TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Chemie
Labor für Synthetische Biochemie

Development and optimization of genetic code expansion tools to study and validate protein-protein interactions

MARIE-KRISTIN VON WRISBERG

Vollständiger Abdruck der, von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitzender:		Prof. Dr. Hubert A. Gasteiger
Prüfer der Dissertation:	1.	Prof. Dr. Kathrin Lang
	2.	Dr. Sabine Schneider
	3.	…….....

Die Dissertation wurde am 08.07.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 24.08.2020 angenommen.

The work for the following thesis was performed from April 2016 to July 2020 in the Lab of Synthetic Biochemistry of Prof. Dr. Kathrin Lang at the Technical University Munich.
The supervisioner of the thesis was Prof. Dr. Kathrin Lang. Mentor for the thesis was Prof. Aymlet Itzen.

Further support for this research work was provided by the SFB 1035.

LIST OF PUBLICATIONS

Some contents and results of the following thesis have been published in the papers or presented posters listed below:

Journal Publications (peer-reviewed)

M. Cigler, T. G. Müller, D. Horn-Ghetko, **M.-K. von Wrisberg**, M. Fottner, R. S. Goody, A. Itzen, M. P. Müller, K. Lang. *Proximity-triggered covalent stabilization of low-affinity protein complexes in vitro and in vivo*. *Angew. Chem. Int. Ed.* **56**(49) (2017), 15737-15741.

S.V. Mayer,[#] A. Murnauer,[#] **M.-K. von Wrisberg**, M.-L. Jokisch and K. Lang. *Photo-induced and rapid labeling of tetrazine-bearing proteins via cyclopropenone-caged bicyclononynes*. *Angew. Chem. Int. Ed.* **58**(44) (2019), 15876-15882. ([#] these authors contributed equally)

Manuscript in preparation

J. Du,[#] **M.-K. von Wrisberg**,[#] B. Gulen, M. Stahl, C. Pett, C. Hedberg, K. Lang, S. Schneider, A. Itzen. Working title**:** *Rab1-AMPylation by Legionella DrrA is allosterically activated by Rab1.* ([#] these authors contributed equally)

Poster presentations (peer-reviewed)

M.-K. von Wrisberg, M. Cigler, M. Stahl, S. Sieber, M. Müller, S. Schneider, A. Itzen and K. Lang. *Proximity-Triggered Covalent Stabilization of Low-Affinity Protein Complexes in vitro and in vivo.* EMBO Workshop „Enzymes, biocatalysis and chemical biology: The new frontiers", 09.09-12.09.2018, 27100 Pavia (Italy).

M.-K. von Wrisberg, M. Fottner, S. Schneider and K. Lang. *Structural insights into the selectivity of a novel aminoacyl tRNA synthetase*. "The 2020 International Symposium on Chemical Biology" Campus Biotech (NCCR in Chemical Biology), 22.01-24.01.2020, 1211 Geneva (Switzerland).
Review

ABSTRACT

Today's understanding of molecular biology has its origins in 1958 when pioneer Francis Crick postulated the central dogma of molecular biology.

"[...] once sequential information has passed into protein it cannot get out again"
(Crick, 1958, p. 153)[1].

Since then hundreds of cellular mechanisms have been researched at the molecular level, thereby furthering and shaping our understanding of the building blocks of life, their interaction in homeostasis and pathogenesis, and at the same time significantly advancing the change towards modern medicine. When we refer to the molecular building blocks of life, we think of the macromolecules DNA, RNA and protein. DNA represents the primary information storage, RNA works as information carrier and proteins are the executive authority for the regulation of cellular processes. Therefore, proteins are usually the focus of the elucidation of cellular molecular mechanisms. Over the last 60 years, the field of biochemistry has been revolutionized. New techniques have brought new possibilities; especially the combination of biological and chemical methods has given a new dimension to the work with and on proteins. The expansion of the genetic code (GCE) is one of these technical innovations. GCE enables the site-specific incorporation of functionalities (dubbed unnatural amino acids (uAAs)) beyond the 20 proteinogenic amino acids into proteins in living cells. This is typically achieved through amber suppression, where the uAA is recognized by an orthogonal aminoacyl-tRNA synthetase (aaRS) and charged onto the respective orthogonal tRNA that directs its incorporation in response to an amber stop codon within the mRNA during protein translation. This has allowed in recent years to genetically encode various designer amino acids with tailored chemical and physical properties into proteins in bacteria and eukaryotes. Together with developments in chemistries that are amenable to and selective within physiological settings, these strategies have started to have a big impact on biological studies, as they enable diverse applications, including approaches for imaging and probing proteins through bioorthogonal reactions, studying protein-interaction patterns in *in vivo* settings via incorporation of crosslinker amino acids and re-engineering proteins functions and biological pathways.

This thesis addresses two broad topics: (i) Expansion and optimization of components of the GCE toolbox for efficient amber suppression in *Escherichia coli* and mammalian cells and (ii) application of GCE-tools for stabilizing protein interactions via chemical crosslinking.
The first two chapters (chapter I +II) encompass the establishment of a novel aaRS/tRNA pair from *Methanomethylophilus alvus* (*Ma*-PylRS/tRNA pair) for amber suppression in living cells. Besides development of an efficient expression- and evolution-system using this novel aaRS/tRNA pair, structural elucidation of *Ma*-PylRS contributed new insights in substrate specificity of this enzyme, laying the foundation for increasing its substrate scope. Moreover, new crystallization data for a novel uAA-bound *Methanosracina mazei* PylRS variant was obtained, revealing new insights into uAA binding and offering a starting point for rational mutagenesis in order to increase the enzyme substrate spectrum.
The second part of the thesis (chapter III) focuses on the application of GCE-based chemical crosslinking methods for studying protein interactions. Within chapter III 4.1 a proximity-enhanced protein crosslinking (PEPC) approach between bromo-alkyl bearing uAAs and

cysteines to enable the stabilization of low-affinity protein-protein interactions is described. This approach allowed the covalent trapping of the transient protein complex between human Rab1b and the ATase domain of the *Legionella* virulence factor DrrA. In collaboration with the group of Prof. Aymelt Itzen (UKE, Hamburg) we could thereby reveal a formerly unknown interaction site which may allosterically regulate DrrA-ATase activity. As part of the project a tailored *E. coli* strain, which circumvents the enzymatic degradation of uAAs for PEPC, was created. Furthermore, the portfolio of uAAs for PEPC was expanded by two candidates that are inert towards enzymatic degradation.

Likewise, a new class of PEPC-uAAs that facilitates identification of crosslinked partners and crosslinking sites via mass spectrometric (MS) analysis was established (chapter III 4.2). Thereby, a MS^2-cleavable linker is integrated into the uAA side chain. Unknown interfaces of protein samples crosslinked in this way can be easily identified and characterized based on the characteristic ion fragmentation pattern of the MS^2 label. On the basis of the Rab1b-DrrA-ATase complex we demonstrated the advantages of the simplified data evaluation enabled by the MS^2-cleavable uAAs.

Furthermore, within a collaboration with Prof. Sieber's group (TUM, Garching) (chapter III 4.3) we set the basis for a method to covalently trap and stabilize labile enzyme-substrate complexes. For this we encoded 2,3-diamonopropionic acid (DAP) into the active site of the serine protease ClpP from *Staphylococcus aureus* (*Sa*ClpP) and showed that we can capture *Sa*ClpP-inhibitor complexes through a stable amide bond. The covalently trapped *Sa*ClpP complexes will further structural elucidation to gain insights into functional snapshots of *Sa*ClpP inhibition and activation.

ZUSAMMENFASSUNG

Das moderne Verständnis vom Begriff der molekularen Biologie fand ihren Anfang 1958 mit dem Pionier Francis Crick, der den Informationsfluss in allen lebenden Zellen als zentrales Dogma der Molekularbiologie postulierte.

„Wenn (sequenzielle) Information einmal in ein Protein übersetzt wurde, kann sie dort nicht wieder herausgelangen“[a] (Crick, 1958, S. 153)[1].

Seitdem wurden hunderte zelluläre Mechanismen auf molekularer Ebene erforscht, deren Kenntnis unser heutiges Verständnis von den Bausteinen des Lebens, ihr Zusammenspiel bei Homöostase und Pathogenese geformt und zugleich den Wandel der modernen Medizin maßgeblich vorangebracht hat. Spricht man von molekularen Bausteinen des Lebens, so sind damit die Makromoleküle DNA, RNA und Proteine gemeint. Die DNA ist der Informationsspeicher, die RNA der Informationsüberträger und die Proteine die Exekutive zur Regulation der zellulären Prozesse. Daher sind meist die Proteine im Fokus der Aufklärung zellulärer molekularer Mechanismen. In den letzten 60 Jahren hat sich das Feld der Biochemie revolutioniert. Neue Techniken brachten neue Möglichkeiten, insbesondere die Kombination von biologischen und chemischen Methoden, wodurch die Arbeit mit und an Proteinen eine neue Dimension erlangte.

Eine der Neuerungen in diesem Feld war die Erweiterung des genetischen Codes (GCE). GCE ermöglicht den positionsspezifischen Einbau von weiteren Funktionalitäten neben den 20 proteinogenen Aminosäuren (bezeichnet als unnatürliche Aminosäuren (uAA)) in Proteine in lebenden Zellen. Dies kann durch Amber Suppression bewerkstelligt werden, wobei eine orthogonale Aminoacyl-tRNA Synthetase (aaRS) die uAA erkennt und auf die entsprechende orthogonale tRNA lädt, welche den Einbau der uAA respektive eines in die mRNA eingebauten Amber-Codons während des zellulären Translationsprozesses vermittelt. Dies erlaubte die genetische Codierung von einer Vielzahl von Designer-Aminosäuren mit maßgeschneiderten chemischen und physikalischen Eigenschaften in bakterielle und eukaryotische Proteine. Im Zusammenspiel mit der Entwicklung von chemischen Reaktionen, die zugänglich und selektiv unter physiologischen Bedingungen sind, haben diese Strategien eine große Bedeutung für biologische Studien bekommen unter anderem als bioorthogonale Labeling-Methode für bildgebende Verfahren, für die *in vivo* Untersuchung und Validierung von Protein-Protein Interaktionen mittels Einbau von Crosslinker-Aminosäuren oder der gezielten Modellierung von Proteinfunktionen und biologischen Reaktionswegen.

Diese Arbeit befasst sich mit zwei Aspekten: (i) Erweiterung und Optimierung von Komponenten zur Erweiterung des genetischen Codes für die effiziente Amber Suppression in *Escherichia coli* und Säugerzellen und (ii) der Anwendung von GCE-Methoden für die Stabilisierung von Proteininteraktionen über chemische Vernetzung.

Die ersten beiden Kapitel (Kapitel I + II) umfassen die Etablierung eines neuen orthogonalen aaRS/tRNA Paares aus dem Organismus *Methanomethylophilus alvus* (*Ma*-PylRS/tRNA Paar) für die Amber Suppression in lebenden Zellen. Neben der Entwicklung eines effizienten Expressions- und Evolutionssystems für das neue aaRS/tRNA Paar wurden durch

[a] Übersetzung durch den Autor der vorliegenden Arbeit

Strukturaufklärung von *Ma*-PylRS neue Erkenntnisse über die Substratspezifität des Enzyms gewonnen, welche die Grundlage für einen erweiterten Substratbereich liefern. Des Weiteren konnten neue Kristallisationsdaten für eine Enzyme-Mutante des aaRS/tRNA Paares mit Herkunft aus dem Organismus *Mathanosarcina mazei* erhoben werden, welche durch Co-Kristallisation mit ihrer uAA Einblicke in das Substartbindeverhalten liefern und Ausgangspunkt für die zielgerichtete Mutagenese zur Erweiterung des Enzyme-Substratspektrums darstellen.

Im zweiten Teil der Arbeit (Kapitel III) stand die Anwendung von GCE-Methoden für die Untersuchung von Protein-Interaktionen auf Basis der chemischen Vernetzung im Fokus. In Kapitel III 4.1 wird ein Ansatz der proximitätsvermittelten Protein Vernetzung (PEPC) beschrieben, der die Reaktion zwischen einer uAA mit Bromoalkyl-Kette und Cysteinen für die Stabilisierung von niederaffinen Interaktionen nutzt. Diese Methode ermöglichte die kovalente Verbrückung des transienten Proteinkomplexes zwischen der human GTPase Rab1b und der ATase-Domäne des *Legionella* Virulenzfaktors DrrA. In Zusammenarbeit mit der Gruppe von Prof. Aymelt Itzen (UKE, Hamburg) konnten wir dabei eine vormals unbekannte Interaktionsstelle aufdecken, welche eine allosterische Regulation auf die DrrA-ATase Aktivität ausüben könnte. Im Rahmen des Projekts wurde ein maßgeschneiderter *Escherichia coli* Expressionsstamm erschaffen, der den enzymatischen Abbau von uAAs für PEPC verhindert. Zusätzlich wurde das Portfolio an uAAs für PEPC um zwei Kandidaten erweitert, welche stabil gegenüber enzymatischem Abbau sind.

Ebenfalls wurde eine neue Klasse an uAAs etabliert, welche die Identifizierung von vernetzten Interaktionspartnern mittels massenspektrometrischer (MS) Analyse erleichtert (Kapitel III 4.2). Dafür wird eine MS^2-spaltbare Gruppe in die Seitenkette der uAA integriert. Unbekannte Interaktionsstellen von derart vernetzten Proteinen könnten leicht anhand des charakteristischen Spaltmusters des MS^2-Labels identifiziert werden. Auf Grundlage des Rab1b-DrrA-ATase Komplexes konnten wir die Vorteile der vereinfachten Datenauswertung basierend auf den MS^2-spaltbaren uAAs demonstrieren.

Außerdem haben wir in Zusammenarbeit mit der Gruppe von Prof. Sieber (TUM, Garching) (Kapitel III 4.3), die Grundlage für eine Methode zur kovalenten Verbrückung und Stabilisierung von instabilen Enzym-Substrat Komplexen geschaffen. Dafür haben wir 2,3-Diaminopropionsäure (DAP) im aktiven Zentrum der Serin-Protease ClpP aus *Staphylococcus aureus* (*Sa*ClpP) eingebaut und konnten zeigen, dass *Sa*ClpP-Inhibitor Komplexe durch die Bildung einer stabilen Amid-Bindung eingefangen werden können. Die kovalent verbundenen *Sa*ClpP Komplexe sollen der zukünftigen Strukturaufklärung dienen, um Einblicke in funktionale Momentaufnahmen der *Sa*ClpP Inhibition und Aktivierung zu erlangen.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviations of natural amino acids

A, Ala	alanine
C, Cys	cysteine
D, Asp	aspartic acid
E, Glu	glutamic acid
F, Phe	phenylalanine
G, Gly	glycine
H, His	histidine
I, Ile	isoleucine
K, Lys	lysine
L, Leu	leucine
M, Met	methionine
N, Asn	asparagine
P, Pro	proline
R, Arg	arginine
Q, Gln	glutamine
S, Ser	serine
T, Thr	threonine
V, Val	valine
W, Trp	tryptophane
Y, Tyr	tyrosine
O, Pyl	pyrrolysine
U, Sec	selenocysteine

Abbreviations of antibiotics

Amp	ampicillin
Cam	chloramphenicol
Kan	kanamycin
Pen	penicillin
Spec	spectinomycin
Str	streptinomycin
Tet	tetracyclin

Abbreviation for DNA/RNA bases

A	adenine
T	thymine
G	guanine
C	cytosine
U	uracil

Abbreviations of unnatural amino acids

AcK	N^6-acetyl-*L*-lysine
AzGGK	N^6-((2-azidoacetyl)glycyl)-L-lysine
BCNK	N^6-((bicyclo[6.1.0]non-4-yn-9-ylmethoxy)carbonyl)-L-lysine
BocK	N^6-(*tert*-butoxycarbonyl)-L-lysine
BrCnK	N^6-(n-bromo-n-oyl)-L-lysine
BrC6K	N^6-(6-bromohexanoyl)-L-lysine
BrC7K	N^6-(7-bromoheptanoyl)-L-lysine
BrCOnK	N^6-((n-bromo-n-oxy)carbonyl)-L-lysine
BrCO6K	N^6-((4-bromobutoxy)carbonyl)-L-lysine
BrCO7K	N^6-(((5-bromopentyl)oxy)carbonyl)-L-lysine
BrCN6K	N^6-((5-bromobutyl)carbamoyl)-L-lysine
BrCN7K	N^6-((5-bromopentyl)carbamoyl)-L-lysine
BrCOS8K	N^6-((2-((3-bromopropyl)thio)ethoxy)carbonyl)lysine
CpK	N^6-(((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl)-L-lysine
DAP	(*S*)-2,3-diaminopropanoic acid
DiaziK	N^6-((2-(3-methyl-3*H*-diazirin-3-yl)ethoxy)carbonyl)-L-lysine
DiZHSeC	(*S*)-2-amino-4-((3-((3-(3-methyl-3*H*-diazirin-3-yl)propyl)amino)-3-

	oxopropyl)selanyl)butanoic acid
GGK	N^6-glycylglycyl-*L*-lysine
HpGGK	N^6-(((2-nitrobenzyl)oxy)carbonyl)glycylglycyl-L-lysine
mTetK	N^6-(((3-(6-methyl-1,2,4,5-tetrazin-3-yl)benzyl)oxy)carbonyl)-L-lysine
pDAP	(2*S*)-2-amino-3-(((2-((1-(6-nitrobenzo[*d*][1,3]dioxol-5-yl)ethyl)thio)ethoxy)carbonyl)amino)propanoic acid
TCO-Lys	N^{ε}-((((E)-cyclooct-2-en-1-yl)oxy)carbonyl)-l-lysine
TetK	N^6-((2-(6-methyl-1,2,4,5-tetrazin-3-yl)ethoxy)-carbonyl)-L-lysine
TM-439	N^6-((2-(vinylsulfinyl)ethoxy)carbonyl)lysine
uAA	unnatural amino acid

Abbreviations of organisms and cells

A. thaliana	*Arabidopsis thaliana*
B. subtilis	*Bacillus subtilis*
C. elegans	*Caenorhabditis elegans*
E. coli	*Escherichia coli*
L. pneumophila	*Legionella pneumophila*
M. barkeri (Mb)	*Methanosarcina barkeri*
M. alvus (Ma)	*Methanomethylophilus alvus*
M. jannaschii (Mj)	*Methanocaldococcus jannaschii*
M. mazei (Mm)	*Methanosarcina mazei*
M. tuberculosis	*Mycobacterium tuberculosis*
HEK293T cells	human embryonic kidney 293T cells
S. cerevisiae	*Saccharomyces cerevisiae*
S. aureus	*Staphylococcus aureus*

Abbreviations of DNA, RNA, proteins, enzymes and related compounds

α-His-HRP	monoclonal anti-His6 antibody-HRP conjugate
α-Rab1b-Y77AMP	monoclonal anti-AMPylated Rab1b antibody from rabbit
α-Strep-HRP	StepMAB-Classic, HRP conjugate (monoclonal anti-StrepTagII antibody-HRP conjugate)
aaRS	aminoacyl-tRNA synthetase
ATase/AMPyD	AMPylation domain
AspN	endoproteinase Asp-N
bp	base pair
BSA	bovine serum albumin
CAT	chloramphenicol acetyltransferase
DNA	desoxyribonucleic acid
DrrA	*L. pneumophila* effector protein DrrA
ds	double-stranded
Ent	enterokinase recognition and cleavage site
GEF	guanosin exchange factor
GluC	endoproteinase GluC
HRP	horseradish peroxidase
kb	kilobase pair

miRNA	micro RNA
PylRS	pyrrolysl tRNA synthetase
Rab1b	ras-related protein Rab1b
Rab8A	ras-related protein Rab8A
RISC	RNA induced silencing complex
RNA	ribonucleic acid
*Sa*ClpP	caseinolytic protease P of *Staphylococcus aureus*
sfGFP	superfolder green fluorescent protein
siRNA	short interfering RNA
Srt	sortase
ss	single-stranded
TEV	tobacco etch virus protease
Throm	thrombin
tRNA	transfer RNA
TyrRS	tyrosyl tRNA synthetase
Ub	ubiquitin

Abbreviations for chemicals and chemical groups

2DPBA	(2-(diphenylphosphino)benzoic acid
ACN	acetonitrile
AcOH	acetic acid
ApCpp	Adenosine-5'-[(α,ß)-methyleno]triphosphate
APS	ammonium persulfate
ATP	adenosine triphosphate
AV170	methyl 4-((3-(3,4,5-trimethoxyphenyl)propanoyl)oxy)benzoate
ß-ME	ß-mercaptoethanol
BCA	bicinchoninic acid
BCN	bicyclo[6.1.0]nonyne
Boc	*tert*-butoxycarbonyl
$CaCl_2$	calcium chloride
$CDCl_3$	deuterated chloroform
D3	3-(non-8-en-1-yl)-4-(pent-4-yn-1-yl)oxetan-2-one (ß-lactone inhibitor)
DCM	dichloromethane
DIPEA	diisopropylethylamine
DMBO	di(methoxybenzo)bicyclo[6.1.0]nonyne
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DMSO-d_6	deuterated dimethylsulfoxide
DSBU	bis(2,5-dioxopyrrolidin-1-yl) 4,4'-(carbonylbis(azanediyl))dibutyrate
DSSO	bis(2,5-dioxopyrrolidin-1-yl) 3,3'-sulfinyldipropionate
dNTP	equimolar desoxynucleotide mix
DTT	dithiothreitol
Et_2O	diethyl ether
EtOH	ethanol
EtOAc	ethyl acetate

FA	formic acid
Fmoc	fluorenylmethyloxycarbonyl
GDP	guanosine diphosphate
GTP	guanosine triphosphate
H_2O	water
HCl	hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IAA	iodoacetamide
IPTG	isopropyl-β-D-thiogalactopyranoside
KCl	potassium chloride
KI	potassium iodide
LiCl	lithium chloride
Me	methyl
MeOH	methanol
MES	2-(*N*-morpholino)ethanesulfonic acid
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulfate
NaCl	sodium chloride
$NaHCO_3$	sodium hydrogencarbonate
NAM	nicotinamide
NaOH	sodium hydroxide
NEt_3	triethylamine
Nor	norbornene
PEI	polyethylenimine
PMSF	phenylmethanesulfonyl fluoride
SDS	sodium dodecyl sulfate
TCEP	tris(2-carboxyethyl)phosphine
TAMRA	tetramethylrhodamine
TBABr	tetrabutylammonium bromide
TBAF	tetrabutylammonium fluoride
TCO	*trans*-cyclooctene
TEMED	*N,N,N′,N′*-tetramethylethane-1,2-diamine
TFA	trifluoroacetic acid
THF	tetrahydrofuran

General abbreviations

2xYT	
AB	antibiotics
AI	auto-induction
bb	backbone
cfu	colony forming units
CB	Coomassie brilliant blue
CV	column volume
DA	Diels-Alder cycloaddition
DMEM	Dulbecco's Modified Eagle Medium

et al.	and others
eq	equivalent
ESI	electrospray ionization
FBS	fetal bovine serum
GCE	genetic code expansion
HD/X	hydrogen deuterium exchange
HPLC	high performance liquid chromatography
HF/HiFi	high fidelity
iEDDAC	inverse-electron demand Diels-Alder cycloaddition
ITC	isothermal titration calorimetry
LB	lysogeny broth
LC-MS	liquid chromatography mass spectrometry
MRSA	methicillin resistant *S. aureus*
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MQ	Milli-Q
MW	molecular weight
NEB	New England Biolabs
NMR	nuclear magnetic resonance
non-AI	non-inducing
NRPS	non-ribosomal peptide sythesis
OD_{600}	optical density at 600 nm
ori	origin of replication
pABA	p-aminobenzamidine agarose
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pK_a	negative decimal logarithm of the acid dissociation constant K_a
POI	protein of interest
QC	quikchange mutagenesis
RC	restriction cloning
ResQ	RESOURCE™ Q anion exchange chromatography column
rt	room temperature
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SLIM	site-directed ligase-independent mutagenesis
SLIC	sequence- and ligation-independent cloning
SOB	super optimal broth
SOC	super optimal broth with catabolite repression
T_a	annealing temperature
T_m	melting temperature
TBS	Tris buffered saline
UV	ultraviolet
WB	western blot
TXR	Texas red
XL-MS	crosslinking mass spectrometry

Abbreviations of units

Å	angstrom
Da	Dalton
d	day
Hz	Hertz
H	hour
g	acceleration of gravity (g = 9,80665 $\frac{m}{s^2}$ (relative centrifugal acceleration))
kDa	kilodalton
kV	kilovolt
L	liter
M	molar
mA	milliampere
min	minutes
MHz	megaHertz
mL	milliliter
mM	millimolar
mmol	millimol
nm	nanometer
nM	nanomolar
nmol	nanomol
ppm	parts per million
rpm	revolutions per minute
s	second
µM	micromolar
µmol	micromole
V	volt
(v/v)	volume per volume
(w/v)	weight per volume
(w/w)	weight per weight

LIST OF FIGURES

LIST OF TABLES

LIST OF SCHEMES

1. INTRODUCTION

1.1. Nature of Molecular Biology

"[...] not so much a technique as an approach, an approach from the viewpoint of the so-called basic sciences with the leading idea of searching below the large-scale manifestations of classical biology for the corresponding molecular plan" (Astbury, 1961, p. 1124)[2].

With this sentence William Astbury[b] explained in *Nature* in 1961 his understanding of the term *molecular biology.* He devised a really smart definition while keeping the term relatively abstract so that it retains room for personal interpretation of the reading public. Moreover, it is quite hard to find an appropriate description for this interdisciplinary research field which satisfies the myriads of aspects comprised in molecular biology. Thousands of researchers have contributed to the continuous advancement in this field, including geneticists, structural biologists, chemists and biochemists up to bioinformaticists.

1.1.1. Central Dogma of Molecular Biology

An essential contribution to the modern understanding of molecular biology was made by pioneers Francis Crick and James Watson through their discovery of the double-helix model as molecular structure of DNA in 1953[4] when they already speculated:

" It follows that in a long molecule many different permutations are possible, and it therefore seems likely that the precise sequence of the bases is the code which carries the genetical information" (Watson & Crick, 1953, p.946)[4]

The inspiration[c] for the hypothesis of Watson and Crick came from a DNA X-ray diffraction image ("photo 51") that visualized the double-helical conformation of the DNA-macromolecule and was obtained in the lab of Rosalind Franklin.[5] Not long after in 1958, the statement of Watson and Crick was further completed by a theory of Crick about the flow of information within a biological system which became famous as *Central Dogma of Molecular Biology*.[1] The regular flow starts with the DNA, which can either be copied (replicated) to DNA or transcribed into RNA. The information provided by RNA can then serve as template for protein synthesis (translation) (Figure 1).

Its message of the general transfer of biological information within a living system is still valid today but gives the impression of a very simple and one-dimensional system. I included some more aspects in the flow of information to illustrate an updated central dogma from a more modern point of view. Beside the general transfer processes including DNA replication, transcription and translation, the "updated dogma" incorporates the knowledge gained in the

[b] A biochemist who published a DNA X-ray diffraction pattern which provisionally indicated a repetitive motif within the DNA structure. (3. Astbury, W. T.; Bell, F. O., X-Ray Study of Thymonucleic Acid. *Nature* **1938,** *141* (3573), 747-748.)

[c] Actually, the DNA X-ray diffraction image (known as "photo 51" from 1952) of Rosalind Franklin first visualized the double-helical conformation and was the inspiration of Watson & Crick in rewiring the molecular structure of DNA in 1953. She published "photo 51" in 5. Franklin, R. E.; Gosling, R. G., Molecular Configuration in Sodium Thymonucleate. Ibid.**1953,** *171* (4356), 740-741.

last 60 years about e.g. special transfer and processing mechanisms of biological information *in vivo* and *in vitro* as well as regulatory processes between the molecules of life (Figure 1). In 1965 the direct translation from DNA to protein was already successfully achieved *in vitro*.[6, 7] RNA can further undergo reverse transcription (e.g. retrovirus)[8] back to DNA or replicate itself (e.g. virus or RNA silencing)[9]. After translation proteins interfere with DNA transcription (e.g. histones)[10, 11], replication (e.g. DNA polymerases)[12], RNA processing[13-15], protein modification (e.g. ubiquitinylation)[16, 17] and metabolite production[18]. The metabolome *vice versa* interacts with the proteome (Figure 1). However, maintenance of the integrity within a living system resembles a highly complex network which requires an exact coordination of all involved components and mechanisms and all this is part of molecular biology.

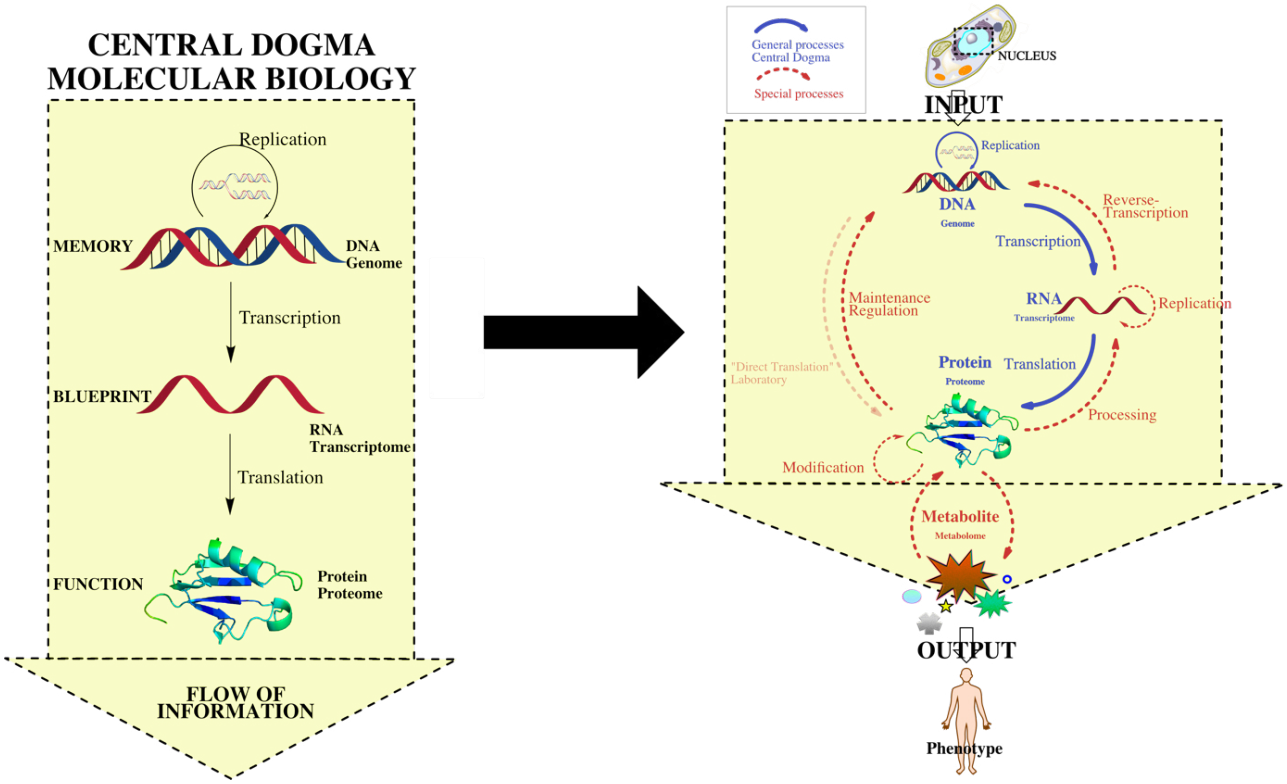


Figure 1: Illustration of the central dogma of molecular biology by Francis Crick and an "updated" version inspired by the achievements of the last 60 years in research field of molecular biology.

The following part focusses in detail on one of these biological transfer processes – the translation of proteins. This process, which converts the mRNA-coded information into a finished protein represents the basis for understanding the concepts of genetic code expansion (GCE).

1.1.2. Proteins – The working units of the cell

Proteins are the executive units within the cells responsible for all kinds of regulatory mechanisms to maintain homeostasis and cell integrity. Dysfunction within this molecular machinery is tipping over the balance of the sensible and fine-tuned cellular processes resulting in pathogenesis. Thus, it is extremely important that the process of protein synthesis is reliably free from defects (error rate around 10^{-4} per codon[19, 20]), but still fast to cover the cellular demand. Proteins consist of 20 proteinogenic amino acids which are organized in a polypeptide chain (primary structure) connected via peptide bonds between the carboxyl-group of the

preceding residue and the amino group of the following amino acid. The polypeptide chain folds into a distinct conformation depending on its sequence - the secondary structure – which is organized in the tertiary structure and further associates with other proteins to complexes which is then called quaternary structure.[21]

In principle the concept of protein biosynthesis is the same in all kingdoms of life. The DNA contains the sequential information of the protein stored in form of nucleobase-coded (adenine = A, thymine = T, guanine = G, cytosine = C) sequences following the rules of Watson & Crick base pairing (A and T; G and C). The DNA-code is copied into the nucleotide sequence of the mRNA, using uracil (= U) instead of T, and finally translated into the protein sequence consisting of the 20 proteinogenic amino acids.[22] The transfer of the four-nucleotide alphabet into a 20 amino acid alphabet requires a special translation which is enabled by tRNAs (Figure 2B). The tRNA molecule is a single chain of 73 – 93 ribonucleotides folded into a L-shaped cloverleaf structure with around half of the nucleotides undergoing base pairing forming double helical regions stabilized via hydrogen bonds. These translator molecules rewire the base-code (genetic code) by converting each mRNA-base-triplet (codon) into the respective amino acid which is attached at the free 3'-end of the CCA-sequence of the tRNA (aminoacylated tRNA). The accessible tRNA anticodon recognizes a complementary triplet-codon through base pairing and is located opposite to the amino acid binding site of the tRNA molecule.[23] The universal genetic code is made of four nucleobases leading to 64 triplet-codons (Figure 2A). The code is degenerated as for most amino acids synonymous codons (synonyms) exist.[22] Most synonyms only differ in the last base of the triplet-codon, which is related to the "wobble" hypothesis: the base pairing between the first base of the tRNA anticodon and the third base of the codon does not follow the general Watson & Crick rules (steric freedom = "wobble"; U and A or G, G and U or C).[23]

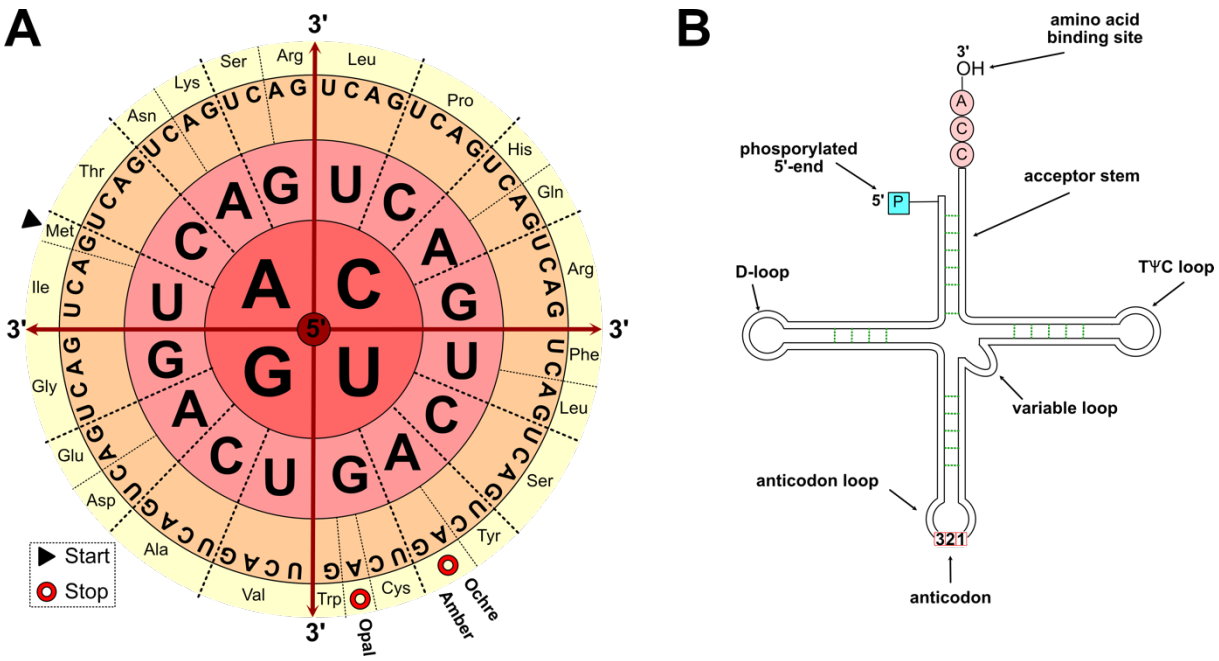


Figure 2: The genetic triplet code of the mRNA is recognized through the tRNA anticodon. A: Universal translation of the nucleobase triplets into amino acids shown as schematic illustration of a codon wheel to read from 5' to 3' direction. B: Schematic structure of a tRNA molecule. The anticodon is indicated as three red rectangles for the three (1, 2, 3) anticodon positions. The green prints symbolize hydrogen bonds of the base pairing nucleotides.

All components (mRNA, aminoacylated tRNAs) are brought together at the ribosome, a ribonucleoprotein complex consisting of two subunits (30S and 50S in prokaryotes, 40S and 60S in eukaryotes) harboring three distinct tRNA binding sites (Figure 3). The A- (aminoacyl-) site allows the entry of new aminoacylated tRNA molecules, the P- (peptidyl-) site harbors the tRNA attached to the nascent peptide chain and deacylated tRNAs are released at the E- (exit-) site. It reads the mRNA in 5' to 3' direction thereby guiding the protein synthesis from the N-terminus to the C-terminus while catalyzing peptide bond formation between the amino acid at the 3'-end of the tRNA molecule in the A-site with the peptidyl-tRNA at its P-site. The translational process is initiated by the start-codon AUG in the mRNA recognized by a special initiator tRNA decoding methionine (in prokaryotes it is a formylated methionyl-tRNA). Many more accessory proteins contribute to protein translation such as elongation factors (EFs): EF-Tu guides the aminoacylated tRNA to the A-site of the ribosome, while EF-G mediates ribosome translocation promoted through enthalpy release during GTP hydrolysis. The termination of the translation process and the release of the polypeptide chain is triggered by stop-codons (UAG = amber, UAA = ochre, UGA = opal) which are recognized by release factors (RF-1 binds UAA or UAG; RF-2 binds UAA or UGA). The whole complex finally dissociates in a GTP consuming reaction which involves EF-G and the ribosome recycling/release factor (RRF).[23]

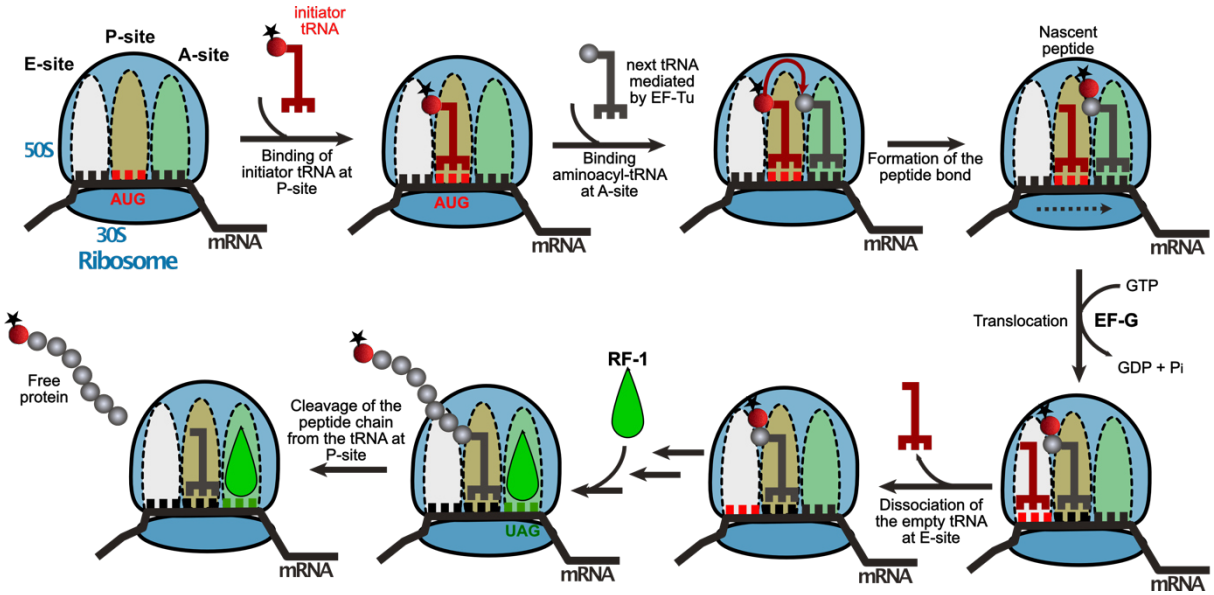


Figure 3: Scheme of the ribosomal protein translation of prokaryotes. The ribosome consisting of two subunits (50S and 30S) is colored in blue and harbors the E-, P- and A-site for tRNA interactions. In red is shown the formylated (black star) initiator methionyl-tRNA which recognizes the AUG-codon, other tRNAs are colored in grey. Red sphere represents the methionine amino acid of the initiator tRNA, grey spheres symbolize other amino acids. Red arrow indicates the formation of the peptide bond between the two amino acids of the tRNAs within the A- and P-site. Dashed black arrow symbolizes the translocation of the ribosome in 3' direction of the mRNA. RF-1 is colored in green and binds to stop-codons to terminate the process.

To ensure that each codon is translated into the correct amino acid the tRNA molecules have to be charged with the correct residues. The enzyme family of aminoacyl-tRNA synthetases (aaRS) fulfills this function and has to accomplish two requirements: 1. recognition of the correct tRNAs and 2. aminoacylation with the correct amino acid. One cell has at least 20 aaRS and each of them recognize their cognate tRNA molecules (for some amino acids several tRNAs exist). The aaRS can be grouped in class I (mainly monomers) and II (mainly dimers) enzymes depending on the location of the tRNA binding site within the aaRS resulting in

different conformations of the CCA-arm upon tRNA binding. Furthermore, the ATP binding occurs in different conformations and the attachment of the amino acid occurs at the 2'-hydroxyl group for class I aaRS whereas class II enzymes couple it to the 3'-hydroxyl group of the 3'end. Most aaRS harbor two active centers (activation and editing site, Figure 4C). In the activation site the esterification of the carboxyl group of the amino acid with the 3'(2')-hydroxyl group of the tRNA takes place under adenosine triphosphate (ATP) consumption (Figure 4A). The first step is the activation of the amino acid by formation of a mixed anhydride termed aminoacyl adenylate (aminoacyl-AMP; Figure 4B) from ATP between the phosphoryl group of adenosine monophosphate AMP and the carboxyl group of the amino acid. The second step is the transfer of the aminoacyl-AMP to the 3'(2')-hydroxyl group of the respective tRNA molecule. Both reaction steps are carried out at the same enzyme. The resulting amino acid-tRNA esters (aminoacyl-tRNA) are the activated intermediates for the protein synthesis which are necessary due to thermodynamically unfavorable nature of peptide bond formation. To guarantee that only correctly charged aminoacyl-tRNAs contribute to translation, the aaRS are highly specific towards certain amino acids. In most aaRS specificity is achieved by a second active center (editing site) which can hydrolyze mischarged amino acids from the tRNA. The quality control of aminoacylation resembles a two-checkpoint system. The aaRS activation site harbors certain coordination sites especially adapted for their respective substrate (often moieties of the amino acid side chain are involved). Moreover, amino acids which are larger than the original substrate are sterically discriminated. After aminoacylation the aminoacyl-tRNA has to pass a second quality checkpoint (editing site). Normally, correctly aminoacylated tRNAs are sterically hindered to fit into the editing site whereas tRNAs which are mischarged with amino acids smaller than the original one can enter and undergo hydrolyzation (Figure 4C). The tRNA recognition cannot be generalized for all aaRS but it always requires the participation of different tRNA features. Common recognition motifs involve the tRNA-anticodon, the anticodon loop, the acceptor stem or the variable loop (Figure 2B).[23]

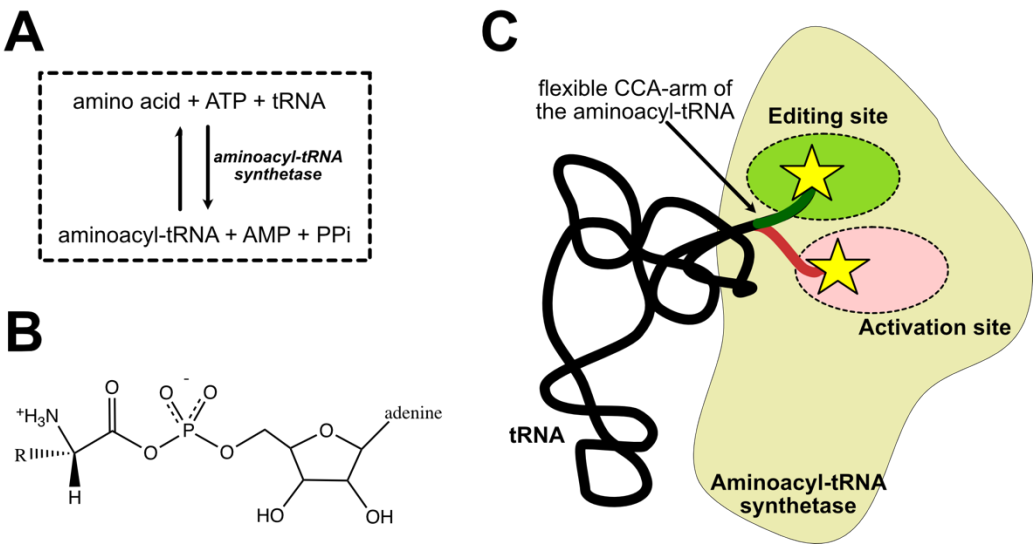


Figure 4: Aminoacyl-tRNA synthetases connect the tRNA and its respective amino acid. A: The aaRS catalyzes the esterification of the tRNA and its amino acid under ATP consumption. B: Activated amino acid (aminoacyl-AMP) in form of a mixed anhydride between the phosphoryl group of AMP and the carboxyl group of the amino acid. C: Schematic illustration of the proof-reading mechanism of the aaRS towards aminoacyl-tRNAs. (R = side chain of an amino acids, PP_i = pyrophosphate)

1.2. Genetic code expansion

The elucidation of molecular mechanisms is accomplished by studying proteins in their natural environment to address questions concerning their exemplary localization, function and interaction. This requires the manipulation of the target protein with biological and chemical approaches and can be achieved by GCE methods. As already described above, the universal genetic code is composed of triplet-codons formed by four different nucleobases. Altogether, 64 possible combinations decode the 20 canonical amino acids and three stop-signals, thus the genetic code is degenerate. The degeneracy offers the potential to reassign some codons for additional building blocks (e.g. unnatural amino acids (uAA)) without the loss of encoding any existing one. That phenomenon can also be found in nature where some archaea, bacteria and eukaryotes exhibit a 21. (selenocysteine, Sec, U) and a 22. (pyrrolysine, Pyl, O) proteinogenic amino acid to expand the standard genetic code. For the incorporation of these amino acids certain aaRS/tRNA pairs (SecRS/$tRNA_{ACU}$ and PylRS/$tRNA_{CUA}$) were evolved and original stop-codons were reassigned to decode the non-canonical amino acids.[24] This principle is the basis of the genetic code expansion (GCE) approach which drastically emerged in the last 20 years. GCE aims to enlarge the standard genetic code by new unnatural amino acids using aaRS/tRNA pairs and specific codon reassignment. Protein manipulation is no longer restricted to functionalities of the 20 natural amino acids but can be widely expanded. This paves the way for the revolution of protein studies and the emergence of new chemistries.

1.2.1. Methods to expand the genetic code

The modification of proteins can be performed in two different manners, (i) residue-specific or (ii) site-specific, and both approaches harbor their own requirements, benefits and fields of application.

Residue-specific incorporation approaches for uAAs are proteome-wide occurring processes which rely on a quite simple method without altering the genetic information or translational components of the host organism. The incorporation of an uAA analog competes with the natural substrate for the aaRS and is consequently introduced into proteins during translation. This approach has three requirements for *in vivo* applications (i) the uAA must be cell-permeable to be present during protein translation, (ii) the uAA must be accepted by the host aaRS which normally recognizes only the respective natural amino acid and (iii) the corresponding natural amino acid is encoding the genetic information of the protein.[25] Because the uAA is a surrogate of the natural amino acid the incorporation levels can be fine-tuned by adjusting the available ratio of both components enabling partial or nearly quantitative replacement.[26] For *E. coli* the most common method is termed selective pressure incorporation (SPI) using an auxotrophic expression host (Figure 5). Auxotrophic cell strains lack certain elements involved in biosynthetic pathways of amino acid genesis so that they depend on additional feeding of the respective amino acids. To initiate the incorporation of the uAA analog the cells are cultivated in defined medium containing the respective uAA instead (or partially replaced) of the natural amino acid. The endogenous (or engineered) aaRS/tRNA pair which usually is responsible for the depleted natural amino acid now accepts the uAA due to its

promiscuity and leads to incorporation during the translational process thereby enabling a global replacement and the production of alloproteins[d].

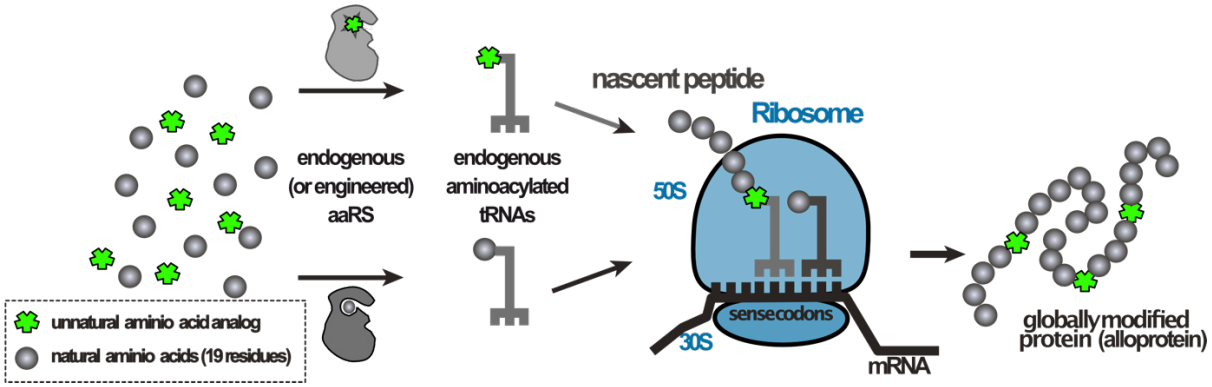


Figure 5: Residue-specific incorporation of an uAA using selective pressure. Auxotrophic strains are cultivated in medium where the natural amino acid (grey spheres) is depleted and replaced by its surrogate uAA (green rectangular star). Due to the selective pressure the uAA is translationally incorporated by the host's endogenous aaRS/tRNA pair which normally accepts the depleted natural amino acids and leads to proteome-wide replacement in response to the corresponding codon.

Representatives of aaRS/tRNA pairs which are often applied for SPI of uAAs are the methionyl-aaRS/tRNA pair[27-29], the leucyl-[30] and isoleucyl-aaRS/tRNA pairs[31] or the phenylalanyl-aaRS/tRNA pair[32]. Exemplary for this is the report of residue-specific incorporation of the leucine analog trifluoroleucine by the LeuRS/tRNA pair using a leucine-auxotrophic *E. coli* host to stabilize coil-coil protein domains such as leucin-zipper domains.[30] Proteins for crystallographic studies are often produced in methionine-auxotrophic strains to replace endogenous methionine by selenomethionine within the protein. The structural elucidation takes advantage of the anomalous signal from the selenium during data acquisition which enables the solution of the phase problem and enormously facilitates the structure determination.[33] Also protein-protein interaction studies were performed using two photo-crosslinking unnatural amino acids (photoleucine and photomethionine).[34]

However, these methods have drawbacks, especially their limited flexibility in the acceptance of uAA structures. Due to use of endogenous aaRS/tRNA pairs specific for natural amino acids usually uAA analogs bearing similar structural features are recognized and those which immensely differ are excluded or require aaRS engineering. In addition, the modification does not occur site-specifically but globally in target protein and entire cellular proteome that involves the danger of toxic site-effects and precludes exclusively labeled proteins or protein sites. The site-specific incorporation of uAAs into proteins circumvents difficulties associated with multi-labeling and substrate flexibility through the use of orthogonal aaRS/tRNA pairs for *in vivo* and *in vitro* applications. The most common approach within the field is the suppression of the amber codon (amber suppression) and is the central part of the following chapter 1.2.2.

For completeness it should be mentioned that also cell-free protein synthesis (CFPS) methods were developed. These *in vitro* approaches are commonly performed in cell extracts (e.g. *E. coli* or rabbit reticulocytes) containing the translational components but have to be supplemented with cofactor, salts, nucleotides, amino acids, DNA template and an energy-regeneration system.[35] There are several ways to adapt the system for uAA incorporation. One example focuses on the depletion of all cognate tRNAs for the respective natural amino acid (which

[d] Protein that contains uAAs

should be replaced by the uAA), as well as the corresponding aaRS and then the addition of artificially aminoacylated tRNAs with the uAA.[35, 36] CFPS is especially interesting for the production of cytotoxic[37] or membrane proteins[38] and the incorporation of non-cell permeable or toxic[39] uAAs, but also has its limitations concerning post-translational modifications or protein-folding issues.[40]

1.2.2. Amber suppression

1.2.2.1. Principle & requirements

Today, amber suppression is a widely used GCE method to introduce unnatural amino acids site-specifically into proteins in living organisms[41] but also in CFPS[42, 43]. The dawn of this method was in the lab of Peter Schultz with the incorporation of O-methyl L-tyrosine during recombinant protein expression in living *E. coli* cells.[44] During the last 20 years the tools for amber suppression immensely progressed with the result that nowadays more than 200 different uAAs can be incorporated into several species beginning from single-cell organisms like *E. coli*[44], *Bacillus subtilis*[45], *Mycobacterium tuberculosis*[46], yeast[47] or mammalian cells[48, 49] up to multicellular organisms e.g. *Caenorhabditis elegans*[50] or the plant *Arabidopsis thaliana*[51, 52]. The approach uses an orthogonal aaRS/tRNA pair which acts independently from endogenous aaRS/tRNA pairs. It accepts the desired uAA, charges it onto its cognate tRNA and leads to its incorporation during the translational process in response to the amber stop-codon (UAG) on the mRNA (Figure 6A). Orthogonal aaRS/tRNA pairs can be manipulated to afford the recognition of a certain uAA and to read-through the amber codon. The major obstacle of the amber suppression approach is the competition between TAG-suppression by the aminoacylated suppressor tRNA and the TAG-mediated translational termination by RF-1.

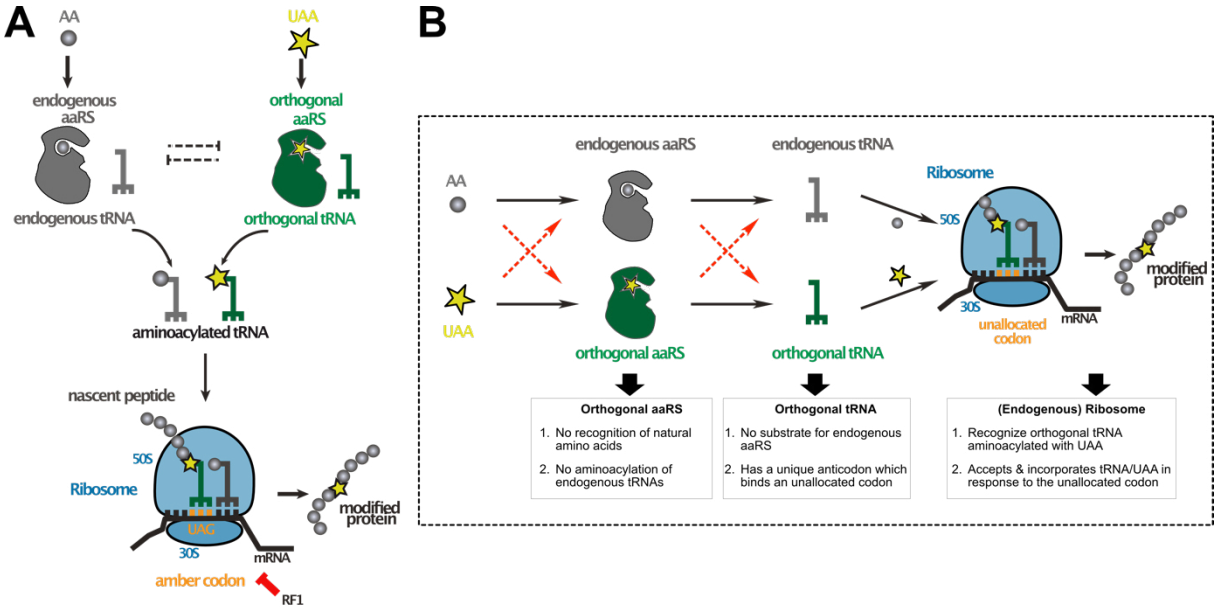


Figure 6: Amber suppression enables the site-specific incorporation of uAAs during translation. A: An orthogonal aaRS (green) exclusively accepts and aminoacylates its cognate orthogonal tRNA (green) with the uAA (yellow star). The charged orthogonal tRNA is now a substrate for the ribosomal translation and recognizes the amber codon on the mRNA thus enabling the incorporation of the uAA into the growing polypeptide chain. There are no cross reactions between the orthogonal pair and endogenous aaRS/tRNA pairs (grey) which recognize natural amino acids (grey sphere) but the competition with RF1. B: Requirements for the orthogonality of an aaRS/tRNA pair in amber suppression approaches.

GCE via amber suppression to incorporate uAAs into proteins requires three central prerequisites (i) an orthogonal aaRS/tRNA pair which recognizes solely the uAA, (ii) an unused (blank) codon which can be reassigned to decode the uAA and (iii) the compatibility with the translational machinery of the host cell (Figure 6B). To fulfill orthogonality of an aaRS/tRNA pair it is important that the aaRS does not aminoacylate any endogenous tRNA beside its cognate tRNA and that it does not accepts one of the natural amino acids. Moreover, the orthogonal tRNA must not be a substrate for the endogenous aaRS and has to display a unique anticodon which binds an unallocated codon (e.g. amber codon). Additionally, the aminoacylated orthogonal tRNA has to be accepted by the endogenous ribosome which can introduce the uAA into the nascent peptide chain during translation. In summary, the orthogonal aaRS/tRNA pair acts independently and without cross-reaction from the endogenous aaRS/tRNA pairs and can be seen as a complementation of the existing aaRS/tRNA sets. To satisfy all these requirements tools have been developed to adapt the components involved in amber suppression approaches as well as the expression host itself.

1.2.2.2. Unnatural amino acids & biocompatibility

The first central point for incorporating an uAA during *in vivo* amber suppression is the amino acids itself. *In vivo* approaches require biocompatibility of the uAA meaning that it does not cause any cytotoxic effects. For chemically synthesized uAAs a challenging issue is the uptake of the uAA into the living organism, as it needs to pass through the cellular barrier (e.g. cell wall or cell membranes). Amino acids are normally too large for passive diffusion across the cellular membrane and require at least the passive transport through the pore of a transmembrane proteins (channel protein) or the active uptake by cellular transporters (ATP-binding cassette transporters[53]).[54, 55] An active intake based on the *E. coli* dipeptide transporter DppA is reported by using a propeptide strategy for *O*-phosphotyrosine (pTyr) and 4-phosphomethyl-L-phenylalanine (Pmp) as dipeptide with N-terminal lysine (Lys-pTyr and Lys-Pmp).[56] DppA can transport dipeptides which display a natural amino acid (e.g. lysine) at the N-terminal position nearly regardless of the C-terminal residue.[57-59] Once entered into the cytosol the dipeptide is cleaved by unspecific intracellular peptidases yielding the free amino acids.[56] Alternatively, the uAA uptake in *E. coli* could be increased by engineering periplasmatic binding proteins, which recognize the uAAs and enable active cell import.[60] Furthermore, it is advantageous if the uAA displays good solubility which can also be improved by formulation as dipeptide[56]. Further it is reported that the protection of the carboxyl group of the UAA with a methyl ester improves the uAA uptake and is cleaved off by cytosolic esterases.[61, 62] An alternative approach to the chemical synthesis of uAAs is the *in vivo* production by engineering biosynthetic pathways of the host organisms as reported for *p*-aminophenylalanine[63] and the terminal-alkyne bearing amino acids propargyl-glycine (Pra)[64].

The second important element is the base triplet which should encode the uAA. The standard genetic code provides three codons which are unassigned to an amino acid but responsible for the termination of the translational process (blank codons/stop-codons). As for termination one stop-codon is actually sufficient, they are good candidates for reassignment towards encoding an uAA. Amber suppression makes use of the least used stop-codon (UAG = amber) which normally terminates translation through binding of RF-1.[65, 66] But also other non-sense suppression approaches using the other two stop codons (UGA = opal[67, 68]; UAA = ochre[69, 70])

were successfully applied to decode uAAs. The combination of amber and ochre suppression was shown for the incorporation of the two uAAs, N^{ε}-Boc-L-lysine (BocK) and *p*-acetylphenylalanine (pAcF), into the same protein in *E. coli*[71] and further in mammalian cells using *O*-methyltyrosine (OMeY) and BocK[72]. Even the simultaneous incorporation of three distinct uAAs into the same protein using the suppression of all three stop-codons and three orthogonal aaRS/tRNA pairs was reported.[73] The termination of the protein translation was a bit problematical in this case as all stop-codons were allocated to uAA incorporation and the consequence was a heterogeneous C-terminus from partial uAA incorporation. This was overcome by the genetic introduction of a C-terminal cleavage system (TEV-site in frame with the TEV protease) followed by multiple consecutive stop-codons in frame with the POI to generate a homogenous protein sample.[73]

In addition to the non-sense codon suppression, efforts have been made to reassign rare sense-codons, which is always challenging due to the degeneracy of the genetic code, except for tryptophan and methionine which are each encoded by one single base-triplet. The AGG-codon which usually is allocated to tryptophan is a common candidate in sense-codon suppression. One strategy to reassign AGG-codons is based on the competition between an orthogonal, engineered pyrrolysyl-aaRS/$tRNA_{CCU}$ (PylRS/PylT) pair and the endogenous arginyl-aaRS/tRNA pair combined with the use of minimal medium in order to reduce arginine availibilty and co-expression of antisense RNA to silence the host arginine-$tRNA_{CCU}$ (*argW* gene). In suppression tests with BocK as uAA and GFP-149AGG as model protein 83 % of produced protein contained the uAA instead of arginine. Nevertheless, the expression yielded a heterologous mixture of proteins which have either the uAA or arginine incorporated at the respective site.[74] Another idea for AGG-reassignment focused on an engineered *E. coli* strain (AGG-21.2) harboring replacements of all AGG codons in essential genes and the knock-out of genes which encode the tRNAs recognizing the AGG codon (*agrU* and *argW*).[75] The UUU-codon (normally encodes phenylalanine) was further used for the incorporation of the uAA L-3-(2-naphthyl)alanine (Nal) in *E. coli* by using a mutant phenylalanyl-aaRS/$tRNA_{AAA}$ (yPheRS/$tRNA_{AAA}$) pair from yeast. The authors took advantage of the fact that phenylalanine is encoded by two codons (UUU and UUC) which are read by a single tRNA (anticodon GAA) enabled through G-U wobble base-pairing. The endogenous $tRNA_{GAA}$ was outcompeted by the mutant yeast $tRNA_{AAA}$ pair for binding of the UUU-codon thus NaI was incorporated instead of phenylalanine.[76] In some organisms there are no special aaRS for glutamine as it is synthesized on their cognate tRNAs. The precursor is a "mischarged" glutamate-$tRNA^{Gln}$ produced by the glutamyl-aaRS (GluRS).[77] Therefore GluRS recognizes tRNA molecules for glutamate ($tRNAs^{Glu}$) as well as glutamine ($tRNA^{Gln}$ = $tRNA_{UUG}$, $tRNA_{CUG}$). This phenomenon was used by the lab of Dieter Söll to liberate one codon of the GluRS/$tRNA^{Gln}$ system of *Methanothermobacter thermautotrophicus* by enzyme engineering. Compared to the wild type GluRS the mutant accepts solely the $tRNA_{CUG}$ isoacceptor thereby liberating the CAA-codon.[78]

Encoding uAAs by quadruplet codons is another approach what would open the possibility of an almost unlimited pool of new unallocated codons. Incorporation of uAAs in response to a four-base-pair codon was already successfully applied in *in vitro* experiments.[79, 80] The major problem which arises from the enlarged codon size are frameshifts caused by the competitive translation of the triplet codon.[74] To address that issue, Jason Chin and colleagues evolved an orthogonal ribosome (ribo-Q1) which efficiently incorporates uAAs encoded by quadruplet codons in combination with an orthogonal mRNA (o-mRNA), extended anticodon tRNAs and

respective aaRS in *E. coli*.[81] Ribo-Q1 was based on the former developed ribo-X (an orthogonal ribosome optimized for amber codon decoding on an orthogonal mRNA[82]) and operates independently to the host translational machinery. It exclusively recognizes a special o-RNA which encodes the POI containing the respective quadrupole codons and is no substrate of the endogenous ribosome. The orthogonal ribosome contains mutations within its 16S rRNA (ribosomal RNA) which are complementary to mutations of the Shine Dalgarno sequence of the o-RNA and responsible for the specific interaction. The advantage of using those parallel system for the incorporation of uAAs into a POI is that the translation of the host proteome remains unaffected of any uAA suppression effects.[81, 82]

Instead of reassigning or depleting existing codons in order to reuse them as "free" codons for GCE, alternative approaches deal with the creation of new orthogonal sense-codons via the expansion of the genetic alphabet by unnatural bases (UBs). The Romesberg lab reported unnatural base pairs (UBP), which undergo DNA replication and RNA transcription *in vitro*.[83-86] By using a semi-synthetic organism (SSO, chromosomally encodes a version of nucleoside triphosphate transporter 2 (*Pt*NTT2) for the supply with the UBs for DNA and RNA [87, 88]) with an expanded genetic alphabet consisting of six nucleotides the approach could be transferred to living cells. Besides the natural A-T/U and G-C nucleobase-pairing, a third UBP (X-Y pair) is present and undergoes replication, transcription and translation.[89-92]. The Romesberg lab further showed that the combination of an expanded genetic alphabet and an expanded genetic code is possible by using a SSO host (67-codon organism) bearing three new codons (AXC, GXT, AGX) which they used for multiple site-specific incorporation of two uAAs and one natural amino acid in combination with the respective aaRSs and their cognate anti-codon-adapted tRNAs (GYT, AYC, XCT) into GFP. Surprisingly, they reported that dual-incorporation of uAAs via their AXC- GXT-codon-system yields 4-fold more protein compared to approaches using a combined amber and ochre codon suppression.[93]

1.2.2.3. Orthogonal aaRS/tRNA pairs

In 2001 the lab of Peter Schultz reported the first orthogonal aaRS/tRNA pair applied for the incorporation of the uAA acid *O*-methyl-L-tyrosine in *E. coli*. They used the *Methanococcus jannaschii* tyrosyl-aaRS/tRNA (*Mj*-TyrRS/tRNA) pair with a mutated tRNA anticodon for the suppression of the TAG-codon in the heterologous *E. coli* host.[44, 94] The *Mj*-aaRS does not recognize any *E. coli* tRNA[95] but the *Mj*-tRNA is poorly accepted by *E. coli* aaRS[94]. This could be circumvented by using a *Mj*-tRNA with an altered acceptor stem to achieve orthogonality. The modified *Mj*-TyrRS/tRNA pair does not interfere with the *E. coli* endogenous aaRS/tRNA pairs and was therefore the first established orthogonal aaRS/tRNA pair for amber suppression with uAAs in living prokaryotic cells.[44] Since then additional orthogonal aaRS/tRNA pairs have been discovered by transplanting aaRS/tRNA pairs into heterologous expression hosts having distinct orthogonality towards respective host cells (Figure 7). The aforementioned *Mj*-TyrRS/tRNA pair is orthogonal to bacteria but not to eukaryotic organisms. In contrast, the *E. coli* tyrosyl-aaRS/tRNA (*Ec*-TyrRS/tRNA) pair is commonly used as orthogonal set in eukaryotic cells but not suitable for amber suppression in bacteria. A special position is occupied by the *Methanosarcina barkeri* (or *Methanosarcina mazei*) derived pyrrolysyl-aaRS/tRNA (*Mb/Mm*-PylRS/tRNA) pair as it displays orthogonality in both eukaryotes and prokaryotes.[96] Recently, a novel PylRS/tRNA (*Ma*-PylRS/tRNA) pair from the organism *Methanomethylophilus alvus* was found to exhibit

the same orthogonal properties as the *Mb/Mm*-PylRS/tRNA pair. By altering the *Ma*-tRNA (*Ma*-tRNA^{evolved}) even orthogonality between both *Mb/Mm*- and *Ma*-PylRS/tRNA pairs was realized.[97, 98] Combinatory experiments showed, that the *Mj*-TyrRS/tRNA pair, *Ec*-TyrRS/tRNA pair, *Mb*-PylRS/tRNA pair and *Ma*-PylRS/tRNA^{evolved} pair can act independently from each other within the respective expression host.[72, 73, 97, 99]

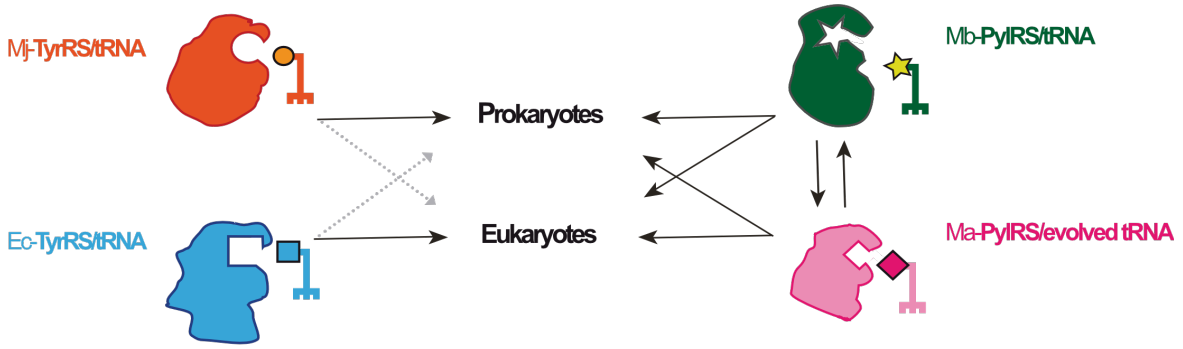


Figure 7: Orthogonality of aaRS/tRNA pairs commonly used in amber suppression approaches. The *Mj*-TyrRS/tRNA pair is orthogonal to the endogenous aaRS/tRNA set of bacteria whereas the *Ec*-TyrRS/tRNA pair displays orthogonality to eukaryotic aaRS/tRNA pairs. *Mb*-PylRS/tRNA and *Ma*-PylRS/tRNA are orthogonal in both, prokaryotes and eukaryotes and by evolving the *Ma*-tRNA they are even orthogonal to each other. Grey dashed arrows represent non-orthogonal relations, black arrows symbolize orthogonal combinations.

The PylRS/tRNA pair from *Methanosarcina* species was rapidly established as gold standard in amber suppression approaches not only because of its dual orthogonality in bacteria and mammalian cells but for some more beneficial features. Compared to the *Mj*-TyrRS/tRNA pair or *Ec*-TyrRS/tRNA pair, it displays no activity towards one of the 20 canonical natural amino acids but pyrrolysine (22nd amino acid, non-canonical). Therefore, the use of PylRS/tRNA pairs does not require efforts to prevent the recognition of natural amino acids when applied in hosts where pyrrolysine is absent. Moreover, the *Mb/Mm*-tRNA (*Mb/Mm*-PylT) anticodon harbors the CUA motif for suppression of the amber codon by nature. Compared to bacterial aaRS genes (encoded by two separate genes each for one domain) the genetic information of the *Mb/Mm*-PylRS is organized in one single *pylRS* gene coding two distinct domains which are both necessary to achieve full catalytic activity of the enzyme *in vivo*.[100, 101] The N-terminal domain (NTD) is involved in *Mb/Mm*-PylT binding via interaction with the variable loop ensuring binding specificity but is prone to aggregation at higher concentrations resulting in low expression levels in the host organism due to aggregation and misfolding.[71, 102, 103] *Mb*-PylRS lacks 35 amino acid in the N-terminal domain but does not generally differ from *Mm*-PylRS otherwise (see ClustalX aligment Figure 18). Interestingly, Chin and co-workers recently identified several new functional PylRS enzymes which are organized in one single catalytic domain. One of those PylRS lacking the NTD is the aforementioned *Ma*-PylRS/tRNA pair.[97, 104] The C-terminal domain (CTD) of *Mb/Mm*-PylRS harbors both the substrate recognition and catalytic active sites.[101, 105] Even if it was previously not possible to crystallize the full-length enzyme due to the negative aggregating effects of the NTD, the crystal structures of the *Mm*-PylRS-CTD revealed insights into the substrate binding site.[103, 106] The hydrophobic binding pocket with a cavity for the pyrrolysine side chain is formed by seven amino acid residues (A302, L305, Y306, L309, N346, C348, and W417) (Figure 8B).[107] The interaction is non-specific for pyrrolysine as the enzyme tolerates alterations in the substrate side chain as shown for N^{ε}-Boc-L-lysine (BocK), N^{ε}-propargyl-L-lysine (PrK)[61] or N^{ε}-(1-methylcycloprop-2-enecarboxamido)-lysine (CpK) (Figure 8A)[108]. This broad promiscuity together with the

absence of an editing site within the enzyme makes the PylRS very attractive for the incorporation of uAAs.[107]

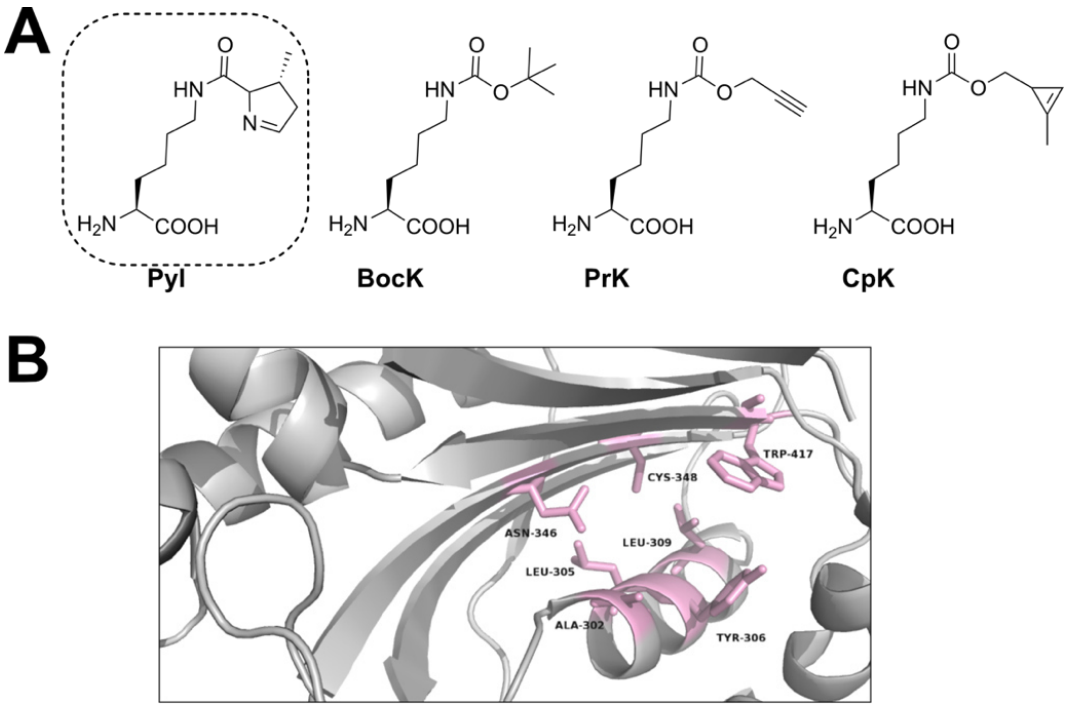


Figure 8: *Mb/Mm*-PylRS is a promiscuous enzyme. A: Structures of the original substrate pyrrolysine (= Pyl) and of the uAAs BocK, PrK and CpK which are also recognized by the *Mb/Mm*-PylRS..B: View into the binding pocket of the crystal structure of the *Mm*-PylRS-CTD (PDB: 2zin) with the seven amino acids (pink) forming the hydrophobic pocket for the substrate side chain.

Another remarkable feature of the *Mb/Mm*-PylRS is that the tRNA recognition works independently from the tRNA anticodon. Therefore, it is possible to change the anticodon sequence without disturbing the enzyme-tRNA interaction.[97, 109-112] According to that the *Mb/Mm*-PylT displays some special structural features including an enlarged anti-codon stem and a small D-loop.[113] Although *Mb/Mm*-PylRS is natively promiscuous, not all uAAs are recognized without previous alterations of the enzyme. Many evolutionary approaches have been arising which are used to develop tailor-made aaRS mutants for the desired uAA. The classical directed evolution approach for *Mb/Mm*-PylRS and *Mj*-TyrRS is based on a two-step selection process in *E. coli* where uAA incorporation is coupled to cell survival mediated by an essential or toxic protein expression (Figure 9).[114] The genetic input for the evolution comes from a aaRS library. The library consists of the aaRS gene with randomized positions commonly in the active site of the enzyme, which are usually generated by NNK-primers[e] (see 5.2.5). The number of randomized sites per library is limited by the transformation efficiency of *E. coli* (for plasmid DNA around ~ 10^8 – 10^{10} cfu per µg DNA)[115] which should not be exceeded by the library size (for NNK-based libraries: size = 32^N ; N = number of randomized positions; see also 5.2.5).

[e] degenerated primers: "N" stands for adenine, thymine, cytosine or guanine bases, "K" stands for guanine and thymine bases

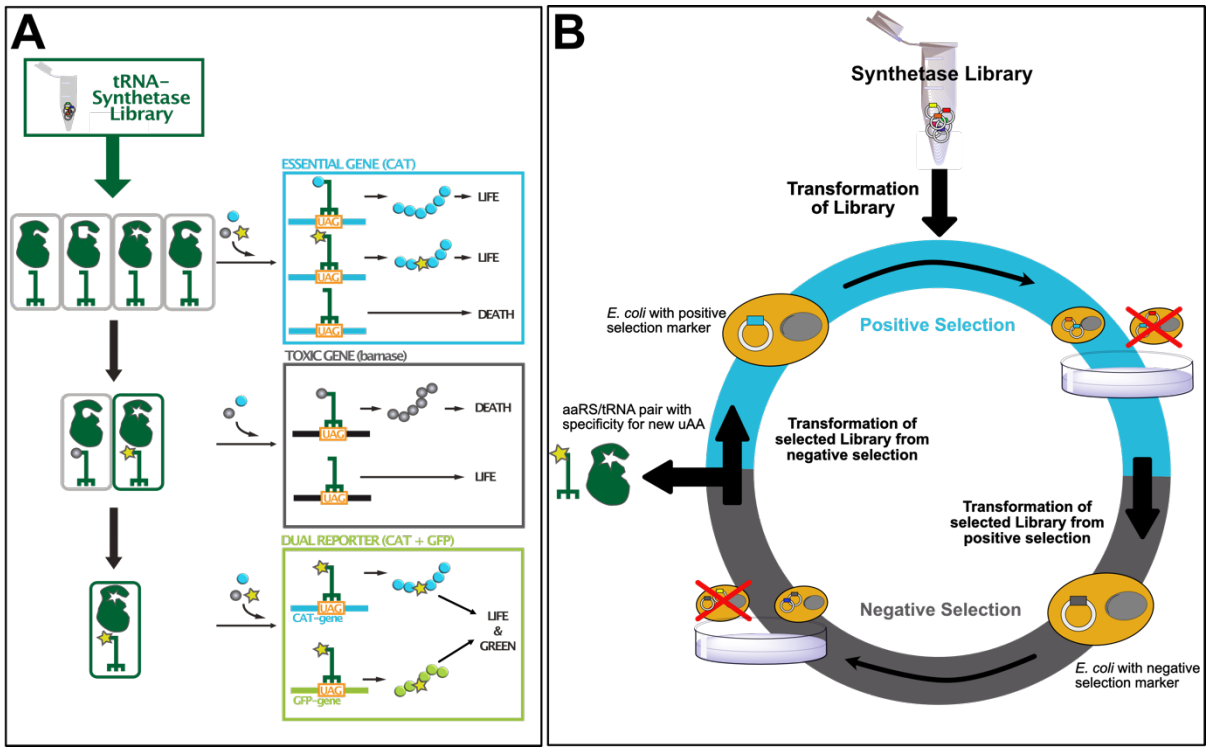


Figure 9: Classical directed evolution approach of an aaRS for new uAAs in *E. coli*. A: Illustration of positive and negative selection mechanisms based on the read-through of TAG-suppressed chloramphenicol acetyltransferase (CAT) and barnase. Optional read-out step using a dual reporter consisting of TAG-suppressed CAT and GFP. B: Scheme of the practical working steps required in the directed evolution of an aaRS.

The evolutionary experiment is initiated by the transformation of the aaRS library into *E. coli* cells containing a plasmid which encodes a selection marker. Traditionally, the chloramphenicol acetyltransferase (CAT) gene with an internal TAG-codon is used for positive selection. Cells are cultivated on medium supplemented with chloramphenicol and the respective uAA. Only colonies which successfully read-through the CAT gene either by incorporation of the uAA or a natural amino acid mediated by a respective aaRS variant survive. The aaRS library DNA of the resulting colonies serves for subsequent negative selection to exclude those aaRS variants which recognize natural amino acids. A negative selection round is performed in cells containing a plasmid which encodes the toxic barnase gene with an internal TAG-codon. In contrast to the positive selection the cells are grown on medium lacking the uAA. Here cells survive only in absence of barnase consequently these cells contain aaRS variants which do not accept any natural amino acid. To further select and easily identify best hits of the resulting colonies after negative selection a read-out step using a dual reporter can be added.[114] This is a second positive selection round based on CAT but combined with a GFP-fluorescence reporter bearing a TAG-codon at position 150. Cells which survive in presence of the uAA as well as chloramphenicol and simultaneously display a green color are promising candidates for harboring successfully evolved aaRS mutants (Figure 9A).

Over recent years many modifications of this evolution method have been developed using other negative selection markers[116] and fluorescence assisted cell sorting (FACS).[117, 118] Moreover, a function-based evolution platform for *E. coli* was reported for the evolution of acylated lysine derivates. The cell survival is coupled to the production and activity of a novel β-lactamase mutant harboring a TAG-substitution for an essential lysine. Produced enzyme harboring an acylated lysine derivate is inactive and has to be processed to lysine by *E. coli*

cobB to reinstall enzymatic activity.[119] However, conventional directed evolution approaches remain time consuming, labor-intensive and are strictly limited to the initial pool of genetic information presented within the library.

With the approach of phage-assisted continuous evolution (PACE) developed by Liu, and co-workers in 2017, a new era of evolution was initiated.[120] They used PACE to evolve a chimeric PylRS with a 45-fold improved efficiency for BocK towards the parental enzyme. PACE connects aaRS activity for tRNA-aminoacylation to the production of pIII[f] (positive selection) or pIII-neg (negative selection) which in turn mediates phage infectivity and propagation. A so-called selection phage (SP) carries all phage genes except of gene III which is substituted by the evolving aaRS gene. Centerpiece of PACE is a special *E. coli* culture machine consisting of a two-chambered flow-apparatus. This allows the continuous supply of new *E. coli* hosts for phage infection as well as the removal of cells harboring non-infectious phages. Another feature is offered by the mutagenesis plasmid encoding inducible mutagenic proteins. Compared to many conventional library-based methods that only consider positions in the aaRS binding site and which are restricted to the initial library-diversity, in PACE the entire aaRS gene is subjected to a continuous mutagenesis which dynamically alters the genetic pool during the evolution experiment.[120] With PACE the number of evolution rounds per experiment can be massively enhanced. Moreover, critical working procedures like DNA isolation and transformation between the individual selection steps are avoided.

Evolution approaches are not restricted to the enzyme component of the orthogonal pair but also focus on the tRNA in order to achieve catalytically more efficient aaRS/tRNA pairs. Concerning the tRNA, GCE requests no cross-interaction with endogenous aaRS but compatibility to the endogenous ribosome and the translational factors (e.g. EF-Tu). Söll and co-workers recently evolved an optimized $tRNA^{Pyl}$ which universally increases the efficiencies of uAA incorporation by altering regions (T- and acceptor stem residues) which are known for EF-Tu interactions.[121-124]

1.2.2.4. Recoded organisms

Since the development of GCE many efforts were taken not just to improve the orthogonal aaRS/tRNA pairs but also to adopt the respective host cells. *E. coli*-based ATM (Altered Translational Machinery) strains are one class of such manipulated hosts. They are created by substituting the genes of a native endogenous aaRS/tRNA pair (e.g. *E. coli* tyrosyl-RS/tRNA pair) by a respective, orthogonal aaRS/tRNA pair (e.g. *Saccharomyces cerevisiae* TrpRS/tRNA pair) within *E. coli* thereby liberating the endogenous aaRS/tRNA pair in order to engineer it as new orthogonal TAG-suppressor for GCE.[125, 126] The Chatterjee lab demonstrated this approach for the tryptophanyl-RS/tRNA pair (ATMW strain)[126] and the tyrosyl-RS/tRNA pair (ATMY strain)[125] of *E. coli*. They further showed that these ATM strains offer a platform for the evolution of the liberated aaRS/tRNA pair.[125, 126]

More fundamental genomic alterations were made by the Church lab that started a genome-wide codon-replacement in *E. coli* in 2011[127] yielding an optimized amber suppression host. Multiplex genome engineering (MAGE) was used to introduce site-specific TAG-to-TAA substitutions across the genome leading to partially recoded strains. These strains were used to combine the recoded regions via hierarchical conjugative assembly genome engineering

[f] pIII is an essential protein required for phage propagation and infectivity

(CAGE) in one single organism.[127, 128] They finished the project in 2013 by publishing a recoded *E. coli* strain (C321.ΔA) whose genome has 321 TAG-to-TAA replacements and a deletion of the RF1 gene (= *prfA*).[128] In this C321.ΔA strain the former TAG-stop codon is reassigned to a sense codon (e.g. for incorporation of uAAs) and translational termination via RF1 is eliminated. A second generation of recoded organisms offers the *E. coli* strain with a 57-codon genome (*rE.coli*-57) presented by the Church lab in 2016 a herculean effort, which is however not completed yet (63 % of the recoded genome is validated so far).[129] To result *rE.coli*-57, seven codons (AGA, AGG, AGC, AGT, TTA, TTG, TAG) have to be substituted by synonymous alternatives and respective tRNA genes as well as the *prfA* gene have to be deleted, thereby creating seven unallocated codons which can potentially be reassigned for uAA incorporation[129].

Replicon excision for enhanced genome engineering through reprogrammed recombination (REXER) is a method developed in the Chin lab to replace large genomic segments (up to 100 kb) by λ-mediated recombination with synthetic dsDNA segments derived via CRISPR/Cas9-initiated excision from a bacterial artificial chromosome (BAC). Iterative rounds of REXER enables genome stepwise interchange synthesis (GENESIS).[130, 131] With a step-wise conjugation-based approach recoded sections of donor cells were finally assembled to the complete genome of a recipient cell.[130, 131] These techniques lead to the generation of a recoded *E. coli* strain with a 61-codon genome (4 Mb). The recoded strain has synonymous codon compression of TCG, TCA and TAG and deletions of their respective suppressor tRNAs genes (*serU, serT*), as well as the RF1 gene (*prfA*) and is named Syn61.[130] Syn61 offers, according to the three gained blank codons, an excellent platform for GCE studies using multi-site incorporation of uAAs. However, the way to fully de-novo synthesized complex genomes has not been achieved yet. Many aspects of codon usage, gene regulation and processing are not sufficiently understood so that global alterations might cause unpredictable effects.[132-134] With the innovations of the past few years in the back it seems realistic to produce more complex organisms with fully synthetic genomes, as it was already possible for a minimal genome of *Mycoplasma mycoides*[135].

1.2.3. Applications of genetic code expansion

GCE approaches enable the elucidation of molecular mechanisms by studying proteins in their native environment to address questions concerning protein localization, interactions or activity. The site-specific incorporation of uAAs into target proteins is a powerful tool in many biological studies and enables exemplary protein labeling, protein crosslinking, protein regulation or the investigation of post-translational modifications (PTMs) of proteins. Therefore, several classes of uAAs have been developed which contain appropriate functionalities within their side chains and which can be introduced into POIs by GCE approaches.

