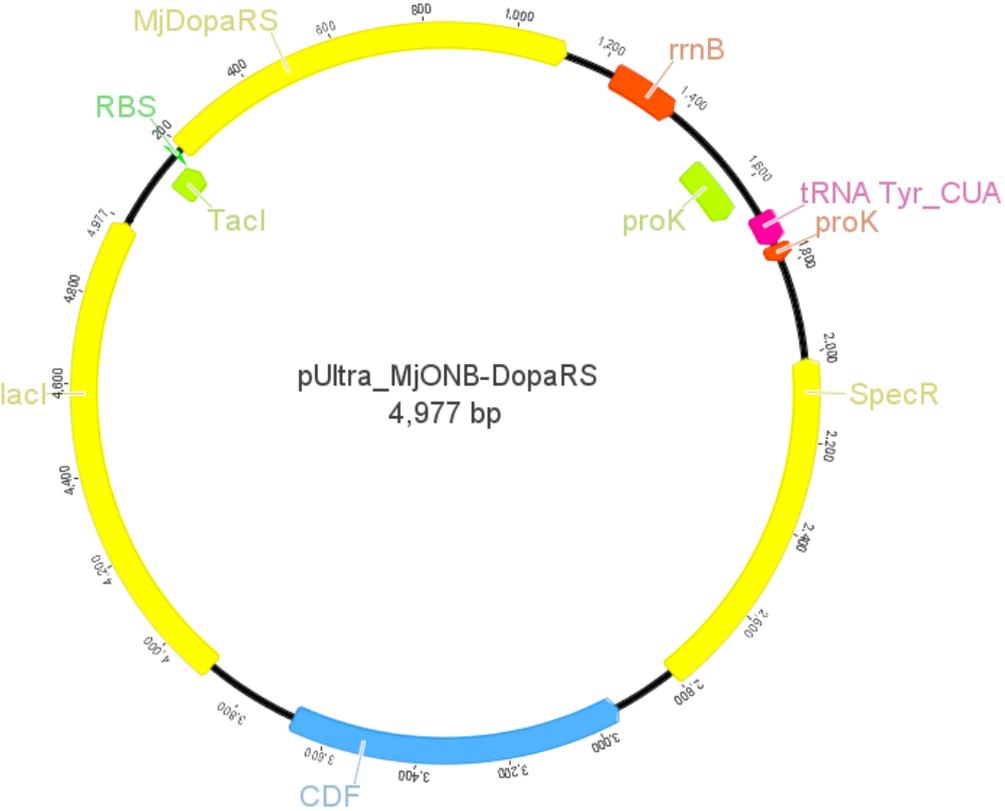


Figure 205: pUltra orthogonal translation system setup.

6.1.7. Nucleotide Sequences

6.1.7.1. tRNAs

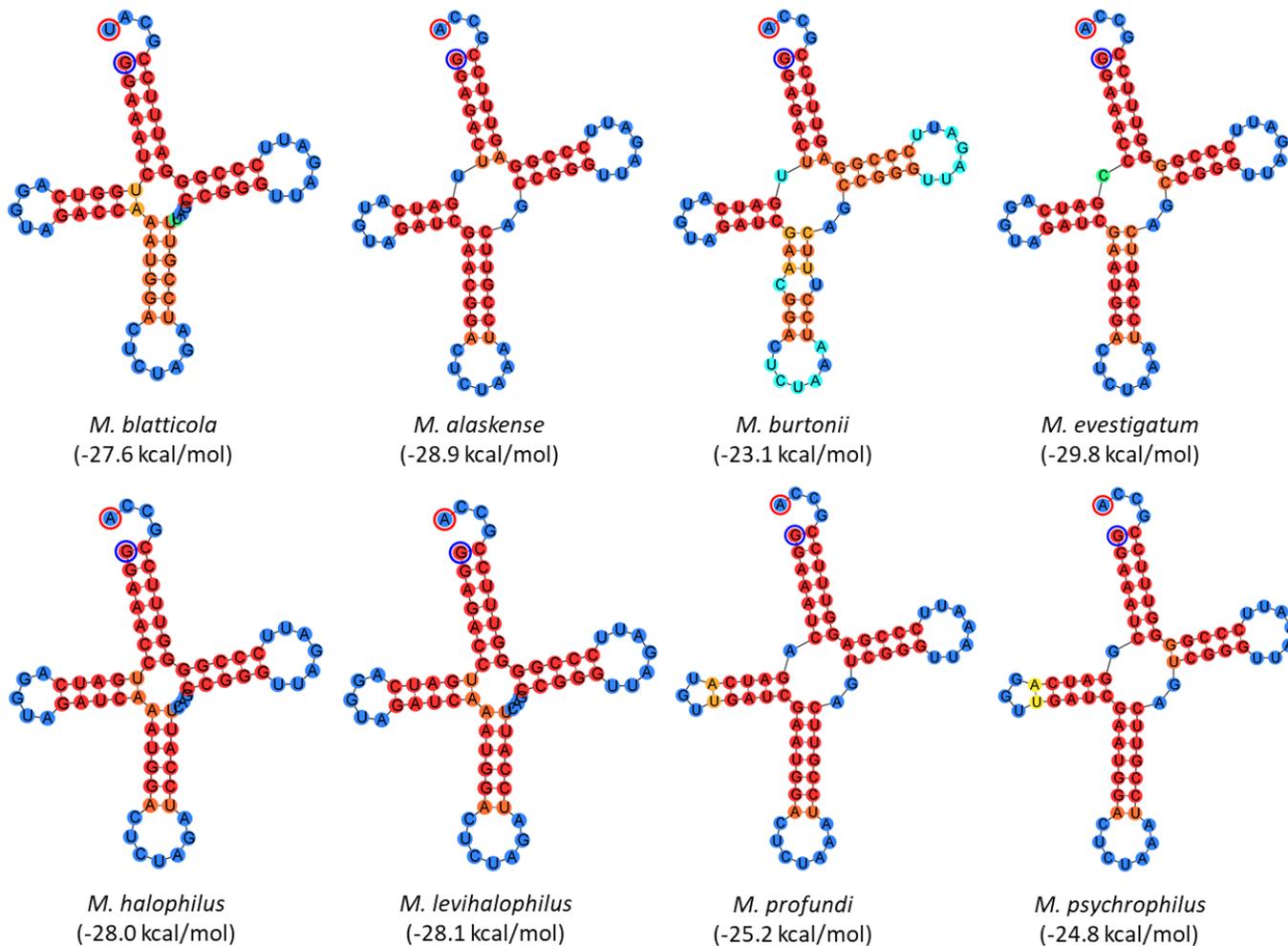


Figure 206: Unique tRNA^{Pyl} sequences. ΔG values in brackets, calculated at 37 °C. The nucleotide binding probability is indicated by color; red = high, green = middle and blue = low. The fold and free energy prediction was performed with geneious which uses the ViennaRNA Package.³¹⁵ The nucleotide with blue a circle is the 5' end and with a red circle the 3' end.

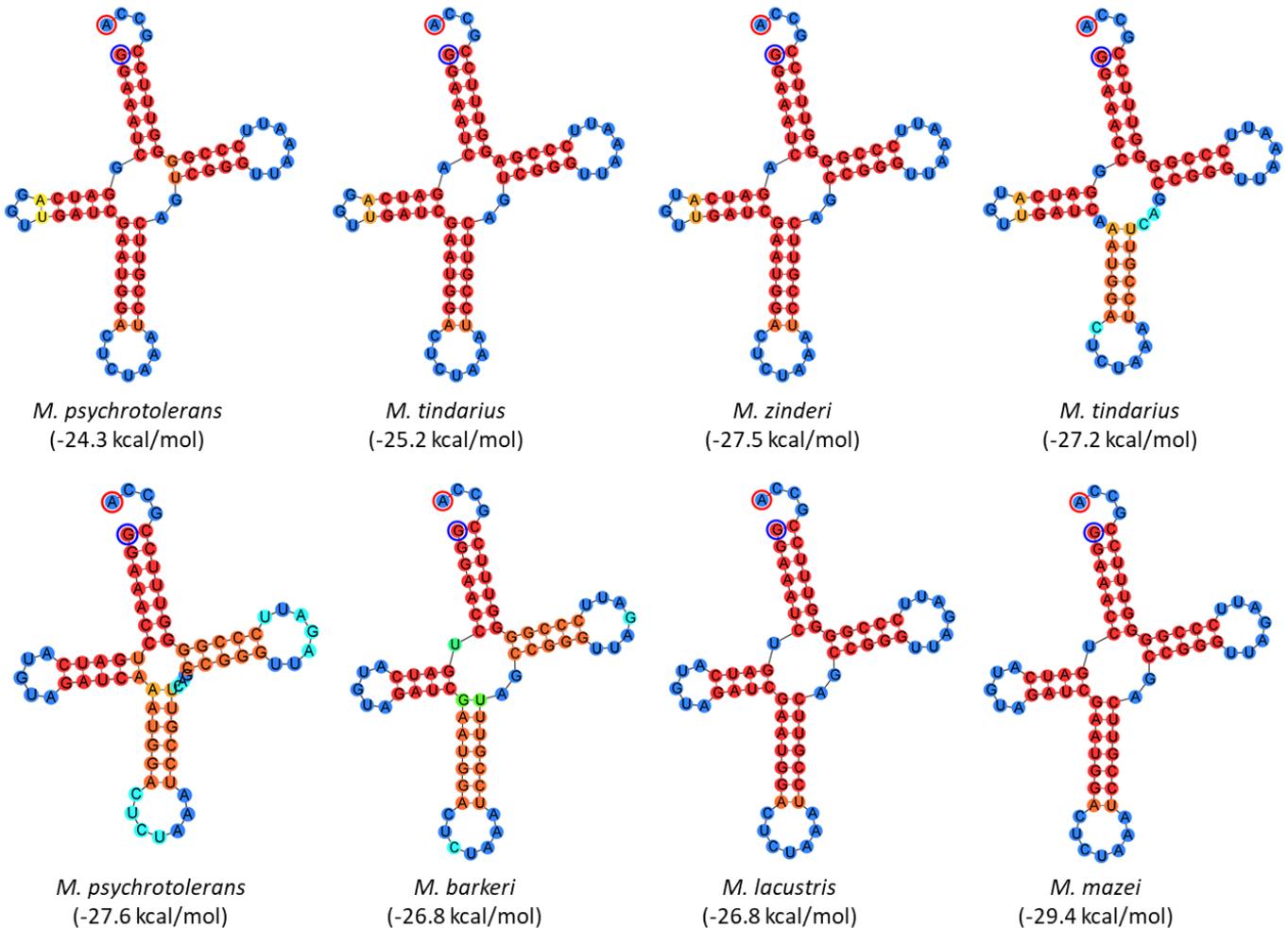


Figure 207: Unique tRNA^{Pyl} sequences. ΔG values in brackets, calculated at 37 °C. The nucleotide binding probability is indicated by color; red = high, green = middle and blue = low. The fold and free energy prediction was performed with geneious which uses the ViennaRNA Package.³¹⁵ The nucleotide with blue a circle is the 5' end and with a red circle the 3' end.

6.1.7.2. Solubility Tags

Containing the N-terminal ATG start codon, all tags were N-terminally fused to the indicated PylRS sequences. Tags 4-9 contain a GGSH-linker between solubility-tag and *MbSacRS*.

GGSH-linker:

GGCGGCTCTCAT

1) InfB(1-21)-tag

ATGACAGATGTAACGATTAAA

2) 10xD-tag

ATGGATGATGACGACGATGACGATGATGATGAC

3) 10xR-tag

ATGAGACGTCGTAGACGTCGCCGTCGTCGTCGT

4) GB1-tag

ATGCAATACAAACTCATTCTCAATGGCAAGACGCTCAAGGGTGAAACCACGACCGAAGC
GGTGAATGCGGCCACCGCGGAGAAAGTGTTCAAACAGTATGCGAACGACAATGGCGTC
AATGGCGAATGGACCTATGACGATGCGACCAAACCTTCACCGTAACCGAA

5) Fh8-tag

ATGCCGAGCGTTCAAGAAGTTGAAAACTGCTGCATGTTCTGGATCGTAATGGTGATGG
TAAAGTTAGCGCAGAAGAAGTAAAGCATTGCGGATGATAGCAAATGTCCGCTGGATA
GCAATAAAATCAAGGCCTTTATCAAAGAGCACGATAAAAACAAAGATGGCAAGCTGGAT
CTGAAAGAAGTGGTTAGCATTCTGAGCAGC

6) SmbP-tag

ATGAGCGGTCATACCGCACATGTTGATGAAGCAGTTAAACATGCCGAAGAAGCAGTTGC
ACACGGTAAAGAAGGCCATACCGATCAGCTGCTGGAACATGCAAAGAAAGTCTGACC
CATGCCAAAGCAGCCAGCGAAGCCGGTGGTAATACCCATGTTGGTCATGGTATTAACA
TCTGGAAGATGCCATCAAACATGGTGAAGAGGGTCATGTTGGTGTGCGACCAAACAG
CACAAGAAGCAATTGAACATCTGCGTGCAAGCGAACATAAAAGCCAT

7) Trx-tag

ATGAGCGATAAAATTATTCACCTGACTGACGACAGTTTTGACACGGATGTAICTCAAAGC
GGACGGGGCGATCCTCGTCGATTTCTGGGCAGAGTGGTGCGGTCCGTGCAAATGATC
GCCCCGATTCTGGATGAAATCGCTGACGAATATCAGGGCAAACCTGACCGTTGCAAACCT
GAACATCGATCAAACCCTGGCACTGCGCCGAAATATGGCATCCGTGGTATCCCGACTC
TGCTGCTGTTCAAACCGGTGAAGTGGCGGCAACCAAAGTGGGTGCACTGTCTAAAGG
TCAGTTGAAAGAGTTCCTCGACGCTAACCTGGCG

8) SUMO-tag

ATGGGTTCTGACTCCGAAGTCAATCAAGAAGCTAAGCCAGAGGTCAAGCCAGAAGTCAA
GCCTGAGACTCACATCAATTTAAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCAAGAT
CAAAAAGACCACTCCTTTAAGAAGGCTGATGGAAGCGTTCGCTAAAAGACAGGGTAAGG
AAATGGACTCCTTAAGATTCTTGTACGACGGTATTAGAATCCAAGCTGATCAGACCCCTG
AAGATTTGGACATGGAGGATAACGATATTATTGAGGCTCACCGCGAACAGATT

9) NusA-tag

ATGAACAAAGAAATTTTGGCTGTAGTTGAAGCCGTATCCAATGAAAAGGCGCTACCTCG
CGAGAAGATTTTCGAAGCATTGGAAAGCGCGCTGGCGACAGCAACAAAGAAAAATATG
ACAAGAGATCGACGTCCGCGTACAGATCGATCGCAAAGCGGTGATTTTGACACTTTC
CGTCGCTGGTTAGTTGTTGATGAAGTCACCCAGCCGACCAAGGAAATCACCTTGAAGC

CGCACGTTATGAAGATGAAAGCCTGAACCTGGGCGATTACGTTGAAGATCAGATTGAGT
 CTGTTACCTTTGACCGTATCACTACCCAGACGGCAAACAGGTTATCGTGCAGAAAGTG
 CGTGAAGCCGAACGTGCGATGGTGGTTGATCAGTTCGCGTGAACACGAAGGTGAAATCA
 TCACCGGCGTGGTGAAAAAAGTAAACCGCGACAACATCTCTCTGGATCTGGGCAACAAC
 GCTGAAGCCGTGATCCTGCGCGAAGATATGCTGCCGCGTGAAAACCTCCGCCCTGGCG
 ACCGCGTTCGTGGCGTGCTCTATTCCGTTCCGCCGGAAGCGCGTGGCGCGCAACTGTT
 CGTCACTCGTTCOAAGCCGGAATGCTGATCGAACTGTTCCGTATTGAAGTGCCAGAAA
 TCGGCGAAGAAGTGATTGAAATTAAGCAGCGGCTCGCGATCCGGGTTCTCGTGCGAA
 AATCGCGGTGAAAACCAACGATAAACGTATCGATCCGGTAGGTGCTTGCGTAGGTATGC
 GTGGCGCGCGTGTTCAGGCGGTGTCTACTGAACTGGGTGGCGAGCGTATCGATATCGT
 CCTGTGGGATGATAACCCGGCGCAGTTCGTGATTAACGCAATGGCACCCGGCAGACGTT
 GCTTCTATCGTGGTGGATGAAGATAAACACACCATGGATATCGCCGTTGAAGCCGGTAA
 CCTGGCGCAGGCGATTGGCCGTAACGGTCAGAACGTGCGTCTGGCTTCGCAGCTGAG
 CGGTTGGGAACTCAACGTGATGACCGTTGACGACCTGCAGGCTAAGCATCAGGCGGAA
 GCGCACGCAGCGATCGACACCTTCACCAAATATCTCGACATCGACGAAGACTTCGCGA
 CTGTTCTGGTAGAAGAAGGCTTCTCGACGCTGGAAGAATTGGCCTATGTGCCGATGAAA
 GAGCTGTTGAAATCGAAGGCCTTGATGAGCCGACCGTTGAAGCACTGCGCGAGCGTG
 CTA AAAATGCACTGGCCACCATTGCACAGGCCAGGAAGAAAGCCTCGGTGATAACAAA
 CCGGCTGACGATCTGCTGAACCTTGAAGGGGTAGATCGTGATTTGGCATTCAA ACTGGC
 CGCCCGTGGCGTTTTGTACGCTGGAAGATCTCGCCGAACAGGGCATTGATGATCTGGCT
 GATATCGAAGGGTTGACCGACGAAAAAGCCGGAGCACTGATTATGGCTGCCCGTAATAT
 TTGCTGGTTCGGTGACGAAGCG

6.1.7.3. *PyIRS used in this Study*

The following section contains DNA sequences used in this study. For clarity, important regions of combined constructs are highlighted by color.

1) *MbPyIRS(T13I:I36V:Y349F)*

ATGGACAAAAACCGCTGGACGTTCTGATTAGCGCAATTGGTCTGTGGATGAGCCGTAC
 CGGCACCCTGCATAAAATCAAACATCATGAAGTTAGCCGCAGCAAAGTCTATATTGAAAT
 GGCATGTGGTGTATCATCTGGTGGTGAATAATAGCCGTAGCTGTCGTACCGCACGTGCAT
 TTCGTCATCACAAATATCGTAAAACCTGTAAACGTTGCCGTGTTAGCGACGAAGATATTA
 ACAATTTTCTGACCCGTAGCACCGAAAGCAAAAATTCAGTTAAAGTTCGTGTTGTGAGCG
 CTCCGAAAGTTAAAAAGCAATGCCGAAAAGCGTTAGTCGTGCACCGAAACCTCTGGAA
 AATAGCGTTAGCGCAAAGCAAGCACCAATACCAGCCGTAGCGTTCCGAGTCCGGCAA
 AAAGCACCCCGAATAGCAGCGTTCGGCAAGCGCACCGGCACCGAGCCTGACCCGTT
 CACAGCTGGATCGTGTTGAAGCACTGCTGAGCCCTGAAGATAAAATCAGCCTGAATATG
 GCAAAACCGTTTTCGTGAACCTGGAACCGGAACTGGTTACCCGTCGTAAAAATGATTTTCA

GCGTCTGTATACCAACGATCGCGAAGATTATCTGGGTAAACTGGAACGTGATATTACCA
AATTTTTCTGGATCGCGGTTTTCTGGAAATCAAAGCCCGATTCTGATTCCGGCAGAAT
ATGTTGAACGTATGGGCATTAATAACGATACCGAACTGAGCAAACAAATCTTCCGCGTT
GATAAAAATCTGTGTCTGCGTCCGATGCTGGCACCGACCCTGTATAACTATCTGCGCAA
ACTGGATCGTATTCTGCCTGGTCCGATTAATACTTTGAAGTTGGTCCGTGCTATCGCAA
AGAAAGTGATGGTAAAGAACACCTGGAAGAGTTTACGATGGTAACTTTTGTGAGATGG
GTAGCGGTTGTACCCGTGAAAATCTGGAAGCACTGATTAAGAGTTTCTGGACTATCTG
GAAATTGACTTTGAAATTGTTGGCGATAGCTGCATGGTTTTTGGTGATACCCTGGATATT
ATGCATGGTGATCTGGAACCTGAGTAGCGCAGTTGTTGGTCCGGTTAGCCTGGATCGCG
AATGGGGTATTGATAAACCGTGGATTGGTGCAGGTTTTTGGTCTGGAACGTCTGCTGAAA
GTTATGCACGGCTTTAAAAACATTAACCGTGAAGCCGTTCCGAGAGCTATTACAATGGT
ATTAGCACCAACCTGTAA

2) *MmPylRSopt*(Y384F)

ATGGATAAAAACCACTAAACACTCTGATCTCTGCTACTGGTCTGTGGATGAGTCGTACC
GGAACCATTATAAAATCAAACACCACGAGGTTAGCCGTTGAAAATCTATATTGAGATG
GCGTGTGGCGATCATCTGGTTGTGAACAATAGCCGCTCTTCTCGTACAGCACGTGCACT
GCGTCACCACAAATATCGTAAAACCTGTAAACGTTGCCGTGTGTCCGATGAGGATCTGA
ACAAATTCCTGACAAAAGCCAATGAGGACCAAACAAGCGTGAAAGTGAAAGTCGTTAGC
GCTCCTACCCGTACTAAAAAGCAATGCCGAAATCCGTTGCTCGTGCCCCTAAACCACT
GGAAAACACTGAAGCAGCACAGGCACAGCCGTCTGGAAGCAAATTCTCTCCGGCCATT
CCTGTTTCTACCCAGGAGTCCGTTTCTGTTCCAGCAAGTGTGAGCACCAGCATTAGCAG
TATTAGCACCGGTGCCACCGCTAGCGCCCTGGTTAAAGGCAATACCAATCCGATTACAA
GCATGTCTGCCCCGTTCAAGCATCAGCTCCAGCACTGACAAAATCCCAAACCGATCGT
CTGGAGGTTCTGCTGAATCCGAAAGACGAAATCAGCCTGAATCCGGCAAACCGTTTCG
TGAACCTGGAGAGCGAACTGCTGTCACGTCGTAAAAAGACCTGCAACAAATCTATGCCG
AAGAACGTGAGAACTATCTGGGGAAACTGGAACGTGAAATCACCCGCTTTTTCTGGAT
CGTGGCTTTCTGGAGATCAAATCCCGATTCTGATTCCTCTGGAGTATATCGAGCGTAT
GGGCATCGACAATGATACCGAACTGAGCAAACAAATTTCCGTGTGGATAAAAATTCT
GTCTGCGCCCTATGCTGGCACCAATCTGTATAACTATCTGCGCAAACCTGGACCGTGCC
CTGCCTGATCCTATCAAATCTTCGAGATCGGCCCGTGTATCGTAAAGAGTCCGACGG
TAAAGAACATCTGGAGGAGTTTACCATGCTGAACTTTTGCCAAATGGGTTTCAGGTTGTAC
TCGTGAGAACCTGGAAAGCATCATACCGATTTTCTGAACCACCTGGGCATTGACTTCA
AAATTGTGGGCGACAGCTGTATGGTGTGGCGACACCCTGGATGTCATGCACGGCGA
CCTGGAACCTGTCTAGTGCCGTTGTTGGACCAATTCCGCTGGACCGTGAGTGGGGTATC
GACAAACCGTGGATCGGAGCAGGATTCGGTCTGGAACGCCTGCTGAAAGTGAAACACG
ACTTCAAAAACATCAAACGTGCCGCCCGTTCTGAATCGTATTATAACGGGATTTCTACCA
ACCTGTAA

3) *MburPyIRS*(Y346F)