1.2.3.1. Labeling of proteins using bioorthogonal reactions

On example for the application of GCE is the site-specific incorporation of uAAs into POIs bearing chemical moieties within their side chains which offer unique bioorthogonal functionalities. These functionalities can undergo bioorthogonal reactions with externally added chemical probes resulting in the labeling of the POI (Figure 10A). A chemical reaction

has to fulfill important requirements to be termed bioorthogonal: (i) Chemoselectivity and specificity (no cross reactions with cell components), (ii) compatibility with biological systems (non-cytotoxic, works under physiological condition, no toxic byproducts, accessibility) and (iii) optimally fast reaction kinetics. For this purpose, several types of bioorthogonal reactions have been developed which display rate constants between 10^{-5} M^{-1} s^{-1} (ketone/hydroxylamine condensation) and 10^{5} M^{-1} s^{-1} (tetrazine-*trans*-cyclooctenes (TCO) cycloaddition) (Figure 10B).[136] In 2000, the first reported bioorthogonal reaction was developed by the Bertozzi lab as they used the Staudinger ligation to site-specifically label glycans on the cell surface. They metabolically introduced azide moieties into the surface exposed glycans and reacted them with biotinylated triarylphosphines resulting in a covalently attached biotin moiety.[137]

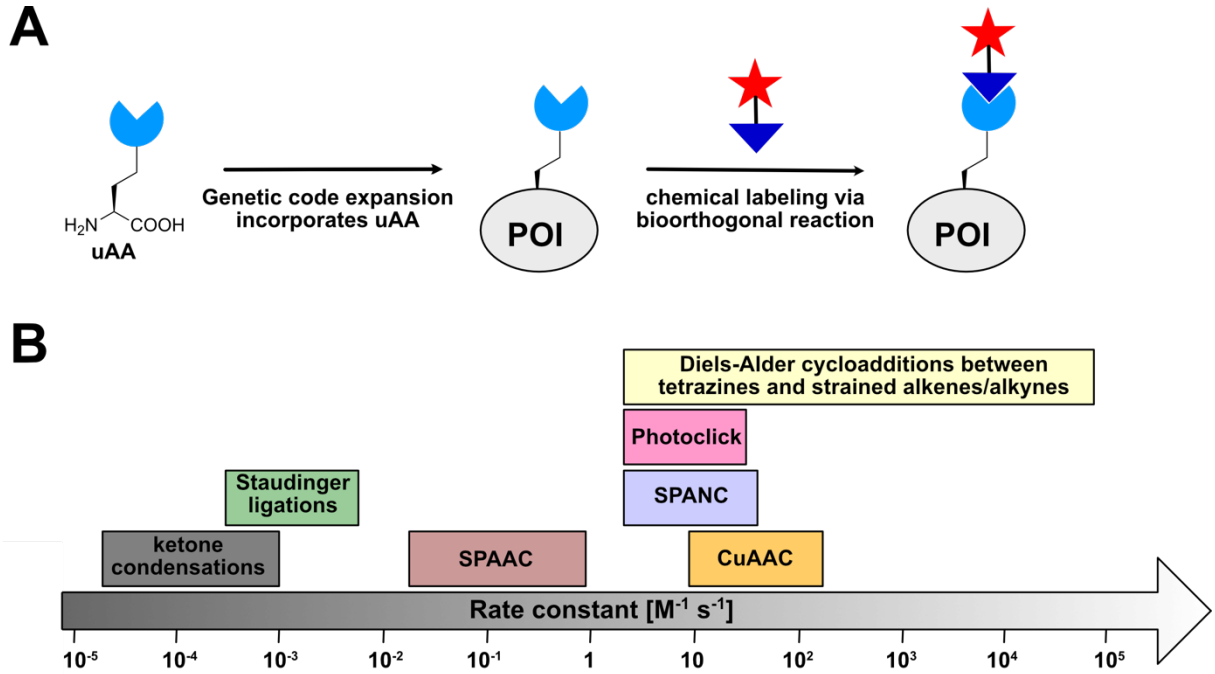


Figure 10: Protein labeling by incorporated uAAs displaying chemical handles for bioorthogonal reactions. A: The uAA bearing an exclusive chemical moiety (light blue) is site-specifically incorporated into proteins (POI, grey) via GCE. The chemical moiety of the uAA can then chemoselectively react in a bioorthogonal reaction with a chemical probe (red star) bearing a respective chemical functionality (dark blue triangle). B: Overview of rate constants of some bioorthogonal reactions used in GCE approaches. (SPAAC = strain-promoted alkyne-azide cycloaddition, SPANC = strain-promoted alkyne-nitrone cycloaddition, CuAAC = Cu^{1}-catalyzed alkyne-azide cycloaddition, Photoclick = photo-click cycloaddition)

The toolbox for bioorthogonal reactions as well as the portfolio of compatible uAAs (Figure 11) has developed immensely over the past 20 years. Beside Staudinger ligation of azides and phosphines, commonly used reactions are photo-inducible dipolar cycloadditions between alkenes and tetrazoles (the tetrazole forms a reactive nitrile imine upon light activation; photo-click chemistry), condensations of ketones and hydrazides/alkoxyamines or two types of azide-alkyne cycloadditions. One type uses the activation of Cu(I) to form triazoles from terminal alkynes and azides (= CuAAC). Due to toxic effects of Cu(I) catalysts for living cells, a second type of azide-alkyne reaction was developed using the ring-strain of an alkyne as driving force for the reaction instead of Cu(I) (= SPAAC).[136]

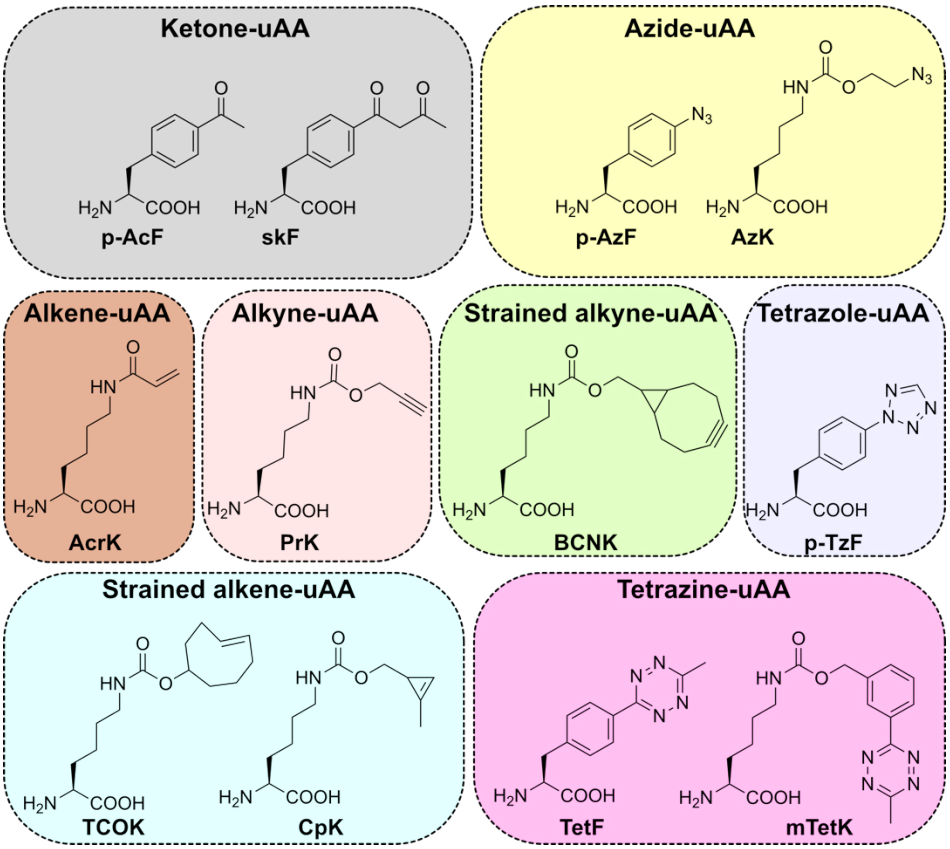


Figure 11: Selection of uAAs introduced via GCE into proteins for bioorthogonal labeling reactions.

The so far fastest bioorthogonal reactions occur between tetrazines and strained dienophiles (alkenes or alkynes) via inverse-electron-demand Diels-Alder cycloaddition (=iEDDAC). iEDDAC has been widely used to label proteins with fluorophores for imaging studies in *E. coli* and mammalian cells.[49, 138, 139] iEDDAC of tetrazines and strained alkyne was reported for super-resolution imaging of cell-surface proteins (epidermal growth factor receptor (EGFR)) as well as for intracellular proteins, β-actin and vimentin (cytoskeletal proteins), in living HEK cells. An uAA bearing a strained alkyne moiety (bicyclo[6.1.0]nonyne-lysine = BCNK, Figure 11) was site-specifically incorporated into EGFR, β-actin and vimentin which underwent reactions with external added tetrazine-fluorescein-conjugates.[140] The Lang lab published an innovative version of the iEDDAC reaction by giving it a light-controllable character (photo-iEDDAC). Therefore, they designed new chemical probes which displayed cyclopropenone-caged strained alkyne moieties (photo-DMBO[g]) which were unreactive towards tetrazines until the masked strained alkyne moiety was decaged via irradiation with UV light. Together with a site-specifically introduced tetrazine bearing uAA (mTetK, Figure 11) into *E. coli* protein (OmpC) and the photo-DMBO probe they achieved spatio-temporally controlled labeling *in vitro* and *in vivo*.[141] Beside for labeling approaches, tetrazine reactions were reported for inhibitor-conjugate/POI pair formation to control the enzymatic activity of the POI using bioorthogonal ligand tethering (BOLT) in combination with inhibitors (iBOLT). An uAA bearing a bioorthogonal functionality (e.g. BCNK with an alkyne moiety) is site-specifically introduced into the POI and reacts with an inhibitor probe displaying a respective bioorthogonal counterpart (e.g. tetrazine moiety) to form the inhibitor-conjugate/POI pair. Through the

[g] cyclopropenone-caged dibenzoannulated bicyclo[6.1.0]nonyne

covalent attachment of the inhibitor to the POI a selective inhibition of the enzyme is achieved. The installation of photoswitchable moieties (e.g. azobenzene moieties) within the inhibitor-conjugate linker enables a light-controlled mediation between an active and an inhibited enzyme state (photo-BOLT).[142]

1.2.3.2. Encoding uAAs bearing biophysical probes

The decoration of target proteins with biophysical probes was also achieved without the need of bioorthogonal reactions by the site-specific incorporation of uAAs which directly bear biophysical probes within their side chains for spectroscopic studies of proteins (Figure 12). The direct incorporation of a biophysical probe can only be realized if the biophysical group is compatible with a respective, orthogonal aaRS/tRNA pair (the size of the uAA is limited by the properties of the respective aaRS/tRNA pair).

The uAA OCF_3Phe (Figure 12A) is reported as label for studying protein dynamics of the thioesterase domain of human fatty acid synthase (FAS-TE) in nuclear magnetic resonance (NMR) spectroscopic experiments (^{19}F NMR studies).[143] Furthermore, PheCN (Figure 12A) incorporated into proteins has the potential to offer excellent probes for resonance Raman spectroscopy of protein dynamics.[144] An alternative, smart strategy to decorate proteins with fluorescent moieties without the need of externally added fluorophore-probes, is the direct, site-specific incorporation of fluorescent uAAs into proteins (Figure 12B). Representatives of such uAAs display for example coumarin (HceG[145]) or naphthalene (Anap[146]) moieties which are installed at the amino acid side chain. Schultz and co-workers used Anap to visualize the subcellular compartments of the endoplasmic reticulum (ER), Golgi apparatus and nucleus by fluorescence in different mammalian cell lines (HeLa, HEK293T, CHO). Therefore, Anap was site-specifically installed into ER-resident (Grp94), Golgi-resident (GalT1) proteins and nuclear protein histone H3 using GCE and the localization was monitored using fluorescence microcopy with two-photon exitation.[146]

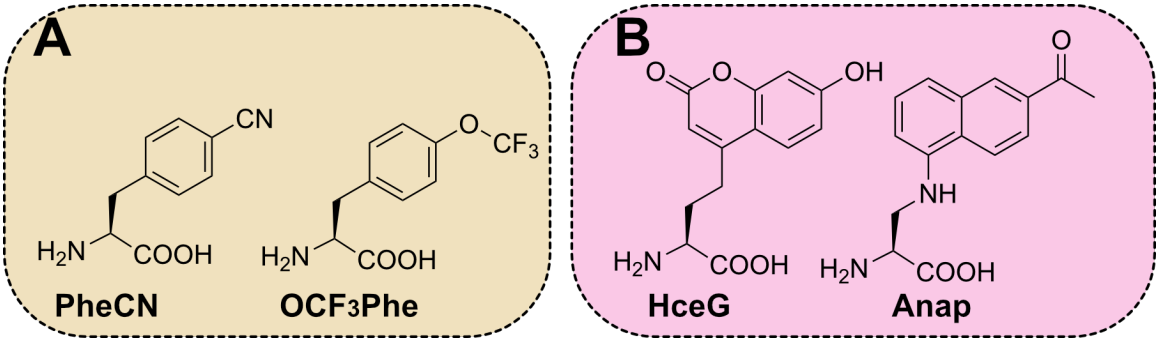


Figure 12: Structures of uAAs which serve as biophysical probes in biological studies for A: spectroscopy studies and B: fluorescent labeling.

1.2.3.3. Studying post-translational protein modifications

Cellular proteins are subjected to reversible, modificatory processes which install post-translational modifications (PTMs) onto target proteins. PTMs are key features in regulatory mechanisms and mediate protein function, localization and degradation.[147] PTM studies are often limited by the accessibility of homogenously modified protein samples. GCE offers a method to generate adequate protein samples for PTM studies by the site-specific incorporation of uAAs which mimic PTMs into the target protein. Several uAAs mimicking PTMs (Figure 13) were already successfully applied to investigate the effect of certain protein modifications

such as phosphorylation, methylation, acetylation, crotonylation and ubiquitination on target proteins.

The chromatin condensation and therefore gene transcription is organized by histone proteins which are tightly regulated by the so-called histone code consisting of a network of certain PTMs[148] and GCE-based approaches offer a toolbox to study those modifications. One reported example uses acetyl-lysine (AcK) for the examination of the role of lysine acetylations at certain positions in histone H3 for the chromatin regulation.[149] Furthermore, the Schultz lab established a GCE-based system for *E. coli* and mammalian cells which allows the site-specific incorporation of crotonylated lysine (Kcr) which is a useful tool for the examination of the histone code.[150] But not only histone function is affected by PTM. For example the role of lysine methylation on enzymatic function of the molecular chaperone Hsp90 was studied by a GCE-based approach with a *Mb*-PylRS/tRNA pair variant using met-ncbK which is converted to mono-methylated lysine (MeK) upon protein purification.[151]

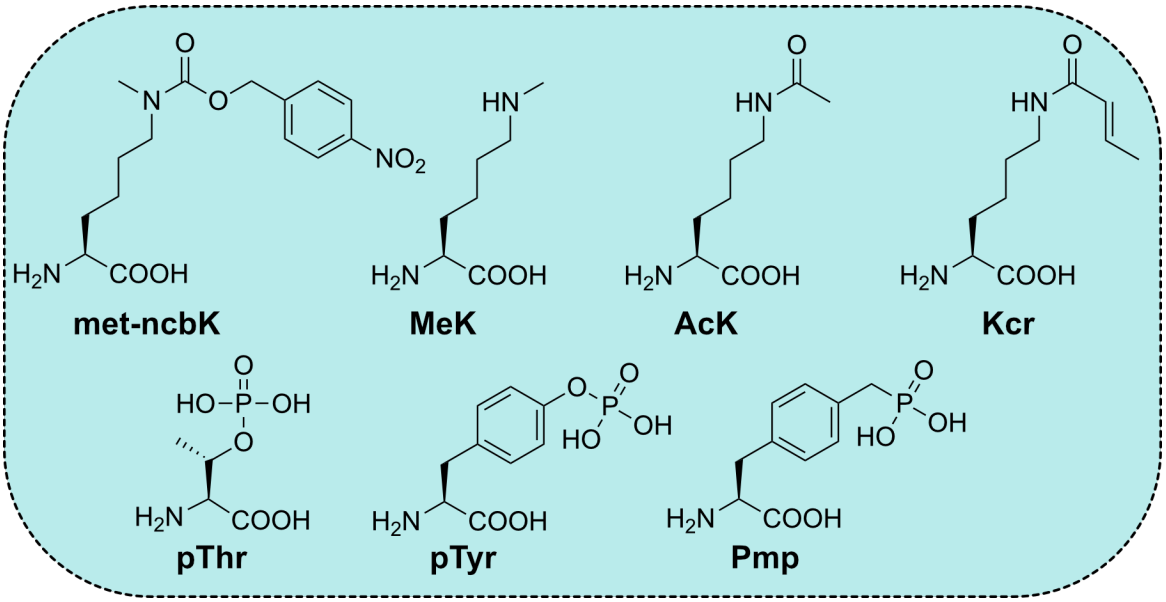


Figure 13: Structures of uAAs applied in GCE-based approaches to study PTMs (PTM = post-translational modification)

Another very important PTM of proteins is the phosphorylation of threonine and tyrosine residues.[152] Different uAAs were developed to allow the investigation of phosphorylation of e.g. threonine and tyrosine residues on the protein function such as pThr, pTyr and Pmp (Figure 13). The generation of proteins sample with stable phosphorylations *in vivo* is rather challenging due to the abundance of endogenous phosphatases which might hydrolyze the modification from the uAA. Söll and colleagues reported a platform for the incorporation of pTyr into proteins using an engineered *Mj*-TyrRS/tRNA pair and an altered *E. coli* host. The authors achieved stability of pTyr by using an altered *E. coli* expression host harboring genomic deletions of five phosphatases to circumvent dephosphorylation of pTyr *in vivo*.[153] Another approach to prevent the pTyr degradation *in cellulo* was published by Luo *et al.* by using the non-hydrolysable analog Pmp. Furthermore, the group presented a strategy to improve the cellular availability of pTyr and Pmp by using a pro-peptide approach (Lys-pTyr, Lys-Pmp).[56] Phosphorylated threonine (pThr) was incorporated by a evolved phosphoseryl-tRNA synthetase/tRNA pair and used to investigate the effect of post-translational phosphorylation of threonine residues on cyclin-dependent kinase 2 (Cdk2) activity.[154] Recently an approach that

combines GCE and sortase-mediated transpeptidation (termed sortylation) to study site-specific lysine ubiquitination of target proteins was developed by the Lang lab and is described in 3.1.[155]

1.2.3.4. Investigation of protein-protein interactions via crosslinking

Protein-protein interactions are fundamental for the integrity of biological processes and involved in signaling and regulatory mechanisms.[156] Studying protein-protein interactions is often hampered by not knowing certain interaction partners and the often transient nature of protein-protein complexes. GCE-based chemical and photo-crosslinking approaches were developed to identify protein interaction partner as well as to stabilize low-affinity protein complexes in order to address biological questions. Therefore, crosslinking uAAs are incorporated into interaction interfaces of target proteins. These uAAs are decorated with functional groups, which react with residues of the interacting protein partner to from covalently trapped protein-protein complexes.

Photo-crosslinking uAAs display photo-reactive moieties such as benzophenones (pBpa[157]), diazirines (AbK = DiaziK[158] or TfmdPhe[159]) or arylazide moieties (e.g. AzPhe[160]) which form reactive species upon irradiation by UV light and can unspecifically react with molecules in vicinity (Figure 14A).[161, 162] Arylazides moieties require UV light around 250 nm to undergo conversion to reactive nitrenes which preferably react with heteroatom-H or C-H bonds. Diazirines and benzophenones are commonly activated by UV light of longer wavelength (around 350 nm) to form the respective carbenes and ketyl diradicals. Activated diazirine as well as benzophenone favor the insertion into C-H bonds in vicinity and benzophenone preferentially reacts with methionine (γ-CH bonds). The benefit of photocrosslinking using benzophenones is the reversibility of the formation of the reactive ketyl diradical whereas the conversion to carbenes or nitrenes is irreversible. Photo-reactive uAAs which are derivatives of phenylalanine (e.g. pBpa, TfmfPhe, AzPhe) are limited in their side chain size to 6 – 8 Å and can only serve for the trapping of short-distance protein interactions. To expand the application to longer-range distances (12 - 15 Å) photo-reactive lysine derivates were designed (e.g. DiaziK).[163] One obstacle of the photo-reactive carbene and nitrene species for crosslinking in cellular environments is the side reaction with water molecules, while ketyl diradicals are inert towards water.[164, 165] The application of photo-crosslinking uAAs is widely used to identify unknown protein interaction partners within the proteome and is commonly combined with mass spectrometric (MS) studies.[157-159, 166, 167] Moreover, bifunctional uAAs bearing a photo-crosslinking moiety as well as a pulldown handle such as alkyne moieties (e.g. BPKyne[168], PrDiAzK[169]; Figure 14B) were designed to enable an affinity enrichment of photo-crosslinked proteins. Drawback of photo-crosslinking strategies using uAAs is a high background due to the unspecific crosslinking reaction and low crosslinking efficiency due to side reactions with water molecules. Moreover, it often requires abrasive irradiation with UV light to trigger the crosslinking reaction which is incompatible for experiments with living cells. To circumvent need of UV light treatment and random crosslinking events a second class of uAAs for chemical crosslinking was established which allows the control over the crosslinking site to generate homogenously linked protein-protein complexes (Figure 15).

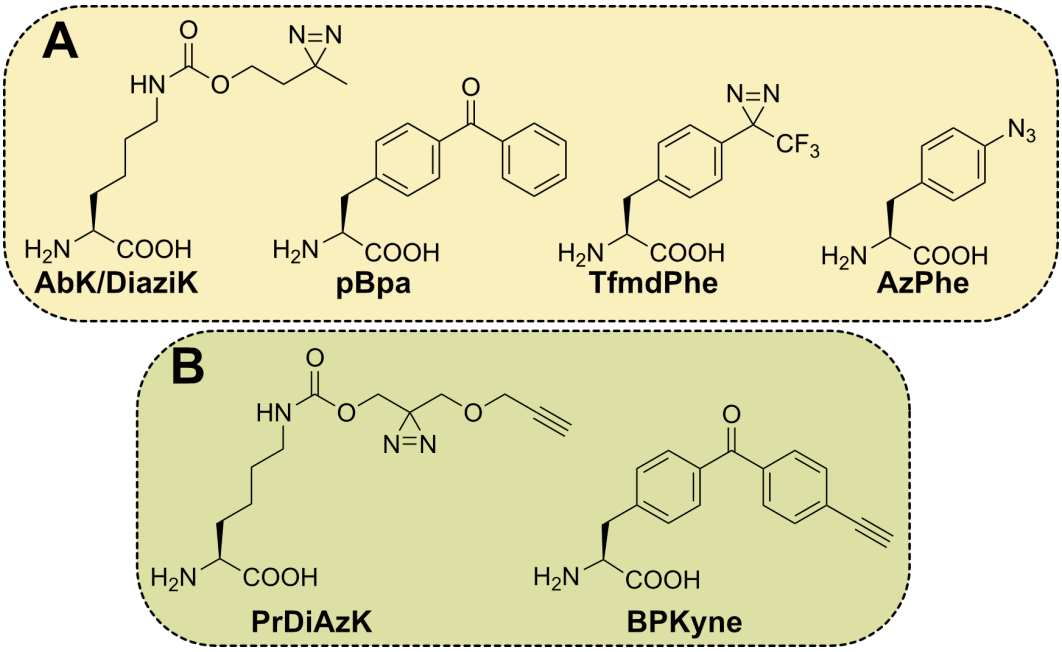


Figure 14: Structures of uAAs applied in GCE-based approaches for photo-crosslinking A: uAAs bearing a photo-reactive functionality within their side chain. B: Bifunctional uAAs for photo-crosslinking and affinity enrichment.

The group of chemical crosslinking uAAs displays electrophilic (e.g. BrCnK, FAcK, AcrK; Figure 16) moieties within their side chains which are inert under physiological conditions but undergo proximity-triggered covalent crosslinking with adjacent nucleophilic residues within the interacting protein partner (e.g. cysteine, aspartate or glutamate; Figure 15; the principle of proximity-enhanced protein crosslinking (PEPC) is discussed in detail in 4.1.1. This typically requires fine-tuning of the electrophilic group of the uAA to ensure reaction specificity.[170] The reaction is realized by the proximity effect which causes a massive increasement of the effective local concentration of both interaction partners leading to a reaction rate enhancement of > 10^5 M.[171]

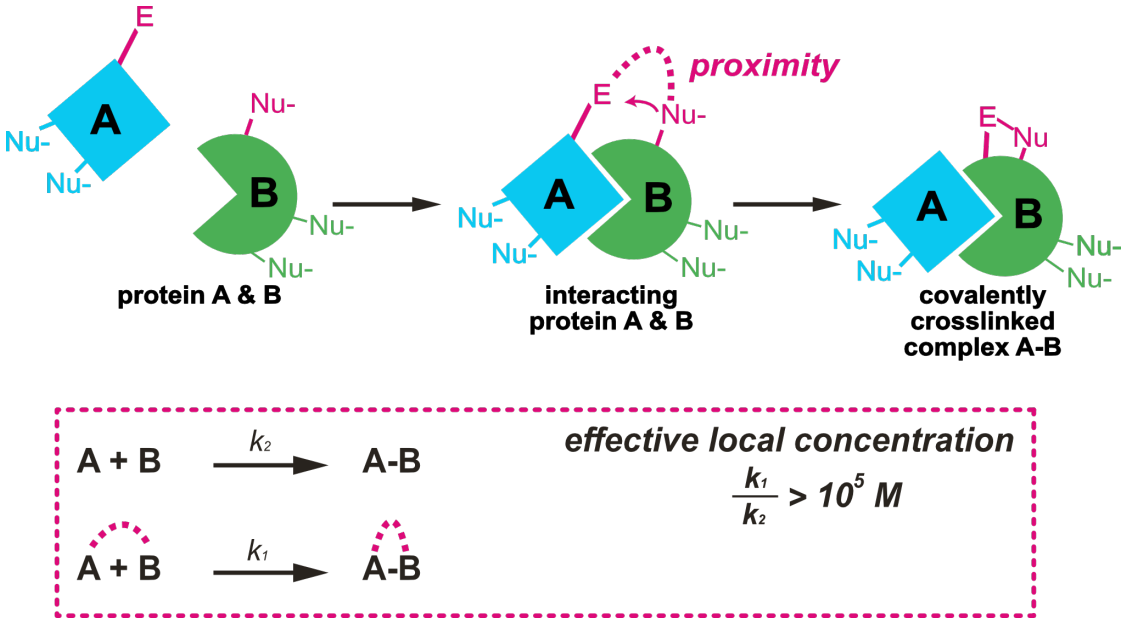


Figure 15: PEPC approach using electrophilic uAAs. An electrophilic uAA (pink) is site-specifically introduced into protein A (blue) via GCE. The electrophilic uAA can react with nucleophilic amino acids (Nu-, pink) closed by due to a massive rate enhancement triggered by the increasement of the local concentration through proximity. In this way covalent crosslinking with an interacting protein B (green) can be achieved. (E = electrophilic moiety, Nu^- = nucleophilic moiety)

One class of electrophilic uAAs for chemical crosslinking uses reactive halide moieties which undergo S_N2-like reactions with nucleophilic groups (e.g. BrCnK, FAcK), preferably cysteines. Cigler *et al.* demonstrated the successful trapping of the low-affinity complex of GDP-bound Rab1b and the guanine exchange factor domain of DrrA for structural elucidation. They used electrophilic bromoalkyl bearing uAAs (BrCnK) which undergo reaction with nucleophilic residues in proximity.[172] Site-specific N^{ε}-fluoroacetyllysine (FAcK) incorporation into calmodulin (CaM) harboring certain cysteine substitutions was reported to result in intramolecular protein crosslinking.[173]
Moreover, a second kind of uAAs bearing Michel acceptors within the side chains (e.g. AcrK; Figure 16) was shown to react with nucleophilic residues such as proximal lysines and was reported for the covalent crosslinking of a model system consisting of the known interaction between the extracellular domain (ECD) of the oncogenic receptor tyrosine kinase ErbB2 and the antigen binding fragment (Fab) of Herceptin.[174]

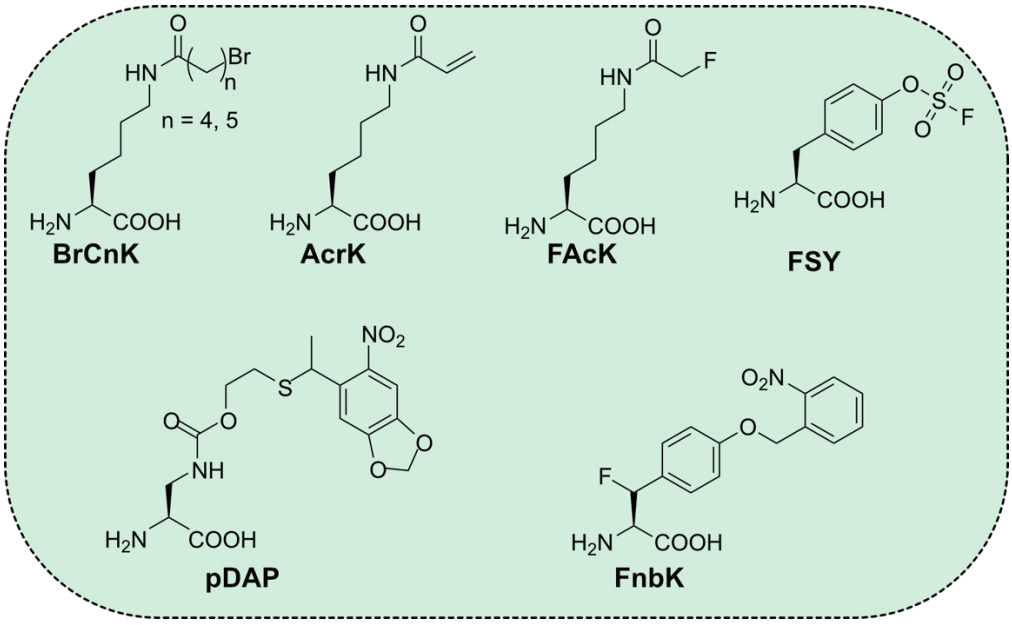


Figure 16: Structures of uAAs applied in GCE-based approaches for chemical crosslinking of proteins.

In addition to electrophilic halides and Micheal acceptors, uAAs were developed which provide quinone methides (e.g. FnbY; Figure 16)[175], aryl fluorosulfates (e.g. FSY; Figure 16)[176], aryl carbamates[177] or aryl isothiocyanates[178] for respective crosslinking reactions with nucleophilic residues other tan cysteines. The aryl fluorosulfate uAA FSY undergoes proximity-enabled sulfur-fluoride (SuFEx) reaction with lysine, tyrosine and histidine residues under physiological conditions. FSY is reported for *in vivo* SuFEx reactions in bacteria and mammalian cells and is site-specifically incorporated by an evolved PylRS/tRNA pair.[176] Other uAAs combine the PEPC approach with an optical controlled activation step. The uAA FnbY has a nitrobenzyl photocage which prevents its crosslinking ability until light-triggered photocage release restores a *p*-quinone methide moiety upon spontaneous elimination of the fluoride moiety at the β-position. The resulting *p*-quinone methide moiety is prone to Michael additions with many nucleophilic amino acids such as cysteine, histidine, tyrosine, lysine methionine or tryptophan.[175] The uAA pDAP (Figure 16) is a special case in the class of chemical crosslinking uAAs as it harbors a nucleophilic moiety for crosslinking reactions in contrast to the above discussed uAAs. Like FnbY, pDAP requires UV light-mediated decaging

to DAP in order to reinstall the reactive, nucleophilic ß-amino group. The ß-amino group nucleophilically attacks electrophilic moieties (e.g. carbonyl groups) of molecules in its proximity and can be used to replace an active site cysteine or serine residue of an enzyme to covalently trap acyl-enzyme-substrate intermediates. It was successfully applied to capture enzyme-acyl intermediates for crystallographic studies.[179] The approach of trapping protein-interactor complexes via DAP is elaborately explained in 4.3.1.

1.2.3.5. Photo-control of proteins

Another useful GCE-based application is offered by amino acids which allow light-mediated protein control.[180] For this purpose, amino acids masked with nitrobenzyl photocages (e.g. ONBY, photoLys, photoCys; Figure 17A) or uAAs containing azobenzene moieties (F_4AzoF[181], Cl-PSCaa[182]; Figure 17B) were designed, which are commonly introduced into critical residues within the active site of proteins. The photocages of these amino acids are irreversibly released by light to uncover the functional group of the amino acid thereby restoring the native enzyme/protein function. This was utilized exemplarily to control enzyme activity or protein localization in a light-dependent manner. A photocaged tyrosine (ONBY) was reported to be site-specifically incorporated into the active site of β-galactosidase thereby preventing its enzymatic activity but allowed the photochemical activation of the enzyme through removal of the photocage.[183] Another example for photo-control of protein localization mediated by uAAs was shown by Chin and co-workers who replaced essential lysines by photocaged lysines (photoLys) within the nuclear localization sequence (NLS) of proteins thereby inhibiting their nuclear import. After light-triggered decaging of the uAA the native NLS function can be reinstalled and proteins acquire the localization into the nucleus.[184] Moreover, the Chin lab used the replacement of the active site cysteine of the Tobacco Etch Virus (TEV) protease by a photocaged cysteine (photoCys) to control the enzymatic activity towards TEV site cleavage which was monitored using a FRET-sensor substrate (ECFP-TevS-YFP-V5) containing a TEV site between the FRET-pair.[185]

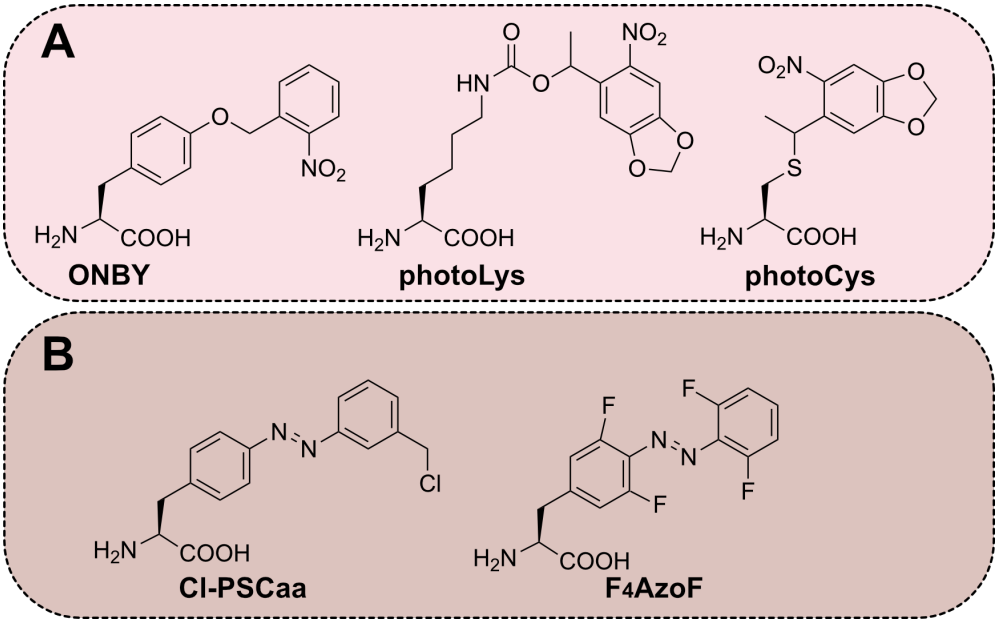


Figure 17: Structures of uAAs applied in GCE-based approaches to control protein function. A: Photocaged uAAs which allow light-dependent protein control. B: uAAs for reversible protein control using azobenzene photoswitching.

Whereas photocages can only be removed once, the reversible, photo-control of protein function can be obtained via incorporation of azobenzene containing uAAs (e.g. F_4AzoF[181], Cl-PSCaa[182]; Figure 17B). The azobenzene moiety changes its conformation by light-induced cis/trans isomerization (photoswitching). By introducing such an uAA at a certain position into a protein the photoswitching of the azobenzene moiety can regulate its function through affecting local conformations of the target protein. Using firefly luciferase (FLuc) as model protein the installation of F_4AzoF at a position within an allosteric region (W417) lead to reversible cycles of an inactive and active enzyme state.[181] The photoswitching properties were also combined with a PEPC approach (e.g. Cl-PSCaa) to span regulatory protein regions (e.g. central helices) by photoswitchable thioether linkages in order to photochemically control the conformation of relevant protein regions and thereby the enzymatic activity.[182]

In summary many efforts were taken to develop custom-made tools for biological studies by GCE-approaches which greatly contribute to the evaluation of biological problems. GCE-based methods allow the modification of proteins within their native environment for a desired application as labeling approaches, activity studies or protein-protein interaction studies for crystallographic experiments or the identification of protein interfaces.

2. CHAPTER I - ESTABLISHMENT OF M. ALVUS

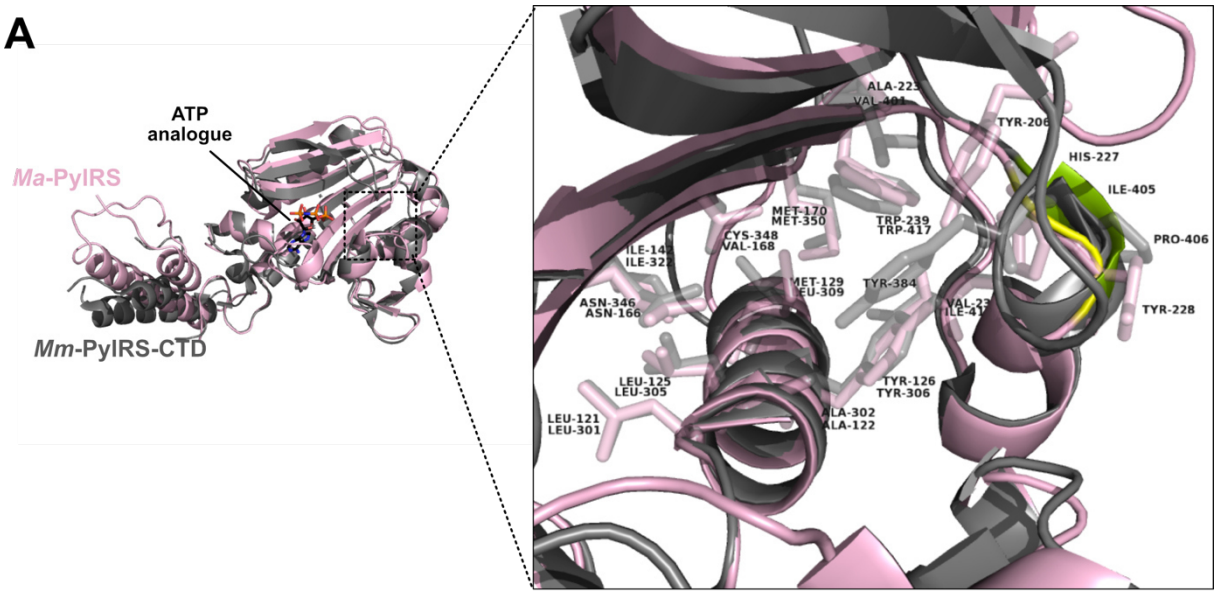
2.1. Background

Orthogonal aaRS/tRNA pairs are elemental for genetic code expansion approaches using amber suppression and were already discussed in 1.2.2.3. Due to its universal orthogonality to endogenous aaRS/tRNA pairs of bacteria and eukaryotes[41], its broad substrate spectrum[107] and anti-codon blindness[110] the aaRS/tRNA pair of choice is, in most cases, the pyrrolysyl-tRNA synthetase/$tRNA_{CUA}$ (PylRS/PylT) pair derived from archaea *Methanosarcina mazei* (*Mm*) and *Methanosarcina barkeri* (*Mb*).

Since 2018, a novel PylRS/tRNA pair from the archaea species, *Methanomethylophilus alvus* (*Ma*), has attracted attention for genetic code expansion approaches.[97, 98, 186] Like the archaeal *Mm*- or *Mb*-PylRS/PylT pair its activity is independent of the tRNA anti-codon.[97] Moreover the wild type enzyme has a relatively high degree of substrate promiscuity[98] and shows orthogonality in bacteria and mammalian cells.[98, 187] Thus, the *Ma*-PylRS/tRNA pair is suited as substituent for the *Mm*- or *Mb*-PylRS/PylT system. Furthermore, *Ma*-PylRS reaches comparable amber suppression efficiencies *in vivo* as the *Mm* enzyme[97, 187] and even higher efficiencies for some lysine derivatives tested in *E. coli*[186]. It was shown that the alteration of the *Ma*-PylRS substrate recognition can be achieved by transplanting mutations established on *Mm- or Mb*- PylRS mutants for certain uAAs.[97]

The similar properties of *Mm*- and *Ma*-PylRS are related to the high structural homology of the *Ma*-PylRS to the catalytically active C-terminal domain (CTD) of *M. mazei* which is visualized in superimposition of the crystal structures of both wild type enzymes in Figure 18A.[188] Moreover, the sequence alignment of *Ma*-PylRS and the *Mm*- and *Mb*-CTD (Figure 18C) predicts sequence identities of 35 % for the *Ma* and *Mb* and 37 % for *Ma* and *Mm* enzyme. The close-up view into the substrate binding pocket of both enzymes shows residues which contribute to substrate recognition (Figure 18A). An overview of the residues influencing substrate recognition of *Ma*-PylRS and its respective equivalent residues in *Mm-/Mb*-PylRS are listed in Figure 18B. These residues can be addressed to either first layer residues which directly interact with the amino acid substrate or second layer residues which play an indirect role in substrate recognition by influencing the first layer residues.[189] Important second order layer residues of *Ma* are highlighted in yellow and their *Mm* counterparts in green. Mutating H227 to proline and Y228 to isoleucine in *Ma*-PylRS mutants can positively affect the incorporation of bulky uAAs such as N^{ε}-((((E)-cyclooct-2-en-1-yl)oxy)carbonyl)-l-lysine (TCO-Lys).[189] Interestingly, the corresponding second order layer residues in *Mm* at position 405 and 406 are originally isoleucine and proline (Figure 18). Despite all the described similarities also clear differences characterize the relationship between *Ma*-PylRS and *Mm*-PylRS. Unlike the *Mm*- or *Mb*-enzyme the *Ma*-PylRS lacks the N-terminal domain (NTD) (Figure 19B).[188] In *Mm*-PylRS this domain is responsible for binding of the cognate tRNA by interaction with parts of the T-arm and variable loop[105, 190] (Figure 19A) and is further absolutely required for full enzymatic activity *in vivo*[100]. The presence of the NTD has however been known to increase aggregation and limit the solubility of active *Mm*-PylRS[103, 107, 190] whereas *Ma*-PylRS show high solubility fostered by the missing NTD.[189] Consequently, *Ma*-PylRS is beneficial for

amber suppression *in vitro* as recently highlighted by Seki *et al.* who demonstrated its suitability in cell-free translation systems.[189] The visualization of the PylRS domain organization and the concept of the PylRS-tRNA interaction for both, *M. mazei* (Figure 19A) and *M. alvus* (Figure 19B), is purely a schematic representation, as crystal structures of complete tRNA-PylRS complexes are missing.



B

M. mazei PylRS	*M. barkeri* PylRS	M. alvus PylRS
M276	M241	M96
L301	L266	L121
A302	A267	A122
L305	L270	L125
Y306	Y271	Y126
L309	L274	M129
I322	I287	I142
N346	N311	N166
C348	C313	V168
M350	M315	M170
Y384	Y349	Y206
S399	S364	S221
V401	V366	A223
I405	V370*	H227
P406	S371*	Y228
L407	L372	L229
I413	I378	V235
W417	W382	W239
G421	G386	G243

*determined from sequence alignment

C

```
CLUSTAL O(1.2.4) multiple sequence alignment

Ma-PylRS      ------------------------------------------------------------      0
Mb-PylRS      MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTARAF      60
Mm-PylRS      MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARAL      60

Ma-PylRS      ------------------------------------------------------------      0
Mb-PylRS      RHHKYRKTCKRCRVSDEDINNFLTRSTESKNSVKVRVVSAPK-VKKAMPKSVSRAPKPLE      119
Mm-PylRS      RHHKYRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLE      120

Ma-PylRS      ------------------------------------------------------------      0
Mb-PylRS      NSVSAKASTNTSRSVPSP----------------------------------AKSTPNSS      145
Mm-PylRS      NTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSISTGATASALVKGNTNPITSMS      180

Ma-PylRS      -----MTVKYTDAQIQRLREYGNGTYEQKVFEDLASRDAAFSKEMSVASTDNEKKIKGMI      55
Mb-PylRS      VPASAPAPSLTRSQLDRVEALLSPEDKIS----L-NMAKPFRELEPELVTRRKNDFQRLY      200
Mm-PylRS      APVQASAPALTKSQTDRLEVLLNPKDEIS----L-NSGKPFRELESELLSRRKKDLQQIY      235
                    :   * :* :*:.   .   : .    * .    * :      : .::.:: :

Ma-PylRS      ANPSRHGLTQLMNDIADALVAEGFIEVRTPIFISKDALARMTITEDKPLFKQVFWIDEKR      115
Mb-PylRS      TNDREDYLGKLERDITKFFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNL      260
Mm-PylRS      AEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNF      295
              ::  .. * :* .:*:  :* .**:*:::**:*  : : ** * :*. * **:* :*::

Ma-PylRS      ALRPMLAPNLYSVMRDLRDHTDGPVKIFEMGSCFRKESHSGMHLEEFTMLNLVDMGPRGD      175
Mb-PylRS      CLRPMLAPTLYNYLRKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVNFCQMGSGCT      320
Mm-PylRS      CLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQMGSGCT      355
              .*******.**. :*.*      .*:****:* *:****..  *******:*: :**

Ma-PylRS      ATEVLKNYISVVMKAAGLPDYDLVQEESDVYKETIDVEINGQEVCSAAVGPHYLDAAHDV      235
Mb-PylRS      RE-NLEALIKEFLDYL-EIDFEIVGDSCMVYGDTLDIMHGDLELSSAVVGPVSLDREWGI      378
Mm-PylRS      RE-NLESIITDFLNHL-GIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGI      413
                 *:  *. .:.     *:.:* :.. ** :*:*:  .. *:.**.***  **   .:

Ma-PylRS      HEPWSGAGFGLERLLTIREKYSTVKKGGASISYLNGAKINGS        277
Mb-PylRS      DKPWIGAGFGLERLLKVMHGFKNIKRASRSESYYNGISTNL-        419
Mm-PylRS      DKPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL-        454
              .:** **********.: . :..:*:.. * ** ** . *
```

```
Percent Identity  Matrix - created by Clustal2.1


  1: Ma-PylRS     100.00   35.32   37.17
  2: Mb-PylRS      35.32  100.00   79.24
  3: Mm-PylRS      37.17   79.24  100.00
```

Figure 18: Comparison of *Mm-/Mb*-PylRS-CTD and *Ma*-PylRS structures. A: Superimposition of the crystal structures of wild type *Mm*-PylRS (C-terminal domain = CTD; grey; PDB: 2q7e) and *Ma*-PylRS (full-length; pink; PDB: 6ezd). Close-up view into the substrate binding pockets of *Mm*- (grey) and *Ma*-PylRS (pink) crystal structures. Residues forming the substrate binding pocket are labeled. Second layer residues of *Mm*-PylRS binding pocket are highlited in green, those of *Ma*-PylRS in yellow. B: Binding pocket residues of *Mm*-PylRS and its corresponding residues in *Mb*- and *Ma*-PylRS. Colored are the fomer marked second order layer residues of *Mm*- (green) and *Ma*-PylRS (yellow). C: Percent Identity Matrix and multiple sequence alignment of the CTD of *Mm-/Mb*- PylRS and full-length *Ma*-PylRS derived by Clustal2.1[191].

Unfortunately, the *Mm*-PylRS/PylT pair and the *Ma*-PylRS/PylT pair have no mutual orthogonality to each other as the *Mm*-PylRS accepts the *Ma*-tRNA.[97, 98] Phylogenetic analysis

of different PylRS genes without NTD revealed their origin in one distinct clade, termed *Methanomassiliicoccales* the seventh order of methanogens[188, 192]. Although the *Mb*-/*Mm*-PylRS is a member of the clade of *Methanosarcinales*[193, 194] the evolutionary divergence between the species, *Mb/Mm* and *Ma*, is not sufficient to produce mutually orthogonal PylRS. To overcome this limitation researchers used rational[98] and evolutionary[97] approaches of the *Ma*-PylT to finally achieve orthogonality of the *Ma*- and *Mm*-PylRS/tRNA pairs thereby paving the way for site-specific dual-incorporation of two different uAAs at distinct sites into one protein *in vivo*.[97, 98]

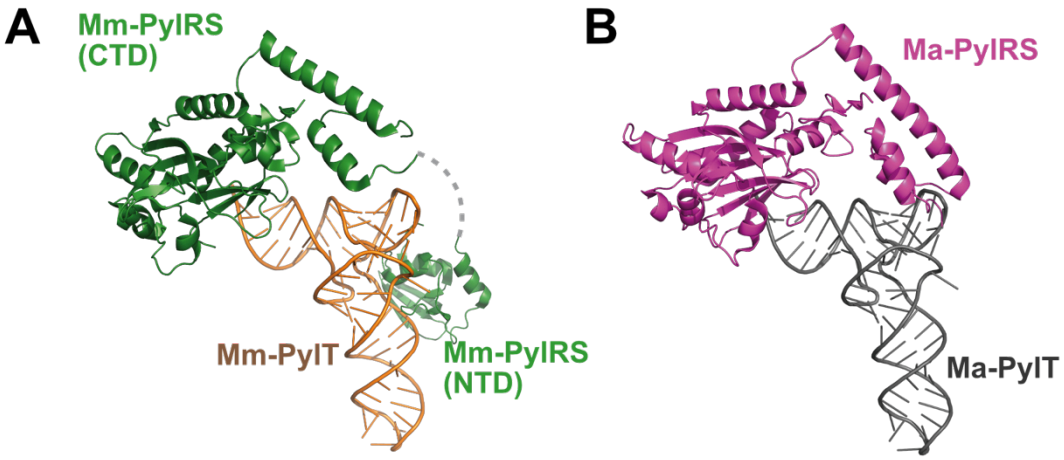


Figure 19: Schematic illustration of the PylRS-tRNA interaction and the PylRS architecture. A: Conceptual model of the *Mm*-PylRS (CTD & NTD; green) binding to its cognate tRNA (*Mm*-PylT; yellow). Grey dotted line should indicate that the domains (CTD and NTD) are connected in the native protein. Crystal structures obtained from the PDB: CTD from PDB: 2e3c; parts of NTD & *Mm*-PyT from PDB:5ud5. B: Conceptual model of the *Ma*-PylRS (pink) binding to its cognate tRNA (grey). Crystal structures obtained from the PDB: *Ma*-PylRS PDB: 6jp2; no crystal structure for *Ma*-PylT is available therefore we used *Mm*-PylT (PDB: 5ud5) to conceptually mimic *Ma*-PylT. (crafted illustration by using the mentioned PDB-files)

2.2. Aim

Our lab is already well equipped with plasmids for amber suppression using the *M. barkeri* PylRS/tRNA pair and the *M. jannaschii* TyrRS/tRNA pair. To further expand our toolbox for genetic code expansion we aim to establish the novel *M. alvus* PylRS/tRNA system. We assume that, although there is an obvious similarity between *Mb* and *Ma*-PylRS/tRNA pairs, the *Ma* system will offer us a new slightly different platform which might improve the incorporation efficiency of poorly accepted uAAs.[186] Additionally, the *Ma*-PylRS/tRNA pair might give us access to the incorporation of uAAs which, until now, cannot be introduced using the *Mb* system. Another interesting aspect is the simultaneous usage of *Ma*- and *Mb*-PylRS for dual incorporation of uAAs at distinct sites in one protein *in vivo*. This is made possible by altering the suppressor tRNA from *M. alvus* to achieve orthogonality for both PylRS/tRNA pairs.[97] Moreover, the *Ma*-PylRS system is perfectly suited for *in vitro* approaches such as cell-free protein synthesis due to its solubility properties which could be an interesting point for future projects. [189] For our project we decided to use the evolved *Ma*-PylT version (tRNA(6)$_{CUA}$) which was published in 2018.[97] It was shown that the *Ma*-tRNA(6)$_{CUA}$ is not recognized by *Mm*-PylRS, but is a good substrate for *Ma*-PylRS.[97] Choosing *Ma*-tRNA(6)$_{CUA}$ for our system opens the possibility to combine it with our existing *Mb*-PylRS/tRNA pair,

thereby enabling approaches for multiple site-specific incorporation of uAAs into one protein. Hence, hereinafter *Ma*-PylT will be used as an acronym for *Ma*-tRNA(6)$_{CUA}$ and all experiments will be performed with this tRNA variant.
Our goal is the establishment of a robust plasmid system for amber suppression with the *Ma*-PylRS/tRNA pair in *E. coli* and mammalian cells. Furthermore, we want to construct beneficial *Ma*-PylRS mutants and identify the substrate spectrum of *Ma*-PylRS and its mutants using existing uAAs from our lab. Last but not least, we are interested in the adaption of our directed evolution approach for *Mb*-PylRS to *Ma*-PylRS in order to use it for the evolution of *Ma*-PylRS mutants for new uAAs.

2.3. Results & Discussion

2.3.1. Plasmids for GCE with *M. alvus*

We decided to implement the *M. alvus* system on our established pBK/pPylT plasmid system for *E. coli*, as some *Ma*-PylRS mutants were already encoded on our pBK-plasmid (MW-Ma7, MW-Ma8; MW-Ma10, MW-Ma11; Table 1) which were cloned by my colleague Maximilian Fottner (PhD student, group of Prof. Lang, TUM).
From this starting further *Ma*-PylRS mutants (MW-Ma9, MW-Ma12 - MW-Ma19; Table 1) were cloned as well as the wild type enzyme (MW-Ma6; Table 1) using site-directed mutagenesis. In parallel, our standard reporter plasmid pPylT_sfGFP-V2P-N150TAG-PylT for amber suppression using *Mb*-PylRS was modified by replacing the *Mb*-PylT with the *Ma*-PylT (MW-79; Table 1). *Ma*-PylT consists of 72 bp with an expanded variable loop (A41, U42, A43, G44) and an additional G56A mutation.[97] The V2P mutation in pPylT_sfGFP-V2P-N150TAG-PylT provokes complete cleavage of the starting-methionine by methionine aminopeptidase in the host cell.[195] This leads to a single species in ESI-MS analysis compared to sfGFP-N150uAA-His6 which shows a mixture of processed (- methionine) and unprocessed (+ methionine) sfGFP-N150uAA-His6. The results of the functionality tests of our successfully designed pBK/pPylT-based amber suppression system in *E. coli* are discussed in 2.3.2.
For amber suppression in mammalian cells, plasmids MM30 and MM31 (Table 1) were obtained from the lab of Jason Chin (MRC). MM30 encodes the *Ma*-PylRS gene and four copies of the *Ma*-tRNA. MM31 harbors the corresponding sfGFP-N150TAG reporter gene and again four copies of the *Ma*-tRNA. To increase usable reporter systems for mammalian cell experiments with *Ma*-PylRS, a second reporter plasmids MM32 (Table 1) was generated by restriction cloning, which carries a mCherry-TAG-GFP fusion construct and four copies of the *Ma*-tRNA. Furthermore, a pET19 helper plasmid (MM28; Table 1) was cloned for easy mutagenesis and transfer of *Ma*-PylRS mutants into the pPB1-backbone of MM30. The application of the *Ma*-plasmids for eukaryotes is shown in 2.3.3.
In order to derive plasmids for the directed evolution of *Ma*-PylRS the existing *Mb*-D65, *Mb*-D7L and *Mb*-sfGFP-CAT plasmids wre adaped for *Mb*-PylRS evolution (Table 55) by substitution of the *Mb*-PylT with the *Ma*-PylT of each vector backbone. A detailed explanation about the role of each plasmid in directed evolution approaches is given in 5.2.5.2. The integrity of the *Ma*-D65, *Ma*-D7L and *Ma*-sfGFP-CAT plasmids is verified in 2.3.2.

A more detailed information about cloning and vector components of all plasmids can be found in the appendix 7.1.1 Table 64.

Table 1 Plasmids for amber suppression with the *M. alvus* system

Code	Plasmid	Application
MW-Ma6	pBK_Ma-PylRSwt	Amber Suppression Bacteria
MW-Ma7	pBK_Ma-PylRS-MF3 (Y126G)	Amber Suppression Bacteria
MW-Ma8	pBK_Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)	Amber Suppression Bacteria
MW-Ma9	pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)	Amber Suppression Bacteria
MW-Ma10	pBK_Ma-PylRS-L121M-L125I-Y126F-M129A	Amber Suppression Bacteria
MW-Ma11	pBK_Ma_PylRS-M129A	Amber Suppression Bacteria
MW-Ma12	pBK_Ma-PylRS-H227I-Y228P	Amber Suppression Bacteria
MW-Ma13	pBK_Ma-PylRS-MF3 (Y126G)-H227I-Y228P	Amber Suppression Bacteria
MW-Ma14	pBK_Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)-H227I-Y228P	Amber Suppression Bacteria
MW-Ma15	pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)-H227I-Y228P	Amber Suppression Bacteria
MW-Ma16	pBK_Ma-PylRSwt-H227A	Amber Suppression Bacteria
MW-Ma17	pBK_Ma-PylRS-MF3 (Y126G)-H227A	Amber Suppression Bacteria
MW-Ma18	pBK_Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)-H227A	Amber Suppression Bacteria
MW-Ma19	pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)-H227A	Amber Suppression Bacteria
MW79	pPylT-sfGFP-V2P-N150TAG-MaPylT	Amber Suppression Bacteria
MW-E1	*Ma*-D65	Directed Evolution
MW-E2	*Ma*-D7L	Directed Evolution
MW-E3	*Ma*-pPylT_sfGFP150TAG CAT	Directed Evolution
MM28	pET19sub_Ma-PylRSwt	Cloning
MM30	pPB1_Ma-PylRSwt_4xMaPylT	Amber Suppression Mammalian cells
MM31	pPB2_sfGFP-His6_4xMaPylT	Amber Suppression Mammalian cells
MM32	pPB3_mCherry-TAG-GFP_4xMaPylT	Amber Suppression Mammalian cells

2.3.2. Amber Suppression with *M. alvus* in *E. coli*

To investigate the function of the pBK/pPylT amber suppression system for the *Ma*-PylRS/tRNA pair, test expressions were performed using co-transformed *E. coli* NEB 10-beta cells with the respective plasmids (Figure 20A). The four *Ma*-PylRS variants (*Ma*-PylRSwt; *Ma*-MF3; *Ma*-MF9; *Ma*-MF18; Table 1) were derived by transplanting the known mutation from *Mb*-PylRS mutants (*Mb*-MF3; *Mb*-MF9; *Mb*-MF18; Table 9) to the corresponding

positions in *Ma*-PylRS. uAAs which are known to be accepted by the respective *Mb*-PylRS mutants served as substrates. The structures of these uAAs are listed in the appendix 7.1.3. Expressions without uAA addition served as negative control experiments. The expressions with wild type *Ma*-PylRS efficiently produced sfGFP (mass of approx. 27 kDA) when introducing N^6-(*tert*-butoxycarbonyl)-L-lysine (BocK) and N^6-(((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl)-L-lysine (CpK) (appendix 7.1.3) at TAG-position 150. The absence of sfGFP expression in cultures not supplemented with BocK or CpK produced just truncated sfGFP species (approx. 17 kDa). This verifies the selectivity of *Ma*-PylRS, as it does not recognize any natural amino acid (Figure 20A).

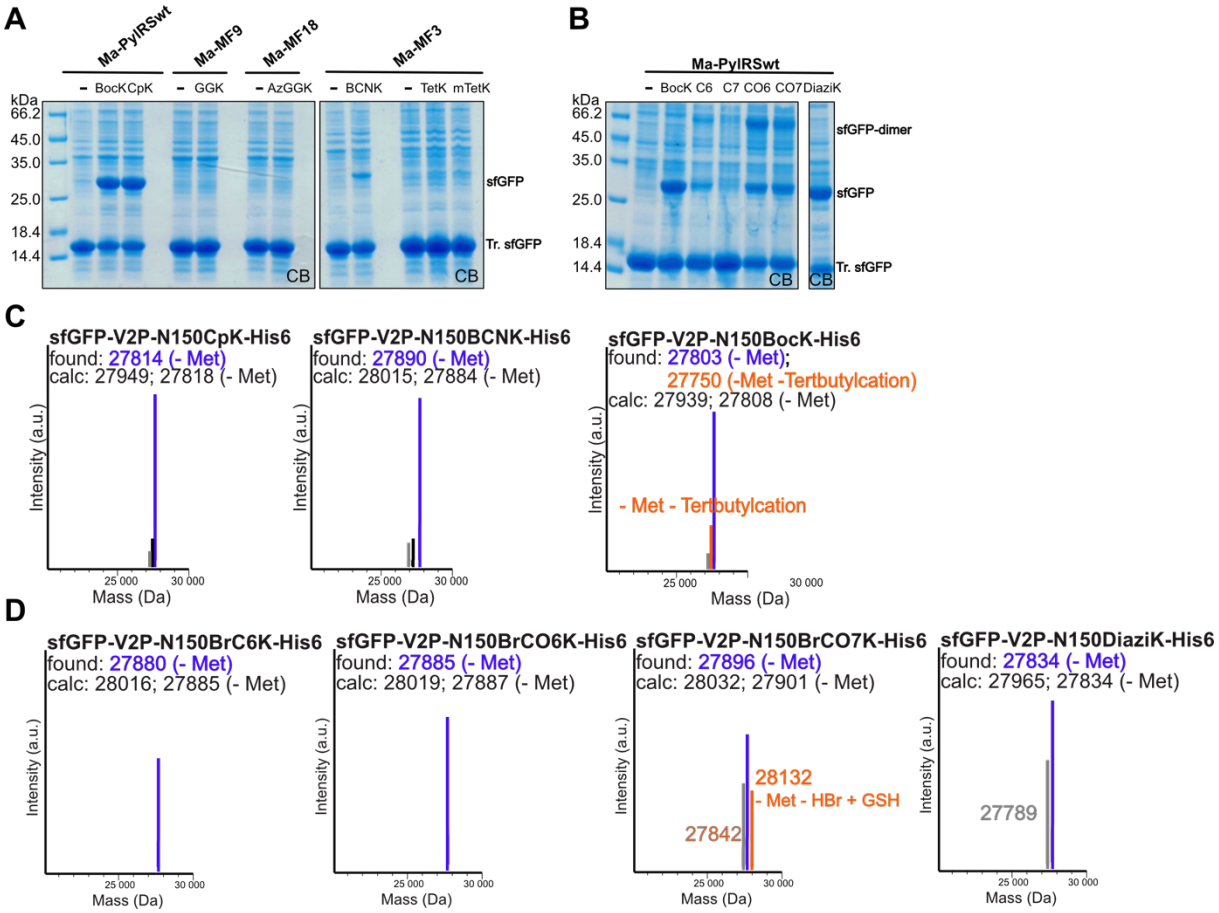


Figure 20: Expression test of the pBK/pPylT system for amber suppression with *Ma*-PylRS/tRNA of sfGFP-V2P-N150TAG-His6 using different *Ma*-PylRS mutants and uAAs. A: SDS gels showing the cell lysates of the test expressions with *Ma*-PylRS variants. Expression of sfGFP-V2P-N150TAG-His6 using *Ma*-PylRS and *Ma*-MF3 were performed in 2x YT medium. Expressions of sfGFP-V2P-N150TAG-His6 using *Ma*-MF9 and *Ma*-MF18 were carried out in AI medium without phenylalanine. B: SDS gel of cell lysates from test expressions with *Ma*-PylRS and further uAAs. C: ESI-MS data (average masses) of purified sfGFP-V2P-N150TAG-His6 obtained in different test expressions from A using *Ma*-PylRS (BocK, CpK) and *Ma*-MF3 (BCNK). D: ESI-MS data (average masses) of purified sfGFP-V2P-N150TAG-His6 derived from the test expressions in C. (CB = Coomassie brilliant blue; Met = methionine; Tr. = truncated; C6 = BrC6K; C7 = BrC7K; CO6 = BrCO6K; CO7 = BrCO7K; "-" = no uAA added; GSH = glutathione)

ESI-MS analysis of the purified sfGFP-V2P-N150uAA-His6 confirmed the correct incorporation of BocK (27803 Da) and CpK (27814 Da) into the protein (Figure 20C). The N-terminal methionine of sfGFP-V2P-N150uAA-His6 is fully cleaved off. For sfGFP-V2P-N150BocK-His6 even the mass corresponding to the loss of the tert-butyl cation (27750 Da) could be detected, which is a frequently observed phenomena occurring during the ESI-MS process (internal data of the group of Prof. Lang, TUM). During a second round of test

expressions, we found out that *Ma*-PylRS is able to recognize uAAs with long alkyl side chains such as BrC6K, BrC7K, BrCO6K, BrCO7K and N^6-((2-(3-methyl-3*H*-diazirin-3-yl)ethoxy)carbonyl)-L-lysine (DiaziK) (Figure 20B, appendix 7.1.3). This was quite suprising as for *Mb*-PylRS three mutations (Y271M, L274A, C313A) are necessary for the recognition of BrC6K, BrC7K, BrCO6K, BrCO7K and DiaziK (internal data of the group of Prof. Lang, TUM & [172, 196]). Beside the bands for full-length and truncated sfGFP, the SDS gel nicly shows the sfGFP dimer resulting from the intermolecular reaction of the incorporated BrC6K, BrC7K, BrCO6K or BrCO7K with a nucleophilic residue (E173) in a second sfGFP monomer.[172] This was initially observed and investigated by Marko Cigler[172]. Furthermore, the ESI-MS data confirmed the successful introduction of the BrC6K (27880 Da), BrCO6K (27885 Da), BrCO7K (27896 Da) and DiaziK (27834 Da) into sfGFP-V2P-N150TAG-His6 (Figure 20D). As expected *E. coli* cells with *Ma*-MF3 mutant were able to produce sfGFP in the presence of N^6-((bicyclo[6.1.0]non-4-yn-9-ylmethoxy)carbonyl)-L-lysine (BCNK) (appendix 7.1.3) as observed for *Mb*-MF3 (Figure 20A).[197] ESI-MS of sfGFP-V2P-N150BCNK-His6 (27890 Da) shows the correct mass (Figure 20C). The known substrates N^6-((2-(6-methyl-1,2,4,5-tetrazin-3-yl)ethoxy)-carbonyl)-L-lysine (TetK) and N^6-(((3-(6-methyl-1,2,4,5-tetrazin-3-yl)benzyl)oxy)carbonyl)-L-lysine (mTetK)[141] (appendix 7.1.3) of *Mb*-MF3 were however not accepted by the *M. alvus* counterpart as only truncated sfGFP was detected in cell lysates on the SDS gel (Figure 20A). Also expressions with *Ma*-MF9 and *Ma*-MF18 did not display full-length sfGFP bands in their SDS gels whereas truncated expression product was clearly visible. Consequentely, in contrast to *Mb*-MF9 (internal data of the group of Prof. Lang) and *Mb*-MF18[155], N^6-glycylglycyl-*L*-lysine (GGK) and N^6-((2-azidoacetyl)glycyl)-L-lysine (AzGGK) (appendix 7.1.3) are also no suitable substrates for their respective *Ma*-PylRS mutants (*Ma*-MF9 and *Ma*-MF18). This demonstated that the simple transplantation of mutations from *Mb*- to *Ma*-PylRS does not automatically lead to the same substrate specificity of the enzyme. Inspired by the recently published influence of second order layer mutations (H227I, Y228P)[189] on the substrate recognition of *Ma*-PylRS we introduced them into our *Ma*-PylRS mutants. The new mutants were again tested for the incorporation of the former not accepted uAAs (Figure 21A). The H227I and Y228P mutations in *Ma*-PylRS had no significant influence on the amber suppression with BocK. Also *Ma*-MF3-H227I-Y228P was still only able to recognize BCNK, but not TetK or mTetK. Similar results were obtained for *Ma*-MF9-H227I-Y228P and GGK. In contrast, the H227I and Y228P point mutations in *Ma*-MF18 enabled strong sfGFP production in presence of AzGGK. In a small scale expression with sfGFP-V2P-N150TAG-His6 and either *Ma*-MF18-H227I-Y228P or *Mb*-MF18 better yielding full-length sfGFP expression could be obtained for *Ma*-MF18-H227I-Y228P (Figure 21D). The western blot in Figure 21F shows phenylalanine (Phe) misincorporation in sfGFP-V2P-N150TAG-His6, known for *Mb*-MF18 even in medium without phenylalanine (AI -Phe). Simultanously visualized is the Phe-misincorporation by *Ma*-MF18-H227I-Y228P in AI +Phe which is drastically reduced in AI -Phe and even lower compared to *Mb*-MF18 in AI -Phe. Masses obtained from ESI-MS of purified sfGFP-V2P-N150AzGGK-His6 from expressions with *Ma*-MF18-H227I-Y228P in AI +Phe (27844 Da) and AI -Phe (27854 Da) confirmed the AzGGK incorporation and that Phe-misincorporation (27729 Da) could only be detected in presence of Phe (Figure 21E). Unfortuantely, *Ma*-MF18-H227I-Y228P did not accept further AzGGK derivates (HpGGK, MpGGK, HpGK, MpGK) (further information and usage of AzGGK derivatives is explaned in 3. Chapter II and appendix 7.1.3). Encouraged by the positive effects

of the H227I and Y228P mutations in *Ma*-MF18 for AzGGK incorporation we wanted to test the *Ma*-MF9-H227I-Y228 with N^6-acetyl-*L*-lysine (AcK) (appendix 7.1.3) which leads to moderate amber suppression efficiencies using *Mb*-MF9 (Figure 21B). We obtained sligthly higher sfGFP-V2P-N150AcK-His6 expression for samples with *Ma*-MF9-H227I-Y228 compared to *Mb*-MF9. No misincpoporation of natural amino acids could be observed with ESI-MS (sfGFP-V2P-N150AcK-His6 27751 Da) (Figure 21C). Furthermore, we demonstrated that H227A mutations in our *Ma*-PylRS mutants (*Ma*-PylRSwt; *Ma*-MF3; *Ma*-MF9; *Ma*-MF18) lead to complete aminoacylation inactivity of all enzymes and tested substrates (appendix Figure 75) indicating the importance of position 227 for substrate recognition of all examined *Ma*-PylRS mutants.

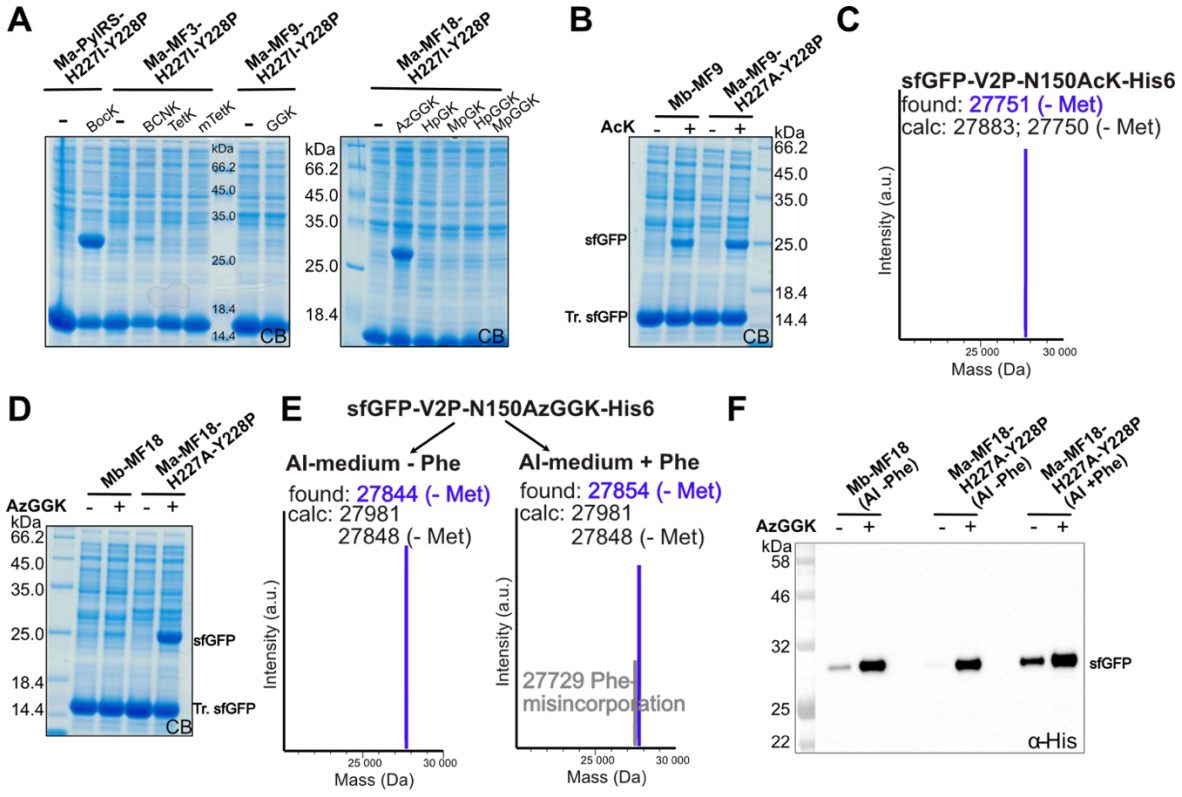


Figure 21: Amber suppression with *Ma*-PylRS mutants of sfGFP-V2P-N150TAG-His6 A: SDS gels showing the cell lysates of the test expressions with different *Ma*-PylRS variants bearing further H227I and Y228 mutations. Expression of sfGFP-V2P-N150TAG-His6 using *Ma*-PylRS-H227I-Y228P and *Ma*-MF3-H227I-Y228P were performed in 2x YT carried out in AI medium without phenylalanine (AI -Phe). B: SDS gel of cell lysates of amber suppression experiments for AcK into sfGFP-V2P-N150TAG-His6 using *Ma*-MF9-H227I-Y228P and *Mb*-MF9 in AI -Phe. C: ESI-MS data (average masses) of purified sfGFP-V2P-N150AcK-His6 obtaind from expression using *Ma*-MF9-H227I-Y228P. D: SDS gel of the comparison of cell lysates of amber suppression experiments for AzGGK in sfGFP-V2P-N150TAG-His6 using *Ma*-MF18-H227I-Y228P and *Mb*-MF18 in AI -Phe. E: ESI-MS data (average masses) of purified sfGFP obtaind from expression using AzGGK and *Ma*-MF18-H227I-Y228P in either AI -Phe or +Phe. F: Western Blot (α-His) of cell lysates obtained from expressions with *Mb*-MF18 and *Ma*-MF18-H227I-Y228P for the incorporation of AzGGK into sfGFP-V2P-N150TAG-His6 either in AI -Phe or +Phe. (CB = Coomassie brilliant blue; Met = methionine; Phe = phenylalanine; Tr. = truncated; "-" = no uAA added; AI -Phe = auto-inducing medium without phenylalanine; AI +Phe = auto-inducing medium witho phenylalanine)

During the work on the *M alvus* system a new uAA for proximity-enhanced crosslinking, N^6-((5-bromopentyl)carbamoyl)-L-lysine (BrCN7K), was developed in our lab which is explained in detail in section 4.2 and which is accepted by our published *Mb*-TEMPOH-I (also known as BrCnKRS, Table 9). Due to its similarity in molecule structure and reaction properies to BrC7K and BrCO7K, we wanted to determine if it is also recognized by *Ma*-PylRS and *Ma*-

PylRS-H227I-Y228P. The expression test detected full-length sfGFP after 6 hours and overnight only in samples supplemented with BrCN7K for all three (*Ma*-PylRS, *Ma*-PylRS-H227I-Y228P, *Mb*-TEMPOH-I) tested synthetases (Figure 22A). Largest sfGFP production was observed for samples with *Mb*-TEMPOH-I. Surprisingly, bands for sfGFP dimer were only present in the cell lysate samples of expressions using *Mb*-TEMPOH-I. This was further confirmed by the elution fractions of the Ni-purified protein samples (Figure 22A). The sfGFP dimer was only detected in the sample derived from the expression with *Mb*-TEMPOH-I whereas samples from expressions with *Ma*-PylRS and *Ma*-PylRS-H227I-Y228P showed the sfGFP monomer band. Masses obtained form ESI-MS showed the correct sfGFP-V2P-N150BrCN7K-His6 mass (27883 Da) for the sample derived from the expression using *Mb*-TEMPOH-I (Figure 22B). For sfGFP samples from expression with either *Ma*-PylRS or *Ma*-PylRS-H227I-Y228P only masses corresponding to sfGFP-V2P-N150BrCN7K-His6 without HBr were detected (Figure 22B). The absence of the bromo-moiety in BrCN7K in sfGFP explains the inability to form a sfGFP dimer during expressions with *Ma*-PylRS and *Ma*-PylRS-H227I-Y228P. Due to the loss of its bromo-moiety the uAA is inert towards nucleophilc attack of natural amino acids such as cysteine or glutamate. A hypothesis that BrCN7K alteration occurs during the ribosomal synthesis or post-translationally on sfGFP can be rejected, as it is not observed in samples originating from expressions with *Mb*-TEMPOH-I. Obviously, removal of the bromo-moiety from BrCN7K is happening during the aminoacylation by *Ma*-PylRS. The foundation of this process still needs to be investigated. Interestingly, BrCO7K is not subjected to the bromine removal as the intact mass for sfGFP-V2P-N150BrCO7K-His6 could be detected (Figure 20D).

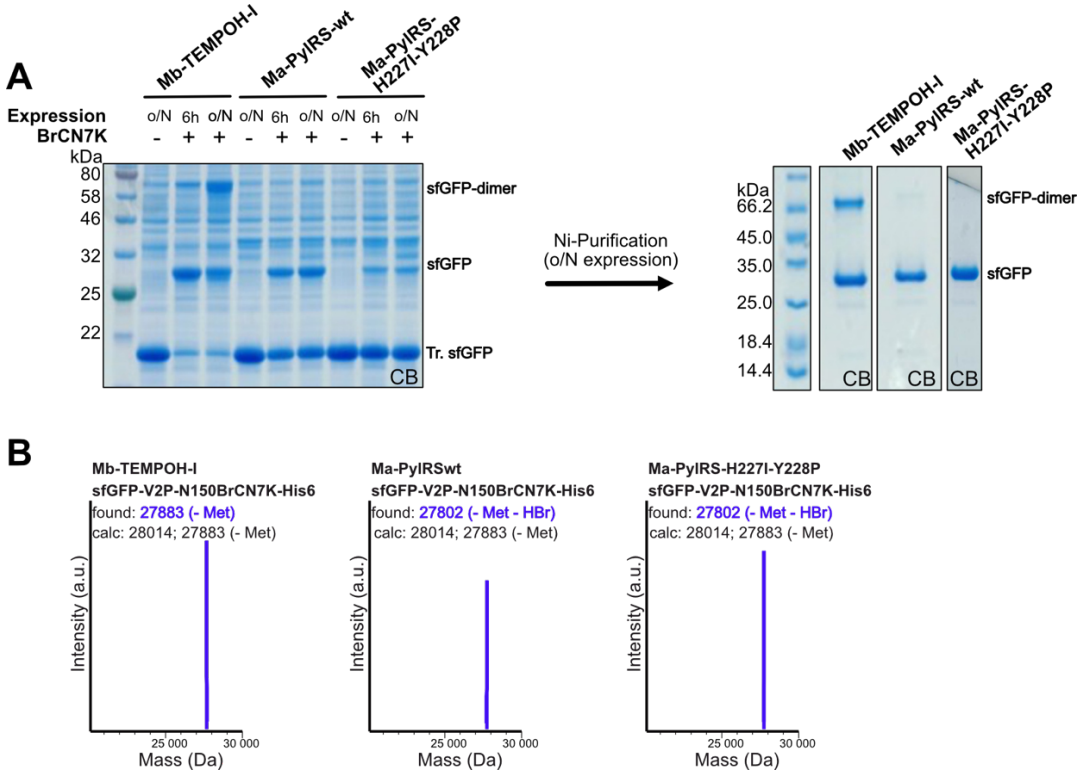


Figure 22: Comparison of expressions for the incorporation of BrCN7K into sfGFP-V2P-N150TAG-His6. A: SDS gels of cell lysates of the incorporation of BrCN7K into sfGFP-V2P-N150TAG-His6 using either *Ma*-PylRS, *Ma*-PylRS-H227I-Y228P or *Mb*-PylRS in LB and SDS gel of the purified sfGFP obtained from these expressions. B: ESI-MS data (exact masses) of purified sfGFP obtaind from expression using BrCN7K and either *Ma*-PylRS, *Ma*-PylRS-H227I-Y228P or *Mb*-TEMPOH-I. (CB = Coomassie brilliant blue; Met = methionine; Tr. = truncated; o/N = overnight; "-" = no uAA added)

To test the functionality of the new *Ma*-D65, *Ma*-D7L and *Ma*-sfGFP-CAT plasmids each selection step of the directed evolution approach was simulated for a positive and negative outcome. MW-Ma6 mimics the role of a *Ma*-PylRS library in our experiment. BocK was used as positive control as it is accepted by *Ma*-PylRS (encoded by MW-Ma6; Figure 20A) and GGK served as negative control as it is not supposed to be incorporated by *Ma*-PylRS. We co-transformed each of the new plasmids into *E. coli* NEB 10-beta cells with MW-Ma6 followed by spreading the cells on plates with or without uAA. For *E. coli* containing plasmids MW-Ma6 and *Ma*-D65 positive selection plates (Table 57) were used. Cells transformed with MW-Ma6 and *Ma*-D7L were plated on negative selection plates (Table 58) and cells harboring the combination of MW-Ma6 and *Ma*-sfGFP-CAT were cultivated on read-out plates (Table 59). The experiment provided the expected results which are summarized in Table 2.

Table 2 Output of the functionality test of the plasmids for directed evolution of *Ma*-PylRS

***Ma*-PylRS co-transformed with:**	**- uAA**	**+ BocK [1 mM]**	**+ GGK [1 mM]**
***Ma*-D65**	No colonies	colonies	No colonies
***Ma*-D7L**	colonies	No colonies	colonies
***Ma*-sfGFP-CAT**	No colonies	Green colonies	No colonies

E. coli transformation with MW-Ma6 and *Ma*-D65 serves as a simulation of a positive selection in directed evolution using chloramphenicol (Cam) as selection marker. The obtained results confirm that only cells survived which incorporated BocK in response to the amber stop codon in the chloramphenicol transferase (CAT), harbored on *Ma*-D65. As GGK is not recognized by *Ma*-PylRS no colonies grew on plates containing GGK. Plates without uAA did also not provide a substrate for *Ma*-PylRS to enable a readthrough of CAT, thus resulting in cell death. The simulation of a negative selection round was performed with cells containing MW-Ma6 and *Ma*-D7L. Here, cells only survive if barnase (encoded by *Ma*-D7L) is not produced due to missing amber suppression of the barnase TAG-codons. Cells cultivated on plates supplemented with BocK showed no colony growth due to production of full-length barnase induced by TAG-codon suppression with BocK by *Ma*-PylRS. Lots of colonies were visible for plates without uAA or with GGK added. In both cases no suitable substrate for *Ma*-PylRS was present leading to the survival of the cells. Finally, the combination of MW-Ma6 and *Ma*-sfGFP-CAT was examined to mimic the read-out step (or second positive selection step) of the directed evolution using again Cam as selection marker. Surviving green colonies indicated successful incorporation of an uAA into CAT and sfGFP. This was the case for cells incubated on plates containing the *Ma*-PylRS substrate BocK. For plates without substrates of *Ma*-PylRS (-uAA, +GGK) no colony growth was obtained. This shows that all *Ma*-plasmids for directed evolution provide the relevant *Ma*-PylT to functionally complement the system for *Ma*-PylRS as well as the corresponding positive and negative selection markers. Additionally, three *Ma*-PylRS libraries were constructed (Lib1-Ma-MF18IP, Lib2-Ma-MF18IP, EP-Lib-Ma-MF18IP; see Table 56)

2.3.3. Amber Suppression with *M. alvus* in HEK cells

To complete the establishment of the *Ma*-PylRS/tRNA system for our lab we tested the functionality of our plasmids designed for experiments in mammalian cells. The plasmid system is a fluorescence-based reporter system for amber suppression consisting of one plasmid providing *Ma*-PylRS (MM30) and the other two encoding either a single- (MM31) or a dual-reporter (MM32) containing TAG-codons at respective sites. After co-transfection into HEK293T cells of one reporter combined with the *Ma*-PylRS bearing plasmids the fluorescence was monitored by microscopy (Figure 23). The dual-fluorescence-reporter MM32 contains a TAG-codon between the two open reading frames of the fluorescent proteins mCherry and GFP. The mCherry gene is settled downstream of the EF1 promotor and the GFP gene is located in frame downstream of mCherry. The mCherry fluorescence indicates successful transfection of cells as this is independent of amber suppression events. Simultaneous GFP expression reflects readthrough at the TAG-codon as effect of successful amber suppression by *Ma*-PylRS and the respective uAA. No GFP fluorescence, *vice versa*, is a hint for terminated translation at the TAG-codon due to inadequate amber suppression. The single reporter MM31 encodes sfGFP-N150TAG-His6 and allows the readout of successful amber suppression by the green fluorescence of the transfected cells. Figure 23A displays the results for the co-transfection of MM30 and MM32 into HEK cells in medium with or without BocK. Strong mCherry fluorescence was detected for both transfections indicating that the transfections worked well. GFP expression was just obtained in the sample supplemented with *Ma*-PylRS substrate BocK which is incorporated at the amber codon thereby enabling further GFP translation. Without BocK, *Ma*-PylRS is lacking an adequate substrate to charge its cognate tRNA whicht renders amber suppression impossible. This is in accordance with our expectations for the dual-reporter system. Fluorescence microscopy of amber suppression experiments in HEK293T cells transfected with MM30 and the MM31 single reporter plasmid produced the output shown in Figure 23B. As expected, strong GFP fluorescence can just be observed in HEK cells cultivated with BocK due to full-length GFP expression enabled by the incorporation of BocK at the internal TAG-codon of GFP. For HEK cells in medium without uAA substrate for the *Ma*-PylRS no fluorescent GFP could be detected. Finally, we reproduced with success the amber suppression by *Ma*-PylRS in the GFP-N150TAG-reporter for our HEK cell culture using the obtained plasmids from the Chin lab (MRC). Moreover, we could confirm the compatibility of our cloned dual-reporter plasmid to the MM30 vector for amber suppression readout. In the past our lab found out that cloning approaches, like site-directed mutagenesis, work badly in mammalian plasmids. As this is however necessary for the creation of new *Ma*-PylRS mutants we constructed the pET19sub_Ma-PylRSwt (MM28) plasmid as cloning and transfer platform for *Ma*-PylRS mutants into pPB1-backbone. The pET19-backbone was already used as helper plasmids for site-directed mutagenesis of *Mm*-PylRS mutants in the lab of Prof. Lang making a suitable vector for *Ma*-PylRS cloning approaches. In the pPB1 plasmid (MM30) the *Ma*-gene is flanked by a NheI (upstream) and a SpeI (downstream) restriction site. As restriction cloning is the method of choice in our lab for altering mammalian plasmids, we introduced the *Ma*-PylRS gene in the pET19sub-backbone and located respective restriction sites up- (NheI) and downstream (SpeI) of the *Ma*-PylRS gene, which can be used to provide suitable inserts for ligation reactions with the digested pPB1 backbone to obtain pPB1-plasmids with desired *Ma*-PylRS mutants. Correct assembly of MM28 was confirmed by sanger sequencing.

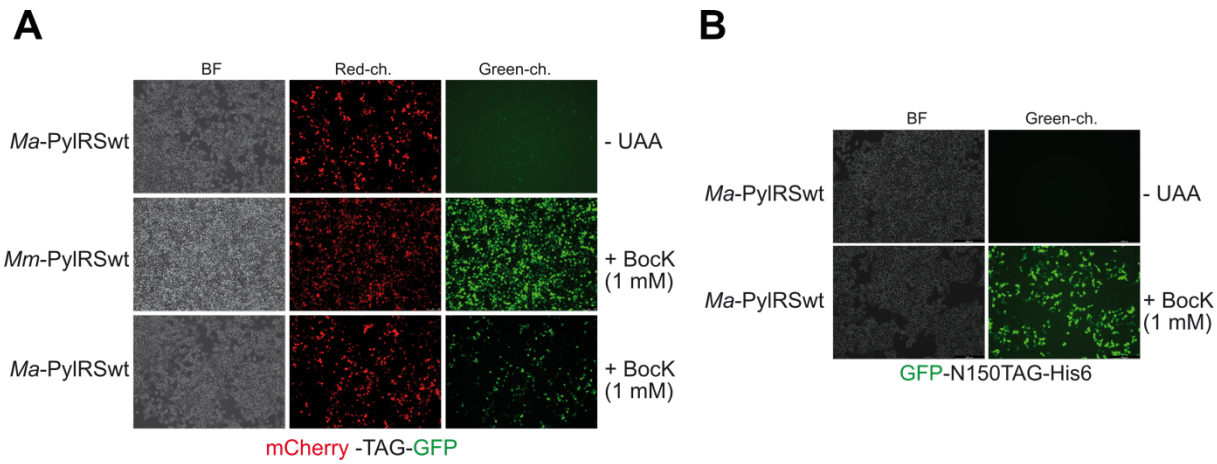


Figure 23: Fluorescence microscopy of amber suppression experiments in HEK293T cells using *Ma*-PylRS and BocK 36 hours post-transfection. A: Fluorescence output of HEK cells co-transfected with MM30 and MM32 growing in medium with or without BocK. B: Fluorescence output of HEK cells co-transfected with MM30 and MM31 growing in medium with or without BocK. (BF = bright field, ch. = channel)

2.4. Summary & Outlook

In summary, we successfully implemented the *Ma*-PylRS/tRNA pair on our pBK/pPylT-vector system for amber suppression in *E. coli*. We could show that the *Ma*-PylRS and its mutants incorporate a wide range of uAAs from our lab. Additionally, we found out that the wild type *Ma*-PylRS is highly promiscuous as it accepts several lysine-based uAAs bearing long alkyl side chains (BrC6K, BrCO6K, BrCO7K, DiaziK) which the *Mb*-PylRS wild type does not recognize. Here structure elucidation of co-crystalized *Ma*-PylRS and e.g. BrCO6K would be interesting because it might help us to understand why the wild type *Ma*-PylRS compared to the *Mm*-PylRS is able to work with these uAAs. We demonstrated that *Ma*-PylRS mutants improved the incorporation of distinct uAAs (AcK, AzGGK) compared to its *Mb* counterparts. Moreover, we used the *Ma*-PylRS system for amber suppression in HEK293T cells with success and prepared a plasmid platform for easy cloning of *Ma*-PylRS mutants and transfer into mammalian vectors. Finally, plasmids for directed evolution approaches with *Ma*-PylRS were successfully implemented as well as *Ma*-PylRS libraries.