ATGGAAAAACAGCTGCTGGATGTTCTGGTTGAACTGAATGGTGTGGCTGAGCCGTAG
CGGTCTGCTGCATGGTATTCGTAATTTTGAATCACCACCAAGCACATCCATATCGAAAC
CGATTGTGGTGCACGTTTTACCGTTCGTAATAGCCGTAGCAGCCGTAGTGCACGTAGCC
TGCGTCATAACAAATATCGTAAACCGTGTAACGTTGTCGTCCGGCAGATGAGCAGATT
GATCGTTTTGTGAAAAAGACCTTCAAAGAGAAACGTCAGACCGTTAGCGTTTTTAGCAG
CCCGAAAAACATGTTCCCAAAAAACCGAAAGTGGCCGTGATTAAGCTTTAGCATT
GCACCCCGAGTCCGAAAGAAGCAAGCGTTAGCAATAGCATTCCGACACCGAGCATTAG
CGTTGTTAAAGATGAAGTTAAAGTGCCCGAGGTGAAATATACCCCGAGCCAGATTGAAC
GTCTGAAAACCTGATGAGTCCGATGATAAAATTCCGATTCAGGATGAACTGCCGGAA
TTTAAAGTTCTTGAAAAAGAACTGATTCAGCGTCGTGATGACCTGAAAAAATGTAT
GAAGAAGATCGCGAAGATCGTCTGGGTAACTGGAACGTGATATTACCGAATTTTTTGT
GGATCGCGGTTTCCTGGAAATCAAAGCCCGATTATGATCCCGTTTGAATATATTGAAC
GCATGGGCATCGATAAAGATGATCACCTGAATAAGCAGATCTTCGTGTTGATGAAAGC
ATGTGTCTGCGTCCGATGCTGGCACCGTGTCTGTATAACTATCTGCGCAAACCTGGATAA
AGTTCTGCCGGATCCGATTCGCATTTTTGAAATTGGTCCGTGCTATCGTAAAGAAAGTGA
TGGTAGCAGCCATCTGGAAGAGTTTACCATGGTGAATTTTTGTCAGATGGGTAGCGGTT
GTACCCGTGAAAATATGGAAGCACTGATTGATGAATTTCTGGAACATCTGGGCATTGAG
TATGAAATTGAAGCCGATAATTGCATGGTGTGGCGATACCATTGATATTATGCATGGT
GATCTGGAACCTGAGCAGCGCAGTTGTTGGTCCGATTCCGCTGGATCGTGAATGGGGTG
TTAATAAACCGTGGATGGGTGCAGGTTTTGGTCTGGAACGTCTGCTGAAAGTTCGTCAT
AACTATACCAATATTCGTCGTGCAAGCCGTTCCAGAGCTGTATTATAACGGCATTAAATACC
AATCTGTAA

4) *MIPyIRS*(Y443F)

ATGGACAAAAACCGCTGAATACCCTGATTAGCACCACCGAACTGTGGATTAGCCGTAC
CGGTATTCTGCATAAAATTCGTCATCATGAAGTGAGCAAACGCAAGATCTATATCGAAAT
GGAATGCGGTGATCAGATTGTGGTGAATAATAGCCGTAGCTGTTCGTACCGCACGTGCA
CTGCGTCATCACAAATATCGTAAAACCTGTAACATTGCCGTGTGAGCGACGAAGATATT
AATCGTTTTCTGACCAAACCGAGCGCAGATAAAAATCGTGTTAAAGTTATGGTTGTGAGC
GCACCGAAAGTTCGTAAAGTATGATGCCGAAAAGCGTTAGCATTGCACCGAAACCGCTGG
AAACCATGGCACCGGAACAGGCAATGCTGAGCGAAAGCCAGCCGGTTCCGACCGCAC
CGGTTAGTGCACCGGCACCGGCAAGTGTCCGGCACCGACATCAGCCCCTGCACCGG
TTTCAGCACCGACACCGACCGAGCGCTCCGACTCCGGCATCAGCTCCGGCTCCAGTGAG
TGCTCCGACGCCGACCTCAGCACCAGCGCCTGCATCTGCTCCAGCGCCAGCAAGCGT
GCCGACACCGGTGAGTACCCCGACCACCATTAGCGCAAGCGCAATGCCTGCAAGCACC
AGTGCACAAGGTATGGCAAGCAGCAATACCCATCCGGCAGCACCGGTTCCAGGTTAAAG

CCCCTGTGCAGGTAAAAGCGCCAGTTCAGGCAAGCGCTCCGGCACTGACCAAATCACA
GGTTGATCGTCTGGAAGGTCTGCTGAGCCCGAAAGATGAAATTAGCCTGGATAGCGGC
ACCCCGTTTCGTAAACTGGAAAGCGAACTGCTGAGTCGTCTGCTGTAAGATCTGAAACA
AATTTATGCCGAAGAACGCGAACATTATCTGGGCAAACCTGGAACGTGATATTACCAAATT
TTTCGTGGATCGCGGTTTCTGGAAATCAAAGCCCGATTCTGATTCCGATGGAATATAT
TGAACGTATGGGCATCGATAACGATAAAGAACTGAGCAAGCAGATTTTTTCGCGTGGATA
ATAACTTTTGTCTGCGTCCGATGCTGGCACCGAATCTGTATAACTATCTGCGTAAATTAG
ATCGTGCCCTGCCGGATCCGATTAATAATCTTTGAAATTGGTCCGTGCTACCGCAAAGAA
AGTGATGGTAAAGAACACCTGGAAGAGTTTACCATGCTGAATTTTTGTCAGATGGGTAG
CGGTTGTACCCGTGAAAATCTGGAAGCAATTATCAAAGACTTCCTGGATTACCTGGGCA
TTGATTTTGAATCGTTGGTGATAGCTGCATGGTTTATGGTGCAACCCTGGATGTTATGC
ATGGTGATCTGGAACCTGAGTAGCGCAGTTGTTGGTCCGGTTTCTATGGATCGTGATTGG
GGTGTTAATAAACCGTGATTGGTGCAGGTTTTGGCCTGGAACGCCTGCTGAAAGTTAT
GCACGACTTTAAAAACATTAACGTGCAAGCCGTAGCGAGAGCTATTACAATGGTATTA
GTACCAATCTGTAA

5) *MpPylRS*(Y373F)

ATGGAACGTAAACCGCTGGATAGCCTGATTAGCAAAAATGGTCTGTGGGTTAGCCGTAA
TGGTGATCTGCATGGTATTCGTAGCTGTGAAACCAGCCAGAAAAATCTGCGTATTACCAT
GGATTGTGGTGAAGTTACCCAGGTTTCGTAATAGCCGTAGCAGCCGTGCAGCACGTAGC
CTGCGTAACCATAAATATCATAAACCGTGCAAAAAATGCCGTCTGGCAGAAGAACGTAT
CAAAGATTTTTCCAATAAAACCGCACGCAAAGATGAAGTTCGTGTTACCGTTAAAACCGT
TCAGAGCAGCCGTTTTAATAGCGTGAAAAGCGATATTCCGGATGCAAGCATGCTGAGCG
AAAGCCTGCAACCGAGCAACATTATTGTTAAACCGCAGATTCATAATAGCCCGACCGAA
CGTCAGCCGAGCGCAACCGCACAGACCTCAAAAAGCCAGCCGAAAACCTATCAGCCGA
AAGTTACCAAACCGGTTAATCATAGCAGTCAGAAACATGCCAAACCGGAAAAAATGAAT
TTACCCAGACACAGAAAAACCGCATTCTGAGCCTGCTGGCACCGGATGATATGATTAGC
TTTAGCAAAGAAAAACGCAGCTTTGCAGAACTGGAAAGCGTTCTGCTGACCCAGCGTAA
AAAAGATCTGCGTGCAATGTATGAAGATAGCCGTGAAAATATGCTGGGTAAACTGGAAC
GTACCATCACCGATTTTTTTGTGGATATGGGTTTCTGGAAAGTAAAAGCCCGATTCTGA
TTCCGTTTGAATACATGGAACGCATGGGTGTTGGTGAAGATAAAGAACTGAGCCGTGAG
ATCTTTCGTGTGGGTGATAATATGTGTCTGCGTCCGATGTTGGCACCGGGTCTGTATAA
TCATCTGCGCAAATTTGATAATGTTCTGCCGGATCCGATTCGCATTTTTGAAATTGGTCC
GTGCTATCGTAAAGAAAGTGATGGTAATAGCCACCTGGAAGAGTTTACCATGCTGAATTT
TTGTCAGATGGGTAGCCGTTGTACCCGTGAGACCCTGGAAGTCTGATTGGTGATTTTC
TGGATTTCTGGATATCGAGTATGAAATTGTTGCCGATAGCTGCATGGTTTATGGTGATA
CCATTGATGTGATGCACCGTAATATGGAACCTGAGCAGCGCAGTTGTTGGTCCGATTCCG
ATGGATATGGATTGGGGTGTTAACAACCGTGATTGGTGCAGGTTTTGGTCTGGAACG

TCTGCTGAAAGCAAAACACAACCTTTAAAAACATTCGTAGCGTTGCACGTAGCGAGAGCT
ATTACAATGGTATTTGTACCAGCCTGTAA

6) *MtPyIRS*(Y408F)

ATGGACAAAAACCGCTGAATACCCTGATTAGCGCAACCGGTCTGTGGATGAGCCGTAC
CGGTAAACTGCATAAAATTCGTATCATGAAGTGAGCAAACGCAAGATCTATATCGAAAT
GGAATGTGGTGAACGTCTGGTGGTGAATAATAGCCGTAGCTGTCGTGCAGCACGTGCA
CTGCGTCATCACAAATATCGTAAAATTTGCAAACACTGCCGTGTGAGTGATGAGGATCT
GAACAAATTTCTGACCCGTACCAATGAGGATAAAAGCAATGCAAAGTTACCGTTGTTAG
CGCACCGAAAATTCGTAAAGTTATGCCGAAAAGCGTTGCACGTACCCCGAAACCGCTG
GAAAATACCGCACCGGTTTCAGACCCTGCCGAGCGAAAGCCAGCCTGCACCGACCACAC
CGATTAGCGCCAGCACACAGCACCCGGCAAGCACCAGTACCACCGCTCCGGCACCGG
CATCAACCACGGCACCCAGCACCCAGCCTCTACAACCGCACCCAGCAAGCGCCTCAACCAC
CATTAGCACCGCAATGCCTGCAAGCACCTCAGCACAGGGCACCCACCAATTTAACT
ATATTAGCGGTGGTTTTCCGCGTCCGATTCCGGTTCAGGCAAGCGCACCCGGCACTGAC
CAAAGCCAGATTGATCGTCTGCAAGGTCTGCTGAGCCCGAAAGATGAAATTAGCCTGG
ATAGCGGCACCCCGTTTTCGTAAACTGGAAAGCGAACTGCTGAGTCGTTCGTAAAGAT
CTGAAGCAGATTTATGCGGAAGAACGCGAACATTATCTGGGCAAACCTGGAACGTGAAAT
CACCAAATTTCTTTGTGGATCGTGGTTTTCTGGAAATCAAAGCCCGATTCTGATTCCGAT
GGAATATATTGAACGCATGGGCATCGATAACGATAAAGAACTGAGCAAACAAATCTTCC
GCGTGGATAATAACTTTTTGTCTGCGTCCGATGCTGGCACCGAATCTGTATAACTATCTG
CGTAAACTGAATCGTGCCCTGCCGGATCCGATTAATCTTTGAAATTGGTCCGTGCTA
CCGCAAAGAAAGTGATGGTAAAGAACACCTGGAAGAGTTTACCATGCTGAATTTTTGTC
AGATGGGTAGCGGTTGTACCCGTGAAAATCTGGAAGCAATTATCAAAGACTTCCTGGAT
TACCTGGGCATTGATTTTGAATCGTTGGTGATAGCTGCATGGTTTATGGTGATACCCTG
GATGTTATGCATGGTGATCTGGAAGTGAAGTAGCGCAGTTGTTGGTCCGGTGCCGATGG
ATCGTGATTGGGGTATTAACAAACCGTGGATTGGTGCAGGTTTTGGTCTGGAACGCCTG
CTGAAAGTAATGCACAACCTTCAAAAACATTAACGTGCAAGCCGTAGCGAGAGCTATTAT
AACGGTATTAGTACCAACCTGTAA

7) *MtPyIRS*(TM-1)(Y349F)

ATGGACAAAAACCGCTGGATGTTCTGATTAGCGCAACCGGTCTGTGGATGAGCCGCA
CCGGTACACTGCATAAAATCAAACATCATGAAGTGAGCAAAGCGCAAGATCTATATTGAAA
TGGCATGTGGTGTATCGTCTGGTTGTGAATAATAGCCGTAATAGTCGTACCGCACGTGCA
TTTCGTATCACAAATATCGTAAAACCTGCAAACACTGCAAAGTGAGCGACGAAGATATT
AGCAATTTTCTGACCAAACCGGCAGAAAATTCAACCAGCGTTAAAGTGAAAGTTATCAGC
ACCCCGAAAGTGAAAAAGCAATGCCGAAAAGCGTTAGCCGTGCACCGAAACCTCTGG
AAACCAGCGTGAGCGCACAGACCAGCGCAAATAAAAGCGGTAGCGTTAGCGTTTCATGC

AAAAAGCGCACCGAATAGCAGCAGCAGTACCAGCACACCCGGCACCCGGCACTGACCCGT
 AGCCAGCTGGATCGTATTGAAGCACTGCTGAGTCCGGAAGATAAAATCAGCCTGGATG
 CAGCAAACCGTTTCGTGAACTGGAAAGCGAACTGCTGGAACGTCGTAAAGGTGATCTG
 CAACGTATTTATGCATATGAGCGCGAAAACCTATCTGGGTAACTGGAACGTGATATTACC
 AAATTTTTCGTGGATCGCGGTTTCTGGAAATCAAAGCCCGATTCTGATTCCGGCAGA
 ATATGTTGAACGTATGGGTATTGATAGCGATAGCGAACTGAGCAAACAGGTTTTTCGTGT
 GGATAAAAATCTGTGTCTGCGTCCGATGCTGGCACCGAATCTGTATAACTACCTGCGCA
 AACTGGATCGTGTGCTGCCGGATCCGATTAATCTTTGAAATTGGTCCGTGCTACCGC
 AAAGAAAGTGATGGTAAAGAACACCTGGAAGAGTTTACCATGCTGAATTTTTGTCAGATG
 GGTAGCGGTAGTACCCGTGAAAATCTGGAAGCACTGATTCGTGAATTTCTGGATTATCT
 GGGCATCGATTTTGAATCGTTGGTGATAGCTGTATGGTGTATGGTGATACCCTGGATG
 TTATGTATGGCGATCTGGAAGTGAAGTACCGCAGTTGTTGGTCCGGTTCCGCTGGATCGC
 GAATGGGGCATTGATAAACCGTGGATTGGTGCAGGCTTTGGTCTGGAACGCCTGCTGA
 AAGTAATGCATGGCTTTAAAACATTAAACGTGCAGCACGTAGCGAGAGCTATTATAACG
 GTATTAGCACCGCCTGTAA

6.1.7.4. Reporter Constructs

- 1) SUMO-sfGFP(1x amber)-His₆ reporter construct with amber codons at position R2

ATGGGCAGCAGCGACTCCGAAGTCAATCAAGAAGCTAAGCCAGAGGTCAAGCCAGAAG
 TCAAGCCTGAGACTCACATCAATTTAAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCA
 AGATCAAAAAGACCACTCCTCTGCGTCTGCTGATGGAAGCGTTCGCTAAAAGACAGGGT
 AAGGAAATGGACTCCTTAAGATTCTTGTACGACGGTATTAGAATCCAAGCTGATCAGAC
 CCCTGAAGATTTGGACATGGAGGATAACGATATTATTGAGGCTCATCGCGAACAGATTG
 GTGGCATGTAGAAAGGCGAAGAGCTGTTCACTGGTGTGCTCCCTATTCTGGTGGAAC
 GGATGGTGATGTCAACGGTCATAAGTTTTCCGTGCGTGGCGAGGGTGAAGGTGACGCA
 ACTAATGGTAACTGACGCTGAAGTTCATCTGTACTACTGGTAACTGCCGGTACCTTG
 GCCGACTCTGGTAACGACGCTGACTTATGGTGTTCAGTGCTTTGCTCGTTATCCGGACC
 ATATGAAGCAGCATGACTTCTTCAAGTCCGCCATGCCGGAAGGCTATGTGCAGGAACG
 CACGATTTCTTTAAGGATGACGGCACGTACAAAACGCGTGCGGAAGTGAATTTGAAG
 GCGATACCCTGGTAAACCGCATTGAGCTGAAAGGCATTGACTTTAAAGAAGACGGCAAT
 ATCCTGGGCCATAAGCTGGAATACAATTTAACAGCCACAATGTTTACATCACCGCCGAT
 AAACAAAAAATGGCATTAAAGCGAATTTAAAATTCGCCACAACGTGGAGGATGGCAG
 CGTGCAGCTGGCTGATCACTACCAGCAAACACTCCAATCGGTGATGGTCCTGTTCTGC
 TGCCAGACAATCACTATCTGAGCACGAAAGCGTTCTGTCTAAAGATCCGAACGAGAAA
 CGCGATCATATGGTTCTGCTGGAGTTCGTAACCGCAGCGGGCATCACGCATGGTATGG
 ATGAACTGTACAAAAGCGCTCATCATCATCATCACTAA

- 2) SUMO-sfGFP(3x amber)-His₆ reporter construct with amber codons at positions R2, N39 and K101

ATGGGCAGCAGCGACTCCGAAGTCAATCAAGAAGCTAAGCCAGAGGTCAAGCCAGAAG
TCAAGCCTGAGACTCACATCAATTTAAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCA
AGATCAAAAAGACCACTCCTCTGCGTCTGATGGAAGCGTTCGCTAAAAGACAGGGT
AAGGAAATGGACTCCTTAAGATTCTTGTACGACGGTATTAGAATCCAAGCTGATCAGAC
CCCTGAAGATTTGGACATGGAGGATAACGATATTATTGAGGCTCATCGCGAACAGATTG
GTGGCATGTAGAAAGGCGAAGAGCTGTTCACTGGTGTCTCCCTATTCTGGTGGA
ACTGGATGGTGTCAACGGTCATAAGTTTTCCGTGCGTGGCGAGGGTGAAGGTGACGCA
ACTTAGGGTAAACTGACGCTGAAGTTCATCTGTACTACTGGTAAACTGCCGGTACCTTG
GCCGACTCTGGTAACGACGCTGACTTATGGTGTTCAGTGCTTTGCTCGTTATCCGGACC
ATATGAAGCAGCATGACTTCTTCAAGTCCGCCATGCCGGAAGGCTATGTGCAGGAACG
CACGATTTCTTTTAGGATGACGGCACGTACAAAACGCGTGCGGAAGTGAAATTTGAAG
GCGATACCCTGGTAAACCGCATTGAGCTGAAAGGCATTGACTTTAAAGAAGACGGCAAT
ATCCTGGGCCATAAGCTGGAATACAATTTTAAACAGCCACAATGTTTACATCACCGCCGAT
AAACAAAAAATGGCATTAAAGCGAATTTTAAATTCGCCACAACGTGGAGGATGGCAG
CGTGCAGCTGGCTGATCACTACCAGCAAACACTCCAATCGGTGATGGTCCTGTTCTGC
TGCCAGACAATCACTATCTGAGCACGCAAAGCGTTCTGTCTAAAGATCCGAACGAGAAA
CGCGATCATATGGTTCTGCTGGAGTTCGTAACCGCAGCGGGCATCACGCATGGTATGG
ATGAACTGTACAAAAGCGCTCATCATCATCATCACTAA

- 3) SUMO-sfGFP(5x amber)-His₆ reporter construct with amber codons at positions R2, N39, K101, E132 and D190

ATGGGCAGCAGCGACTCCGAAGTCAATCAAGAAGCTAAGCCAGAGGTCAAGCCAGAAG
TCAAGCCTGAGACTCACATCAATTTAAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCA
AGATCAAAAAGACCACTCCTCTGCGTCTGATGGAAGCGTTCGCTAAAAGACAGGGT
AAGGAAATGGACTCCTTAAGATTCTTGTACGACGGTATTAGAATCCAAGCTGATCAGAC
CCCTGAAGATTTGGACATGGAGGATAACGATATTATTGAGGCTCATCGCGAACAGATTG
GTGGCATGTAGAAAGGCGAAGAGCTGTTCACTGGTGTCTCCCTATTCTGGTGGA
ACTGGATGGTGTCAACGGTCATAAGTTTTCCGTGCGTGGCGAGGGTGAAGGTGACGCA
ACTTAGGGTAAACTGACGCTGAAGTTCATCTGTACTACTGGTAAACTGCCGGTACCTTG
GCCGACTCTGGTAACGACGCTGACTTATGGTGTTCAGTGCTTTGCTCGTTATCCGGACC
ATATGAAGCAGCATGACTTCTTCAAGTCCGCCATGCCGGAAGGCTATGTGCAGGAACG
CACGATTTCTTTTAGGATGACGGCACGTACAAAACGCGTGCGGAAGTGAAATTTGAAG
GCGATACCCTGGTAAACCGCATTGAGCTGAAAGGCATTGACTTTAAATAGGACGGCAAT
ATCCTGGGCCATAAGCTGGAATACAATTTTAAACAGCCACAATGTTTACATCACCGCCGAT
AAACAAAAAATGGCATTAAAGCGAATTTTAAATTCGCCACAACGTGGAGGATGGCAG

CGTGCAGCTGGCTGATCACTACCAGCAAACACTCCAATCGGT**TAG**GGTCCTGTTCTGC
 TGCCAGACAATCACTATCTGAGCACGCAAAGCGTTCTGTCTAAAGATCCGAACGAGAAA
 CGCGATCATATGGTTCTGCTGGAGTTCGTAACCGCAGCGGGCATCACGCATGGTATGG
 ATGAACTGTACAAAAGCGCT**CATCATCATCATCACTAA**

4) His₆-SUMO-sfGFP(1x amber)-strep reporter construct with amber codons at position R2

ATGGGCAGCAGC**CATCATCATCATCACGGTTCTGACTCCGAAGTCAATCAAGAAGC**
TAAGCCAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTCACATCAATTTAAAGGTGTCCG
ATGGATCTTCAGAGATCTTCTTCAAGATCAAAAAGACCACTCCTCTGCGTCGTCTGATGG
AAGCGTTCGCTAAAAGACAGGGTAAGGAAATGGACTCCTTAAGATTCTTGTACGACGGT
ATTAGAATCCAAGCTGATCAGACCCCTGAAGATTTGGACATGGAGGATAACGATATTATT
 GAGGCTCATCGCAACAGATTGGTGGCATG**TAG**AAAGGCCGAAGAGCTGTTCACTGGTG
 TCGTCCCTATTCTGGTGGAACTGGATGGTGTGATGTCAACGGTCATAAGTTTTCCGTGCGT
 GGCGAGGGTGAAGGTGACGCAACTAATGGTAAACTGACGCTGAAGTTCATCTGTACTAC
 TGGTAAACTGCCGGTACCTTGGCCGACTCTGGTAACGACGCTGACTTATGGTGTTCAGT
 GCTTTGCTCGTTATCCGGACCATATGAAGCAGCATGACTTCTTCAAGTCCGCCATGCCG
 GAAGGCTATGTGCAGGAACGCACGATTTCTTTAAGGATGACGGCACGTACAAAACGC
 GTGCGGAAGTGAAATTTGAAGGCGATACCCTGGTAAACCGCATTGAGCTGAAAGGCATT
 GACTTTAAAGAAGACGGCAATATCCTGGGCCATAAGCTGGAATACAATTTTAAACAGCCA
 CAATGTTTACATCACCGCCGATAAACAAAAAATGGCATTAAAGCGAATTTTAAAATTCG
 CCACAACGTGGAGGATGGCAGCGTGCAGCTGGCTGATCACTACCAGCAAACACTCCA
 ATCGGTGATGGTCCTGTTCTGCTGCCAGACAATCACTATCTGAGCACGCAAAGCGTTCT
 GTCTAAAGATCCGAACGAGAAACGCGATCATATGGTTCTGCTGGAGTTCGTAACCGCAG
 CGGGCATCACGCATGGTATGGATGAACTGTACAAAAGCGCTTGGAGCCACCCGCAGTT
 CGAAAAATAA