With these achievements in hand our lab is well prepared for the application of the new *M. alvus* system for genetic code expansion in *E. coli* and mammalian cells in future experiments and evolution approaches. Currently, our lab is working on the application of *Ma*-PylRS for azobenzene uAAs. However, some aspects, like the phenomena of the “missing bromo moiety” during the amber suppression with *Ma*-PylRS or mutants and BrCN7K, still remain unclear and demand further investigation.

3. CHAPTER II – ENZYME CRYSTALLIZATION

3.1. Background

For years many groups and scientists have studied the effects and mechanisms of post-translational modifications (PTM) of proteins. The ubiquitylation of proteins plays a central role in this field because it is one of the most abundant post-translational modifications in eukaryotes and associated with intracellular regulatory processes *inter alia* protein degradation[198], endocytosis[199] or cell cycle control[200,201, 202] Ubiquitylation is the covalent attachment of one or more ubiquitin (Ub) proteins to the ε-amino groups of lysines in the target protein. Studying ubiquitylation is rather challenging as it occurs in a huge diversity considering modification pattern, topology and further output.[203]

Nature evolved a complex network for the regulation of ubiquitylation events where special enzymes, divided into the class of E1, E2 or E3 ligases or deubiquitinases (DUBs)[204], have a major role. There are two known Ub-dependent E1-enzymes[205-207], more then 40 E2-[208] and even over 600 E3-enyzmes[209]. During ubiquitylation the C-terminal carboxylate of Ub reacts with the ε-amino group of lysine in a substrate protein forming a stable isopeptide bond.[202] This event requires a three-staged cascade mediated by special E1 (Ub-activating-enzyme), E2 (Ub-carrier/conjugating-enzyme) and E3 (Ub-ligase) enzymes. E1 activates Ub by coupling the C-terminal carboxylate of Ub onto the thiol group of an active site cysteine through a thioester under consumption of ATP.[210] Subsequently, Ub is transferred from the E1-enzyme to an active site cysteine of the E2-enzyme via transthioesterification.[211, 212] In the last step an E3-Ub-ligase mediates the transition of Ub from the E2-enyzme onto the ε-amino group of lysine in its target protein.[213] However, studying these processes requires methods to produce homogenous and site-specific Ub-conjugated proteins. The limited availability of those approaches currently poses the most severe bottleneck. The generation of targeted ubiquitylated protein samples is very complicated due to the often unknown E3-Ub-ligases and thus, the investigation of the consequences of ubiquitylation on certain target proteins remains challenging.

Recently our group published a new E1/E2/E3-indpendent approach, termed sortylation, which enables post-translational and site-specific decoration of proteins with ubiquitin *in vitro* and *in vivo* (shown in Figure 24).[155] The chemoenzymatic method consists of two fundamental steps: first GCE to site-specifically introduce AzGGK into proteins which is converted via (2-(diphenylphosphino)benzoic acid (2DPBA) induced Staudinger reduction to GGK[214]. GGK serves as target in sortase-mediated transpeptidation with Ub which bears a sortase recognition motif (LALTG) at its C-terminal end. This results in an Ub-protein conjugate with a native isopeptide bond between Ub and the target protein with just two point mutations in the linker region (R72A; R74T). The AzGGK incorporation is mediated by an evolved *Mb*-PylRS mutant (*Mb*-MF18 = AzGGKRS).[155]

Taking this approach to the next level, our lab has designed photocaged GGK derivatives (Figure 25) that should elegantly allow light-controlled instead of chemical induced sortylation using AzGGK (Figure 24). In theory the photocaged GGK derivative (e.g. HpGGK) is introduced site-specifically into proteins using amber suppression. After deprotection by UV light the free GG-moiety of the side chain is available and undergoes sortase-mediated

ubiquitylation as already described for AzGGK after conversion via Staudinger reduction (Figure 24). Combining photocages with uAAs in genetic code expansion approaches was already extensively applied to convey light control with a spatial and temporal resolution.[215-217]

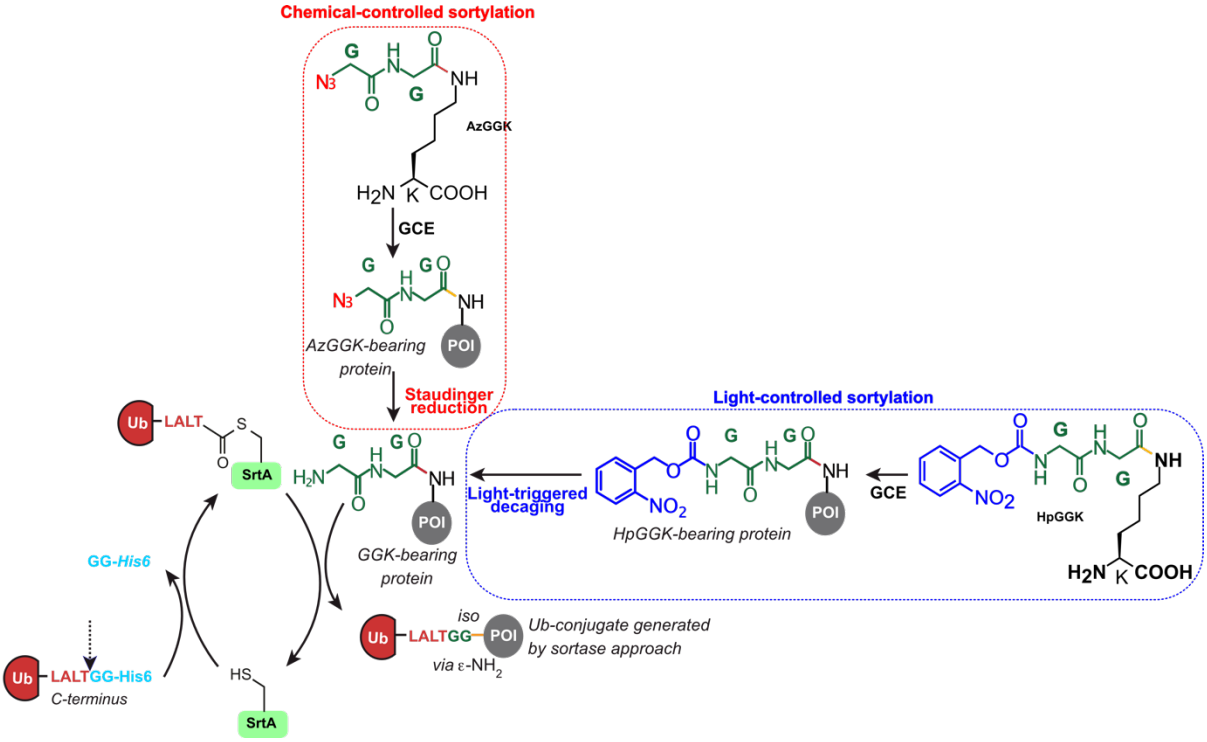


Figure 24: Scheme for the sortase-mediated ubiquitylation reaction activated either in a chemical (red dashed rectangle) or light-controlled (blue dashed rectangle) manner. An unnatural amino acid bearing a masked glycine dipeptide (GG, green) in its side chain is site-specifically incorporated into proteins via GCE. After demasking the GG-moiety of the incorporated uAA on the protein it undergoes SrtA-mediated transpeptidation with a modified Ub, which displays a C-terminal sortase recognition motif (LALTG). Two strategies for the conversion of the masked unnatural amino acid to the GG-bearing side chain are presented. AzGGK can be chemically converted by *in vivo* Staudinger reduction. Photocaged GGK-derivatives (HpGGK) need light-induction to reveal the GG-motif of the side chain. Ubiquitylated proteins derived from this approach exhibit a native isopeptide bond between Ub and the POI and two point mutations (R72A; R74T) in the linker region.[155] (Ub = ubiquitin; SrtA = sortase A; POI = protein of interest)

The critical step is the development of an aminoacyl-tRNA-synthetase mutant that specifically recognizes one or more of the new photocaged GGK derivatives. In order to increase the probability for finding a suited enzyme, Maximilian Fottner (PhD, group of Prof. Lang) synthesized a broad portfolio of different GGK derivatives shown in Figure 25. All of them are lysine based uAAs which are coupled to a nitrobenzyl-based photocage at the amino group of the terminal side chain glycine. We used two slightly different versions of the nitrobenzyl photocage with either a hydrogen (HpGK. HpGGK) or a methyl group (MpGK, MpGGK) at its benzylic CH_2-group. The photocage release is activated by UV light and undergoes a Norris Type II reaction to yield the free amine group[218]. The side chain provides either a single glycine (HpGK, MpGK) or a glycine dipeptide (HpGGK, MpGGK) which is linked via amide bond to the ε-amino group of lysine.

Unfortunately, so far neither screening of our aaRS-collection (including AzGGKRS = *Mb*-MF18) nor directed evolution with our existing *Mb*-PylRS libraries resulted in a synthetase which accepts one of the new uAAs from Figure 25. To get insights into the binding of AzGGK to *Mb*-MF18 and guide our attempts in finding a PylRS variant specific for the photocaged GGK derivatives, we aimed at solving a crystal structure of the AzGGK-bound synthetase.

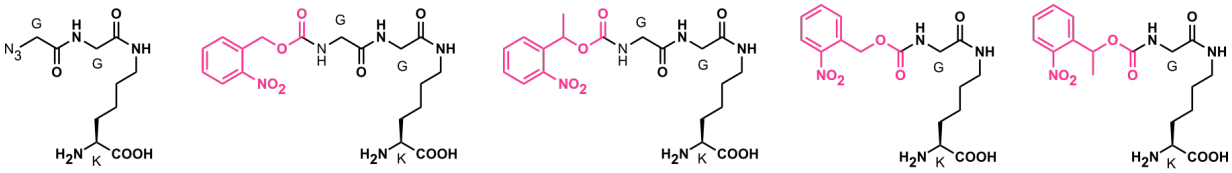


Figure 25: Moleculare structures of AzGGK and its photocaged derivates (HpGGK, MpGGK, HpGK, MpGK). The photocage is marked in pink. K indicates the lysine-derived basic structure, G indicates glycine moieties of the side chain.

3.2. Aim

Aim of the crystallization project is the gain of insights into the binding mode of pyrolysyl-tRNA synthetase mutants to their substrate AzGGK. Therefore, the goal is to co-crystallize the enzyme together with its substrate AzGGK. Crystallization will be performed for both of our synthetase mutants for AzGGK, the *Mm-/Mb*-MF18 and the new *Ma*-MF18-H227I-Y228P described in 3.3.1 and 3.3.2. Supported by the new knowledge gained from the crystal structures we will design tailor-made synthetase mutants and libraries for photocaged GGK-derivatives in order to bring the sortase-mediated ubiquitylation approach to the next level by putting it under light-control. Simultaneously, we want to contribute to the portfolio of crystal structure of *Mm*- and *Ma*-PylRS mutants. Especially, the co-crystallization of *Ma*-MF18-H227I-Y228P and AzGGK would complete the *Ma*-PylRS structure collection as there is, to the best of our knowledge, no crystal structure available which shows the enzyme or one of its mutants together with its respective substrate. The project will be comprised of cloning of respective expression vectors, expression and purification of *Mm*-MF18 and *Ma*-MF18-H227I-Y228P, set-up of crystallization experiments, structure determination, rational mutagenesis and test expression of *Mm*-MF18 mutants for GGK derivatives as well as the *Mm*-MF18-based library creation and directed evolution.
This project is a collaboration together with an expert in structural biology, Dr. Sabine Schneider (LMU), who was responsible for crystallographic data acquisition and data evaluation.

3.3. Results & Discussion

3.3.1. Crystallization of *M. mazei* PylRS

The project was started with the design of an expression vector for *Mm*-MF18 to obtain the protein in appropriate amounts. We decided to use the catalytic active C-terminal domain (CTD) of *Mm*-MF18 for crystallization as this already provided successful crystal structures in the past. [103, 109] The N-terminal domain (NTD) of the PylRS is known to be prone to aggregation, which could negatively affect crystallization.[107, 190] The T7-driven pET28 backbone was equipped with the gene encoding the CTD of *Mm*-MF18 from residue 185-454 in frame with a N-terminal His6-tag followed by a thrombin cleavage site for possible removal of the His6-tag (Figure 26A). The plasmid was suitable for efficient production of *Mm*-MF18-CTD in *E. coli* Rosetta (DE3) as indicated by a big overexpression band at approximately 33 kDa in the SDS gel from the cell lysate sample induced with Isopropyl-β-D-

thiogalactopyranoside (IPTG) (Figure 26A). Subsequent Ni^{2+}-affinity purification yielded *Mm*-MF18-CTD, which was further applied to size exclusion chromatography (SEC) to finally obtain pure fractions of *Mm*-MF18-CTD for the crystallization experiments (Figure 26A). In order to get co-crystals of *Mm*-MF18-CTD with AzGGK the enzyme was loaded *in vitro* with its substrate directly before setting up the crystallization experiments. Therefore, the purified *Mm*-MF18-CTD was incubated with AzGGK in the presence of 1 mM ATP. Excess of AzGGK and adenosine triphosphate (ATP) was removed and the *Mm*-MF18-CTD-AzGGK was adjusted to the desired concentration using centrifugal filter devises. During the crystallization screens many crystals for *Mm*-MF18-CTD were obtained which displayed different morphologies depending on the crystallization conditions (Figure 26B). They were examined with UV light as white glow of the crystal is a hint for the presence of protein (Figure 26B).

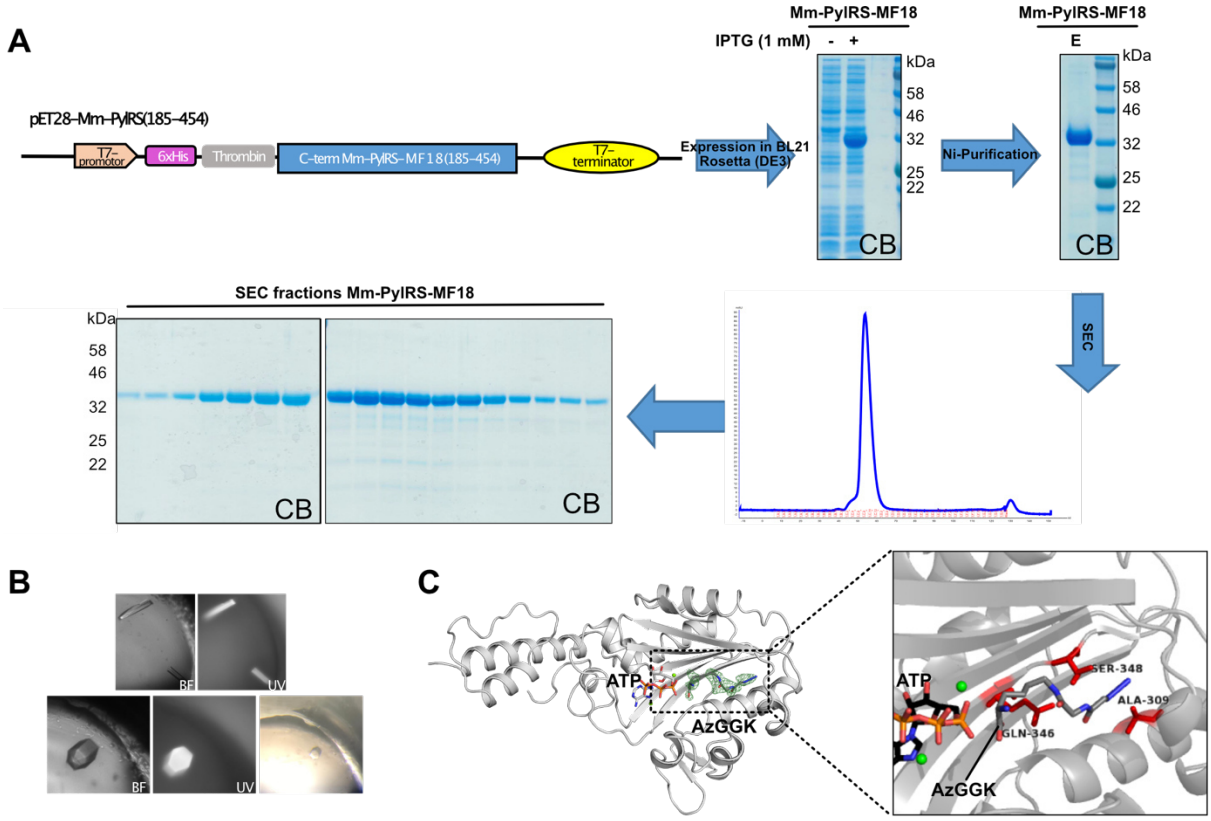


Figure 26: Approach for the crystallization of *Mm*-MF18 (*Mm*-PylRS-MF18). After expression of the C-terminal domain (CTD) of *Mm*-MF18 in *E. coli* Rosetta (DE3) cells proteins are purified in a two-step purification process consisting of a Ni-affinity purification and subsequent SEC. A: pET28-plasmid construct for the expression of N-terminally His6-tagged *Mm*-MF18 (CTD; aa 185-454) in *E. coli* Rosetta (DE3) as well as the SDS gels from the cell lysates derived from expression and the SDS gels showing the eluates from the two purification steps. Further the SEC ÄKTA chromatogram (280 nm) is illustrated. The construct of *Mm*-MF18 (blue) displays a N-terminal His6-tag (pink) with a thrombin cleavage site (grey) between His6-tag and *Mm*-MF18. The expression is controlled by a T7 promotor (orange) and T7 terminator (yellow). B: Images of *Mm*-MF18 crystals with different morphology obtained in different crystallization set-ups. Shown is the BF image and the corresponding UV image. C: Obtained crystal structure of *Mm*-MF18 (grey) in complex with ATP and its substrate AzGGK and a zoom into the enzyme binding pocket. Red color indicates residues (L309A; N346Q; C348S) which are mutated in *Mm*-MF18 compared to the wild type enzyme. Green spheres represent coordinated Mg^{2+} ions. (IPTG = Isopropyl-β-D-thiogalactopyranoside; E: Elution from Ni-purification; SEC = size exclusion chromatography; CB = Coomassie brilliant blue; BF = bright field; UV = ultraviolet; ATP = adenosine triphosphate; aa = amino acid)

Crystals of *Mm*-MF18-CTD in complex with AzGGK which finally enabled structure determination with a resolution range of 1.81 - 1.75 angstrom (Å) were obtained in 0.2 M ammonium sulphate, 15% (w/v) PEG4000 at 18°C. A summary of the data collection and

refinement statistics can be seen in appendix 7.3 Table 68. Figure 26C illustrates the crystal structure of *Mm*-MF18-CTD co-crystallized with AzGGK (AzGGK abundance in the crystals of 30 – 40 %) and ATP. The close-up view into the substrate binding pocket clarifies the location and orientation of AzGGK within the enzyme. The interactions which have an impact on AzGGK binding in *Mm*-MF18 are highlighted in Figure 27A. Mutation L309A, N346Q and C348S (*Mm* numbering, all colored in yellow in Figure 27A) within *Mm*-MF18 contribute to substrate binding. Additional residues (Y306, W417) which might positively influence AzGGK stabilization in *Mm*-MF18 are marked in green in Figure 27A. Hydrogen bonds are supposed to span distances up to 3.5 Å, π-interactions up to 6.5 Å.[219] The L309A mutation is necessary to enlarge the substrate pocket for the long side chain of AzGGK to prevent steric clashes with the leucine side chain. In order to stabilize the azide group at the end of the AzGGK side chain the C348 was mutated to serine. The distance between the hydroxy group of the serine side chain and the azide group of AzGGK is about 3.5 - 3.6 Å (measured to the first and third nitrogen-atom of the azide group). This distance enables the formation of hydrogen bond between the hydroxy group and the zwitterionic azide functionality. Furthermore, the tryptophan at position 417 might stabilize the azide through its delocalized π-electron system. The N346Q mutations brings the amide group of its side chain closer to AzGGK (distance 3.1 Å) through elongating the side chain length by one methyl group compared to asparagine. In this way the amino group of glutamine side chain can undergo hydrogen bonding with the second carbonyl group of peptide bond between the two glycine moieties. Moreover, the peptide backbone of the AzGGK side chain is positively affected by interactions with the π-electrons of the aromatic ring of the tyrosine at position 306 which is located in a distance of around 4 Å. Our crystal structure nicely explains the occurrence of the mutations found in *Mm*-MF18 and their importance for AzGGK recognition.

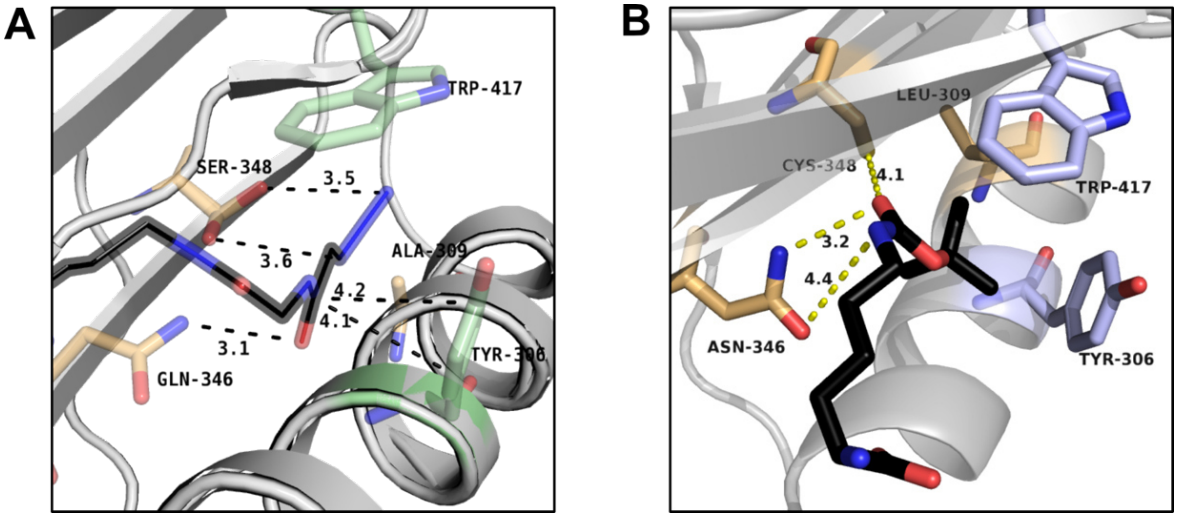


Figure 27: Binding and stabilization of respective substrates in the enzyme binding pocket of *Mm*-MF18 and the wild type *Mm*-PylRS. Mutated residues of Mm-MF18 and the corresponding residues in the wild type enzyme are marked in yellow. Further residues influencing substrate binding are shown in green for *Mm*-MF18 and AzGGK (black) and in blue for *Mm*-PylRS and BocK (black). Distances (in Å) between the respective substrates and distinct residues are illustrated by black dashed lines. A: Pocket of *Mm*-MF18 with its substrate AzGGK obtained from our crystal structure. B: Pocket of *Mm*-PylRS with its substrate BocK obtained from PDB: 2zin.

Figure 27B contrasts the binding effects in the wild type *Mm*-PylRS to its BocK substrate. Due to the shorter side chain length, BocK fits into the pocket even when leucine is present at position 309. Furthermore, the amino group of the side chain of asparagine at position 346 is

with a distance of 3.2 Å close enough to form a hydrogen bond with the N^{ε}-carbonyl group of the BocK side chain.[220] Possibly, a second hydrogen bond between the carbonyl of N346 and the amino moiety in the BocK side chain contributes to stabilization. Also, the cysteine at position 348 with a distance of 4.1 Å might coordinate BocK at its N^{ε}-carbonyl moiety. In conclusion, both structures allow to rationalize how their respective substrates are stabilized inside their binding pockets.

3.3.2. Crystallization of *M. alvus* PylRS

For the *Ma*-PylRS crystallization approach we decided to use the same expression system that was already proven for the crystallization with *Mm*-MF18-CTD (see 3.3.1). Five different *Ma*-PylRS mutants were cloned into the pET28-vector (Table 3) in order to have operational plasmids for future crystallization experiments with other *Ma*-PylRS mutants. In this study only the construct bearing *Ma*-MF18-H227I-Y228P (MW-Ma5) was used, which is schematically shown in Figure 28A. Compared to the crystallization approach with *Mm*-MF18 the whole enzyme for *Ma*-PylRS (amino acids 1-275) was used as it has no aggregation-prone N-terminal domain.[188]

Table 3 Plasmids for crystallization experiments with *M. alvus* PylRS

Code	Plasmid	Application
MW-Ma1*	pET28_His-Throm-Ma-PylRSwt	Crystallization
MW-Ma2*	pET28_His-Throm-Ma-PylRS-MF3 (Y126G)	Crystallization
MW-Ma3*	pET28_His-Throm-Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)	Crystallization
MW-Ma4*	pET28_His-Throm-Ma-PylRS-MF18 (M129A, N166Q, V168S)	Crystallization
MW-Ma5	pET28_His-Throm-Ma-PylRS-MF18 (M129A, N166Q, V168S)-H227I-Y228P	Crystallization

* The expression of the constructs MW-Ma1 to MW-Ma4 from Table 3 is shown in the appendix Figure 77A.

As for *Mm*-MF18-CTD good overexpression of *Ma*-MF18-H227I-Y228P was obtained in *E. coli* Rosetta (DE3) cells, which were induced with IPTG (Figure 28A). Also, the Ni^{2+}-affinity purification worked well, which is indicated by the SDS gel band for *Ma*-MF18-H227I-Y228P (*Ma*-MF18IP) at approximately 31 kDa (Figure 28A). In the beginning the Ni-elution was directly subjected to SEC, because crystallization with *Ma*-MF18IP was set up without further cleavage of the N-terminal His6-tag as before with *Mm*-MF18-CTD. The AzGGK-loading occurred in the same manner as for *Mm*-MF18-CTD directly before preparing the crystallization plates. Unfortunately, the crystals which were obtained for *Ma*-MF18IP did not yield sufficient resolution for the data evaluation process in structure elucidation. Therefore, we decided to cleave off the His6-tag via thrombin digest in-between the two purification steps which worked successfully (Figure 28A). Thrombin with a molecular weight of 37 kDa has similar size as *Ma*-MF18IP which renders proper separation via SEC impossible. Thus, thrombin was removed prior to SEC using p-aminobenzamidine agarose (pABA). This finally resulted in pure tag-less *Ma*-MF18IP, which was also confirmed by ESI-MS (30933 Da; (calc.: 30939 Da)) (Figure 28A). ÄKTA chromatograms of the purification steps are shown in appendix Figure 77B and C. New crystallization experiments were then started with the tag-free *Ma*-MF18IP. Finally, the structure could be solved with resolution range of 1.55 - 1.50 Å from crystals of *Ma*-MF18IP in complex with ATP (Figure 29B), which were obtained in a

manual grid screen in 0.15 M MES, pH 6.0 acid 18 % (w/v) PEG 8000. To our best knowledge, we were able to derive the first crystal structure of a *Ma*-PylRS mutant (data collection and refinement statistics can be seen in appendix 7.3 Table 68). Unfortunately, there was no free electron density corresponding to AzGGK. Several modified uAA-loading processes for *Ma*-MF18IP were tested, using higher AzGGK concentration, prolonged incubation time and non-hydrolysable ATP analogous (ApCpp). Nevertheless, so far, no *Ma*-MF18IP crystal structure in complex with AzGGK could be obtained. To our knowledge, all published crystal structures of *M. alvus* synthetases (PDB: 6ezd & 6jp2) do not contain a co-crystallized substrate so far. It therefore seems to be in the nature of the protein that a co-crystallization is difficult to achieve. The remaining options are to refine the uAA-loadding method or to try the uAA-capture artificially.

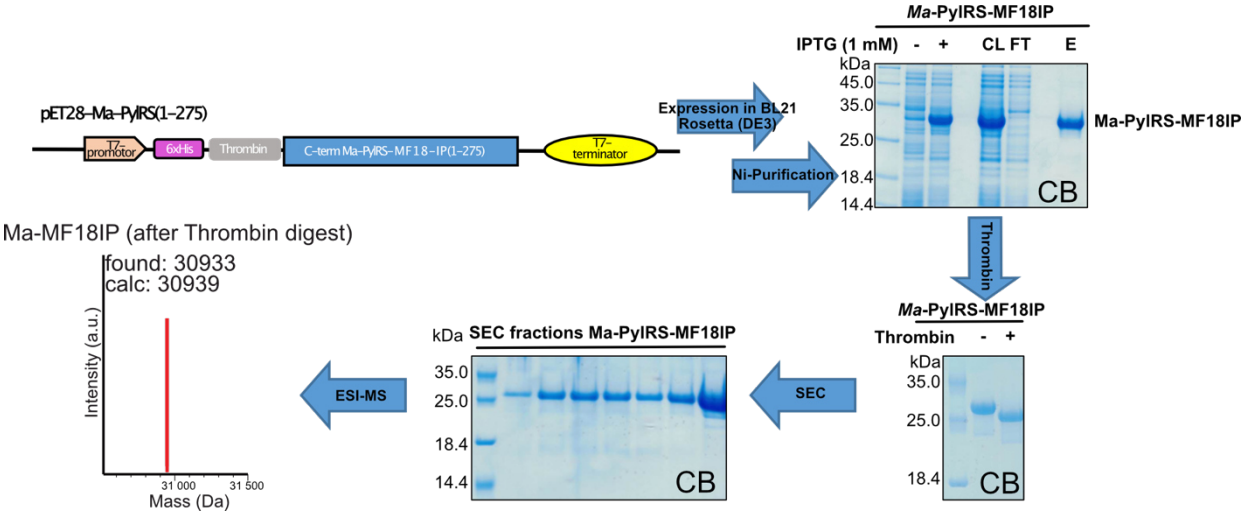


Figure 28: Approach for the crystallization of *Ma*-MF18-H227I-Y228P (*Ma*-PylRS-MF18IP). After expression of *Ma*-MF18IP in *E. coli* Rosetta (DE3) cells, proteins are purified in a two-step purification process consisting of a Ni-affinity purification and subsequent SEC with an in-between thrombin digest. pET28-plasmid construct for the expression of N-terminally His6-tagged *Mm*-MF18 (CTD; aa 185-454) in *E. coli* Rosetta (DE3) as well as the SDS gels from the cell lysates derived from expression, the Ni-purification, the thrombin digest and the SEC. The construct of *Ma*-MF18IP (blue) displays a N-terminal His6-tag (pink) with a thrombin cleavage site (grey) between His6-tag and *Mm*-MF18. The expression is controlled by a T7 promotor (orange) and T7-terminator (yellow). Furthermore, the ESI-MS data (average mass) of purified and thrombin digested *Ma*-MF18IP is shown. (IPTG = Isopropyl-β-D-thiogalactopyranoside; SEC = size exclusion chromatography; CB = Coomassie brilliant blue)

Figure 29B illustrates the *Ma*-MF18IP structure in complex with ATP. The close-up view into the binding pocket contains a modeled AzGGK and the mutation sites compared to the wild type enzyme are colored in brown (Figure 29C). We suspect that the M129A, N166Q and V168S mutations in *Ma*-MF18IP have similar influence on AzGGK binding as the L309A, N346Q and C348S in *Mm*-MF18. The additional two mutations of the second order layer residues, H227I and Y228P, which are absolutely necessary for AzGGK incorporation (see 2.3.2), cannot be clearly addressed by the obtained structure. Both mutations are located in a hairpin (ß5-ß6) motif which seems to form a kind of lid over the substrate binding pocket.[189] In *Ma*-MF18IP the Y228P mutation might stabilize the hairpin in a certain steric conformation which is required for AzGGK binding. Moreover, Y228P could affect the main-chain conformation of surrounding residues such as L229.[189] Proline is known as "helixbreaker"[221] and is the only amino acid with no hydrogen in its peptide bond (imino acid) which in turn cannot contribute to hydrogen bond formation. Aside from that, proline has a limited set of phi/psi angles (dihedral angles) compared to the other natural amino acids.[222, 223] Consequently,

it has a special role in the secondary structure of proteins and is often related to cis-configurated peptide bonds.[224] For the H227I mutation Seki *et al.* proposed, beside the influence on first order layer residues, an effect on the dynamic of the open-closed equilibrium of the ß5-ß6 hairpin (as mentioned above, this hairpin acts like a lid over the substrate pocket) in *Ma*-PylRS which might lead to an alteration of enzymatic activity.[189] It is conceivable that similar phenomena could also apply to our structure. An exact understanding of these mutations could be possibly gained through a structure of *Ma*-MF18IP with bound AzGGK.

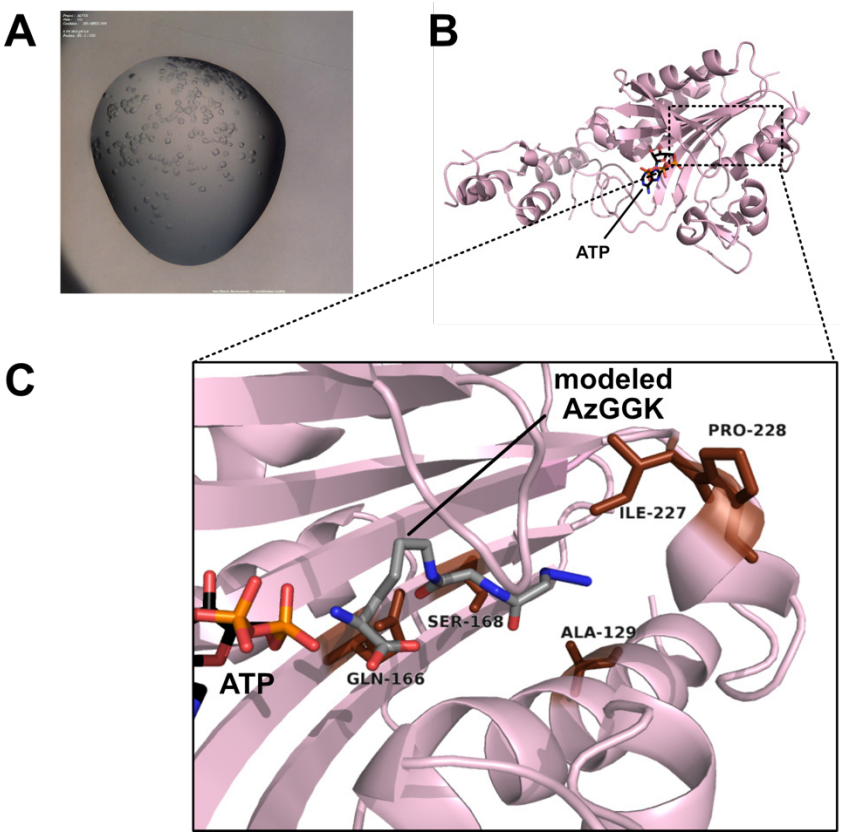


Figure 29: Crystal structure of *Ma*-MF18-IP A: Image (BF) of a drop containing Ma-MF18IP crystal obtained during the crystallization experiment. B: Obtained crystal structure of *Ma*-MF18IP (pink) in complex with ATP and C: a zoom into the enzyme binding pocket. Brown color indicates residues (M129A; N166Q; V168S; H227I; Y228P) which are mutated in *Ma*-MF18IP compared to the wild type enzyme. The AzGGK is modeled manually into the binding pocket. (BF = bright field, ATP = adenosine triphosphate)

3.3.3. Rational designed PylRS mutants & libraries

Before starting with the design of synthetase mutants for GGK derivatives the decaging properties of the new uAAs were tested exemplary on free Boc-HpGGK (Figure 30). Through UV irradiation HpGGK should be converted to GGK by release of the photocage (Figure 30A). Figure 30B shows the UV traces at 193 nm from Boc-HpGGK before and after UV light exposure (365 nm) for 120 min. The blue colored chromatogram belongs to the sample without UV treatment and has a distinct peak at 8.4 min which can be addressed to the Boc-HpGGK. The photocage confers a high UV activity to the molecule that is expressed in the clearly visible peak of the UV trace. The chromatogram after 120 min of UV light (green) lacks the peak at 8.4 min completely indicating the full conversion of Boc-HpGGK to Boc-GGK which itself has

no noticeable UV activity according to the missing photocage. The experiment demonstrates that the nitrobenzyl photocage can be removed completely from the uAA by UV light at 365 nm. The exposure time necessary for full conversion of the uAA on protein will have to be established separately for each protein again as this is always affected by position-specific effects.

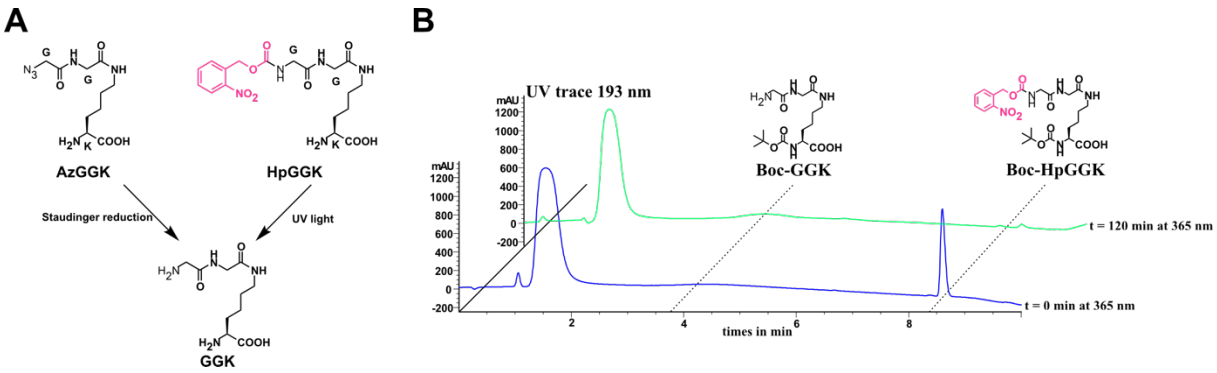


Figure 30: Demasking AzGGK and its photocaged derivative HpGGK to reveal the free GG-bearing side chain. A: Scheme of the two demasking strategies: AzGGK undergoes Staudinger reduction to yield GGK whereas light-triggered release of the photocage (pink) converts HpGGK to GGK. B: Decaging Boc-HpGGK by UV light at 365 nm. Shown are the UV traces (193 nm) obtained during ESI-MS of Boc-HpGGK samples (1 mM in PBS) before (blue chromatogram) and after (green chromatogram) irradiation for 120 min.

Based on the knowledge from our new crystal structure the goal was to rationally design synthetase mutants for the incorporation of photocaged GGK derivatives. Therefore, we concentrated on the obtained *Mm* structure because AzGGK was present in its pocket. *Mm*-PylRS and *Mb*-PylRS only differ in their N-terminal regions where *Mb*-PylRS lacks 35 amino acids (aa) but their CTDs are equal concerning amino acid composition and substrate acceptance. Consequently, *Mm*-PylRS can be substituted by *Mb*-PylRS and *vice versa*. *Mb*-MF18 served as basis for our new mutants. HpGGK was modeled into the binding pocket of the *Mm*-MF18 structure (Figure 31A) allowing the identification of the most promising residues (*Mm*-nomenclature: A309, I322, M350, S348, I413 colored yellow; corresponding *Mb*-nomenclature: A274, I287, M315, S313, I378), which might be good candidates for mutation to enable its recognition. We hypothesized that mutations to smaller residues like glycine or alanine (e.g. A271G, M315G, I378A, S313G, I287A) would enlarge the pocket in order to prevent steric clashes and to make space for the bulky photocage of HpGGK. Mutations to residues like asparagine, glutamine (I378N or Q) or cysteine (M315C) were hypothesized to provide favorable interaction effects toward HpGGK to stabilize it inside the pocket. A collection of *Mb*-MF18 variants was created with up to four mutations by altering the respective sites (A274, I287, M315, S313, I378) in *Mb*-MF18 and which are listed in Table 4.
Then the performance of the each *Mb*-MF18 mutant was tested in test expressions with AzGGK and GGK derivatives for the incorporation in sfGFP-V2P-N150TAG-His6 (Figure 31B & appendix Figure 76). Unfortunately, the western blots show that none of the mutants was able to recognize neither AzGGK nor a photocaged GGK derivative. Therefore, our structural guided rational approach seemed not to yield the right positions and required mutations to boost the synthetase towards the recognition of our desired uAAs. It should also be noted that none of the mutations provoked misincorporation of a natural amino acid.

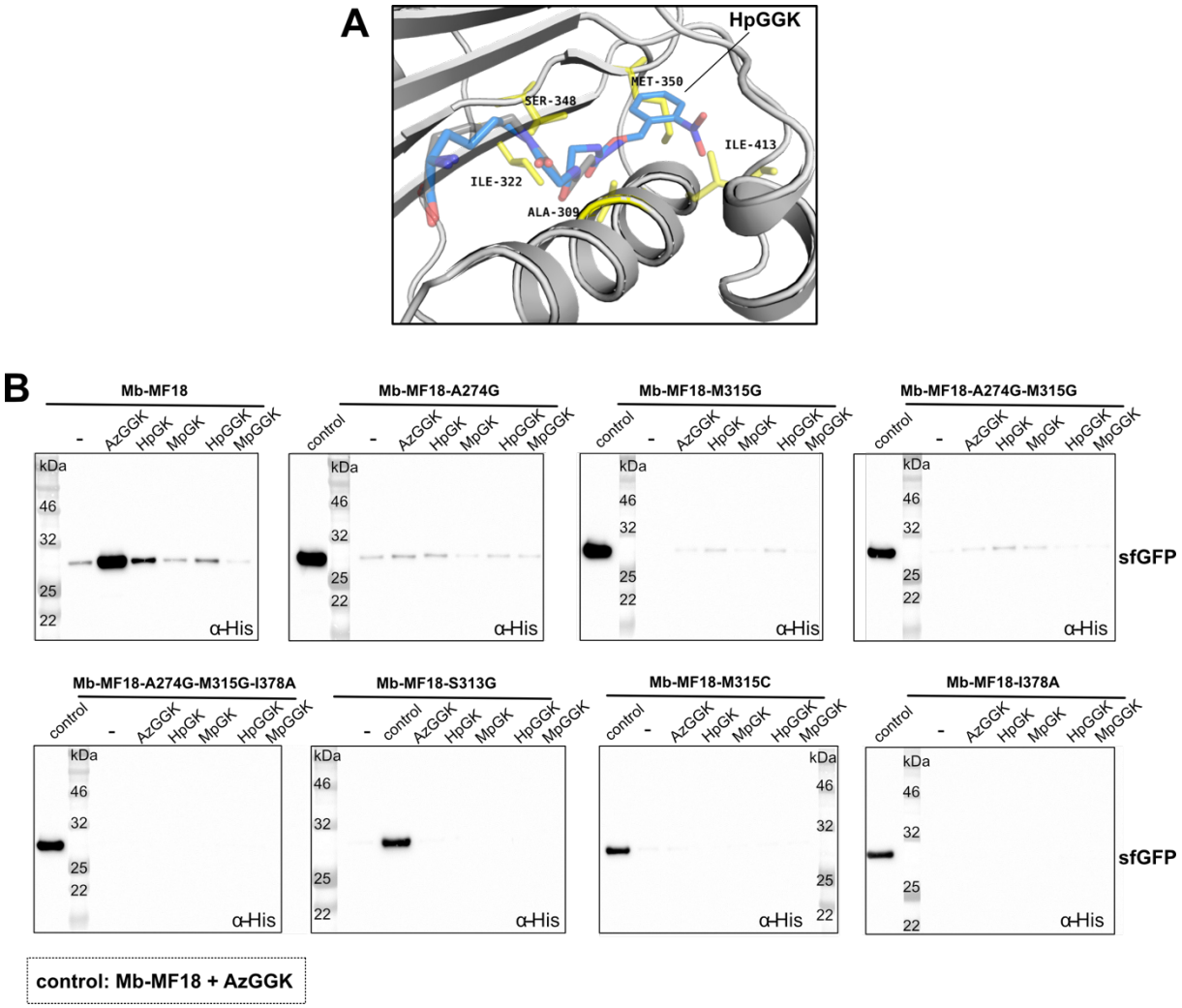


Figure 31: *Mb*-MF18 mutants for incorporation of GGK-derivatives. A: View into the crystal structure of the *Mm*-MF18 substrate binding pocket (grey) with modeled HpGGK (blue) inside. In yellow are shown potential residues candidates for mutation to enable HpGGK recognition. B: Western blots (α-His) of *E. coli* cell lysates obtained from test expressions with GGK derivatives and sfGFP-V2P-N150TAG-His6 using *Mb*-MF18 mutants. A cell lysate from an expression using AzGGK and *Mb*-MF18 served as positive control for successful sfGFP production. As negative control (-control) an expression without uAA addition was used.

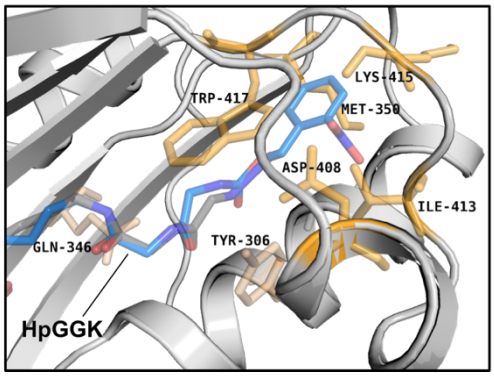
Table 4 *Mb*-PylRS-MF18 mutants for GGK derivative incorporation

Code	Plasmid
MW-S7	pBK_Mb-PylRS-L274A-N311Q-C313S (MF18)
MW-S8	pBK_Mb-MF18(A274G, N311Q, C313S)-A274G
MW-S9	pBK_Mb-MF18(L274A, N311Q, C313S)-M315G
MW-S10	pBK_Mb-MF18(L274A, N311Q, C313S)-I378A
MW-S11	pBK_Mb-MF18(L274A, N311Q, C313S)-S313G
MW-S12	pBK_Mb-MF18(L274A, N311Q, C313S)-I378N
MW-S13	pBK_Mb-MF18(L274A, N311Q, C313S)-I378Q
MW-S14	pBK_Mb-MF18(L274A, N311Q, C313S)-M315C
MW-S15	pBK_Mb-MF18(L271A, N311Q, C313S)-A274G-M315G
MW-S16	pBK_Mb-MF18(L274A, N311Q, C313S)-A274G-I378A
MW-S17	pBK_Mb-MF18(L274A, N311Q, C313S)-M315C-I378Q

MW-S18	pBK_Mb-MF18(L274A, N311Q, C313S)-M315C-I378N
MW-S19	pBK_Mb-MF18(L274A, N311Q, C313S)-A274G-M315C-I378A
MW-S20	pBK_Mb-MF18(L274A, N311Q, C313S)-A274G-M315G-I378A
MW-S21	pBK_Mb-MF18(L274A, N311Q, C313S)-A274G-M315G-I287A-I378A

In parallel, another approach started with the creation of *Mb*-MF18-based libraries specially designed for the acceptance of GGK derivatives. Again, the residues (*Mb*-nomenclature: Y271, I287, Q311, M315, D373, I378, K380,W382 ; *Mm*-nomenclature: Y306, I322, Q346, M350, D408, I413, K415,W417) which were supposed to be important for the recognition of our desired substrates based on the *Mm*-MF18-CTD structure with modeled HpGGK (Figure 32A) were chosen for randomization. In summary, three different *Mb*-MF18-based libraries with defined randomized positions and one randomly randomized error prone library of *Mb*-MF18 (EP-Lib-Mb-MF18) (Figure 32B) were cloned. Lib-Mb-MF18-1 and Lib-Mb-MF18-2 were derived using standard NNK-primers. Lib-Mb-MF18-3 was obtained via randomization with trimer-based NNK-primers and shows an especially good randomization (see 5.2.5.1). That was expected as the trimer-based primer synthesis has no codon bias and stop-codons can be excluded.

A



B

Mb-MF18 Libraries	Randomization sites
Lib-Mb-MF18-1	Y271 + Q311
Lib-Mb-MF18-2	I287 + M315 + I378
Lib-Mb-MF18-3	M315 + D373 + I378 + K380 + W382
EP-Lib-Mb-MF18	random

Figure 32: *Mb*-MF18-based libraries for directed evolution with HpGGK. A: View into the crystal structure of the *Mm*-MF18 substrate binding pocket (grey) with modeled HpGGK (blue) inside. In yellow are shown potential residues (yellow + ocher) for randomization in libraries designed for directed evolution with HpGGK. B: Overview of the created *Mb*-MF18-based libraries with their respective randomization sites.

With all libraries directed evolution approaches on HpGGK were performed. The evolution with the EP-Lib-Mb-MF18 yielded no colonies after negative selection. For the other three evolutions using Lib-Mb-MF18-1, Lib-Mb-MF18-2 and Lib-Mb-MF18-3 the first positive and negative selection rounds yielded colonies. Most clones were obtained for the evolution with Lib-Mb-MF18-3. The overall number of surviving clones was quite low after negative selection in all three evolution approaches. Maybe this can be traced back to the fact that *Mb*-MF18 is known for phenylalanine misincorporation. Possibly many colonies die from phenylalanine incorporation into barnase thereby losing also mutants which might simultaneously recognize HpGGK and could be useful when expressing them in medium without phenylalanine. So negative selection provides definitively a set screw which should be further optimized. During the dual-positive selection (with sfGFP-CAT as selection plasmid) no colonies could be obtained for Lib-Mb-MF18-1 and Lib-Mb-MF18-2. The evolution with Lib-Mb-MF18-3 yielded a few clones, but none of them showed sfGFP expression. Nonetheless, the clones were picked and the readout performed, but no GFP-fluorescence could be measured, implying that

the cultures did not contain a *Mb*-MF18 mutant which accepts HpGGK. In summary, our efforts did not produce a *Mb*-MF18 mutant specific for HpGGK. Hence, we are still searching an aminoacyl-tRNA synthetase which is able to recognize any of our photocaged GGK derivatives.

3.4. Summary & Outlook

The project yielded two new crystal structures of pyrolysyl-tRNA sythethase mutants, *Mm*-MF18 and *Ma*-MF18-H227I-Y228P. The crystallization of *Ma*-MF18-H227I-Y228P offered a structure in complex with ATP but without AzGGK in its pocket. For *Mm*-MF18 we obtained co-crystallization of the enzyme with ATP and its substrate AzGGK. This provided insights into the interactions of the binding pocket residues with AzGGK thereby contributing to the understanding of the role of the mutations in *Mm*-MF18 necessary for AzGGK recognition. Based on the knowledge gained from the *Mm*-MF18 structure with AzGGK we identified residues which might be relevant for HpGGK binding by modeling HpGGK into the binding pocket. Mutations on these residues were used to rationally design *Mb*-MF18 mutants for amber suppression with photocaged GGK derivatives. Additionally, we created tailor-made *Mb*-MF18 libraries for directed evolution with HpGGK. Unfortunately, we were not successful in obtaining or evolving suitable *Mb*-MF18 mutants which recognize HpGGK or another derivative to date.

In order to find a synthetase which accepts HpGGK further directed evolution approaches with other libraries based on either *Mb*-MF18 or *Ma*-MF18-H227I-Y228P have to be performed. Another promising aspect could be a *Mb*-MF18 or *Ma*-MF18-H227I-Y228P mutant design based on the prediction of candidates by computational biochemistry approaches such as molecular modeling and molecular dynamics. Moreover, new crystallization experiments with *Ma*-MF18-H227I-Y228P and AzGGK should be performed to finally obtain the desired crystal structure of the synthetase bearing AzGGK inside the binding pocket.

4. Chapter III - Chemical Crosslinking

4.1. Stabilization of Rab1b-DrrA Complexes

Parts of this project are published in the paper from Cigler & Müller *at al.*[172]. A manuscript for the Rab1b-DrrA-ATase project (Du, v. Wrisberg *et al.*) is in preparation and will be submitted soon.

4.1.1. Background

Rab1b (Figure 33A) is a small human GTPase and has a central role in regulating vesicular trafficking between the endoplasmatic reticulum (ER) and the Golgi apparatus.[225, 226] Rab1b is active in its GTP-bound form and inactive when bound to GDP. Depending on the bound nucleotide Rab1b can adopt different conformational states and works thereby as a molecular switch.[226]

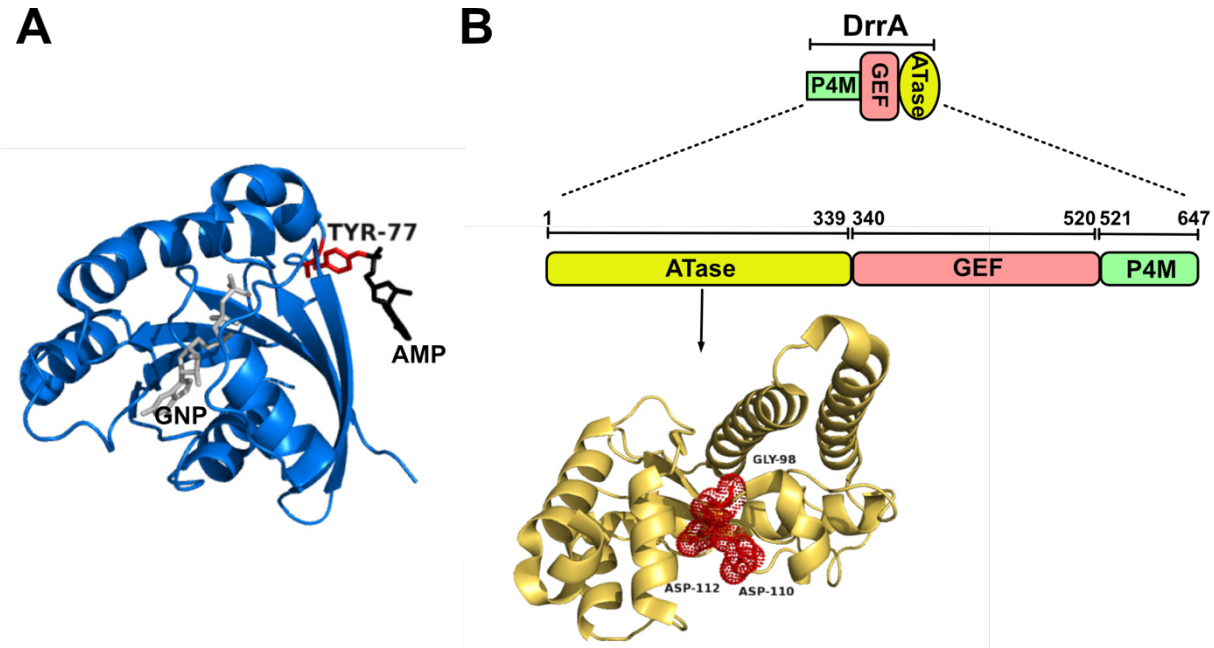


Figure 33: Schematic domain organization of DrrA and crystal structures of DrrA-ATase and Rab1b. A: Crystal structure of small GTPase Rab1b (blue; PDB: 3nkv) bound to GNP (grey) and covalently modified with AMP (black) at Y77 (red). B: Architecture of the *Legionella* virulence factor DrrA. The protein consists of three domains, an ATase domain (yellow), a GEF domain (salmon-pink) and a P4M domain (green). Start and end points of each domain are marked by the respective residues. The crystal structure of the ATase domain (PDB: 3NKU) with its active site residues (red) is shown below. (ATase = AMPylation domain, GEF = guanosine exchange actor, AMP = adenosine monophosphate, GNP = guanosine 5′-[β,γ-imido]triphosphate, P4M = phosphatidylinositol-4-phosphate binding)

There are five structural loops involved in the conformational change of Rab1b known as G1-G5. These important structural elements directly play a part in the binding of either GTP or GDP.[227] The loops G2 (switch I) and G3 (switch II) are especially important for the discrimination between GDP and GTP through the interaction with its γ-phosphate. Moreover, they mainly contribute to the conformational switching of Rab1b during GDP/GTP exchange. The role of G1 is the recognition of the GDP/GTP ß-phosphate while G4 and G5 are involved

in binding of the guanosine base.[227] The switching between the two Rab1b states is mediated by accessory proteins GAPs and GEFs (Figure 34). GTPase activating proteins (GAPs) induce the hydrolysis of GTP by Rab1b. Guanosine exchange factors (GEFs) feature the release of GDP from Rab1b and the binding of new GTP.[226] Inactive Rab1b is located in the cytosol and bound to its guanosine dissociation inhibitor (GDI) until it is activated and recruited to membranes of the ER or Golgi apparatus where it regulates vesicular docking and fusion (Figure 34).[225, 228]

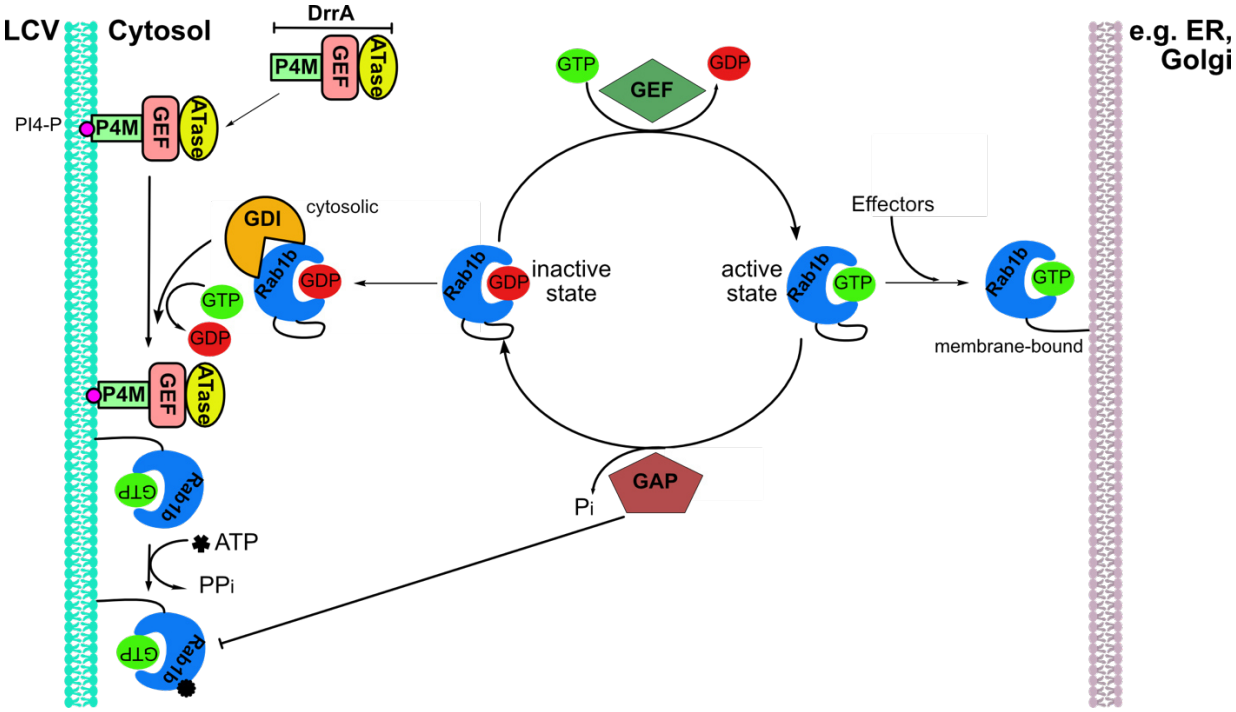


Figure 34: Schematic illustration of the Rab1b cycle and the intervention of DrrA. The cycle of Rab1b encompasses inactive (GDP-bound) and active (GTP-bound) states of Rab1b which can be interchanged by GEF and GAP protein stimulation. Inactive Rab1b in the cytosol is bound to GDI. Active Rab1b is recruited and bound to membranes. During infections, *Legionella* protein DrrA is accumulated at the PI4-M-rich LCV membrane through the interaction of its P4M domain with PI4-P. DrrA recruits cytosolic Rab1b and its GEF domain catalyzes the activation of Rab1b mediated by the exchange from GDP to GTP as well as the membrane anchoring at the LCV. Furthermore, the DrrA-ATase activity AMPylates Rab1b at position Y77. Rab1b is thereby fixed in a constitutive active from and resistant against GAPs. (GDI = guanine nucleotide dissociation inhibitor, GEF = guanosine exchange factor, GAP = GTPase activating proteins, GDP = guanosine diphosphate, GTP = guanosine triphosphate, PI4-P = phosphatidylinositol-4-phosphate, ATP = adenosine triphosphate, AMP = adenosine monophosphate, LCV = *legionella* containing vacuole, ER = endoplasmatic reticulum, P_i = phosphate, PP_i = diphosphate)

During infections with the pathogen *Legionella pneumophila,* the vesicular trafficking of the host cell is dysregulated by virulence factors of the bacterium. After phagocytosis, the host cell normally kills bacteria in the phagosome via the endocytic pathway, a process which is regulated by host Rab proteins. *Legionella* can prevent the phagosome maturation through its secreted effector proteins which interfere with those Rab proteins. Even more, it manipulates the host in a way that it provides all necessary components to the bacterium for its growth and replication.[229, 230] This enables the intracellular survival of the pathogen and its replication within a compartment called the *Legionella* containing vacuole (LCV).[229, 230] One of these *Legionella* virulence factors is the multidomain protein DrrA. It is an important manipulator of Rab1b and consists of three domains: GEF domain for the exchange from GDP to GTP in Rab1b[231], an AMPylation domain (ATase; active site motif G-X_{11}-D-X-D motif corresponding

to G98, D110, D112) for installing adenosine monophosphate (AMP) at position Y77 in Rab1b[232] and a P4M (phosphatidylinositol-4-phosphate binding) domain for the association with the PI4-P (phosphatidylinositol-4-phosphate)-rich LCV membrane[233] (Figure 33B). DrrA recruits cytosolic Rab1b to the LCV where it is activated by the GEF domain. Furthermore, the DrrA-ATase is responsible for decorating Rab1b with AMP at Y77, keeping it thereby in a constitutively active state which is resistant towards the stimulation of GAP (Figure 34).[232] Through these manipulations by DrrA, Rab1b contributes to the persistence of the *Legionella* infection.

The small human GTPase Rab1b and its effectors are the research field of the Itzen lab (UKE, formerly TUM). To investigate the Rab1b-DrrA-GEF interface our lab developed in collaboration with the Itzen lab and M. Müller (MPI Dortmund) a GCE-based crosslinking approach to stabilize the low-affinity ternary complex of GDP-bound Rab1b and DrrA-GEF for successful structural elucidation.[172] Nevertheless, the interaction between Rab1b and the ATase domain of DrrA remains structurally unknown and is a blind spot in the Rab1b-DrrA biology of *Legionella* infections. To remedy this, we dedicated our attention towards the stabilization of the Rab1b-DrrA-ATase complex using the same approach as proven for Rab1b-DrrA-GEF.

The basic principle of our planned approach involves the site-specific incorporation of an electrophilic uAA into one protein partner/protein A, which is inert under physiological conditions does not react with components of the host translational machinery but reacts with proximal natural nucleophilic residues of an interaction partner of protein A (protein B, Figure 35). Other nucleophilic amino acids remain unaffected. As electrophilic uAAs serve bromoalkyl-bearing lysine derivates with different linker lengths (BrCnK), which can be incorporated into proteins using GCE (Figure 35).[172, 234]

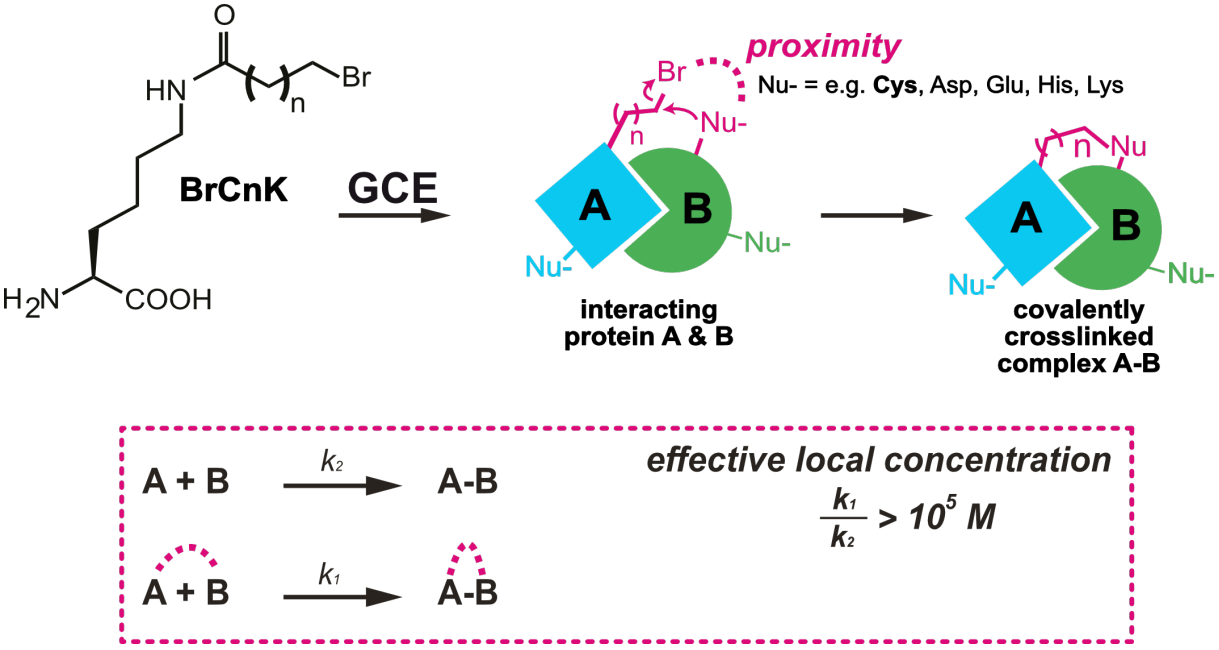


Figure 35: Scheme of the proximity-triggered crosslinking approach for protein complexes using bromoalkyl bearing uAAs and GCE. An electrophilic uAA (pink) with a long alkyl side chain attached to a bromide is site-specifically introduced into protein A (blue) via GCE. The electrophilic uAA can react with nucleophilic amino acids (Nu-, pink) closed by due to a massive rate enhancement triggered by the increasement of the local concentration through proximity. In this way covalent crosslinking with an interacting protein B (green) can be achieved.

The reaction is enabled through the massive rate enhancement by raising the effective local concentration of the reaction partners mediated by proximity.[170] This effect can lead to an increase of the effective local concentration up to 10^5 M [171]. The reaction between the two proteins A and B can be assumed as either unimolecular or bimolecular. For a bimolecular reaction of the separate proteins A and B the reaction rate relies on random encounter of both components A and B and thus is directly dependent on their concentrations. For two interacting proteins A and B their interface can be seen figuratively as connection/link so that they do not longer act separately. The proximity effect drastically increases the local concentrations of both proteins A and B thereby mimicking a kind of unimolecular reaction. Figure 35 shows the schematic reaction mode of proximity-triggered crosslinking between proteins A and B. Therefore, a specific protein crosslinking between the protein partners is ensured as the reaction occurs only at protein sites in proximity. The required proximity between two proteins can normally just be obtained at distinct interaction interfaces.

4.1.2. Aim

Aim of the project is the stabilization of the transient Rab1b-DrrA-ATase complex in order to facilitate its structural elucidation. Therefore, we plan to use the same *in vivo* chemical crosslinking approach as Cigler & Müller *et al.*[172] for the stabilization of the Rab1b-DrrAGEF complex. This approach uses bromoalkyl bearing uAAs which are introduced into proteins via GCE. For our project we require the establishment of a robust expression system providing all components for amber suppression as well as identification of suitable sites for uAA incorporation into one protein partner and corresponding cysteines into the other. The determination of suitable sites for crosslinking remains especially challenging as there is no structural information about the low-affinity Rab1b-DrrA-ATase interface available. Additionally, we want to implement a well applicable purification process for the crosslinked complex which yields pure samples for crystallization experiments.
The project is conceived as collaboration together with the group of Prof. Itzen (UKE) which in parallel works on the trapping of Rab1b and DrrA-ATase using an *in vitro* chemical crosslinking approach with a non-hydrolysable ATP analog bearing a reactive chloroacetamide moiety (TReND1).

4.1.3. Results & Discussion

4.1.3.1. Establishing a co- & triple-expression system for GCE

The beginning of the projects encompassed the development of an expression system which provided the simultaneous production of Rab1b, DrrA-ATase, the synthetase and its cognate tRNA within one *E. coli* cell to realize our planned *in vivo* proximity-enhanced protein crosslinking (PEPC) approach. The synthetase mutant BrCnKRS (TEMPOH-I) which accepts the uAAs BrCnK (BrC6K, BrC7K) was already established in Cigler & Müller *et al.*[172]. A Rab1b variant with a reduced hydrolytic activity was chosen (Rab1b-Q67A)[235] that accumulates Rab1b-Q67A-GTP. It was assumed that Rab1b in its GTP-bound state is the target for DrrA AMPylation. DrrA-ATase constructs consisited of the single ATase domain ($DrrA_{1-339}$ or $DrrA_{16-352}$) or the dual domain version of ATase domain together with a catalytically inactive GEF domain ($DrrA_{1-647}$ or $DrrA_{16-533}$ both containing N451A, R453A, D480A, S483A

mutations which prevent the catalytic activity of the GEF domain[236]). Our experiments started with DrrA versions 1-339 or 1-647 but during the ongoing project the constructs were modified to more stable versions $DrrA_{16\text{-}352}$ and $DrrA_{16\text{-}533}$.

The initial plasmid system was based on pNHD-Duet-T7/pKWI, also published by Cigler & Müller *et al.*[172], where the Rab1b and $DrrA_{340\text{-}533}$ were substituted by the respective Rab1b-Q67A and DrrA-variants. Unfortunately, no robust expression of the wild type proteins in *E. coli* BL21 (DE3) cells could be established (data not shown).

A new system had to be created which was based on araBAD- instead of T7-driven expression due to the good experiences with amber suppression in *E. coli* NEB 10-beta cells. Finally, two system versions were cloned: one using a co-transformation approach, the other based on a triple-transformation with compatible plasmids concerning oris and selection markers. The dual-plasmid system consisted of a pBAD-Duet plasmid bearing Rab1b and DrrA both under control of separate araBAD-promotors and a pEVOL plasmid which encodes *Mm*-TEMPOH-I and *Mm*-PylT (PylT). The pEVOL plasmid was first described for amber suppression using the *Mj*-aaRS/tRNA pair.[237] It features two copies of the *Mm*-TEMPOH-I, one constitutively expressed and one regulated by an araBAD-promotor leading to a high rate of *Mm*-TEMPOH-I for amber suppression during the protein expression. The triple component system required a pBAD plasmid for Rab1b expression, a pBAD-RSF1031 plasmid providing the DrrA gene and the pEVOL plasmids harboring the information for *Mm*-TEMPOH-I and PylT. By coding Rab1b and DrrA on separate vectors, a flexible combination of different mutants was achieved. A detailed description of plasmid backbones and cloning can be found in the appendix 7.1.1 and Table 64. The Rab1b-Q67A gene is fused to a C-terminal His6-tag. DrrA constructs were cloned in different versions. $DrrA_{1\text{-}339}$, $DrrA_{16\text{-}352}$ and $DrrA_{1\text{-}647}$ are either tag-less ($DrrA_{1\text{-}339}$, $DrrA_{16\text{-}352}$) or N-terminally linked to a StrepII-tag bearing either an enterokinase ($DrrA_{1\text{-}339}$, $DrrA_{1\text{-}647}$) or a TEV cleavage site ($DrrA_{16\text{-}352}$) in-between. The StrepII-tag on DrrA and the His6-tag on Rab1b enabled independent detection and purification of both proteins. $DrrA_{16\text{-}533}$ constructs have a C-terminal TEV site followed by a His6-tag. The new plasmids were tested for the expression of wild type Rab1b-Q67A-His6 and Strep-$DrrA_{1\text{-}339}$ in *E. coli* NEB 10-beta and yielded good expression bands in the SDS gels for all tested combinations (Figure 36).

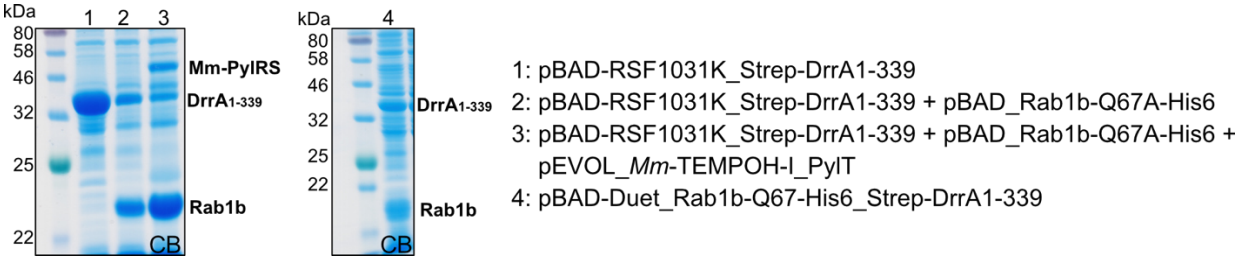


Figure 36: Establishment of the functionality of the new co- and triple-plasmid systems for the Rab1b-DrrA-ATase project. SDS gels of *E. coli* cell lysates obtained from test expressions with the new plasmids for the crosslinking approach of Rab1b and DrrA-ATase. The expression cultures were induced with arabinose. (CB = Coomassie brilliant blue)

4.1.3.2. Expanding the toolbox for chemical crosslinking

In parallel to the cloning efforts our toolbox for amber suppression with BrCnK substrates was continuously optimized. Therefore, two new modified versions of BrC6K and BrC7K (Figure 37A) as well as a new tailor-made *E. coli* strain for expression with BrCnK were created. Our previous experiments had shown that BrCnK is prone to *E. coli* CobB-mediated cleavage of the

amide-bond converting them into lysine.[172] CobB is the predominant NAD-dependent protein deacetylase in *E. coli* and belongs to the sirtuin family. Its role is the removal of acetyl groups from acetylated ε-amino groups of lysines on proteins. This requires the recognition and cleavage of the amide bond by CobB.[238] The enzymatic degradation of BrCnK can be prevented by inhibiting CobB through the addition of nicotinamide (NAM) to the expression cultures due to dependence of CobB on its cofactor NAD. A more elegant way would be a modified uAA which itself is inert towards CobB but displays the same properties as BrCnK concerning reactivity and recognition by PylRS. This was accomplished by the simple replacement of the amide bond between the ε-amino group of lysine and the bromoalkyl chain by a carbamate linkage yielding N^6-((4-bromobutoxy)carbonyl)-L-lysine (BrCO6K) and N^6-(((5-bromopentyl)oxy)carbonyl)-L-lysine (BrCO7K) (Figure 37A). Due to the alteration in the linkage, CobB should no longer be able to recognize BrCO6K or BrCO7K. The establishment of chemical synthesis routes for these uAAs was done by Anton Murnauer during this master's thesis.[196]

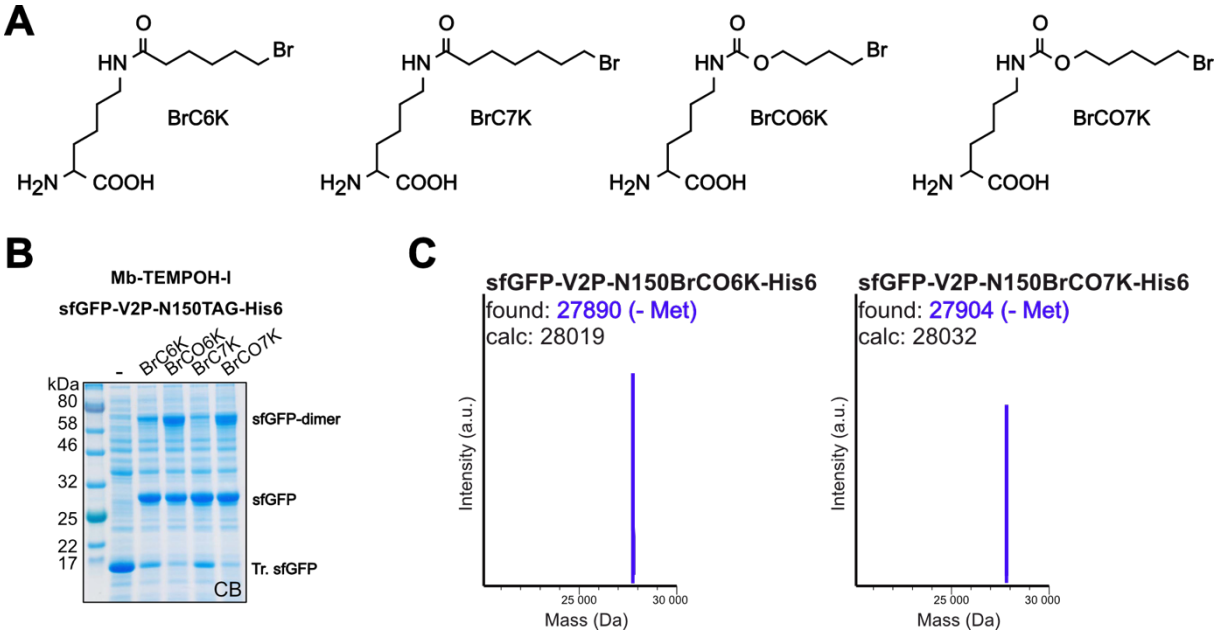


Figure 37: Establishment of two new uAAs for PEPC. A: Structures of the established uAAs, BrC6K and BrC7K, and the new uAAs, BrCO6K and BrCO7K, designed for PEPC in our lab. B: SDS gel of the *E. coli* cell lysates obtained from test expression of sfGFP-N150TAG-His6 using *Mb*-TEMPOH-I (BrCnKRS) and the four different uAA from A. C: ESI-MS data (average masses) of purified sfGFP-N150BrCO6K-His6 and sfGFP-N150BrCO7K-His6 from the test expression in B. (CB = Coomassie brilliant blue, Met = methionine, Tr. = truncated)

Both new uAAs were substrate of *Mb*-TEMPOH-I and efficiently incorporated into sfGFP-V2P-N150TAG-His6 leading to protein bands for the sfGFP-monomer and the crosslinked sfGFP-dimer in *E. coli* expressions (Figure 37B). The sfGFP-dimer formation linked via BrCO6K or BrCO7K were more efficient than for their BrC6K and BrC7K counterparts, confirming thereby the functionality for PEPC. Additionally, the ESI-MS data of purified sfGFP-V2P-N150BrCO6K-His6 and sfGFP-V2P-N150BrCO7K-His6 showed the correct masses for each of the incorporated uAAs (Figure 37C and appendix 7.4.1 Figure 78).

The incorporation of BrCO6K and BrCO7K and their stability was further tested in expressions using HEK292T cells. Therefore, the reporter mCherry-TAG-GFP was used and obtained fluorescence images are shown in Figure 38A. All transfections followed production of mCherry in comparable amounts but sfGFP expression could just be obtained in cultures

supplemented with the respective uAA for the transfected *Mm*-PylRS variant (*Mm*-PylRS for BocK; *Mm*-TEMPOH-I for BrCO6K and BrCO7K). Incorporation of BrCO6K led to equivalent levels of GFP fluorescence as observed for BocK. BrCO7K produced slightly less green fluorescent HEK cells. Amber suppression of sfGFP-N150TAG-His6 using BrCO6K and BrCO7K in HEK cells also yielded green fluorescent cells and further ESI-MS of purified sfGFP-N150BrCO6K-His6 and sfGFP-N150BrCO6K-His6 confirmed their correct incorporation (Figure 38B). The masses corresponded to sfGFP species where the N-terminal methionine was cleaved off and a N-terminal acetylation was added, two common modifications in mammalian cells. Taken together, these experiments showed that BrCO6K and BrCO7K are well suited for the application in GCE and PEPC approaches in both bacteria and mammalian cells and provide similar or even better results compared to their counterparts BrC6K and BrC7K.

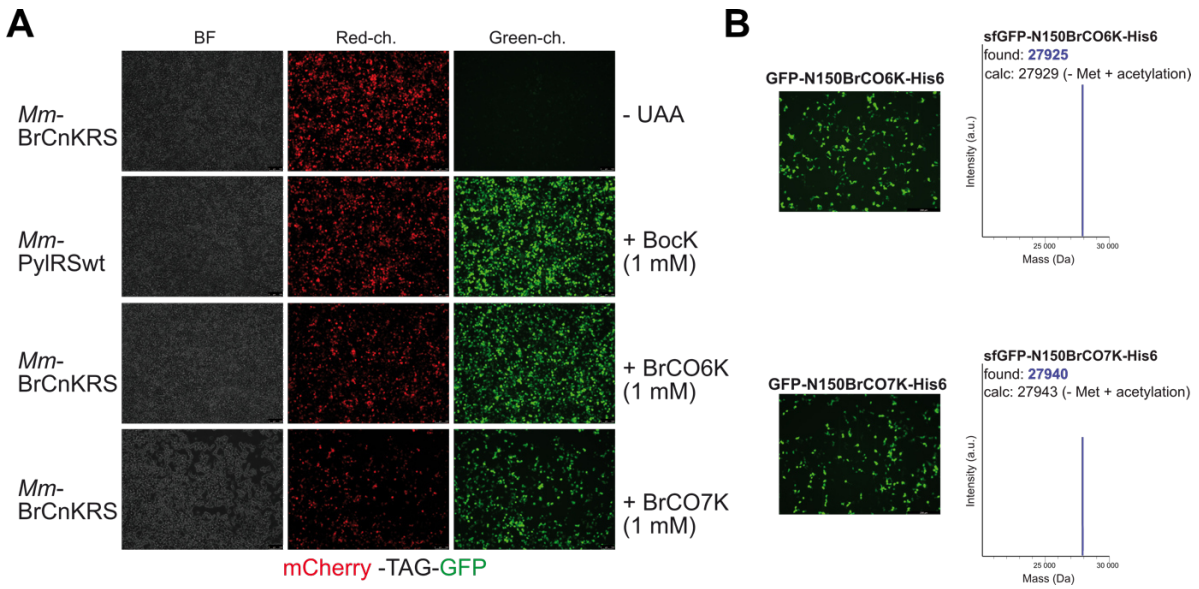


Figure 38: Fluorescence microscopy of amber suppression experiments in HEK293T cells using *Mm*-PylRS variants and BocK, BrCO6K or BrCO7K 36 hours post-transfection. A: Fluorescence images of HEK co-transfected with mCherry-TAG-GFP-plasmid and either *Mm*-PylRSwt or *Mm*-BrCnKRS (*Mm*-TEMPOH-I) growing in medium with or without respective uAA. B: Fluorescence images of HEK cells co-transfected with sfGFP-N150TAG-His6-plasmid and *Mm*-BrCnKRS growing in medium with BrCO6K or BrCO7K. ESI-MS data (average masses) of purified sfGFP-N150BrCO6K-His6 or sfGFP-N150BrCO6K-His6 is shown next to the corresponding images. (BF = bright field, ch. = channel)

Another approach that we pursued in order to prevent BrCnK cleavage was the creation of an *E. coli* expression host which lacks the enzyme CobB and was started before the new BrCO6K/BrCO7K were available. An *E. coli* NEB 10-beta ΔcobB knock-out strain was derived by replacing *cobB* through a kanamycin-resistance (Kan^R) cassette via RedET-mediated homologous recombination.[239-242] All components were provided by the Quick & Easy Gene deletion kit purchased from Gene Bridges GmbH. Figure 39, 5.2.2.9 and appendix 7.4.1 Figure 79 illustrate in detail the workflow, mechanism and components of RedET-mediated homologous recombination for chromosomal gene substitutions in *E. coli*.

With the new strain in hand, the incorporation of BrC6K and BrC7K into proteins was performed, followed by ESI-MS analysis of purified proteins in order to monitor the uAA integrity. Rab1b-T72TAG-His6 and Rab1b-Y37TAG-His6 are known for offering CobB accessible positions of the uAA[172] and were chosen as model proteins. The expression of the

proteins took place in three different set-ups for both tested uAAs. One using *E. coli* NEB 10-beta, one using *E. coli* NEB 10-beta supplemented with NAM and one using the new *E. coli* NEB 10-beta ΔcobB strain. The chemical co-transformation of plasmids into *E. coli* NEB 10-beta ΔcobB was achieved in the same way as for *E. coli* NEB 10-beta without noticeable changes concerning transformation efficiency. The growth behavior of the expression cultures monitored by OD_{600} showed no differences between the strains or according to the addition of NAM (data not shown). Thus, the deletion of *cobB* in *E. coli* NEB 10-beta did not affect fitness of the cells in our experiments. Purified amber suppressed Rab1b proteins were obtained for all expressions in equivalent yields (Figure 40A and appendix 7.4.1 Figure 80A). We monitored the integrity of the proteins via ESI-MS and observed cleavage of BrCnK in expressions using *E. coli* NEB 10-beta without NAM indicated by a mass peak which corresponded to the incorporation of lysine instead of BrC6K or BrC7K (Figure 40B and appendix 7.4.1 Figure 80B-D). This was prevented in expressions using NAM or *E. coli* NEB 10-beta ΔcobB. In summary, the new *E. coli* NEB 10-beta ΔcobB represents an ideal alternative to NAM addition in order to prevent enzymatic degradation of BrC6K and BrC7K on proteins.

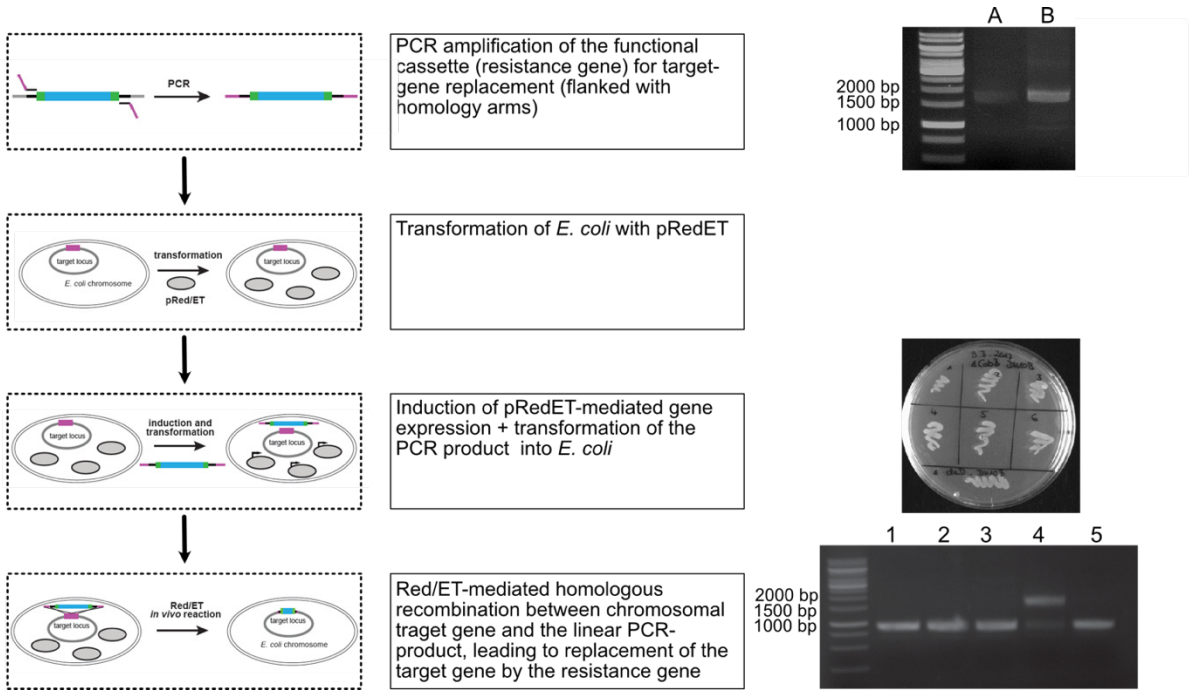


Figure 39: Workflow for the creation of an *E. coli* knock-out (KO) strain by Red/ET-mediated homologous recombination.[239-242] The four main working steps are illustrated schematically on the left together with a short explanation in the middle. On the upper right is shown the agarose gel of the PCR-amplified functional Kan^R-cassette (1750 bp) for KO of *cobB* in *E. coli* strains (A = PCR product for KO in *E. coli* NEB 10-beta; B = PCR product for *E. coli* BL21 (DE3). Below, colonies of *E. coli* NEB 10-beta obtained from homologous recombination and cultured on a LB agar plate containing kanamycin. Bottom, agarose gel showing the products of the colony-PCR using *E. coli* NEB 10-beta colonies (1-5) obtained after homologous recombination as templates. A DNA-band at 1750 bp indicates successful replacement of *cobB* with the Kan^R-cassette (clone 4). Colonies with unsuccessful homologous recombination events produce DNA-bands at 900 bp. *cobB* = 840 bp; Kan^R = 1637 bp. (Kan^R – kanamycin resistance, bp = base pair)

In a second expression test using the same set-up and Rab1b-Q67A-T72TAG-His6, the newly designed uAAs, BrCO6K and BrCO7K, were investigated towards CobB cleavage exemplary on BrCO6K. The expressions and Ni-purifications yielded similar Rab1b-Q67A-T72BrCO6K-His6 amounts for all cultures (Figure 40C). As expected, the ESI-MS data of purified proteins samples detected only correct masses corresponding to the incorporation of BrCO6K into

Rab1b-Q67A-His6 at position 72 for all expressions indeed for the sample using *E. coli* NEB 10-beta without NAM (Figure 40D). Masses which would correspond to incorporated lysine could not be observed. This verified the assumption that the carbamate-linkage in BrCO6K and BrCO7K mediates resistance to CobB cleavage in *E. coli*. Moreover, these results highlighted the advantages of BrCO6K and BrCO7K compared to its amide bond bearing counterparts BrC6K and BrC7K. By implementing these two beneficial changes, the CobB resistant uAAs and the *E. coli* NEB 10-beta ΔcobB strain, we could extend and improve our GCE toolbox for PEPC approaches.