5) His₆-SUMO-sfGFP(3x amber)-strep reporter construct with amber codons at positions R2,
 N39 and K101

ATGGGCAGCAGC**CATCATCATCATCACGGTTCTGACTCCGAAGTCAATCAAGAAGC**
TAAGCCAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTCACATCAATTTAAAGGTGTCCG
ATGGATCTTCAGAGATCTTCTTCAAGATCAAAAAGACCACTCCTCTGCGTCGTCTGATGG
AAGCGTTCGCTAAAAGACAGGGTAAGGAAATGGACTCCTTAAGATTCTTGTACGACGGT
ATTAGAATCCAAGCTGATCAGACCCCTGAAGATTTGGACATGGAGGATAACGATATTATT
 GAGGCTCATCGCAACAGATTGGTGGCATG**TAG**AAAGGCCGAAGAGCTGTTCACTGGTG
 TCGTCCCTATTCTGGTGGAACTGGATGGTGTGATGTCAACGGTCATAAGTTTTCCGTGCGT
 GGCGAGGGTGAAGGTGACGCAACT**TAG**GGTAAACTGACGCTGAAGTTCATCTGTACTA
 CTGGTAAACTGCCGGTACCTTGGCCGACTCTGGTAACGACGCTGACTTATGGTGTTCAG
 TGCTTTGCTCGTTATCCGGACCATATGAAGCAGCATGACTTCTTCAAGTCCGCCATGCC

GGAAGGCTATGTGCAGGAACGCACGATTTCCCTTTTAGGATGACGGCACGTACAAAACG
 CGTGCGGAAGTGAAATTTGAAGGCGATACCCTGGTAAACCGCATTGAGCTGAAAGGCA
 TTGACTTTAAAGAAGACGGCAATATCCTGGGCCATAAGCTGGAATACAATTTTAAACAGCC
 ACAATGTTTACATCACCGCCGATAAACAAAAAATGGCATTAAAGCGAATTTTAAATTC
 GCCACAACGTGGAGGATGGCAGCGTGCAGCTGGCTGATCACTACCAGCAAAACACTCC
 AATCGGTGATGGTCCTGTTCTGCTGCCAGACAATCACTATCTGAGCACGCAAAGCGTTC
 TGTCTAAAGATCCGAACGAGAAACGCGATCATATGGTTCTGCTGGAGTTCGTAACCGCA
 GCGGGCATCACGCATGGTATGGATGAACTGTACAAAAGCGCTTGGAGCCACCCGCAGT
 TCGAAAAATAA

- 6) His₆-SUMO-sfGFP(5x amber)-strep reporter construct with amber codons at positions R2, N39, K101, E132 and D190

ATGGGCAGCAGCCATCATCATCATCACGGTTCTGACTCCGAAGTCAATCAAGAAGC
 TAAGCCAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTCACATCAATTTAAAGGTGTCCG
 ATGGATCTTCAGAGATCTTCTTCAAGATCAAAAAGACCACTCCTCTGCGTCGTCTGATGG
 AAGCGTTCGCTAAAAGACAGGGTAAGGAAATGGACTCCTTAAGATTCTTGACGACGGT
 ATTAGAATCCAAGCTGATCAGACCCCTGAAGATTTGGACATGGAGGATAACGATATTATT
 GAGGCTCATCGCAACAGATTGGTGGCATGTAGAAAGGCGAAGAGCTGTTCACTGGTG
 TCGTCCCTATTCTGGTGGAACTGGATGGTGTGCAACGGTCATAAGTTTTCCGTGCGT
 GGCGAGGGTGAAGGTGACGCAACTTAGGGTAAACTGACGCTGAAGTTCATCTGTACTA
 CTGGTAAACTGCCGGTACCTTGGCCGACTCTGGTAACGACGCTGACTTATGGTGTTCAG
 TGCTTTGCTCGTTATCCGGACCATATGAAGCAGCATGACTTCTTCAAGTCCGCCATGCC
 GGAAGGCTATGTGCAGGAACGCACGATTTCCCTTTTAGGATGACGGCACGTACAAAACG
 CGTGCGGAAGTGAAATTTGAAGGCGATACCCTGGTAAACCGCATTGAGCTGAAAGGCA
 TTGACTTTAAATAGGACGGCAATATCCTGGGCCATAAGCTGGAATACAATTTTAAACAGCC
 ACAATGTTTACATCACCGCCGATAAACAAAAAATGGCATTAAAGCGAATTTTAAATTC
 GCCACAACGTGGAGGATGGCAGCGTGCAGCTGGCTGATCACTACCAGCAAAACACTCC
 AATCGGTAGGGTCCTGTTCTGCTGCCAGACAATCACTATCTGAGCACGCAAAGCGTTC
 TGTCTAAAGATCCGAACGAGAAACGCGATCATATGGTTCTGCTGGAGTTCGTAACCGCA
 GCGGGCATCACGCATGGTATGGATGAACTGTACAAAAGCGCTTGGAGCCACCCGCAGT
 TCGAAAAATAA

- 7) 6-R11-1-amilCP-His₆

ATGGGTATCATTAGAAGAATGGTGGTGTAGTGTGATCGCTAAACAAATGACCTACAA
 GGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCGAAGGCGATGGAAAAG
 GTAAGCCCTACGAGGGGGAGCAGACGGTAAAGCTCACTGTCACCAAGGGCGGACCTCT
 GCCATTTGCTTGGGATATTTTATCACCACAGTGTGACGACGGAAGCATAACCATTCACCAA
 GTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTCCCGGAGGGCTATACATGGG

AGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATGATTCCAGCATC
 CAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGAACCTTCTCCCAATGGA
 CCTGTCATGCAGAAGAAGACACAGGGCTGGGAACCCAACACTGAGCGTCTCTTTGCAC
 GAGATGGAATGCTGCTAGGAAACAACCTTTATGGCTCTGAAGTTAGAAGGAGGCGGTAC
 TATTTGTGTGAATTTAAACTACTTACAAGGCAAAGAAGCCTGTGAAGATGCCAGGGTAT
 CACTATGTTGACCGCAAACCTGGATGTAACCAATCACAACAAGGATTACACTTCGGTTGA
 GCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCCGGCAGCCATCATCATCATC
 ATCACTAA

6-R11-1 contains an amber codon and flanking glycine residues.¹⁵¹ Parts of the amilCP coding sequence were derived from the iGEM Registry of Standard Biological Parts (http://parts.igem.org/Part:BBa_K592009).

8) PDZ(F325amber)-His₆

ATGGGTGGTGAAGAAGATATTCCTCGCGAACCGCGTTCGATTGTTATTCATCGTGGTAG
 CACCGGTCTGGGTTAGAAATATCATTGGTGGTGAAGATGGCGAAGGCATTTTTATTAGCT
 TTATTCTGGCAGGCGGTCCGGCAGATCTGAGCGGTGAACTGCGTAAAGGTGATCAGAT
 TCTGAGCGTTAATGGTGTGATCTGCGTAATGCAAGCCATGAACAGGCAGCAATTGCAC
 TGAAAAATGCAGGTCAGACCGTTACCATTATCGCACAGTATAAACCGGAAGAATATAGC
 CGTTTTGAAGCAGGCAGCCATCATCATCATCACTAA

Details of this construct have been published before.¹²⁸

9) ELP(5x amber)-sfGFP-His₆ (as example for all other ELP constructs)

ATGAGCAAAGGTGGGCCGGGCGTGGGTGTCCCGGGTGGCGGTGTTCCGGGTGCAGG
 CGTTCGGGTTAGGGCGTGCCGGGCGGCGGTGTTCCGGGTGCTGGTGTGCCGGGCTA
 GGGTGTCCCGGTGGCGGTGTGCCGGGCGCAGGTGTCCCGGGTTAGGGTGTCCGG
 GCGGCGGTGTCCCGGTGCAGGTGTGCCGGGCTAGGGTGTGCCGGGCGGCGGTGTC
 CCGGTGCAGGTGTGCCGGGCTAGGGTGTGCCGGGCTGGCTGGGCCCGGGCGGTGG
 CGGTCGTAAAGGCGAAGAGCTGTTCACTGGTGTTCGTCCTATTCTGGTGGAACTGGAT
 GGTGATGTCAACGGTCATAAGTTTTCCGTGCGTGGCGAGGGTGAAGGTGACGCAACTA
 ATGGTAACTGACGCTGAAGTTCATCTGTACTACTGGTAACTGCCGGTACCTTGGCCG
 ACTCTGGTAACGACGCTGACTTATGGTGTTCAGTGCTTTGCTCGTTATCCGGACCATAT
 GAAGCAGCATGACTTCTTCAAGTCCGCCATGCCGGAAGGCTATGTGCAGGAACGCACG
 ATTTCTTTAAGGATGACGGCACGTACAAAACGCGTGCGGAAGTGAATTTGAAGGCGA
 TACCCTGGTAAACCGCATTGAGCTGAAAGGCATTGACTTTAAAGAAGACGGCAATATCC
 TGGGCCATAAGCTGGAATACAATTTAACAGCCACAATGTTTACATCACCGCCGATAAAC
 AAAAAATGGCATTAAAGCGAATTTTAAATTCGCCACAACGTGGAGGATGGCAGCGTG
 CAGCTGGCTGATCACTACCAGCAAACACTCCAATCGGTGATGGTCCTGTTCTGCTGCC

AGACAATCACTATCTGAGCACGCAAAGCGTTCTGTCTAAAGATCCGAACGAGAAACGCG
 ATCATATGGTTCTGCTGGAGTTCGTAACCGCAGCGGGCATCACGCATGGTATGGATGAA
 CTGTACAAAGGCAGCCATCATCATCATCAC

The yellow part is the starter sequence from the Meyer and Chilkoti publication which seems to experience good expression strength.³²⁷

6.1.7.5. Genomes used for In Silico Analysis

All sequences can be found in the NCBI or JGI database. The tRNA^{Pyl} was mostly not annotated and was searched for with a homology model of known tRNA^{Pyl} using Geneious.