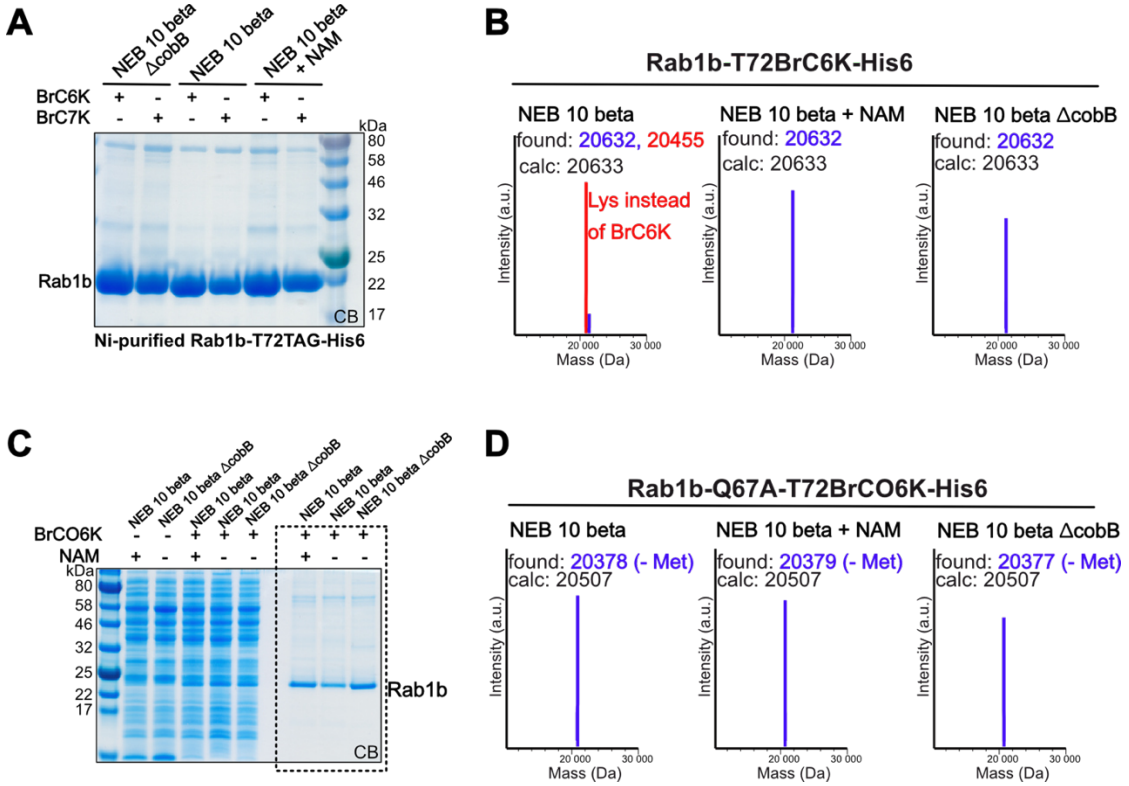


Figure 40: Amber suppression of Rab1b-T72TAG-His6 variants with uAAs for PEPC using different expression conditions and *E. coli* strains. A: SDS gel of Ni-purified Rab1b-T72BrC6K-His6 and Rab1b-T72BrC7K-His6 obtained from expression with either *E. coli* NEB 10-beta (+/- NAM) or *E. coli* NEB 10-beta ΔcobB. B: ESI-MS data (average masses) of purified Rab1b-T72BrC6K-His6 obtained from expression with either *E. coli* NEB 10-beta (+/- NAM) or *E. coli* NEB 10-beta ΔcobB. C: SDS gel of cell lysates and Ni-purified (black dashed rectangle) Rab1b-Q67A-T72BrCO6K-His6 obtained from expression with either *E. coli* NEB 10-beta (+/- NAM) or *E. coli* NEB 10-beta ΔcobB. D: ESI-MS data (average masses) of purified Rab1b-Q76A-T72BrC6K-His6 obtained from expression with either *E. coli* NEB 10-beta (+/- NAM) or *E. coli* NEB 10-beta ΔcobB. (CB = Coomassie brilliant blue, NAM = nicotinamide)

4.1.3.3. Identification of crosslink positions in Rab1b and DrrA

Next step for the Rab1b-DrrA-ATase project was the identification of Rab1b sites which are in proximity to DrrA-ATase. Therefore, an *in vivo* screening using the photo-crosslinking uAA DiaziK was performed in which we tested all existing Rab1b-TAG-His6 mutants in our lab originally created for the Rab1b-DrrA-GEF project. Upon irradiation with UV light at 365 nm, the photoreactive diazirine group of DiaziK forms highly reactive carbene species which undergo unselective reactions with close by molecules.[243, 244] It should be noted that, uAA containing diazirine moieties were already used for crosslinking proteins before.[245]

DiaziK was incorporated into Rab1b-TAG-His6 mutants during simultaneous expression with $DrrA_{1\text{-}339}$. Cell lysates with or without UV light treatment at 365 nm were analyzed via western

blot (α-His) shown in Figure 41A-B and appendix 7.4.1 Figure 81. Successful amber suppression of Rab1b-TAG mutants was observed for all expression cultures supplemented with DiaziK. No Rab1b production was obtained in samples lacking DiaziK. Samples expressing Rab1b-R69DiaziK-His6 and Rab1b-T72DiaziK-His6 showed an additional band at approximately 60 kDa only upon irradiation which matches well the mass of the crosslink between Rab1b-DrrA$_{1-339}$ (approx. 20 kDa) and DrrA$_{1-339}$ (approx. 40 kDa) (Figure 41A). Expressions without DrrA$_{1-339}$ were performed to confirm that the band corresponds to a crosslink between Rab1b and DrrA$_{1-339}$. No crosslink formation was observed in samples lacking the DrrA component (Figure 41B). The crosslink was also obtained for samples using Rab1b-R69DiaziK-His6 and DrrA$_{1-647}$ upon UV treatment (Figure 41C).

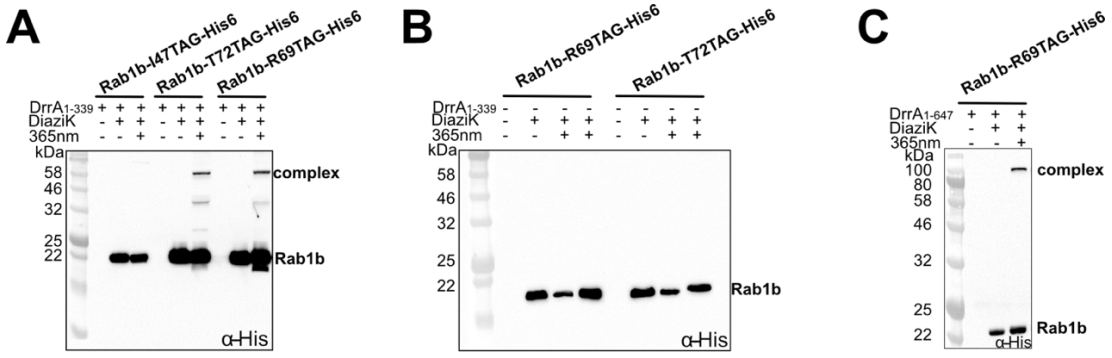


Figure 41: Identification of suitable positions in Rab1b for crosslink formation with DrrA-ATase using a screen with DiaziK. A: Western blot (α-His) of cell lysates from test expressions with triple-transformed (Rab1b-TAG-His6 variant, DrrA$_{1-339}$, *Mb*-TEMPOH-I) *E. coli* NEB 10-beta using DiaziK. Samples were irradiated with UV light (365 nm) for 10 min to induce crosslink formation. B: Western blot (α-His) of cell lysates from test expressions with co-transformed (Rab1b-TAG-His6 variant, *Mb*-TEMPOH-I) *E. coli* NEB 10-beta using DiaziK. Samples were irradiated with UV light (365 nm) for 10 min to induce crosslink formation. C: Western blot (α-His) of cell lysates from test expressions with triple-transformed (Rab1b-R69TAG-His6, DrrA$_{1-647}$, *Mb*-TEMPOH-I) *E. coli* NEB 10-beta using DiaziK. Samples were irradiated with UV light (365 nm) for 10 min to induce crosslink formation. (365 nm = UV light at 365 nm for 10 min).

Consequently, positions 69 and 72 in Rab1b must be located at the interaction interface between Rab1b and DrrA$_{1-339}$ proteins and were used for BrC6K incorporation in Rab1b-Q67A-His6. Previous experiments showed that BrC6K is able to react with different nucleophilic amino acids such as cysteine or glutamate in proximity.[172] Trusting that there might be nucleophilic residues within the Rab1b-DrrA-ATase interface, BrCnK uAAs were used for amber suppression of Rab1b-Q67A-R69TAG-His6 co-expressed with DrrA$_{1-339}$ and gratifyingly crosslink formation was observed in western blot (α-His) analysis (Figure 43). To identify the corresponding site in DrrA$_{1-339}$ contributing to crosslink formation, MS/MS experiments were performed. Therefore, the Rab1b-Q67A-R69BrC6K-His6-DrrA$_{1-339}$ complex (approx. 60 kDa) was successfully purified via Ni-affinity chromatography and SEC in yields between 0.2 – 0.5 mg per liter expression (Figure 42A-B). Ni-purification served for the removal of unreacted DrrA$_{1-339}$. The Ni-elution containing unreacted Rab1b-Q67A-R69BrC6K-His6 (P2) and the complex (P1) was further separated by SEC shown as two peaks (P1 and P2) in the chromatogram (Figure 42B).

The ESI-MS data of the protein eluates detected three species of Rab1b-Q67A-R69BrC6K-His6 corresponding to a mono-, di- and tri-AMPylated protein (Figure 42C). This was also confirmed by MS/MS which identified Y5 and Y37 in Rab1b as additional AMPylation sites beside Y77. The poly-AMPylation of Rab1b might be an artefact provoked by the overexpression of DrrA-ATase.

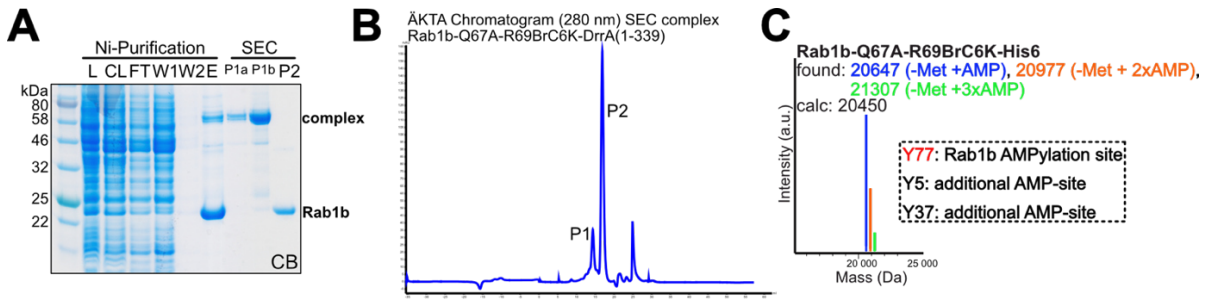


Figure 42: Purification of the Rab1b-Q67A-R69BrC6K-His6-DrrA$_{1-339}$ complex. A: SDS gel of the Ni-purification and SEC fractions of Rab1b-Q67A-R69BrC6K-His6-DrrA$_{1-339}$ complex. B: Chromatogram (280 nm) of the Rab1b-Q67A-R69BrC6K-His6-DrrA$_{1-339}$ complex purification containing two peaks (P1 & P2). C: ESI-MS data (average masses) for Rab1b-Q67A-R69BrC6K-His6 of the purified protein sample after Ni-purification and SEC. (CB = Coomassie brilliant blue, L = Lysate, CL = Cleared lysate, FT = Flow-through, W1/W2 = Wash 1/Wash2, E = Elution, P1 = SEC peak 1, P2 = SEC peak 2, P1a/P1b = SEC fractions a and b of peak 1, Y = tyrosine)

The MS/MS data evaluation for identifying the crosslinked peptide and thereby the crosslinking site within DrrA$_{1-339}$ was extremely laborious and exhausting but could finally identify two hits, D82 and E38, in DrrA$_{1-339}$ as presumed crosslinking sites to Rab1b-Q67A-R69BrC6K-His6. The MS/MS data of both hits are shown in appendix 7.4.1 Figure 84 and Figure 85A. By replacement of the D82 and E38 with alanine or cysteine, position D82 was verified as the correct crosslinking site in DrrA-ATase (Figure 43 and appendix 7.4.1 Figure 85B). The substitution D82C drastically increased BrC6K- and BrC7K-mediated crosslink formation whereas the mutation to alanine completely abolished crosslinking (Figure 43). Compared to that, mutating E38 did not affect the crosslink formation (appendix 7.4.1 Figure 85B). Additionally, the improvement for *in vivo* PEPC by the D82C mutation in DrrA was demonstrated for Rab1b-Q67A-R69BrC6K-His6 and DrrA$_{1-647}$-D82C (complex is approx. 95 kDa; appendix 7.4.1 Figure 87). Moreover, increased complex formation between Rab1b-Q67A-R69uAA-His6 and DrrA-ATase-D82C was confirmed for all BrCnK and BrCOnK uAAs (appendix 7.4.1 Figure 86B). Same results were achieved for expressions using Rab1b-Q67A-T72TAG-His6 and DrrA$_{1-339}$-D82C (appendix 7.4.1 Figure 86A).

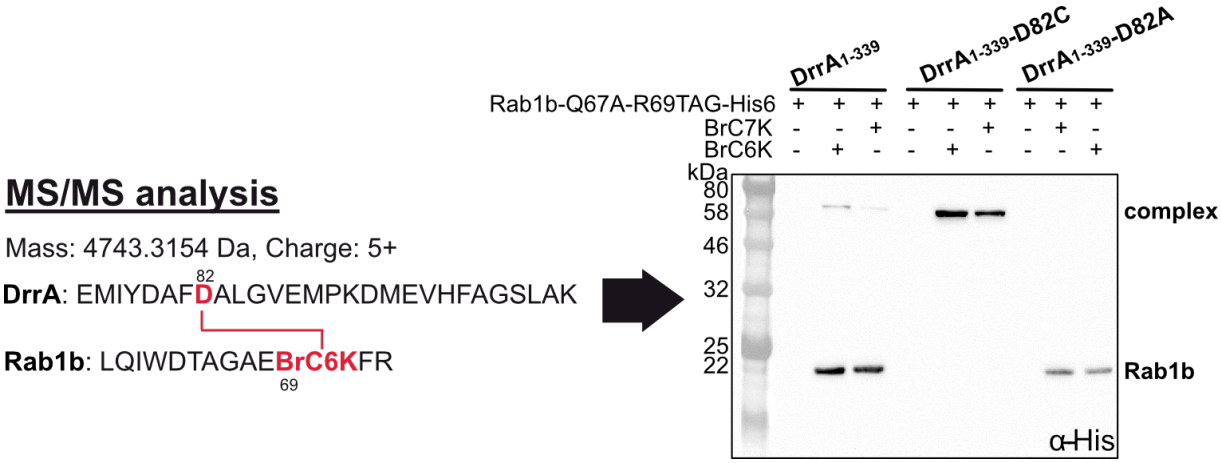


Figure 43: Identification of the respective crosslink site in DrrA$_{1-339}$ to Rab1b-Q67A-R69CrC6K-His6. MS/MS analysis of the identified crosslinked peptide Rab1b-Q67A-R69CrC6K-His6 and DrrA$_{1-339}$. The crosslink to D82 in DrrA$_{1-339}$ is schematically indicated in red. Western blot (α-His) of cell lysates from test expressions using BrC6K or BrC7K in co-transformed *E. coli* NEB 10-beta with pBAD-Duet_Rab1b-Q67A-R69TAG-His6_Strep-DrrA$_{1-339}$-D82X and pEVOL-*Mm*-TEMPOH-I-PylT. (X = alanine, cysteine or aspartate)

In parallel, the group of Prof. Itzen developed a DrrA-ATase construct with improved expression and stability properties. This DrrA-ATase version encompasses amino acids residues 16 – 352. The new $DrrA_{16-352}$ including the D82C mutation was transferred into our expression vectors and tested towards *in vivo* crosslink formation in amber suppression experiments together with Rab1b-Q67A-R69TAG-His6 and BrC6K (Figure 44A-B). As expected, the expressions using the new $DrrA_{16-352}$-D82C yielded higher protein amounts for single $DrrA_{16-352}$ and the respective Rab1b-DrrA complex compared to expression with $DrrA_{1-339}$-D82C. Therefore, from here forward we based all of our new constructs on the updated DrrA-ATase version.

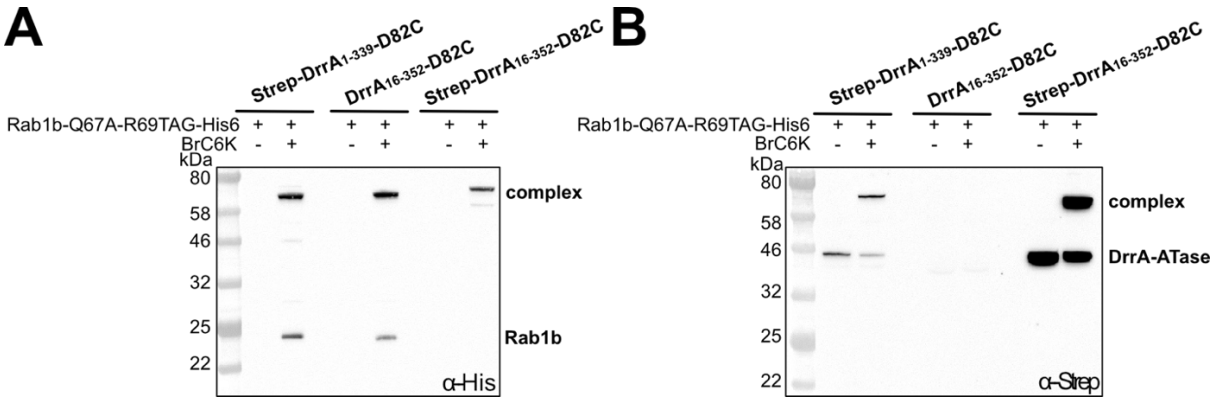


Figure 44: Establishment of different DrrA-ATase-D82C variants for crosslinking experiments with Rab1b-Q67A-R69BrC6K-His6. A: Western blot (α-His) of cell lysates from test expressions using BrC6K in co-transformed *E. coli* NEB 10-beta with pBAD-Duet_Rab1b-Q67A-R69TAG-His6_DrrA-Atase-D82C-variant and pEVOL-*Mm*-TEMPOH-I-PylT. B: Corresponding western blot (α-Strep) to the experiment from A.

Using the aforementioned two step purification approach (Figure 45), Rab1b-Q67A-R69BrC6K-His6-$DrrA_{16-352}$-D82C complex was obtained in yields of 2.8 – 3.0 mg per liter expression culture. More purifications of different versions of Rab1b-Q67A-R69TAG-His6-DrrA-ATase complexes are shown in the appendix 7.4.1 Figure 93. The D82C mutation in $DrrA_{16-352}$ improved the crosslink formation drastically so that neither the Ni-elution (Figure 45A) nor the SEC chromatogram (Figure 45C) or SDS gel (Figure 45B) detected single Rab1b-Q67A-R69BrC6K-His6, assuming that almost complete complex formation took place. Similar results were obtained for the purifications of *in vivo* crosslinked large Rab1b-Q67A-R69BrC6K-His6-$DrrA_{1-647}$-D82C complex (appendix 7.4.1 Figure 94A).

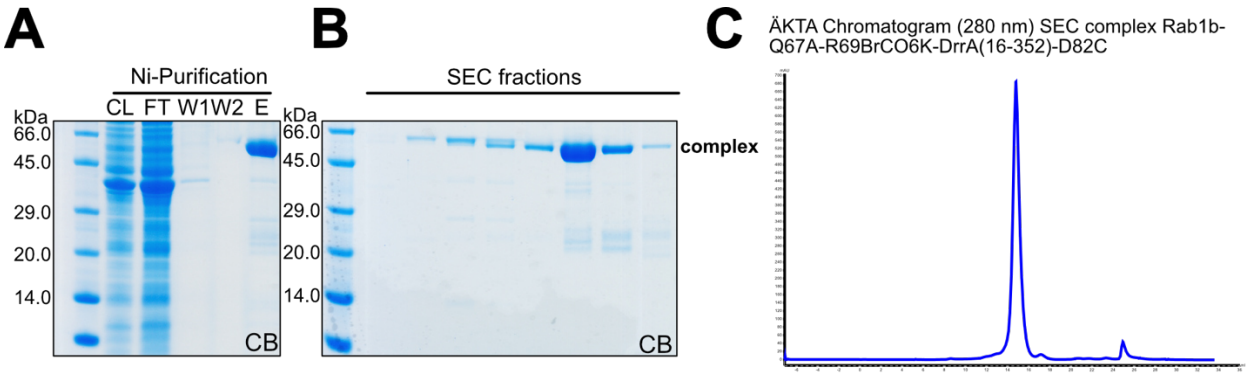


Figure 45: Purification of the Rab1b-Q67A-R69BrCO6K-His6-$DrrA_{16-352}$-D82C complex. A: SDS gel of the Ni-Purification of Rab1b-Q67A-R69BrCO6K-His6-$DrrA_{16-352}$-D82C complex. B: SDS gel of SEC fractions of Rab1b-Q67A-R69BrCO6K-His6-$DrrA_{16-352}$-D82C complex. C: ÄKTA chromatogram (280 nm) of the SEC of Rab1b-Q67A-R69BrCO6K-His6-$DrrA_{16-352}$-D82C complex. (CB = Coomassie brilliant blue, CL = Cleared lysate, FT = Flow-through, W1/W2 = Wash 1/Wash2, E = Elution)

The purified Rab1b-Q67A-R69BrC6K-His6-DrrA$_{16-352}$-D82C complex (approx. 60 kDa) was used for crystallization experiments together with Dr. Sabine Schneider. Unfortunately, none of the screened conditions led to crystal formation of the complex. Later, the crystal structure of the Rab8A-DrrA-ATase complex (further discussed in 4.1.3.4) solved by the Itzen group showed that the His6-tag of Rab1b seemed to disturb proper protein packing and prevented crystal formation. Interestingly, the position D82 in DrrA-ATase is located on the opposite to its catalytic site. This was surprising as we would have expected that the Rab-DrrA interface is located near the catalytic site of DrrA-ATase. Due to the AMPylation activity towards Rab proteins.

During the wait of the MS/MS data evaluation, many efforts were taken to experimentally localize the previously unknown crosslinking site in DrrA-ATase to Rab1b-Q67A-R69TAG-His6. Several DrrA$_{1-339}$ mutants, bearing aspartate-to-cysteine or -alanine mutations, were expressed with BrC6K and Rab1b-Q67A-R69TAG-His6 and analyzed (appendix 7.4.1 Figure 83C). Many more combinations of TAG-codon containing Rab1b and cysteine bearing DrrA-ATase were examined (appendix 7.4.1 Figure 83A-B). Additionally, big *in vivo* screens of different TAG-position in DrrA$_{16-352}$ for crosslinking to Rab1b-Q67A-His6 were performed using DiaziK and BrC6K (appendix 7.4.1 Figure 82A-B) in order to identify sites in proximity to Rab1b-Q67A-His6. None of those efforts identified crosslinking sites in the protein partners.

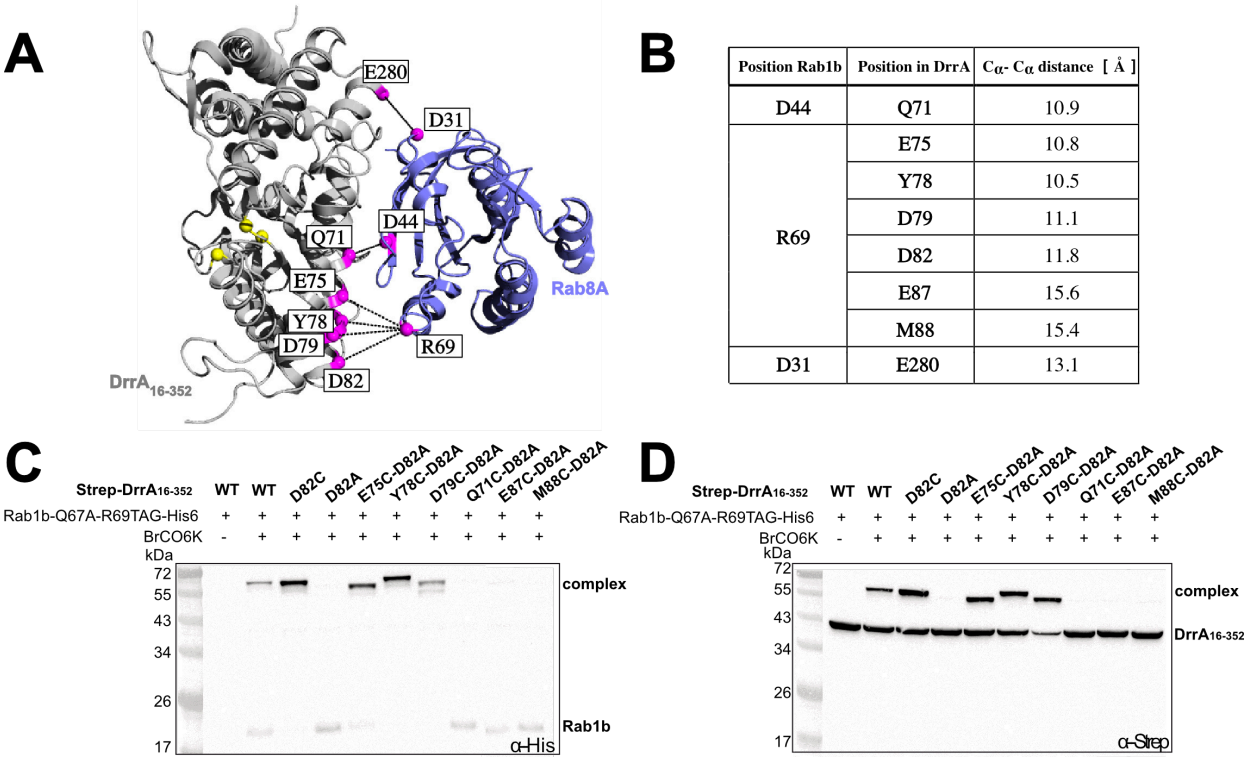
4.1.3.4. Verification of the Rab1b-DrrA interface

The group of Prof. Itzen was able to trap a Rab1b-DrrA-ATase complex via an *in vitro* crosslinking approach (unpublished data, manuscript in preparation) using a Rab1b, DrrA$_{16-352}$-L197C mutant and TReND1. TReND1 covalently bridges both proteins between residues Y77 in Rab1b and C197 in DrrA$_{16-352}$. Using this approach, they were able to solve the crystal structure of the Rab1b homolog Rab8A covalently trapped to DrrA$_{16-352}$ (unpublished data, manuscript in preparation, Figure 46A). Rab8A displays good crystallization properties compared to Rab1b and can be easily substituted for use in *in vitro* experiments. In the case of our *in vivo* GCE approach, the replacement of Rab1b by Rab8A wasn't an option due to the bad expression yields for Rab8A which would become even worse through amber suppression. Surprisingly, the solved complex reveals an unexpected interaction site of Rab8A and DrrA-ATase since Rab8A is located at the opposite site of the catalytic center (residues D110 and D112 in DrrA, yellow) of DrrA-ATase (Figure 46A).

These results raised the question whether the complex is an artifact of the chemical crosslinking with TReND1. After considering the positions R69 in Rab8A and D82 in DrrA-ATase (distance 11.8 Å) (which correspond to our identified crosslinked sites using BrCnK or BrCOnK), it became clear that their locations fitted perfectly with our crosslink data as BrC6K and its derivates are able to bridge distances between 7 – 14 Å. Thus, the fidelity of the obtained structure was confirmed by the high specificity of our approach that only provides crosslinks mediated by the proximity effect.

To further eliminate the possibility of dealing with an artificial complex the interaction interface was verified using our PEPC approach. Based on the Rab8A-DrrA$_{16-352}$ crystal structure we chose positions at the interface of Rab8A and DrrA-ATase with Cα-Cα distances between 10 – 16 Å for pairwise substitution with BrCO6K (in Rab1b) and cysteine (in DrrA$_{16-352}$). Due to their close homology, Rab8A sites could be easily transplanted to the sequence of Rab1b (Figure 46B). To circumvent undesired crosslinking to D82 in DrrA-ATase, a respective

alanine substitution was introduced for experiments using Rab1b-Q67A-R69TAG-His6. Depending on the distance between the substituted residue pairs, the crosslink formation between Rab1b-Q67A-TAG-His6 mutants and DrrA$_{16-352}$ cysteine mutants was observed. Specific crosslinks were obtained for residue pairs with distances up to 13 Å whereas no crosslinking took place for positions located > 15 Å from each other (Figure 46C-D and appendix 7.4.1 Figure 88A). The crosslink between the residue D31 (in Rab1b) and E280 (in DrrA-ATase) was confirmed in an indirect way through introducing an alanine at position E280 which inhibits the respective crosslink formation with BrCO6K at position 31 (appendix 7.4.1 Figure 88B). With this, a complete verification of the non-conventional Rab1b-DrrA-ATase interface was achieved. This newly discovered interface is speculated to be associated with an allosteric mechanism for the regulation of the DrrA-AMPylation activity towards Rab1b. The group of Prof. Itzen currently works on the elucidation of that hypothesis towards the role in Rab1b-DrrA biology.



Position Rab1b	Position in DrrA	Cα- Cα distance [Å]
D44	Q71	10.9
R69	E75	10.8
	Y78	10.5
	D79	11.1
	D82	11.8
	E87	15.6
	M88	15.4
D31	E280	13.1

Figure 46: Validation of the Rab1b-DrrA interface. A: Crystal structure of the Rab8A-DrrA$_{16-352}$ complex solved by the group of Prof. Itzen. Important residues for the catalytic activity of DrrA are marked as yellow spheres. Pink spheres present residues in the interface of Rab8A and DrrA$_{16-352}$ which might form distinct crosslinks via PEPC approach. B: Table summarizes the distances between distinct position in the interface of Rab1b and DrrA$_{16-352}$ based on measurements using the Rab8A-DrrA$_{16-352}$ complex assuming Rab1b would have the same location as its homolog Rab8A. C: Western blot (α-His) of cell lysates of amber suppression using Rab1b-Q67A-R69BrCO6K-His6 in combination with different Strep-DrrA$_{16-352}$ cysteine mutants. Combinations for pairwise BrC6K (within Rab1b) and Cys-mutations (within DrrA) were choosen based on the Rab8A-DrrA$_{16-352}$ complex. D: Corresponding western blot (α-Strep) to the experiment from C.

4.1.3.5. Characterization of the Rab1b-DrrA-ATase crosslink *in vitro*

Complementary to the *in vivo* experiments with the crosslink between Rab1b-Q67A-R69TAG-His6 and DrrA-ATase-D82C, an *in vitro* characterization was part of the project. The Ni-purifications of Rab1b constructs (approx. 20 kDa) provided pure protein in adequate yields (Rab1b-Q67A-His6: 6 – 8 mg, Rab1b-Q67A-R69BrC6K-His6 and Rab1b-R79BrC6K-His6: 2 – 4 mg per liter expression culture) (appendix 7.4.1 Figure 90). Pure DrrA$_{16-352}$ constructs

(approx. 40 kDa) were obtained via Strep-purification in yields between 2 – 4 mg per liter expression (appendix 7.4.1 Figure 91). The Strep-purification of DrrA$_{1-647}$ constructs (approx. 75 kDa) turned out to be challenging as the yields were quite low (0.3 – 1.2 mg per liter expression) and the elution fractions always contained an undefined double band (appendix 7.4.1 Figure 92A-B). Inspired by the improved stability of DrrA$_{16-352}$ modified DrrA-ATase-GEF constructs from residues 16 - 533 was tested towards its expression and purification behavior. The domain modification enhanced the yields for Ni-purified DrrA$_{16-533}$ constructs (approx. 60 kDa) up to 12 – 18 mg per liter cultures and the unrelated band became negligible (appendix 7.4.1 Figure 92C-D). The integrity of DrrA$_{16-533}$ was confirmed by *in vivo* complex formation of Rab1b-Q67A-R69BrC6K-His6 and DrrA$_{16-533}$-D82C (approx. 80 kDa) and purification (yield: 1 mg per liter expression) (appendix 7.4.1 Figure 94B).

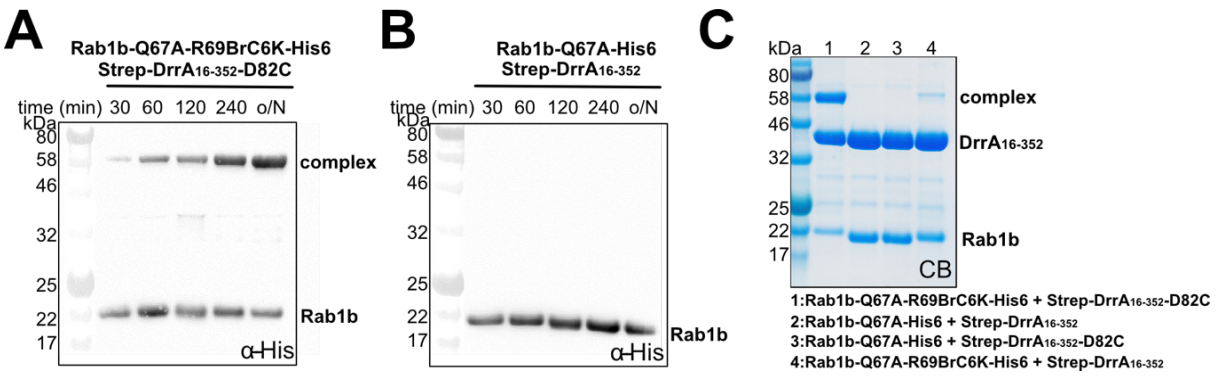


Figure 47: *In vitro* crosslinking of purified Rab1b-Q67A-His6 variants and DrrA$_{16-352}$ variants. A: Western blot (α-His) of the time dependent crosslink formation of samples containing purified Rab1b-Q67A-R69BrC6K-His6 and DrrA$_{16-352}$-D82C. B: Western blot (α-His) of the time dependent crosslink formation of samples containing purified Rab1b-Q67A-His6 and DrrA$_{16-352}$. C: SDS gel and legend of the crosslink experiment using different combinations of purified Rab1b-Q67A-His6 variants and DrrA$_{16-352}$ variants. (CB = Coomassie brilliant blue, o/N = overnight)

Time-dependent crosslink formation was performed with purified proteins using different combinations of Rab1b and DrrA$_{16-352}$ or DrrA$_{16-533}$ mutants. As expected, an efficient crosslink formation over time was only observed for samples containing Rab1b-Q67A-R69BrC6K-His6 and DrrA$_{16-352}$-D82C or DrrA$_{16-533}$-D82C respectively (Figure 47A and Figure 48A). Minimal crosslink formation to the D82 of DrrA was detected in samples with Rab1b-Q67A-R69BrC6K-His6 and DrrA$_{16-352}$ or DrrA$_{16-533}$ after incubation overnight (Figure 47B, Figure 48B and appendix 7.4.1 Figure 89). As expected, the combinations of wild type Rab1b-Q67A-His6 and DrrA variants did not produce any crosslink (Figure 47, Figure 48 appendix 7.4.1 Figure 89). This supports that both DrrA constructs (DrrA$_{16-352}$ and DrrA$_{16-533}$) have the same crosslinking properties towards Rab1b. Furthermore, highly specific crosslinking occurred exclusively between position R69BrC6K in Rab1b and a cysteine (or aspartate) at position 82 in DrrA.

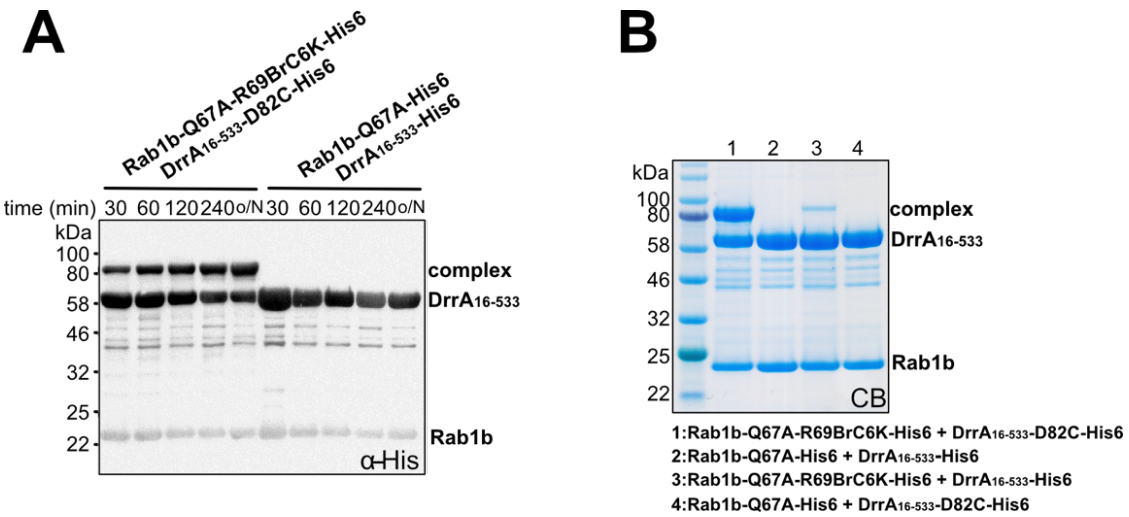


Figure 48: *In vitro* crosslinking of purified Rab1b-Q67A-His6 variants and DrrA$_{16\text{-}533}$ variants. A: Western blot (α-His) of the time-dependent crosslink formation of samples containing purified Rab1b-Q67A-R69BrC6K-His6 and DrrA$_{16\text{-}533}$-D82C or purified Rab1b-Q67A-His6 and DrrA$_{16\text{-}533}$. B: SDS gel and legend of the crosslink experiment using different combinations of purified Rab1b-Q67A-His6 variants and DrrA$_{16\text{-}533}$ variants. (CB = Coomassie brilliant blue, o/N = overnight)

Further *in vitro* experiments revealed the formation of a ternary complex consisting of two Rab1b proteins covalently attached to DrrA$_{16\text{-}533}$ (complex II, approx. 100 kDa). DrrA has multiple distinct interaction sites towards Rab1b where two of them were known to produce crosslinks in PEPC approaches. One of them was published by Cigler & Müller *et. al.* and is located within the DrrA-GEF domain. The covalent stabilization occurred via crosslink between Rab1b-R79BrC6K and D512C in DrrA-GEF.[172] The second site is our newly discovered interaction interface between Rab1b and DrrA-ATase. Samples containing Rab1b-R79BrC6K-His6, Rab1b-Q67A-R69BrC6K-His6 and double DrrA-ATase-GEF mutant (DrrA$_{16\text{-}533}$-D82C-D512C) with two respective cysteine mutation (D82C for the crosslink to Rab1b-Q67A-R69BrC6K-His6 in the ATase domain, D512C for the crosslink to Rab1b-R79BrC6K-His6 in the GEF domain) were able to produce independent dual crosslinking of Rab1b mutants to the same DrrA$_{16\text{-}533}$ at distinct sites (Figure 49).

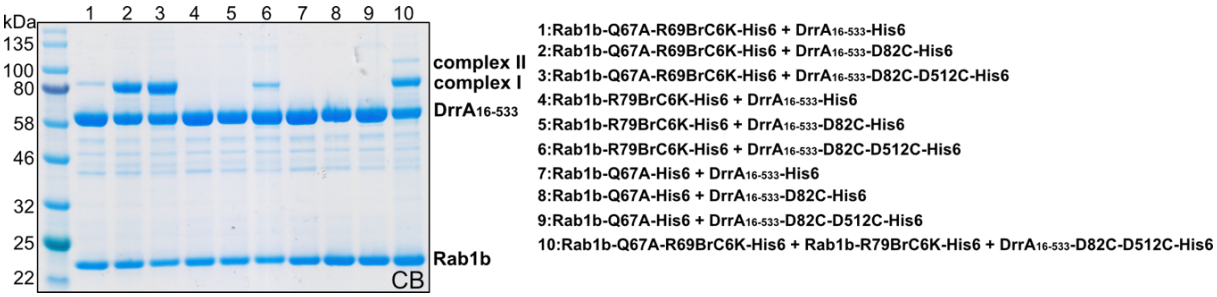


Figure 49: SDS gel and legend of the *in vitro* double crosslink experiment using different combinations of purified Rab1b-Q67A-His6 variants and DrrA$_{16\text{-}533}$ variants. Complex I = one Rab1b variant and one DrrA variant; complex II = two Rab1b variants and one DrrA variant. (CB = Coomassie brilliant blue)

Additionally, the AMPylation activity of DrrA-ATase mutants towards different Rab1b variants was investigated. The results were analyzed using western blots (α-Rab1b-AMP) and ESI-MS (Figure 50). It was shown that Rab1b-Q67A-R69BrC6K-His6 was AMPylated by DrrA$_{16\text{-}352}$ as well as by their cysteine mutant DrrA$_{16\text{-}352}$-D82C (Figure 50B+F). DrrA$_{16\text{-}352}$-D82C displayed AMPylation activity towards Rab1b-Q67A-His6 and Rab1b-Q67A-R69BrC6K-His6 (Figure 50B+F). This confirmed that incorporation of BrC6K at position of

Rab1b as well as the Q67A mutation did not interfere with the AMPylation reaction. The cysteine mutation at position D82C in DrrA$_{16-352}$ did not affect the catalytic activity of DrrA$_{16-352}$ towards Rab1b proteins either. Moreover, the AMPylation state and activity of the Rab1b-Q67A-R69BrC6K-His6-DrrA$_{16-352}$-D82C complex was examined. Western blot (α-Rab1b-AMP) of the complex confirmed that Rab1b-Q67A-R69BrC6K-His6 within the complex is in the AMPylated state (Figure 50A). Furthermore, the Rab1b-Q67A-R69BrC6K-His6-DrrA$_{16-352}$-D82C complex was able to AMPylate additional Rab1b *in vitro* (Figure 50C+E). This finding indicated that the complex formation did not interfere with the AMPylation activity of the bound DrrA$_{16-352}$ and was in conformity with the location of the non-conventional binding site at the opposite of the DrrA$_{16-352}$ catalytic site.

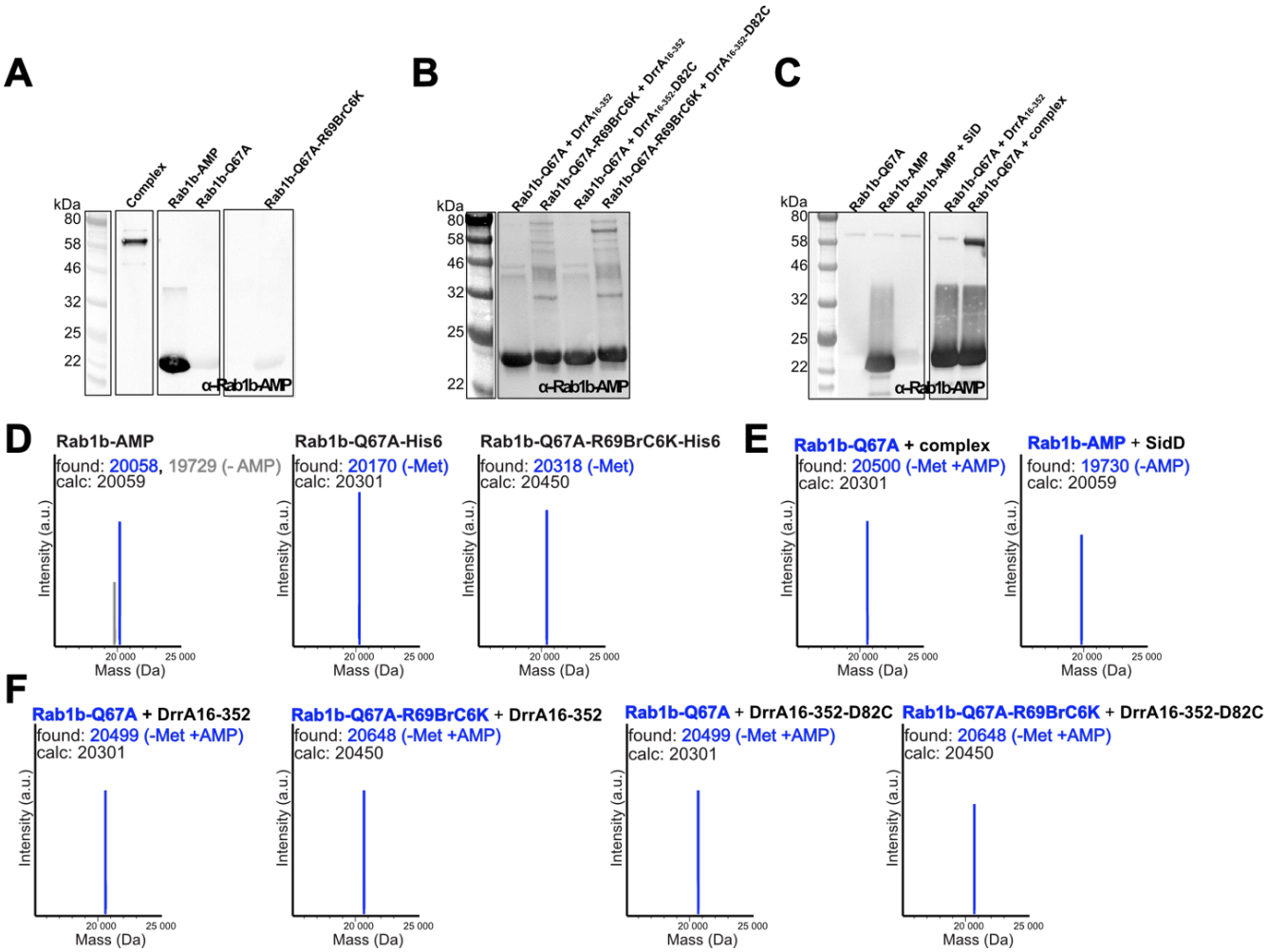


Figure 50: AMPylation assay with purified Rab1b variants, DrrA variants, Rab1b-DrrA complex and SidD *in vitro*. A: Western blot (α-Rab1b-AMP) of purified single proteins (Rab1b-AMP, Rab1b-Q67A-His6, Rab1b-Q67A-R69BrC6K-His6, Rab1b-Q67A-R69BrC6K-His6-DrrA$_{16-352}$ complex). B: Western blot (α-Rab1b-AMP) of different samples containing different combinations of Rab1b and DrrA variants. C: Western blot (α-Rab1b-AMP) of samples containing single Rab1b variants or combinations of Rab1b variants with SidD or Rab1b-Q67A-R69BrC6K-His6-DrrA$_{16-352}$ complex or DrrA$_{16-352}$. D: ESI-MS data (average masses) of single Rab1b variants obtained from the experiment in A. E: ESI-MS data (average masses) for Rab1b variants obtained from samples of the experiment in C. F: ESI-MS data (average masses) for Rab1b variants obtained from samples of the experiment in B. (complex = Rab1b-Q67A-R69BrC6K-His6-DrrA$_{16-352}$-D82C)

4.1.4. Summary & Outlook

During the work on the stabilization of the low-affinity Rab1b-DrrA-ATase complex we successfully established a robust co- (pEVOL/pBAD-Duet) and triple- (pEVOL/pBAD/pBADRSF1031K) component expression system for amber suppression experiments with the two proteins Rab1b and DrrA. Moreover, we expanded the toolbox of uAAs for proximity-enhanced crosslink by the two new candidates, BrCO6K and BrCO7K, which are inert towards CobB processing. Further, we identified a suited site (R69) in Rab1b for the incorporation of BrC6K/BrCO6K leading to complex formation even with wild type DrrA-ATase. Using a MS/MS approach we could determine the respective crosslinking position to Rab1b-R69 in DrrA-ATase (D82) and mutation to cysteine drastically improves crosslink formation efficiency. We could show that the crosslink formation also occurs in experiments using DrrA-ATase-GEF. Using *in vitro* approaches, we showed that our mutated proteins (Rab1b-R69BrC6K/BrCO6K and DrrA-ATase-D82C) show wild type properties concerning AMPylation events. Furthermore, the obtained complex shows AMPylation activity towards Rab1b. Based on the crystallographic data obtained by our collaboration partner, the found complex likely represents an unknown allosteric activation site. We were able to confirm this new unexpected interaction interface by substitution of specific residues against TAG-codons for BrC6K/BrCO6K incorporation and cysteine in their corresponding binding partner. These findings affirm the existence of this allosteric interaction site and disprove the assumption of artificial occurrence. The evaluation of the role in the biology of the Rab1b-DrrA regulation is ongoing.

During the work we could successfully find and stabilize a Rab1b-DrrA-ATase complex which revealed a formerly unknown allosteric site. Nevertheless, the originally desired complex could not be trapped yet. Therefore, new experiments should be performed to finally obtain the complex between Rab1b and DrrA-ATase, where Rab1b binds to the conventional site of DrrA-ATase for structure elucidation. This would contribute to a deeper understanding of the Rab1b-DrrA biology.

4.2. MS^2-cleavable uAAs

The idea for pursuing a class of uAAs containing MS^2-cleavable crosslinkers came up during the Rab1b-DrrA-ATase project (see 4.1) when we realized the complexity of the MS^2 data evaluation for crosslink site detection due to the low signal response of these interlinked peptides compared to total signal output in MS. During the time-consuming and often exhausting phase of MS/MS data analysis we realized a great demand for easier detection of crosslinked peptides in MS/MS data sets. This marked the birth of a new category of uAAs designed to fulfill two requirements: an efficient crosslinking of proteins via PEPC and an efficient detection of crosslinking sites between those proteins using MS analysis. The chemical work on this project was supported by Marko Cigler and Anton Murnauer, both members of our group. Parts of the following chapter were already mentioned in the dissertation of Marko Cigler.[246]

4.2.1. Background

In recent years, the application of MS-based techniques has become an increasingly powerful tool to study whole proteomes but also specific protein interactions and dynamics. Many methods developed such as hydroxyl radical footprinting MS (HRF-MS)[247, 248] and hydrogen/deuterium exchange MS (HDX-MS)[249] enable high-resolution MS footprinting of proteins in a high-throughput and sensitive fashion.[250] Moreover, crosslinking MS (XL-MS) is extensively used to examine proteins and protein-protein interfaces.[251-254] Bifunctional chemical crosslinkers covalently bridge protein probes in XL-MS in residue- but not site-specific manner.[255] The obtained data after sample digestion and liquid chromatography-electrospray ionization tandem MS (LC/ESI-MS/MS short MS/MS or MS^2) analysis is scoured for the signals corresponding to the interlinked peptides to determine interface regions and crosslinking sites.[256, 257] This process remains challenging and requires extensive bioinformatic tools and computer-assisted data evaluation due to the low-abundance of the interlinked peptide signals compared to the total signal output. This is further complicated by the aforementioned semi-specific reaction of the bifunctional crosslinker with residues within the examined proteins resulting in great heterogeneity between the crosslinked peptides.[258] XL-MS achieved the next level by the implementation of bifunctional chemical crosslinking reagents bearing MS^2-cleavable groups within their linker regions (e.g. disuccinimidyl sulfoxide (DSSO)[259] or disuccinimidyl dibutyric urea (DSBU)[260]; Figure 51A) which provide characteristic ion fragmentation patterns of the interlinked peptide during the fragmentation in subsequent MS^2 analyses.[259, 261] This immensely facilitates the data evaluation and identification of the crosslinked products. To say it simple, the signals corresponding to the crosslink-bearing peptides are identified and distinguish from other signals via their characteristic pattern.
The XL-MS workflow for the identification of crosslinked peptides and their respective crosslinking sites encompasses three major steps: peptide mass fingerprint (MS^1), fragmentation of crosslinked peptides (MS^2) and fragmentation of the peptide backbone (MS^2 or MS^3).[258, 262] The MS-induced fragmentation of these crosslinkers (DSSO and DSBU) is achieved by collision-induced dissociation (CID)[263] for e.g. C-S bond of sulfoxides[258, 259] or by high-energy collision-induced dissociation (HCD)[263, 264] for C-N bonds of urea motifs[260, 261].

The fragmentation process happens in a special collision cell harboring a chemically inert gas (N_2, He, Ar). The precursor ions obtained in MS^1 are guided into the cell via acceleration in an electric field where they collide with the energized gas molecules. Energy-mediated bond cleavage of the precursor ions is finally enabled by transferring the kinetic energy accumulated from the collision events.[263] This generates product ions (α and ß ion pairs) which are detected in MS^2 as certain ion doublets with a distinct mass difference (26u doublets).[265] This provides the information for the identification of the interlinked peptides. To further locate the crosslinking sites within the peptides, an additional analysis of the peptide backbone is necessary.[258] Depending on the kind of crosslinker used this can occur simultaneously during MS^2 or in a separate MS^3 process of the former identified product ions. Commonly, C-N bonds require higher energies for fragmentation in MS^2 (HCD) as does peptide backbone fragmentation (peptide sequencing). [260, 261] The more labile C-S bonds of sulfoxides are fragmentated using low-energy CID. Here, the peptide backbone fragmentation requires the more energetic process of a subsequent, third MS experiment (MS^3).[259]

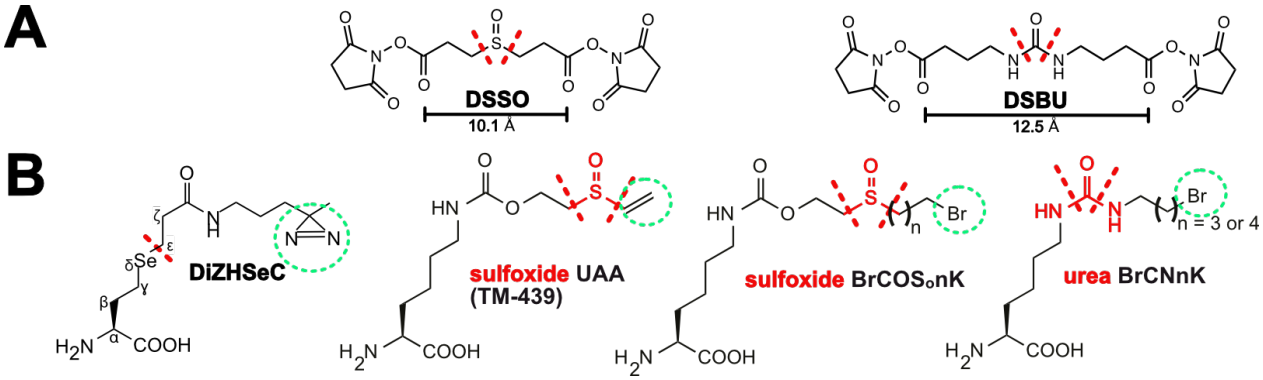


Figure 51: Molecular structures of uAAs and chemical crosslinkers for MS-approaches. Red dashed lines denote cleavable bonds in the molecule during MS^2 process. Green dashed circles represent chemical moieties for chemical or photo-crosslinking. A: Structures of the chemical crosslinkers DSSO[259] and DSBU[260] used in MS approaches for the identification of crosslinked peptides. B: Published uAA (DiZHSeC)[266], which provides a label for MS analysis, and new designed uAAs (TM-439, BrCOSonK, BrCNnK) which bear MS^2-cleavable moieties and PEPC-competent functionalities in their side chains.

An alternative but less popular method to render MS data evaluation easier is the introduction of MS signal enhancers via GCE as already published for the DiZHSeC (Figure 51B).[266] After site-specific incorporation of DiZHSeC into proteins its diazirine moiety can undergo crosslinking upon irradiation with UV light. The inducible Se_δ-C_ε bond cleavage and formation of a stable acrylamide serves as easy identifiable labels for MS analysis.

Nevertheless, heterogeneity issues within the crosslinked protein samples remain problematic for both approaches as the crosslinking reactions proceed unselectively. The aforementioned chemical crosslinkers DSSO and DSBU use *N*-hydroxysuccinimide (NHS) esters which react primarily with the ε-amino group of lysine.[267] While the modification is residue specific, it occurs protein-wide ending up in a mixture of inter- and intra-linked proteins as well as unlinked species (dead-end).[259] Photo-crosslinking with the diazirine moiety[243] of DiZHSeC also resultes in an non-homogeneously crosslinked sample.

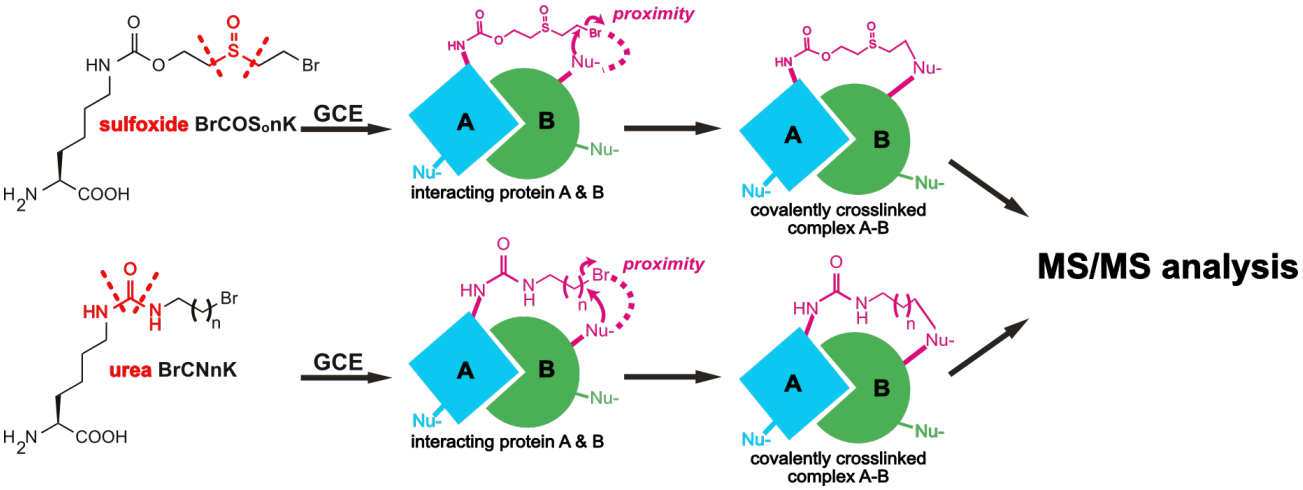


Figure 52: PEPC using uAAs which bear MS^2-cleavable moieties within their side chains. An electrophilic uAA containing a MS^2-cleavable moieties (red) is site-specifically introduced into protein A (blue) via GCE. The uAA can react with nucleophilic amino acids (Nu-, pink) triggered by proximity. In this way covalent crosslinking with an interacting protein B (green) can be achieved. The protein complex A-B can then be used for MS/MS analysis to identify the crosslinked residues. Red dashed lines denote cleavable bonds in the molecule during the MS^2 process.

To tackle down the aforementioned challenges, we wanted to combine the properties of MS^2-cleavable groups with the site-specific incorporation of uAAs to provide homogeneous crosslinked samples for an optimized MS/MS data analysis. The idea is the implementation of MS^2-cleavable groups (sulfoxide moiety or urea-motif) into the side chains of uAAs used for PEPC (Figure 51B). Inspired by the DSSO and DSBU crosslinkers, we developed a new class of uAAs harboring two distinct features: one functional group for PEPC (bromo-moiety or Michel acceptor) and one for MS^2-mediated fragmentation (Figure 51B). Their structures are mainly oriented on the BrCOnK scaffold. The uAAs can be further sub-divided in urea-based (BrCNnK) and sulfoxide-based (TM-439 or BrCOSonK) uAAs. We suggest a two-step approach consisting of GCE to incorporate the new uAA into proteins for crosslinking followed by MS analysis for the identification of crosslinking sites. The approach benefits from the formation of homogenous crosslinked protein samples via PEPC reaction (for more details see 4.1.1) (Figure 52) and their use for subsequent XL-MS (Figure 53). Proteins with unreacted uAA can be easily removed during purification processes. The homogeneity of the sample together with the MS^2-cleavable moiety provide ideal conditions for the identification of the interlinked peptides as well as the position of the crosslink.

The proposed MS analysis workflow for the GCE-based approach using either BrCNnK (Figure 53A) or BrCOSonK (Figure 53B) encompasses the same stages as aforementioned for experiments using DSSO or DSBU. After GCE and sample preparation of crosslinked proteins, MS^1 reveals fingerprinting of the obtained peptides. For samples derived from BrCNnK-incorporation, HCD-induced MS^2 enables crosslink and peptide backbone cleavage in one step whereas crosslink fragmentation results in characteristic ion doublets (26u doublets; α and ß ions). For BrCOSonK-crosslinked samples, CID only leads to crosslink fragmentation in MS^2 shown as a distinct ion pattern. A further MS^3 experiment is required for sequencing of isolated peptides from MS^2. The obtained crosslink fragmentation and peptide sequencing data serves for easy identification of interlinked peptides and corresponding sites.

In summary, our approach would produce homogenously crosslinked protein samples which provide characteristic ion fragmentation pattern in MS^2 for easy identification of interlinked peptides and crosslinking sites by combining the features of PEPC and XL-MS.

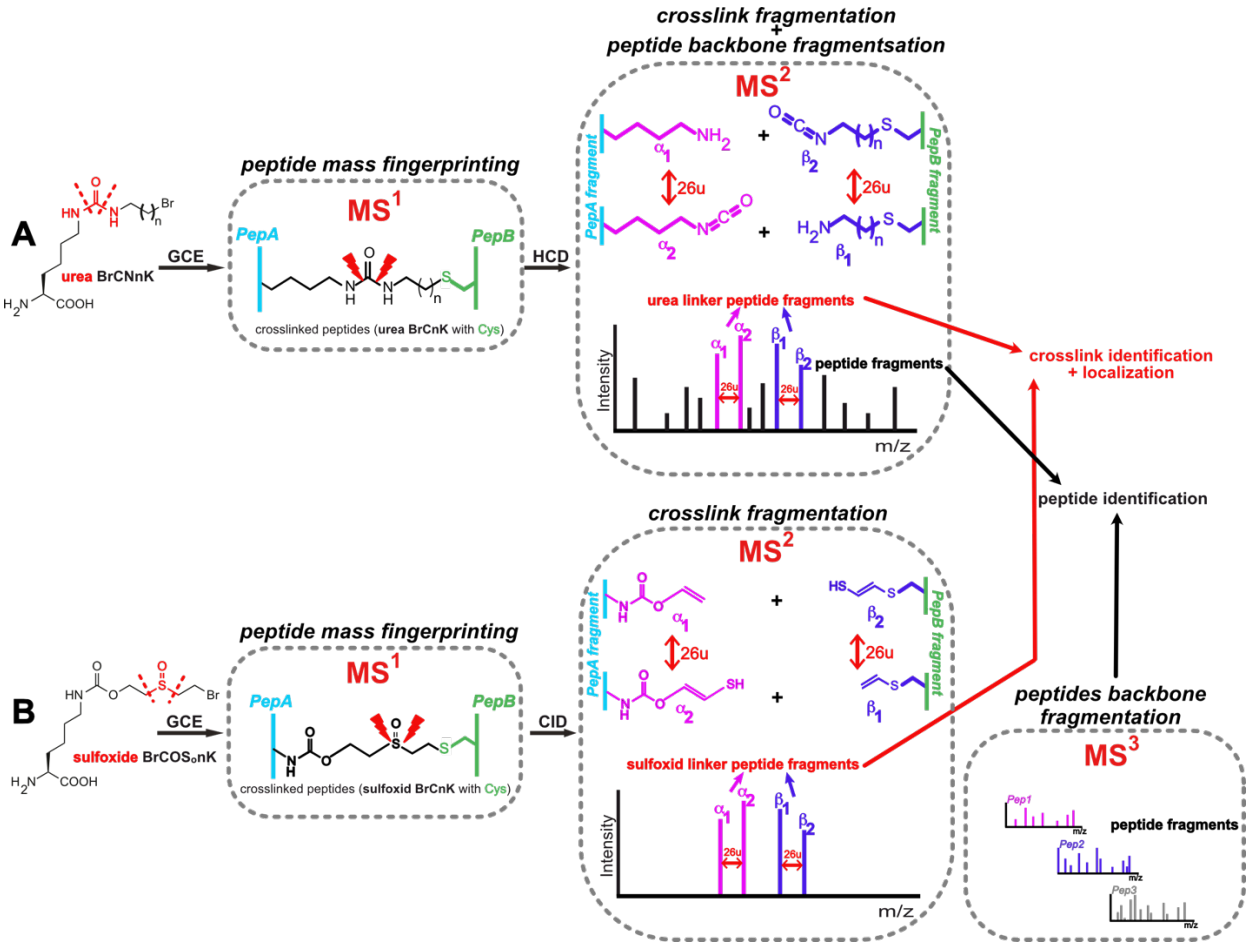


Figure 53: Proposed MS analysis workflows for crosslinked protein complexes via uAAs bearing MS/MS-cleavable moieties. A: Workflow for the approach using uAAs bearing side chains with urea motif. After incorporation of the uAA via GCE and crosslinking, the purified and digested protein sample is used in MS^1 for peptide mass fingerprinting. The HCD process in MS^2 cleaves the labile urea bond between interlinked peptides leading to a characteristic fragment ion pattern (26u doublets; blue and pink), as well as the peptide backbone (black). This enables the identification of the crosslinked peptides as well as the respective crosslinking sites. B: Workflow for the approach using uAAs decorated with sulfoxide moieties. After incorporation of the uAA via GCE, the crosslinked protein sample was purified and digested for usage in MS^1 for peptide mass fingerprinting. The CID process in MS^2 fragments the labile sulfoxide-motif characteristically between interlinked peptides leading to distinct fragment ion pattern (26u doublets; blue and pink) which allows crosslink identification. The peptide backbone is nearly unaffected. Subsequent MS^3 enables the fragmentation of the peptide backbones to identify the corresponding crosslinked peptides. (CID = collision-induced dissociation, HCD = High-energy collision-induced dissociation, MS = mass spectrometry, PepA = peptide A, PepB = peptide B, Red dashed lines denote cleavable bonds in the molecule during MS^2 process)

4.2.2. Aim

Aim of the project is the establishment of our new class of MS^2-cleavable uAAs for GCE which should facilitate MS-based crosslink identification of protein samples. On the one hand, we want to develop an amber suppression system for the efficient incorporation of the MS^2-cleavable uAAs *in vivo* and validate their applicability for PEPC approaches. On the other hand, we are interested in their cleavage properties during MS^2 analysis towards the desired fragmentation pattern for a simplified crosslink detection. The identified Rab1b-Q67A-

R69uAA-DrrA$_{16-352}$-D82C complex (see 4.1) will serve us as proof-of-principle target for *in vivo* crosslink formation, downstream processing, MS analysis and data elucidation.
The project is a collaboration with the group of Prof. Sinz (Martin-Luther University Halle-Wittenberg, Germany) who performed the MS data evaluation of protein samples measured within the group of Prof. Sieber (TUM).

4.2.3. Results & Discussion

4.2.3.1. Lysine derivatives bearing sulfoxide moieties

BrCOS8K (Figure 54A) was the first successfully synthesized uAA in the class of MS^2-cleavable uAAs by Marko Cigler. Due to synthesis issues concerning the oxidation step to a sulfoxide (discussed in the dissertation of Marko Cigler[246]) the uAA contains a thioether moiety in its bromoalkyl side chain. It was for an *in vitro* on-protein oxidation reaction using H_2O_2 to install the sulfoxide group after incorporation, a method which was already successfully applied before.[268, 269]
Two *Mb*-PylRS mutants of our lab-collection were able to recognize BrCOS8K in test expressions with sfGFP-V2P-N150TAG-His6 (Figure 54B), *Mb*-TEMPOH-I and *Mb*-SM29. *Mb*-SM29 was discarded for further experiments because the expression using *Mb*-TEMPOH-I yielded much higher sfGFP expression levels. Because *Mb*-TEMPOH-I also accepts BrCOnK variants, it was not surprising that it also recognized the structurally related BrCOS8K. Interestingly, the SDS gels could not detect a band corresponding to the crosslinked sfGFP-dimer which normally is present in expression with PEPC-related uAAs, e.g. BrCOnK. Only the purification of sfGFP-V2P-N150BrCOS8K-His6 visualized both monomer- and dimer band (Figure 54C). ESI-MS analysis of sfGFP-V2P-N150BrCOS8K-His6 showed three dominant species (Figure 54D). The biggest peak (27933 Da) corresponded to the sfGFP-V2P-N150BrCOS8K-His6 mass without the N-terminal methionine. A slightly less intense species (27850 Da) belonged to a mass without methionine and HBr which could be a product of an intramolecular crosslinking reaction or formation of an allyl thioether by elimination of the bromide. The third mass (28162 Da) represented a GSH adduct. The formation of two unreactive species (28162 Da and 27950 Da) towards PEPC explained the less efficient sfGFP-dimer. The on-protein oxidation of purified sfGFP-V2P-N150BrCOS8K-His6 did not yield the desired sulfoxide of BrCOS8K (data not shown).
BrCOS8K was not longer included in following studies because of the high amounts of by-product formation during amber suppression (GSH adduct, loss of HBr) and the failure in oxidation experiments *in vitro.*

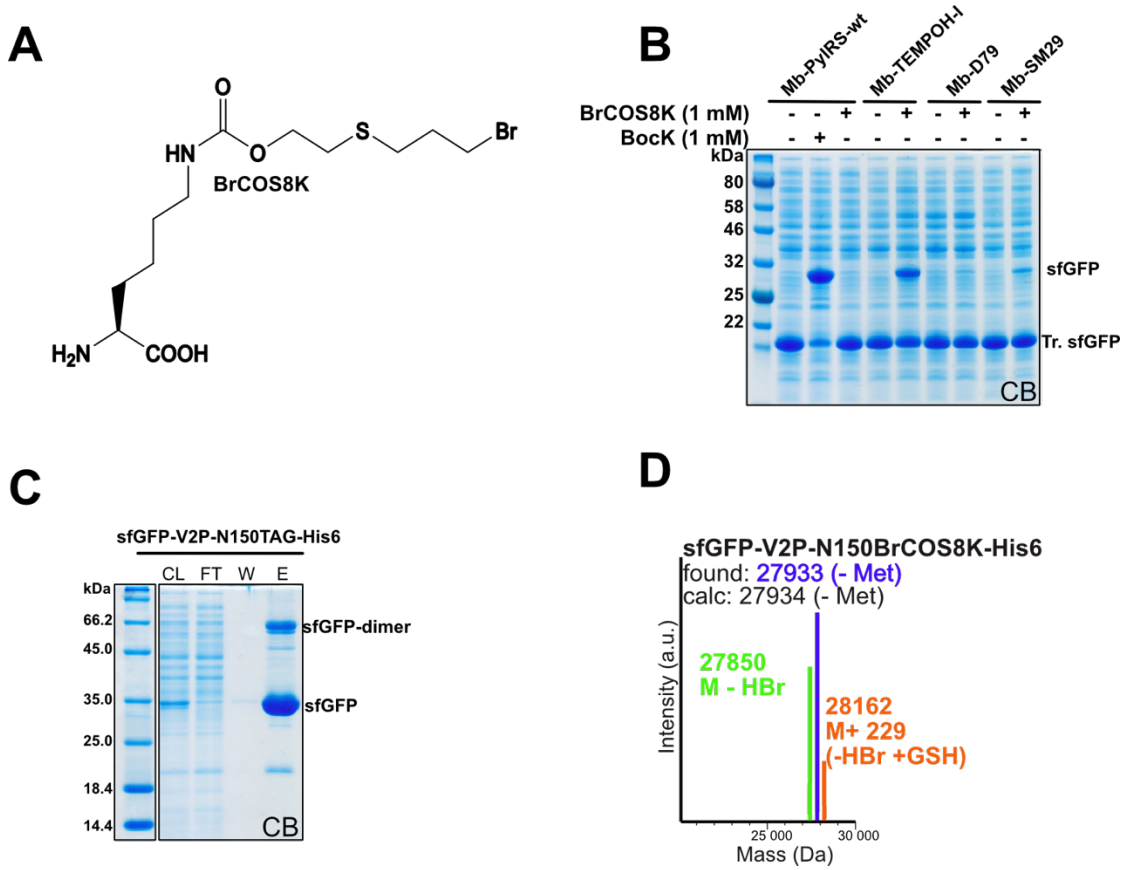


Figure 54: Amber suppression of sfGFP-V2P-N150TAG-His6 using BrCOS8K in *E. coli*. A: Structure of BrCOS8K bearing a sulfur in its side chain intended for *in vitro* oxidation to a sulfoxide on proteins. B: SDS gel of cell lysates of the incorporation test for BrCOS8K in sfGFP-V2P-N150TAG-His6 using different *Mb*-PylRS mutants. C: SDS gel of the Ni-purification of sfGFP-V2P-N150BrCOS8K-His6 obtained from expression with *Mb*-TEMPOH-I. C: ESI-MS data (average masses) obtained from the purified sfGFP-N150BrCOS8K-His6 in C. (CB = Coomassie brilliant blue, CL = Cleared lysate, FT = Flow-through, W = Wash, E = Elution, GSH = glutathione)

The second compound, TM-439, of sulfoxide-based MS^2-cleavable uAAs (Figure 55A) was synthesized by Anton Murnauer (PhD student in our group). Inspired by Shang *et al.*[268] TM-439 bears a Michael acceptor (mild electrophile) and PEPC reaction should be mediated via a Michael addition. Using test expression with sfGFP-V2P-N150TAG-His6 three PylRS mutants (*Mb*-TEMPOH-I, *Mb*-SM29, *Ma*-PylRS-H227I-Y228P) were identified to accept TM-439 and produced good amounts of sfGFP (Figure 55A). Unfortunately, the crosslinked sfGFP-dimer band was absent in the SDS gel of the test expressions (Figure 55A) as well as in the elution of the Ni-purified sfGFP-V2P-N150TM-439-His6 derived from the expression with *Mb*-TEMPOH-I (Figure 55B). Furthermore, the ESI-MS data of the purified sfGFP-V2P-N150TM-439-His6 could only detect a mass (27838 Da) corresponding to a respective uAA where the sulfoxide moiety is reduced to the corresponding thioether (Figure 55C). Without the sulfoxide no proper Michael acceptor is available for the reaction with nucleophiles and the uAA is "dead" for PEPC. It was assumed that the reduction occurred inside the *E. coli* cells as the TM-439 stock itself and a sample of TM-439 incubated in PBS at 37 °C was stable as judged by ESI-MS (292 Da) (data not shown). A class of *E. coli* enzymes called methionine sulfoxide reductase (Msr) family, MsrA, MsrB, MsrC and MsrP, which normally catalyze the reduction of oxidized methionines[270] supposed to be responsible for the reduction of the sulfoxide moiety in TM-439. Several *E. coli* knock-out strains with single or multiple deletion in *msrA, msrB,*

msrC and *msrP* (= *yedY*) were used for amber suppression with TM-439. None of those host strains was able to prevent the reduction of the uAA as the ESI-MS data of all purified sfGFP-V2P-N150TM-439-His6 confirmed the incorporation of reduced TM-439 (Figure 55D-E). These results could either indicate that this enzyme class does not affect TM-439 or that individual knock-outs were not effective in suppressing the reduction or that our uAA is not a substrate of these enzymes. The Msr class encompasses many more components than MsrA, MsrB, MsrC and MsrP e.g. enzymes which recognize free methionine (fRMsr[271] or fSMsr[272]). *E. coli* strains lacking fSMsr and fRMsr proteins might be a promising point to figure out if the reduction of free TM-439 causes the problem. Nevertheless, several other enzymes are involved in the oxidative damage repair in *E. coli* and are related to reduction of sulfoxide moieties, exemplary the DmsABC[273] complex or BisC[274], which might interfere with TM-439 integrity.

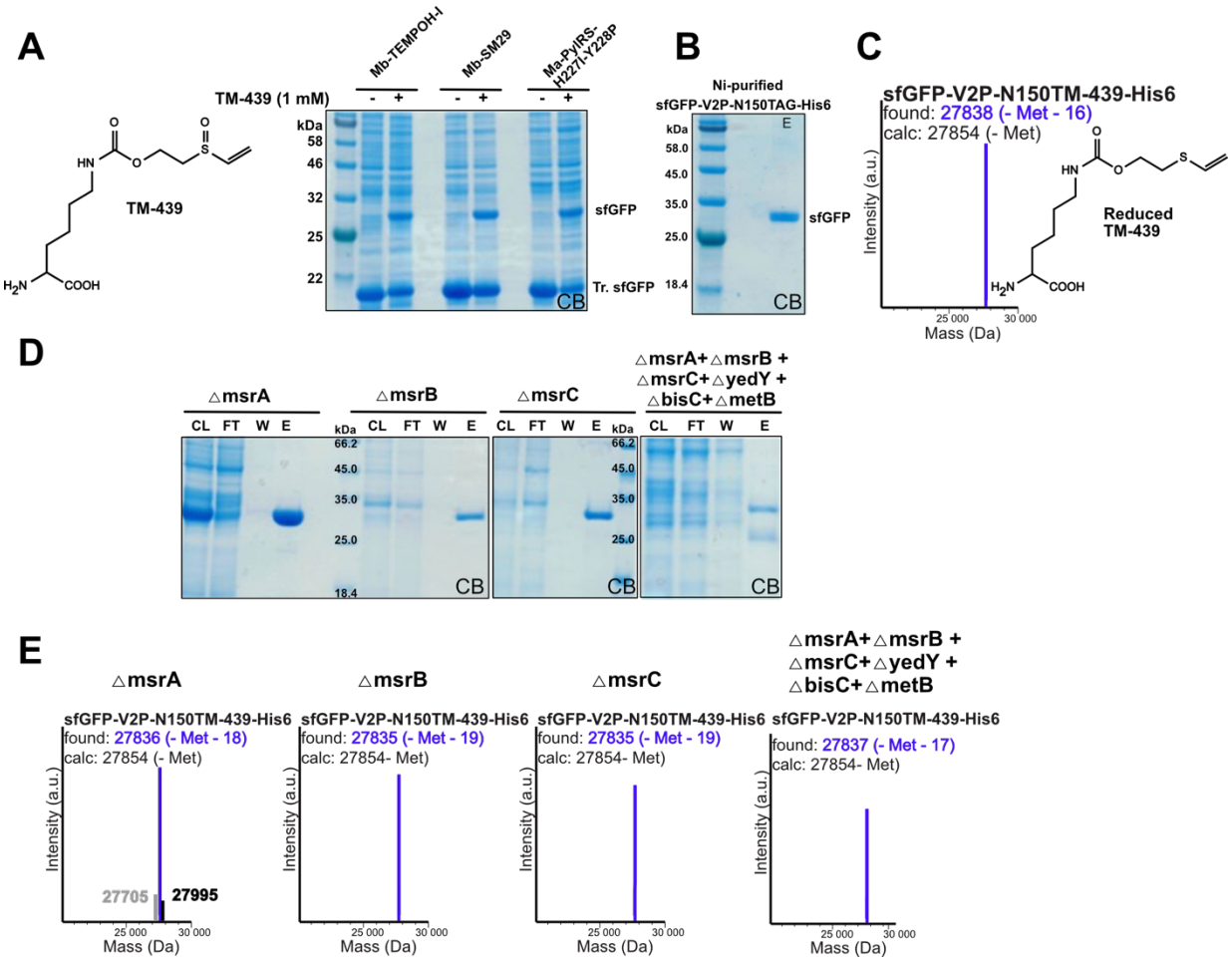


Figure 55: Amber suppression of sfGFP-V2P-150TAG-His6 using TM-439 in *E. coli*. A: Structure of TM-439 and SDS gel of cell lysates (*E. coli* NEB 10-beta) of the incorporation test for TM-439 in sfGFP-V2P-N150TAG-His6 using different *Mb*-PylRS mutants. B: SDS gel of the Ni-purification of sfGFP-V2P-N150TM-439-His6 obtained from expression with *Mb*-TEMPOH-I. C: ESI-MS data (exact masses) obtained from the purified sfGFP-V2P-N150TM-439-His6 in B and the structure of reduced TM-439. D: SDS gel of the Ni-purification of sfGFP-V2P-N150TM-439-His6 obtained from expressios with *Mb*-TEMPOH-I using different *E. coli* strains containing knock-outs of proteins from the Msr-family. E: ESI-MS data (average masses) obtained from the purified sfGFP-V2P-N150TM-439-His6 in D. (CB = Coomassie brilliant blue, CL = Cleared lysate, FT = Flow-through, W = Wash, E = Elution)

Additionally, the incorporation of TM-439 was tested in HEK293T cells which differ from *E. coli* concerning the reducing environment. TM-439 was incorporated using *Mm*-TEMPOH-I (= *Mm*-BrCnKRS) and produced dual fluorescent HEK cells using the mCherry-TAG-sfGFP

reporter (Figure 56). Purified sfGFP-N150TM439-His6 obtained from amber suppression in HEK cells transfected with plasmids bearing *Mm*-TEMPOH-I and sfGFP-N150TAG-His6 was used for ESI-MS. Unfortunately, only the mass of the reduced sfGFP-N150TM439-His6 with N-terminal cleaved methionine and an additional acetylation was found. The reduction of the sulfoxide occurred also in the mammalian cell environment as before in the *E. coli* experiments.

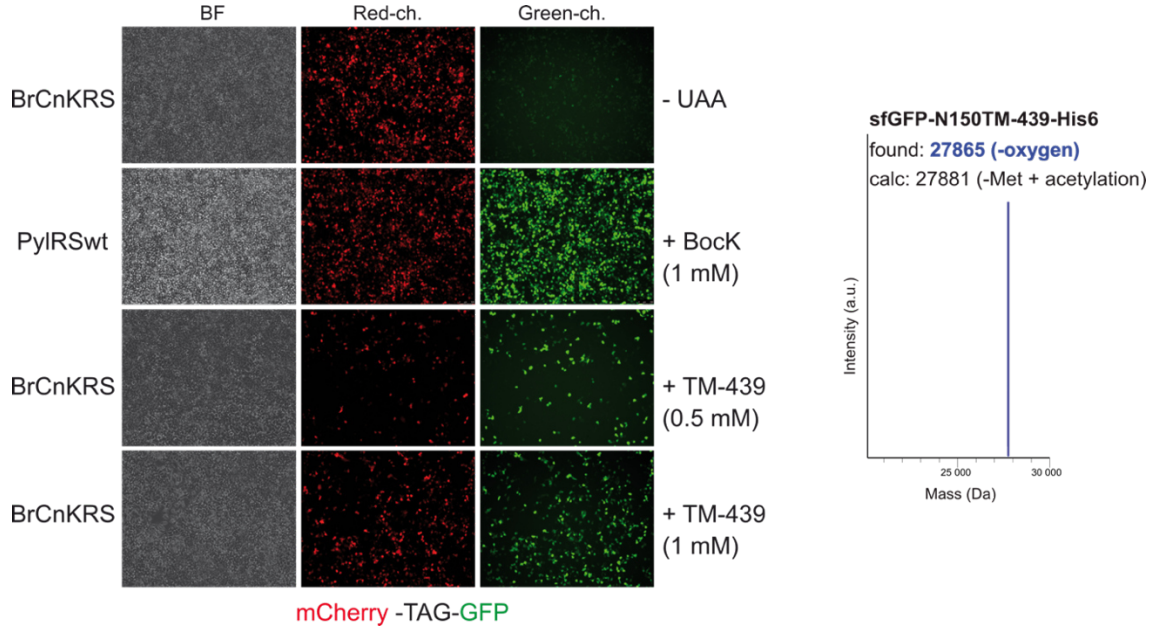


Figure 56: Fluorescence microscopy of amber suppression experiments in HEK293T cells using *Mm*-BrCnKRS (= TEMPOH-I) and TM-439 36 hours post-transfection. Fluorescence output of HEK cells co-transfected with a reporter plasmid containing mCherry-TAG-GFP and the plasmid encoding the respective *Mm*-PylRS variant. Cells were growing in medium with or without the respective uAA. ESI-MS data (exact masses) of purified sfGFP-N150TM-439-His6 obtained from the 36h-expression in HEK293T cells co-transfected with the sfGFP-N150TAG-His6-reporter plasmid and the plasmid encoding *Mm*-BrCnKRS. (BF = bright field, ch. = channel)

4.2.3.2. Lysine derivatives bearing urea motifs

Two more MS^2-cleavable uAAs for PEPC (BrCN6K and BrCN7K) were synthesized and discussed by Marko Cigler[246]. BrCN6K resembled BrC6K in its bromoalkyl chain length but tuned out to undergo intramolecular cyclization under elimination of the bromo-group after the deprotection step with HBr which was monitored by ESI-MS (data not shown) (Figure 57A). Although the cyclized uAA is no longer useful for our purposes (and was therefore not pursued any further), it did get incorporated into sfGFP-N150TAG-His6 by the wild type *Mb*-PylRS (Figure 57A).

For the second candidate BrCN7K no intramolecular reaction during synthesis was observed. The incorporation test in sfGFP-N150TAG-His6 revealed two *Mb*-PylRS mutants (*Mb*-TEMPOH-I and *Mb*-SM30) able to accept BrCN7K, producing thereby single sfGFP-N150BrCN7K-His6 and crosslinked sfGFP-dimers (Figure 57B). Also, the elution fractions of purified sfGFP-N150BrCN7K-His6 from 6 hour and o/N expression with *Mb*-TEMPOH-I contained both bands for single sfGFP and dimer (Figure 57C). Beside the correct mass of sfGFP-N150BrCN7K-His6 (28033 Da), ESI-MS analyses detected a mass related to N-terminal methionine cleavage (27899 Da) and one for a GSH adduct (28262 Da) for both samples (Figure 57D). The GSH adduct mass immensely increased in the sample after o/N sfGFP-N150BrCN7K-His6 expression. Furthermore, the correct mass of the double crosslinked sfGFP-dimer (55907 Da or 55757 Da for the N-terminal methionine cleaved species) was

observed in both samples (Figure 57D). This confirmed that BrCN7K is useful for PEPC approaches and that even in long-time expressions, where GSH addition is an issue, adequate amounts of crosslinked protein can be produced.

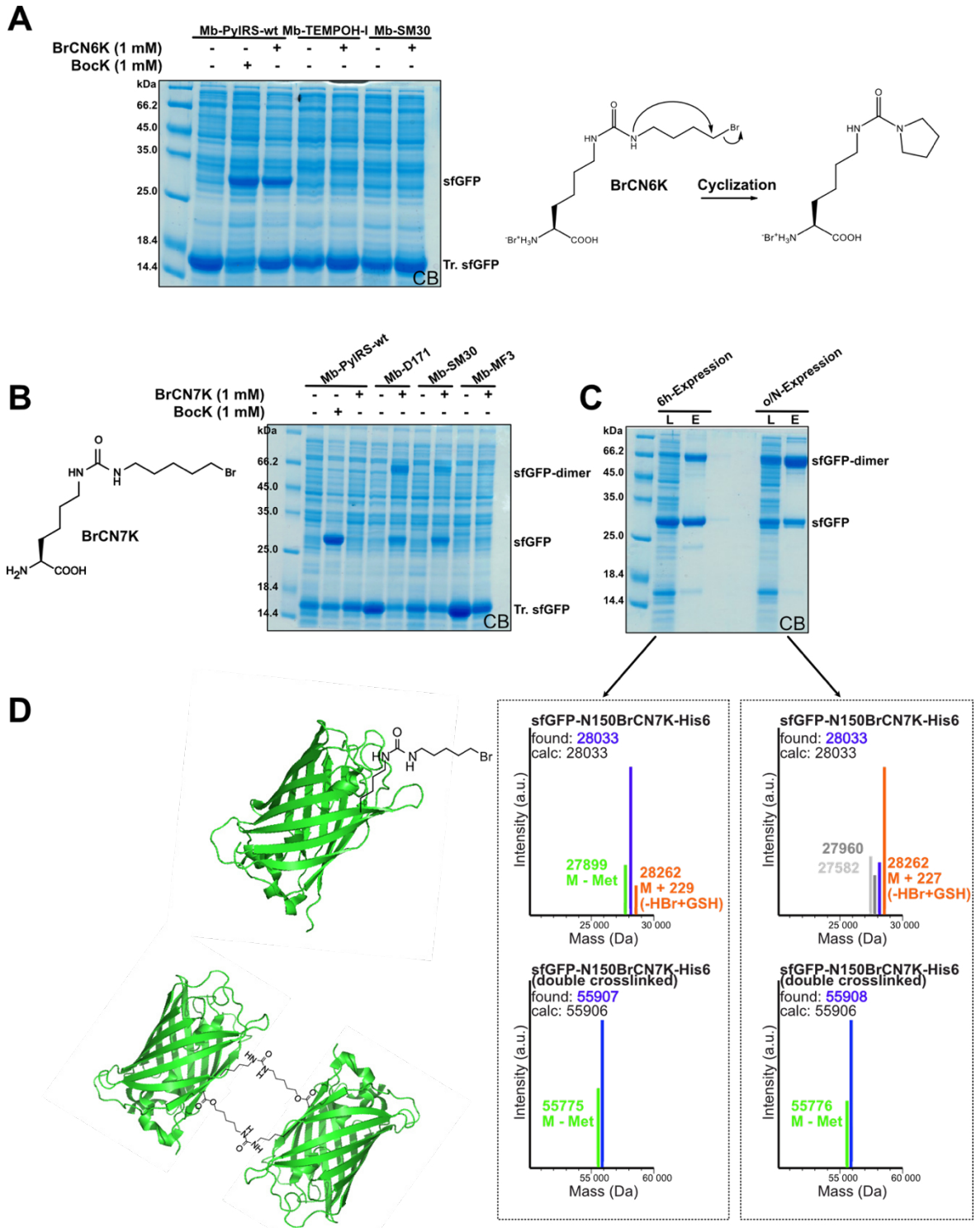


Figure 57: Amber suppression of sfGFP-N150TAG-His6 using BrCN6K or BrCN7K in *E. coli*. A: SDS gel of cell lysates of the incorporation test for BrCN6K in sfGFP-N150TAG-His6 using different *Mb*-PylRS mutants. Proposed intramolecular cyclization reaction of free BrCN6K. B: Structure of BrCN7K bearing a urea-moiety in its side chain and SDS gel of cell lysates of the incorporation test for BrCN7K in sfGFP-N150TAG-His6 using different *Mb*-PylRS mutants. C: SDS gel of the Ni-purification of sfGFP-N150BrCN7K-His6 obtained from either 6 h- or o/N-expression with *Mb*-TEMPOH-I. D: ESI-MS data (average masses) obtained from the purified sfGFP-N150BrCN7K-His6 samples in C. The sfGFP-monomer with incorporated BrCN7K and the double crosslinked sfGFP-dimer are schematically illustrated (PDB: 2b3q). (CB = Coomassie brilliant blue, CL = Cleared lysate, FT = Flow-through, W = Wash, E = Elution, GSH = glutathione)

The application of BrCN7K was further examined in HEK293T cell experiments. Initial experiments provoked cell death upon BrCN7K addition (data not shown). The last deprotection step in BrCN7K synthesis, entailed Boc-deprotection with HBr and the solvent was removed under reduced pressure and BrCN7K was obtained as oily substance, but no further purification was carried out. So, the toxicity likely stems from leftover HBr in the BrCN7K batch. While this didn't affect the experiments in *E. coli*, it did have a major impact on the cell viability of HEK293T cells. The BrCN7K was thus purified via HPLC to remove the excess of salts and used in a new round of amber suppression experiments in HEK cells. Now, the incorporation of BrCN7K into the reporter mCherry-TAG-GFP led to efficient readthrough and production of high GFP amounts without any toxic effects to the cells (Figure 58A). When comparing the obtained sfGFP levels in expressions with 1 mM of uAA, the GFP fluorescence were even higher for BrCN7K using *Mm*-TEMPOH-I (= BrCnKRS) than for the sample with BocK and *Mm*-PylRS. The SDS gel of Ni-elution fractions of sfGFP-N150BrCN7K-His6, expressed in co-transfected HEK cells with *Mm*-TEMPOH-I and supplemented with BrCN7K, displayed the two characteristic band for the sfGFP-monomer and -dimer (Figure 58B). The ESI-MS data of purified sfGFP-N150BrCN7K-His6 showed the correct mass for sfGFP-N150BrCN7K-His6 with a cleaved N-terminal methionine and an acetylation (27941 Da) (Figure 58C). The mass of the GSH adduct (28262 Da) was present as well.

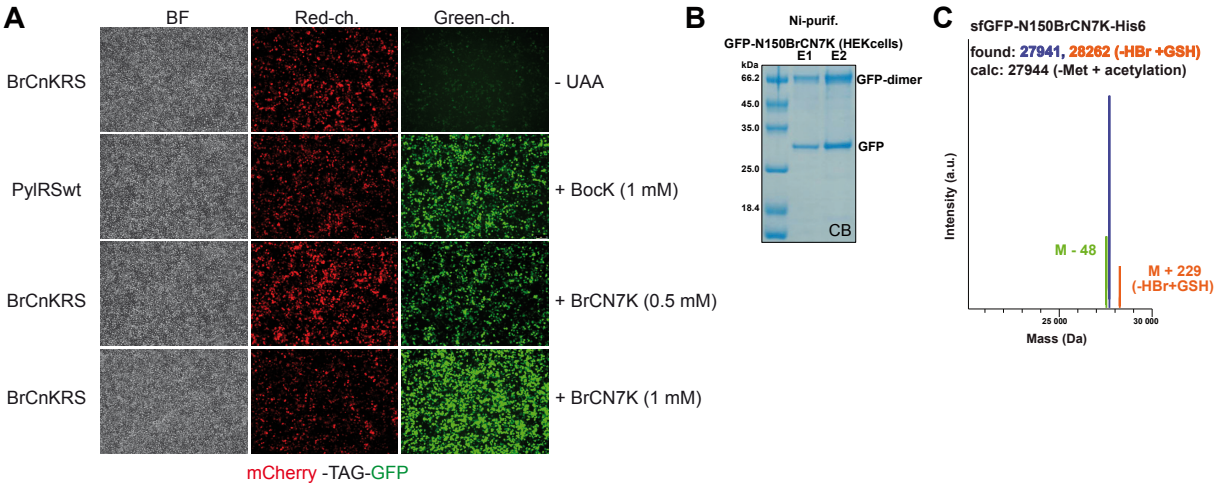


Figure 58: Amber suppression with BrCN7K in HEK cells. A: Fluorescence microscopy of amber suppression experiments in HEK293T cells using *Mm*-BrCnKRS (= *Mm*-TEMPOH-I) and BrCN7K 36 hours post-transfection. Fluorescence output of HEK cells co-transfected with a reporter plasmid containing mCherry-TAG-GFP and the plasmid encoding the respective *Mm*-PylRS variant. Cells were growing in medium with or without the respective uAA. B: SDS gel of the elution fractions (E1 + E2) of the Ni-purification of sfGFP-N150BrCN7K-His6 from HEK293T cells. C: ESI-MS data (average masses) of purified sfGFP-N150BrCN7K-His6 obtained from the 36h-expression in HEK293T cells co-transfected with the sfGFP-N150TAG-His6-reporter plasmid and the plasmid encoding *Mm*-BrCnKRS. (BF = bright field, ch. = channel, CB = Coomassie brilliant blue, GSH = glutathione)

As BrCN7K was sensitive to GSH, the stability of free BrCN7K was particularly investigated on a small molecule level. The completely deprotected uAA was incubated in PBS either supplemented with a 10-fold excess of GSH or without GSH. The samples were monitored by ESI-MS over time and the area underneath the chromatograms was used to quantify the results (Figure 59). The data confirmed that BrCN7K is labile in aqueous environments. The BrCN7K peak decreased within 24 hours to less than the half (Figure 59A). This was even worse in the presence of GSH (Figure 59B) where after 24 hours no BrCN7K was left. The emerging

BrCN7K-GSH adduct runs within the buffer peak and is thus not visible as separate peak within the chromatograms.

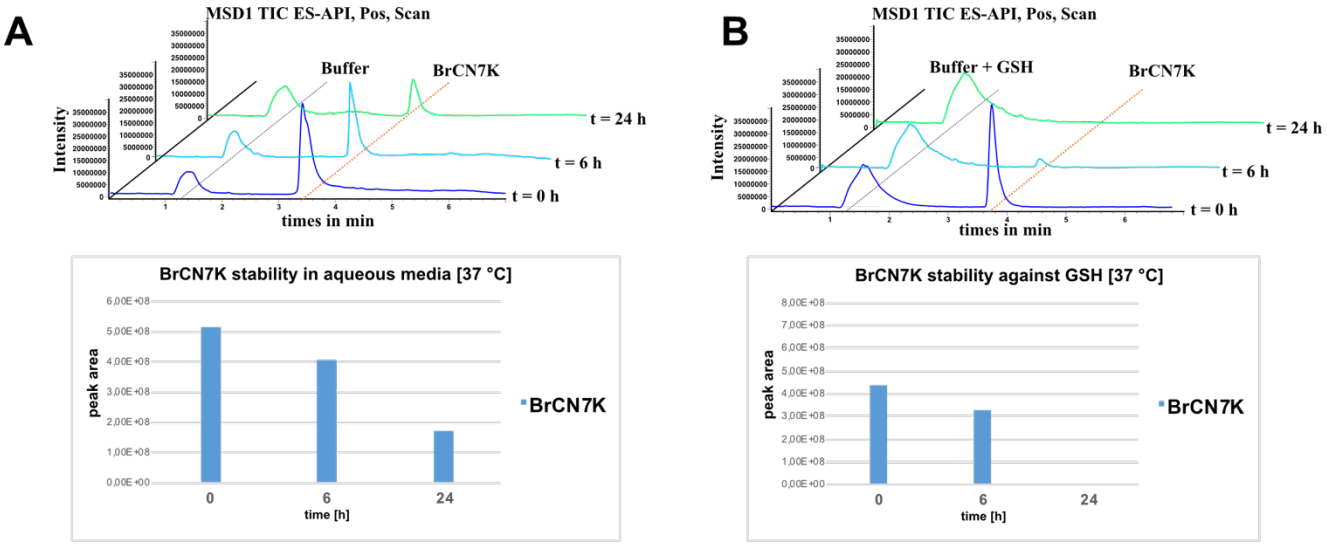


Figure 59: Stability test of free BrCN7K. A: At the top are shown the profiles of positive scans of the ESI-MS at different time points of samples containing 1 mM BrCN7K in PBS. The results were summarized in the diagram below. The bars represent the peak areas of BrCN7K (blue) and the buffer (orange) overtime. B: At the top are shown the profiles of positive scans of the ESI-MS at different time points of samples containing 1 mM BrCN7K and a 10-fold excess of GSH (10 mM) in PBS. The results were summarized in the diagram below. The bars represent the peak areas of BrCN7K (blue) and the buffer (orange) overtime. (GSH = glutathione)

To further address the crosslinking ability of BrCN7K in presence of GSH *in vitro,* we used as a model system the protein pair Rab1b-Q67A-R69TAG-His6 and DrrA$_{16-352}$-D82C, which we previously stabilized with BrC6K/BrCO6K. The single proteins, Rab1b-Q76A-R69BrCN7K-His6 (appendix 7.4.2 Figure 96A-B) and Strep-DrrA$_{16-352}$-D82C (appendix 7.4.1 Figure 91D-E) were purified and the crosslinking reaction was performed in the presence of different concentrations of GSH. The ESI-MS data of purified Rab1b-Q76A-R69BrCN7K-His6 already showed high amounts of GSH-adducts (appendix 7.4.2 Figure 96B). Nevertheless, crosslink formation over time between Rab1b-Q67A-R69BrCN7K-His6 and Strep-DrrA$_{16-352}$-D82C could be nicely detected for all samples and was not influenced by additional GSH (even upon addition of a 2000-fold excess of GSH compared to DrrA$_{16-352}$-D82C (10 µM DrrA$_{16-352}$-D82C versus 20 mM GSH) (Figure 60). Control experiments using other combinations of Rab1b and Strep-DrrA$_{16-352}$ variants confirmed that crosslink formation was exclusively observed between Rab1b-Q67A-R69BrCN7K-His6 and Strep-DrrA$_{16-352}$-D82C (appendix 7.4.2 Figure 97).

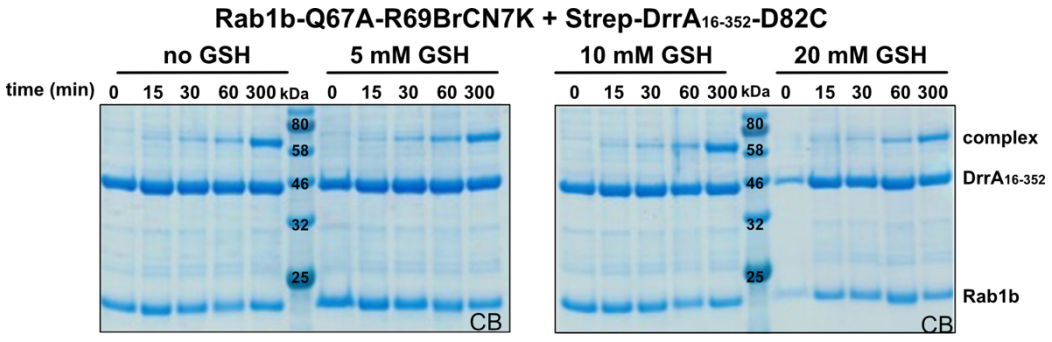


Figure 60: SDS gels of the time-dependent *in vitro* crosslinking of purified Rab1b-Q67A-R69BrCN7K-His6 and Strep-DrrA$_{16-352}$-D82C in the presence of different GSH concentrations. (CB = Coomassie brilliant blue, GSH = glutathione)

Our next goal was to investigate the fragmentation properties of BrCN7K in MS^2. Once more, we used the Rab1b-Q67A-R69TAG-His6-$DrrA_{16-352}$-D82C complex as the model system in our study. Simultaneous expression of Rab1b-Q67A-R69TAG-His6 and $DrrA_{16-352}$-D82C using *Mm*-TEMPOH-I and BrCN7K for amber suppression in *E. coli* yielded efficient *in vivo* complex formation where almost all Rab1b-Q67A-R69BrCN7K was reacted (Figure 61A). No unspecific bands could be observed compared to the sample using BrCO7K which also provided crosslinked Rab1b-Q67A-R69BrCN7K-His6-$DrrA_{16-352}$-D82C but some additional unrelated bands (Figure 61A). The purification of Rab1b-Q67A-R69BrCN7K-His6-$DrrA_{16-352}$-D82C complex worked without any problems and yielded pure protein complex for further XL-MS analysis (Figure 61B and appendix Figure 95C). The ESI-MS data of the purified Rab1b-Q67A-R69BrCN7K-His6-$DrrA_{16-352}$-D82C complex showed the correct mass for the mono-AMPylated and crosslinked proteins (59304 Da) a known modification at Y77 of Rab1b which is catalyzed by DrrA (Figure 61C, see 4.1.3.3).
Additionally, the sfGFP-dimer and monomer obtained in the aforementioned Ni-purifications of the test expressions with sfGFP-N150TAG-His6 were separated using SEC (Figure 95A-B) and used as a second sample for XL-MS experiments to investigate the fragmentation of BrCN7K in MS^2 experiments.

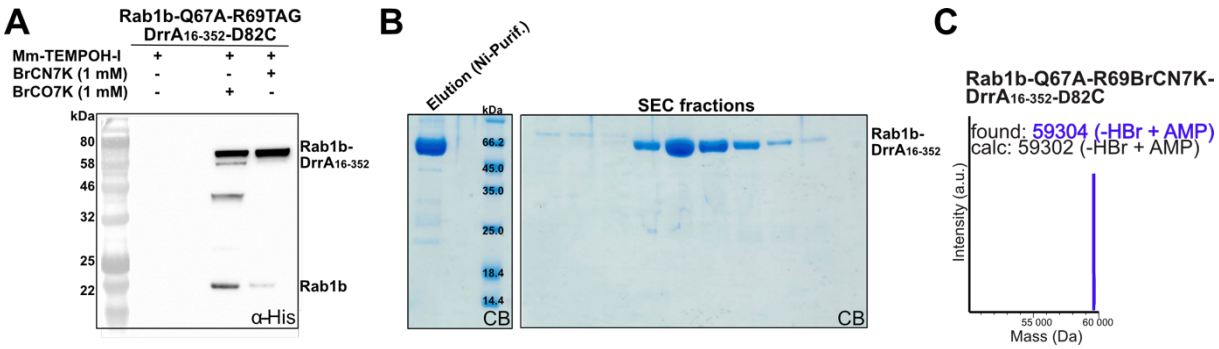


Figure 61: *In vivo* crosslinking of Rab1b and DrrA using BrCN7K. A: Western blot (α-His) of *E. coli* cell lysates from expression with co-transformed pBAD-Duet-Rab1b-Q67A-R69TAG-His6_$DrrA_{16-352}$-D82C and pEVOL-*Mm*-TEMPOH-I_PylT using either no uAA, BrCN7K or BrCO7K. B: SDS gels of the elution from the Ni-purification and SEC fractions of Rab1b-Q67A-R69BrCN7K-His6_$DrrA_{16-352}$-D82C complex. C: ESI-MS data (average masses) of purified Rab1b-Q67A-R69BrCN7K-His6_$DrrA_{16-352}$-D82C complex.

The MS^2 spectra obtained for Rab1b-Q67A-R69BrCN7K-His6-$DrrA_{16-352}$-D82C (one exemplary spectrum is illustrated in Figure 62) nicely showed the expected characteristic ion fragmentation pattern (26u doublets) formed by the MS-induced cleavage of the urea-motif of BrCN7K (marked in yellow) and enabled a fast recognition of the crosslinked peptide. The exact molecular structure and m/z values of these ion pairs are shown in Figure 74. A complete list of identified crosslinked peptide products obtained by MS^2 analysis of Rab1b-Q67A-R69BrCN7K-$DrrA_{16-352}$-D82C complex is annotated in the appendix 7.4.2 Figure 99. Furthermore, ions (y-ions and b-ions) resulting from the peptide backbone fragmentation were present and served for the identification of the crosslinking site. The y-ions were derived from C-terminal peptide fragments (labeled in blue) and b-ions belonged to N-terminal peptide fragments (labeled in red). For the Rab1b-Q67A-R69BrCN7K-His6-$DrrA_{16-352}$-D82C sample XL-MS clearly identified the crosslink position to 69 of Rab1b and 82 in DrrA and was drastically simplified using BrCN7K compared to the exhausting initial MS data evaluation of the Rab1b-Q67A-R69BrC6K-His6-$DrrA_{16-352}$ sample aforementioned in 4.1.3.3.

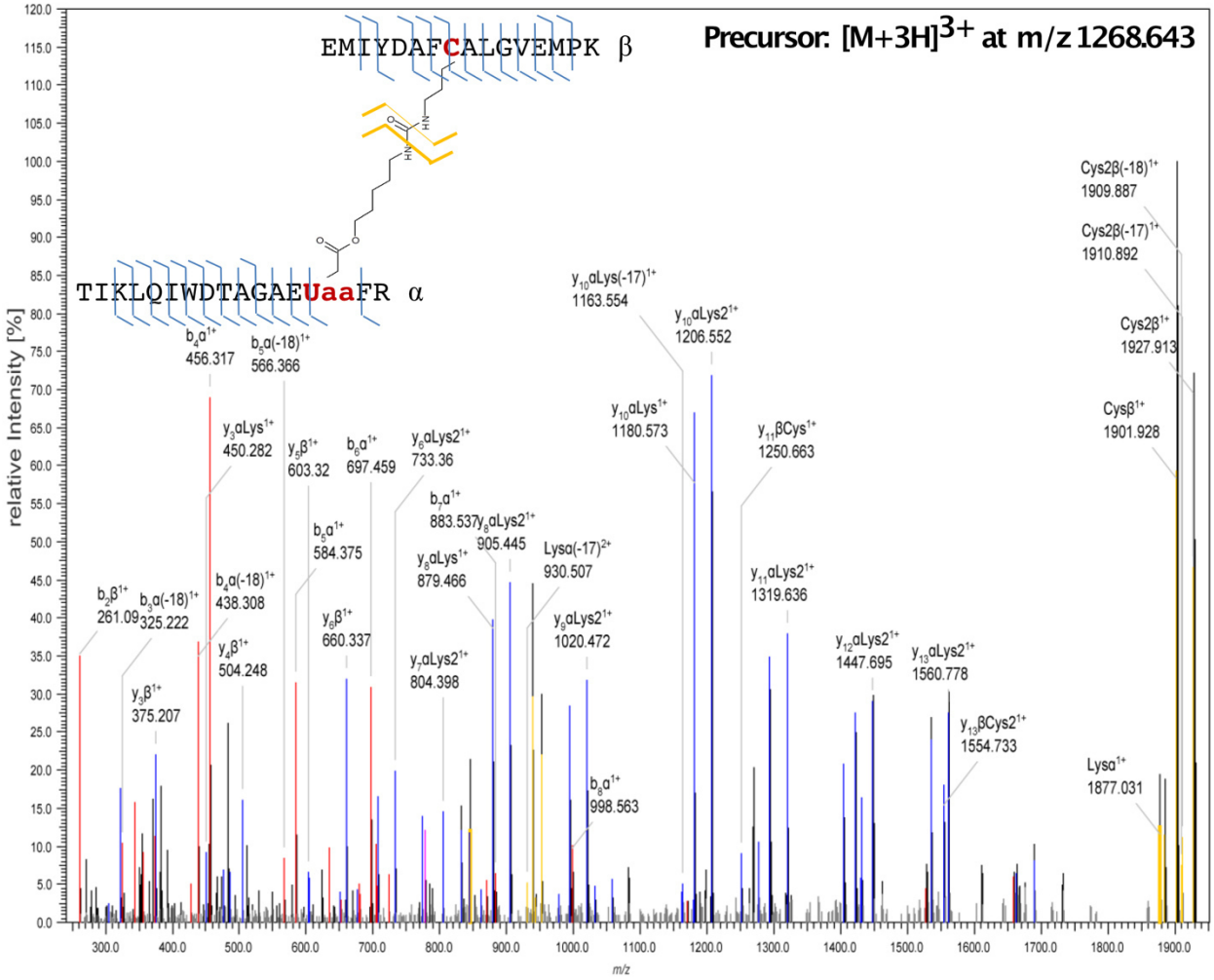


Figure 62: MS^2 analysis of the purified and digested Rab1b-Q67A-R69BrCN7K-His6_DrrA$_{16-352}$-D82C complex. MS^2 fragment spectrum of the Rab1b-Q67A-R69BrCN7K-His6_DrrA$_{16-352}$-D82C complex obtained from the data evaluation using MeroX (version 2.0.1.4). Cleavage products of the urea linker are colored in yellow. The y-ions (derived from C-terminal peptide fragments) are labeled in blue and b-ions (derived from N-terminal peptide fragments) are labeled in red which both belong to the cleavage of the peptide backbone (the lowered number indicates the number of residues encompassed in the fragment).

The crosslink formation between BrCN7K and a glutamate (E173 in sfGFP; identified by Cigler, Müller *et al.* for the sfGFP-dimer with BrC6K[172]) turned out to remain more tricky to identify from the MS^2-data evaluation because no crosslinked peptide ion products were found using the same standard settings in MeroX as for Rab1b-Q67A-R69BrCN7K-His6-DrrA$_{16-352}$-D82C complex. The reason was found in the large size of the crosslinked peptides (around 6.5 kDa) which arise during the trypsin digest of the sfGFP-dimer sample and which served for MS-analysis. Thus, the automatic signal annotation is sometimes imperfect, and the monoisotopic peak of the precursor ions cannot be detected rendering a software-assisted data evaluation not possible. A manual search for the masses of the expected product ions in the MS^2 successfully identified the characteristic fragmentation products belonging to the urea-linker (Figure 74). The signals could be also obtained by modifying the MeroX settings to higher precursor precision (1000 ppm instead of 10 ppm) combined with a special search for peptide ions linked to E173 with a charge state of 6+. An exemplary MS^2 spectrum and the identified crosslinked peptide products are shown in appendix 7.4.2 Figure 98. To address the peptide size problem an optimization of the enzymatic digest during the sample preparation would be useful e.g. through double digest with trypsin and endoproteinase AspN or GluC.

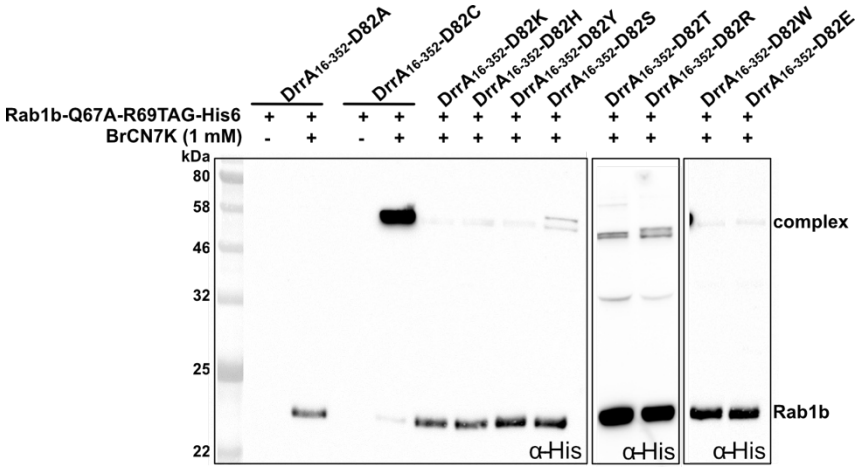


Figure 63: Western blot (α-His) of *E. coli* cell lysates expressing Rab1b-Q67A-R69BrCN7K-His6 and DrrA$_{16-352}$-D82X. (X = A, alanine; C, cysteine; K, lysine; H, histidine; Y, tyrosine, S, serine; T, threonine; R, arginine; W, tryptophan; E, glutamate)

Until now, two crosslinked protein samples, one linked via a thioether (BrCN7K to cysteine) the other by an ester (BrCN7K to glutamate), were investigated and resulted in characteristic fragmentation pattern of the urea-motif in MS^2. Further complexes linked via an ether bond (e.g. BrCN7K to tyrosine) or an amine linkage (e.g. BrCN7K to lysine) were also pursued for their potential use in XL-MS. Therefore, D82 in DrrA$_{16-352}$ was substituted to different nucleophilic amino acids and *in vivo* complex formation with Rab1b-Q67A-R69BrCN7K-His6 was examined (Figure 63). The crosslinking was significantly weaker than the one seen for D82C mutant of DrrA$_{16-352}$. However, minimal crosslink bands could be detected. The complexes Rab1b-Q67A-R69BrCN7K-His6-DrrA$_{16-352}$-D82Y, Rab1b-Q67A-R69BrCN7K-His6-DrrA$_{16-352}$-D82K and Rab1b-Q67A-R69BrCN7K-His6-DrrA$_{16-352}$-D82T were purified (appendix 7.4.2 Figure 100) and their XL-MS analyses are currently ongoing.

4.2.4. Summary & Outlook

In summary we designed two classes of MS^2-cleavable uAAs for PEPC in GCE approaches. One class consists of lysine derivatives bearing a sulfoxide moiety (TM-439) or a precursor (BrCOS8K) in their side chain for distinct fragmentation in CID for MS^2. The other class presents lysine derivatives with a urea-motif in their side chains (BrCN6K, BrCN7K), which after HCD, also provide characteristic ion fragmentation patterns in MS^2. Additionally, both classes of uAA are equipped with electrophilic moieties for PEPC, like a bromoalkyl-moiety (BrCN6K, BrCN7K, BrCOS8K) or a Michael-acceptor (TM-439). BrCN6K was excluded during the study as it underwent an undesired intramolecular cyclization reaction.
We found a *Mb*-PylRS mutant (*Mb*-TEMPOH-I) which accepts three of the MS/MS-cleavable uAAs (BrCOS8K, TM-439, BrCN7K). ESI-MS data of the incorporation of BrCOS8K into GFP showed, beside the correct mass, masses for GSH-adducts and of species without bromo-moiety. The incorporation of TM-439 using different *E. coli* strains as well as in HEK cells always provided masses for the reduced TM-439. This renders both uAAs inapplicable for our approach. BrCN7K was efficiently incorporated in *E. coli* and HEK cells and provided good crosslinks for the sfGFP-dimer and the Rab1b-Q67A-R69BrCN7K-DrrA$_{16-352}$-D82C complex. We could show that although BrCN7K is prone to GSH addition, the crosslink formation is still efficient enough even after prolonged expressions (o/N). The MS^2 analysis of thioether-linked

Rab1b-Q67A-R69BrCN7K-DrrA$_{16-352}$-D82C complex showed the expected fragmentation pattern for the crosslinked peptides making a fast and easy data evaluation of the crosslinking residues possible. The MS data evaluation for the ester-linked sfGFP-dimer turned out to be more complex due to the size of the peptides developed during the sample preparation. Nevertheless, we were able to identify the characteristic product ions for the urea-motif and detected crosslinked peptides and corresponding sites in the sfGFP-dimer. The crosslinking of protein samples using our new MS^2-cleavable uAAs facilitates the crosslink identification in MS^2 data evaluation and is highly useful due to the high natural abundance of aspartate and glutamate residues within proteins.

In order to expand our approach, future studies will focus on the fragmentation properties in MS^2 data evaluation of protein complexes which are linked via an ether bond (e.g. BrCN7K to tyrosine) or an amine linkage (e.g. BrCN7K to lysine) instead of a thioether or ester.

4.3. Trapping *Sa*ClpP with interactors

4.3.1. Background

The caseinolytic protease P is the central regulation machinery involved in protein quality control, protein turnover, stress response or gene regulation to ensure maintenance of cell integrity and homeostasis.[275-278] While ClpP was first discovered in *E. coli.*[279], it is not surprising that ClpP homologs exist in most bacterial and eukaryotic organisms apart from some archaea and fungi species.[280-282] To release energy-dependent proteolytic activity ClpP requires the association with subunits of Clp ATPases (ClpA or ClpX).[278] In organisms like in the pathogen *Staphylococcus aureus, Sa*ClpP furthermore acts as a key virulence regulator as it is reported that *S. aureus* strains with deletions of *clpX* and *clpP* show significantly lower levels of secreted virulence factors.[283, 284] *Sa*ClpP/*Sa*ClpX indirectly monitors the expression of major virulence factor genes, such as *hla,* by interfering with the global transcriptional virulence regulating *sar/agr* network. Due to its role in pathogenesis, *Sa*ClpP is a target of great interest in development of drugs for bacterial infections as alternative to conventional antibiotics.[283-285] Mechanistically, *Sa*ClpP is a serine protease that is able to enzymatically degrade protein or peptide substrates.[281, 286] *Sa*ClpP is organized in a cylindrical homotetradecameric complex consisting of two heptameric rings. Inside the complex is a chamber in which the fourteen active site serines, one per each *Sa*ClpP monomer (Figure 64B), are sequestered.[287] To obtain full and specific proteolytic activity, *Sa*ClpP depends on ATP-driven chaperones of the AAA+ family, ClpX and ClpC, which together with ClpP, form the complexes ClpXP or ClpCP.[288, 289] The chaperone recognizes, binds and prepares the substrate for further ClpP degradation by previous substrate unfolding.[290, 291]

The tetradecameric *Sa*ClpP structure exhibits three important characteristics: (i) Central interaction between the opposite heptameric rings is mediated by the handle domain (helix E).[292] (ii) N-terminal loops at the ends of the cylindrical complex are responsible for the interaction with chaperones.[288, 292] (iii) The inner side of the cylinder encompasses the active site residues and hydrophobic surfaces which provide the interface of the two big ring subunits.[293]

*Sa*ClpP, the "breathing enzyme", exists in three different conformational states (Figure 64A) which are associated with its proteolytic status. The extended conformation represents the proteolytic active state and is marked by a stretched helix E conformation whereas the inactive counterpart, known as compressed state, shows a kinked helix E conformation.[294-296] In a stretched conformation, helix E leads to a functional and active arrangement of the catalytic triad (Ser98, His123, Asp172) of each monomer in the inner core of *Sa*ClpP.[294, 297] The so-called compact state of *Sa*ClpP is an intermediate between the extended and compressed conformation.[292, 296, 298]

Due to the importance of *Sa*ClpP as potential target in the treatment of bacterial infections the development of *Sa*ClpP inhibitors has been an important research focus in the Siebr lab at TUM and several promising *Sa*ClpP inhibitor candidates were reported in recent years. Amon these are e.g. ß-lactone inhibitors[285], hydroxyamide-based inhibitors[299] or phenyl ester derivatives[300] which affect *Sa*ClpP in *S. aureus* and methicillin resistant *S. aureus* (MRSA).[301] More structural information that record the conformational effects of an inhibitor on *Sa*ClpP would provide

beneficial insights and contribute to an improved inhibitor design or the development of alternative strategies for *Sa*ClpP inhibition. In order to obtain *Sa*ClpP structures in complex with its inhibitors an effective and irreversible trapping approach of the enzyme-inhibitor intermediate is needed.

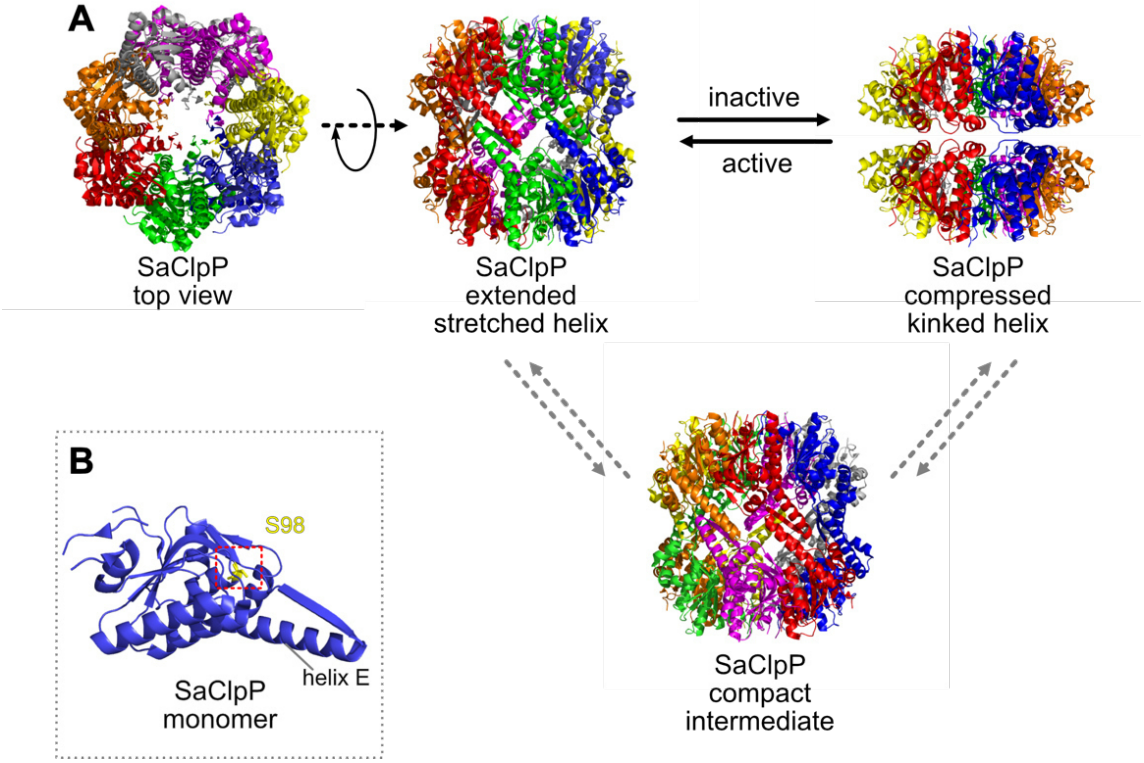


Figure 64: Architecture of the caseinolytic protease P from *S. aureus*. A: Top and side views of *Sa*ClpP in the extended (PDB: 3v5e), compressed (PDB: 3qwd) and compact (PDB: 3v5i) conformational states. B: Structure of *Sa*ClpP monomer (isolated from PDB: 3v5e) showcasing its catalytically active serine (S98) in yellow and a stretched helix E conformation.

An elegant way of covalently linking acyl-enzyme intermediates was recently published by the Chin group. Thereby the unnatural amino acid (uAA), 2,3-diaminopropanoic acid (DAP), was used to trap different intermediates of the valinomycin synthetase thioesterase domain bound to its substrates during the non-ribosomal peptide synthesis (NRPS) pathway.[179] The DAP-trapping strategy for proteins relies on the replacement of an active site serine or cysteine residue by DAP which reacts, similar to cysteine and serine, via nucleophilic attack at an electrophilic moiety in proximity (e.g. a carbonyl group within a substrate) and provides an amide-linked product (Figure 65B). In contrast to thioester or ester linkages, which would occur from the reaction with cysteine or serine, amide bonds are extremely stable.[179, 302] Therefore, DAP is ideally suited to stably trap intermediates of serine- or cysteine-catalyzed enzymatic reactions. The high similarity of DAP to serine or cysteine hampered the evolution of a specific orthogonal aaRS. To make the installation of DAP possible, or more precisely the evolution of a specific aaRS, a photo-protected version of DAP (pDAP) was used to successfully evolve an aaRS for pDAP. The pDAP can be deprotected after the incorporation into the protein leading to DAP at the desired position.[179] The deprotection strategy for pDAP is shown in Figure 65A. In the first step, UV light mediates the release of the photocage. The intermediate then undergoes a spontaneous fragmentation event and finally reveals DAP.[179]

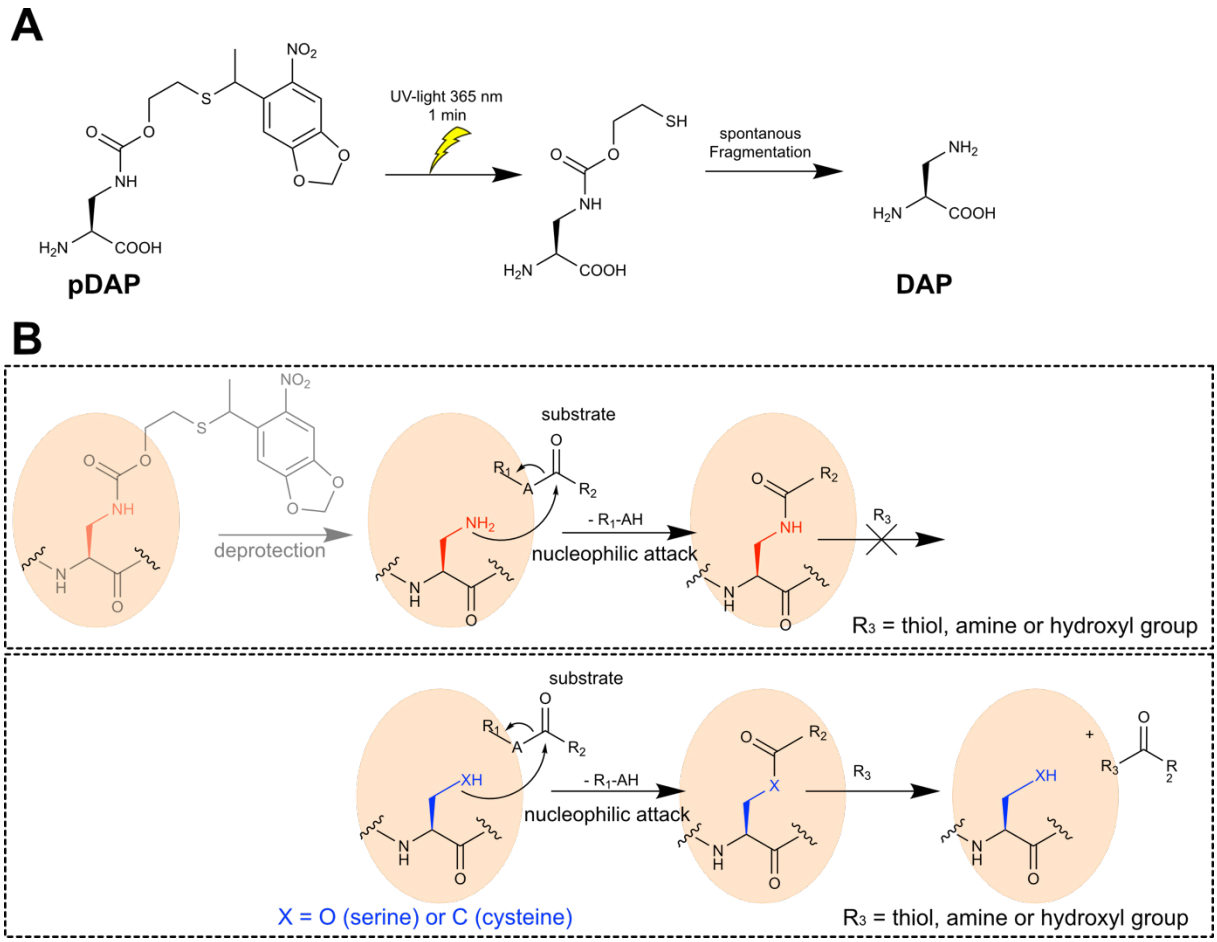


Figure 65: Covalently trapping acyl–enzyme intermediates with DAP. A: Light-mediated conversion of pDAP to DAP. Treatment with UV light (365 nm) triggers the release of the photocage moiety of pDAP. The emerged intermediate then spontaneously fragments to DAP which contains the desired reactive amino group. B: DAP, serine or cysteine attack the carbonyl group of the substrates yielding acyl-enzyme intermediates under release of R1-AH. The attack of a nucleophilic R_3 group leads to dissociation of acyl-enzyme intermediate in the case of serine or cysteine to regenerate the active site. In contrast the DAP-related acyl-enzyme intermediate is stable towards the attack of R_3. (Red indicates DAP side chain, blue indicates serine or cysteine side chain, orange spheres symbolize the active site of the protein, $R_1/R_2/A$ = various chemical groups, R_3 = nucleophilic moiety)

The ß-amino group of DAP incorporated into peptides exhibits a relatively low pK_a of 6.3 - 7.5[303] so that it can act as nucleophile in contrast to e.g. the ε-amino group of free lysine with a pK_a of 10.5. [304, 305] The reactive ß-amino group undergoes amide bond formation with electrophilic moieties (e.g. carbonyl groups) of substrates in its vicinity through nucleophilic attack. As already mentioned, amide bonds are very stable under aqueous conditions in contrast to other linkages (e.g. thioester or ester). Thus, complexes linked via DAP might be ideally suited for long-term experiments in aqueous environments such as crystallization. The hydroxyl group of serine alone is a poor nucleophile (pK_a of 14[306, 307]) and not able to nucleophilically attack its substrates like it is the case in *Sa*ClpP. To obtain the required nucleophilicity, it depends on the assistance of the two remaining residues of the catalytic triad, histidine and aspartate, which convert the hydroxyl group to the more reactive alkoxide by a mechanism known as charge relay system.[306, 307] We assume that, in contrast to the hydroxyl group of serine, the ß-amino group of DAP alone is a sufficiently strong nucleophile to react with electrophilic sites of molecules in proximity. Therefore, we hypothesized that the substitution of S98 of S*a*ClpP with DAP will enable us to stably trap the *Sa*ClpP-interactor complexes and enable their structural studies.

4.3.2. Aim

This project is part of a collaboration together with the group of Prof. Sieber (TUM) and was performed together with Konstantin Eckl (PhD student). The overall goal of the project is the structural elucidation of conformational effects caused by interactor binding to *Sa*ClpP via co-crystallization experiments of covalently trapped interactors (e.g. inhibitors or peptides) to *Sa*ClpP. Using genetic code expansion, we aim to incorporate pDAP into *Sa*ClpP in place of its active site serine (S98). After expression, purification and conversion of *Sa*ClpP-S98pDAP to *Sa*ClpP-S98DAP, the incorporated DAP should be able to covalently trap *Sa*ClpP inhibitors in its proximity through nucleophilic attack of its ß-amino group leading to irreversible acyl-enzyme-intermediates. Figure 66 illustrates the planned strategy of the *Sa*ClpP-pDAP project.

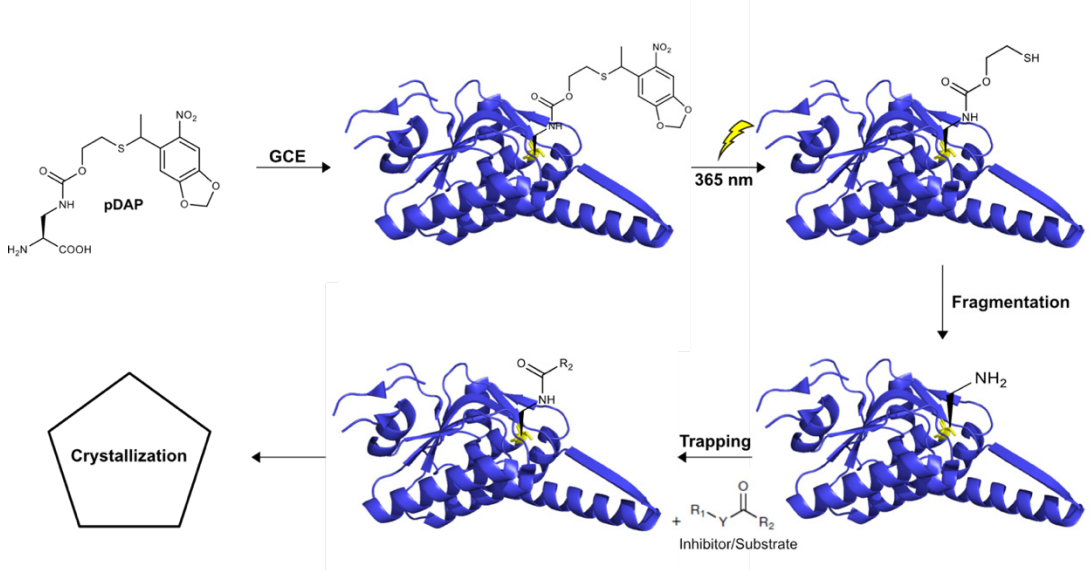


Figure 66: Approach for trapping the interaction partners of *Sa*ClpP: pDAP will be incorporated into *Sa*ClpP-S98TAg using genetic code expansion. After protein purification, photocage-release of pDAP is triggered by UV light (365 nm). The emerging DAP-intermediate undergoes spontaneous fragmentation ending up in DAP which acts as a nucleophile and can irreversibly react with electrophilic site of e.g. inhibitors in vicinity. The covalently trapped *Sa*ClpP-S98DAP-interactor complex is used for crystallization experiments for structure elucidation. Structure of *Sa*ClpP monomer was obtained from PDB 3v5e. (GCE = genetic code expansion)

Interim aims also include the establishment of an amber suppression system which efficiently incorporates pDAP in *Sa*ClpP-S98TAG, establishment of efficient purification protocols of *Sa*ClpP-S98pDAP, the establishment of decaging conditions for *Sa*ClpP-pDAP and the confirmation of the *Sa*ClpP-S98DAP integrity.

4.3.3. Results & Discussion

The chemical synthesis of pDAP was done by Konstantin Eckl and is shown in the appendix Scheme 17. The experiments for *Sa*ClpP-S98pDAP to *Sa*ClpP-S98DAP conversion, the trapping of *Sa*ClpP-S98DAP with inhibitors as well as the analytical SEC experiments were performed in collaboration with Konstantin Eckl.

4.3.3.1. Establishing an amber suppression system for pDAP

The first step towards the incorporation of pDAP into proteins is the creation of a suitable orthogonal aaRS/tRNA pair in a plasmid system which provides all necessary components,

aaRS, tRNA and the protein of interest. We decided to use the pBK/pPylT-vector system which harbors the *Mb*-PylRS-PylT pair for amber suppression and is well established in our lab. pBK- and pPylT-plasmids are well suited for simultaneous application within one cell as they can exist side by side due to their compatible origins of replication (ori) and antibiotic resistances. The pBK-plasmid has an pBR322 ori and an ampicillin resistance marker whereas pPylT-vectors uses tetracycline as selection marker and a p15A ori for replication (appendix 7.1.1). Moreover, this system is ideal for experiments which effort the expression of solely one protein of interest (POI). To site-specifically incorporate pDAP with our system we cloned the published mutations (Y271C, N311Q, Y349F and V366C[179]) into the wild type *Mb*-PylRS gene on our pBK-vector by successive site-directed mutagenesis rounds. The *Mb*-PylRS gene is regulated by the constitutive glnS promotor. The gene for *Sa*ClpP was installed on the pPylT-vector under control of an araBAD promotor together with the constitutively expressed *Mb*-PylT (lpp promotor). Initially, we cloned a C-terminal His6-tag in frame with *Sa*ClpP to make an easy and fast His-tag affinity purification possible. Nevertheless, we later changed the C-terminal affinity-tag to a StrepII-tag in order of stability issues during protein purification, which are discussed in detail in 4.3.3.2. The TAG-codon was introduced in place of the active site serine at position S98 of *Sa*ClpP. Beside the pPylT-plasmids bearing *Sa*ClpP-S98TAG, we further cloned *Sa*ClpP wild type and *Sa*ClpP-S98A mutant for both affinity-tag variants (His6-tag bearing variants into the pPylT-backbone; StrepII-tag variants into a pBAD-backbone), which could serve as reference proteins for later experiments. *Sa*ClpP-S98A is a catalytically inactive mutant of *Sa*ClpP and might serve as a negative control. In summary, we set up an amber suppression system using the *Mb*-PylRS-PylT pair. The pBK-plasmid encodes the *Mb*-PylRS mutant (Y271C, N311Q, Y349F and V366C = pDAPRS) for incorporation of pDAP. The other two essential components, *Mb*-PylT and *Sa*ClpP, are located on the compatible pPylT-plasmid. An overview about cloning procedure and plasmids is shown in the appendix 7.1.1 Table 64 Now, with a theoretically functional amber suppression system for incorporating pDAP into *Sa*ClpP in hand we tested its performance in *E. coli*.

4.3.3.2. Expression & Purification of *Sa*ClpP-S98pDAP

The recombinant expression of *Sa*ClpP and *Sa*ClpP-S98A as well as the amber suppression of SaClpP-S98TAG was performed in our standard host *E. coli* NEB 10-beta. To test the functionality of our plasmids constructs we set up expressions of single *Sa*ClpP-His6 proteins as well as amber suppression of *Sa*ClpP-S98TAG-His6 followed by Ni^{2+}-affinity purification of each protein. *Sa*ClpP is expressed at 25 °C, as established by the Sieber group.[308, 309] As in our experience amber suppression using M*b*-PylRS/tRNA pair works best at 37 °C we decided to perform expression of all three *Sa*ClpP variants at this temperature.

Single expression of *Sa*ClpP proteins (wild type and S98A mutant) showed nice over expression bands in the SDS-PAGE only for induced cell lysate samples. The subsequent purification via Ni-slurry behaved as expected and yielded purified *Sa*ClpP variants with a molecular weight around 22 kDa (Figure 67A).

To test amber suppression of *Sa*ClpP-S98TAG, the pBK-pDAPRS and the respective pPylT-*Sa*ClpP-S98TAG were co-transformed into *E. coli* NEB 10-beta. Additionally, we used the site-specific incorporation of BocK (accepted by *Mb*-PylRS and typically serves as gold standard for us) as control. The SDS gel of cell lysates obtained from the amber suppression experiment did not clearly show the presence of *Sa*ClpP (Figure 67B). This was not surprising, as the

protein yields during amber suppression are reduced compared to non-amber suppressed proteins due to the competition between translational termination mediated RF1 binding and amber-mediated read-through at the TAG-position during the protein expression[310]. To verify whether amber suppression of *Sa*ClpP-S98TAG produced full-length protein, western blot (α-His) analysis was performed, which clearly showed that comparable amounts of intact protein were present in samples where BocK or pDAP were added. In contrast, no protein bands were visible in cell lysates without added uAAs (Figure 67B). This indicated that amber suppression of the position S98 of *Sa*ClpP worked successfully. Purified proteins were obtained after Ni^{2+}-purification, however the yield of *Sa*ClpP-S98pDAP was significantly worse compared to *Sa*ClpP-S98BocK (Figure 67B). This was quite surprising since the western blot suggested relatively similar amounts of produced *Sa*ClpP.

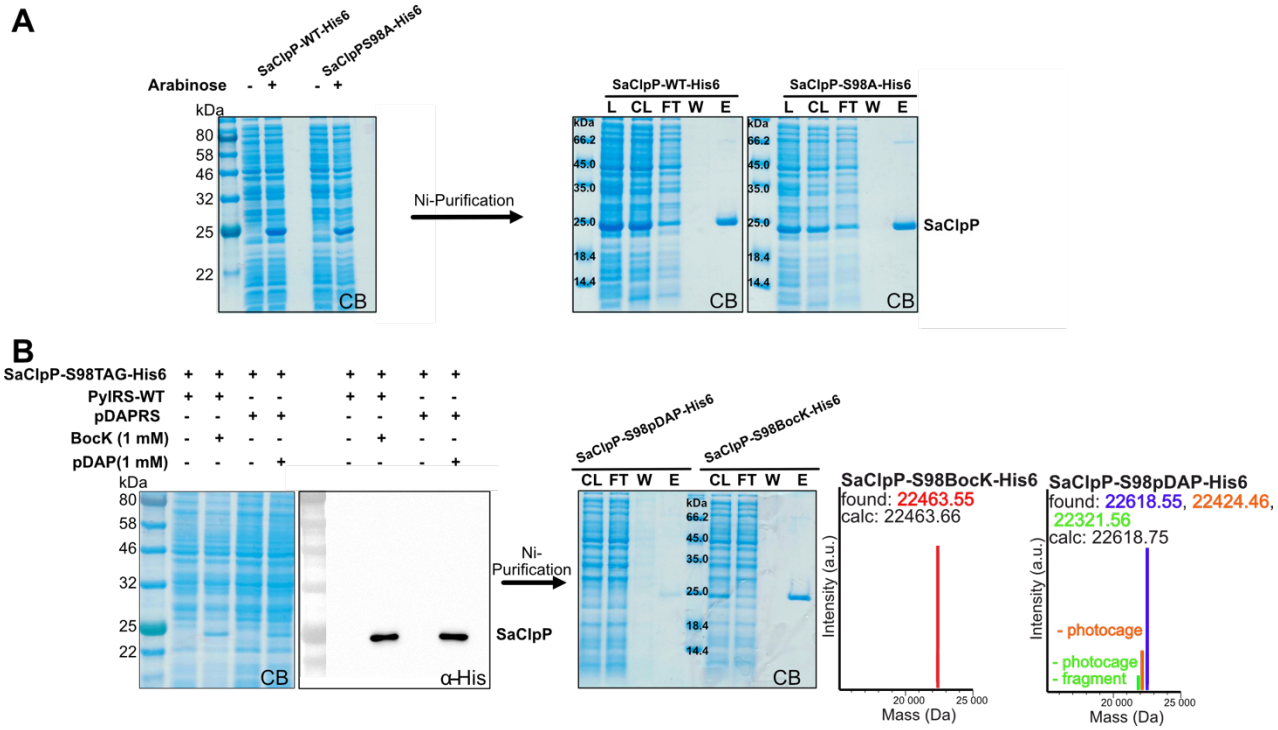


Figure 67: Expression and Ni^{2+}-affinity purification of *Sa*ClpP variants in *E. coli* NEB 10-beta. A: SDS gels of cell lysates of the expression and purification of *Sa*ClpP-His6 and *Sa*ClpP-S98A-His6. B: SDS gel and western blot (α-His) of cell lysates of amber suppression of *Sa*ClpP-S98TAG-His6 with BocK and pDAP. The purified proteins were analyzed by ESI-MS (exact masses). (CB: Coomassie brilliant blue; CL: Cleared lysate; FT: Flow through; W: Wash; E: Elution)

Nevertheless, ESI-MS analysis of the intact proteins confirmed the site-specific incorporation of BocK (22463 Da) and pDAP (22618 Da) into *Sa*ClpP. The spectrum of the *Sa*ClpP-S98pDAP sample harbors three mass peaks (Figure 67B). The main peak (blue; 22618 Da (calc.: 22618 Da) belongs to *Sa*ClpP-S98pDAP-His6. The additional masses correspond to species lacking the photocage moiety (orange; 22424 Da (calc.: 22423 Da)) and to decaged DAP (green; 22321 Da (calc.: 22321 Da). The decaging had likely occurred during the expression and purifications procedures due to daylight exposure. During the subsequent concentration and re-buffering process of *Sa*ClpP-His6, *Sa*ClpP-S98A-His6, *Sa*ClpP-S98pDAP-His6 and *Sa*ClpP-S98BocK-His6, protein precipitation occurred. We supposed that this is a consequence caused by the His6-tag of *Sa*ClpP and not of the incorporation of pDAP as the same was observed for *Sa*ClpP-His6 and *Sa*ClpP-S98A-His6 as well. Encouraged by the

experiences of the Sieber group, we attempted to circumvent this problem by switching to a StrepII-tag-based protein purification. A simple affinity-tag replacement at *Sa*ClpP should not affect the functionality of our plasmid system, which we confirmed by expressions and purifications of *Sa*ClpP with replaced affinity-tags (Figure 68).

The SDS gels in Figure 68A illustrate the over expression of *Sa*ClpP proteins (wild type and S98A mutant) in cell lysates at 30 °C which were further successfully purified using the ÄKTA system with a Strep-affinity column (protein band at 22 kDa). We lowered the expression temperature to obtain higher protein yields in order to have enough protein which then can be used to analyze the solubility behavior during the buffer exchange and concentration step. As hoped for the proteins did not tend to precipitate anymore over the re-buffering and concentration steps, establishing the usage of StrepII-constructs as standards for all following experiments. Expression of *Sa*ClpP-StrepII and *Sa*ClpP-S98A-StrepII yielded 2.88 $\frac{mg}{L}$ and 3.1 $\frac{mg}{L}$ protein, related to the culture volume. Due to the fact that we used another *E. coli* expression host (*E. coli* NEB 10-beta cells instead of *E. coli* BL21 (DE3) cells), an araBAD-driven instead of a T7-driven plasmid system and other expression conditions our yields differ from the published ones which are between 10 – 80 $\frac{mg}{L}$ for *Sa*ClpP variants expressed in *E. coli* BL21 (DE3) at 37 °C.[296]

To obtain best incorporation of pDAP into *Sa*ClpP-S98TAG-StrepII, we needed to find the sweet spot between efficient *Sa*ClpP expression (normally at 25 °C) and efficient amber suppression (normally at 37 °C). Therefore, we screened different expression temperatures (25 °C, 30 °C, 37 °C), arabinose amounts (0.02 %, 0.01 %) for induction and induction time point OD_{600} = 0.5; OD_{600} = 0.7) (Figure 68B). While the arabinose amount and the OD_{600} at induction did not influence the protein yields of *Sa*ClpP-S98pDAP-StrepII, expression at 30 °C provided the highest protein amounts. Consequently, the expression conditions for *Sa*ClpP-S98pDAP-StrepII were determined at 30 °C after induction at an OD_{600} = 0.7 with 0.02 % arabinose. *Sa*ClpP-S98pDAP-StrepII was successfully purified via Strep-Tactin® XT Superflow (Figure 68C) with yields between 0,71 – 1,2 $\frac{mg}{L}$ and the intact protein MS (Figure 68D) showed the correct masses for *Sa*ClpP-S98pDAP-StrepII (main peak blue 22835 Da (calc.: 22385 Da)) and its related species (orange 22642 Da (- photocage) (calc.: 22642 Da); green 22538 Da (fully converted) (calc.: 22538 Da)). The *Sa*ClpP-S98pDAP-StrepII yield is clearly below the one of non-amber suppressed *Sa*ClpP variants (around 50 %). This is not unexpected and a consequence of competition at the TAG-codon between RF1 and pDAP-*Mb*-PylT during protein translation *in vivo*. Nevertheless, with this optimized plasmid system and protein purification protocol we were able to access *Sa*ClpP-S98pDAP-StrepII for initial decaging and trapping experiments.

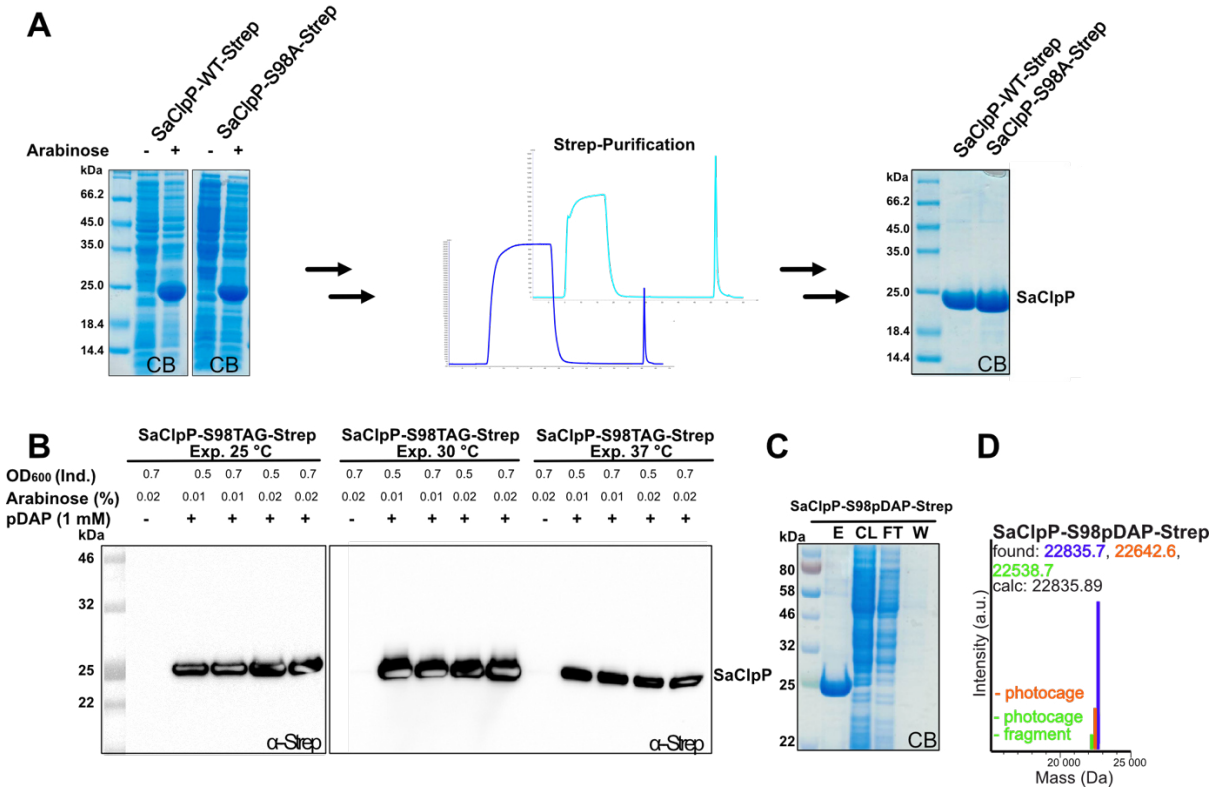


Figure 68: Expression and Strep-affinity purification of *Sa*ClpP variants in *E. coli* NEB 10-beta. A: SDS gels of cell lysates of the expression at 30 °C and of purified of *Sa*ClpP-StrepII and *Sa*ClpP-S98A-StrepII after ÄKTA purification (ÄKTA chromatograms at 280 nm). B: Western blot (α-Strep) of cell lysates of the amber suppression of *Sa*ClpP-S98pDAP-StrepII using different temperatures, inductions time points and arabinose amounts. C: SDS gel of the Strep-purification of *Sa*ClpP-S98pDAP-StrepII using Strep-Tactin® XT Superflow. D: ESI-MS (exact masses) of purified *Sa*ClpP-S98pDAP-StrepII. (CB: Coomassie brilliant blue; CL: Cleared lysate; FT: Flow through; W: Wash; E: Elution)

4.3.3.3. Conversion of pDAP to DAP

The initial decaging experiments were carried out on small molecule level on fully protected pDAP. To investigate the photocage release mediated by UV light, pDAP solutions in MeOH were exposed to 365 nm for different periods of time and the course of the reactions was followed by ESI-MS (UV trace at 254 nm, Figure 69B). The sample without UV exposure showed one dominant peak with a retention time around 6 min which was assigned to the UV active pDAP (the UV activity of pDAP arises from the presence of the photocage moiety). With longer exposure time to 365 nm, the pDAP peak dissappears and a new peak arises with a retention time of 4 min which corresponds to the released photocage. Due to photocage removal the remaining DAP-intermediate has no significant UV activity and is no longer visible at 254 nm. After 5 min irradiation time no peak for pDAP can be detected leading to the assumption that all pDAP has successfully released the photocage.

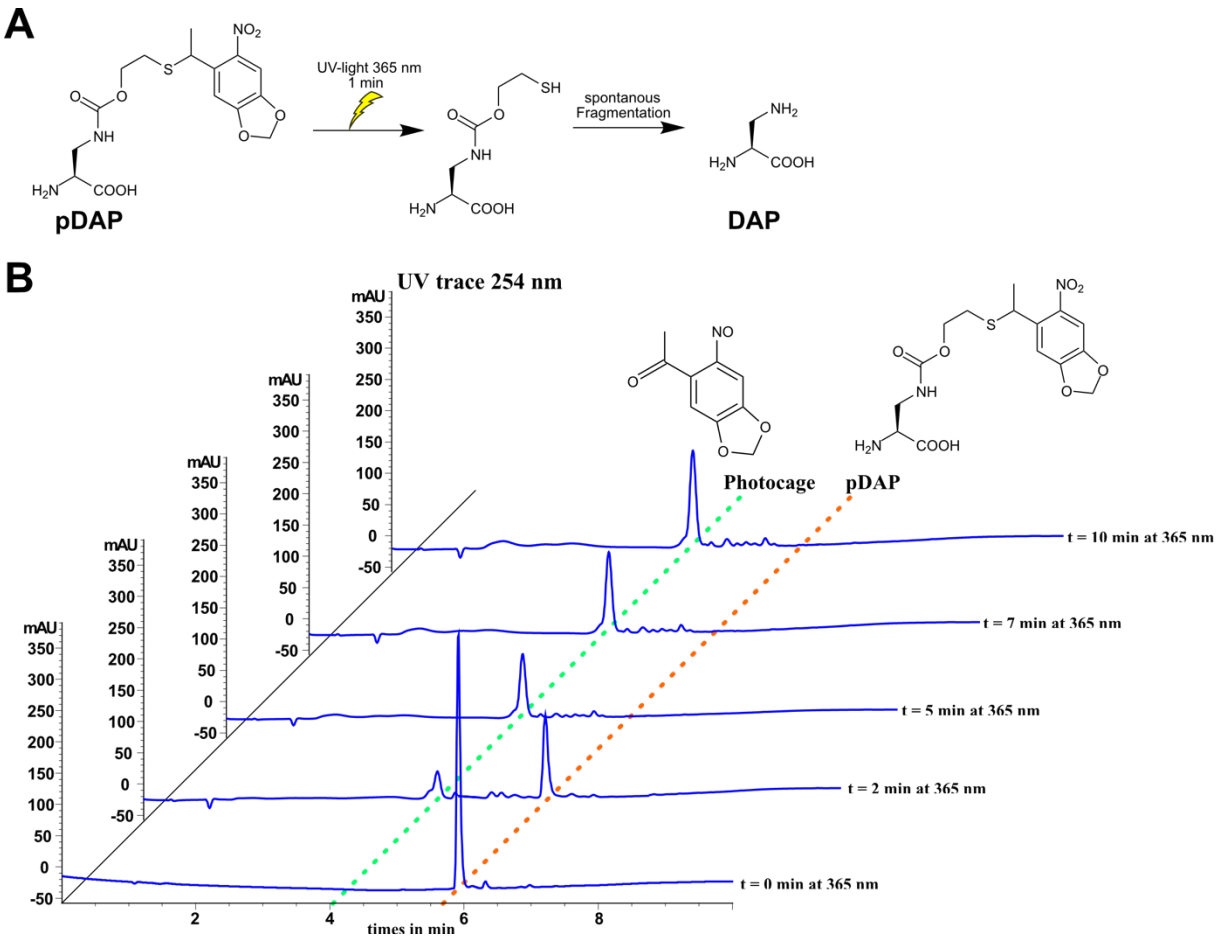


Figure 69: Conversion of pDAP to DAP. A: Scheme of the conversion of pDAP to DAP. The photocage of pDAP is released triggered by UV light (365 nm). The emerging intermediates spontaneously fragments to DAP. B: Schematic course of the photo decaging reaction illustrated by the UV traces at 254 nm from LC-ESI-MS analysis of pDAP samples after irradiation with UV at 365 nm for different time intervals. Orange dashed line marks the disappearing peak of the pDAP while the green dashed line marks the arising peak of the released photocage. The formed DAP is not UV active and therefore not detected in the respective chromatograms.

Next step was the establishment of pDAP conversion on purified *Sa*ClpP-S98pDAP-StrepII. Figure 70A shows the three species of *Sa*ClpP which appear during the conversion process of *Sa*ClpP-S98pDAP-StrepII to *Sa*ClpP-S98DAP-StrepII. ESI-MS (LTQ) of non-irradiated samples only has one main peak for *Sa*ClpP-S98pDAP-StrepII (blue 22835 Da) and two minor side peaks for the intermediate (orange 22642 Da) and *Sa*ClpP-S98DAP-StrepII (green 22538 Da). The peak allocation for irradiated (9 min; 365 nm) *Sa*ClpP-S98pDAP-StrepII samples showed a different pattern directly after exposure. Now, the main peak belongs to the *Sa*ClpP-S98pDAP intermediate (orange 22642 Da) indicating an almost complete removal of the pDAP photocage. After further prolonged incubation (12 hours at 4°C & 12 hours at 32 °C) of the irradiated *Sa*ClpP-S98pDAP sample the peak pattern of the ESI-MS changes again. The main peak belongs to *Sa*ClpP-S98DAP (green 22538 Da) which arose through spontaneous fragmentation of the former *Sa*ClpP-S98pDAP intermediate. Figure 70B shows the ESI-MS data obtained during the reaction course *Sa*ClpP-S98pDAP-StrepII conversion experiment. *Sa*ClpP-S98pDAP-StrepII samples were irradiated for different periods of time (5 min, 7 min, 9 min) in PZ buffer pH 7.0 In all samples the pDAP photocage could be removed, with best decaging efficiencies after 9 min UV exposure, but small amounts of *Sa*ClpP-S98pDAP were still detected. The following fragmentation event was further monitored at 15 °C over time. Almost no fragmentation occurred after 13 hours incubation. Even after a further 24 hours

incubation, a maximum conversion rate of around 50 % (9 min sample) could be obtained. In contrast to the published approach for sfGFP-N150pDAP (1 min at 365 nm; 10 hours incubation at 37 °C for full conversion to sfGFP-N150DAP) we need longer UV light exposure to obtain photocage release and much longer incubation times for the spontaneous fragmentation event.[179] This effect could be the result of the relatively shielded and inaccessible pDAP location at position 98 in *Sa*ClpP in contrast to exposed position N150 at the protein surface of sfGFP.

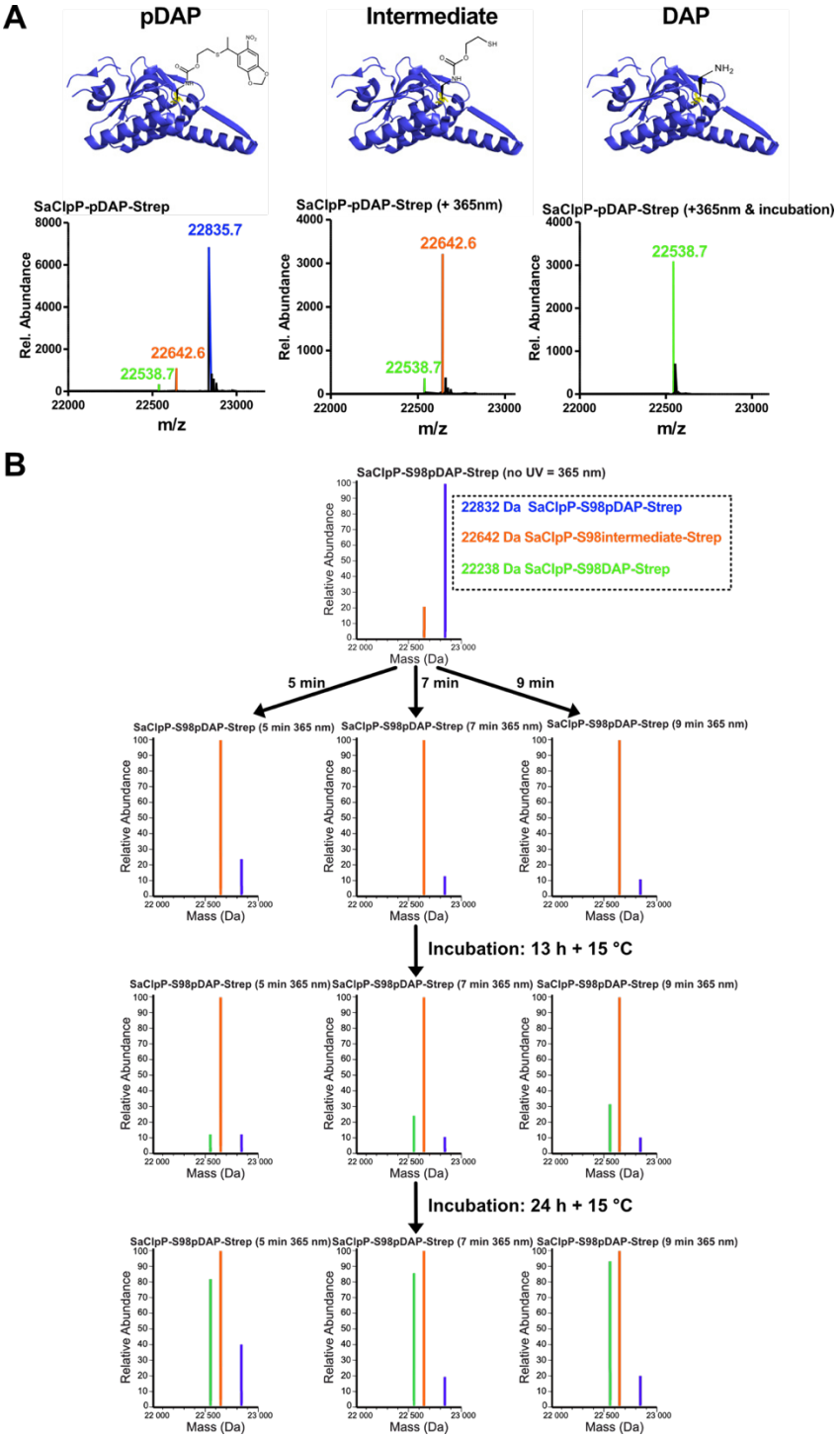


Figure 70: Photo decaging and conversion of *Sa*ClpP-S98pDAP-StrepII to *Sa*ClpP-S98DAP-StrepII. A: *Sa*ClpP-S98pDAP-StrepII species during the conversion process with the corresponding ESI-MS data (exact masses). B: ESI-MS data (exact masses) obtained from different decaging experiments with *Sa*ClpP-S98pDAP-StrepII before and after irradiation (5min, 7 min, 9 min UV light exposure (365 nm)) as well as during incubation. Structure of *Sa*ClpP monomer was obtained from PDB: 3v5e.

From these first experiments we concluded that the critical step of the pDAP to DAP conversion on *Sa*ClpP-S98pDAP is the spontaneous fragmentation event after photocage release. Therefore, we investigated the influence of pH value and temperature for the fragmentation process in order to improve the conversion process. We irradiated purified *Sa*ClpP-S98pDAP-StrepII in PZ buffer (pH 7.0) for 10 min at 365 nm followed by incubation at 4 °C or 15 °C overnight. The results obtained from the ESI-MS data (not shown) are summarized in Figure 71A. Incubation o/N at 4°C causes almost no fragmentation. Only a marginal improvement could be obtained for the sample at 15 °C indicating that higher temperatures are favorable for the fragmentation. We used the 4 °C sample for a second incubation step (o/N at 32 °C) and reached an almost complete fragmentation (around 85 %) of *Sa*ClpP-S98pDAP-StrepII to *Sa*ClpP-S98DAP-StrepII.

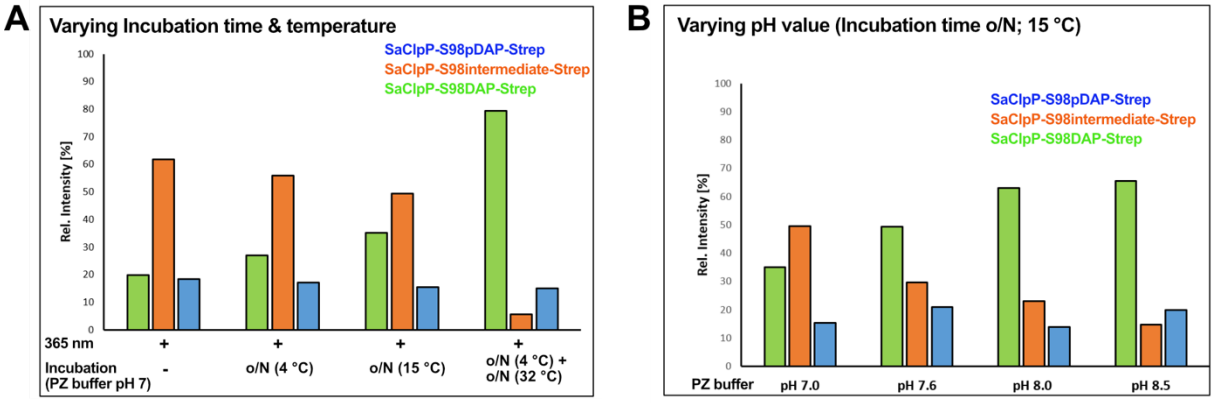


Figure 71: Investigation of the role of pH and temperature for the spontaneous fragmentation event for *Sa*ClpP-S98pDAP intermediate to *Sa*ClpP-S98DAP. A: Summarized ESI-MS data of the incubation of irradiated *Sa*ClpP-S98pDAP-StrepII at different temperatures (4°C, 15 °C, 32 °C). B: Summarized ESI-MS data of the Incubation of irradiated *Sa*ClpP-S98pDAP-StrepII in buffers with different pH values (pH 7.0; pH 7.5; pH 8.0; pH 8.5). (o/N = overnight)

Furthermore, we monitored the fragmentation behavior of the *Sa*ClpP-S98pDAP-StrepII intermediate at different pH values and summarized the resulting ESI-MS data in Figure 71B. The decaged *Sa*ClpP-S98pDAP-StrepII was incubated o/N at 15 °C in PZ buffer of varying pH. With the increasing pH of the buffer, the fragmentation also becomes more efficient. Highest conversion rate of pDAP intermediate to DAP was observed in PZ buffer pH 8.5. Ultimately, spontaneous fragmentation to desired *Sa*ClpP-DAP favors higher pH values (pH around 8.0 – 8.5) and higher incubation temperatures (around 32 °C). The final experiments for the optimization of the conversion process still have to be done as this was not possible due to technical problems with the ESI-MS equipment.

4.3.3.4. Trapping of *Sa*ClpP-interactors

The next important step of the *Sa*ClpP project was the evaluation of covalent trapping of interactors, such as small molecule inhibitors, by *Sa*ClpP-S98DAP. To achieve this, *Sa*ClpP-S98DAP was incubated with two different known *Sa*ClpP inhibitors, D3 and AV170 (Figure 72A) which covalently modify S98 in *Sa*ClpP.[300, 311] D3[285] is a ß-lactone-based inhibitor while AV170[300] belongs to the inhibitor class of phenyl ester derivates. We monitored alterations of the protein mass caused by the covalent addition of the inhibitor via ESI-MS. For the *Sa*ClpP-S98DAP sample incubated with D3 no mass shift occurred indicating that no modification of *Sa*ClpP-S98DAP with D3 took place (data not shown).

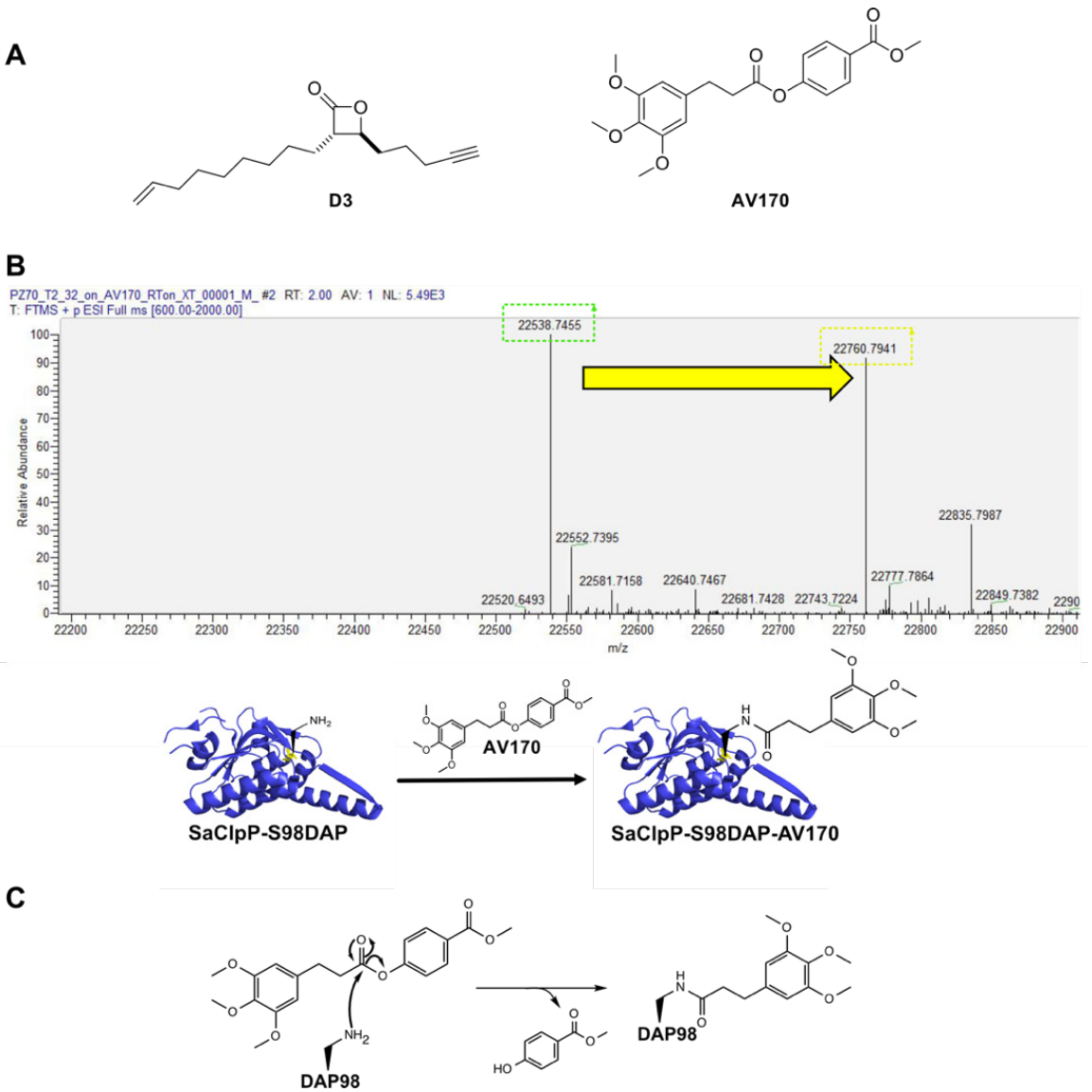


Figure 72: Trapping small molecule inhibitors to *Sa*ClpP-S98DAP. A: Structures of the small molecule inhibitors D3 and AV170 used in this study. B: ESI-MS data (exact masses) of the observed mass shift (visualized by the big yellow arrow) after incubation of *Sa*ClpP-S98DAP with AV170 (*Sa*ClpP-S98DAP-StrepII found: 22538 Da and calc.: 22538 Da; *Sa*ClpP-S98DAP-AV170-StrepII found: 22760 Da and calc.: 22760 Da) and *Sa*ClpP species resulting from the experiment (structure of *Sa*ClpP monomer was obtained from PDB: 3v5e). C: Proposed reaction mechanism for the trapping of AV170 to *Sa*ClpP-S98DAP.

Nevertheless, a distinct mass shift of + 222 Da, was detected for the sample with AV170, which corresponds to the crosslinking reaction between DAP and AV170 (*Sa*ClpP-S98DAP: 22538 Da (green); *Sa*ClpP-S98DAP-AV170: 22760 Da (yellow) (Calc.: 22760 Da)) (Figure 72B). This modification was stable for at least 24 hours at 32 °C (data not shown). Figure 72C illustrates the likely reaction mechanism between DAP and AV170. The nucleophilic ß-amino group of DAP (at position 98) attacks the electrophilic carbon of the ester group in AV170, followed by the release of the phenolic rest. Thus, *Sa*ClpP-S98DAP could successfully trap AV170 *in vitro* showing that our approach is suited for the trapping of *Sa*ClpP with interactors.

4.3.3.5. *Sa*ClpP-S98DAP integrity

In parallel, we set up studies to investigate the oligomeric state of *Sa*ClpP-S98pDAP and *Sa*ClpP-S98DAP using analytical SEC runs with *Sa*ClpP as reference for intact homotetradecameric assembly. Figure 73 shows the resulting ÄKTA chromatograms at

280 nm. The *Sa*ClpP reference is colored in red and has the big peak for the tetradecameric state at a retention volume of 11.5 mL. Heptameric *Sa*ClpP species run normally at a retention volume of approximately 13 mL as known from published data using the same experimental set-up.[312] The chromatogram of *Sa*ClpP-S98pDAP (blue) shows no distinct peak for neither the tetradecameric conformation nor the heptameric but a broad peak over 7 – 13 mL retention volume. This might be caused by aggregates of a heterogenous mixture of *Sa*ClpP-S98pDAP oligomers. The incorrect oligomeric state might be a consequence of the incorporation of bulky pDAP instead of the relatively small serine at the central position 98. This could lead to misfolding events and disruption of interactions which might be important for correct oligomer assembly. The green line marks the chromatogram of *Sa*ClpP-S98DAP. After conversion from pDAP to the much smaller DAP, the former broad peak from SaClpP-S98pDAP disappears and a peak around 12 mL retention volume appears. It is conceivable that this peak represents the compressed state of the tetradecameric *Sa*ClpP. This conformation was already observed for a *Sa*ClpP-D172N mutant with a retention volume of 12 mL.[312] The removal of the photocage and the fragmentation to the small DAP seem to at least partially reduce the aggregates and to reinstall the oligomeric assembly of *Sa*ClpP-S98DAP. This has to be analyzed in more detail in further experiments.

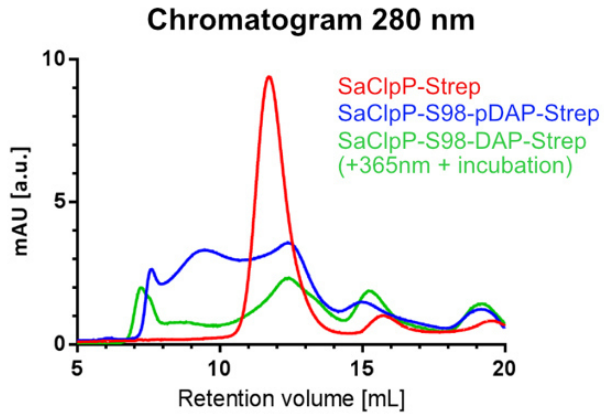


Figure 73: Analysis of the oligomeric state of *Sa*ClpP. The chromatograms at 280 nm of analytical SEC runs are shown for *Sa*ClpP-S98pDAP (blue line), *Sa*ClpP-S98DAP (green line) and *Sa*ClpP (red line).

4.3.4. Summary & Outlook

In summary, we were able to establish an amber suppression system for the incorporation of pDAP into *Sa*ClpP. Furthermore, we could demonstrate that the position 98 of the active site serine is suited for replacement with pDAP using genetic code expansion. We successfully expressed and purified *Sa*ClpP-S98pDAP-StrepII. Through experiments on small molecule level, we determined that the exposure time of 10 min at 365 nm is sufficient to remove the photocage on purified *Sa*ClpP-S98pDAP. The limiting step for full conversion to *Sa*ClpP-S98DAP is represented by the spontaneous fragmentation of the DAP-intermediate to DAP. We figured out that the fragmentation event is pH and temperature dependent. Best conversion conditions were obtained at pH of 8.5 and higher temperatures (e.g. 32 °C). Using this approach, we have demonstrated that *Sa*ClpP-S98DAP could be covalently modified with the small molecule inhibitor AV170 and that the stapled *Sa*ClpP-AV170 complex is stable at least over 24 hours at 32°C. We were also able to partially rescue the oligomeric assembly of *Sa*ClpP-

S98pDAP by conversion from the bulky pDAP to DAP. The results so far encourage us to intensively pursue the project.