Table 12: Genomes Used for extraction of PyIRS and tRNA^{Pyl} sequences

Organism	Strain	Accession number
<i>Methanimicrococcus blatticola</i>	DSM 13328	GCF_004363215.1
<i>Methanococcoides alaskense</i>	DSM 17273	SAMN18250245
<i>Methanococcoides burtonii</i>	DSM 6242	GCF_000013725.1
<i>Methanococcoides methylutens</i>	DSM 2657	GCF_000765475.1
<i>Methanococcoides vulcani</i>	SLH 33	GCF_900111645.1
<i>Methanohalobium evestigatum</i>	Z-7303	GCF_000196655.1
<i>Methanohalophilus euhalobius</i>	WG1_MB	SAMN08777283
<i>Methanohalophilus halophilus</i>	Z-7982	GCF_001889405.1
<i>Methanohalophilus levihalophilus</i>	DSM 28452	GCF_017874375.1
<i>Methanohalophilus mahii</i>	DSM 5219	GCF_000025865.1
<i>Methanohalophilus portucalensis</i>	FDF-1T	GCF_002761295.1
<i>Methanohalophilus profundus</i>	SLHTYRO	GCF_004137855.1
<i>Methanobolus bombayensis</i>	DSM 7082	GCF_017873415.1
<i>Methanobolus halotolerans</i>	SY-01	GCF_004745425.1
<i>Methanobolus profundus</i>	Mob M	GCF_900114835.1
<i>Methanobolus psychrophilus</i>	R15	GCF_000306725.1
<i>Methanobolus psychrotolerans</i>	YSF-03	GCF_002243045.1
<i>Methanobolus tindarius</i>	DSM 2278	GCF_000504205.1
<i>Methanobolus vulcani</i>	PL 12/M	GCF_900100715.1
<i>Methanobolus zinderi</i>	DSM 21339	GCF_013388255.1
<i>Methanomethylovorans hollandica</i>	DSM 15978	GCF_000328665.1
<i>Methanosalsum zhilinae</i>	DSM 4017	GCF_000217995.1
<i>Methanosarcina acetivorans</i>	C2A	GCF_000007345.1
<i>Methanosarcina barkeri</i>	227	GCF_000970065.1
<i>Methanosarcina flavescens</i>	E03.2	GCF_001304615.2
<i>Methanosarcina horonobensis</i>	HB-1	GCF_000970285.1

<i>Methanosarcina lacustris</i>	Z-7289	GCF_000970265.1
<i>Methanosarcina mazei</i>	S-6	GCF_000007065.1
<i>Methanosarcina siciliae</i>	T4/M	GCF_000970085.1
<i>Methanosarcina soligelidi</i>	SMA-21	GCF_000744315.1
<i>Methanosarcina spelaei</i>	MC-15	GCF_002287235.1
<i>Methanosarcina thermophila</i>	TM-1	GCF_000969885.1
<i>Methanosarcina vacuolata</i>	Z-761	GCF_000969905.1
<i>Methermicoccus shengliensis</i>	DSM 18856	GCA_013330515.1
<i>Candidatus Methanomethylophilus alvus</i>	Mx1201	GCF_000300255.2
<i>Candidatus Methanomassiliicoccus intestinalis</i>	MGYG-HGUT-02160	GCF_902383905.1
<i>Methanomassiliicoccus luminyensis</i>	MGYG-HGUT-02161	GCF_902383895.1

6.1.8. Determination of the PyIRS Variable Regions (Linker)

The variable region was determined after aligning the PyIRS within each family. The conserved region at the N- and C-terminus where the linker is between, is given below. Numbers in brackets are the residue numbers of the motif corresponding to a reference sequence.

1) *Methanosarcina* PyIRS Alignment

Reference sequence is PyIRS of *M. barkeri*. Conserved motif at N-terminus KPLE (116-119). Conserved motif at C-terminus APSLT (152-156).

2) *Methanococcoides* PyIRS Alignment

Reference sequence is PyIRS of *M. burtonii*. Conserved motif at N-terminus KSFS (114-117). Conserved motif at C-terminus PEVKY (147-151).

3) *Methanohalophilus* PyIRS Alignment

Reference sequence is PyIRS of *M. halophilus*. Conserved motif at N-terminus SQARV (91-95). Conserved motif at C-terminus DYTPAQKKRI (138-147).

4) *Methanobolus* PyIRS Alignment

Reference sequence is PyIRS of *M. psychrophilus*. Conserved motif at N-terminus VQSSRF (99-104). Conserved motif at C-terminus KNEFT (175-179).

6.1.9. Similarity matrix of Methanosarcina PyIRS

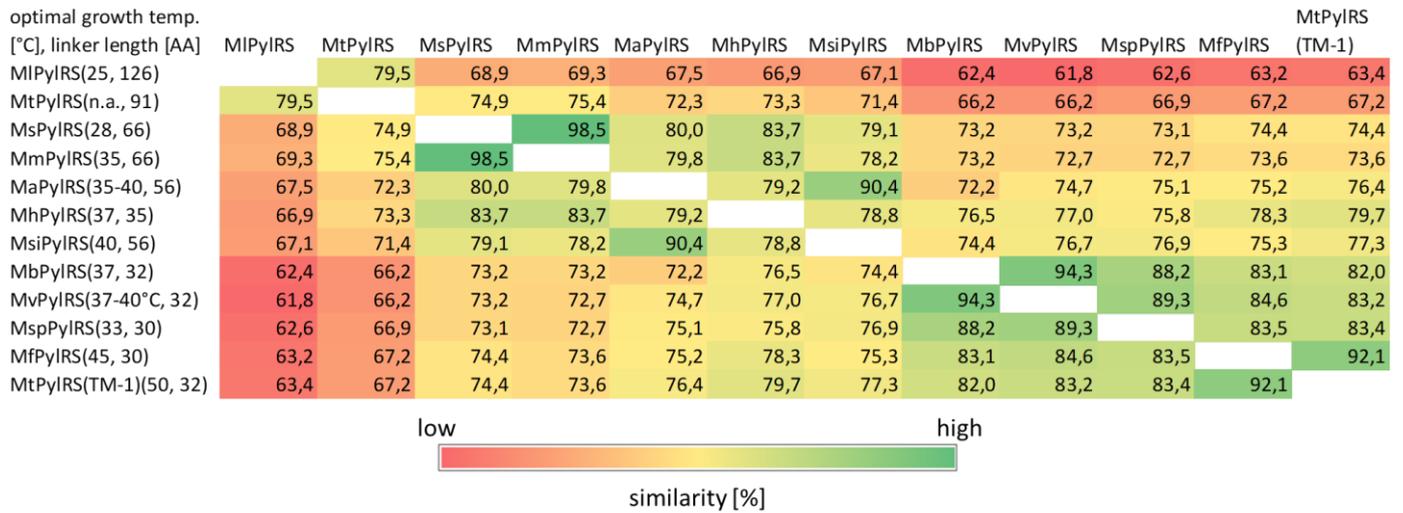


Figure 208: Similarity matrix of the complete PyIRS N- and C-terminus, without the linker.

6.1.10. Descriptive Statistical Data of PyIRSs

PyIRS without linker/AA content [%]	Methano												standard deviation		min	average	
	sarcina acetivorans	Methano barkeri	Methano salsarius	Methano horonobensis	Methano lacustris	Methano mazei	Methano siciliae	Methano spelaei	Methano soligelidii	Methano thermophilum	Methano salsarius (TM-1)	Methano vacuolata	range	max			
A	5.4	4.1	5.2	4.4	3.9	4.6	4.1	4.7	4.4	4.1	5.2	4.4	1.5	0.47	5.4	3.9	4.5
C	2.5	2.3	2.3	2.3	2.3	2.1	2.1	2.3	2.1	2.3	1.8	2.3	0.7	0.17	2.5	1.8	2.2
D	5.4	6.2	5.7	6.4	6.7	6.2	6.2	6.5	6.4	5.9	6.5	6.2	1.3	0.35	6.7	5.4	6.2
E	8.9	7.8	8.0	7.7	7.5	7.7	8.3	7.8	7.7	7.8	8.0	8.0	1.4	0.35	8.9	7.5	7.9
F	3.8	3.9	3.9	3.6	3.6	3.6	3.4	3.6	3.9	3.6	3.6	3.9	0.5	0.16	3.9	3.4	3.7
G	6.0	5.4	5.7	5.7	5.4	5.4	5.7	4.9	5.7	5.7	5.7	5.2	1.1	0.28	6.0	4.9	5.5
H	2.2	2.3	2.1	2.3	2.6	2.6	2.3	1.8	2.6	2.6	2.1	2.1	0.8	0.25	2.6	1.8	2.3
I	6.0	5.9	5.7	7.2	7.2	6.7	6.5	6.7	6.7	7.0	6.2	6.2	1.5	0.48	7.2	5.7	6.5
K	7.3	8.3	8.8	9.3	8.8	9.0	8.3	8.8	9.0	9.0	8.8	8.5	2.0	0.50	9.3	7.3	8.7
L	11.1	10.3	10.6	10.3	10.1	10.8	10.6	10.3	10.8	10.6	10.9	10.3	1.0	0.29	11.1	10.1	10.6
M	2.8	3.1	2.8	3.1	3.4	2.6	3.6	3.1	2.6	3.4	2.8	3.1	1.0	0.31	3.6	2.6	3.0
N	4.7	4.4	4.4	3.9	4.1	4.6	3.6	4.4	4.4	4.9	3.4	3.6	1.5	0.46	4.9	3.4	4.2
P	5.4	4.7	4.7	4.6	4.7	4.6	4.4	4.7	4.6	4.7	4.9	4.4	1.0	0.25	5.4	4.4	4.7
Q	0.9	1.0	1.0	1.5	1.3	1.5	1.0	1.0	1.5	1.3	1.0	1.0	0.6	0.22	1.5	0.9	1.2
R	7.6	8.3	7.8	7.5	8.0	7.7	8.3	7.5	7.7	8.3	7.8	8.0	0.8	0.29	8.3	7.5	7.9
S	7.3	7.5	7.5	7.2	7.0	7.0	7.5	7.5	6.7	6.7	7.8	8.5	1.8	0.48	8.5	6.7	7.4
T	4.1	4.4	3.4	4.6	3.9	4.9	5.2	4.1	4.9	3.6	3.6	3.9	1.8	0.56	5.2	3.4	4.2
V	5.4	6.2	6.5	4.6	5.7	4.6	5.2	5.7	4.6	4.7	5.7	6.5	1.9	0.69	6.5	4.6	5.5
W	0.6	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.2	0.06	0.8	0.6	0.8
Y	2.5	3.1	3.4	2.8	3.1	2.8	3.1	3.9	2.8	3.1	3.6	3.1	1.4	0.37	3.9	2.5	3.1
sum	99.9	100.0	100.3	99.8	100.1	99.8	100.2	100.1	99.9	100.1	100.2	100.0					
unpolar	40.5	39.0	40.2	38.6	39.4	38.3	38.6	39.6	38.4	38.9	40.1	39.6	2.2	0.71	40.5	38.3	39.3
polar	28.0	28.1	27.7	28.0	27.1	28.3	28.2	28.1	28.1	27.6	26.9	27.6	1.4	0.42	28.3	26.9	27.8
basic	17.1	18.9	18.7	19.1	19.4	19.3	18.9	18.1	19.3	19.9	18.7	18.6	2.8	0.68	19.9	17.1	18.8
acidic	14.3	14.0	13.7	14.1	14.2	13.9	14.5	14.3	14.1	13.7	14.5	14.2	0.8	0.26	14.5	13.7	14.1

Figure 209: Amino acid distribution in percent and descriptive indicators of *Methanosarcina* PyIRS variants. The data describes the amino acid distribution for the PyIRS sequence without the variable region (linker). The coloring indicates green for high values and red for low values. Color scales are different for amino acids and range/standard deviation.