Next steps in this work would encompass the final refinement and establishment of the conversion conditions of *Sa*ClpP-S98pDAP to *Sa*ClpP-S98DAP and the transfer to large scale set-ups. Furthermore, screening of more inhibitors or peptide substrates which can be trapped by *Sa*ClpP-S98DAP has to be done. Finally, promising candidates of covalently trapped *Sa*ClpP-S98DAP-interactor complexes will be used in crystallization experiments for structural elucidation.

5.MATERIALS & METHODS

5.1. Chemical Work

5.1.1. Chemicals & Solutions

Chemicals were obtained in the purest available quality from the suppliers Acros Organics, Alfa Aesar, Applichem, Carbolution, Carl Roth, Iris Biotech, Jena Bioscience, Sigma Aldrich, Thermo Fisher Scientific, New England Biolabs and Tokyo Chemical Industry and used without further purification. Anhydrous solvents were used in chemical reactions if not used in combination with water. Technical grade solvents, obtained from the *Materialverwaltung*, were distilled before usage. For protection during the reaction, an argon-filled balloon was placed on the reaction vessel.

Table 5 Standard solutions for chemical works

Solution	Amount/Concentration
Citric acid	20 % (w/v)
NaCl	saturated
$NaHCO_3$	saturated
NaOH	1 M
Na_2SO_3	20 % (w/v)

5.1.2. Thin Layer Chromatography

The reaction course, educt conversion and product formation of chemical reactions were monitored by thin layer chromatography (TLC) using Silica gel 60 F_{254} TLC plates (Merck, Germany). For the visualization of spots on the TLC plate a UV-lamp at 254 nm wavelength and the staining solutions from Table 6 were used.

Table 6 Thin layer chromatograpy solutions

Staining solution	Composition	Target	Detection
Potassium permanganate	1.5 g $KMnO_4$ 10 g K_2CO_3 1.25 mL 10% NaOH 200 mL H_2O	Olefins, oxidizable groups	yellow spots on purple background
Ninhydrin	0.72 g ninhydrin 200 mL EtOH	Amines	pink to blue spots on light background

5.1.3. Silica Flash Column Purification

For flash chromatography, technical grade solvents, like DCM and pentane, which were obtained from *Materialverwaltung* were distilled before use. EtOAc and MeOH were purchased in analytical grade purity from Fisher Scientific. The separation of reactions, containing several compounds, occurred over silica gel with a grain distribution of 40-60 µm and a pore size of 60 Å (Silica gel, Ultra Pure, 40 - 60µm, 60Å, Acros Organics, Germany).

5.1.4. High Performance Liquid Chromatography

The Shimadzu LC-20AT Prominence system, consisting of two pumps (LC-20AT), a degassing unit (DGU-20A_{3R}), a diode array detector (SPD-M-20A), a fraction collector (FRC-10A) and a communications bus module (CBM-20A) was used for HPLC purification of small molecules. As columns served a Phenomenex Luna C18, 5 µm (4.6 x 250 mm) column to separate compounds on an analytical scale; a Phenomenex Luna C18, 5 µm (10 x 250 mm) column and a Phenomenex Luna C18, 10 µm (21.2 x 250 mm) served for preparative scale runs. Solvent A (MilliQ-H_2O+ 0.1 % FA) and solvent B (ACN + 0.1 % FA) could be used without filtration.

For preparative purification of BrCO6K and BrCN7K a gradient from 5 % to 80 % solvent B in 10 min was used, followed by 80 % solvent B for 3.5 min. Then solvent B was decreased from 80 % to 5 % within 30 sec and remained at 5 % until the end of run (total 23 min). Runs were performed with a flow rate of 10 $\frac{mL}{min}$ with the Luna 10 µm column.

For preparative purification of HpGGK a gradient from 5 to 15 % solvent B in 17 min, followed by a gradient from 15 to 80 % solvent B in 30 sec was used. Solvent B remained at 80 % for 3.5 min before decreasing solvent B from 80 to 5 % within 30 sec. Until the end of run (total 30 min) solvent B stayed at 5 %. Runs were performed with a flow rate of 10 $\frac{mL}{min}$ with the Luna 10 µm column.

5.1.5. Liquid Chromatography Mass Spectrometry

An Agilent 1260 Infinity Series LC system with an Agilent 6210 ESI Single Quadrupole mass spectrometer was used for recording LC-MS spectra. As solvents served solvent A (MilliQ-H_2O + 0.1 % FA) and B (ACN + 0.1 % FA) (unfiltered). For the analysis of small molecules a Jupiter C18 5 µm (2 x 150 mm) capillary column (Phenomenex, Torrence, USA) at rt was used with a flow rate of 0.5 $\frac{mL}{min}$ and with a gradient from 5 - 95 % B in 6 min. Measurements were done in positive and negative ion mode and the spectra were analyzed with OpenLab ChemStation Edition Software C.01.07.SR3 [465].

5.1.6. Nuclear Magnetic Resonance Spectroscopy

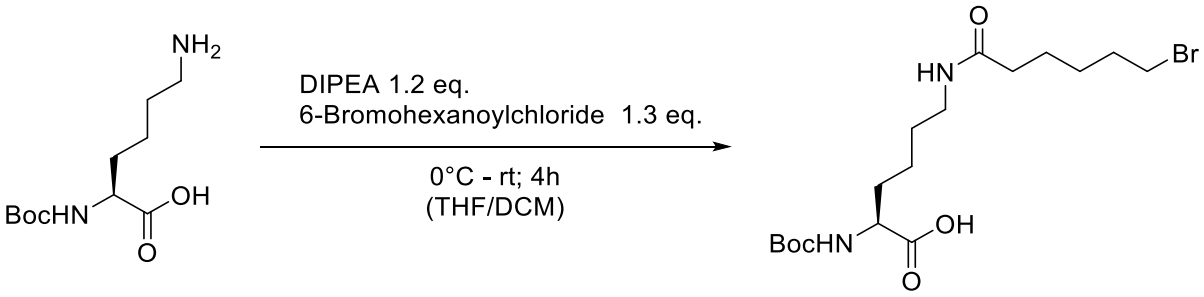
Bruker Systems AVHD300 (300; 75 Hz) or AVHD400 (400; 100 Hz) or AVHD500 (500 Hz) were used to measure ^{1}H and ^{13}C spectra of small molecules. The chemical shift δ [ppm] was adjusted to the residual proton peak of the used solvent. The characterization of the signal

multiplicity is organized in s (singlet), d (doublet), t (triplet), q (quartett), h (heptet), m (multiplet) and their combinations.

5.1.7. Chemical Synthesis

5.1.7.1. Synthesis of BrC6K

Synthesis of *α*-Boc-BrC6K – Amide coupling



Scheme 1 Synthesis of Boc-BrC6K – Amide coupling

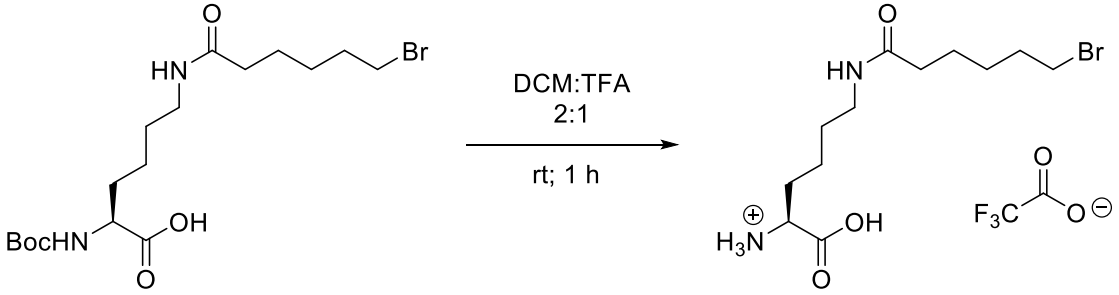
Experimental procedure

α-Boc-Lys-OH (1.00 eq.) was suspended in DCM/THF (1:1; 3.7 mL per mmol *α*-Boc-Lys-OH) and DIPEA (1.20 eq.) was added. The suspension was cooled to 0 °C and 6-Bromohexanoylchloride (1.30 eq.) was added dropwise. The reaction was stirred for 4 hours and during it warmed up to rt. EtOAc (7.4 mL per mmol *α*-Boc-Lys-OH) was added to the reaction and the organic phases were washed with 10 % citric acid (4 times; 5 mL per mmol *α*-Boc-Lys-OH) and brine (2 times; 8 mL per mmol *α*-Boc-Lys-OH). The organic phases were dried over Na_2SO_4 and the solvent was removed under reduced pressure. The crude product was purified by flash column chromatography (DCM:MeOH + 0.25 % acidic acid = 99:1 → 93:7). *α*-Boc-BrC6K was isolated as brown oily substance in 34 % yield.

^{1}H NMR (500 MHz, MeOD) δ = 4.05 (dd, ^3J = 9.1, 4.7 Hz, 1H), 3.45 (t, ^3J = 6.7 Hz, 2H), 3.17 (t, ^3J = 6.9 Hz, 2H), 2.19 (t, ^3J = 7.4 Hz, 2H), 1.93 – 1.74 (m, 3H), 1.71 – 1.59 (m, 3H), 1.59 – 1.36 (m, 15H) ppm.

MS(ESI): m/z calc. for $C_{17}H_{31}O_5N_2Br$: 422.14 (for ^{79}Br), 424.14 (for ^{81}Br); found: 445.1 $[M+Na]^+$ (for ^{79}Br), 447.1 $[M+Na]^+$. (for ^{81}Br)

Synthesis of BrC6K – Boc deprotection



Scheme 2 Synthesis of BrC6K – Boc deprotection

Experimental procedure

α-Boc-BrC6K (1.00 eq.) was dissolved in DCM (7 ml per mmol *α*-Boc-BrC6K) and then TFA (3.5 mL per mmol *α*-Boc-BrC6K) was added (DCM:TFA corresponds to a 2:1 ratio). The

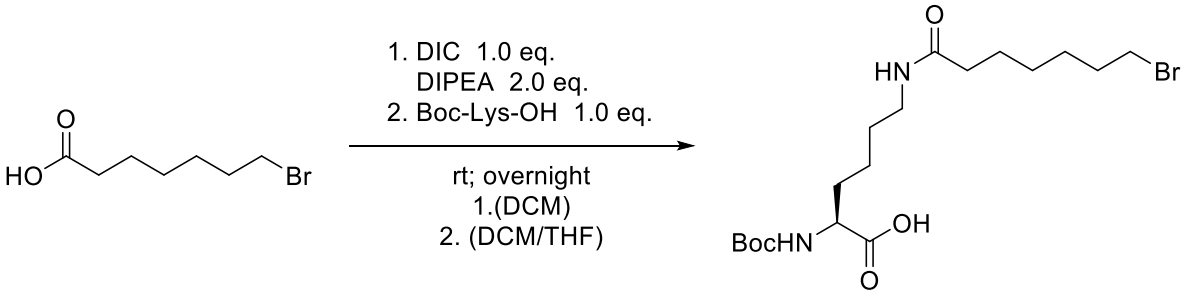
reaction was stirred for at least 1 hour until TLC showed full conversion. The solvent was removed under reduced pressure. The crude product was dissolved in MeOH (1.8 mL per mmol *α*-Boc-BrC6K) and washed with ice cold diethyl ether (3 times; 10 mL). The product was separated by centrifugation (4°C, 4200 rpm, 5 min) and the supernatant was discarded. BrC6K was isolated as white powder in 94 % yield.

^{1}H NMR (300 MHz, DMSO) δ = 8.18 (d, ^3J = 3.1 Hz, 3H), 7.77 (t, ^3J = 4.9 Hz, 1H), 3.87 (q, ^3J = 5.8 Hz, 1H), 3.50 (t, ^3J = 6.7 Hz, 2H), 3.02 (q, ^3J = 6.2 Hz, 2H), 2.05 (t, ^3J = 7.3 Hz, 2H), 1.85 – 1.65 (m, 4H), 1.58 – 1.20 (m, 8H).

MS(ESI), m/z calc. for $C_{12}H_{23}BrN_2O_3$: 322.09 (for ^{79}Br), 324.09 (for ^{81}Br); found: 323.1 $[M+H]^+$ (for ^{79}Br), 325.1 $[M+H]^+$ (for ^{81}Br)

5.1.7.2. Synthesis of BrC7K

Synthesis of *α*-Boc-BrC7K – Amide coupling



Scheme 3 Synthesis of Boc-BrC7K – Amide coupling

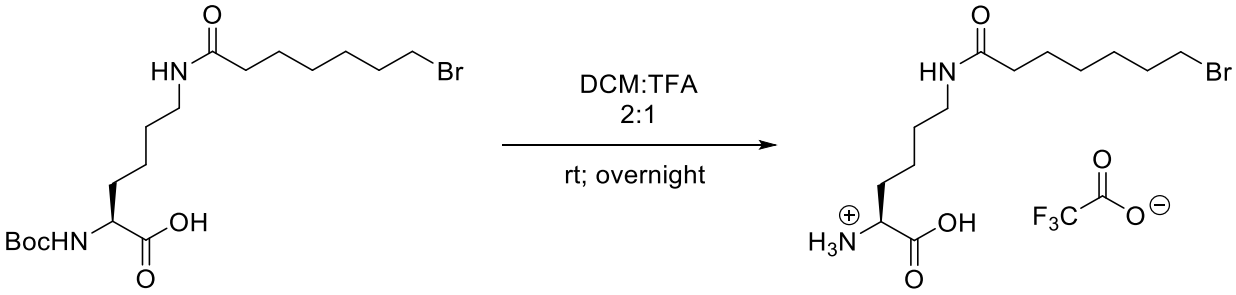
Experimental procedure

7-Bromoheptanoic acid (2.00 eq.) was dissolved in anhydrous DCM (14 mL per mmol 7-Bromoheptanoic acid) and DIPEA (2.00 eq.) as well as DIC (1.00 eq.) were added. The reaction was stirred for 1.5 hours until TLC showed full conversion. The solution was given to a suspension of *α*-Boc-Lysine-OH (1.00 eq.) in anhydrous THF (26 mL per mmol *α*-Boc-Lysine-OH) and the reaction was stirred overnight. The organic phase was washed with 10% citric acid (3 times; 30 mL per mmol *α*-Boc-Lysine-OH) and the combined aqueous phases were extracted with DCM (30 mL per mmol *α*-Boc-Lysine-OH). The combined organic phases were washed with brine (46 mL per mmol *α*-Boc-Lysine-OH), dried over Na_2SO_4 and the solvent removed under reduced pressure. The crude product was purified via flash column chromatography (DCM:MeOH + 0.25% acidic acid = 99:1 → 93:7) and. *α*-Boc-BrC7K was isolated as colorless oil in 46 % yield.

^{1}H NMR (300 MHz, DMSO-d_6) δ = 7.72 (t, ^3J = 5.6 Hz, 1H), 6.98 (d, ^3J = 8.0 Hz, 1H), 3.86 – 3.76 (m, 1H), 3.51 (t, ^3J = 6.7 Hz, 2H), 2.99 (q, ^3J = 6.3 Hz, 2H), 2.03 (t, ^3J = 7.4 Hz, 2H), 1.78 (p, ^3J=6.8 Hz, 2H), 1.68 – 1.20 (m, 21H) ppm.

MS(ESI), m/z calc. for $C_{18}H_{33}BrN_2O_5$: 436.16 (for ^{79}Br), 438.16 (for ^{81}Br); found: 435.0 $[M-H]^-$ (for ^{79}Br), 437.0 $[M-H]^-$ (for ^{81}Br).

Synthesis of BrC7K – Boc deprotection



Scheme 4 Synthesis of BrC7K – Boc deprotection

Experimental procedure

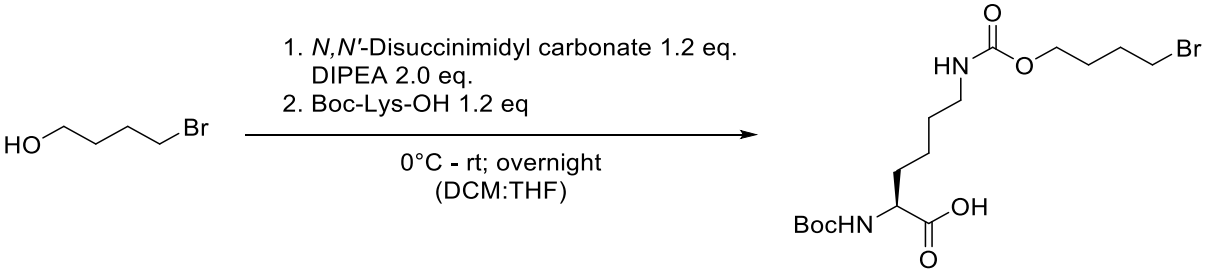
α-Boc-BrC7K (1.00 eq.) was dissolved in DCM (6.7 mL per mmol *α*-Boc-BrC7K) and the solution was cooled on ice. Then TFA (3.3 mL per mmol *α*-Boc-BrC7K) was added and the reaction was stirred for 1 hour at rt until TLC showed full conversion. The solvent was removed under reduced pressure. The crude product was dissolved in MeOH (0.5 mL per mmol *α*-Boc-BrC7K) and washed with cold diethyl ether (-20 °C, 3 times 10 mL). The product was separated by centrifugation (4°C, 4200 rpm, 5 min) and the supernatant was discarded. The product was dried on air and BrC7K isolated as colorless powder in 69 % yield.

^{1}H NMR (300 MHz, Deuterium Oxide) δ = 3.79 (t, ^3J = 6.4 Hz, 1H), 3.21 (t, ^3J = 6.7 Hz, 2H), 2.94 (t, ^3J = 6.7 Hz, 2H), 1.99 (t, ^3J = 7.3 Hz, 2H), 1.78 – 1.60 (m, 2H), 1.55 (p, ^3J = 6.7 Hz, 2H), 1.37 – 1.25 (m, 4H), 1.23 – 0.98 (m, 6H) ppm.

MS(ESI), m/z calc. for $C_{13}H_{25}BrN_2O_3$: 336.10 (for ^{79}Br), 338.1 (for ^{81}Br); found: 337.1 $[M+H]^+$ (for ^{79}Br), 339.1 $[M+H]^+$ (for ^{81}Br).

5.1.7.3. Synthesis of BrCO6K

Synthesis of *α*-Boc-BrCO6K – Carbamate coupling



Scheme 5 Synthesis of Boc-BrCO6K – Carbamate coupling

Experimental procedure

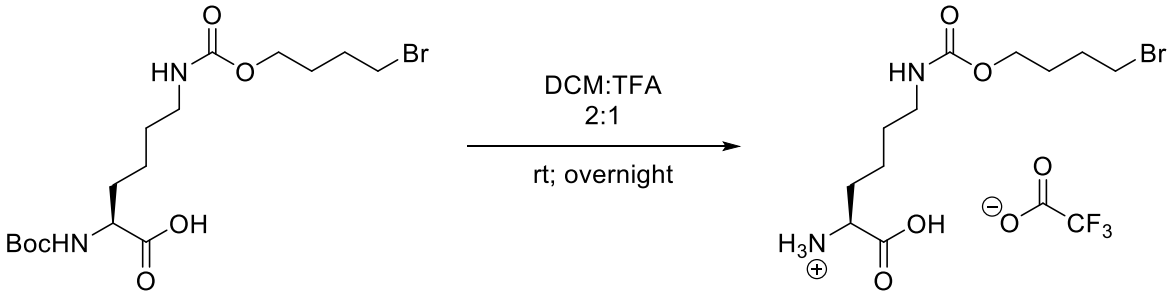
4-Bromobutan-1-ol (1.00 eq.) was dissolved in anhydrous DCM (6.1 mL per mmol 4-Bromobutan-1-ol) and cooled to 0 °C in an ice bath. *N*,*N*′-Disuccinimidyl carbonate (1.50 eq.) and DIPEA (1.50 eq.) were added at 0 °C. The reaction was allowed to warm up to rt and was stirred for 4 h until TLC showed full conversion of the starting material. Then anhydrous THF (6.1 mL per mmol 4-Bromobutan-1-ol) and *α*-Boc-Lys-OH (1.20 eq.) were added and the reaction stirred overnight at rt. Then DCM (7.6 mL per mmol 4-Bromobutan-1-ol) was added and the organic phase washed with 10 % citric acid (2 times; 15.3 mL per mmol 4-Bromobutan-1-ol), brine (15.3 mL per mmol 4-Bromobutan-1-ol) and the combined organic phase dried over Na_2SO_4. The solvent was removed under reduced pressure

and the crude product purified by flash column chromatography (DCM:MeOH = 97:3 + 0.25 % acidic acid). α-BocBrCO6K was isolated in 57 % yield as a light yellow oil.

^{1}H NMR (500 MHz, DMSO-d_6) δ = 7.07 (t, ^3J = 5.7 Hz, 1H), 7.00 (d, ^3J = 7.9 Hz, 1H), 3.95 (t, ^3J = 6.5 Hz, 2H), 3.88 – 3.77 (m, 1H), 3.55 (t, ^3J = 6.6 Hz, 2H), 2.93 (q, ^3J = 6.6 Hz, 2H), 1.84 (p, ^3J = 6.8 Hz, 2H), 1.70 – 1.49 (m, 4H), 1.42 – 1.23 (m, 13H) ppm.

MS(ESI), m/z calc. for $C_{16}H_{29}BrN_2O_6$: 424.12 (for ^{79}Br), 426.12 (for ^{79}Br); found: 447.1 $[M+Na]^+$ (for ^{79}Br), 449.1 $[M+Na]^+$ (for ^{81}Br).

Synthesis of BrCO6K – Boc deprotection



Scheme 6 Synthesis of BrCO6K – Boc deprotection

Experimental procedure

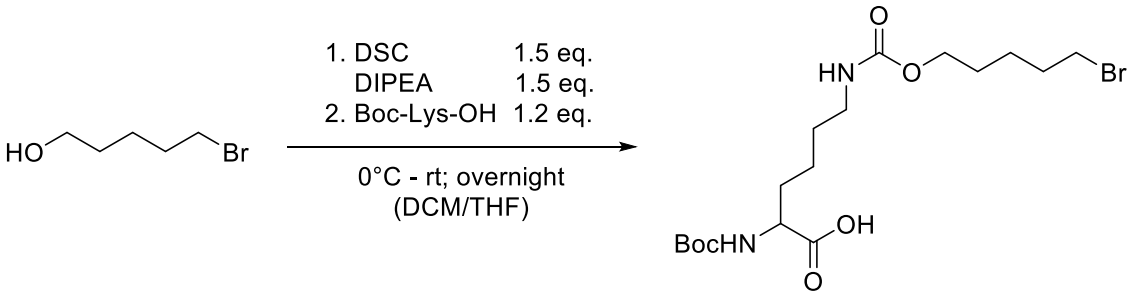
α-Boc-BrCO6K (1.00 eq.) was dissolved in DCM (10.9 mL per mmol α-Boc-BrCO6K) and TFA (10.9 mL per mmol α-Boc-BrCO6K) was added. The reaction was stirred overnight at rt and the solvent was removed under reduced pressure. The crude product was dissolved in methanol (0.5 mL per mmol α-Boc-BrCO6K) and the solution given dropwise to cold diethyl ether (-20 °C, 25 mL per mmol α-Boc-BrCO6K). The suspension was centrifuged at 4200 rpm and 4 °C, the supernatant was discarded. The solid was resuspended and again washed twice with cold diethyl ether. The product was dried on air and BrCO6K was isolated in 72 % yield as a light brown solid. For mammalian cell applications BrCO6K was further purified via HPLC (5 - 80% acetonitrile in water over 10 mins).

^{1}H NMR (500 MHz, DMSO-d6) δ = 8.18 (d, 3J = 3.4 Hz, 3H), 7.05 (br-s, 1H), 3.95 (t, 3J = 6.5 Hz, 2H), 3.87 (q, 3J = 5.8 Hz, 1H), 3.53 (t, 3J = 6.7 Hz, 2H), 2.96 (q, 3J = 5.6 Hz, 2H), 1.84 (p, 3J = 6.9 Hz, 2H), 1.80 – 1.69 (m, 2H), 1.65 (p, 3J = 6.9 Hz, 2H), 1.44 – 1.26 (m, 4H).

MS(ESI), m/z calc. for $C_{11}H_{21}BrN_2O_4$: 324.07 (for ^{79}Br), 326.07 (for ^{81}Br); found: 325.1 $[M+H]^+$ (for ^{79}Br), 327.1 $[M+H]^+$ (for ^{81}Br).

5.1.7.4. Synthesis of BrCO7K

Synthesis of α-Boc-BrCO7K – Carbamate coupling



Scheme 7 Synthesis of Boc-BrCO7K – Carbamate coupling

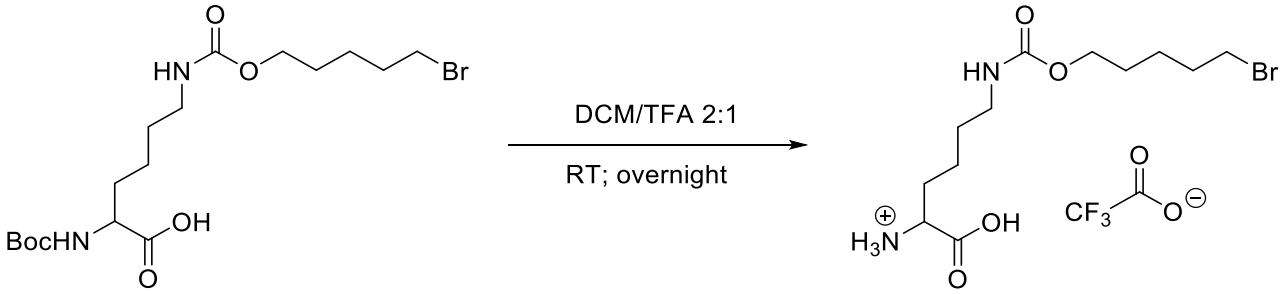
Experimental procedure

5-Bromopentan-1-ol (1.00 eq.) was dissolved in anhydrous DCM (6.7 mL per mmol 5-Bromopentan-1-ol) and cooled to 0 °C using an ice bath. *N,N´*-Disuccinimidyl carbonate (1.50 eq.) and DIPEA (1.50 eq.) were added at 0 °C. The reaction was allowed to warm up to rt and was stirred for 5 h until TLC showed full conversion. Then anhydrous THF (6.7 mL per mmol 5-Bromopentan-1-ol) and *α*-Boc-Lysine-OH (1.20 eq.) were added and the reaction stirred overnight at rt. DCM (8.4 mL per mmol 5-Bromopentan-1-ol) was added, the organic phase was washed with 10 % citric acid (2 times; 16.7 mL per mmol 5-Bromopentan-1-ol) and brine (16.7 mL per mmol 5-Bromopentan-1-ol). The organic phase was dried over Na_2SO_4 and the solvent was removed under reduced pressure. The crude product was purified by flash column chromatography (DCM:MeOH = 96:4 + 0.25% acidic acid) and Boc-BrCO7K was isolated as yellow oil.

^{1}H NMR (300 MHz, MeOD-d_4) δ = 4.10 – 3.99 (m, 3H), 3.45 (t, ^3J = 6.7 Hz, 2H), 3.10 (t, ^3J = 6.7 Hz, 2H), 1.95 – 1.75 (m, 4H), 1.73 – 1.39 (m, 21H) ppm.

MS(ESI), m/z calc. for $C_{17}H_{31}BrN_2O_6$: 438.14 (for ^{79}Br), 440.14 (for ^{81}Br); found: 439.1 $[M+H]^+$ (for ^{79}Br), 441.1 $[M+H]^+$ (for ^{81}Br).

Synthesis of BrCO7K – Boc deprotection



Scheme 8 Synthesis of BrCO7K – Boc deprotection

Experimental procedure

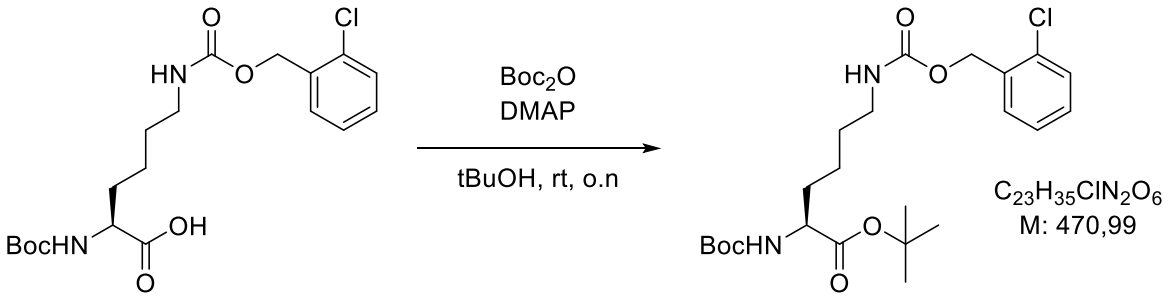
α-Boc-BrCO7K (1.0 eq.) was dissolved in DCM (10.9 mL per mmol *α*-Boc-BrCO7K) and TFA (5.45 mL per mmol *α*-Boc-BrCO7K) was added. The reaction was stirred overnight and the solvent was removed under reduced pressure. The crude product was dissolved in methanol (0.5 mL per mmol *α*-Boc-BrCO7K) and the solution added dropwise to cold diethyl ether (-20 °C, 2 mL per mmol *α*-Boc-BrCO7K). The suspension was centrifuged at 4200 rpm and 4 °C, the supernatant was discarded. The solid was resuspended and again washed twice with cold diethyl ether. The product was dried on air and BrCO7K was isolated in 50 % yield over two steps.

^{1}H NMR (300 MHz, Deuterium Oxide + TFA) δ = 2.93 (m, 3H), 2.34 (t, ^3J = 6.7 Hz, 2H), 2.02 (t, ^3J = 6.5 Hz, 2H), 0.98 – 0.80 (m, 2H), 0.80 – 0.69 (m, 2H), 0.60 – 0.27 (m, 8H).

MS(ESI), m/z calc. for $C_{12}H_{23}BrN_2O_4$: 338.08 (for ^{79}Br), 340.08 (for ^{81}Br); found: 339.1 $[M+H]^+$ (for ^{79}Br), 341.1 $[M+H]^+$ (for ^{81}Br).

5.1.7.5. Synthesis of BrCN7K

tert-butyl esterification of Boc-Lys(2-Cl-Cbz)-OH



Scheme 9 Synthesis of BrCN7K – Esterification of Boc-Lys(2-Cl-Cbz)-OH

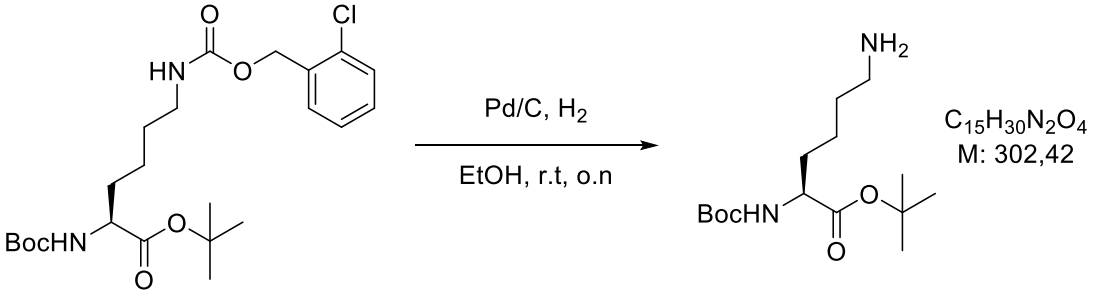
Experimental procedure

Boc-Lys(2-Cl-Cbz)-OH (10.8 g, 26.0 mmol, 1.0 eq.) and DMAP (1.27 g, 10.4 mmol, 0.4 eq.) were dissolved in 105 ml *tert*-butanol and the flask was flushed with argon. Then Boc_2O (7.95 g, 36.4 mmol, 1.4 eq.) was added carefully to the reaction mixture and the reaction was stirred at room temperature overnight. The solvent was removed under reduced pressure and the residue was dissolved in DCM (50 mL). The organic phase was washed with 10 % citric acid (2 x 30 mL) and brine (30 mL). The organic phase was dried over Na_2SO_4 and the solvent was removed under reduced pressure. The crude residue was purified by column chromatography with 2 % MeOH in DCM. The product was dried under vacuum and obtained as a white foam (81%).

^{1}H-NMR (500 MHz, $CDCl_3$, 300 K): δ = 7.44 – 7.40 (m, 1H), 7.40 – 7.34 (m, 1H), 7.26 – 7.23 (m, 2H), 5.21 (s, 2H), 5.05 (d, ^3J = 8.6 Hz, 1H), 4.90 – 4.81 (m, 1H), 4.20 – 4.11 (m, 1H), 3.24 – 3.15 (m, 2H), 1.83 – 1.73 (m, 1H), 1.66 – 1.51 (m, 3H), 1.46 (s, 9H,), 1.43 (s, 9H,), 1.40 – 1.33 (m, 2H, H12) ppm.

MS (ESI+): m/z calc. for $C_{23}H_{35}ClN_2O_6$: 471.2 $[M+H]^+$; found: 471.2 $[M+H]^+$

Hydrogenolysis of 2-Cl-Cbz protection group



Scheme 10 Synthesis of BrCN7K – Hydrogenolysis of 2-Cl-Cbz protection group

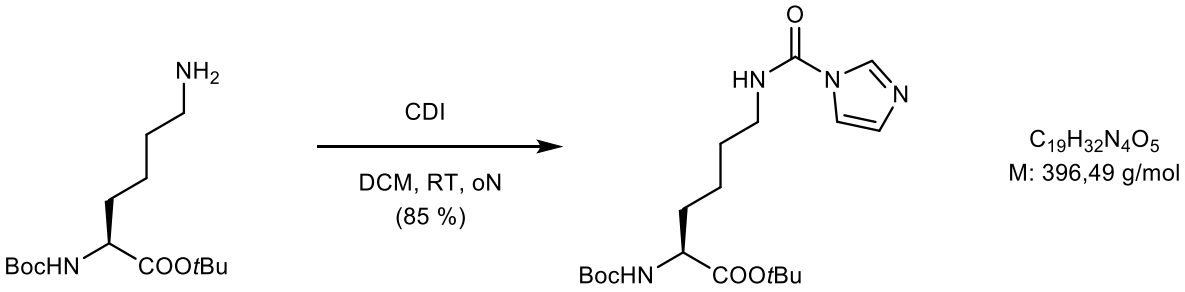
Experimental procedure

Boc-Lys(2-Cl-Cbz)-O*t*Bu (10.2 g, 21.6 mmol, 1.0 eq.) was dissolved in EtOH and palladium on carbon (15 mg/mmol) was added to the solution. The flask was flushed with argon and evacuated three times. Then the flask was flushed with hydrogen and the reaction was stirred overnight at room temperature. The next day the reaction mixture was filtered over Celite and the solvent was removed under reduced pressure. The product was used without further purification.

^{1}H NMR (500 MHz, $CDCl_3$) δ = 5.19 (d, J = 8.7 Hz, 1H), 4.15 – 4.09 (m, 1H), 2.99 (t, J = 7.9 Hz, 2H), 1.89 – 1.68 (m, 4H), 1.67 – 1.52 (m, 2H), 1.45 (s, 9H), 1.42 (s, 9H), 1.42 – 1.40 (m, 2H).

MS (ESI+): m/z calc. for $C_{15}H_{30}N_2O_4$: 303.2 $[M+H]^+$; found: 303.2 $[M+H]^+$

<u>*tert*-butyl N^2-(*tert*-butoxycarbonyl)-N^6-(1*H*-pyrrole-1-carbonyl)-*L*-lysinate</u>



Scheme 11 Synthesis of BrCN7K – Formation of *tert*-butyl N^2-(tert-butoxycarbonyl)-N^6-(1*H*-pyrrole-1-carbonyl)-*L*-lysinate

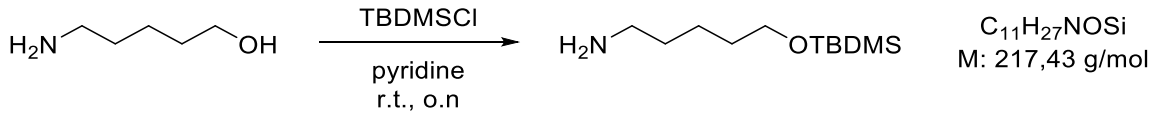
Experimental procedure

Boc-Lys-O*t*Bu (3.27 g, 10.8 mmol, 1.0 eq.) was dissolved in DCM and the solution was added over 1 hour to the solution of CDI (3.5 g, 21.6 mmol, 2.0 eq.) in DCM at 0 °C. During this time, the initially white suspension gets clear. The stirring was continued overnight at room temperature, after which the mixture was washed with brine and the separated aqueous phase extracted twice with DCM. The combined organic phases were dried over Na_2SO_4 and concentrated under reduced pressure. Purification by column chromatography on silica, packed in DCM and eluted with 2 → 4 % MeOH in DCM yielded the product as white foam (85%).

^{1}H-NMR (300 MHz, $CDCl_3$, 300 K): δ = 8.18 (s, 1H), 7.45 (s, 1H), 7.07 (dd, J = 1.6 Hz, 0.9 Hz), 6.41 (br s, 1H), 5.18 (d, J = 6.9 Hz, 1H), 4.22 – 4.11 (m, 1H), 3.55 – 3.32 (m, 2H), 1.87-1.48 (m, 2H), 1.66 – 1.50 (m, 2H), 1.46 (s, 9H), 1.42 (s, 9H), 1.52– 1.38 (m, 2H) ppm.

MS (ESI+): m/z calc. for $C_{19}H_{32}N_4O_5$: 397.2 $[M+H]^+$; found: 397.2 $[M+H]^+$

<u>TBDMS protection of 5-amino-1-pentanol</u>



Scheme 12 Synthesis of BrCN7K – TBDMS protection of alcohols

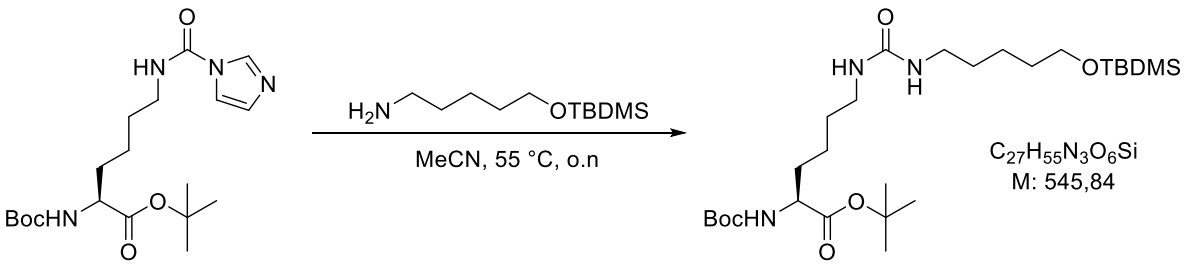
Experimental procedure

TBDMSCl (5.26 g, 34.9 mmol, 1.2 eq.) was added to a solution of 5-amino-1-pentanol (3.0 g, 29.1 mmol, 1.0 eq.) in pyridine and the mixture was stirred overnight at room temperature. The mixture was then concentrated under reduced pressure and the obtained residue was dissolved in DCM. The DCM phase was washed with saturated $NaHCO_3$ solution, dried over Na_2SO_4 and concentrated *in vacuo*. Purification of the crude product by column chromatography on silica (packed in 10 % MeOH in DCM and eluted with 2 % Et_3N/10 % MeOH/DCM, v/v) afforded the desired product as yellow liquid (94 %).

^{1}H NMR (300 MHz, $CDCl_3$) δ = 3.58 (t, J = 6.4 Hz, 2H), 2.75-2.63 (m, 2H), 2.31 (s, 2H), 1.58-1.26 (m, 6H), 0.86 (s, 9H), 0.01 (s, 6H) ppm.

MS (ESI +): m/z calc. for $C_{11}H_{27}NOSi$: 218.2 $[M+H]^+$; found: 218.2 $[M+H]^+$

Urea bond formation



Scheme 13 Synthesis of BrCN7K – Urea bond formation

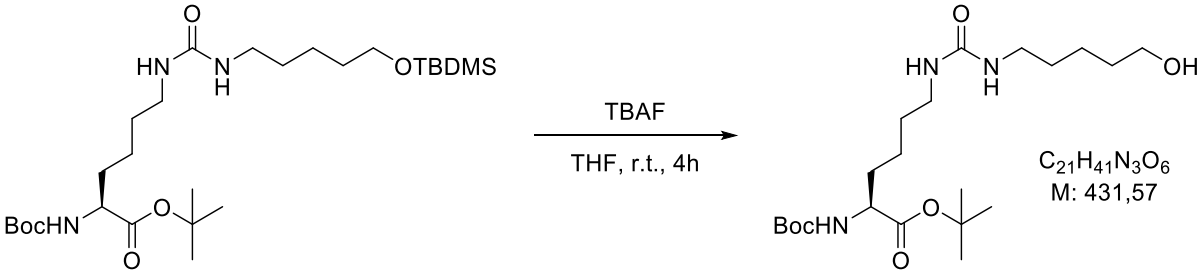
Experimental procedure

TBDMS-protected 5-amino-1-pentanol (1.58 g, 7.25 mmol, 1.35 eq.) and *tert*-butyl N^2-(*tert*-butoxycarbonyl)-N^6-(1*H*-pyrrole-1-carbonyl)-*L*-lysinate (2.09 g, 5.30 mmol, 1.0 eq.) were dissolved in 32 ml MeCN and stirred for 24 hours at 55 °C. Upon completion, as judged by TLC (4 % MeOH in DCM), the mixture was diluted with 50 ml DCM. The organic phase was washed with 1 M HCl (2 x 40 mL) and saturated $NaHCO_3$ solution (40 ml). The combined organic phases were dried over Na_2SO_4, filtered, and the solvent was removed under reduced pressure. The crude product was purified by flash column chromatography, packed in 2 % MeOH in DCM and eluted with 2 → 4 % MeOH in DCM. This yielded the product as pale yellow oil (96 %).

^{1}H NMR (300 MHz, $CDCl_3$) δ = 5.12 (d, J = 7.6 Hz, 1H), 4.58 (t, J = 5.6 Hz, 1H), 4.50 (t, J = 5.6 Hz, 1H), 4.18-4.06 (m, 1H), 3.58 (t, J = 6.4 Hz, 2H), 3.25-3.05 (m, 4H), 1.88-1.45 (m, 8H), 1.45 (s, 9H), 1.43 (s, 9H), 1.41-1.30 (m, 4H), 0.87 (s, 9H), 0.03 (s, 6H) ppm.

MS (ESI +): m/z calc. for $C_{27}H_{55}N_3O_6Si$: 546.4 $[M+H]^+$; found: 546.4 $[M+H]^+$

TBDMS deprotection



Scheme 14 Synthesis of BrCN7K – TBDMS deprotection

Experimental procedure

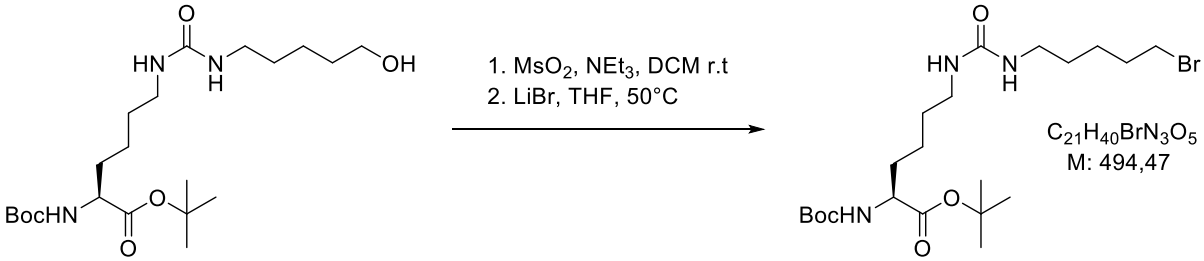
tert-butyl N^2-(*tert*-butoxycarbonyl)-N^6-((5-((*tert*-butyldimethylsilyl)oxy)pentyl)carbamo-yl)-*L*-lysinate (2.38 g, 4.35 mmol, 1.0 eq.) was dissolved in 44 ml dry THF. Then a 1 M TBAF solution in THF (5.05 ml, 1.16 eq.) was added and the reaction was stirred for 4 hours at room temperature. After completion of the reaction a, judged by TLC (5 % MeOH in DCM), the solvent was removed under reduced pressure. The remaining residue was dissolved in 40 ml DCM and the organic phase was washed with 30 ml 10 % citric acid, and 30 ml saturated $NaHCO_3$ solution, dried over Na_2SO_4, filtered, and the solvent was removed under reduced

pressure. The crude product was purified by flash column chromatography, packed in 2 % MeOH in DCM and eluted with 4 → 6 % MeOH in DCM. *tert*-butyl N^2-(*tert*-butoxycarbonyl)-N^6-(5-hydroxypentyl)carbamoyl)-*L*-lysinate was obtained as pale yellow oil (93 %).

^{1}H NMR (300 MHz, $CDCl_3$) δ = 5.15 (d, J = 8.3 Hz, 1H), 4.50 – 4.30 (m, 2H), 4.17 – 4.07 (m, 1H), 3.41 (t, J = 6.3 Hz, 2H), 3.24 – 3.08 (m, 4H), 1.88 (p, J = 7.0 Hz, 2H), 1.80 – 1.70 (m, 1H), 1.69 – 1.48 (m, 10 H), 1.46 (s, 9H), 1.44 (s, 9H), 1.44 – 1.32 (m, 2H) ppm.

MS (ESI+): m/z calc. for $C_{21}H_{41}N_3O_6$: 432.5 $[M+H]^+$; found: 432.3 $[M+H]^+$

Formation of protected N^6-((5-bromopentyl)carbamoyl)-*L*-lysine (BrCN7K)



Scheme 15 Synthesis of BrCN7K – Formation of protected N^6-((5-bromopentyl)carbamoyl)-*L*-lysine

Experimental procedure

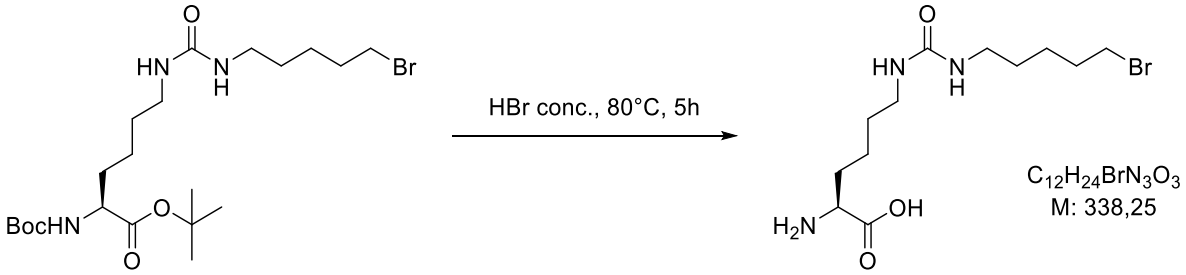
tert-butyl N^2-(*tert*-butoxycarbonyl)-N^6-(5-hydroxypentyl)carbamoyl)-*L*-lysinate (1.80 g, 4.16 mmol, 1.0 eq.) was dissolved in 15 ml DCM, then NEt_3 was added to the reaction mixture and it was cooled down to 0 °C. Methanesulfonic anhydride (2.18 g, 12.5 mmol, 3.0 eq.) in 7 ml DCM was added carefully to the reaction mixture and the mixture was stirred for 2 hours at room temperature. After completion the reaction mixture was washed with water (2 x 20 ml), the organic phase dried over Na_2SO_4, filtered and the solvent was removed under reduced pressure.

The remaining residue was dissolved in 21 ml dry THF, LiBr (2.17 g, 25.0 mmol, 6.0 eq.) was added and the reaction was stirred at 50 °C overnight. The next day the solvent of the reaction mixture was removed, the residue dissolved in 40 ml EtOAc, the organic phase was washed with water (3 x 20 ml) dried over Na_2SO_4, filtered and the solvent was removed under reduced pressure. The crude product was purified by flash column chromatography, packed in 2 % MeOH in DCM and eluted with 2 → 5 % MeOH in DCM. This way the product was isolated as a colourless oil (90%).

^{1}H NMR (300 MHz, $CDCl_3$) δ = 5.11 (d, J = 8.0 Hz, 1H), 4.48 (br-s, 1H), 4.22 – 4.05 (m, 1H), 3.40 (t, J = 6.7 Hz, 2H), 3.29 – 3.03 (m, 4H), 1.87 (p, J = 6.9 Hz, 2H), 1.81 – 1.69 (m, 2H), 1.68 – 1.59 (m, 2H), 1.58 – 1.49 (m, 4H), 1.46 (s, 9H), 1.44 (s, 9H), 1.42 – 1.31 (m, 2H) ppm.

MS (ESI+): m/z calc. for $C_{21}H_{40}BrN_3O_5$: 494.4 $[M+H]^+$; found: 494.1 $[M+H]^+$

Boc and *tert*-butyl deprotection of BrCN7K



Scheme 16 Synthesis of BrCN7K – Boc and *tert*-butyl deprotection

Experimental procedure

Protected BrCN7K (3.33 g) was dissolved 20 mL conc. HBr and stirred at 80 °C for 5 hours. The solvent was removed under reduced pressure and the crude product purified via HPLC (5-80% acetonitrile in water over 10 mins). This way the product was isolated as a white solid. The UAA was stored at -20 °C and when needed prepared as 100 mM stock solution in 0.2 M NaOH.

^{1}H NMR (400 MHz, DMSO) δ = 8.23 (br-s, 3H), 3.88 – 3.80 (m, 1H), 3.49 (t, J = 6.7 Hz, 2H), 2.94 (t, J = 6.3 Hz, 4H), 1.85 – 1.68 (m, 4H), 1.46 – 1.18 (m, 8H).

^{13}C NMR (101 MHz, DMSO) δ = 171.3, 158.7, 52.3, 39.4, 39.2, 35.5, 32.2, 30.0, 29.8, 29.3, 25.2, 22.0.

MS (ESI+): m/z calc. for $C_{12}H_{24}BrN_3O_3$: 338.2 $[M+H]^+$; found: 338.1 $[M+H]^+$ (for ^{79}Br), 340.1 (for ^{81}Br) $[M+H]^+$

5.1.8. Decaging of pDAP

Decaging of the pDAP triggered by UV light at 365 nm performed with 1 mM pDAP in MeOH. The sample was irradiated over different time intervals during it was cooled on ice. The photocage release was monitored via ESI-MS.

5.1.9. Decaging of Boc-HpGGK

Samples of 1 mM Boc-HpGGK in PBS were irradiated with UV light at 365 nm. During the irradiation the sample was stirred and cooled on ice. At certain time points samples were analyzed using ESI-MS.

5.1.10. Stability of TM-439 and BrCN7K towards GSH

For testing the stability of TM-439 and BrCN7K towards GSH samples with 1 mM of each uAA in PBS were incubated at 37 °C supplemented with 10 mM GSH. The reaction over time was monitored using ESI-MS.

5.2. Biochemical Work

5.2.1. Material

For microbiological experiments all utilized glassware, equipment, medium, solutions, stocks and materials were sterilized before use. Disposables were already purchased sterile. Media, solutions, stocks and glassware were sterilized. Compounds which are heat stable and glassware were autoclaved at 121 °C for 20 min at 2 bar. Temperature sensitive compounds were sterile filtered through a membrane of 0.2 µm pore size.

5.2.1.1. Organisms

Table 7 Bacterial strains

Organism	Genotype	Use	Origin
Escherichia coli K12 TB1	F$^-$ *ara* Δ*(lac-proAB) [Φ80dlac* Δ*(lacZ)M15] rpsL*(Str^R) *thi hsdR*	Transformation, Expression	NEB (Ipswich, USA)
Escherichia coli NEB 10-beta	Δ*(ara-leu) 7697 araD139 fhuA* Δ*lacX74 galK16 galE15 e14-*ϕ*80*d*lacZ*Δ*M15 recA1 relA1 endA1 nupG rpsL* (StrR) *rph spoT1* Δ*(mrr-hsdRMS-mcrBC)* DH10B derivative	Cloning, Transformation, Expression	NEB (Ipswich, USA)
Escherichia coli NEB 10-beta ΔcobB	Δ*(ara-leu) 7697 araD139 fhuA* Δ*lacX74 galK16 galE15 e14-*ϕ*80*d*lacZ*Δ*M15 recA1 relA1 endA1 nupG rpsL* (StrR) *rph spoT1* Δ*(mrr-hsdRMS-mcrBC)* Δ*cobB*::*kan* (Kan^R) DH10B derivative	Transformation, Expression	Lang Lab, TUM (Garching, Germany)
Escherichia coli MegaX DH10B T1R	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC), Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ$^-$ *rps*L *nup*G *ton*A	Cloning, Transformation, Expression	Invitrogen (Thermo Fisher Scientific, Waltham, USA)
Escherichia coli BL21	*fhuA2 [lon] ompT gal [dcm] ΔhsdS*	Transformation, Expression	NEB (Ipswich, USA)
Escherichia coli Rosetta (DE3)	F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE (Cam^R) BL21 (DE3) derivative	Transformation, T7-Expression	Novagen (Merck Group, Darmstadt, Germany)
Escherichia coli MB6565 (ΔmetB; ΔmsrA; ΔmsrB; ΔmsrC; ΔbisC; ΔyedY)	F- *[araD139]$_{B/r}$ Δ(argF-lac)169 λ$^-$ e14-flhD5301 Δ(fruK-yeiR)725(fruA25) relA1 rpsL150(strR) rbsR22 Δ(fimB-fimE)632(::IS1)deoC1 ΔmetB ΔmsrA::spec (SpecR) ΔmsrB ΔmsrC ΔbisC ΔyedY*	Transformation, Expression	Mehmet Berkmen & Bruno Manta (Ipswich, USA)
Escherichia coli KEIO JW4178 (ΔmsrA)	F-, *Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, rph-1, Δ(rhaD-rhaB)568, hsdR514, msrA::kan* (KanR) BW25113 derivative	Transformation, Expression	Jung Lab, LMU (Munich, Germany)

Escherichia coli KEIO JW1767 (ΔmsrB)	F-, *Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, rph-1, Δ(rhaD-rhaB)568, hsdR514,msrB::kan* (KanR) BW25113 derivative	Transformation, Expression	Jung Lab, LMU (Munich, Germany)
Escherichia coli KEIO JW1821 (ΔmsrC)	F-, *Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, rph-1, Δ(rhaD-rhaB)568, hsdR514, msrC::kan* (KanR) BW25113 derivative	Transformation, Expression	Jung Lab, LMU (Munich, Germany)

Table 8 Mammalian cells

Cell line	Organism	Specification	Origin
HEK293T	*Homo sapiens, human*	Permanent cell line derived from embryonic human kidney by transformation with sheared human adenovirus type 5 DNA. Express a version of the SV40 Large T antigen. Epithelial-like cells	Feige Lab, TUM (Garching, Germany)

5.2.1.2. Plasmids

A detailed information of each plasmid backbone and cloning procedure can be found in the appendix 7.1.1 and Table 64.

Table 9 Synthetase-bearing plasmids for bacterial cell experiments

Code	Plasmid	Antibiotic	Origin
MW-S26	pBK_Mb-PylRS (D4)	Kan	Lang Lab (TUM)
MW-S34 Mb-TEMPOH-I	pBK_Mb-PylRS-Y271M-L274A-C313A (Mb-TEMPOH-I) (BrCnKRS)	Kan	Lang Lab (TUM)
Mb-MF3	pBK_Mb_PylRS-Y271G-C313V (MF3)	Amp	Lang Lab (TUM)
Mb-D79	pBK_Mb-PylRS-C313V	Kan	Lang Lab (TUM)
Mb-SM29	pBK_Mb-PylRS-L271A-C313V-Y349F	Kan	Lang Lab (TUM)
Mb-SM30	pBK_Mb-PylRS-L271A-C313S-Y349F	Kan	Lang Lab (TUM)
Mb-MF9	pBK_Mb-PylRS-L266M-L270I-Y271F-L274A-C313F (MF9)	Amp	Lang Lab (TUM)
MW-S7 Mb-MF18	pBK_Mb-PylRS-L274A-N311Q-C313S (MF18)	Amp	Lang Lab (TUM)
MW-S8	pBK_Mb-MF18(A274G, N311Q, C313S)-A274G	Amp	Site-directed Mutagenesis
MW-S9	pBK_Mb-MF18(L274A, N311Q, C313S)-M315G	Amp	Site-directed Mutagenesis
MW-S10	pBK_Mb-MF18(L274A, N311Q, C313S)-I378A	Amp	Site-directed Mutagenesis
MW-S11	pBK_Mb-MF18(L274A, N311Q, C313S)-S313G	Amp	Site-directed Mutagenesis

MW-S12	pBK_Mb-MF18(L274A, N311Q, C313S)-I378N	Amp	Site-directed Mutagenesis
MW-S13	pBK_Mb-MF18(L274A, N311Q, C313S)-I378Q	Amp	Site-directed Mutagenesis
MW-S14	pBK_Mb-MF18(L274A, N311Q, C313S)-M315C	Amp	Site-directed Mutagenesis
MW-S15	pBK_Mb-MF18(L271A, N311Q, C313S)-A274G-M315G	Amp	Site-directed Mutagenesis
MW-S16	pBK_Mb-MF18(L274A, N311Q, C313S)-A274G-I378A	Amp	Site-directed Mutagenesis
MW-S17	pBK_Mb-MF18(L274A, N311Q, C313S)-M315C-I378Q	Amp	Site-directed Mutagenesis
MW-S18	pBK_Mb-MF18(L274A, N311Q, C313S)-M315C-I378N	Amp	Site-directed Mutagenesis
MW-S19	pBK_Mb-MF18(L274A, N311Q, C313S)-A274G-M315C-I378A	Amp	Site-directed Mutagenesis
MW-S20	pBK_Mb-MF18(L274A, N311Q, C313S)-A274G-M315G-I378A	Amp	Site-directed Mutagenesis
MW-S21	pBK_Mb-MF18(L274A, N311Q, C313S)-A274G-M315G-I287A-I378A	Amp	Site-directed Mutagenesis
MW-S45 pDAPRS	pBK_Mb-PylRS-Y271C-N311Q-Y349F-V366C	Kan	Site-directed Mutagenesis
MW-S59	pEVOL_Mm-PylRSwt-PylT	Cam	Restriction Cloning
MW-S60 Mm-TEMPOH-I	pEVOL_Mm-PylRS-Y306M-L309A-C348A-PylT (Mm-TEMPOH-I) (BrCnKRS)	Cam	Restriction Cloning
MW-Ma6 Ma-PylRS	pBK_Ma-PylRSwt	Amp	Site-directed Mutagenesis
MW-Ma7 Ma-MF3	pBK_Ma-PylRS-MF3 (Y126G)	Amp	Lang Lab (TUM)
MW-Ma8 Ma-MF9	pBK_Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)	Amp	Lang Lab (TUM)
MW-Ma9 Ma-MF18	pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)	Amp	Site-directed Mutagenesis
MW-Ma10	pBK_Ma-PylRS-L121M-L125I-Y126F-M129A	Amp	Lang Lab (TUM)
MW-Ma11	pBK_Ma_PylRS-M129A	Amp	Lang Lab (TUM)
MW-Ma12	pBK_Ma-PylRS-H227I-Y228P	Amp	Site-directed Mutagenesis
MW-Ma13	pBK_Ma-PylRS-MF3 (Y126G)-H227I-Y228P	Amp	Site-directed Mutagenesis
MW-Ma14	pBK_Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)-H227I-Y228P	Amp	Site-directed Mutagenesis
MW-Ma15	pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)-H227I-Y228P	Amp	Site-directed Mutagenesis

MW-Ma16	pBK_Ma-PylRSwt-H227A	Amp	Site-directed Mutagenesis
MW-Ma17	pBK_Ma-PylRS-MF3 (Y126G)-H227A	Amp	Site-directed Mutagenesis
MW-Ma18	pBK_Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)-H227A	Amp	Site-directed Mutagenesis
MW-Ma19	pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)-H227A	Amp	Site-directed Mutagenesis
MW-Ma1	pET28_His-Throm-Ma-PylRSwt	Kan	Restriction Cloning
MW-Ma2	pET28_His-Throm-Ma-PylRS-MF3 (Y126G)	Kan	Restriction Cloning
MW-Ma3	pET28_His-Throm-Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)	Kan	Restriction Cloning
MW-Ma4	pET28_His-Throm-Ma-PylRS-MF18 (M129A, N166Q, V168S)	Kan	Restriction Cloning
MW-Ma5	pET28_His-Throm-Ma-PylRS-MF18 (M129A, N166Q, V168S)-H227I-Y228P	Kan	Restriction Cloning
MW-S4	pET28_His-Throm-Mm-PylRS-MF18(L309A, N346Q, C348S)(185-454)	Kan	Lang Lab (TUM)

Table 10 Rab1b/DrrA-bearing plasmids for bacterial cell experiments

Code	Plasmid	Antibiotic	Origin
TM10	pPylT Rab1b-His6 WT	Tet	Lang Lab (TUM)
TM11	pPylT Rab1b-His6 K10TAG	Tet	Lang Lab (TUM)
TM12	pPylT Rab1b-His6 D16TAG	Tet	Lang Lab (TUM)
TM13	pPylT Rab1b-His6 S17TAG	Tet	Lang Lab (TUM)
TM14	pPylT Rab1b-His6 T72TAG	Tet	Lang Lab (TUM)
TM15	pPylT Rab1b-His6 R79TAG	Tet	Lang Lab (TUM)
TM16	pPylT Rab1b-His6 N98TAG	Tet	Lang Lab (TUM)
TM17	pPylT Rab1b-His6 Y37TAG	Tet	Lang Lab (TUM)
TM18	pPylT Rab1b-His6 I47TAG	Tet	Lang Lab (TUM)
TM19	pPylT Rab1b-His6 R69TAG	Tet	Lang Lab (TUM)
MW-P41	pBAD_Rab1b-Q67A-His6	Amp	SLIM
MW-P42	pBAD_Rab1b-Q67A-R69TAG-His6	Amp	Site-directed Mutagenesis
MW-P43	pBAD_Rab1b-Q67A-T72TAG-His6	Amp	Site-directed Mutagenesis
MW-P46	pBAD_Rab1b-Q67A-D44TAG-His6	Amp	Site-directed Mutagenesis
MW-P48	pBAD_Rab1b-Q67A-D31TAG-His6	Amp	Site-directed Mutagenesis
MW-P54	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)	Kan	DNA Assembly
MW-P55	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)-D82C	Kan	Site-directed Mutagenesis
MW-P56	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)-D82A	Kan	Site-directed Mutagenesis

MW-P57	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)-D22C	Kan	Site-directed Mutagenesis
MW-P58	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)-D25C	Kan	Site-directed Mutagenesis
MW-P59	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)-R24C	Kan	Site-directed Mutagenesis
MW-P93	pBAD-RSF1031K_StrepII-Ent-DrrA(1-647)-N451A-R453A-D480A-S483A	Kan	DNA Assembly
MW-P94	pBAD-RSF1031K_StrepII-Ent-DrrA(1-647)-N451A-R453A-D480A-S483A-D82C	Kan	Site-directed Mutagenesis
MW-P60	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)	Kan	Site-directed Mutagenesis
MW-P62	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-D82C	Kan	Site-directed Mutagenesis
MW-P61	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-Q71C	Kan	Site-directed Mutagenesis
MW-P65	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-I123TAG	Kan	Site-directed Mutagenesis
MW-P68	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-D131TAG	Kan	Site-directed Mutagenesis
MW-P70	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-Q138TAG	Kan	Site-directed Mutagenesis
MW-P75	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-E179TAG	Kan	Site-directed Mutagenesis
MW-P76	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-A180TAG	Kan	Site-directed Mutagenesis
MW-P77	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-S184TAG	Kan	Site-directed Mutagenesis
MW-P78	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-K241TAG	Kan	Site-directed Mutagenesis
MW-P79	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-E316TAG	Kan	Site-directed Mutagenesis
MW-P88	pBAD-RSF1031K_DrrA(16-533)-TEV-His6 (N451A-R453A-D480A-S483A)	Kan	DNA Assembly
MW-P89	pBAD-RSF1031K_DrrA(16-533)-D82C-TEV-His6 (N451A-R453A-D480A-S483A)	Kan	Site-directed Mutagenesis
MW-P92	pBAD-RSF1031K_DrrA(16-533)-D82C-D512C-TEV-His6 (N451A-R453A-D480A-S483A)	Kan	Site-directed Mutagenesis
MW-P100	pBAD-Duet_Rab1b-Q67A-His6-StrepII-Ent-DrrA(1-339)	Amp	DNA Assembly
MW-P101	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)	Amp	Site-directed Mutagenesis
MW-P102	pBAD-Duet_Rab1b-Q67A-T72TAG-His6-StrepII-Ent-DrrA(1-339)	Amp	Site-directed Mutagenesis

MW-P104	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)-D82A	Amp	Site-directed Mutagenesis
MW-P105	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)-D82C	Amp	Site-directed Mutagenesis
MW-P106	pBAD-Duet_Rab1b-Q67A-T72TAG-His6-StrepII-Ent-DrrA(1-339)-D82C	Amp	Site-directed Mutagenesis
MW-P107	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)-E38A	Amp	Site-directed Mutagenesis
MW-P108	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)-E38C	Amp	Site-directed Mutagenesis
MW-P109	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)-D165C	Amp	Site-directed Mutagenesis
MW-P110	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)-D172C	Amp	Site-directed Mutagenesis
MW-P111	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)-D165A-D172A	Amp	SLIM
MW-P123	pBAD-Duet_Rab1b-Q67A-His6-DrrA(16-352)	Amp	DNA Assembly
MW-P130	pBAD-Duet_Rab1b-Q67A-D31TAG-His6-DrrA(16-352)	Amp	Site-directed Mutagenesis
MW-P131	pBAD-Duet_Rab1b-Q67A-D31TAG-His6-DrrA(16-352)-D131A	Amp	Site-directed Mutagenesis
MW-P132	pBAD-Duet_Rab1b-Q67A-D31TAG-His6-DrrA(16-352)-E58A	Amp	Site-directed Mutagenesis
MW-P133	pBAD-Duet_Rab1b-Q67A-D31TAG-His6-DrrA(16-352)-E280A	Amp	Site-directed Mutagenesis
MW-P124	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-DrrA(16-352)	Amp	Site-directed Mutagenesis
MW-P125	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-DrrA(16-352)-D82C	Amp	Site-directed Mutagenesis
MW-P144	pBAD-Duet_Rab1b-Q67A-His6-StrepII-TEV-DrrA(16-352)	Amp	DNA Assembly
MW-P145	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)	Amp	Site-directed Mutagenesis
MW-P150	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82C	Amp	Site-directed Mutagenesis
MW-P151	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A	Amp	Site-directed Mutagenesis
MW-P154	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A-Q71C	Amp	Site-directed Mutagenesis
MW-P155	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A-Y78C	Amp	Site-directed Mutagenesis
MW-P156	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A-E75C	Amp	Site-directed Mutagenesis

MW-P157	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A-D79C	Amp	Site-directed Mutagenesis
MW-P158	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A-E87C	Amp	Site-directed Mutagenesis
MW-P159	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A-M88C	Amp	Site-directed Mutagenesis
MW-P164	pBAD-Duet_Rab1b-Q67A-His6-StrepII-Ent-DrrA(1-647)(N451A-R453A-D480A-S483A)	Amp	DNA Assembly
MW-P165	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-647)(N451A-R453A-D480A-S483A)	Amp	Site-directed Mutagenesis
MW-P166	pBAD-Duet_Rab1b-Q67A-T72TAG-His6-StrepII-Ent-DrrA(1-647)(N451A-R453A-D480A-S483A)	Amp	Site-directed Mutagenesis
MW-P167	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-647)-D82C (N451A-R453A-D480A-S483A)	Amp	Site-directed Mutagenesis
MW-P160	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-DrrA(16-533)-TEV-His6 (N451A-R453A-D480A-S483A)	Amp	Site-directed Mutagenesis
MW-P161	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-DrrA(16-533)-D82C-TEV-His6 (N451A-R453A-D480A-S483A)	Amp	Site-directed Mutagenesis
MW-P162	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-DrrA(16-533)-D82C (N451A-R453A-D480A-S483A)	Amp	Site-directed Mutagenesis

Table 11 *Sa*ClpP-bearing plasmids for bacterial cell experiments

Code	Plasmid	Antibiotic	Origin
MW-P1	pBAD_*Sa*ClpPwt-StrepII	Amp	Sieber Lab (TUM)
MW-P6	pBAD_*Sa*ClpP-S98A-StrepII	Amp	Site-directed Mutagenesis
MW-P8	pPylT_*Sa*ClpP-S98TAG-StrepII	Tet	Site-directed Mutagenesis
MW-P11	pPylT_*Sa*ClpPwt-His6	Tet	DNA Assembly
MW-P12	pPylT_*Sa*ClpP-S98A-His6	Tet	Site-directed Mutagenesis
MW-P14	pPylT_*Sa*ClpP-S98TAG-His6	Tet	Site-directed Mutagenesis

Table 12 Further plasmids for bacterial cell experiments

Code	Plasmid	Antibiotic	Origin	Comment
MW63	pBAD_sfGFPwt-His6	Amp	Lang Lab (TUM)	Reporter
MW64	pBAD_sfGFP-N150TAG-His6	Amp	Lang Lab (TUM)	Reporter Amber Suppression
MW65	pBAD_sfGFP-V2P-N150TAG-His6	Amp	Site-directed Mutagenesis	Reporter Amber Suppression
MW71	pPylT_sfGFP-N150TAG-His6	Tet	Lang Lab (TUM)	Reporter Amber Suppression
MW72	pPylT_sfGFP-N150TAG-E173C-His6	Tet	Site-directed Mutagenesis	Reporter Amber Suppression
MW78	pPylT_sfGFP-V2P-N150TAG-His6	Tet	Lang Lab (TUM)	Reporter Amber Suppression

MW79	pPylT_sfGFP-V2P-N150TAG-His6-MaPylT	Tet	Restriction Cloning	Reporter Amber Suppression
MW80	pPylT_sfGFP-V2P-N150TAG-E173C-His6	Tet	Site-directed Mutagenesis	Reporter Amber Suppression
MW61	pRed/ET-plasmid (thermosensitive)	Amp	Gene Gene Bridges GmbH (Heidelberg, Germany)	Knock-out strain
MW62	pFLPe-709 (thermosensitive)	Amp	Gene Gene Bridges GmbH (Heidelberg, Germany)	Knock-out strain
MW59	pBAD-RSF1031K_Gro7	Kan		Expression Chaperone
MW-E4	*Mb*-D65	Tet	Lang Lab (TUM)	Directed Evolution
MW-E5	*Mb*-D7L	Cam	Lang Lab (TUM)	Directed Evolution
MW-E6	*M*b-pPylT_sfGFP150TAG CAT	Tet	Lang Lab (TUM)	Directed Evolution
MW-E1	*Ma*-D65	Tet	Restriction Cloning	Directed Evolution
MW-E2	*Ma*-D7L	Cam	Restriction Cloning	Directed Evolution
MW-E3	*Ma*-pPylT_sfGFP150TAG CAT	Tet	Restriction Cloning	Directed Evolution

Table 13 Plasmids for mammalian cell experiments

Code	**Plasmid**	**Antiobiotic**	**Origin**	**Comment**
MM1	SE323_Mm-PylRSwt_4xMmPylT	Amp	Lang Lab (TUM)	Amber Suppression
MM2	SE323_Mm-TEMPOH-I_4xMmPylT (Y306M, L309A, C348A)	Amp	Lang Lab (TUM)	Amber Suppression
MM9	SE348_mCherry-TAG-GFP-4xMmPylT	Amp	Lang Lab (TUM)	Reporter (Microscopy)
MM10	SE362_sfGFP-N150TAG-6His-4xMmPylT	Amp	Lang Lab (TUM)	Reporter ESI-MS
MM19	SE348_Rab1b-Q67A-R69TAG-His6_4xMmPylT	Amp	Restriction Cloning	Amber Suppression
MM25	pAcGFP-DrrA(8-533)-R70A-Q71A-K74A	Kan	Itzen Lab (UKE)	Expression
MM26	pAcGFP-DrrA(8-533)-D110A-D112A	Kan	Itzen Lab (UKE)	Expression
MM27	pAcGFP-DrrA(8-533)-D110A-D112A-D82A	Kan	Site-directed Mutagenesis	Expression
MM33	pAcGFP-DrrA(8-533)-D110A-D112A-D82C	Kan	Site-directed Mutagenesis	Expression
MM35	pAcGFP-DrrA(8-533)-R70A-Q71A-K74A-D82A	Kan	Site-directed Mutagenesis	Expression
MM36	pAcGFP-DrrA(8-533)-R70A-Q71A-K74A-D82C	Kan	Site-directed Mutagenesis	Expression

MM30	pPB1_Ma-PylRSwt_4xMaPylT	Amp	Chin Lab (MRC)	Amber Suppression
MM31	pPB2_sfGFP-His6_4xMaPylT	Amp	Chin Lab (MRC)	Reporter ESI-MS
MM32	pPB3_mCherry-TAG-GFP_4xMaPylT	Amp	Restriction Cloning	Reporter (Microscopy)

5.2.1.3. Gene-fragments

The gene string (*Ma*-PylT-cassette) was obtained from Genewiz (Leipzig, Germany).
DNA-Sequenz (5'→3'):

TGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTG
GCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGC
CATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATC
ATTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCC
TATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCG
CTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAA
CATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCTCCTTGCATGCACCATTCCTTGCGGCGGCGG
TGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCTGGCGAAAGGGGGATG
TGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTTAAA
AAAAATCCTTAGCTTTCGCTAAGgATCTGCAGTGGCGAGAGACCGGGGTGTCGAACCCCGCTATGCTAGGTTTTAGAGACCCG
CTGGTCGCCGGACCGTCCCCCGCGGCCGCGAATTCAGCGTTACAAGTATTACACAAAGTTTTTTATGTTGAGAATATTTTTTTG
ATGGGGCGCCACTTATTTTTGATCGTTCGCTCAAAGAAGCGGCGCCAGGGTTGTTTTTCTTTTCACCAGTGAGACGGGCAACA
GAACGCCATGAGCGGCCTCATTTCTTATTCTGAGTTACAACAGTCCGCACCGCTGCCGGTAGCTCCTTCCGGTGGGCGCGGGG
CATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTATTAACGAAGCGCTTAAAAAAAT
TACGCCCCGCCCTGCCACTCATCGCAGTTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGA
CGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTC
ATCACCGAAACGCGCGAGGCAGCAGATCAATTCGCGCGCGAAGGCGAAGCGGCATGCATAATGTGCCTGTCAAATGGACG

5.2.1.4. Marker & Standards

Table 14 DNA-marker and protein-standards

Marker & Standard	Supplier
BluRay Prestained Protein Marker (10-180 kDa)	Jena Bioscience GmbH (Jena, Germany)
Color Prestained Protein Standard, Broad Range (11–245 kDa)	NEB Inc. (Ipswich, USA)
Color Prestained Protein Standard, Broad Range (1o–250 kDa)	NEB Inc. (Ipswich, USA)
High Range DNA Ladder (0.5-10 kb)	Jena Bioscience GmbH (Jena, Germany)
Protein Marker III (6.5-200 kDa)	Applichem (Darmstadt, Germany)
peqGOLD Protein Marker I	VWR Life Science Competence Center (Erlangen, Germany)

5.2.1.5. Enzymes & Proteins & Additives & Buffers

Table 15 Antibodies and further enzymes

Enzyme & Proteins	Buffer/Comment	Supplier
α-Rab1b-Y77AMP (from rabbit). (= α-Rab1b-AMP)	1:200 = working dilution	Itzen Lab, UKE (Hamburg, Germany)
AflIII	10x NEB 3.1 buffer	NEB Inc. (Ipswich, USA)
Antarctic phosphatase	10x Antarctic phosphatase buffer	NEB Inc. (Ipswich, USA)
Anti-His6-Peroxidase, mouse monoclonal antibody (50 $\frac{U}{mL}$) (= α-His-HRP)	1:5000 = working dilution	Roche Applied Science (Penzberg, Germany)
Anti-GFP monoclonal antibody (= α-GFP) (from mouse) (B-2; sc-9996)	1:1000	Santa Cruz Biotechnology (Dallas, USA)
AscI	10x NEB 3.1 buffer	NEB Inc. (Ipswich, USA)
AsiSI	10x Cutsmart	NEB Inc. (Ipswich, USA)
BamHI-HF	10x Cutsmart	NEB Inc. (Ipswich, USA)
BglII	10x NEB 3.1 buffer	NEB Inc. (Ipswich, USA)
BsaI-HF	10x Cutsmart	NEB Inc. (Ipswich, USA)
BSA, protein standard (1 $\frac{mg}{mL}$ or 2 $\frac{mg}{mL}$)	-	Sigma Aldrich (St. Louis, USA)
BspEI	10x NEB 3.1 buffer	NEB Inc. (Ipswich, USA)
BsrGI-HF	10x Cutsmart	NEB Inc. (Ipswich, USA)
BsiWI-HF	10x Cutsmart	NEB Inc. (Ipswich, USA)
DNAseI powder from bovine pancreas	-	Sigma Aldrich (St. Louis, USA)
DpnI	10x Cutsmart	NEB Inc. (Ipswich, USA)
EcoRV-HF	10x Cutsmart	NEB Inc. (Ipswich, USA)
Goat-anti rabbit-HRP (AB007)	1:40 000 = working dilution	Itzen Lab, UKE (Hamburg, Germany)
Goat-anti mouse-HRP (32430)	1:500 = working dilution	Thermo Fisher Scientific (Waltham, USA)
HindIII-HF	10x Cutsmart	NEB Inc. (Ipswich, USA)
KasI	10x Cutsmart	NEB Inc. (Ipswich, USA)
LysC (0.5 $\frac{\mu g}{\mu L}$)	-	Promega (Madison, USA)
MfeI-HF	10x Cutsmart	NEB Inc. (Ipswich, USA)
NEBuilder HiFi DNA Assembly Master Mix	-	NEB Inc. (Ipswich, USA

NcoI	10x NEB 3.1 buffer	NEB Inc. (Ipswich, USA)
NcoI-HF	10x Cutsmart	NEB Inc. (Ipswich, USA)
NdeI	10x Cutsmart	NEB Inc. (Ipswich, USA)
NheI	10x Cutsmart	NEB Inc. (Ipswich, USA)
NotI-HF	10x Cutsmart	NEB Inc. (Ipswich, USA)
NsiI-HF	10x Cutsmart	NEB Inc. (Ipswich, USA)
PflF1	10x Cutsmart	NEB Inc. (Ipswich, USA)
PstI-HF	10x Cutsmart	NEB Inc. (Ipswich, USA)
Q5 HF DNA polymerase	5x Q5 reaction buffer	NEB Inc. (Ipswich, USA)
SalI-HF	10x Cutsmart	NEB Inc. (Ipswich, USA)
SphI	10x Cutsmart	NEB Inc. (Ipswich, USA)
StrepMAB-Classic, HRP conjugate (=α-Strep-HRP)	1:5000 = working dilution	IBA GmbH (Göttingen, Germany)
T4 DNA ligase	10x T4 DNA Ligase buffer	NEB Inc. (Ipswich, USA)
T4 Polynucleotide Kinase	10x T4 DNA Ligase buffer	NEB Inc. (Ipswich, USA)
Taq DNA Polymerase	10x ThermoPol reaction buffer	NEB Inc. (Ipswich, USA)
TEV protease	-	Lang Lab (München, Germany)
Thrombin	-	Enzyme Research Laboratories (South Bend, USA)
Gibco™ Trypsin-EDTA (0.05 %)	Mixture of proteases derived from porcine pancreas	Thermo Fisher Scientific (Waltham, USA)
Trypsin (0.5 $\frac{\mu g}{\mu L}$)	-	Promega (Madison, USA)
XbaI	10x Cutsmart	NEB Inc. (Ipswich, USA)
Xho I	10x Cutsmart 10x NEB 3.1 buffer	NEB Inc. (Ipswich, USA)
XmaI	10x Cutsmart	NEB Inc. (Ipswich, USA)

Table 16 Commercial buffers and additives for enzymes

Buffer / additive for PCR	**Supplier**
10x Antarctic phosphatase reaction buffer	NEB Inc. (Ipswich, USA)
10x Cutsmart	NEB Inc. (Ipswich, USA)
DMSO	NEB Inc. (Ipswich, USA)
dNTP (10 mM each)	NEB Inc. (Ipswich, USA)
10x NEB 2.1 buffer	NEB Inc. (Ipswich, USA)
10x NEB 3.1 buffer	NEB Inc. (Ipswich, USA)

5x Q5 enhancer	NEB Inc. (Ipswich, USA)
5x Q5 reaction buffer	NEB Inc. (Ipswich, USA)
10x T4 DNA ligase buffer	NEB Inc. (Ipswich, USA)
10x ThermoPol reaction buffer	NEB Inc. (Ipswich, USA)

5.2.1.6. Kits

Table 17 Commercial kits

Kit	Supplier
HiSpeed Plasmid Midi Kit	Qiagen (Hilden, Germany)
HiSpeed Plasmid Maxi Kit	Qiagen (Hilden, Germany)
GenElute™ Bacterial Genomic DNA Kit	Sigma Aldrich (St. Louis, USA)
JBS Error-Prone Kit	Jena Bioscience GmbH (Jena, Germany)
Monarch PCR DNA Cleanup Kit	NEB Inc. (Ipswich, USA)
Monarch DNA Gel Extraction Kit	NEB Inc. (Ipswich, USA)
Pierce™ BCA Protein Assay Kit	Thermo Fisher Scientific (Waltham, USA)
peqGOLD Plasmid Miniprep Kit (Classic Line)	VWR Life Science Competence Center (Erlangen, Germany)
Quick and Easy Gene Deletion Kit	Gene Bridges GmbH (Heidelberg, Germany)

5.2.1.7. Antibiotics

Antibiotics were dissolved, and sterile filtered using a 0.2 µm membrane filter and stored at - 20 °C. The addition of antibiotics to appropriate solutions (e.g. agar, medium, etc.) occurred after the solutions had a temperature of max. 50 °C.

Table 18 Antibiotics for bacterial and mammalian cell culture

Solution	Stock	Working concentration	Comment
Ampicillin	100 $\frac{mg}{mL}$	100 $\frac{\mu g}{mL}$	Dissolved in water
Chloramphenicol	50 $\frac{mg}{mL}$	50 $\frac{\mu g}{mL}$	Dissolved in 96 % (v/v) EtOH
Kanamycin	50 $\frac{mg}{mL}$	50 $\frac{\mu g}{mL}$	Dissolved in water
Spectinomycin	50 $\frac{mg}{mL}$	50 $\frac{\mu g}{mL}$	Dissolved in water
Tetracycline	17 $\frac{mg}{mL}$	17 $\frac{\mu g}{mL}$	Dissolved in 70 % (v/v) EtOH/water
100x AB (VWR)	10 000 U penicillin 10 $\frac{mg}{mL}$ streptomycin	100 U penicillin 0.1 $\frac{mg}{mL}$ streptomycin	For mammalian cell culture

5.2.1.8. Medium for bacterial & eukaryotic cell culture

Table 19 Medium for bacterial & eukaryotic cell culture

Medium	Components	Comment	Supplier/Reference
AI medium	0.25 % (w/v) Aspartate 0.5 % (v/v) Glycerol 1x 17- or 18-AA Mix* 1x Salt-Mix (M-Mix) 40 $\frac{\mu g}{mL}$ Leucine 0.05 % (w/v) L-Arabinose 2 mM Mg_2SO_4 0.05 % (w/v) Glucose 1x Trace Metal Mix	Bacterial cell culture Using uAAs, which are lysine derivates harboring an amide linkage use 4 mM NAM. *Using PylRS-mutants, known for Phe-misincorporation use 17 AA-Mix.	[313]
Non-AI medium	0.25 % (w/v) Aspartate 1x 17- or 18-AA Mix* 1x Salt-Mix (M-Mix) 40 $\frac{\mu g}{mL}$ Leucine 2 mM Mg_2SO_4 0.5 % (w/v) Glucose 1x Trace Metal Mix	Bacterial cell culture *Using PylRS-mutants known for Phe-misincorporation use 17 AA-Mix.	[313]
2x YT medium	16 $\frac{g}{L}$ Trypton 10 $\frac{g}{L}$ Yeast extract 5 $\frac{g}{L}$ NaCl	Bacterial cell culture	Carl Roth GmbH (Karlsruhe, Germany)
LB agar	10 $\frac{g}{L}$ Trypton 5 $\frac{g}{L}$ Yeast extract 5 $\frac{g}{L}$ NaCl 15 $\frac{g}{L}$Agar-Agar	Bacterial cell culture	Carl Roth GmbH (Karlsruhe, Germany)
LB medium	10 $\frac{g}{L}$ Trypton 5 $\frac{g}{L}$ Yeast extract 5 $\frac{g}{L}$ NaCl	Bacterial cell culture	Carl Roth GmbH (Karlsruhe, Germany)
SOB medium	20 $\frac{g}{L}$ Trypton 5 $\frac{g}{L}$ Yeast extract 0.96 $\frac{g}{L}$ $MgCl_2$ 0.5 $\frac{g}{L}$ NaCl 0.186 $\frac{g}{L}$ KCl	Bacterial cell culture	Carl Roth GmbH (Karlsruhe, Germany)
SOC medium	20 $\frac{g}{L}$ Trypton 5 $\frac{g}{L}$ Yeast extract 0.96 $\frac{g}{L}$ $MgCl_2$ 0.5 $\frac{g}{L}$ NaCl 0.186 $\frac{g}{L}$ KCl 20 mM Glucose	Bacterial cell culture	Prepare SOB medium and add glucose
Gibco™ DMEM	+ 4.5 $\frac{g}{L}$ D-Glucose + 110 $\frac{mg}{L}$ Sodium Pyruvate	Mammalian cell culture	Thermo Fisher Scientific (Waltham, USA)

FBS		Mammalian cell culture	Biochrom GmbH (Berlin, Germany)
Gibco™ Opti-MEM	+ HEPES + L-Glutamine + 2.4 $\frac{g}{L}$ Sodium Bicarbonate	Mammalian cell culture Reduced serum medium	Thermo Fisher Scientific (Waltham, USA)

5.2.1.9. Solutions and Stocks

Table 20 Solutions & Stocks for biological works

Solution & Stock	Stock	Comment
APS stock	10 % (w/v) APS in water	SDS-PAGE
L-Arabinose stock	20 % (w/v) in water	Bacterial cell culture
ApCpp stock	500 mM in 100 mM Tris-buffer (pH 7.4)	Crystallization
ATP stock	500 mM in 100 mM Tris-buffer (pH 7.4)	Crystallization Protein Assay
5 % (w/v) Aspartate	5 % (w/v) in water, pH 7.5	Bacterial cell culture
Biotin stock	1 $\frac{mg}{mL}$ in water	Bacterial cell culture
EDTA stock	500 mM in water	Buffer
Glucose stock	20 % (w/v) in water	Medium
40 % (w/v) Glucose	40 % (w/v) in water	Bacterial cell culture
Glycerol	10 % (v/v) in water	Bacterial cell culture
IPTG stock	1 M in water	Bacterial cell culture
Leucine stock	4 $\frac{mg}{mL}$ in water, pH 7.5	Bacterial cell culture
Mg_2SO_4 stock	1 M in water	Bacterial cell culture
NAM stock	1 M in water	Bacterial cell culture
PEI	1 $\frac{mg}{mL}$ in H_2O (pH 7.0 -7.6)	Mammalian cell culture
PMSF stock	200 mM in DMSO	Protein Puification
Sodium azide stock	20 % (w/v) in water	-
Solvent A	MilliQ-H_2O+ 0.1 % FA	HPLC & LC-MS
Solvent B	ACN + 0.1 % FA	HPLC & LC-MS
Thiamin stock	1 $\frac{mg}{mL}$ in water	Bacterial cell culture
50x M-Mix (Salt-Mix)	1.25 M Na_2HPO_4 1.25 M KH_2PO_4 2.5 M NH_4Cl 0.25 M Na_2SO_4	Bacterial cell culture [313]
25x 18 AA-Mix	5 $\frac{mg}{mL}$ Glutamic sodium salt 5 $\frac{mg}{mL}$ Aspartic acid 5 $\frac{mg}{mL}$ Lysine-HCl 5 $\frac{mg}{mL}$ Arginine-HCl 5 $\frac{mg}{mL}$ Histidine-HCl-H_2O 5 $\frac{mg}{mL}$ Alanine 5 $\frac{mg}{mL}$ Proline 5 $\frac{mg}{mL}$ Glycine	Bacterial cell culture [313]

	5 $\frac{mg}{mL}$ Threonine 5 $\frac{mg}{mL}$ Serine 5 $\frac{mg}{mL}$ Glutamine 5 $\frac{mg}{mL}$ Asparagine-H_2O 5 $\frac{mg}{mL}$ Valine 5 $\frac{mg}{mL}$ Leucine 5 $\frac{mg}{mL}$ Isoleucine 5 $\frac{mg}{mL}$ Phenylalanine 5 $\frac{mg}{mL}$ Tryptophan 5 $\frac{mg}{mL}$ Methionine	
25x 17 AA-Mix	same as 25x 18 AA-Mix but without 5 $\frac{mg}{mL}$ Phenylalanine	Bacterial cell culture
5000x Trace Metal Mix	20 mM $CaCl_2$ * 2 H_2O 10 mM $MnCl_2$ * 4 H_2O 10 mM $ZnSO_4$ * 7 H_2O 2 mM $CoCl_2$ * 6 H_2O 2 mM $CuCl_2$ 2 mM $NiCl_2$ 2 mM Na_2MoO_4 * 2 H_2O 2 mM Na_2SeO_3 2 mM H_3BO_3 50 mM $FeCl_3$	Bacterial cell culture 313
5x M9 salts	15 $\frac{g}{L}$ KH_2PO_4 64 $\frac{g}{L}$ Na_2HPO_4 2.5 $\frac{g}{L}$ NaCl 5.0 $\frac{g}{L}$ NH_4Cl	Bacterial cell culture

5.2.1.10. Buffer

Buffers, used with the ÄKTA pure system (GE Healthcare), were filtered through a membrane with 0.45 µm pore size and degassed before use.