linker/AA content [%]	Methano	standard	max	min	average												
	sarcina	deviation															
A	32.1	15.6	13.3	20.0	24.6	15.2	33.9	19.4	16.7	19.8	12.5	15.6	21.4	6.68	33.9	12.5	19.9
C	0.0	0.0	0.0	0.0	0.0	0.0	1.8	0.0	0.0	0.0	0.0	0.0	0.0	1.8	0.0	0.0	0.2
D	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
E	0.0	0.0	0.0	5.7	1.6	3.0	1.8	3.2	3.0	1.1	0.0	3.1	5.7	1.71	5.7	0.0	1.9
F	0.0	0.0	0.0	0.0	0.0	1.5	0.0	0.0	1.5	2.2	0.0	0.0	2.2	0.77	2.2	0.0	0.4
G	3.6	0.0	3.3	2.9	0.8	4.5	1.8	0.0	4.5	3.3	3.1	0.0	4.5	1.65	4.5	0.0	2.3
H	0.0	0.0	3.3	0.0	0.8	0.0	1.8	0.0	0.0	0.0	3.1	0.0	3.3	1.21	3.3	0.0	0.8
I	0.0	0.0	0.0	0.0	0.8	6.1	1.8	3.2	6.1	4.4	0.0	0.0	6.1	2.35	6.1	0.0	1.9
K	1.8	6.3	3.3	2.9	1.6	3.0	3.6	6.5	3.0	1.1	6.3	3.1	5.4	1.78	6.5	1.1	3.5
L	1.8	0.0	0.0	5.7	0.8	1.5	1.8	3.2	3.0	1.1	0.0	0.0	5.7	1.65	5.7	0.0	1.6
M	1.8	0.0	0.0	2.9	3.2	1.5	1.8	0.0	1.5	1.1	0.0	0.0	3.2	1.12	3.2	0.0	1.2
N	1.8	9.4	3.3	2.9	0.8	4.5	0.0	6.5	4.5	2.2	6.3	6.3	9.4	2.62	9.4	0.0	4.0
P	17.9	15.6	6.7	14.3	23.8	9.1	16.1	16.1	9.1	16.5	6.3	21.9	17.5	5.40	23.8	6.3	14.5
Q	5.4	0.0	3.3	8.6	4.8	6.1	5.4	3.2	6.1	4.4	3.1	3.1	8.6	2.06	8.6	0.0	4.5
R	1.8	3.1	3.3	0.0	0.0	0.0	0.0	0.0	0.0	1.1	0.0	3.1	3.3	1.35	3.3	0.0	1.0
S	17.9	31.3	40.0	20.0	15.9	24.2	16.1	32.3	22.7	17.6	37.5	28.1	24.1	8.07	40.0	15.9	25.3
T	8.9	9.4	10.1	8.6	11.1	10.6	7.1	3.2	9.1	20.9	12.5	6.3	17.7	4.07	20.9	3.2	9.8
V	5.4	9.4	10.1	5.7	9.5	9.1	5.4	3.2	9.1	2.2	9.4	9.4	7.9	2.66	10.1	2.2	7.3
W	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0
Y	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	0.0	0.0	1.1	0.30	1.1	0.0	0.1
sum	100.2	100.1	100.0	100.2	100.1	99.9	100.2	100.0	99.9	100.1	100.1	100.0					
unpolar	59.0	40.6	30.1	48.6	62.7	44.0	60.8	45.1	47.0	47.3	28.2	46.9	34.5	10.31	62.7	28.2	46.7
polar	37.6	50.1	60.0	43.0	33.4	49.9	32.2	45.2	46.9	49.5	62.5	43.8	30.3	8.88	62.5	32.2	46.2
basic	3.6	9.4	9.9	2.9	2.4	3.0	5.4	6.5	3.0	2.2	9.4	6.2	7.7	2.80	9.9	2.2	5.3
acidic	0.0	0.0	0.0	5.7	1.6	3.0	1.8	3.2	3.0	1.1	0.0	3.1	5.7	1.71	5.7	0.0	1.9
length	56	32	30	35	126	66	56	31	66	91	32	32	96.0	28.53	126.0	30.0	54.4

Figure 210: Amino acid distribution in percent and descriptive indicators of *Methanosarcina* PyIRS variable regions (linker). The coloring indicates green for high values and red for low values. Color scales are different for amino acids and range/standard deviation.

PyIRS without linker/AA content [%]	Methanoco				standard deviation	max	min	average		
	Methanoco	Methanoco	ccoides	Methanoco						
A	2.8	2.6	2.6	2.3	0.5	0.2	2.8	2.3	2.6	
C	2.3	2.3	2.3	2.3	0.0	0.0	2.3	2.3	2.3	
D	6.2	6.7	6.2	6.5	0.5	0.2	6.7	6.2	6.4	
E	9.3	9.3	8.6	9.1	0.7	0.3	9.3	8.6	9.1	
F	3.9	4.4	4.2	4.2	0.5	0.2	4.4	3.9	4.2	
G	6.0	5.2	6.2	6.2	1.0	0.4	6.2	5.2	5.9	
H	2.6	2.6	2.1	2.1	0.5	0.2	2.6	2.1	2.4	
I	7.5	7.3	7.3	7.0	0.5	0.2	7.5	7.0	7.3	
K	7.8	8.0	7.3	7.5	0.7	0.3	8.0	7.3	7.7	
L	9.6	9.3	9.6	9.6	0.3	0.1	9.6	9.3	9.5	
M	2.8	3.4	2.9	2.9	0.6	0.2	3.4	2.8	3.0	
N	4.1	3.9	4.4	3.6	0.8	0.3	4.4	3.6	4.0	
P	4.9	4.9	4.9	4.9	0.0	0.0	4.9	4.9	4.9	
Q	2.3	2.1	2.6	2.3	0.5	0.2	2.6	2.1	2.3	
R	9.1	8.8	9.1	9.4	0.6	0.2	9.4	8.8	9.1	
S	5.7	6.2	6.2	6.5	0.8	0.3	6.5	5.7	6.2	
T	3.9	3.6	3.9	3.6	0.3	0.2	3.9	3.6	3.8	
V	4.9	5.4	5.5	5.7	0.8	0.3	5.7	4.9	5.4	
W	0.8	0.8	0.8	0.8	0.0	0.0	0.8	0.8	0.8	
Y	3.4	3.1	3.4	3.4	0.3	0.1	3.4	3.1	3.3	
sum		99.9	99.9	100.1	99.9					
unpolar		37.2	38.1	37.8	37.4	0.9	0.3	38.1	37.2	37.6
polar		27.7	26.4	29.0	27.9	2.6	0.9	29.0	26.4	27.8
basic		19.5	19.4	18.5	19.0	1.0	0.4	19.5	18.5	19.1
acidic		15.5	16.0	14.8	15.6	1.2	0.4	16.0	14.8	15.5

Figure 211: Amino acid distribution in percent and descriptive indicators of *Methanococcoides* PyIRS variants. The data describes the amino acid distribution for the PyIRS sequence without the variable region (linker). The coloring indicates green for high values and red for low values. Color scales are different for amino acids and range/standard deviation.

linker/AA content [%]	Methanoco				range	standard deviation	max	min	average
	Methanoco ccoides alaskense	Methanoco ccoides burtonii	Methanoco ccoides s	Methanoco ccoides methyluten vulcani					
A	3.3	3.3	14.3	17.9	14.6	6.5	17.9	3.3	9.7
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
D	3.3	3.3	3.6	10.7	7.4	3.2	10.7	3.3	5.2
E	6.7	6.7	7.1	3.6	3.5	1.4	7.1	3.6	6.0
F	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
G	0.0	0.0	3.6	3.6	3.6	1.8	3.6	0.0	1.8
H	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
I	10.0	10.0	0.0	0.0	10.0	5.0	10.0	0.0	5.0
K	10.0	10.0	10.7	10.7	0.7	0.4	10.7	10.0	10.4
L	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
M	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
N	3.3	3.3	10.7	10.7	7.4	3.7	10.7	3.3	7.0
P	16.7	16.7	17.9	17.9	1.2	0.6	17.9	16.7	17.3
Q	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
R	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S	23.3	23.3	7.1	7.1	16.2	8.1	23.3	7.1	15.2
T	10.0	6.7	21.4	14.3	14.7	5.5	21.4	6.7	13.1
V	13.3	16.7	3.6	3.6	13.1	5.8	16.7	3.6	9.3
W	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Y	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
sum	99.9	100.0	100.0	100.1					
unpolar	43.3	46.7	35.8	39.4	10.9	4.1	46.7	35.8	41.3
polar	36.6	33.3	42.8	35.7	9.5	3.5	42.8	33.3	37.1
basic	10.0	10.0	10.7	10.7	0.7	0.4	10.7	10.0	10.4
acidic	10.0	10.0	10.7	14.3	4.3	1.8	14.3	10.0	11.3
length	56.0	57.0	58.0	59.0	3.0	1.1	59.0	56.0	57.5

Figure 212: Amino acid distribution in percent and descriptive indicators of *Methanococcoides* PyIRS variable regions (linker). The coloring indicates green for high values and red for low values. Color scales are different for amino acids and range/standard deviation.

PyIRS without linker/AA content [%]	Methanohalo philus euehalobius		Methanohalo philus halophilus		Methanohalo philus levihalophilus		Methanohalo philus mahii		Methanohalo philus portucalensis		Methanohalo philus profundi		range	standard deviation	max	min	average
A	3.3	3.0	3.6	3.0	3.0	3.0	3.0	0.6	0.2	3.6	3.0	3.2					
C	2.2	2.2	2.2	2.2	2.2	2.2	2.2	0.0	0.0	2.2	2.2	2.2					
D	6.3	6.9	6.1	7.2	6.9	6.9	6.9	1.1	0.4	7.2	6.1	6.7					
E	6.6	6.6	9.1	6.6	6.6	6.6	6.6	2.5	0.9	9.1	6.6	7.0					
F	3.9	3.6	4.7	3.6	3.6	3.6	3.6	1.1	0.4	4.7	3.6	3.8					
G	5.8	5.8	6.9	5.5	5.8	5.8	5.8	1.4	0.4	6.9	5.5	5.9					
H	2.5	2.5	1.9	2.2	2.5	2.2	2.2	0.6	0.2	2.5	1.9	2.3					
I	7.4	6.9	6.9	7.4	7.2	7.2	7.2	0.5	0.2	7.4	6.9	7.2					
K	8.0	7.7	8.5	7.4	7.4	7.4	7.4	1.1	0.4	8.5	7.4	7.7					
L	9.6	10.2	9.6	9.9	9.9	9.9	9.9	0.6	0.2	10.2	9.6	9.9					
M	3.9	3.6	3.6	3.6	3.9	3.9	3.9	0.3	0.2	3.9	3.6	3.8					
N	5.5	6.1	4.4	5.5	5.0	5.5	5.5	1.7	0.5	6.1	4.4	5.3					
P	3.9	3.9	3.3	3.6	3.9	3.9	3.9	0.6	0.2	3.9	3.3	3.8					
Q	2.2	2.2	1.4	2.8	2.2	2.5	2.5	1.4	0.4	2.8	1.4	2.2					
R	8.5	8.3	8.8	8.3	9.1	8.8	8.8	0.8	0.3	9.1	8.3	8.6					
S	5.8	6.1	6.1	6.1	6.1	6.1	6.1	0.3	0.1	6.1	5.8	6.1					
T	5.8	5.2	4.1	6.1	5.8	5.5	5.5	2.0	0.7	6.1	4.1	5.4					
V	4.4	5.0	4.4	4.7	4.7	4.7	4.7	0.6	0.2	5.0	4.4	4.7					
W	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.0	0.0	0.8	0.8	0.8					
Y	3.6	3.6	3.6	3.6	3.6	3.6	3.6	0.0	0.0	3.6	3.6	3.6					
sum	100.0	100.2	100.0	100.1	100.2	100.1											
unpolar	37.2	37.0	36.9	36.6	37.0	37.0	0.6	0.2	37.2	36.6	37.0						
polar	30.9	31.2	28.7	31.8	30.7	31.2	3.1	1.0	31.8	28.7	30.8						
basic	19.0	18.5	19.2	17.9	19.0	18.4	1.3	0.4	19.2	17.9	18.7						
acidic	12.9	13.5	15.2	13.8	13.5	13.5	2.3	0.7	15.2	12.9	13.7						

Figure 213: Amino acid distribution in percent and descriptive indicators of *Methanohalophilus* PyIRS variants. The data describes the amino acid distribution for the PyIRS sequence without the variable region (linker). The coloring

indicates green for high values and red for low values. Color scales are different for amino acids and range/standard deviation.

linker/AA content [%]	Methanohalo philus euhalobius	Methanohalo philus halophilus	Methanohalo philus levihalophilus	Methanohalo philus mahii	Methanohalo philus portucalensis	Methanohalo philus profundi	range	standard deviation	max	min	average
	A	14.6	11.9	11.1	7.3	14.3	16.7	9.4	3.0	16.7	7.3
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
D	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
E	9.8	7.1	8.9	9.8	7.1	7.1	2.7	1.2	9.8	7.1	8.3
F	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
G	0.0	0.0	2.2	0.0	0.0	0.0	2.2	0.8	2.2	0.0	0.4
H	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
I	0.0	2.4	2.2	2.4	0.0	0.0	2.4	1.2	2.4	0.0	1.2
K	24.4	28.6	17.8	26.8	26.2	28.6	10.8	3.7	28.6	17.8	25.4
L	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
M	2.4	2.4	0.0	2.4	2.4	2.4	2.4	0.9	2.4	0.0	2.0
N	4.9	2.4	4.4	2.4	7.1	2.4	4.7	1.7	7.1	2.4	3.9
P	14.6	11.9	17.8	12.2	11.9	14.3	5.9	2.1	17.8	11.9	13.8
Q	0.0	0.0	2.2	0.0	0.0	2.4	2.4	1.1	2.4	0.0	0.8
R	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S	7.3	11.9	13.3	17.1	7.1	9.5	10.0	3.5	17.1	7.1	11.0
T	7.3	7.1	6.7	4.9	7.1	4.8	2.5	1.1	7.3	4.8	6.3
V	14.6	14.3	13.3	14.6	16.7	11.9	4.8	1.5	16.7	11.9	14.2
W	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Y	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
sum	99.9	100.0	99.9	99.9	99.9	100.1					
unpolar	46.2	42.9	44.4	38.9	45.3	45.3	7.3	2.4	46.2	38.9	43.8
polar	19.5	21.4	28.8	24.4	21.3	19.1	9.7	3.3	28.8	19.1	22.4
basic	24.4	28.6	17.8	26.8	26.2	28.6	10.8	3.7	28.6	17.8	25.4
acidic	9.8	7.1	8.9	9.8	7.1	7.1	2.7	1.2	9.8	7.1	8.3
length	41.0	42.0	45.0	41.0	42.0	42.0	4.0	1.3	45.0	41.0	42.2

Figure 214: Amino acid distribution in percent and descriptive indicators of *Methanohalophilus* PyIRS variable regions (linker). The coloring indicates green for high values and red for low values. Color scales are different for amino acids and range/standard deviation.