Table 21 Buffer compositions

Buffer	**Components**	**Comment**
5x SLIM buffer (H-buffer)	750 mM NaCl 125 mM Tris-HCl, pH 9 100 mM EDTA	Hybridization SLIM
Q5 Reaction Buffer (5x)	100 mM Tris-HCl, pH 8.8 50 mM KCl 5 mM $MgCl_2$ 50 mM $(NH_4)SO_4$ 0.5 % (v/v) Triton X-100 0.5 $\frac{mg}{mL}$ BSA	PCR
10x DNA Loading buffer	40 % (w/v) Sucrose 0,15 % (w/v) Orange G 0.05 % (w/v) Xylen cyanol 0.05 % (w/v) Bromophenol blue	Agarose gelelectrophoresis

50x TAE buffer	2 M Tris-HCl, pH 8.8 0.1 M EDTA 1 M AcOH	Agarose gelelectrophoresis
4x SDS Loading buffer	240 mM Tris-HCl, pH 6.8 40 % (v/v) Glycerol 8% (w/v) SDS 0.04 % (w/v) Bromophenol blue 5 % (v/v) β-ME or 400 mM DTT	SDS-PAGE
20x SDS-MES Running buffer	1 M MES, pH 7.3 1 M Tris base, pH 7.3 2 % (w/v) SDS	SDS-PAGE Do not adjust pH
4x Resolving buffer	1.5 M Tris-HCl, pH 8.8 0.4 % (w/v) SDS	SDS-PAGE
4x Stacking buffer	0.5 M Tris-HCl, pH 6.8 0.4 % (w/v) SDS	SDS-PAGE
Semi-Dry transfer buffer (Bjerrum&Schiefer-Nielsen transfer buffer)	48 mM Tris base, pH 8.9 - 9.4 39 mM Glycerol 1.3 mM = 0.0375 % (w/v) SDS 20 % (v/v) MeOH	Do not adjust pH Western Blot
10x PBS (pH 7.4)	1370 mM NaCl 270 mM KCl 100 mM Na_2HPO_4 20 mM K_2PO_4	-
10x TBS	200 mM Tris-HCl, pH 7.6 1500 mM NaCl	Western Blot
1x TBS-T	20 mM Tris-HCl. pH 7.6 150 mM NaCl 0.1 % (v/v) Tween 20	Western Blot
Blocking solution	5 % (w/v) skimmed milk powder in 1x TBS-T	Western Blot
Buffer X	7 M Urea 2 M Thiourea 20 mM HEPES, pH 7.5	Denaturation of proteins for MS/MS. LC-MS grade
Buffer E	80 % (v/v) ACN 0.5 % (v/v) formic acid	Elution buffer for MS/MS samples. LC-MS grade
His-lysis buffer (Mammalian cells)	50 mM Tris-HCl, pH 8 150 mM NaCl 1 % (w/v) Triton X-100 20 mM Imidazole 1x Protease inhibitor solution (VWR)	If freeze-thaw lysis was used, 1 % (w/v) Triton X-100 were omitted from the buffer
His-wash buffer (Mammalian cells)	50 mM Tris-HCl, pH 8 150 mM NaCl 1 % (w/v) Triton X-100 20 mM Imidazole	If freeze-thaw lysis was used, 1 % (w/v) Triton X-100 were omitted from the buffer
His-elution buffer (Mammalian cells)	50 mM Tris-HCl, pH 8 150 mM NaCl 1 % (w/v) Triton X-100 300 mM Imidazole	If freeze-thaw lysis was used, 1 % (w/v) Triton X-100 were omitted from the buffer
RIPA lysis buffer (Mammalian cells)	50 mM Tris-HCl, pH 150 mM NaCl 1 % (w/v) Triton X-100	

	1 mM EDTA 0.1 % (w/v) SDS 0.5 % (w/v) sodium doxychlorate 1x Protease inhibitor solution (VWR)	
His-lysis buffer (*E. coli*)	20 mM Tris-HCl, pH 8 300 mM NaCl 30 mM Imidazole 1 per 30 mL tablet c0mplete™ protease inhibitor or 1 mM PMSF 0.1 $\frac{mg}{mL}$ DNAseI	Per 1 g cell pellet use 10 mL buffer. For sfGFP use 1mM PMSF; for all other proteins use c0mplete™ protease inhibitor
His-wash buffer (*E. coli*)	20 mM Tris-HCl, pH 8 300 mM NaCl 30 mM Imidazole	
His-elution buffer (*E. coli*)	20 mM Tris-HCl, pH 8 300 mM NaCl 300 mM Imidazole	
GFP-SEC buffer (*E. coli*)	20 mM Tris-HCl, pH 8 300 mM NaCl	
GFP-storage buffer (*E. coli*)	20 mM Tris-HCl, pH 8 300 mM NaCl	GFP-storage buffer (*E. coli*)
His-lysis buffer PylRS (*E. coli*)	20 mM Tris-HCl, pH 7.4 300 mM NaCl 30 mM Imidazole 1 per 30 mL tablet c0mplete™ protease inhibitor or 1 mM PMSF 0.1 $\frac{mg}{mL}$ DNAseI	
His-wash buffer PylRS (*E. coli*)	20 mM Tris-HCl, pH 7.4 300 mM NaCl 30 mM Imidazole	
His-elution buffer PylRS (*E. coli*)	20 mM Tris-HCl, pH 7.4 300 mM NaCl 300 mM Imidazole	
PylRS-SEC buffer (*E. coli*)	10 mM HEPES, pH 7.4 300 mM NaCl 5 mM $MgCl_2$ 1 mM DTT	Also used for loading of uAA onto PylRS
PylRS-thrombin buffer (*E. coli*)	10 mM HEPES, pH 7.4 300 mM NaCl 5 mM $MgCl_2$ 0.75 mM DTT	
Rab1b/DrrA-lysis buffer (*E. coli*)	50 mM HEPES, pH 8 500 mM LiCl 20 mM Imidazole 2 mM β-ME 1 mM $MgCl_2$ 1 per 30 mL tablet c0mplete™ protease inhibitor 0.1 $\frac{mg}{mL}$ DNAse I	Per 1 g cell pellet use 10 mL buffer. Working with Rab1b needed the addition of 0.01 mM GTP or GDP. For Rab1b use GDP; for Rab1b-Q67A use GTP
Rab1b/DrrA-wash buffer (*E. coli*)	50 mM HEPES, pH 8 500 mM LiCl 20 mM Imidazole 2 mM β-ME 1 mM $MgCl_2$	Working with Rab1b needed the addition of 0.01 mM GTP or GDP. For Rab1b use

		GDP; for Rab1b-Q67A use GTP
Rab1b/DrrA-elution buffer (*E. coli*)	50 mM HEPES, pH 8 500 mM LiCl 500 mM Imidazole 2 mM β-ME 1 mM $MgCl_2$	Working with Rab1b needed the addition of 0.01 mM GTP or GDP. For Rab1b use GDP; for Rab1b-Q67A use GTP
Rab1b-storage buffer (*E. coli*)	20 mM HEPES, pH 8 150 mM NaCl 2 mM DTT 5 mM $MgCl_2$ 0,01 mM GTP or GDP	For Rab1b use GDP; for Rab1b-Q67A use GTP
DrrA-storage buffer (*E. coli*)	20 mM HEPES, pH 8 150 mM NaCl 2 mM DTT 1 mM $MgCl_2$	
SEC-buffer Rab1b-DrrA-complex (*E. coli*)	20 mM HEPES, pH 8 150 mM NaCl 2 mM DTT 5 mM $MgCl_2$ 0,01 mM GTP or GDP	For Rab1b use GDP; for Rab1b-Q67A use GTP
Storage buffer Rab1b-DrrA-complex (*E. coli*)	20 mM HEPES, pH 8 150 mM NaCl 2 mM DTT 5 mM $MgCl_2$ 0,01 mM GTP or GDP	For Rab1b use GDP; for Rab1b-Q67A use GTP
ResQ-buffer I (*E. coli*)	20 mM HEPES, pH 8 2 mM DTT 5 mM $MgCl_2$	Working with Rab1b needed the addition of 0.01 mM GTP or GDP. For Rab1b use GDP; for Rab1b-Q67A use GTP
ResQ-buffer II (*E. coli*)	20 mM HEPES, pH 8 1000 mM NaCl 2 mM DTT 5 mM $MgCl_2$	Working with Rab1b needed the addition of 0.01 mM GTP or GDP. For Rab1b use GDP; for Rab1b-Q67A use GTP
Strep-lysis buffer (*E. coli*)	100 mM Tris-HCl, pH 8 150 mM NaCl 1 mM EDTA 1 per 30 mL tablet c0mplete™ protease inhibitor	Per 1 g cell pellet use 10 mL buffer. For *Sa*ClpP do not use protease inhibitors!
Strep-wash buffer (*E. coli*)	100 mM Tris-HCl, pH 8 150 mM NaCl 1 mM EDTA	
Strep-elution buffer (*E. coli*)	100 mM Tris-HCl, pH 8 150 mM NaCl 1 mM EDTA 50 mM Biotin/2.5 mM Desthiobiotin	For Strep-Tactin/StrepTrap HP use desthiobiotin. For Strep-TactinXT use biotin.
*Sa*ClpP-storage buffer	20 mM HEPES, pH 7.0	

(*E. coli*)	100 mM NaCl	
PZ-buffer	25 mM HEPES, pH 7.6 200 mM KCl 1 mM DTT 5 mM $MgCl_2$ 10 % (v/v) Glycerol	
Xlink buffer AMP-assay buffer	20 mM HEPES, pH 8 150 mM NaCl 2 mM β-ME 1 mM $MgCl_2$ 0,01 mM GTP	For experiments with $DrrA_{16-533}$ additional GDP (final 10 µM) was added

5.2.1.11. Unnatural amino acids stocks

Table 22 Unnatural amino acids and usage

Unnatural amino acid (uAA)	Stock solution [mM]	Working concentration bacteria [mM]	Working concentration mammalian cells [mM]	Solvent
AcK	100	2	-	H_2O
AzGGK	100 or 200	4	-	H_2O
BCNK	50	2	1	100 mM NaOH/DMSO (1:1)
BocK	100	1 - 2	1	100 mM NaOH
BrC6K	100	2	1	1 M NaOH
BrC7K	100	2	1	1 M NaOH
BrCO6K	100	2	1	1 M NaOH
BrCO6K (HPLC)	100	2	1	0.3 M NaOH
BrCO7K	100	2	1	1 M NaOH
BrCN6K	100	1	-	H_2O
BrCN7K (as HBr-salt)	100	1	1	H_2O
BrCN7K (HPLC)	100	1	1	0.2 M NaOH
BrCOS8K	100	2	-	2 M NaOH
CpK	100	2	-	100 mM NaOH
DAP	100	1	-	70 mM TFA in H_2O/DMSO (1:3)
DiaziK	100	2	-	100 mM TFA
GGK	100	4	-	H_2O
HpGGK	100	4	0.5 - 1	15 % (v/v) DMSO/H_2O
mTetK	100	2	-	200 mM TFA in DMSO

pDAP	100	1	-	70 mM TFA in H_2O/DMSO (1:3)
TetK	100	2	-	200 mM TFA (H_2O)
TM-439	100	1	1	H_2O
mTriK	100	2	1	H_2O

5.2.2. Bacterial Cell Culture Methods

5.2.2.1. Solid culture of *E. coli*

For culturing *E. coli* on plates, e.g. if single colonies are needed, LB-agar plates with appropriate antibiotics were used. After plating or streaking *E. coli* cells on LB agar plates, an incubation time of 12- 16 hours at 37 °C is sufficient to get visible colonies. Plates could be stored at 4 °C up to one month.

5.2.2.2. Liquid culture of *E. coli*

In general *E. coli* cells were cultivated in liquid cultures at 37 °C, 200 rpm in sterile Erlenmeyer flasks. Depending on further application, bacterial cells were grown in either LB, 2xYT, TB, non-AI or AI medium containing appropriate antibiotics. In order to ensure good oxygen entry, the medium volume should not exceed $\frac{1}{3}$ of the total volume. The cell number and cell growth can be monitored by OD_{600}. Cultures with high densities ($OD_{600} > 1$) had to be diluted 1:10 with corresponding medium before measurement.
For stationary pre-cultures, either one colony of appropriate *E. coli* was picked from a culture plate, or a fresh transformation-cell-mixture was directly inoculated in 50 mL liquid medium (containing antibiotics) and grown overnight (around 12 hours) at 37 °C, 200 rpm.

5.2.2.3. Preparation of *E. coli* stocks for long-term storage

For long-term storage of *E. coli* cells (with or without containing a plasmid) cryo-cultures were used. Therefore, stationary pre-cultures of the appropriate cells were prepared in LB medium at 37 °C. For cryo-conservation 1 mL of the respective culture was mixed in special cryo-tubes with 100 % glycerol to a final concentration of 16.6 % (v/v) glycerol. The cell suspension was flash frozen in liquid nitrogen and stored at -80 °C.

5.2.2.4. Preparation of competent *E. coli*

Chemically competent *E. coli* cells

For the preparation of chemically competent *E. coli,* an overnight culture was prepared in either LB or 2xYT medium at 37 °C, 200 rpm for 12 – 16 h. A main culture (100 mL), in either LB or 2xYT medium, was inoculated 1:100 from the overnight culture and grown at 37 °C, 200 rpm until the OD_{600} reached 0.5 - 0.7. The culture was cooled on ice for 20 min before harvesting the cells by centrifugation at 4200 g, 4 °C, 10 min. The pellet was resuspended carefully in 50 mL ice-cold $CaCl_2$-solution (100 mM in water) and incubated on ice for 30 min. Cells were again centrifuged (4200 g, 4 °C, 10 min) and resuspended in 10 mL of ice-cold $CaCl_2$-solution (100 mM) before mixing with 2 mL 100 % glycerol (final 16.6 % (v/v)). After incubation for

5 min on ice, the cell suspension was aliquoted and flash frozen in liquid nitrogen before storage at -80 °C.

Electro-competent *E. coli* cells

A main culture (500 mL), in LB or 2xYT medium, was inoculated 1:100 from a stationary overnight culture (also in LB or 2xYT), and grown at 37 °C, 200 rpm to an OD_{600} of 0.5 – 0.7. After cooling for 20 min on ice the culture was harvested by centrifugation (4200 rpm, 4 °C, 10 min). The pellet was washed three times with 10 % (v/v) glycerol, by resuspending and then centrifuging (4200 rpm, 4 °C, 10 min). The first and second resuspension step occurred in 500 mL, the third in 20 mL 10 % (v/v) glycerol. After the last washing round the cells were finally resuspended in 2-3 mL 10 % (v/v) glycerol, aliquoted, flash frozen in liquid nitrogen and stored at -80 °C.

Determination of transformation efficiency and integrity of competent *E. coli*

To ensure the integrity and the functionality/quality of the competent *E. coli*, the cells had to be tested. To exclude a contamination a sample of the freshly prepared, untransformed cells was plated on LB agar plates of each antibiotic in the lab and one plate without antibiotics. The plates were incubated overnight at 37 °C. The next day, all plates are supposed to be empty, except for the plate without antibiotic or containing antibiotics which fit to the intrinsic resistance of the respective *E. coli* strain. The functionality was validated by calculating the transformation efficiency (number of colony forming units (cfu) growing per 1 µg of transformed plasmid DNA). Therefore cells, which were flash frozen and stored at -80 °C for 24 h, were single and co-transformed with plasmid DNA (according to 5.2.2.5, depending on the kind of competence), plated on LB agar with appropriate antibiotics and incubated at 37 °C overnight. The next day, grown colonies were counted and the transformation efficiency was determined using the following formula F1:

$$(\boldsymbol{F1})\ \ Transformation\ \ efficiency\ = \left(\frac{grown\ colonies\ on\ the\ plate}{transformed\ DNA\ amount\ [ng]}\right) \times 1000\,\frac{ng}{\mu g} \times \frac{Volume\ of\ the\ transfomation}{plated\ volume}$$

The transformation efficiency for chemically competent *E. coli* should be in the range of 10^5 – $10^7\ \frac{cfu}{\mu g\ DNA}$ and for electro competent cells around 10^7 – $10^8\ \frac{cfu}{\mu g\ DNA}$.

5.2.2.5. Transformation of *E. coli*

Chemical transformation/Heat shock transformation of *E. coli* cells with plasmid DNA

Chemically competent *E. coli* cells (50 - 100 µL) were thawed on ice and mixed with 0.5 - 1.0 µL of the respective plasmid. For co- or triple-transformations of different plasmids, 0.5 - 1.0 µL of each plasmid was added. The cell-DNA mixture was incubated for 20 – 30 min on ice before heat shock at 42 °C for 45 s in a water bath. Cells were again incubated on ice for 5 min before rescuing with 900 µL SOC medium (in special cases SOB is used instead of SOC, if glucose is disadvantageous for further working steps). The cell suspension was grown at 37 °C, 200 rpm for 1 hour. For *E. coli* Rosetta (DE3) cells the incubation in SOC medium occurred for 2 hours. The transformation mixture can either be plated on culture plates or directly inoculated in medium with appropriate antibiotics. In general, single colonies on culture plates were obtained after incubation for 12 – 16 hours at 37 °C. Stationary liquid cultures resulted after 12 – 16 hours at 37 °C, 200 rpm.

Electro-transformation/Electroporation of *E. coli* cells

For electroporation of DNA electro-competent *E. coli* cells (50 - 100 µL) were thawed on ice. 0.5 - 1.0 µL of the respective linear or plasmid DNA was added. The cell-DNA mixture was transferred into a pre-cooled electroporation cuvette (2.0 mm gap size) avoiding bubbles. Electroporation pulse was performed at 2.5 kV. Cells were rescued with 900 µL SOC medium (in special cases SOB is used instead of SOC, if glucose is disadvantageous for further working steps) and incubated at 37 °C, 200 rpm for 1 hour. Then cells were either spread on culture plates or inoculated in medium containing corresponding antibiotics. After 12 - 16 hours of incubation at 37 °C, visible single colonies appeared on plates or stationary pre-cultures were grown.

5.2.2.6. Recombinant protein expression in *E. coli*

Expression of Rab1b/DrrA in *E. coli*

The expression of Rab1b and DrrA (without amber stop codons) from araBAD-driven expression-vectors was carried out in *E. coli* NEB 10-beta cells. Plasmids were chemically transformed (see 5.2.2.5) and directly inoculated in 50 mL LB medium (+ antibiotic). Next day the pre-culture was diluted to an OD_{600} of 0.05 - 0.07 in fresh LB medium (+ antibiotic). Cells were incubated at 37 °C, 200 rpm until they reached an OD_{600} around 0.6 – 0.7 and cooled on ice for 15 min. Then expression was induced using arabinose (final 0.02 % (w/v)) and took place at 37 °C, 200 rpm for 12 -16 hours. Cells were harvested and the pellet could be flash frozen with liquid nitrogen and stored at -20 °C until further use.

Expression of *Sa*ClpP in *E. coli*

The expression of *Sa*ClpP (without amber stop codons) from pBAD-vectors was carried out in *E. coli* NEB 10-beta cells. Therefore, plasmids were chemically transformed (see 5.2.2.5) and directly inoculated in 50 mL LB medium (+ Amp). Next day the pre-culture was diluted to an OD_{600} of 0.05 - 0.07 in fresh LB medium (+ Amp). Cells were incubated at 37 °C, 200 rpm until they reached an OD_{600} around 0.6 – 0.7 and cooled on ice for 15 min. Then expression was induced using arabinose (final 0.02 % (w/v)) and took place at 30 °C, 200 rpm for 12 – 16 hours. Cells were harvested and the pellet could be flash frozen with liquid nitrogen and stored at -20 °C until further use.

T7-driven recombinant expression of PylRS in *E. coli*

PylRS proteins of *M. mazei* or *M. alvus* for protein crystallization were expressed in *E. coli* Rosetta (DE3) cells. For *M. mazei*, only the C-terminal domain (residues 185 – 454) of PylRS was expressed, whereas for *M. alvus* the full-length protein was used. The respective pET28 vectors, harboring the PylRS variant fused to a N-terminal His6-Tag and a thrombin cleavage site, were transformed into chemically competent *E. coli* Rosetta (DE3) cells according to 5.2.2.5. 50 µL, 200 µL and 500 µL of the transformation mixture were directly inoculated into 50 mL 2x YT medium each with kanamycin and chloramphenicol. After incubation for 12 – 16 hours at 37 °C and 200 rpm, the OD_{600} of the pre-cultures was measured and only cultures with an $OD_{600} < 1.0$ were used for preparation of the respective expression cultures. Therefore, the cells were diluted to an OD_{600} around 0.03 in 2x YT medium containing Kan and CAM and grown at 37 °C, 200 rpm to an OD_{600} of 0.6 – 0.7. Cells were cooled on ice for 15 min before expression was induced by the addition of IPTG (1 mM final concentration). Expression took

place at 18 °C, 200 rpm for 10 – 12 hours. Cells were pelleted at 5000 g, 4°C for 15 min. The pellet could be flash frozen with liquid nitrogen and stored at -20 °C.

5.2.2.7. Amber suppression of proteins in *E. coli*

Amber suppression in *E. coli* using auto-induction medium

Plasmids, bearing the POI, PylRS-mutant and corresponding tRNA, were co-transformed in chemically competent *E. coli* strains (see 5.2.2.5), which are compatible with arabinose driven expression systems (e.g. *E. coli* NEB 10-beta, *E. coli* NEB 10-beta ΔcobB, *E. coli* MegaX DH10B or *E. coli* BL21). The transformation was directly inoculated into 50 mL non-AI medium with appropriate antibiotics and incubated overnight at 37 °C, 200 rpm to obtain a respective pre-culture. The pre-culture was diluted to an OD_{600} between 0.04 – 0.07 in AI medium with corresponding antibiotics. The culture was incubated at 37 °C, 200 rpm to an OD_{600} around 0.2, before supplementing the culture with uAA (concentration depends on uAA, see Table 22). In the case of using an uAA with light-sensitive groups the culture flask was protected against light using aluminum foil. Cells were grown and expressed as long as desired (normally 4 hours, 6 hours or overnight) at 37 °C, 200 rpm. After expression the cells were either harvested for further protein purification or used for SDS-PAGE or western blot analysis of the cell lysates.

Amber suppression in *E. coli* using LB/2x YT medium

The co-transformation of plasmids (see 5.2.2.5), encoding the POI, PylRS-mutant and corresponding tRNA, in chemically competent *E. coli* strains (compatible with arabinose driven expression systems) were directly inoculated in 50 mL LB or 2x YT medium with respective antibiotics. After an incubation overnight at 37 °C, 200 rpm a stationary pre-culture was obtained, which was diluted in fresh LB or 2x YT medium, containing appropriate antibiotics, to an OD_{600} between 0.04 – 0.07. The cells were incubated at 37 °C, 200 rpm and uAA (concentration dependent on uAA, see Table 22) was added when the culture reached an OD_{600} around 0.3. In expressions using BrC6K or BrC7K nicotinamide (10 mM final) was added. Induction with 0.02 % (w/v) arabinose occurred when the OD_{600} was around 0.6 - 0.7. Cells were then expressed at 37 °C, 200 rpm for 4 hours, 6 hours or overnight.

Amber suppression of *Sa*ClpP in *E. coli*

The expression with amber suppression for *Sa*ClpP proteins differed from the general procedure in *E. coli*. The co-transformation of pBK_pDAPRS plasmid and the pPylT_*Sa*ClpP-S98X plasmid was performed in chemically competent *E. coli* NEB 10-beta (see 5.2.2.5) and inoculated in 50 mL LB medium containing appropriate antibiotics. The pre-culture was grown overnight at 37 °C, 200 rpm before diluting in fresh LB (with antibiotics) to an OD_{600} between 0.05 – 0.07. The cells were incubated at 37 °C, 200 rpm and supplemented with pDAP (final concentration 1 mM) at an OD_{600} of 0.3. Cells were further grown at 37 °C, 200 rpm to an OD_{600} around 0.6 - 0.7. Then the culture was cooled on ice for 15 min before shifting the incubation temperature to 30 °C. Cells were expressed overnight at 30 °C, 200 rpm. After expression the cells were either harvested for further protein purification or used for SDS-PAGE or western blot analysis of the cell lysates.

5.2.2.8. Cell-harvest and cell-lysis of *E. coli*

Cultures of *E. coli* for protein expression had to be harvested after their respective expression times. Therefore, the culture was centrifuged at 4200 rpm for 15 min, 4 °C. The supernatant was discarded, and the pellet was flash frozen with liquid nitrogen and stored at -20 °C or -80 °C or could be directly used for protein purification.

Cell disruption by sonification

For disrupting bigger cells amounts (from expression cultures larger than 10 mL) we used the Branson Digital Sonifier SFX 250. Cell pellets were resuspended in respective lysis buffer and lysed by ultrasonication at 20 % amplitude for a total pulse time of 2.5 min – 3 min with permanent cooling in an ice-water slurry. During the lysis process, 1 second of sonication is followed by 2 seconds of pausing. After sonification the lysate was centrifuged at 13 000 g, 4 °C for 30 min to obtain the clear supernatant (= cleared lysate) for further work.

Mechanical cell disruption with glass beads

In case of very small volumes of expression cultures (4 - 10 mL), glass beads and mechanical destruction were chosen for cell lysis. Therefore, cell pellets were resuspended in 500 – 1000 µL appropriate lysis buffer and 50 – 80 mg glass beads were added. The cells-glass bead suspension was vortexed for 10 min at full power. The suspension was centrifuged at 13 000 rpm, 4 °C for 20 min to separate the beads together with the cell debris from the supernatant (= cleared lysate), which could be used in following working steps.

5.2.2.9. Creation of *E. coli* knock-out strains

Producing *E. coli* knock-out strains by Red/ET recombination

Metabolic engineered *E. coli* strains for the expression of special proteins or the amber suppression with special uAAs were produced via Quick and Easy Gene Deletion Kit from Gene Bridge. These strains have, depending on the respective application, certain chromosomal deletions of one or several genes. To obtain these chromosomal knock-outs λ-mediated homologous recombination (Red/ET recombination) *in vivo* was used. During the recombination event the gene of interest is replaced by a functional antibiotic resistance cassette flanked by FRT sites (= FRT-PGK-gb2-neo-FRT). The recombination site is mediated via homology arms flanking the FRT sites of the functional cassette. The system provided by Gene Bridges GmbH is suited for all laboratory *E. coli* strains. The approach encompassed four steps: Generation of the linear FRT-PGK-gb2-neo-FRT flanked with homology arms; Transformation of the plasmid encoding proteins necessary for Red/ET recombination; Induction of the Red/ET recombination & transformation of linear FRT-PGK-gb2-neo-FRT flanked with homology arms; Identification of successful recombined *E. coli* clones. For multiple chromosomal deletions, two additional working steps were necessary: Transformation of the FLP expression plasmid; FLP recombination for selection marker removal.

Generation of a linear functional cassette with homology arms for recombination

For guiding the homologous recombination event to the chromosomal target gene of choice, homology regions (50 bp) had to be attached on both sites of the FRT-PGK-gb2-neo-FRT cassette by PCR. Therefore, PCR primers were used which anneal the FRT-PGK-gb2-neo-FRT template (18 – 24 bp) and have additional 50 bp overhangs at their 5'-end which are homologous to the chromosomal target site of choice. In general, the 50 bp are chosen directly

adjacent up- and downstream to the desired insertion site as overhangs (for reverse primer do not forget to take the reverse complement orientation).
The linear functional cassette with homology arms was generated using primers from Table 24 in Q5-PCR according to 5.2.4.3 and yielded a fragment of around 1750 bp which was verified by agarose gel electrophoresis (see 5.2.4.4). After purification using a Monarch PCR DNA Cleanup Kit, the concentration of the linear cassette was adjusted to 100 – 400 $\frac{ng}{\mu L}$ and could be further used for electroporation and recombination.

Table 23 Overview of *E. coli* knock-out strains and the corresponding primers

Recombinant *E. coli* strain	***E. coli* host strain**	**Gene to delete**	**Primer pair for FRT-PGK-gb2-neo-FRT flanked with homology arms (5'→ 3')**	**Analytical primer pair (5'→ 3')***
NEB 10-beta ΔcobB	NEB 10-beta	ΔcobB	G19 G20	G21 G22
BL21-AI ΔcobB	BL21-AI	ΔcobB	G19 G20	G21 G22
BL21 (DE3) ΔcobB	BL21 (DE3)	ΔcobB	G19 G20	G21 G22
NEB 10-beta ΔelaD	NEB 10-beta	ΔelaD	G24 G27	G25 G28
BL21-AI ΔelaD	BL21-AI	ΔelaD	G23 G24	G25 G26
KEIO ΔnfsA + ΔnfsB	KEIO JW0835 (ΔnfsA)	ΔnfsB	G31 G32	G37 G38 & G41** G42**
KEIO ΔnfsA + ΔnfsB + ΔazoR	KEIO ΔnfsA + ΔnfsB	ΔazoR	G33 G34	G35 G36
KEIO ΔnfsA + ΔnfsB + ΔydjA	KEIO ΔnfsA + ΔnfsB	ΔydjA	G46 G47	G48 G49

* Analytical primers anneal the flanking regions of the respective insertion site
** Primer anneal within the gene which should be replaced

Table 24 Primer used for *E. coli* knock-outs

Primer	**Sequence (5'→3')**
G19	TGCGTGGTGCGGCCTTCCTACATCTAACCGATTAAACAACAGAGGTTGCTAATTAACCCTCACTAAAGGGCG
G20	CGCAAATTCAATTAATTGCGTCCCCTTGCAGGCCTGATAAGCGTAGTGCATAATACGACTCACTATAGGGCTC
G21	CGGCCTTCCTACATCTAAC
G22	TCAATTAATTGCGTCCC
G23	TGCCTCTTTTCATCTTGATTCAGCACAACCTTTGCAACTAAAAGGAGCCCAATTAACCCTCACTAAAGGGCG
G24	TTCGACCAGTCAGCATATTTGCTGACTGGTCGAATTAATTAACAATGATGTAATACGACTCACTATAGGGCTC
G25	AAACAGGGCGAAACTTGCC
G26	TGGAAATACCCGAAGATTG

G27	CTCTTTTCATCTTGATTCAGCACAACCTTTGCAACTAAAAGGAGCCCATGAATTAACCCTCACTAAAGGGCG
G28	TACAGGGCGGAATTTATTC
G31	ACCTTGTAATCTGCTGGCACGCAAAATTACTTTCACATGGAGTCTTTATG AATTAACCCTCACTAAAGGGCG
G32	TAGCCGGGCAGATGCCCGGCAAGAGAGAATTACACTTCGGTTAAGGTGATTAATACGACTCACTATAGGGCTC
G33	AACAAGCAACGGGGCATCGCCCATTCAAACATCTATAAGGAAACACCATGAATTAACCCTCACTAAAGGGCG
G34	TCTACAGTCCACATCAAGACCGTGTCCGGTTATGCAGAAACAATGCTGTCTAATACGACTCACTATAGGGCTC
G35	GAAAACAGGTAAGTGGATTGAGATG
G36	CTATCTCTGTGGCAGGATTGTG
G37	CAGGGTTATGCAAATCAGGAG
G38	CTCGCTTACCATTTCTCGTTG
G39	CCAGAAGTGCAGATTCTG
G40	CCATCCCGGGACAATATG
G41	CGCTGTGATGACCTACC
G42	GCCAGCAAAAAACTTACCCC
G43	GGATATCATTTCTGTCGCCTTAAAGCGTCATTCCACTAAGGCATTTGATG AATTAACCCTCACTAAAGGGCG
G44	GTGATGTTTTGCGGCAGACGAGATTTCGGCAGCGTAGCGTTAAAATCTTCTAATACGACTCACTATAGGGCTC
G45	CTGATGGATCTTAGTCAGC
G46	ATCCAGACAGTTTCGCGACAATTATCAGAAATAAGTTACAAACGGCGTCGAATTAACCCTCACTAAAGGGCG
G47	CTCTCGTGTTAATTTTGTGAATAAATATCACGACAGGAGTTAATCAAATGTAATACGACTCACTATAGGGCTC
G48	CTTCGCCTGCTCACTATGC
G49	CGGTACTTGTCAGCAGGAATTG

Transformation of *E. coli* with pRedET

The thermo-sensitive pRedET plasmid encodes the proteins (Redα, Redβ; Redγ; Gam) for the recombination derived from the λ-phage. Additionally, the RecA protein is expressed to compensate the toxic effect of Redγ on *E. coli* (many *E. coli* strains such as DH10B are recA negative ($recA^-$)). RepA is further encoded and awards thermo-sensitivity to pRedET (oriR101). The temperature-sensitive RepA protein is essential for replication and partitioning of pRedET and is not functional at temperatures above 30 °C, so that the plasmid will be lost during culturing at higher temperatures (segregational loss).

To start the knock-out experiment the pRedET plasmid was transformed via electroporation into the respective *E. coli* strain. Therefore, 30 µL of an overnight culture in LB was used to inoculate 1.4 mL fresh LB (depending on the *E. coli* strain the addition of antibiotic could be necessary). The culture was incubated for 2 – 3 hours at 37 °C, shaking, before preparing them for electroporation. Cells were centrifuged at 11 000 rpm, 2 - 4 °C, 1 min and the supernatant was discarded and the pellet was resuspended in 1 mL ice-cold 10 % (v/v) glycerol. The centrifugation was repeated and cells were resuspended again. Cells were centrifuged once

more and the supernatant discarded so that 20 – 30 µL were left for final resuspension. 1 µL of the pRedET plasmid was mixed with the freshly prepared competent cells and electroporated (see 5.2.2.5). Cells were rescued with SOC medium followed by an incubation time of 70 min at 30 °C, shaking. All cells were plated on LB agar (+ Amp) and incubated at 30 °C for at least 15 hours.

Homologous recombination in the chromosomal *E. coli* DNA *in vivo*

One single *E. coli* colony containing pRedET was picked to inoculate an overnight culture in LB (+ Amp) incubated at 30 °C. Next day, 30 µL of pre-cultures were diluted in 1.4 mL LB (+ Amp) (this was done in duplicates) and incubated at 30 °C for around 2 hours until the OD_{600} reached 0.3. To induce expression of recombination proteins, 50 µL of 10 % (w/v) L-arabinose were added to each of the cultures (final 0.3 – 0.4 (w/v)), which were grown at 37 °C for 1 hour. Then cells were prepared for electroporation as described before and 200 – 400 ng of the linear functional cassette with homology arms were electroporated. Immediately 1 mL SOC medium was added and cells were grown at 37 °C for 3 hours, allowing recombination to occur. All cells were plated on LB agar (+ Kan) and incubated overnight at 37 °C.

Identification of *E. coli* with successfully modified genome

The next day up to 10. – 15 *E. coli* colonies were picked and streaked ontoa new LB agar plates (+ Kan) and inoculated in LB medium (+ Kan) to obtain stationary cultures. Both, plates and liquid cultures, were incubated at 37 °C overnight. The genomic DNA from these cultures was extracted using the GenElute™ Bacterial Genomic DNA Kit. To identify successfully modified colonies Q5-PCRs on each of the isolated genomic DNAs were performed according to 5.2.4.3 with primer pairs flanking the chromosomal insertion site (primer see Table 24). The PCR products were analyzed by agarose gel electrophoresis (see 5.2.4.4). Samples, showing a DNA band with the correct size of the integrated FRT-PGK-gb2-neo-FRT (around 1750 bp), were purified with Monarch PCR DNA Cleanup Kit and sequenced (see 5.2.4.9). Cryo-cultures were made for long-term storage of successfully engineered *E. coli* strains (see 5.2.2.3).

Optional: Removal of the selection marker using FLP recombination

For the generation of *E. coli* strains with multiple genomic deletions it was necessary to remove the Kan^R-selection marker before starting the next deletion round. This was realized by FLP-mediated site-directed recombination between the FRT sites integrated at the ends of the functional cassette. The 709-FLPe plasmid encodes the *flpe* gene for the FLPe recombinase (FLPe has an improved thermal stability and enhanced activity at 37 – 40 °C compared to wild type) under control of the heat labile repressor cl857. A temperature shift from 30 °C to 37 °C induced a transient expression of FLPe recombinase. Furthermore 709-FLP carries the *repA* gene and the oriR101 which is regulated by the temperature-sensitive RepA so that cultivation at 30 °C is necessary for plasmid maintenance as explained for pRedET.

The 709-FLPe was transformed into freshly prepared electro-competent *E. coli* (which were identified as successfully modified) as described for the pRedET-transformation. The cells were rescued with SOC medium and incubated at 30 °C for 1.5 – 2 hours before plating on LB agar plates containing 50 $\frac{\mu g}{mL}$ (half-strength) ampicillin and incubating for 24 hours at 30 °C (it was important to omit the antibiotic (Kan) the flanked cassette confers resistance to). A single colony was inoculated in 1 mL LB medium (important: also omit antibiotic of the FLPe-plasmid) and incubate at 30 °C, shaking for 2 – 3 hours. The temperature was shifted to 37 °C

and further incubated for 5 hours. In this time FLPe recombination occurred and segregational loss of the plasmids started. Finally, the culture was spread on LB agar plates without antibiotics and incubated at 37 °C overnight. Next day up to 25 colonies were picked and streaked on plates, first with antibiotic against which the eliminated marker conferred resistance to (Kan) and secondly a plate without antibiotic. Plates were incubated at 37 °C overnight. To identify desired *E. coli* the growth behavior on the LB plates with and without antibiotic were compared. Successfully recombinant clones showed no longer resistance to kanamycin, consequently they just grew on LB agar without antibiotics. These colonies were used for inoculation of cultures for genomic DNA extraction and further PCR analysis as already described. PCR products smaller than 100 bp (or no PCR product) marked *E. coli* which successfully excised genomic selection marker by FLPe recombination and could be used as host for additional gene deletion.

5.2.3. Mammalian Cell Culture Methods

5.2.3.1. Maintenance of HEK cells

HEK293T cells are an adherent cell line and were cultivated in tissue culture vessels (10 cm^2 dishes). They were grown at 37 °C with an atmosphere containing 5 % CO_2 in a special incubator. DMEM with 10 % FBS and 1x AB (= DMEM++) served as standard medium which was pre-warmed to 37 °C before use to avoid cell stress and apoptosis due to high temperature differences. To prevent confluency and to ensure cell maintenance, HEK293T cells should be passaged regularly (in a 1:5 ratio every two days (Mondays & Wednesdays) and in a 1:10 ratio on Fridays). HEK293T cells should always be handled with care to prevent their detachment from the dish.

5.2.3.2. Passaging and seeding of HEK cells

For passaging of HEK293T cells, first the old medium was removed and the cells were washed with 5 mL 1x PBS once. To detach cells from the dish 1 mL trypsin-EDTA solution was added and incubated for 2 – 3 min at 37°C, 5 % CO_2. The dish was carefully tapped on the side to detach all cells before resuspending them in 9 mL DMEM++.
In the case of seeding, the cell count was determined by an automated cell counter (Invitrogen, Thermo Fisher Scientific), as average of two samples. Therefore, 10 µL trypan blue were mixed with 10 µL cell suspension and mixture was transferred to a cell counting chamber for each sample. Table 25 depicts recommended seeding numbers which result in 70 – 80 % confluency on the next day. According to the cell count, HEK293T cells were added to new culture vessels containing fresh DMEM++.

Table 25 Conditions and applications for HEK cell experiments in different scales

Culture vessel	Cell seeding number	Medium volume	Application
24-well plate	100 000	0.5 mL	microscopy
6-well plate	600 000	2 mL	Microscopy Western Blot
10 cm^2 dish	3 500 000	10 mL	Microscopy Western Blot Protein Purification

For passaging (1:10), 1 mL of the cell suspension was transferred to a new 10 cm^2 dish and filled up with 9 mL DMEM++. For splitting in a 1:5 ratio, 2 mL cell suspension and 8 mL DMEM++ were used. An even cell distribution was reached by gently rocking back and forth the culture vessels, before incubating them at 37 °C, 5 % CO_2.

5.2.3.3. Transfection & protein expression in HEK cells

Standard amber suppression experiment

DNA used in mammalian cell culture was pure and endotoxin-free. Additives, like uAAs, HCl or NaOH, were sterile filtered. For amber suppression two plasmids had to be reversely co-transfected, one encoding a PylRS-mutant, the other carrying the amber-suppressed protein of interest. Both plasmids harbor four copies of the corresponding PylT.

One day before transfection cells were seeded according to Table 25 resulting in 60 – 80 % confluency and healthy cells ready for transfection. PEI (= polyethylenimine) was used as transfection reagent. PEI enables rapid delivery of plasmid DNA to cells by forming positively charged polyplexes. The medium did not have to be changed after transfection procedure because PEI is not cytotoxic. The uAA was added directly to the fresh DMEM++ before transfection. The amount of plasmid DNA necessary for different cultures sizes is listed in Table 26. The DNA amount consisted of a 3:1 ratio of plasmid encoding the amber-suppressed protein of interest to plasmid bearing the PylRS. To prepare the transfection solution, first PEI and then OptiMEM was added to the DNA, followed by vortexing and an incubation time of 15 min at 37 °C. In the meantime old medium was replaced with fresh DMEM++ containing the uAA. Depending on the uAA used, neutralization with 1 M HCl or 1 M NaOH could be necessary. This was indicated by a color-shift of the phenol red indicator (basic pH = purple; acidic pH = yellow). Finally, the transfection solution was added dropwise and cells were incubated for 24 – 48 h at 37 °C, 5 % CO_2 for recombinant protein expression.

Table 26 Conditions for the reverse transfection of HEK cells for standard amber suppression

Culture vessel	DNA amount	PEI	OptiMEM
24-well plate	0.5 µg	2.25 µL	100 µL
6-well plate	2 µg	9 µL	200 µL
10 cm^2 dish	10 µg	27 µL	1000 µL

5.2.3.4. Lysis of HEK cells and western blot analysis

Cell lysate of HEK293T cells for western blot analysis was prepared 24 – 48 hours after transfection from one well of a 6-well-plate. First cells were washed twice with 1x PBS (1 mL), before scraping off the cells with a cell scraper in 500 µL 1x PBS and transferring them into a reaction tube. Alternatively, cells could be detached from the plate using trypsin (according to 5.2.3.2), but this required two additional washing steps with 1x PBS to remove trypsin. The cells were harvested by centrifugation for 2.5 min at 500 g, 4°C (repeated once after turning the tube for 180°, so that a proper pellet is formed) and supernatant discarded. Cells were resuspended in 75 µL RIPA lysis buffer and vortexed every 10 min for 30 seconds during the 30 min incubation on ice. Alternatively, freeze-thaw lysis could be used. Therefore, HEK cells were resuspended in lysis buffer (detergent was omitted) and flash frozen in liquid nitrogen followed by thawing at rt and vortexing. This was repeated 5 times. To separate the cell debris

and nuclei, cells were centrifuged at 16 000 g for 15 min, 4 °C. The supernatant (= cleared lysate), which encompassed the cytosolic fraction, was transferred to a new reaction tube. Before loading 20 – 25 µL of the supernatant (25 – 30 µg of total protein) onto SDS-gels (15 %), it was mixed with 4x SDS loading buffer, cooked for 5 min at 98 °C and centrifuged for 3 min at 16 000 g. SDS-PAGE and Western blot was performed as described in 5.2.6.8 and 5.2.6.9.

5.2.3.5. Total protein amount of HEK cell lysates

The quantification of the total protein concentration of cleared lysates was important for further working steps, e.g. for calculating samples volumes for SDS-PAGE to insure even loading of total protein amounts. Therefore, the photometric BCA assay (Pierce™ BCA Protein Assay Kit; Thermo Fisher Scientific) a copper-based quantification method which relies on a colorimetric reaction, was used. Proteins (mainly cysteine or cystine, tyrosine and tryptophan residues, as well as the peptide bonds of the protein backbone) cause the reduction of Cu^{2+} to Cu^{+} in an alkaline solution (biuret reaction), which then forms purple-colored complexes with BCA. A BSA stock solution (2 $\frac{mg}{mL}$ BSA, Sigma-Aldrich) was used for preparing BSA-standards according to the manufacturer's protocol. The readout was performed according to manufacturer's microplate procedure but using 1.5 mL reaction tubes where 2 µL protein samples or BSA-standard-dilution and 40 µL BCA-working solution were mixed. Absorbance measurements were carried out in a photometer at 562 nm.

5.2.3.6. Protein Purification via Ni-NTA from HEK cells

His-tagged proteins produced in HEK293T cells were purified using Ni-NTA beads. Transfected HEK293T cells for protein expression, and further purification, were grown in 10 cm^2 dishes (number of dishes depends on the expression rates of the protein of interest) for 40 – 48 hours. Cells were washed twice with 1x PBS before being scraped off in 1 mL 1x PBS and transferred to a 2 mL reaction tube. Alternatively, cells could be trypsinized (according to 5.2.3.2), which required an additional washing step with 1x PBS before resuspending. The cells were pelleted by centrifugation at 500 g, 4°C for 2.5 min (centrifugation step was repeated after turning the tube for 180° for proper pelleting) and supernatant discarded. Cells were resuspended in 500 µl His-lysis buffer and incubated for 30 min on ice (during the incubation cells were vortexed every 10 min for 30 seconds). Alternatively, freeze-thaw lysis could be used (see 5.2.3.4). The cleared cell lysate was obtained by centrifugation for 15 min at 16 000 g, 4°C. In the meantime, Ni-NTA slurry (50 µL) was washed twice with His-washing buffer (750 µL). The cleared lysate (= supernatant after centrifugation) was transferred to the Ni-NTA slurry and incubated for 1.5 hours, 4°C on a rotating mixer followed by centrifugation for 2.5 min at 3500 g, 4 °C (centrifugation step was repeated after turning the tube for 180° for proper pelleting). The supernatant was discarded and the Ni-NTA beads were washed with 750 µl His-washing buffer. The washing step was repeated, three to four times, with centrifugation steps in-between (2.5 min at 3500 g, 4 °C). After washing, the supernatant was removed and 200 µL His-elution buffer was added. After incubating for 5 min on ice the beads were centrifuged for 2.5 min at 3500 g, 4 °C (centrifugation step was repeated after turning the tube for 180° for proper pelleting). Supernatant (= elution 1) was then transferred to a new reaction tube. If necessary, elution was repeated with 200 µL His-elution buffer (= elution 2). Instead of Ni-NTA beads, spin columns (Qiagen) could also be used. Therefore, the cleared

lysate was applied to the Ni-NTA columns and the procedure followed according to the manufacturer's protocol, except the buffers. As buffers the His-buffers for mammalian cells according to Table 21 were used. Samples could be directly used for SDS-PAGE, Western Blot. For further analysis by LC-MS the elution was filtered through a cellulose acetate membrane (0.22 µm pore size) and then concentrated via an Amicon Ultra 0.5 mL centrifugal filter (MWCO depends on protein size; for GFP use MWCO 10 kDa) to a volume of 20 µL - 50 µL was necessary. 5 – 10 µL concentrated protein solution were injected into the LC-MS. sfGFP produced in mammalian cells commonly shows a mass corresponding to cleavage of the initial methionine and N-terminal acetylation.

5.2.3.7. Microscopy of HEK cells

Morphology of cells, as well as the production of fluorescence proteins by HEK293T cells, was examined by microscopy. Visualization was performed with a Leica DMi8 fluorescence microscope and its corresponding Leica application suite X (LAS X) as well as Fiji software. Green GFP fluorescence was detected using the green channel (Excitation: bandpass filter 470/40 nm; Dichroic: 495 nm; Emission: 525/50 nm = GFP filter cube settings) and mCherry fluorescence was monitored by the red channel (Excitation: bandpass filter 560/40 nm; Dichroic: longpass filter 585 nm; Emission: bandpass filter 630/75 nm = TXR filter cube settings).

5.2.4. Molecular Biological Methods

5.2.4.1. DNA-isolation from *E. coli*

Isolation of genomic DNA from *E. coli*

Genomic DNA from gram-negative bacteria, such as *E. coli,* was isolated using the GenElute™ Bacterial Genomic DNA Kit. Therefore 1.5 mL stationary culture from the respective *E. coli* strain were needed. The procedure was carried out in accordance with the manufacturer's protocol for gram-negative bacteria. The DNA was finally eluted in MilliQ water (150 – 200 µL) instead of the elution buffer and stored at -20 °C.

Isolation of plasmid DNA from *E. coli*

E. coli strains like NEB 10-beta or MegaX DH10B $T1^R$ were mainly used for plasmid DNA isolation. Standard plasmid preparation from 6 mL overnight culture was carried out with the peqGOLD Plasmid Miniprep Kit, following the manufacturer's protocol, with the exception of eluting the plasmid DNA in water instead of TE buffer.

For larger preparation scales, e.g. 50 – 150 mL (Midi) or 150 – 250 mL (Maxi) culture volume, the Qiagen HiSpeed Plasmid Midi and Maxi Kits were used, according to the manufacturer's protocol. Again, the plasmid elution was performed with MilliQ water. The isolated plasmids were stored at -20 °C.

5.2.4.2. Determination of DNA-concentration

The measurement of DNA concentrations was performed using a NanoPhotometer® N60 (spectrophotometer). The principle is based on the characteristic absorption of DNA at 260 nm (one A_{260nm}-unit = 50 $\frac{\mu g}{mL}$ DNA). Moreover, the absorption ratios $\frac{260\ nm}{230\ nm}$ and $\frac{260\ nm}{280\ nm}$ are also calculated to indicate the purity of the DNA sample. $A_{260/280}$ reflects protein and RNA

contaminations ($A_{260/280}$ = 1.8 corresponds to pure nucleic acid), whereas $A_{260/230}$ shows the presence of organic solvents, salts or polysaccharides ($A_{260/230}$ = 2.0 - 2.2 corresponds to pure nucleic acid). The respective solvent of the DNA sample served as blank.

5.2.4.3. Polymerase Chain Reaction (PCR)

Routine PCR using Q5 HF DNA Polymerase

For the amplification of high-quality double-stranded DNA fragments, e.g. for cloning, the Q5 HF DNA Polymerase was used. This enzyme has a very low error rate (1.4 x 10^{-6} mutations per base per doubling) and a high reaction rate with 20-30 $\frac{s}{kb}$. Table 27 shows the pipetting scheme for routine Q5 PCR and Table 28 the respective cycling steps and conditions. If a PCR reaction did not deliver any or to many products using the standard conditions, the optimal T_a was screened. Further variable points were the primer or dNTP concentration and the Q5 enhancer usage.

Table 27 Sample preparation for Q5-standard PCR

Total volume	**25 µL**	**50 µL**
5x Q5 reaction buffer	5 µL	10 µL
Q5 enhancer	5 µL	10 µL
dNTP (10 mM each)	0.25 - 0.5* µL Fragment < 1500 bp = 0.25 µL Fragment > 2500 bp = 0.35 µL Fragment > 4500 bp = 0.5 µL	0.5 - 1.0* µL Fragment < 1500 bp = 0.5 µL Fragment > 2500 bp = 0.75 µL Fragment > 4500 bp = 1.0 µL
Forward primer (10 µM)	0.625 µL	1.25 µL
Reverse primer (10 µM)	0.625 µL	1.25 µL
Template DNA	10 ng	15 – 25 ng
Q5·HF DNA polymerase	0.25 µL	0.5 µL
MilliQ-H_2O (nuclease free)	Fill up to 25 µL	Fill up to 50 µL

* amount of dNTP depends on the length of amplified DNA fragment

Table 28 Protocol for Q5-standard PCR

Step	**Temperature**	**Time**	**# of cycles**
Initial denaturation	98 °C	1 min	1x
Denaturation	98 °C	15 s	32 - 35x**
Annealing	T_a [°C]*	30 s	
Elongation	72 °C	30 s/kb	
Final elongation	72 °C	5 or 10 min***	1x
Storage	12 °C	∞	1x

* T_a values were calculated using NEB T_m Calculator v 1.9.13 (NEB Inc., Ipswich, USA)

** # of cycles depended on DNA size; 35 cycles for DNA < 6000 bp, 32 cycles for ≥ 6000 bp

*** Final elongation was chosen based on elongation time: t < 2 min (5 min), t ≥ 2 min (10 min)

PCR samples were analyzed by agarose gel electrophoresis (see 5.2.4.4). Purification of PCR products, without undesired by-products, was performed using Monarch PCR DNA Cleanup Kit. Samples containing unspecific DNA bands had to be purified with the Monarch DNA Gel Extraction Kit to isolate the desired DNA fragment.

Touchdown-PCR using Q5 HF DNA Polymerase

For PCR reactions that, despite attempts to optimize, did not show the desired result (e.g. a lot of unspecific by-products), a so-called Touchdown-PCR was used. The pipetting scheme followed Table 27 but the cycling protocol differed (Table 29). This method increases the specificity of the PCR reaction by using a high initial T_a, which was gradually lowered per PCR cycle, normally until it reached the calculated T_a or some degrees below. Due to the high initial T_a more specific template molecules are enriched in the batch, before the cycling at low T_a started.

Table 29 Protocol for Q5-Touchdown-PCR

Step	Temperature	Time	# of cycles
Initial denaturation	98 °C	1 min	1x
Denaturation	98 °C	15 s	6x
Annealing	Initial T_a [°C]* = T_a + 3 °C Δ 0.5 °C/cycle or Δ 1 °C/cycle	30 s	
Elongation	72 °C	30 s/kb	
Denaturation	98 °C	15 s	28 x
Annealing	T_a or T_a – 3 °C	30 s	
Elongation	72 °C	30 s/kb	
Final elongation	72 °C	5 or 10 min	1x
Storage	12 °C	∞	1x

*initial T_{Ia} should be several degrees above the estimated T_a

Colony-PCR using Taq DNA Polymerase

Colony-PCR was used for the identification of promising candidates from cloning, which could be sent for sequencing. This screening method is just applicable for cloning of larger construct changes which can be detected by PCR band sizes in agarose gels or due to specific primer annealing, e.g. insertion of cleavage sites, affinity-Tags or genes. Colony-PCR is not suited to identify positive hits e.g. of point mutations derived via site-directed mutagenesis. The Taq-DNA Polymerase has an error rate per base and doubling of 2.7 x 10^{-4} and a velocity of 60 $\frac{s}{kb}$. The resulting DNA products from colony-PCRs served only for analytical agarose gels and not for further cloning steps.

Obtained *E. coli* colonies from respective cloning steps served directly as templates for colony-PCR. Therefore, colonies had to be picked and resuspended in 40 µL MilliQ water (template backup tube), before using them in adequate PCR reactions. The pipetting scheme is shown in Table 30 and the cycling protocol can be seen in Table 31.

Table 30 Sample preparation for colony-PCR using Taq-DNA-Polymerase

Total volume	**25 µL**
ThermoPol reaction buffer (10x)	2.5 µL
DMSO	1.0 µL
dNTPs (10 mM)	0.25 µL
Forward primer (10 µM)	0.25 µL
Reverse primer (10 µM)	0.25 µL
Template (resuspended colony)	5 µL
Taq DNA Polymerase	0.25 µL
MilliQ-H_2O	15.5 µL

Table 31 Protocol for Taq-colony-PCR

Step	**Temperature**	**Time**	**# of cycles**
Initial denaturation	95 °C	5 min	1x
Denaturation	95 °C	45 s	32x
Annealing	T_a [°C]*	30 s	
Elongation	72 °C	60 s/kb	
Final elongation	72 °C	5 min	1x
Storage	12 °C	∞	1x

* T_a values were calculated using NEB T_m Calculator v 1.9.13 (NEB Inc., Ipswich, USA)

Positive clones were identified by agarose gel electrophoresis (see 5.2.4.4) and could be directly inoculated for plasmids isolations using the appropriate template backup tube. The isolated plasmid DNA was finally sequenced (see 5.2.4.9) to confirm the integrity of the DNA sequence.

5.2.4.4. Agarose Gel Electrophoresis

For separating DNA samples, according to their size and conformation, agarose gel electrophoresis was the method of choice. The negatively charged DNA runs through the gel towards the anode due to an electric field, whereby the DNA separates according to its size. This can be used to examine a PCR reaction product or to analyze analytical or preparative DNA digests.

For the agarose gel electrophoresis, 1 % (w/v) agarose gels were used (see Table 32). Therefore, agarose was dissolved in 1x TAE buffer and heated (in a microwave) until the suspension was clear. Then DNA stain Clear G was added and mixture was panned carefully, before it was transferred to the gel chamber of the PerfectBlue systems and fitted with a suitable comb. The gel was cured after 15 - 20 min and was then transferred to the electrophoresis chamber, filled with 1x TAE buffer, followed by sample loading.

For the preparation of DNA samples, 10x DNA loading buffer was mixed in a 1:10 ratio with the DNA probe, before loading onto the gel. The DNA marker/standard (High Range DNA

Ladder, ready-to-use) was applied (10 µL) as size control. The separation occurred at 120 V. Analytical gels were run for 20 – 25 min, whereas preparative gels (e.g. for gel extraction) needed up to 40 min of running time. Samples, containing DNA/plasmids and libraries from directed evolution, required a duration of 1 – 1.5 hours. Visualization of the DNA bands in the gel was realized by the excitation of the fluorescence of DNA stain Clear G (which intercalates into the DNA) with UV light, using the Bio Imaging System GeneGenius.

Table 32 Composition of agarose gels for electrophoresis

Component	Mini gel (S)	Mini gel (L)	Midi gel (S)
Agarose	0.5 g	1.0 g	2.0 g
1x TAE	50 mL	100 mL	200 mL
DNA stain Clear G	3 µL	6 µL	10 µL
Final concentration	1 % (w/v)	1 % (w/v)	1 % (w/v)

5.2.4.5. Restriction Endonucleases & Cloning

Removal of template/parental DNA

Often it is necessary to remove template or parental DNA, e.g. from PCR products. Therefore, the restriction endonuclease DpnI was used because the enzyme only recognizes methylated DNA (methylated restriction site). PCR-amplified DNA does not carry methylations, whereas DNA isolated from dam^+ and/or dcm^+ bacterial strains is highly methylated. Consequently, DpnI selectively removes template plasmid DNA and unmethylated DNA remains intact. DpnI digest was performed according to table Table 33 at 37 °C for 1 – 2 hours. After digest the sample was purified using the Monarch PCR DNA Cleanup Kit.

Table 33 Sample preparation for DpnI digest

Total volume	Digest 25 µL	Digest 50 µL
Template/parental DNA (methylated)	Up to 1 µg in x µL	Up to 1 µg in x µL
DpnI (10 U/µL)	1 µL	1 µL
10x Cutsmart	2.5 µL	5 µL
MilliQ-H_2O (nuclease free)	21.5 µL - x µL	44 µL - x µL

Restriction cloning

Restriction cloning is the classical method for manipulating DNA sequences by inserting, replacing or removing DNA sections using restriction endonucleases. These enzymes cut double-stranded DNA at certain recognition sites (palindromic sequences mainly between 4 - 8 bps in size) leading to either blunt or sticky DNA ends. The whole procedure of restriction cloning consists of three steps: generation of DNA fragments (vector backbone and inserts) by restriction digest, dephosphorylation of the vector backbone and ligation of DNA fragments. Here, the enzyme volume should not extend more 10 % of the total reaction volume.

Analytical restriction digest

For the analysis of plasmids, an analytical restriction digest was used. This helped to identify promising candidates of successful cloning work. Usually, two different restrictions enzymes

were chosen for the test digest, resulting in a specific band size pattern visualized by agarose gel electrophoresis (see 5.2.4.4). The pipetting scheme for an analytical restriction digest is shown in Table 34. The digest was carried out at 37 °C for 1 hour before loading the total volume onto the agarose gel.

Table 34 Sample preparation for analytical digestion with restriction endonucleases

Total volume	**25 µL**
DNA	400 – 500 ng in x µL
Restriction enzyme 1 (10 U/µL)	1 µL
Restriction enzyme 2 (10 U/µL)	1 µL
10x Cutsmart or NEB 3.1 buffer	2.5 µL
MilliQ-H_2O (nuclease free)	20.5 µL – x µL

Preparative restriction digest

DNA fragments for restriction cloning were achieved, either by digesting a plasmid or by digesting PCR-products (respective restriction site could be introduced via PCR primer; located in primer overhangs). Combining restriction enzymes generating non-compatible DNA ends led to unidirectional or guided cloning (just one orientation of the insert integration is possible). For the preparative restriction digest 5 µg vector backbone template and 3 µg DNA insert were used. Table 35 presents the setup of the preparative restriction digest, which was carried out at 37 °C for 1.5 hours.

Table 35 Sample preparation for preparative digestion with restriction endonucleases

Total volume	**Digest backbone 100 µL**	**Digest insert 100 µL**
Vector backbone template	5 µg in x µL	-
DNA insert	-	3 µg in x µL
Restriction enzyme 1 (10 U/µL)	5 µL	3 µL
Restriction enzyme 2 (10 U/µL)	5 µL	3 µL
10x Cutsmart or NEB 3.1 buffer	10 µL	10 µL
MilliQ-H_2O (nuclease free)	80 µL - x µL	84 µL - x µL

To obtain purified backbone and insert fragments for cloning the respective DNA fragments had to be isolated. The isolation method depended on the used template DNA origin. The resulting linear DNA bands from a plasmid digest were separated with agarose gel electrophoresis (see 5.2.4.4) and the desired backbone or insert was isolated via gel extraction using Monarch DNA Gel Extraction Kit. For DNA fragments derived from digested PCR products DNA purification with the Monarch PCR DNA Cleanup Kit was sufficient.

Dephosphorylation of DNA fragments

To prevent re-circularization of the vector backbone, and thus false positive clones, the backbone was dephosphorylated with Antarctic phosphatase overnight at 37 °C. The pipetting scheme is shown in Table 36. The dephosphorylated backbone was purified using the Monarch PCR DNA Cleanup Kit and was than ready for ligation.

Table 36 Sample preparation for dephosphorylation reaction

Total volume	**30 µL**
Linearized vector DNA	16 µL (max. 5 µg vector)
Antarctic phosphatase (5 U/µL)	1 µL
10x Antarctic phosphatase buffer	2 µL
MilliQ-H_2O (nuclease free)	9 µL

Ligation of DNA fragments

Linearized vector backbone and insert were ligated via compatible sticky ends using T4 DNA Ligase in an overnight reaction at 16 °C. Alternatively, it was performed for 2 hours at room temperature. The pipetting scheme for the ligation reaction is summarized in Table 37. The amount of backbone DNA was supposed to be between 50 – 100 ng. The ratio of backbone:insert was determined by the sizes of the used DNA fragments. The smaller the insert compared to the backbone, the larger its excess. Because the ratio is based on pmol dimension of ds DNA, the concentration in $\frac{ng}{\mu L}$ had to be converted into $\frac{pmol}{\mu L}$. This could be performed with the Promega BioMath calculator, according to the following equation F2:

$$\text{(F2)} \qquad \text{pmol (DNA)} = \mu\text{g (DNA)} \times \frac{\text{pmol}}{660\ \text{pg}} \times \frac{10^6\ \text{pg}}{1\ \mu\text{g}} \times \frac{1}{\text{N}}$$

N = number of nucleotides

660 pg/pmol = the average molecular weight of a nucleotide pair

Table 37 Sample preparation for T4-ligation of two fragments

DNA ratio vector:insert [pmol]	**1:3**	**1:5**	**1:7**
Amount of linearized vector	100 ng in x µL	100 ng in x µL	100 ng in x µL
Amount of insert	y µL	y µL	y µL
T4 DNA ligase	1 µL	1 µL	1 µL
10x T4 DNA ligase buffer	2 µL	2 µL	2 µL
MilliQ-H_2O	17 µL - x µL - y µL	17 µL - x µL - y µL	17 µL - x µL - y µL
Total volume	20 µL	20 µL	20 µL

x = volume of vector; y = volume of insert

After ligation 5 µL of the reaction were directly transformed into chemically competent *E. coli* cells (NEB 10-beta or MegaX DH10B strain) according to 5.2.2.5.

5.2.4.6. Site-directed mutagenesis

Site-directed mutagenesis with Q5 HF DNA Polymerase

Site-specific mutations, e.g. point mutations, codon-substitutions or insertions/deletions of small DNA sequences, into double-stranded DNA (plasmids) were introduced by a method based on the Q5 Site-directed Mutagenesis Kit from NEB. The methods consisted of four steps: mutagenesis by PCR amplification, template removal, phosphorylation and ligation. In the NEB kit template removal, phosphorylation and ligation is performed in one working step, called KLD (kinase, ligase, DpnI) reaction.

First a Q5 standard PCR was set up according to 5.2.4.3 with 50 µL sample volume. The primer for the PCR were designed and the T_a determined using the NEBaseChanger webtool. The respective mutations were placed in the primer sequence, so that during the PCR amplification the mutations were implemented in the emerging linear PCR product. Forward and reverse primer did not overlap.

After successful PCR, the linear DNA product was purified with the Monarch PCR DNA Cleanup Kit (elution in 15 µL MilliQ water) and the template DNA was removed by DpnI digest (see Table 38) for 1 hour at 37 °C, followed by a second purification step (elution in 10 µL MilliQ water).

Table 38 Preparation of DpnI digest for site-directed mutagenesis using Q5-DNA Polymerase

Total volume	**20 µL**
DpnI	1 µL
10x Cutsmart	2 µL
Purified PCR product	14 µL
MilliQ-H_2O (nuclease free)	3 µL

The phosphorylation reaction with T4 Polynucleotide kinase was set up according to Table 39 and carried out at 37 °C for 30 min. The ligation with T4 DNA Ligase could be directly performed in the same reaction tube by adding the respective enzyme and fresh buffer (Table 39). Reaction occurred at 16 °C overnight.

Table 39 Preparation of the phosphorylation & ligation reaction for site-directed mutagenesis I

Total volume	**10 µL**
T4 Polynucleotide kinase	0.5 µL
10x T4 DNA ligase buffer	1 µL
Digested & Purified PCR product	2 µL
MilliQ-H_2O (nuclease free)	6.5 µL
➔ 30 min at 37 °C	

Total volume	**20 µL**
Phosphorylation reaction	10 µL
Add 10x T4 DNA ligase buffer	1 µL
T4 DNA Ligase	1 µL
MilliQ-H_2O	8 µL
➔ 16 °C overnight	

Alternatively, the three separated working steps (template removal, phosphorylation and ligation) could be combined in one reaction for 1 hour at room temperature (shown in Table 40).

Table 40 Preparation of the phosphorylation & ligation reaction for site-directed mutagenesis II

Total volume	**10 µL**
T4 Polynucleotide kinase	0.5 µL
T4 DNA ligase	0.5 µL
Dpn1	0.5 µL
10x T4 DNA ligase buffer	1 µL
Purified PCR product	2 µL
MQ-H_2O (nuclease free)	5.5 µL

Finally, 5 µL of the ligation product were transformed, without further purification, into chemically competent *E. coli* cells (NEB 10-beta or MegaX DH10B strain) (see 5.2.2.5). The plasmids DNA from obtained clones had to be verified using sequencing technologies (see 5.2.4.9).

5.2.4.7. Cloning *via* SLIM

Site-directed, ligase-independent mutagenesis (= SLIM)

SLIM is a PCR-based cloning method to easily modify plasmid elements (e.g. insertion of protein-Tags, substitutions of base pairs, deletions, etc.). For this two specially designed primer sets are needed. One primer set binds just upstream of the region to be modified and the other set downstream. Each primer set consisted of one primer containing the sequence to be introduced (= tailed primer; tail-region can be up to 30 bp) and second regular primer (= short). The "tailed"-primer of one set needs to be in forward direction, as in the other primer set it has to be vice versa[314]. SLIM included three working steps: Q5-PCR; DpnI digest and hybridization reaction.

To amplify two DNA-fragments, for the later hybridization, two separate Q5-PCRs were performed, each using one of the specially constructed primer sets. The PCR set-up for 50 µL samples and protocol took place according to 5.2.4.3. The result were two linearized copies of the desired DNA with a tail-region at either 5'- or 3'-end (PCR I and PCR II). These tail-regions served as overhangs for ligase-independent circularization during hybridization reaction. After

PCR the products were purified using Monarch PCR DNA Cleanup Kit, followed by DpnI digest (see 5.2.4.5). Finally, without further purification steps, the hybridization reaction was performed according to Table 41 and Table 42.

Table 41 Sample preparation for SLIM-hybridization reaction

Total volume	**50 µL**
H-buffer (5x)	10 µL
Digested PCR I (forward-tailed)	10 µL
Digested PCR II (reverse-tailed)	10 µL
MilliQ-H_2O	20 µL

Table 42 Protocol for the SLIM-hybridization reaction

Step	**Temperature**	**Time**	**# of cycles**
Initial denaturation	99 °C	3 min	1x
Denaturation	65 °C	5 min	3x
Hybridization	30 °C	40 min	
Storage/Hold	12 °C	∞	1x

The hybridization reaction (ca. 5 µL) was directly transformed into chemically competent *E. coli* cells (e.g. NEB 10-beta) (see 5.2.2.5) to obtain colonies the next day which than could be used for further analysis (e.g. sequencing).

5.2.4.8. Cloning *via* DNA-assembly

Cloning via NEBuilder HiFi DNA Assembly

For some cloning projects restriction cloning was impeded due to lacking suitable restriction site. To circumvent introducing such sites, cloning methods based on DNA fragment assembly via complementary overhangs could be used. These methods are independent of restriction sites and offer a more flexible way of cloning. Also, multiple fragment assembly in one isothermal assembly reaction is possible.

To be assembled DNA fragments have to contain respective complementary overlaps at the 5'-end and the 3'-end of 18 - 25 bp with an T_m between 50 - 60 °C to the appropriate following fragment. These overlaps were introduced by PCR primers containing respective DNA sequences at their 5'-end. DNA fragments were generated either by Q5 PCR (see 5.2.4.3) or by restriction digest (see 5.2.4.5). The term vector refers to the largest DNA fragment, whereas the smaller fragments are named inserts.

After isolation and purification of the DNA fragments using the Monarch PCR DNA Cleanup Kit or Monarch Gel Extraction Kit the DNA concentration in $\frac{ng}{\mu L}$ and amount in $\frac{pmol}{\mu L}$ was determined by Nanodrop spectrophotometer and Promega BioMath calculator (see 5.2.4.2 and

F2). The isothermal assembly reaction was performed at 50 °C for 60 – 75 min according to Table 43.

Table 43 Sample preparation of DNA-assembly reactions

	2-3 fragment assembly	
Total DNA content	0.1 pmol in x µL	0.1 pmol in x µL
DNA ratio vector: insert	1:2	1:3 or 1:5
Vector	50 ng	50 ng
Insert	> 200 bp	< 200 bp
NEBuilder HiFi Master-Mix (2x)	5 µL	5 µL
MilliQ-H_2O	5 µL - x µL	5 µL - x µL
Total volume	10 µL	10 µL

x = total volume of fragments

The assembly reaction was carried out with 50 ng vector. For inserts > 200 bp a 1:2 ratio of vector to insert was chosen. For inserts < 200 bp the ratio was adjusted to 1:3 or 1:56 vector:insert. After incubation 5 µL of the reaction were directly transformed into chemically competent *E. coli* NEB 10 B following the protocol in 5.2.2.5 and plated onto LB agar plates with appropriate antibiotics. Plasmid DNA of resulting clones was isolated via peqGOLD Plasmid Miniprep Kit and could be further analyzed or prepared for sequencing (see 5.2.4.9).

5.2.4.9. DNA-Sequencing

For the verification of, e.g. successful cloning, DNA sequencing was applied. Different companies providing a customer overnight sanger sequencing service. Depending on the company, the sample preparation varied a little. Sequencing primer are listed in Table 44.

GATC Biotech (now part of Eurofins Genomics GmbH, Ebersberg, Germany), required for its LIGHTRUN service 5 µL of purified template (plasmid: 80 - 100 $\frac{ng}{\mu L}$, PCR; product: 20 - 80 $\frac{ng}{\mu L}$) mixed with 5 µL sequencing primer (5 µM). Finally, the 1.5 mL reaction tube was labeled with a respective barcode. The sequencing results could be downloaded from the GATC Biotech or Eurofins Genomics homepage.

For overnight sequencing (Express Delivery service) with GENEWIZ Germany GmbH (Leipzig, Germany), 5 µL of the purified template (plasmid: 30 - 100$\frac{ng}{\mu L}$, PCR product: 10 – 50 $\frac{ng}{\mu L}$) and 5 µL sequencing primer (5 µM) were transferred in a 1.5 mL reaction tube and labeled with an appropriate barcode. The sequencing data was obtained from the GENEWIZ website.

Furthermore, the Mix2Seq Kit from Eurofins Genomics GmbH (Ebersberg, Germany) was used for overnight DNA sample sequencing. In the pre-labeled Mix2Seq tubes, 15 µL purified template (plasmid: 50 - 100$\frac{ng}{\mu L}$, PCR product: 150-300 bp 1 $\frac{ng}{\mu L}$; 300-1000 bp 5 $\frac{ng}{\mu L}$; >1000 bp 10 $\frac{ng}{\mu L}$) and 2 µL sequencing primer (10 µM) were mixed. The sequencing-data was downloaded from the Eurofins Genomics homepage.

Table 44 Sequencing primer

Primer	Sequence (5‘ → 3‘)	Target
S15	TGTGATGGCTTCCATGTCG	pEVOL_PylT
S16	CACTGCGTCTTTTACTGGC	pEVOL_araBAD
S17	ATCAGACCGCTTCTGCG	pEVOL_PylRS
S18	ATTCACCACCCTGAATTGAC	pET301 lac
S19	GGCAAACAAATTCTCGTCC	pBAD
S20	TGAAGCATTTATCAGGGTTA	pET301 Amp
S21	CCATAATCCTTGTTAGATTA	Mm_PylRS(2)
S22	GAAGCGGAATTAATTCGCGA	Mm_PylRS(2)
S27	GCGTCACACTTTGCTATG	ara promotor
S28	GGGAGAACCGTATATAAG	SE323/SE325 Plasmide
S29	AAGACGGCAATATGGTGG	SE323/SE325 Plasmide
S31	CCAGTCTTTCGACTGAG	rrnB1 terminator
S32	CGTAGAGGATCGAGATC	T7 in pET
S33	GAATTCCTCCTGCTAGCCC	pBAD_Duet_I
S34	GATCCGGGCCCTCTAGATG	pBAD_Duet_IIa
S35	GGGCTAGCAGGAGGAATTC	pBAD_Duet_IIb
S36	GCCTGGCAGTTCCCTACTC	pBAD_Duet_IIc
S37	CCTACAAAAGCACGCAAAC	pBK_PylRSrv
S38	CGTCAGAGGATCCTCGG	pBK_PylRSfv-D1
S40	GATGTAGGTGTTCCACAG	Ma-D65 tRNA
S41	CTTGACGGAGTAGCATAGG	Ma-sfGFP-CAT-tRNA
S42	CGTTTTGGATTGTTTGCTG	in glnS-terminator
S43	GCATAATGTGCCTGTCAAATGG	Seq_pPylT_tRNA
S44	AGTTGAAGGATCCTCGG	F_Seq_pBK_PylRS_D4
S45	TGTAAAACGACGGCCAGT	M13 fv
S46	TCAATGGGGTGGAGACTTG	EF-1 rv
S47	GATAAGCGTAGCGCATCAGG	pBK-PylRSrv (MjS3)

5.2.5. Libraries & Directed Evolution

5.2.5.1. Construction of PylRS-libraries

Construction of defined randomized PylRS libraries for directed evolution

The basis for the directed evolution approach is a DNA-library of a certain set of PylRS-genes. As template for the creation of a PylRS-library either the wild type PylRS or a PylRS mutant can serve. The PylRS-libraries in our lab are either randomized at defined amino acid positions in the substrate binding pocket of PylRS which are suspected to have an influence on the binding properties of the enzyme or randomly mutated by error prone mutagenesis. Randomization at defined positions can be achieved by using degenerated primers (NNK-primers). "N" can stand for adenine, thymine, cytosine and guanine bases, whereas "K" just encompasses guanine and thymine bases. The position to be randomized is encoded by this NNK-sequence which results in 32 possible codons that are translated to all 20 natural amino acids and the TAG-stop codon. According to the degeneration of the genetic code, some amino acids are presented by more than one triplet-codon resulting from the NNK-sequence. This might lead to an unequal distribution of amino acids at this position. Regarding the size (numbers of different members within the library) of libraries used in our lab we are strictly limited to the transformation efficiency of *E. coli* (~ 10^8 - 10^9 $\frac{cfu}{\mu g}$). Considering a 3-fold coverage (see F6) as requirement for subsequent directed evolution, up to five different positions can be randomized within one PylRS-library using NNK-primer pairs resulting in a library size of $(4x4x2)^5$ = ~ 3.36 x 10^7 members (3-fold coverage ~ 1 x 10 ^{8}) (see F3 + Table 45). Another way of randomization is based on library-primers synthesized by trimer-oligonucleotides.[315-321] Regarding the DNA-triplet code there exist 64 possible combinations of codons, but just 20 are required to cover all the natural amino acids. These defined 20 codons are used in the trimer-approach enclosing the variability at the randomization site to 20 codons compared to the 32 possible combinations using NNK-primer. By reducing the diversity within the individual randomization sites to 20, the number of library members decreases drastically. Consequently, the achievement of a complete coverage of the library size using *E: coli* is easier to implement. Compared to NNK-primer approach, library, based on trimer-oligonucleotides, with five randomized positions consists of 20^5 = 3.2 x 10^6 members (3-fold coverage 9.6 x 10^6) (see F4 + Table 45). Thus, even for *E. coli*-based evolution approaches, increasing the number of randomized codons up to six within one library is possible (timer-based library with six randomized positions: 20^6 = 6.4 x 10^7 members (3-fold coverage 1.92 x 10^8); F4 + Table 45). Additionally, no stop-codon production leads to unfunctional library-members (incomplete proteins). Moreover, codon bias is circumvented. A quite similar result can be obtained by ordering 20 equivalent primers which only differ in the codon at the randomization site and mixing them in equivalent amounts. For each natural amino acid one respective codon is used, normally the same as used during the trimer-synthesis (see Table 46).

Table 45 Dependence of the library sizes from randomized positions and codon-usage in primers

Number of randomized codons (N)	Library size (NNK-based) 32^N	3-fold coverage# Library size (NNK-based)	Library size (Trimer-based) 20^N	3-fold coverage# Library size (Trimer-based)
1	32	96	20	60
2	1 024	3 072	400	1 200
3	3.28×10^4	9.84×10^4	8 000	2.4×10^4
4	1.05×10^6	3.15×10^6	1.6×10^5	4.8×10^5
5	3.36×10^7	1.01×10^8	3.2×10^6	9.6×10^6
6	$1{,}07 \times 10^9$ *	3.21×10^9 *	6.4×10^7	1.92×10^8

* not suitable for *E. coli*-based directed evolutions due to limited transformation efficiency

3-fold coverage of the library size is requirement for directed evolution approaches

Table 46 Overview of the codon-usage for the different 20 AA in library-primers

Amino acid	One letter code	5'→3' Triplet for Trimer-synthesis
Lys	K	AAA
Asn	N	AAC
Thr	T	ACC
Ile	I	ATC
Met	M	ATG
Gln	Q	CAG
His	H	CAT
Pro	P	CCA
Arg	R	CGT
Leu	L	CTG
Glu	E	GAA
Asp	D	GAT
Ala	A	GCA
Gly	G	GGT
Val	V	GTT
Tyr	Y	TAC
Ser	S	TCT
Cys	C	TGC
Trp	W	TGG
Phe	F	TTC

The equitation to calculate the parameters for a certain library are summarized below.

$$\text{(F3)}\quad \text{Theoretical library size (NNK}-\text{library)} = (4 \times 4 \times 2)^x = 32^x$$

$$\text{(F4)}\quad \text{Theoretical library size (Trimer}-\text{library)} = 20^x$$

x = number of randomized codons within the library

$$\text{(F5)}\quad N = \frac{\ln(1-\text{confidence})}{\ln\left(1-\left(\frac{1}{\text{theoretical size}}\right)\right)}$$

N = Number of clones needed to reach the aimed confidence
Confidence = % of statistical confidence, e.g. 0.9 (90 %)
Theoretical size = Theoretical number of members within the library (F3 or F4)

$$\text{(F6)}\quad coverage = \frac{obtained\ colony\ number}{Theoretical\ library\ size} \geq 3$$

Table 47 Primer used for library constructions

Primer	Sequence (5‘ → 3‘)	Target
Ella1	GGAAAGGTCTCTAGCCTGX01CGTGAATGGGGCX01GATX01CCGX01ATTGGCGCGGGTTTTGGCC	Lib1-*Mb*-MF18 - D373 + I378 + K380 + W382
Ella2	GGAAAGGTCTCAGGCTAACCGGACCCACCACCG	Lib1-*Mb*-MF18 - D373 + I378 + K380 + W382
Ella3	GGAAAGGTCTCTTCGCAAX01GGCAGCGGCTGCACCCGTG	Lib1-*Mb*-MF18 – M315
Ella4	GGAAAGGTCTCTGCGAAAACTGAACCATGGTGAATTCTTC	Lib1-*Mb*-MF18 – M315
L24	GTTTTCGCAANNKGGCAGCGGCTGCAC	Lib2-*Mb*-MF18 – M315
L25	TGAACCATGGTGAATTCTTC	Lib2-*Mb*-MF18 – M315
L26	TCCGATCAAANNKTTTGAAGTGGGCCCG	Lib2-*Mb*-MF18 – I287
L27	CCCGGCAGAATACGATCC	Lib2-*Mb*-MF18 – I287
L28	TGAATGGGGCNNKGATAAACCGTGGATTG	Lib2-*Mb*-MF18 – I378
L29	CGATCCAGGCTAACCGGA	Lib2-*Mb*-MF18 – I378
L30	CCCGACCCTGNNKAACTATGCGCG	Lib3-*Mb*-MF18 – Y271
L31	GCCAGCATCGGACGCAGG	Lib3-*Mb*-MF18 – Y271
L32	CACCATGGTTNNKTTTTCGCAAATG	Lib3-*Mb*-MF18 – N311
L33	AATTCTTCCAGGTGTTCTTTG	Lib3-*Mb*-MF18 – N311
L34	GCTTTGAGGAATCCCATATGATGG	Amplification of *Mb*-PylRS during EP-PCR
L35	GGAAAAGCTTTTACAGGTTCGTGCTAATGCC	
L36	GGAAAAGCTTGTTTCAAACGCTAAATTGCC	Backbone amplification for EP-product
L37	CATCATATGGGATTCCTCAAAGC	
L38	GCTTTGAGGAATCCCATATGATGG	Amplification of *Ma*-PylRS during EP-PCR
L39	GGAAAAGCTTTTACAGGTTCGTGCTAATGCC	
LMa1	GTTNNKCGTGATCTGCGCGATCATA	Lib2-*Ma*-MF18IP – Y126 + M129

LMa2	GCTKNNCAGATTCGGTGCCAGCAT	Lib2-*Ma*-MF18IP – Y126 + M129
LMa3	GNNKGATATGGGTCCGCGTGGT	Lib2-*Ma*-MF18IP – N166 + V168
LMa4	AGKNNCAGCATGGTAAATTCTTCCAGATG	Lib2-*Ma*-MF18IP – N166 + V168
LMa5	TCCGGTTAAANNKTTTGAAATGGGTAGCTG	Lib1-*Ma*-MF18IP – M170
LMa6	CCATCGGTATGATCGCGC	Lib1-*Ma*-MF18IP – M170
LMa7	GCTGTCTGATNNKGGTCCGCGTG	Lib1-*Ma*-MF18IP – I142
LMa8	TGCAGCATGGTAAATTCTTC	Lib1-*Ma*-MF18IP – I142
LMa9	AGCACATGATNNKCATGAACCGTGGTCAG	Lib1-*Ma*-MF18IP – V235
LMa10	GCATCCAGCGGAATCGGA	Lib1-*Ma*-MF18IP – V235

X01 = randomized position via trimer-oligonucleotides according to the codon-usage in Table 46
NNK = randomized position of the degenerated primer

In general, the procedure to create PylRS-libraries consists of successive individual randomization rounds, depending on the number and localization of randomized codons. Randomization sites which can be covered in one primer can be randomized simultaneously in one round. Primer pairs for library creation are listed in Table 47. Success of each individual randomization process was investigated by transformation and DNA-sequencing of a selection of library outcomes.

Construction strategy I - Inverse PCR, BsaI-digest, T4-DNA-ligation

Each randomization round included an inverse PCR followed by BsaI-treatment and T4-DNA ligation. For the PCR special primer pairs with BsaI restriction sites at their 5'-end were designed. For the design of the forward primer the residues which to be randomized were located first. Then after traveling -7 base pairs (upstream) of the randomization site (2 bp = spacer + 4 bps cutting site + 1) the complete sequence from -7 to approximately +25 bps after (downstream) the randomization site was copied. The primer is supposed to terminate with a G or C. Finally, the BsaI restriction site was added to the 5'-end and a random sequence of 5 bps in front of it. The randomization site was either marked by NNK or trimer-oligonucleotide (X01) placeholder. For the reverse primer the reverse complement sequence from -2 to -26 (upstream) of the randomization site was copied and added at the 5'-end the BsaI recognition site together with five random bases. Again, the primer is supposed to end with a G or C. For PCR amplification standard Q5 PCR was chosen (see 5.2.4.3) with an elongation time of 3 minutes. Due to primer size high annealing temperatures were required (sometimes > 72 °C). If standard Q5 PCR did not yield the desired PCR product touchdown-PCR with Q5 were performed using 72 °C as initial annealing temperature (going from 72 °C to 70 °C in 4 cycles (- 0.5 °C/cycle)) (see 5.2.4.3). Introducing the first randomized site required five PCR reactions with 50 µL each. For the second and third randomized position 10 – 15 PCR reaction with 50 µL each were necessary for the individual positions. For any further randomized site at least 20 reactions per position were carried out. The PCR products were analyzed by agarose gel electrophoresis (see 5.2.4.4) and either purified with Monarch PCR DNA Cleanup Kit or gel extracted using Monarch Gel Extraction Kit according to manufacturer's protocol. DNA

concentration was determined using the Nanodrop and the simultaneous digest with BsaI and DpnI which was performed according to Table 48 at 37 °C for 2 – 3 hours.

Table 48 Preparation of DpnI digest in library construction I

Total volume	**Digest 120 µL**
PCR product	Up to 10 µg in x µL
DpnI (10 U/µL)	2 µL
BsaI-HF (10 U/µL)	12 µL
10x Cutsmart	12 µL
MilliQ-H_2O (nuclease free)	96 µL - x µL

The digested DNA was purified using the Monarch PCR DNA Cleanup Kit and was ready for the ligation reaction. Therefore, the DNA concentration in the reaction mixture was adjusted to 5 $\frac{ng}{\mu L}$ to avoid the formation of DNA concatemers during ligation. The reaction (see Table 49) occurred at 16 °C for 36 hours.

Table 49 Preparation of ligation reaction in library construction I

Compound	**Ligation**
Digested PCR	1 µg (Final concentration 5 $\frac{ng}{\mu L}$) in x µL
10x T4 DNA ligase buffer	20 µL
T4 DNA Ligase	5 µL
MilliQ-H_2O	175 µL - x µL
Total volume	200 µL

The DNA of the ligation reaction was purified by Monarch PCR DNA Cleanup Kit. 1 µg of the library DNA from the randomization was transformed in freshly prepared electrocompetent *E. coli* NEB 10-beta cells. Therefore, a pre-culture of *E. coli* NEB 10-beta in 2x YT medium was diluted to an OD_{600} of 0.05 and incubated at 37 °C until the OD_{600} reached 0.4 – 0.5. Cells were cooled for 10 min on ice followed by pelleting via centrifugation (4200 rpm, 4 °C, 10 min). The pellet was resuspended in 100 mL ice-cold 10 % (v/v) glycerol and again centrifuged (4200 rpm, 4 °C, 10 min). This washing step was repeated twice. Then the pellet was finally resuspended in 100 – 200 µL 10 % (v/v) glycerol. 1 µg of the desired library was split into 4x 50 µL of the freshly prepared competent *E. coli* and electroporated according to 5.2.2.5. Transformed cells were rescued with 950 µL SOC medium each and incubated for 1 hour at 37 °C before they were inoculated in 1 L pre-warmed 2xYT medium containing the antibiotic for the respective library. The culture was incubated at 37 °C for 12 - 16 hours. After 15 min incubation a 1 mL sample was taken for the dilution experiment to determine the coverage of the library transformation. Therefore, a dilution series from 10^{-4} up to 10^{-8} was created in 2x YT medium. The 10^{-4}-dilution was directly reached by plating 100 µL of the 1 mL-culture sample on LB agar (+ antibiotic of the transformed library). Further dilutions were

prepared by mixing 100 µL of the previous level of dilution with 900 µL medium and plating of 100 µL of each dilution on respective LB agar. This was continued until the 10^{-8} level was reached. The plates were incubated at 37 °C for 12 – 16 hours. The next day the resulting colonies were counted, and the coverage and confidence were calculated (see F3 - F6). If a 3-fold coverage of the library members from the respective randomization round and a confidence > 90 % was obtained the 1 L-culture was used for plasmid isolation of the library DNA using the HiSpeed Plasmid Midi Kit. The isolated DNA was used was PCR-template for the next randomization round. This was continued until all randomization sites were covered. During each randomization round 10 – 15 colonies resulting from the transformation were used for plasmids miniprep and further DNA sequencing to analyze the randomization on DNA level.