PyIRS without linker/AA content [%]	Methanobolus								range	standard deviation	max	min	average
	Methanobolus bombayensis	Methanobolus halotolerans	Methanobolus profundi	Methanobolus psychrophilus	Methanobolus psychrotolera ns	Methanobolus tindarius	Methanobolus vulcani	Methanobolus zinderi					
A	2.4	3.8	2.7	3.5	2.7	3.2	3.0	2.7	1.4	0.4	3.8	2.4	3.0
C	2.4	2.7	2.4	2.7	2.4	2.4	2.4	2.7	0.3	0.1	2.7	2.4	2.5
D	5.4	4.0	5.1	6.2	4.9	5.9	5.7	5.1	2.2	0.6	6.2	4.0	5.3
E	9.4	9.4	9.2	7.5	10.2	8.6	9.2	9.1	2.7	0.7	10.2	7.5	9.1
F	4.6	5.1	4.6	4.8	5.1	4.9	5.1	5.4	0.8	0.3	5.4	4.6	5.0
G	6.2	6.5	6.7	5.6	6.7	6.2	6.2	5.9	1.1	0.4	6.7	5.6	6.3
H	1.6	1.6	1.9	2.1	1.9	1.9	2.2	1.6	0.6	0.2	2.2	1.6	1.9
I	6.7	5.4	6.7	5.6	8.6	6.7	7.0	5.9	3.2	0.9	8.6	5.4	6.6
K	8.6	8.6	8.6	7.0	8.4	8.4	8.1	8.3	1.6	0.5	8.6	7.0	8.3
L	10.0	8.9	10.0	9.4	9.4	10.2	10.0	8.9	1.3	0.5	10.2	8.9	9.6
M	4.3	4.8	4.0	4.6	4.0	4.3	4.3	4.3	0.8	0.3	4.8	4.0	4.3
N	5.4	4.8	4.9	5.1	4.6	4.6	5.1	4.3	1.1	0.3	5.4	4.3	4.9
P	3.5	4.3	3.8	3.5	3.8	3.8	3.8	4.3	0.8	0.3	4.3	3.5	3.9
Q	1.6	1.9	1.3	2.4	1.9	1.9	1.3	2.2	1.1	0.4	2.4	1.3	1.8
R	8.4	7.5	8.1	8.8	7.8	8.1	8.1	8.3	1.3	0.4	8.8	7.5	8.1
S	8.1	8.3	7.3	8.3	7.5	7.5	7.3	8.9	1.6	0.5	8.9	7.3	7.9
T	3.2	3.0	3.8	4.3	3.0	3.2	3.5	3.0	1.3	0.4	4.3	3.0	3.4
V	4.6	5.9	5.1	5.4	3.5	4.9	4.6	6.2	2.7	0.8	6.2	3.5	5.0
W	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.0	0.0	0.8	0.8	0.8
Y	2.7	2.7	3.0	2.4	2.7	2.4	2.4	2.2	0.8	0.2	3.0	2.2	2.6
sum	99.9	100.0	100.0	100.0	99.9	99.9	100.1						
unpolar	36.9	39.0	37.7	37.6	37.9	38.8	38.6	38.5	2.1	0.7	39.0	36.9	38.1
polar	29.6	29.9	29.4	30.8	28.8	28.2	28.2	29.2	2.6	0.8	30.8	28.2	29.3
basic	18.6	17.7	18.6	17.9	18.1	18.4	18.4	18.2	0.9	0.3	18.6	17.7	18.2
acidic	14.8	13.4	14.3	13.7	15.1	14.5	14.9	14.2	1.7	0.6	15.1	13.4	14.4

Figure 215: Amino acid distribution in percent and descriptive indicators of *Methanobolus* PyIRS variants. The data describes the amino acid distribution for the PyIRS sequence without the variable region (linker). The coloring indicates green for high values and red for low values. Color scales are different for amino acids and range/standard deviation.

linker/AA content [%]	Methanolobus									standard deviation	max	min	average
	Methanolobus bombayensis	Methanolobus halotolerans	Methanolobus profundus	Methanolobus psychrophilus	Methanolobus psychrotolera	Methanolobus tindarius	Methanolobus vulcani	Methanolobus zinderi	range				
A	6.9	10.0	6.9	5.7	2.8	10.1	10.1	12.0	9.2	2.8	12.0	2.8	8.1
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
D	0.0	3.8	1.4	2.9	2.8	2.5	1.3	1.3	3.8	1.1	3.8	0.0	2.0
E	6.9	2.5	5.6	4.3	8.3	6.3	5.1	4.0	5.8	1.7	8.3	2.5	5.4
F	1.4	2.5	0.0	0.0	1.4	0.0	1.3	2.7	2.7	1.0	2.7	0.0	1.2
G	2.8	2.5	1.4	0.0	1.4	2.5	1.3	1.3	2.8	0.9	2.8	0.0	1.7
H	1.4	2.5	4.2	4.3	4.2	3.8	3.8	2.7	2.9	1.0	4.3	1.4	3.4
I	6.9	1.3	4.2	5.7	5.6	6.3	1.3	1.3	5.6	2.3	6.9	1.3	4.1
K	13.9	10.0	9.7	11.4	11.1	12.7	11.4	17.3	7.6	2.3	17.3	9.7	12.2
L	0.0	2.5	0.0	2.9	0.0	1.3	0.0	2.7	2.9	1.3	2.9	0.0	1.2
M	1.4	1.3	0.0	1.4	1.4	0.0	1.3	1.3	1.4	0.6	1.4	0.0	1.0
N	9.7	6.3	6.9	5.7	13.9	12.7	11.4	5.3	8.6	3.2	13.9	5.3	9.0
P	12.5	16.3	13.9	12.9	12.5	12.7	10.1	12.0	6.2	1.6	16.3	10.1	12.9
Q	5.6	8.8	8.3	10.0	8.3	2.5	3.8	8.0	7.5	2.5	10.0	2.5	6.9
R	2.8	5.0	4.2	1.4	2.8	2.5	2.5	1.3	3.7	1.2	5.0	1.3	2.8
S	13.9	6.3	16.7	17.1	12.5	17.7	16.5	9.3	11.4	3.9	17.7	6.3	13.8
T	12.5	10.0	13.9	7.1	5.6	5.1	12.7	9.3	8.8	3.1	13.9	5.1	9.5
V	1.4	7.5	2.8	5.7	5.6	1.3	5.1	6.7	6.2	2.2	7.5	1.3	4.5
W	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Y	0.0	1.3	0.0	1.4	0.0	0.0	1.3	1.3	1.4	0.7	1.4	0.0	0.7
sum	100.0	100.4	100.1	99.9	100.2	100.0	100.3	99.8	13.6	4.6	41.4	27.8	32.9
unpolar	30.5	41.4	27.8	34.3	29.3	31.7	29.2	38.7	12.7	4.5	47.2	34.5	41.5
polar	44.5	35.2	47.2	41.3	41.7	40.5	47.0	34.5	4.2	1.2	21.3	17.1	18.4
basic	18.1	17.5	18.1	17.1	18.1	19.0	17.7	21.3	5.8	1.7	11.1	5.3	7.4
acidic	6.9	6.3	7.0	7.2	11.1	8.8	6.4	5.3	3.7	1.2	5.0	1.3	2.8
length	72.0	80.0	72.0	70.0	72.0	79.0	79.0	75.0	10.0	3.7	80.0	70.0	74.9

Figure 216: Amino acid distribution in percent and descriptive indicators of *Methanolobus* PyIRS variable regions (linker). The coloring indicates green for high values and red for low values. Color scales are different for amino acids and range/standard deviation.

6.1.11. Example of Gene Deletion with *nfsB*

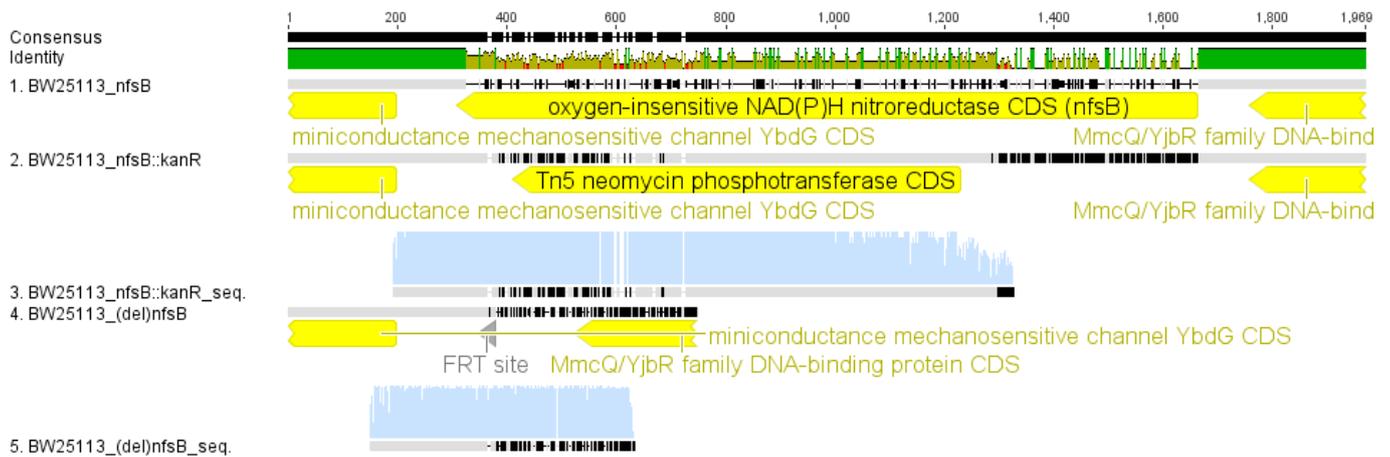


Figure 217: Alignment of *nfsB* containing target locus with 2) locus wherein *nfsB* is replaced by the Neomycin resistance cassette, 3) corresponding sanger sequencing of PCR product, 4) flip out of Neomycin resistance cassette, leaving the FRT site, 5) corresponding sanger sequencing of flip out.

7. Eidesstattliche Erklärung

Hiermit erkläre ich an Eides Statt, dass ich die vorliegende Arbeit eigenständig und ohne fremde Hilfe angefertigt habe. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Es wurden ausschließlich die angegebenen Quellen und Hilfsmittel genutzt. Alle Ergebnisse, die in Kooperation mit anderen Arbeitsgruppen erhalten wurden und in anderen Arbeiten zu finden sind, sind dementsprechend gekennzeichnet. Diese Arbeit wurde in dieser oder ähnlicher Form bisher bei keiner anderen Prüfungsbehörde einer Hochschule, weder im Inland noch im Ausland, eingereicht.

Nikolaj Georg Koch

Berlin, Juni 2022