Construction strategy II – Q5-site-directed mutagenesis

A second approach for the creation of PylRS-libraries was based on the Q5-site directed mutagenesis strategy. This method showed very good results for the introduction of point mutations and insertion or deletion of small DNA sequences. The workflow encompassed Q5-PCR followed by DpnI digest and T4 PNK mediated phosphorylation and T4-ligation as described in 5.2.4.6. The primer pairs for these Q5-PCRs were quite smaller than for the strategy using Inverse PCR, BsaI-digest and T4-DNA-ligation what rendered the PCR easier to handle due to much lower annealing temperatures and secondary primer structures effects. Primers were designed using the NEBaseChanger webtool. The bases at the randomization site were coded by NNK nucleotides. For DNA amplification standard Q5-PCR were used according to 5.2.4.3.

The amount of PCR reaction needed per randomized position was equivalent to strategy I. The PCR product was analyzed and purified before setting up the DpnI digest (see Table 50) which took place at 37 °C for 2 hours.

Table 50 Preparation of DpnI digest in library construction II

Total volume	**50 µL ***
DpnI	2 µL
10x Cutsmart	5 µL
Purified PCR product	X µL
MilliQ-H_2O (nuclease free)	43 µL - x µL

* per 50 µL digest was used for the product of 8 – 10 PCR reaction à 50 µL

Before starting the phosphorylation (see Table 51), the digested DNA was purified using Monarch PCR DNA Cleanup Kit. The T4 PNK reaction was incubated at 37 °C for 1 hour. Then the T4 DNA ligation could be directly performed in the same reaction tube by adding the respective enzyme and fresh buffer (see Table 51). After purification of the ligated DNA the whole procedure for the transformation, dilution experiment and verification of the randomization was carried out as described for above for strategy I.

Table 51 Preparation of phosphorylation and ligation reaction in library construction II

Total volume	**50 µL**
T4 Polynucleotide kinase	0.5 µL per 100 ng blunt ends
10x T4 DNA ligase buffer	5 µL
Digested & Purified PCR product	x µL
MilliQ-H_2O (nuclease free)	Fill up to 50 µL
➔ 60 min at 37 °C	
Total volume	**100 µL**
Phosphorylation reaction	50 µL
Add 10x T4 DNA ligase buffer	5 µL
T4 DNA Ligase	0.5 µL per 100 ng blunt ends
MilliQ-H_2O	Fill up to 100 µL
➔ 16 °C overnight	

Construction of Error Prone PylRS libraries for directed evolution

Random mutagenesis of PylRS gene for libraries was achieved by error prone PCR. The method was developed by Cladwell & Joyce (1992)[322] where certain PCR conditions result in the enhancement of the error rate of the DNA polymerase during PCR. For the method non-proofreading polymerases are used like the Taq DNA polymerase. The parameters which provoke the enhanced mutation rate are the addition on Mn^{2+}, the increase of Mg^{2+} and the use of unbalanced amounts of dNTP during PCR. The number of doublings during the error-prone PCR determines the degree of mutagenesis. In general mutagenesis rates of 0.6 – 2.0 % are archived.[322, 323]

For the creation of error-prone PylRS-libraries the JBS Error-Prone Kit was used. First the PylRS gene had to be amplified, the template was removed using DpnI and the final PylRS-EP-library-bearing plasmids were archived via restriction cloning. The PylRS gene was mutated during amplification with error prone PCR (see Table 52 and Table 53) using the primer pair L34+L35 or L38+L39 from Table 47 which was located outside the PylRS gene and harbored restrictions sites at their 5'-ends. The backbone for the mutated PylRS genes was derived via Q5-Standard PCR according to 5.2.4.3 using the primer pair L36 and L37 (see Table 47).

Table 52 Sample preparation for Error-Prone PCR

Total volume	**50 µL**
10x EP Reaction buffer	5 µL
dNTP EP-Mix	2 µL
Taq DNA Polymerase	0.6 µL
Forward Primer (10 µM)	5 µL

Reverse Primer (10 µM)	5 µL
Template (plasmids encoding respective PylRS variant)	50 – 100 ng
MilliQ-H_2O (nuclease free)	Fill up

Table 53 Protocol for Error-Prone PCR

Step	Temperature	Time	# of cycles
Initial denaturation	94 °C	1 min	1x
Denaturation	94 °C	30 s	25x
Annealing	T_a [°C]*	30 s	
Elongation	72 °C	60 s/kb (90 s)	
Final elongation	72 °C	3 min	1x
Storage	12 °C	∞	1x

The products of the EP-PCR and the Q5-PCR were purified using the Monarch PCR DNA Cleanup Kit and the purified DNA was digested 2 hours at 37 °C with DpnI according to Table 54. The digest was purified with the Monarch PCR DNA Cleanup Kit.

Table 54 Preparation of DpnI digest for Error-Prone PCR products

Total volume	50 µL *
DpnI	2 µL
10x Cutsmart	5 µL
Purified PCR product	X µL
MilliQ-H_2O (nuclease free)	43 µL - x µL

* 2 µL DpnI were used per 100 ng of template DNA

Next step was the restriction cloning according to the procedure described in 5.2.4.5 using HindIII-HF and NdeI. For T4-ligation a backbone to insert ratio of 1:3 was chosen and the final DNA concentration was adjusted to 5 $\frac{ng}{\mu L}$. Before the error prone library was suitable for evolution a final purification with the Monarch PCR DNA Cleanup Kit was performed.

5.2.5.2. Directed Evolution of PylRS in *E. coli* for new uAAs

General requirements for directed evolution approaches

In order to find PylRS mutants which accept new uAAs a directed evolution approach was used. This method is based on libraries of the PylRS wild type or mutants from *M. barkeri* or *M. alvus*, where up two five positions in the binding site were randomized. This approach encompasses a three-step selection process. First a positive selection step, afterwards a negative selection followed by a fluorescence readout step which might be coupled to a second positive

selection. The plasmids used for directed evolution are listed in Table 55 and the libraries used in this work are listed in Table 56.

Table 55 Plasmids used for directed evolution

Plasmid	Application	Selection marker	Resistance
Mb-D65 (10.3 kb)	Positive selection	Chloramphenicol acetyltransferase with TAG-codon at position 111 (CAT111TAG)	Tet
Mb-D7L (7.4 kb)	Negative selection	Barnase with TAG-codons at positions 3 and 45 (barnaseTAG,45TAG)	Cam
Mb-pPylT_sfGFP150TAG (6.2 kb)	Fluorescence readout	sfGFP with TAG-codon at position 150 (sfGFP150TAG)	Tet
Mb-pPylT_sfGFP150TAG-CAT (7.7 kb)	Fluorescence readout coupled with Positive selection	sfGFP with TAG-codon at position 150 and Chloramphenicol transferase with TAG-codon at position 111 (sfGFP150TAG CAT111TAG)	Tet
Ma-D65 (10.3 kb)	Positive selection	Chloramphenicol acetyltransferase with TAG-codon at position 111 (CAT111TAG)	Tet
Ma-D7L (7.4 kb)	Negative selection	Barnase with TAG-codons at positions 3 and 45 (barnaseTAG,45TAG)	Cam
Ma-pPylT_sfGFP150TAG-CAT (7.7 kb)	Fluorescence readout coupled with Positive selection	sfGFP with TAG-codon at position 150 and Chloramphenicol transferase with TAG-codon at position 111 (sfGFP150TAG CAT111TAG)	Tet

Table 56 Libraries for directed evolution of PylRS

Library	Randomized positions	Origin	Creation
Lib1-*Mb*-MF18	M315; D373; I378; K380; W382	*Mb*-MF18-L274A-N311Q-C313S (Amp)	Inverse PCR, BsaI-digest, T4-DNA-ligation (Timer-Primer)
Lib2-*Mb*-MF18	I287; M315; I378	*Mb*-MF18-L274A-N311Q-C313S (Amp)	Q5-site directed mutagenesis (NNK-Primer)
Lib3-*Mb*-MF18	Y271; N311	*Mb*-MF18-L274A-N311Q-C313S (Amp)	Q5-site directed mutagenesis (NNK-Primer)
EP-Lib-*Mb*-MF18	-	*Mb*-MF18-L274A-N311Q-C313S (Amp)	Error Prone PCR
Lib1-*Ma*-MF18IP	I142; M170; V235	*Ma*-MF18-H227I-Y228P (Amp)	Q5-site directed mutagenesis (NNK-Primer)

Lib2-*Ma*-MF18IP	Y126; M129; N166; V168	*Ma*-MF18-H227I-Y228P (Amp)	Q5-site directed mutagenesis (NNK-Primer)
EP-Lib-*Ma*-MF18IP	-	*Ma*-MF18-M129A-N166Q-V168S-H227I-Y228P (Amp)	Error Prone PCR

The plasmids *Mb*-D65, *Mb*-D7L, *Mb*-pPylT_sfGFP150TAG and *Mb*-pPylT_sfGFP150TAG-CAT were used for directed evolutions on libraries based on *Mb*-PylRS, as they carry a version of *Mb*-tRNA$_{CUA}$. *Ma*-D65, *Ma*-D7L and *Ma*-pPylt_sfGFP150TAG CAT harbor an evolved *Ma*-tRNA$_{CUA}$, which makes them suitable for libraries based on *Ma*-PylRS. The plasmids for directed evolution, were chemically transformed into *E. coli* NEB 10-beta cells and plated onto LB agar with appropriate antibiotics (see 5.2.2.5). The resulting colonies were used for inoculation of *E. coli* cultures to generate competent cells for the individual selection steps. Plates were stored at 4 °C for a maximum of one month. The first critical step in the directed evolution process is the transformation of the library into cells, already containing the positive selection plasmid D65, since there is the risk of losing library members due to insufficient transformation efficiency. To ensure covering all possible members of the library, the statistical significance of the coverage was calculated with F5 with an assumption of 90 % confidence. In general, the theoretical library size of a library with x randomized codons was calculated with equation F3, as this library was generated by NNK-primer approach. Libraries with five randomized position have a theoretical size of $32^5 = 3.3x\ 10^7$ library members, so that a 2.3-fold coverage is needed to reach a confidence of 90 %. The theoretical size of libraries which were randomized by using just one codon per amino acid and x randomized positions was calculated using equation F4. Such libraries with five randomized position have a theoretical size of $20^5 = 3.2x\ 10^6$ library members.

Directed evolution – Positive selection

The *Mb*-D65 or *Ma*-D65 plasmids for the positive selection encode constitutively expressed chloramphenicol transferase with a TAG-codon at position 111 and either a mutant *Mb*-tRNA$_{CUA}$ or *Ma*-tRNA$_{CUA}$. Principle of the positive selection is, that only *E. coli* cells expressing PylRS-library members which are able to incorporate either a natural or the desired unnatural amino acid at TAG111 of the chloramphenicol transferase, survive on the positive selection plates with chloramphenicol.

For positive selection, a stationary overnight culture in 2xYT (+ Tet) (50 mL) at 37 °C of *E. coli* NEB 10-beta-*Mb*-D65 or -*Ma*-D65 was derived from a single colony. This culture was diluted in fresh 2xYT medium (+ Tet) (100 mL) to an OD_{600} of 0.05 and cultivated at 37 °C to an OD_{600} around 0.4 – 0.5. Then, cells were ready to make them electro-competent for library transformation. First, cells were cooled for 10 min on ice followed by pelletizing via centrifugation (4200 rpm, 4 °C, 10 min). The pellet was resuspended in 100 mL ice-cold 10 % (v/v) glycerol and again centrifuged (4200 rpm, 4 °C, 10 min). This washing step was repeated twice. Then, the pellet was finally resuspended in 100 – 200 µL 10 % (v/v) glycerol and the cell suspension ready for electro-transformation. 1 µg of the desired library was split into

4x 50 µL of the freshly prepared competent cells and electroporated according to 5.2.2.5. After incubation for 1 hour in SOC medium, the cells were inoculated in 1 L pre-warmed 2xYT medium (+Tet + antibiotic of the used library) and incubated at 200 rpm, 37 °C for 12 – 16 hours. After 15 min incubation the dilution experiment to was set-up to determine the coverage of the library transformation as aforementioned in 5.2.5.1. Therefore, 1 mL of the culture was taken to perform a dilution series from 10^{-4} up to 10^{-8}. The plates were incubated at 37 °C for 12 – 16 hours. The next day the resulting colonies from plates with suitable dilutions were counted and the coverage was calculated (see F6). If the desired coverage of the library was not reached the transformation had to be repeated. For successful library transformation the 1 L-culture was diluted in 100 mL 2xYT medium (+ antibiotics) to an OD_{600} around 0.05 and incubated until they reached an OD_{600} of 0.6. Two 10 mL samples of the cells were taken, one without uAA and the other was supplemented with the uAA (final concentration 2 – 4 mM depending on the uAA) and grown at 37 °C for 3 hours. Before spreading 600 – 1000 µL of the culture, supplemented with uAA, on one positive selection plate (see Table 57) the OD_{600} was determined and should be around 2-3. The plates were incubated at 37 °C for 48 hours and surviving colonies were scraped off using 50 mL 2xYT (+Tet + antibiotic of the used library). The resulting cell suspension was cultivated at 37 °C for 2 hours, pelleted and the plasmid DNA was isolated using Qiagen HiSpeed Plasmid Midi Kit. To separate the library-DNA (3.3 kb) from the positive selection plasmid (10.3 kb) an agarose gel electrophoresis (120 V; 1.75 hours) was performed with 5 µg of the isolated DNA. The DNA band corresponding to the library was extracted with the Monarch DNA Gel Extraction Kit and was used for negative selection.

Table 57 Composition of positive-selection plates in directed evolution

Components	**Comments**
40 mL 5x M9 salts	Plates were always prepared freshly.
200 µL 1 M $MgSO_4$	Agarose were heated until liquid and cooled in a 50 °C water bath. Be sure that agarose is not too hot to prevent the destruction of heat-labile compounds.
60 µL 1 M $CaCl_2$	
200 µL 40 % (w/v) Glucose	Other components were mixed and pre-warmed in a 50 °C water bath.
2 mL Leucine (4 $\frac{mg}{mL}$)	
40 µL 5000x Trace metal mix	Agarose was combined with the other components and plates were immediately poured.
200 µL Biotin (1 $\frac{mg}{mL}$)	
200 µL Thiamin (1 $\frac{mg}{mL}$)	Concentration of the uAA depended on the respective uAA and was between 2 – 4 mM.
X mL uAA (100 mM)	
240 µL Cam	
200 µL Tet	
200 µL Antibiotic of the library	
100 mL 5 % (w/v) Agarose	
Fill up to 200 mL MilliQ-H_2O	

Directed evolution – Negative selection

For the negative selection the *Mb*-D7L or *Ma*-D7L plasmid was used. The plasmid encodes a constitutively expressed mutant of *Mb*-$tRNA_{CUA}$ or *Ma*-$tRNA_{CUA}$ and an arabinose-inducible

barnase gene with two TAG-codons at positions 3 and 45. The negative selection is based on the toxic barnase protein, which is only built when expressed PylRS-library members incorporate a natural amino acids in response to the TAG-codons. *E. coli* harboring these PylRS-library mutants die during the negative selection process, so only PylRS-members survive which do not accept any of the natural amino acids.
For negative selection a stationary overnight culture in 2xYT (+ Cam) (50 mL) at 37 °C of *E. coli* NEB 10-beta-*Mb*-D7L or -*Ma*-D7L was derived from a single colony. The pre-culture was diluted and used for the preparation of electro-competent *E. coli* NEB 10-beta-*Mb*-D7L or -*Ma*-D7L, which were prepared and transformed according to the description for the positive selection. The purified library-DNA (100 - 200 ng; derived from the positive selection) was electroporated in 2x 50 µL of the freshly prepared competent *E. coli* and rescued with 900 µL SOB medium. The cells were incubated for 1 hour at 37 °C and combined before plating 600 µL and 1400 µL onto separate negative selection plates (see Table 58). The plates were incubated at 37 °C for 12 - 16 hours and surviving colonies were scraped off with 50 mL 2xYT (containing appropriate antibiotics). The cell suspension was cultivated at 37 °C for 2 hours and the plasmids DNA was isolated with the Qiagen HiSpeed Plasmid Midi Kit. The separation of library-DNA (3.3 kb) and negative selection plasmid (7.4 kb) occurred via agarose gel electrophoresis (120 V; 1.75 hours). Therefore 5 µg of the isolated plasmid DNA was used and the DNA band at 3.3 kb was extracted with Monarch DNA Gel Extraction Kit and could be further used for readout.

Table 58 Composition of negative selection plates in directed evolution

Volume	Components	Comments
200 mL	LB agar*	LB agar has to be cooled before usage to prevent destruction of antibiotics.
200 µL	Cam	
200 µL	Antibiotic of the library	
2 mL	20 % (w/v) Arabinose	

*For libraries based on PylRS-mutants which are known for Phe-misincorporation autoinduction plates without Phe were used (according to Table 57) instead of LB agar.

Directed evolution – Readout

The dual reporter plasmids *Mb*-pPylT_sfGFP150TAG-CAT or *Ma*-pPylT_sfGFP150TAG-CAT encode an arabinose-inducible sfGFP gene with a TAG-codon at position 150 as well as a constitutively expressed chloramphenicol transferase with a TAG codon at position 111 and a constitutively expressed mutant of either *Mb*-$tRNA_{CUA}$ or *Ma*-$tRNA_{CUA}$. The single reporter pPylT_sfGFP150TAG lacks the chloramphenicol transferase gene with a TAG codon at position 111. The readout by GFP fluorescence helps to identify *E. coli* cells which express PylRS-library members which incorporate the uAA in response to the TAG-codon. In case of the dual reporter this is additionally coupled to a second positive selection step.
Electro-competent *E. coli* NEB 10-beta-*Mb*-pPylT_sfGFP150TAG/-*Mb*-pPylT_sfGFP150TAG-CAT/-*Ma*-pPylT_sfGFP150TAG-CAT were prepared and transformed according to the description for positive selection. 100 – 200 ng of the isolated library-DNA of the negative selection was transformed into 2x 50 µL of the electro-competent *E. coli* and rescued with 900 µL SOB medium before incubation at 37 °C for 1 hour. The cells were harvested by centrifugation (4200 rpm, 5 min) and half of the supernatant was discarded. The

pellet was resuspended in the remaining medium and plated on an autoinduction plate (+Tet + antibiotic of the library) (see Table 59). 1 µL of the pBK_PylRS-wt plasmid was also transformed into 50 µL of the freshly prepared electro-competent *E. coli* and plated onto a separate autoinduction plate (+Tet + Kan + 2 mM BocK) as positive control. The plates were incubated at 37 °C for 48 – 60 hours (depending on individual growth rates). Resulting sfGFP-expressing colonies (green) were picked for inoculation of 1 mL non-inducing medium (see Table 59) in a 96-well Megablock format. Additionally, 4 – 6 colonies of the positive control plate were also picked. The cultures were grown at 37 °C for 24 hours before being used to inoculate two 96-well Megablocks with autoinduction medium (see Table 59), one supplemented with the respective uAA and one without uAA. After incubation for 36 – 48 hours at 37 °C 100 µL of the cultures were transferred to a 96-well microplate containing 50 µL MilliQ-H_2O per well. The sfGFP fluorescence (exited at 480 nm, measured at 527 nm) and the OD_{600} were measured with a Tecan Spark 10M Microplate reader. The microplates were shaken (amplitude 3.5) before measurements. The sfGFP fluorescence was normalized to the OD_{600} for each culture. The resulting values of the cultures with uAA were divided by the resulting values of each associated culture without uAA. Cultures of PylRS members which showed values ≥ 3 were inoculated into 50 mL 2xYT (+Tet +antibiotic of the library) to perform a bigger amber suppression on sfGFP150TAG (take 1 mL of the culture as backup for inoculation of novel cultures). sfGFP was purified using high density Ni^{2+}-agarose (see 5.2.6.2) and analyzed by LC-MS. If the sfGFP mass displayed the right mass for successful uAA incorporation a new overnight culture was inoculated using the 1 mL backup sample. Next day the plasmids DNA was isolated via the Qiagen HiSpeed Plasmid Midi Kit. The separation of library-DNA (3.3 kb) and readout plasmid (7.7 kb or 6.2 kb) occurred via agarose gel electrophoresis (120 V; 1.75 hours). The DNA band at 3.3 kb was extracted with the Monarch DNA Gel Extraction Kit and sent for sequencing (see 5.2.4.9) to analyze the mutations of the evolved PylRS mutant.

Table 59 Compositions of medium & solutions used for readout in directed evolution

Medium/plate	Components	Comments
Autoinduction plate	10 mL 5 % (w/v) Aspartate 10 mL 10 % (v/v) Glycerol 8 mL 1x 17- or 18-AA Mix* 4 mL 50x M 2 mL Leucine (4 $\frac{mg}{mL}$) 500 µL 20% (w/v) Arabinose 400 µL 1 M $MgSO_4$ 50 µL 40% (w/v) Glucose 40 µL 5000x Trace metal mix 200 µL Biotin (1 $\frac{mg}{mL}$) 200 µL Thiamin (1 $\frac{mg}{mL}$) 800 µL India Ink X mL uAA (100 mM)	Prepare freshly for use Agarose were heated until liquid and cooled in a 50 °C water bath. Be sure that agarose is not too hot to prevent the destruction of heat-labile compounds. Other components were mixed and pre-warmed in a 50 °C water bath. Agarose was combined with the other components and plates were immediately poured. Concentration of the uAA depended on the respective uAA and was between 2 – 4 mM. #If using single reporter pPylT_sfGFP150TAG do not use Cam.

	100 µL Cam[#] 200 µL Tet 200 µL Antibiotic of the library 100 mL 5 % Agarose Fill up to 200 mL MilliQ-H_2O	*For libraries based on PylRS-mutants which are known for Phe-misincorporation use 25x 17 AA mix.
Non-inducing medium for directed evolution	5 mL 5 % (w/v) Aspartate 4 mL 1x 17- or 18-AA Mix* 2 mL50x M 1 mL Leucine (4 $\frac{mg}{mL}$) 200 µL 1 M $MgSO_4$ 4 mL 40 % (w/v) Glucose 20 µL 5000x Trace metal mix X mL uAA (100 mM) 100 µL Tet 100 µL Antibiotic of the library Fill up to 100 mL MilliQ-H_2O	Can be used up to 7 days. *For libraries based on PylRS-mutants which are known for Phe-misincorporation use 25x 17 AA mix.
Autoinducing medium for directed evolution	5 mL 5 % (w/v) Aspartate 5 mL 10 % (v/v) Glycerol 4 mL 1x 17- or 18-AA Mix* 2 mL 50x M 1 mL Leucine (4 $\frac{mg}{mL}$) 250 µL 20 % (w/v) Arabinose 200 µL 1M $MgSO_4$ 125 µL 40 % (w/v) Glucose 20 µL 5000x Trace metal mix X mL uAA (100 mM) 100 µL Tet 100 µL Antibiotic of the library Fill up to 100 mL MilliQ-H_2O	Prepare freshly for use. Concentration of the uAA depended on the respective uAA and was between 2 – 4 mM. *For libraries based on PylRS-mutants which are known for Phe-misincorporation use 25x 17 AA mix.

5.2.6. Proteinchemical Methods

5.2.6.1. Determination of protein concentration

Determination of protein concentration using Nanophotometer

The determination of purified protein concentrations was done with a nanophotometer at 280 nm. The absorption at 280 nm is related to the amino acid composition of the respective protein. The absorption characteristic is mainly influenced by aromatic amino acid residues like tyrosine or tryptophan. The protein-specific molar excitation coefficient of the reduced protein was determined with the ProtParam webtool, as well as the average mass. The monoisotopic mass was calculated using the PeptideMass web-service. Table 66 summarizes important molar excitation coefficients. The calculation method is based on the law of Lambert-Beer. This

method is not suitable for protein complexes or proteins, which do not contain aromatic residues. For these cases the Bradford assay had to be used for the determination of protein concentration.

Determination of protein concentration using Bradford assay

The principle of the Bradford assay for the determination of protein concentrations is based on the triphenylmethane dye Brilliant Blue G-250, which forms complexes with basic and aromatic amino acids of the protein. This association between the dye and the protein causes an absorbance shift from 465 nm to 595 nm and can be measured photometrically. The absorption is thereby proportional to the protein concentration.

The assay was set up in plastic cuvettes with a total volume of 1.55 mL and the commercially available Bradford Reagent from Sigma-Aldrich, according to the manufacturer's technical bulletin. 1.5 mL of the pre-warmed Bradford Reagent were mixed with 0.05 mL protein sample and incubated for 15 min at rt, before measuring the absorbance at 595 nm. Simultaneously a calibration curve with BSA as protein standard (0, 0.25, 0.5, 0.75 and 1 $\frac{mg}{mL}$, diluted in respective the buffer) was created, which was used for calculating the concentration of the protein sample.

5.2.6.2. Purification of His-tagged proteins

Purification steps of proteins containing any light-sensitive uAAs (e.g. pDAP, HpGGK, etc.) were performed in the dark (darken flasks with aluminum foil etc.) as much as possible. The respective washing- and elution-buffers for the different proteins are listed in Table 21.

Purification of His-tagged proteins via Ni^{2+}-agarose

Smaller scale expressions of His-tagged proteins in *E. coli* (4 – 200 mL culture volume) were purified using high density nickel agarose slurry. Therefore, the cells were harvested, resuspended in respective His-lysis buffer and lysed according to Table 21. Before the cleared lysate was applied to the Ni^{2+}-beads (0.2 – 0.6 mL depending on the size of the expression), they were washed twice with 20 mL appropriate 1x His-wash buffer by centrifugation at 1000 g; 4 °C, 5 min to remove 20 % EtOH. The cleared lysate was incubated on the washed Ni^{2+}-beads for 1 hour on ice, gently shaking. After the incubation the suspension was transferred to an empty plastic column and the flow through was collected. The beads were washed with 50 mL His-wash buffer and the protein was eluted two or three times with 500 µL of the respective His-elution buffer. The fractions were checked by SDS-PAGE (see 5.2.6.8). The elution fractions were pooled and concentrated or rebuffered in buffer without imidazole using centrifugal units with corresponding cut-off or dialysis (see 5.2.6.7).

Purification of His-tagged proteins via Ni-NTA Spin Kit

Alternatively, small protein amounts (e.g. protein from 4 mL expression cultures) were purified using the Ni-NTA Spin Kit (Qiagen). The procedure was performed according to the manufacturer's protocol, except for cell lysis (see 5.2.2.8) and used buffers. Here, the His-purification buffers for the respective protein (see Table 21) were used instead of those recommended by Qiagen.

Purification of His-tagged proteins via ÄKTApure system

For big expressions with high amounts of His-tagged protein (500 - 2000 mL culture volume) the ÄKTApure system with a HisTrap FF columns (1 mL) or a self-packed Ni-NTA column (CV 10 mL of Ni-NTA slurry from Jena Bioscience) were used. Using a liquid chromatography

system, it was necessary that all buffers were filtered (0.45 µm pore size) and degassed. The buffer compositions stayed the same for the respective proteins (see Table 21). If necessary, the cleared lysate also had to be filtered, before applying it to the system. The purification protocol consisted of five steps: initial equilibration of the column with 10 CV His-washing buffer, sample application using an air sensor, washing with 10 CV His-washing buffer, elution with 3 CV His-elution buffer using an sample collector and final equilibration with 6 CV His-washing buffer. Flow rates depended on the column: HisTrap FF columns (1 mL) 1 $\frac{mL}{min}$; self-packed Ni-NTA column 10 $\frac{mL}{min}$.

5.2.6.3. Purification of StrepII-tagged proteins

Proteins having light-sensitive uAAs (e.g. pDAP, HpGGK, etc.) incorporated were purified in the dark (darken flasks with aluminum foil etc.) as much as possible. Buffer compositions were listed in Table 21.

Purification of StrepII-tagged proteins via Strep-Tactin® XT Superflow

Proteins with StrepII-Tag were purified via Strep-Tactin® XT Superflow, mostly used for small scale expression (4 – 200 mL culture volume). Using the XT-technology the elution was only possible with biotin. The buffers for the purification differed between proteins and are shown in Table 21. The expression culture was harvested, resuspended in respective Strep-lysis buffer and sonicated as described in 5.2.2.8. After centrifugation, to separate cell debris and cleared lysate, the cleared lysate was transferred to the washed Strep-Tactin XT beads (100 – 500 mL) and gently shaken for 1 hour at 4 °C. The Strep-Tactin XT beads were washed once with appropriate Strep-washing buffer (centrifuge at 4°C, 5 min at 1000 g). The suspension was then applied to a gravity-flow plastic column and the flow-through was collected. The beads were washed with 50 – 60 mL of the respective Strep-washing buffer. The StrepII-tagged protein was eluted with 3x 200 µL Strep-elution buffer. Now the protein could be used for further application, e.g. concentration, rebuffering, analysis or storage.

Purification of StrepII-tagged proteins via Strep-Tactin XT Spin columns

For small protein expressions (4 mL culture volume) the protein could be purified via Strep-Tactin XT Spin columns as an alternative. Except for the cell lysis (cells were disrupted using glass beads (see 5.2.2.8)) and the used buffers, the manufacturer's protocol was followed. Used buffers depended on the respective protein and are listed in Table 21.

Purification of StrepII-tagged proteins via ÄKTApure system

The ÄKTApure liquid chromatography system with a StrepTrap HP columns (1 mL; flow rate 1 $\frac{mL}{min}$) was used for large scale expression of StrepII-tagged proteins (500 - 1000 mL). The buffer recipes remained unchanged for each protein (see Table 21) but had to be filtered (0.45 µm pore size) and degassed before use, as well as the cleared lysate if necessary. The purification methods comprised five steps: initial equilibration of the column with 10 CV Strep-washing buffer, sample application using an air sensor, washing with 10 CV Strep-washing buffer, elution with 3 CV Strep-elution buffer using an sample collector and final equilibration with 6 CV Strep-washing buffer.

5.2.6.4. Purification of sfGFP-Dimers

For the purification of crosslinked sfGFP-dimers via uAAs, two purification steps were performed. sfGFP and sfGFP-dimer were isolated using Ni^{2+}-affinity purification (using either Ni^{2+}-slurry or ÄKTA) with appropriate buffers (see Table 21) according to 5.2.6.2. For separation of sfGFP-monomer and -dimer the elution fractions from Ni^{2+}-affinity purify was used for SEC. The HiLoad™ 16/600 Superdex 75 prep grade column (flow rate 1 $\frac{mL}{min}$) was equilibrated with 1 CV GFP-SEC buffer. The sample was applied via 5 mL loop onto the column. The column was washed with 1.5 CV GFP-SEC buffer. Samples were collected in 0.5 mL fractions using a 96-well sample collector. Fractions were analyzed e.g. by SDS-PAGE (see 5.2.6.8).

5.2.6.5. Purification of Rab1b-DrrA complexes

For the purification of different crosslinked Rab1b-DrrA complexes a two-step purification procedure was necessary. The first step consisted of Ni^{2+}-affinity purification (using either Ni^{2+}-slurry or ÄKTA) with appropriate buffers for Rab1b (see Table 21) according to 5.2.6.2 The elution fractions of the Ni^{2+}-purification was used for further SEC. Therefore, these fractions were pooled and diluted with SEC-buffer Rab1b-DrrA-complex (1:1) to reduce the imidazole concentration. Concentration with centrifugal units (see 5.2.6.7) was in some cases necessary to obtain a total volume which was compatible with the loop volume for sample application on SEC column (0.5 or 2 mL). For SEC a Superdex™ 200 Increase 10/300 GL column was used with a flowrate of 0.5 $\frac{mL}{min}$. The column was equilibrated with 1 CV SEC-buffer Rab1b-DrrA-complex before the sample was applied onto the column via a loop. The column was washed with 1.5 CV buffer and the 0.5 mL-fractions were collected using a 96-well sample collector.
As an alternative to SEC, ion-exchange chromatography (IEX) with a RESOURCE™ Q anion exchange column (flowrate 4 $\frac{mL}{min}$) was applied together with respective buffers for ResQ (see Table 21). Here it was important to reduce the NaCl concentration in the sample from the Ni^{2+} elution below 50 mM before applying it to the ResQ run. This was achieved by diluting the Ni^{2+}-eluate with ResQ-buffer I. The procedure included an equilibration with 10 CV ResQ-buffer I, followed by sample application using an air sensor and washing with 10 CV ResQ-buffer I. Then the salt concentration was increased over 20 CV from 100 → 0 % ResQ-buffer I and 0 → 100 % ResQ-buffer II and further 20 CV ResQ-buffer II were run. Finally, the column was washed with 20 CV ResQ-buffer I. During the IEX 0.75 mL fractions were collected using the 96-well sample collector, which were analyzed by SDS-PAGE. Fractions containing the protein complex were pooled. Depending on further application, samples were concentrated and the concentration was determined by Bradford-assay (see 5.2.6.1). Rab1b-DrrA complexes were stored at –80 °C after flash freezing with liquid nitrogen.
For crystallization experiments complex samples of contractions between 8 – 12 $\frac{mg}{mL}$ were used. Crystallization trials of Rab1b-DrrA complex were carried out using the sitting-drop vapour diffusion method and a Protein Crystallization Robot Oryx4 (Douglas Instruments Ltd) or Protein Crystallography Dispenser Phoenix (Art Robbins Instruments). Several commercially available buffers screens (Qiagen) in a 96-well format were tested.

5.2.6.6. Purification of PylRS Proteins for Crystallization

For the purification of C-terminal PylRS domains from *M. mazei* or PylRS from *M. alvus* (both around 30 kDa), for further crystallization set-ups, a two-step purification strategy was used. After expression, according to 5.2.2.6, first a Ni-NTA purification step was performed followed by size exclusion chromatography to obtain pure PylRS. *Mm*PylRS (C-term. domain) could be directly used for crystallization experiments after Ni-NTA and SEC. For *Ma*PylRS an additional working step, between Ni-NTA and SEC, was necessary for cleaving of the N-terminal 6xHis-Tag via thrombin digest. The cell pellet was resuspended in His-lysis buffer PylRS, sonicated and centrifuged (see 5.2.2.8). The cleared lysate was applied for further purification steps.

Purification via Ni-affinity chromatography using ÄKTApure system

The clear lysate was used for affinity chromatography using the ÄKTA pure system with a self-packed Ni-NTA column (CV 10 mL of Ni-NTA slurry from Jena Bioscience) (see 5.2.6.2). For further SEC the volume of the elution had to be reduced to 5 mL using Amicon Ultra 15 mL Centrifugal filters (10,000 NMWL) (see 5.2.6.7). Simultaneously the His-elution buffer PylRS was exchanged to PylRS-SEC buffer.

Thrombin digest for 6xHis-Tag removal of *Ma*-PylRS

The N-terminal 6xHis-Tag of *Ma*PylRS mutants had to be cleaved off for crystallization experiments via thrombin (37 kDa). The elution volume was reduced using Amicon Ultra 15 mL Centrifugal filters (10,000 NMWL) (see 5.2.6.7). Then a buffer exchange from His-elution buffer PylRS to PylRS-thrombin buffer via PD-10 columns was performed, according to the manufacturer's gravity protocol. The thrombin digest occurred overnight at 4 °C in SnakeSkin® Dialysis (10 K MWCO), equilibrated with PylRS-thrombin buffer. The protein solution was filled into the dialysis tube and thrombin was added (for 1 mg of PylRS 5 U of thrombin were used). The dialysis was performed against PylRS-thrombin buffer so that the cleaved off 6xHis-Tag could leave the tube continuously. The next day thrombin was removed by p-aminobenzamidine agarose (= pABA) via gravity flow columns. The pABA was washed with PylRS-SEC buffer before use. The digested protein sample was added to pABA, incubated for 15 min 4°C, smoothly shaking and loaded onto a gravity flow column. Thrombin stuck to the agarose whereas the Tag-free PylRS remained in the flow-through and could be used for further SEC. Depending on the volume of the flow-through a concentrating step to 5 mL using Amicon Ultra 4 ml or 15 mL Centrifugal filters (10,000 NMWL) was necessary.

Purification *via* size exclusion chromatography

For SEC the HiLoad™ 16/600 Superdex 75 prep grade column (flow rate 1 $\frac{mL}{min}$) was equilibrated with 1 CV PylRS-SEC buffer. The prepared elution from the Ni-NTA was injected via 5 mL loop onto the column. For separation the column was washed with 1.5 CV PylRS-SEC buffer. Samples were collected in 1 mL fractions using a 96-well sample collector. Fractions were analyzed by SDS-PAGE (see 5.2.6.8) and respective fractions were pooled and concentrated by centrifugation using Amicon Ultra 4 ml or 15 mL Centrifugal filters (10,000 NMWL) (see 5.2.6.7). Concentration was determined with the Nanodrop spectrophotometer (see 5.2.6.1). Protein in PylRS-SEC buffer was aliquoted, flash frozen in liquid nitrogen and stored at -80 °C until further use.

5.2.6.7. Buffer Exchange & Concentrators

Rebuffering and concentrating purified protein solutions using Centrifugal filters

To increase the concentration of a purified protein solution Amicon Ultra Centrifugal filters with respective cut-off and volume were used. In general, the cut-off should not extend one third of your protein size to prevent undesired passing of the protein through the membrane. The centrifugal filters were centrifuged at 4 °C for 3 – 6 min, so that the volume was reduced due to pushing through the membrane. The centrifugation was performed several times until the desired volume/concentration was reached. Between the centrifugation steps the protein solution was carefully mixed to avoid extreme concentration gradients, especially at the membrane, what might cause precipitation. Depending on the abstract volume the centrifugal units resist different centrifugation velocities according to manufacturer's manual. Centrifugal filter units could be further used for fast buffer exchange by repeated centrifugation steps (4 – 6 times) and refilling the volume with respective buffer, e.g. storage buffer.

Buffer exchange by dialysis

An alternative way of rebuffering was provided by dialysis. Depending on the volume of protein solution and the protein size either SnakeSkin® Dialysis Tubing (10 K MWCO) (> 3 mL sample), ReadyLyzer (MWCO 3.5 kDa) (0.8 – 3 mL sample) or Xpress Micro Dialyzer (MWCO 3.5 kDa) (0.3 – 1 mL sample) were used. The dialysis buffer, e.g. respective storage buffer, depended on protein and further application. A high excess of dialysis buffer (1 – 5 L) was used to ensure proper rebuffering. Before applying the protein solution to the dialysis unit, SnakeSkin® Dialysis Tubing or dialyzers were washed with respective dialysis buffer. Then the protein sample was filled in and the dialysis occurred at 4 °C for at least 4 hours or overnight.

5.2.6.8. SDS-PAGE

For separating cell lysates or protein samples via Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), mini gels (8 cm x 8 cm x 0.1 cm gel size) were used (in general 15 % gels). Gels were prepared regarding to Table 60. Complete polymerization occurred in 15 – 20 min. The separation in an electric field occurred at 110 V until samples entered the separating gel, then the voltage was shifted to 200 V. The SDS-PAGE was terminated when the blue dye of the sample loading buffer left the gel. As running buffer served 1x SDS-MES running buffer.

For the separation of *E. coli* cell lysates, first the OD_{600} of the culture was measured. Then 1 mL samples of the corresponding cultures were harvested by centrifugation (13 000 rpm, 2min, rt). According to their respective OD_{600}, the pellets were resuspended in 1x SDS loading buffer, to ensure equal loading of lysate amounts. 100 µL 1x SDS loading buffer were used for 1.0 OD_{600} units. Samples were cooked at 95 °C for 10 min and centrifuged at 13 000 rpm, rt for 15 min. Samples were separated using 15 % mini gels.

For the preparation of other SDS-samples (e.g. of purified proteins, protein mixtures or cleared lysates) the samples were mixed with 4x SDS loading buffer (final concentration 1x) and cooked at 95 °C for 2 min. Depending on the protein size, 10 %, 12 % or 15 % mini gels were used for separation.

After SDS-PAGE, the gels were either stained with Coomassie fast stain (Instant*Blue* coomassie stain or Quick Coomassie Stain; according to manufacturer's protocol) or further used for analysis via western blot.

Table 60 Preparation scheme of gels for SDS-PAGE

	Resolving gel			**Stacking gel**
	For 2 gels			For 2 gels
	10 %	**12 %**	**15 %**	4 %
Acrylamide/Bis Solution 37.5:1 (40 % w/v)	4.07 mL	4.88 mL	6.11 mL	0.50 mL
4x Resolving gel buffer or 4x Stacking buffer	4.07 mL	4.07 mL	4.07 mL	1.75 mL
ddH_2O	8.14 mL	7.33 mL	6.11 mL	3.25 mL
APS (10 % (w/v))	162.75 µL	162.75 µL	162.75 µL	50 µL
TEMED	16.28 µL	16.28 µL	16.28 µL	5 µL

5.2.6.9. Western Blot

After performing SDS-PAGE (see 5.2.6.8) the unstained mini gel could be used for western blotting. For these experiments, a pre-colored protein marker was applied, in order to visualize the respective bands on the nitrocellulose (NC) membrane. The protein transfer from the SDS gel to the NC membrane was done by either semi-dry blotting or fast blotting with the iBlot™ 2. The respective procedures are described below. After successful protein blotting, the NC membrane was blocked in TBS-T (containing 0.1 % (v/v) Tween20) with 5% (w/v) skimmed milk powder for 1 hour at rt. The Blocking solution was removed and the primary antibody was added in a suitable dilution (1:5000 for α-His-HRP and α-Strep-HRP; 1:200 for α-Rab1b-Y77AMP; 1:1000 for α-GFP) in TBS-T containing 1 % (w/v) skimmed milk powder and incubation occured overnight at 4 °C. Using primary antibodies conjugated to HRP, the blot was directly washed (three times with TBS-T, each for 10 min) and imaged. For unconjugated primary antibodies (e.g. α-Rab1b-Y77AMP, α-GFP), another incubation step with a secondary antibody was required, which was conjugated to HRP. Therefore, fresh TBS-T containing 1 % (w/v) skimmed milk powder was added and mixed in a 1:40 000 for goat-anti-rabbit-HRP and 1:500 for goat-anti-mouse-HRP dilution with the secondary antibody. After 1-hour incubation at rt the blot was washed three times with TBS-T again. The detection of the antibody-protein complexes, conjugated to HRP, was performed with the Amersham ECL Prime Western Blotting Detection Reagent (two component system, consisting of a luminol- and a H_2O_2-solution). The blot was incubated with Western Blotting Detection Reagent (according to manufacturer's protocol) and detection was carried out in a WB Imager Fusion Pulse 6 was. The exposure time was adjusted for the respective experiment.

Semi-dry western blotting

The SDS gel was equilibrated with Semi-Dry transfer buffer for 5 min at rt before blotting. To transfer the separated proteins from the SDS gel to a NC membrane a blotting stack had to be assembled. Therefore, 2 – 3 Whatman papers, as well as the NC membrane, were soaked with Semi-Dry transfer buffer and put onto the anode of the semi-dry blotter with the wetted NC membrane on top. The SDS gel was transferred onto the NC membrane, followed by 2 – 3 more Whatman papers (soaked with Semi-Dry transfer buffer). During stack assembly air bubbles were removed to ensure proper protein transfer. The cathode of the semi-dry blotter was put on top of the stack and fixed strongly. The transfer was carried out at 320 mA for 1 – 1.5 hours (depending on molecular weight of the proteins).

Fast western blotting with the iBlot2 system

For shorter blotting times the iBlot™ 2 Dry Blotting system (Thermo Fisher Scientific) was used according to the manufacturer's manual. The blotting required pre-assembled iBlot™ 2 transfer stacks (with a NC membrane). The transfer was performed with program P0 (20 V for 1 min, 23 V for 4 min, 25 V for 2 min).

5.2.6.10. AMPylation assay with Rab1b & DrrA

The activity of Rab1b proteins can be modified by AMPylation at position Y77. This is realized by the enzymatic activity of different effector proteins (e.g. DrrA or SidD) which can either AMPylate or de-amplyate Rab1b. For the detection of the AMPylation-state of Rab1b proteins (single proteins or in complexes) western blot analysis with a specific antibody (α-Rab1b-Y77AMP = α-Rab1b-AMP (from rabbit)), which binds to Rab1b that is AMPylated at position Y77 (Rab1b-Y77AMP), was used. Additionally, LC-MS was performed to determine the Rab1b masses. The *in vitro* AMPylation was carried out in AMP-assay buffer (see Table 21) at 25 °C with 25 µM or 50 µM Rab1b (in the beginning unAMPylated), 2.5 molar excess of ATP and 0.5 µM or 1 µM effector protein. Samples were taken after 2 hours. For SDS-PAGE and western blot samples the reaction was terminated by adding 4x SDS Loading buffer and boiling at 95 °C for 3 min. For the LC-MS samples a dilution in ACN (1:8) quenched the reaction. SDS-PAGE and western blot was performed according to 5.2.6.8 and 5.2.6.9.

5.2.6.11. *In-vitro* protein crosslinking of Rab1b & DrrA

For *in vitro* formation of Rab1b-DrrA complexes Rab1b proteins (final 7.5 µM or 15 µM) were mixed with DrrA proteins (final 5 µM or 10 µM) in Xlink buffer (see Table 21) and ATP (final 25 µM) were added. Rab1b was used in an 1.5 fold excess (Rab1b:DrrA = 1.5:1). For experiments using $DrrA_{16\text{-}533}$ the Xlink buffer was further supplemented with GDP (final 10 µM). For experiments concerning GSH-influence further GSH was added to the buffer in different concentrations. The reaction was incubated at 25 °C, 200 rpm and samples were taken after different time points. Reaction was stopped by directly adding 4x SDS-Loading buffer and cooking at 95 °C for 3 min. 10 µL were used for SDS-PAGE (see 5.2.6.8) and Western blot (α-His) (see 5.2.6.9).

5.2.6.12. Photo-crosslinking of proteins via DiaziK *in vivo*

For triggering crosslinking via DiaziK *in vivo*, introduced into proteins during amber suppression in *E. coli* (see 5.2.2.7), light-activation was necessary. Therefore, 2 mL of the living cell culture was transferred into wells of a 6-well plate after expression and placed on ice before

irradiating with UV-light (360 nm) for 10 min. Cell samples could be further analyzed via SDS-PAGE (see 5.2.6.8) and Western Blot (see 5.2.6.9).

5.2.6.13. Deprotection of of *Sa*ClpP-pDAP *in vitro*

The photocage of pDAP introduced into *Sa*ClpP was removed by light irradiation at 365 nm. The purified *Sa*ClpP-S98pDAP was transferred into 6-well plates (2 mL per well) and treated with UV-light for different periods of time while being cooled on ice under gently stirring. The removal of the photocage and the subsequent spontaneous fragmentation to final *Sa*ClpP-S98DAP was monitored via electrospray ionization mass spectrometry (ESI-MS) of the intact protein (see 5.2.6.16, LTQ-FT Ultra mass spectrometer). To identify best conditions for the conversion of pDAP to DAP on *Sa*ClpP different irradiation-times, buffer-pH-values, incubation time and temperatures were screened.

5.2.6.14. Modification of *Sa*ClpP-DAP with probes *in vitro*

Purified and deprotected (removed photocage + successfully fragmented; see 5.2.6.13) *Sa*ClpP-S98DAP (4 µM) was mixed with an access of respective small molecule inhibitor (AV170 final 100 µM = 25-fold access (10 mM stock in DMSO) or D3 final 200 µM = 50-fold access (5 mM stock in DMSO), both from the Sieber group) and incubated overnight at rt. Modification of *Sa*ClpP-S98DAP was monitored using ESI-MS (see 5.2.6.16, LTQ-FT Ultra mass spectrometer). Successfully modified probes were further incubated at 32 °C for 24 hours and measured again to check the stability of the modification.

5.2.6.15. Analytical SEC of *Sa*ClpP

The oligomeric state of *Sa*ClpP and its mutants was analyzed by analytical SEC using an ÄKTA™ Purifier 10 chromatography system with a calibrated Superdex™ 200 10/300 GL column at 4 °C. PZ-buffer (pH 7.0) was used as solvent with a flow rate of 0.25 $\frac{mL}{min}$. Protein samples were loaded via 500 µL-sample loop. The elution profile was recorded by monitoring the UV-absorption at 280 nm. Samples were adjusted to a volume of 200 µL and a concentration of 10 µM with PZ-buffer (pH 8.0). To generate fully decaged *Sa*ClpP-S98Dap-Strep, the 10 µM probe was irradiated for 10 min at 365 nm, followed by an incubation at 32 °C for 12 hours. Successful decaging was checked by ESI-MS using LTQ-FT Ultra mass spectrometer (see 5.2.6.16). A run with 200 µL *Sa*ClpP-Strep (10 µM) served as control.

5.2.6.16. ESI-MS of intact proteins

ESI-MS using Agilent mass spectrometer

Measurement of purified full-length proteins was carried out on an Agilent 1260 Infinity Series LC system with an Agilent 6210 ESI Single Quadrupole mass spectrometer and a Jupiter C4 5 µm (2 x 150 mm) capillary column (Phenomenex, Torrence, USA). The mobile phase consisted of solvents A (MilliQ-H_2O + 0.1 % FA) and B (ACN + 0.1 % FA) (unfiltered). Measurements were performed at rt with a flow rate of 0.3 $\frac{mL}{min}$ using a gradient of 5 – 55 % B (ACN + 0.1 % FA) in 5 min. The samples were analyzed in positive mode, following protein UV absorbance at 280 nm. Total protein mass was calculated by deconvolution within the MS Chemstation software (Agilent Technologies). Theoretical masses of proteins were calculated

using ProtParam or PeptideMass webtools and were manually corrected for the mass of the unnatural amino acid mutagenesis.

ESI-MS using LTQ-FT Ultra mass spectrometer

The LTQ-FT Ultra mass spectrometer consisting of an electrospray ionization source (spray voltage 4.0 kV, tube lens 110 V, capillary voltage 48 V, sheath gas 60 a.u., aux gas 10 a.u., sweep gas 0.2 a.u.) coupled to a Dionex UltiMate 3000 HPLC Instrument (both Thermo Fisher Scientific) was used for high-resolution ESI-MS measurements. Desalting was carried out with Massprep desalting cartridges (Waters). Briefly, proteins were loaded in H_2O + 0.1 % (v/v) formic acid (buffer A) and eluted in a 3.5 min gradient (1.5 min from 6 to 95 % buffer B (90:10 ACN:H_2O + 0.1 % (v/v) formic acid, 2 min 95 % buffer B) at a flow rate of 400 $\frac{\mu L}{min}$. Spectra were acquired in full scan (positive ion) mode with a resolution of R = 200.000 from m/z 600 to m/z 2000 and deconvoluted with the Thermo Xcalibur Xtract algorithm (Thermo Fisher Scientific) using basic deconvolution default settings. The instrument and equipment from the group of Prof. Sieber (TUM, Germany) was used.

5.2.6.17. XL-MS (crosslinking MS) analysis of proteins

Instruments, equipment and software for MS/MS measurement and data evaluation were not listed in the software- and equipment-tables. This was performed by Dr. Matthias Stahl (former group of Prof. Sieber (TUM, Germany), now Karolinska Institute (Sweden)) or Dr. Christian Ihling (group of Prof. Sinz (Halle-Wittenberg, Germany)) using their material.

Sample preparation for XL-MS

All reagents used for XL-MS-sample preparation were LC-MS grade chemicals and solvents. Measurement were performed with the instrument Orbitrap Fusion or XL. Protein samples (50 – 300 µg) were lyophilized and resuspended in 200 µL buffer X for denaturation. To dissolve the proteins completely the samples were sonicated for 10 min in an ultrasonic bath. 0.2 µL of 1 M DTT was added to each sample, mixed and incubated for 45 min at rt and 450 rpm to reduce proteins. For alkylation 2 µL of 550 mM IAA was supplemented, mixed and incubated at rt,450 rpm for 30 min in the dark. The reaction was quenched by the addition of 0.8 µL of 1 M DTT and incubation at rt for 30 min at 450 rpm. 1 µL of 1 M LysC (0.5 $\frac{\mu g}{\mu L}$) was added and the samples were incubated at rt for further 2 hours, 450 rpm in the dark. 600 µL of 50 mM TEAB were added to adjust the pH around 8 (if not add more NaOH). Finally, 1.5 µL trypsin (0.5 $\frac{\mu g}{\mu L}$) was added and the digest was carried out at 37 °C overnight at 450 rpm. The reaction was stopped by the addition of 8 µL formic acid, (pH should be 3 or below, if not more formic acid was added). For desalting 50 mg Sep-Pak C18 columns were used which were equilibrated with 1 mL ACN, 1 mL buffer E and 3x 1 mL 0.1 (v/v) TFA. Samples were loaded and washed with 3x 1 mL 0.1 (v/v) TFA and 0.5 mL 0.5 % (v/v) formic acid. Elution of peptide samples into a LoBind tube was performed with 3x 250 µL buffer E. Alternatively, a stage-tip-method was used for desalting with a double C18-membrane (Octadecyl C18 47 mm Extraction disks, Empore Products, USA). The membrane was washed with 70 µL MeOH, 70 µL buffer E (80 % (v/v) ACN, 0.5 % (v/v) formic acid) and three times with 70 µL 0.5 % (v/v) formic acid before the digested protein sample was loaded. The membrane was washed again three times with 70 µL 0.5 % (v/v) formic acid before eluting with two times 30 µL buffer E into a LoBind tube. Samples were lyophilized using a SpeedVac and were stored at -80 °C until measurement.

Before applying the samples onto the instrument, they were dissolved in 30 µL 1 % (v/v) formic acid and filtered using Ultrafree-MC-GV centrifugal filters. The filter was washed with 300 µL 1 % (v/v) formic acid by centrifugation at 13 000 rpm for 2 min before the dissolved sample was filtered (13 000 rpm; 2 min) and collected into a new LoBind tube.

XL-MS acquisition of purified crosslinked Rab1b-DrrA complexes

MS analysis of digested protein samples was performed on an Orbitrap Fusion instrument coupled to an Ultimate3000 Nano-HPLC via an electrospray easy source (all Thermo Fisher Scientific). 7 µL of the sample were loaded on a 2 cm PepMap RSLC C18 trap column (particles 2 µm, 100A, inner diameter 75 µm, Thermo Fisher Scientific) with 0.1 % TFA and separated on a 50 cm PepMap RSLC C18 column (particles 2 µm, 100A, inner diameter 75 µm, Thermo Fisher Scientific) constantly heated to 50 °C. The gradient was run from 5-32 % acetonitrile, 0.1 % formic acid (7 min 5 %, 105 min to 22 %, 10 min to 32 %, 10 min to 90 %, 10 min wash at 90 %, 10 min equilibration at 5 %) at a flow rate of 300 $\frac{\mu L}{min}$. Survey scans (m/z 300-1500) were acquired in the orbitrap with a resolution of 120,000 and a maximum injection time of 50 ms with an automatic gain control (=AGC) target of 4e5. Most intense ions with charge states 4-8 and an intensity threshold of 5e3 were selected for fragmentation by high-energy collisional dissociation (HCD) with a collision energy of 30 %. Fragment spectra were again recorded in the orbitrap with a resolution of 30,000, a maximum injection time of 100 ms and an AGC target of 5e4. The option "inject ions for all available parallelizable time" was enabled. Dynamic exclusion was enabled with an exclusion duration of 120 s. The overall cycle time was 5 s.

Data analysis for crosslink detection of Rab1b-DrrA complexes

Data evaluation was performed by Dr. Matthias Stahl (former group of Prof. Sieber (TUM, Germany), now Karolinska Institute (Sweden)).

Raw mass data files were converted into mzML files using MSConvert from ProteoWizard [324]. Crosslink searches were then performed with the software Kojak, version 1.5.5 (http://www.kojak-ms.org)[325] against a database consisting of Rab1b_Q67A_R69K-DrrA1-339, common contaminant proteins downloaded from the Andromeda configuration in the MaxQuant software package[326] and all of the reverse sequences. The MS^1 and MS^2 resolution were set to 120,000 and 30,000, respectively. Variable modifications: Oxidation on methionine (+15.9949) and AMPylation on tyrosine (+329.0525). Static modifications: Carbamidomethylation (+57.02146) on cysteine. Three modifications were allowed per peptide. The fragment bin offset was set to 0 and the size to 0.03. A maximum of three missed cleavages was allowed. The minimum peptide mass was set to 300 Da. The precursor mass tolerance was fixed to 10. The unnatural amino acid was encoded as lysine in our database, thus we searched for crosslinks (+96.0575) between lysines and cysteines, serines, threonines, aspartates, glutamates, and the protein N-termini. Additionally, we searched for the monolink (+175.9837). Annotated spectra were visualized with the Kojak Spectrum viewer[327]. Further analysis such as hit selection and statistics was performed with custom R scripts (https://github.com/higsch/crosslinkR).

The mass spectrometry proteomics data, the fasta file and Kojak configuration file have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[328] with the dataset identifier PXD019043.

The data remains private until manuscript acceptance. In order to access the repository during the reviewing process, please use the following credentials:
User: reviewer35397@ebi.ac.uk
Password: X9EEG5Jb

XL-MS acquisition of purified crosslinked complexes with MS^2-cleavable uAAs (BrCN7K)
Samples of digested crosslinked protein complexes via MS/MS-cleavable uAAs were analyzed via LC-MS/MS using an Dionex UltiMate 3000 HPLC system (Thermo Fisher Scientific) equipped with an Acclaim C18 PepMap100 (75 µm ID x 2 cm) trap column (Thermo Fisher Scientific) and an Aurora UHPLC (C18 1.6 µM 75 µm ID x 25 cm) separation column (IonOpticks) coupled to an EASY-source equipped Q Exactive Plus mass spectrometer (Thermo Fisher Scientific). The sample was reconstituted in 1 % (v/v) formic acid in H_2O and filtered before loading 4 µL of the peptide sample onto the trap column with a flow rate 5 $\frac{\mu L}{min}$ in aqueous 0.1% TFA Sample was subsequently transferred to the separation column at a flow rate of 0.4 $\frac{\mu L}{min}$. Samples were separated using a 152 min gradient (buffer A: H_2O with 0.1% (v/v) formic acid; buffer B: ACN with 0.1% (v/v) formic acid). In the first 7 min buffer B raised from 0 % to 5 %, over the next 105 min buffer B was increased from 5 % to 22 %. In the next 10 min buffer B was further increased from 22 % to 35 % and finally from 35 % to 90 % buffer B within another 10 min period before holding 90 % buffer B for 10 min followed by decreasing buffer B within 0.1 min from 90 % to 5 % and before staying at 5 % buffer B until the end of the run (total: 152 min). Samples were ionized at spray voltage of 1.8 kV. The instrument was used in a TopN data dependent mode of 10. The orbitrap mass analyzer with a resolution of R = 140,000 and an automatic gain control (= AGC) target of 3.0e6 and with a maximum injection time of 100 ms was used to perform MS full scans with a scan range from to 300 – 2000 m/z and with 60 % RF lens amplitude. Peptides with intensities higher than 4.0e3 and charge state 3 – 8 were fragmented with stepped higher-energy collisional dissociation (= HCD) (27 %; 30 %; 33%). Recording of the MS^2 scans was done in the orbitrap. The AGC target was set to 2.0e4 and the isolation window to 1.6 m/z and a maximum injection time of 250 ms. Dynamic exclusion duration was set to 60 s. The method was developed by Anja Fux.

Data analysis for crosslink detection of complexes with MS^2-cleavable uAAs (BrCN7K)
Data evaluation of MS/MS measurements with crosslinked protein complexes using MS/MS-cleavable uAAs was done by Dr. Christin Ihling (group of Prof. Sinz (Martin-Luther University Halle-Wittenberg, Germany)).
Raw mass files were exported to mzml-files using Proteome Discoverer (Version 2.4, Thermo Fisher Scientific). Crosslink searches were then performed with the software MeroX (Version 2.0.1.4) with the following settings: Protease sites: after K or R, max. 3 missed cleavages; Variable modifications: oxidation of M; Xlinks from K (stands for the uAA BrCN7K) to C (for Rab1b-DrrA complexes) or K to E (for sfGFP-dimer) corresponding crosslinker fragments are shown in Figure 74; Precursor precision MS^1: 10 ppm for Rab1b-DrrA complexes and 1000 ppm for sfGFP-dimer; Precursor precision MS^2: 15 ppm; Minimum charge state: 2+ (for Rab1b-DrrA complexes) 6+ (for sfGFP-dimer); FDR cutoff: 1 %.

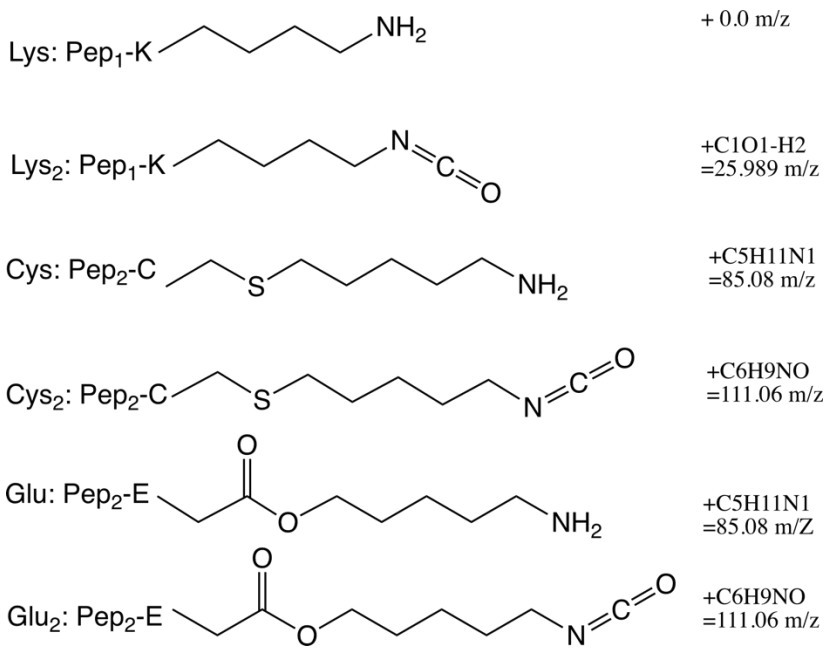


Figure 74: Molecular structures and corresponding m/z values for the crosslink fragments used in the settings of the MeroX analysis for Rab1b-DrrA complexes and sfGFP-dimer crosslinked via BrCN7K.

5.2.6.18. Protein crystallization

Loading uAA onto PylRS mutants *in vitro*

To co-crystallize the purified PylRS proteins with its respective uAA substrate a uAA-loading step was performed *in vitro*. The uAA loading was carried out directly before starting the crystallization. For sample preparation first ATP/ApCpp (1 mM final), followed by the uAA (2 mM, 5 mM or 10 mM final) was added to PylRS-SEC buffer. At last PylRS (1 $\frac{mg}{mL}$ final) was added carefully. The sample was incubated on ice or at RT for 15 min up to 1 hour and during this time the samples were inverted every 15 min. After incubation the samples were centrifuged at 4°C, 10 min at 16 000 g to remove any precipitate. The protein was concentrated to 10 - 20 $\frac{mg}{mL}$ using centrifugal filter devices (Amicon Ultra 0.5 ml Centrifugal filters (10,000 NMWL)) (see 5.2.6.7).

Crystallization of PylRS & data evaluation

Software and instruments, used for data collection and evaluation of crystallization experiments, are not listed in the software- and equipment-table and was performed by Dr. Sabine Schneider (LMU).

High-throughput crystallization trials of *Mm*-PylRS-MF18-CTD and *Ma*-PylRS-MF18IP in complex with AzGGK and/or ATP were carried out using the sitting-drop vapor diffusion method at 4 °C, 18 °C or 20 °C and a Protein Crystallization Robot Oryx4 (Douglas Instruments Ltd) or Protein Crystallography Dispenser Phoenix (Art Robbins Instruments). Several commercially available buffers screens (Qiagen) in a 96-well format were tested. Crystals of *Mm*-PylRS-MF18-CTD in complex with AzGGK were obtained in 0.2 M ammonium sulphate, 15% (w/v) PEG4000, at 18°C. Crystals of *Ma*-PylRS-MF18IP in complex with ATP were grown in a manual grid screen in 0.15 M MES, pH 6.0 acid 18 %(w/v) PEG 8000. For cryoprotection, crystals were soaked in mother liquor supplemented with 30% (w/v) ethylene glycol, flash-cooled and stored in liquid nitrogen. Data collection was carried out at the synchrotron beamline PX I at the Swiss light Source (SLS, Villigen Switzerland, *Mm*-PylRS-

MF18-CTD) and P14 at the Deutsches Elektronen-Synchrotron (DESY, Hamburg Germany, *Ma*-PylRS-MF18IP). The data were processed with XDS[329] and the structures were solved by molecular replacement with PHASER[330] using the coordinates of the previously reported structures (*Mm*-PylRS-CTD PDB code: 2q7e; *Ma*-PylRS PDB code 6ezd). Model building was done with COOT and refinement of the coordinates was carried out with REFMAC5[331, 332]. The structural figures were prepared using PyMOL (The PyMOL molecular graphic system, version 1.5 Schrödinger LLC, 2015, New York). For data processing and structure refinement statistics, see appendix 7.3 Table 68.

5.3. Consumables & Commercial solutions

Table 61 Consumables and commercial solutions

Consumables/Solutions	**Supplier**
Acrylamide/Bis Solution 37.5:1 (40 % w/v)	Serva Electrophoresis GmbH (Heidelberg, Germany)
Amersham ECL Prime Western Blotting Detection Reagent	GE Healthcare Life Science (Marlborough, USA)
Amersham Protran Nitrocellulose Blotting Membrane, 0.2 µm	GE Healthcare Life Science (Marlborough, USA)
Amicon Ultra 0.5 mL Centrifugal filters, 10,000 NMWL	Merck KGaA (Darmstadt, Germany)
Amicon Ultra 0.5 mL Centrifugal filters, 30,000 NMWL	Merck KGaA (Darmstadt, Germany)
Amicon Ultra 4 mL Centrifugal filters, 10,000 NMWL	Merck KGaA (Darmstadt, Germany)
Amicon Ultra 4 mL Centrifugal filters, 30,000 NMWL	Merck KGaA (Darmstadt, Germany)
Amicon Ultra 15 mL Centrifugal filters, 10,000 NMWL	Merck KGaA (Darmstadt, Germany)
Amicon Ultra 15 mL Centrifugal filters, 30,000 NMWL	Merck KGaA (Darmstadt, Germany)
Bradford Reagent for 0.1-1.4 mg/mL protein	Sigma Aldrich (St. Louis, USA)
Cannulas Sterican® (0.8 * 25mm) (0,8*120mm)	B. BRAUN DEUTSCHLANDGMBH & CO. KG (Melsungen, Germany)
cOmplete™, Mini, EDTA-free Protease Inhibitor cocktail (EASYpack)	Roche Applied Science (Penzberg, Germany)
Countess™ Cell Counting Slides	Thermo Fisher Scientific (Waltham, USA)
DNA stain Clear G	Serva Electrophoresis GmbH (Heidelberg, Germany)
Electroporation cuvettes, gap width 0.2 cm	Sigma Aldrich (St. Louis, USA)
Empore Octadecyl C18 47 mm Extraction disks	CDS Analytical (Oxford, USA)
Glass beads, acid washed (≤ 106 µm; - 140 U. S. sieve)	Sigma Aldrich (St. Louis, USA)
High Density Nickel Agarose	Jena Bioscience GmbH (Jena, Germany)
iBoltTM Empty Mini Gel Cassettes	Thermo Fisher Scientific (Waltham, USA)
iBlot™ 2 NC Regular Top & Bottom Stack	Thermo Fisher Scientific (Waltham, USA)
illustra™ Nap™-5 Columns Sephadex™ G-25 Grade	GE Healthcare Life Science (Marlborough, USA)
INTELLI-PLATE 96-3 low profile 3-well sitting drop	Art Robbins Instruments (Sunnyvale, USA)
Instant*Blue* coomassie stain	Expedeon Ltd. (Cambridgeshire, UK)

LC-MS: • Vials (1.5 mL) with PP screw cap ND9, 6 mm hole, silicon white/PTFE red septum, 55° shore A, 1.0 mm (548-0911) • Micro insert 0.1 mL 31 x 3.6 mm (548-0006)	VWR International GmbH (Darmstadt, Germany)
Microplate, 96 well, PS, F-Bottom (chimney well), µCLEAR®, black, medium binding (655096) for fluorescence measurements	Greiner Bio-One (Kremsmünster, Austria)
Microplate, 96 well, UV-Star®, COC, F-Bottom (chimney well), µCLEAR®, black (655809) for absorption measurements	Greiner Bio-One (Kremsmünster, Austria)
Microscopy: • Coverslips, ø 12 mm • Mounting Medium: Shandon Immu-Mount • Slides, ground edges 90 °	Carl Roth GmbH (Karlsruhe, Germany) Thermo Fisher Scientific (Waltham, USA) VWR International GmbH (Darmstadt, Germany)
Ni-NTA Spin Kit	Qiagen (Hilden, Germany)
Nitrile gloves, NITRILE Light	VWR International GmbH (Darmstadt, Germany)
p-aminobenzamidine agarose (saline suspension)	Sigma Aldrich (St. Louis, USA)
Pasteur pipettes	Th. Geyer GmbH & Co. KG (Renningen, Germany)
PD-10 Columns	GE Healthcare Life Science (Marlborough, USA)
Plastic consumables (falcon, reaction tube, serological pipettes, spatula, loops, etc.)	Sarstedt (Nümbrecht, Germany)
PEI (Polyethylenimine, branched: average M_W ~25 000 by LS average M_n ~10,000 by GPC, branched)	Sigma Aldrich (St. Louis, USA)
Petri dish Standard (10 cm^2)	Sarstedt (Nümbrecht, Germany)
Protein LoBind tubes, 2 mL	Eppendorf (Hamburg, Germany)
Protease Inhibitor Cocktail 5 MammCell/Tissue (100x)	Applichem (Darmstadt, Germany)
Quick Coomassie Stain	Serva Electrophoresis GmbH (Heidelberg, Germany)
ReadyLyzer 0.8 (MWCO 3.5 kDa)	Serva Electrophoresis GmbH (Heidelberg, Germany)
ReadyLyzer 3 (MWCO 3.5 kDa)	Serva Electrophoresis GmbH (Heidelberg, Germany)
Rotilabo Blotting paper, thickness 1.5 mm	Carl Roth GmbH + Co KG. (Karlsruhe, Germany)
Sep-Pak Vac 1cc (50 mg) iC18 Cartridges	Waters GmbH (Eschborn, Germany)
SnakeSkin® Dialysis Tubing (10 K MWCO, 16 mm dry I.D., 35 feet)	Thermo Fisher Scientific (Waltham, USA)
Strep-Tactin® XT Superflow® (50 % suspension)	IBA GmbH (Göttingen, Germany)
Strep-Tactin® XT Spin Columns	IBA GmbH (Göttingen, Germany)
Syringe NORM-JECT, Luer connection (1 mL; 2 mL; 5 mL; 10 mL; 20 mL; 50 mL)	Henke Sass Wolf GmbH (Tuttingen, Germany)
Silica gel, Ultra Pure, 40 - 60µm, 60Å	Acros Organics (Fisher Scientific – Schwerte, Germany)
Tips for pipettes	Starlab GmbH (Hamburg, Germany)
Silica gel 60 F254 TLC plates	Merck KGaA (Darmstadt, Germany)

TC-plate 6-well Standard F	Sarstedt (Nümbrecht, Germany)
TC dish 100 (10 cm^2) Standard	Sarstedt (Nümbrecht, Germany)
Ultrafree-MC-GV, centrifugal filters (Durapore - PVDF 0.22 µm)	Merck KGaA (Darmstadt, Germany)
UV cuvettes, micro	Brand GmbH + CO KG (Wertheim, Germany)
VIEWseal™, sealing film, transparent (676070)	Greiner Bio-One (Kremsmünster, Austria)
Xpress Micro Dialyzer MD300, MWCO 3.5 kDa	Scienova GmbH (Jena, Germany)
Xpress Micro Dialyzer MD1000, MWCO 3.5 kDa	Scienova GmbH (Jena, Germany)

5.4. Equipment

Table 62 Equipment

Equipment	Supplier
AccuBlock™ Digital Dry Bath D1110-230V	Labnet International Inc. (Edison, USA)
ÄKTA™ Pure (chromatography system)	GE Healthcare Life Science (Marlborough, USA)
ÄKTA™ Purifier 10 chromatography system	GE Healthcare Life Science (Marlborough, USA)
Analog Vortex Mixer	VWR International GmbH (Darmstadt, Germany)
Automated cell counter Contess™ II FL	Thermo Fisher Scientific (Waltham, USA)
Bio Imaging System GeneGenius (DNA)	SynGene (Cambridge, UK)
BioPhotometer 6131 for OD_{600} measurements	Eppendorf (Hamburg, Germany)
Branson SFX 250 Digital Sonifier	Emerson Electric Co. (St. Louis, USA)
Centrifuges: • Eppendorf 5804 R • Eppendorf 5424 • Heraeus Multifuge X1R • Heraeus Multifuge X3R • VWR Micro Star 17 • VWR Micro Star 17R	 Eppendorf (Hamburg, Germany) Eppendorf (Hamburg, Germany) Thermo Fisher Scientific (Waltham, USA) Thermo Fisher Scientific (Waltham, USA) VWR International GmbH (Darmstadt, Germany) VWR International GmbH (Darmstadt, Germany)
CO_2-Incubator (Serie CB) for cell culture	Binder GmbH (Tuttlingen, Germany)
Cold light source KL 1500 LCD	Olympus (Shinjuku, Japan)
Concentrator 5301	Eppendorf (Hamburg, Germany)
Electroporator MicroPulser	Bio-Rad Laboratories Inc. (Hercules, USA)
Fluorometer Fluoromax-4 with Temperature Controller LFI-3751 (5A, 40 W)	Horiba Jobin Yvon (Edison, USA) Wavelength Electronics (Bozeman, USA)
Forma™ 905 -86°C Upright Ultra-Low Temperature Freezer	Thermo Fisher Scientific (Waltham, USA)
HiLoad™ 16/600 Superdex 75 prep grade	GE Healthcare Life Science (Marlborough, USA)

HisTrap FF columns (1 mL)	GE Healthcare Life Science (Marlborough, USA)
HPLC: LC20-AT	Shimadzu (Kyoto, Japan)
iBlot™ 2 Dry Blotting System	Thermo Fisher Scientific (Waltham, USA)
ImageQuant LAS 4000 and LAS 4000mini	GE Healthcare Life Science (Marlborough, USA)
Incubator INCU-Line IL 115	VWR International GmbH (Darmstadt, Germany)
Incubation shaker Infors HT Multitron Standard	Infors AG (Bottmingen, Germany)
Incubation shaker Infors HT Ecotron	Infors AG (Bottmingen, Germany)
Jupiter C18 5 µm (2 x 150 mm) capillary column	Phenomenex (Torrence, USA)
Jupiter C4 5 µm (2 x 150 mm) capillary column	Phenomenex (Torrence, USA)
Luna C18, 5 µm (4.6 x 250 mm) column	Phenomenex (Torrence, USA)
Luna C18, 5 µm (10 x 250 mm) column	Phenomenex (Torrence, USA)
Luna C18, 10 µm (21.2 x 250 mm) column	Phenomenex (Torrence, USA)
LC-MS: 1260 Infinity 6130 Single Quadrupole	Agilent Technolgies (Santa Clara, USA)
Leica DMi8 fluorescence microscope	Leica Microsystems GmbH (Westlar, Germany)
LTQ-FT Ultra mass spectrometer Dionex Ultimate 3000 HPLC Instrument Massprep desalting cartridges	Thermo Fisher Scientific (Waltham, USA) Thermo Fisher Scientific (Waltham, USA) Waters GmbH (Eschborn, Germany)
Lyophilizer Christ Alpha 2-4 LDplus	Martin Christ Gefriertrocknungsanlagen GmbH (Osterode am Harz, Germany)
MicroCal PEAQ-ITC	Malvern Panalytical GmbH (Kassel Germany)
Microscope Eclipse TS100-F	Nikon (Minato, Japan)
Microwave oven (KMW 4441 DB)	Koenic/Imtron GmbH (Ingolstadt, Germany)
Mini Gel Tank	Thermo Fisher Scientific (Waltham, USA)
NanoPhotometer® N60 (spectrophotometer)	Implen GmbH (Munich, Germany)
NMR: AVHD300, AVHD400, AVHD500	Bruker (Billerica, USA)
NeXtal Tubes Suite (several different versions)	Qiagen (Hilden, Germany)
RESOURCE™ Q anion exchange chromatography column	GE Healthcare Life Science (Marlborough, USA)
Rotavapor: • Rotavapor R-300, R-215 • Membrane pump SC 920	Büchi Labortechnik AG (Flawil, Switzerland) KNF Neuberger GmbH (Freiburg, Germany)
PCR Cycler: T100 Thermal Cycler	Bio-Rad Laboratories Inc. (Hercules, USA)
PerfectBlue™ Gelsystem Mini S, Mini L and Midi S	VWR Life Science Competence Center (Erlangen, Germany)

PerfectBlue™ 'Semi-Dry' Electro Blotter Sedec M	VWR Life Science Competence Center (Erlangen, Germany)
Pipettes	Eppendorf (Hamburg, Germany)
PowerPac™ Power Supply Basic and HC (for SDS-PAGE)	Bio-Rad Laboratories Inc. (Hercules, USA)
Power Supply EV231 (for DNA electrophoresis)	Consort bvba (Turnhout, Belgium)
Precision scale ABJ220-4NM	Kern & Sohn GmbH (Balingen, Germany)
Protein Crystallization Robot Oryx4	Douglas Instruments Ltd (Berkshire, UK)
Protein Crystallography Dispenser Phoenix	Art Robbins Instruments (Sunnyvale, USA)
Prontosil C18 120-5-C18-AQ column	BISCHOFF Analysentechnik u. -geräte GmbH (Leonberg, Germany)
Scanning Spectrophotometer 7205 UV/Visible 72 Series Diode Array	Jenway (Staffordshire, United Kingdom)
Stereo Microscope OLYMPUS SZX10	Olympus (Shinjuku, Japan)
StrepTrap™ HP columns (1 mL)	GE Healthcare Life Science (Marlborough, USA)
Superdex™ 200 Increase 10/300 GL	GE Healthcare Life Science (Marlborough, USA)
Superdex™ 200 10/300 GL	GE Healthcare Life Science (Marlborough, USA)
Syringe pump LA30	Landgraf Laborsysteme GmbH (Langenhagen, Germany)
Tecan Spark 10M Microplate reader	Tecan Group Ltd (Männedorf, Switzerland)
Thermo-Incubation shaker Thriller	VWR Life Science Competence Center (Erlangen, Germany)
Tumbling roll mixer RM5	Ingenieurbüro CAT M.Zipperer GmbH (Ballrechten-Dottingen, Germany)
UV Lamp (handheld), UVGL-58, 1x 6 W, 254/365 nm	UVP (Cambridge, UK)
UV Lamp: VL-215.L, 2x 15 W, 365 nm	Vilber Lourmat (Marne-la-Vallée Cedex 3, France)
Varian Cary 50 UV-Vis Spectrophotometer	Agilent Technolgies (Santa Clara, USA)
Vortex Mixer Vortexer	Heathrow Scientific (Lakeview Pkwy, USA)
Water bath (26 L; 230 V) for cell culture	VWR Life Science Competence Center (Erlangen, Germany)
WB Imager Fusion Pulse 6	Vilber Lourmat (Marne-la-Vallée Cedex 3, France)

5.5. Software & Databases

Table 63 Used software, databases and web-tools

Software	Supplier
Adobe Illustrator CS4 14.0.0	Adobe (San José, USA)
Affinity Designer 1.7.3	Serif (Nottingham, UK)
ChemDraw Professional 16.0.1.4 (77)	Perkin Elmer Informatics Inc. (Waltham, USA)
Clustal Omega/clustalo 1.2.4 (https://www.ebi.ac.uk/Tools/msa/clustalo/, last accessed 21.05.2020)	EMBL-EBI (Hinxton, UK)
Converter Weight to Molar Quantity (http://www.molbiol.ru/ger/scripts/01_04.htmL, last accessed 20.06.2020)	www.molbiol.ru/ger
EvolutionCapt Pulse 6 17.03	Vilber Lourmat (Marne-la-Vallée Cedex 3, France)
ExPASy – Translate tool (https://web.expasy.org/translate/, last accessed 20.06.2020)	SIB Swiss Institute of Bioinformatics (Lausanne, Switzerland)
ExPASy – Compute pI/MW (https://web.expasy.org/compute_pi/, last accessed 20.06.2020)	SIB Swiss Institute of Bioinformatics (Lausanne, Switzerland)
ExPASy – ProtParam (https://web.expasy.org/protparam/, last accessed 20.06.2020)	SIB Swiss Institute of Bioinformatics (Lausanne, Switzerland)
ExPASy – PeptideMass (https://web.expasy.org/peptide_mass/, last accessed 30.05.2020)	SIB Swiss Institute of Bioinformatics (Lausanne, Switzerland)
FluorEssence v2.5.2.0 with Origin v.7.5878	Horiba Jobin Yvon (Edison, USA)
GeneSnap v7.12.06	SynGene (Cambridge, United Kingdom)
ImageReader LAS-4000 Control Software	GE Healthcare Life Science (Marlborough, USA)
ImageJ 1.52 (Fiji software)	Public domain (Wayne Rasband (NIH))
Leica application suite X (LAS X) software	Leica Microsystems GmbH (Westlar, Germany)
MestReNova v11.0.4-18998	Mestrelab Research, S.L. (Santiago de Compostela, Spain)
MicroCal PEAQ-ITC Control Software	Malvern Panalytical GmbH (Kassel Germany)
MicroCal PEAQ-ITC Analysis Software	Malvern Panalytical GmbH (Kassel Germany)
NEBaseChanger v1.2.8 (https://nebasechanger.neb.com/, last accessed 20.06.2020)	NEB Inc. (Ipswich, USA)

NEBcloner v1.3.8 (https://nebcloner.neb.com/#!/redigest, last accessed 20.06.2020)	NEB Inc. (Ipswich, USA)
NEB T_m Calculator v1.9.13 (https://tmcalculator.neb.com/#!/main, last accessed 20.06.2020)	NEB Inc. (Ipswich, USA)
OriginPro 2017G b9.4.1.354	OriginLab Corporation (Northampton, USA)
OpenLab CDS ChemStation Edition C.01.07.SR3 [465]	Agilent Technologies (Santa Clara, USA)
PrimerX (http://www.bioinformatics.org/primerx/, last accessed 03.06.2017)	Bioinformatics.org
Promega BioMath Calculator (https://www.promega.com/a/apps/biomath/, last accessed 10.05.2020)	Promega (Madison, USA)
PyMOL v1.8.6.1	Schrödinger, LLC (Cambridge, USA)
SciFinder	American Chemical Society (Washington, DC, USA)
SnapGene 3.1.4	GSL Biotech LLC (Chicago, USA)
SparkControl v2.1	Tecan Group Ltd (Männedorf, Switzerland)
Thermo Xcalibur Xtract software	Thermo Fisher Scientific (Waltham, USA)

6.REFERENCES

1. Crick, F. H., On protein synthesis. *Symp Soc Exp Biol* **1958,** *12*, 138-63.
2. Astbury, W. T., Molecular biology or ultrastructural biology? *Nature* **1961,** *190*, 1124.
3. Astbury, W. T.; Bell, F. O., X-Ray Study of Thymonucleic Acid. *Nature* **1938,** *141* (3573), 747-748.
4. Watson, J. D.; Crick, F. H., Genetical implications of the structure of deoxyribonucleic acid. *Nature* **1953,** *171* (4361), 964-7.
5. Franklin, R. E.; Gosling, R. G., Molecular Configuration in Sodium Thymonucleate. *Nature* **1953,** *171* (4356), 740-741.
6. Uzawa, T.; Yamagishi, A.; Oshima, T., Polypeptide synthesis directed by DNA as a messenger in cell-free polypeptide synthesis by extreme thermophiles, Thermus thermophilus HB27 and Sulfolobus tokodaii strain 7. *J Biochem* **2002,** *131* (6), 849-53.
7. McCarthy, B. J.; Holland, J. J., Denatured DNA as a direct template for in vitro protein synthesis. *Proc Natl Acad Sci U S A* **1965,** *54* (3), 880-6.
8. Temin, H. M.; Mizutani, S., Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus. *Nature* **1970,** *226* (5252), 1211-1213.
9. Ahlquist, P., RNA-Dependent RNA Polymerases, Viruses, and RNA Silencing. *Science* **2002,** *296* (5571), 1270.
10. Zhang, Y., Transcriptional regulation by histone ubiquitination and deubiquitination. *Genes Dev* **2003,** *17* (22), 2733-40.
11. Berger, S. L., An embarrassment of niches: the many covalent modifications of histones in transcriptional regulation. *Oncogene* **2001,** *20* (24), 3007-13.
12. O'Donnell, M.; Langston, L.; Stillman, B., Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb Perspect Biol* **2013,** *5* (7).
13. Papasaikas, P.; Valcárcel, J., The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem Sci* **2016,** *41* (1), 33-45.
14. Ramanathan, A.; Robb, G. B.; Chan, S. H., mRNA capping: biological functions and applications. *Nucleic Acids Res* **2016,** *44* (16), 7511-26.
15. Hajnsdorf, E.; Kaberdin, V. R., RNA polyadenylation and its consequences in prokaryotes. *Philos Trans R Soc Lond B Biol Sci* **2018,** *373* (1762).
16. Macek, B.; Forchhammer, K.; Hardouin, J.; Weber-Ban, E.; Grangeasse, C.; Mijakovic, I., Protein post-translational modifications in bacteria. *Nat Rev Microbiol* **2019,** *17* (11), 651-664.
17. Swatek, K. N.; Komander, D., Ubiquitin modifications. *Cell Res* **2016,** *26* (4), 399-422.
18. Harris, E. D. In *Biochemical Facts behind the Definition and Properties of METABOLITES*, 2003.
19. Kurland, C. G., Translational accuracy and the fitness of bacteria. *Annu Rev Genet* **1992,** *26*, 29-50.
20. Jakubowski, H.; Goldman, E., Editing of errors in selection of amino acids for protein synthesis. *Microbiol Rev* **1992,** *56* (3), 412-29.
21. Berg, J. M.; Tymoczko, J. L.; Stryer, L., *Zusammensetzung und Struktur von Proteinen - Stryer Biochemie*. 6 ed.; Springer Spektrum Akademischer Verlag: Heidelberg, 2010.
22. Berg, J. M.; Tymoczko, J. L.; Stryer, L., *DNA, RNA und der Fluss der genetischen Information - Stryer Biochemie* 6ed.; Springer Spektrum Akademischer Verlag: Heidelberg, 2010.
23. Berg, J. M.; Tymoczko, J. L.; Stryer, L., *Proteinsynthese - Stryer Biochemie*. 6 ed.; Springer Spektrum Akademischer Verlag: Heidelberg, 2010.
24. Ambrogelly, A.; Palioura, S.; Söll, D., Natural expansion of the genetic code. *Nature Chemical Biology* **2007,** *3* (1), 29-35.

25. Ngo, J. T.; Tirrell, D. A., Noncanonical Amino Acids in the Interrogation of Cellular Protein Synthesis. *Accounts of Chemical Research* **2011,** *44* (9), 677-685.
26. Nowatzki, P. J.; Franck, C.; Maskarinec, S. A.; Ravichandran, G.; Tirrell, D. A., Mechanically Tunable Thin Films of Photosensitive Artificial Proteins: Preparation and Characterization by Nanoindentation. *Macromolecules* **2008,** *41* (5), 1839-1845.
27. Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R., Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proceedings of the National Academy of Sciences* **2002,** *99* (1), 19.
28. Kiick, K. L.; Tirrell, D. A., Protein Engineering by In Vivo Incorporation of Non-Natural Amino Acids: Control of Incorporation of Methionine Analogues by Methionyl-tRNA Synthetase. *Tetrahedron* **2000,** *56* (48), 9487-9493.
29. van Hest, J. C.; Tirrell, D. A., Efficient introduction of alkene functionality into proteins in vivo. *FEBS Lett* **1998,** *428* (1-2), 68-70.
30. Tang, Y.; Ghirlanda, G.; Petka, W. A.; Nakajima, T.; DeGrado, W. F.; Tirrell, D. A., Fluorinated Coiled-Coil Proteins Prepared In Vivo Display Enhanced Thermal and Chemical Stability. *Angew Chem Int Ed Engl* **2001,** *40* (8), 1494-1496.
31. Wang, P.; Tang, Y.; Tirrell, D. A., Incorporation of Trifluoroisoleucine into Proteins in Vivo. *Journal of the American Chemical Society* **2003,** *125* (23), 6900-6906.
32. Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A., Genetically Engineered Fluoropolymers. Synthesis of Repetitive Polypeptides Containing p-Fluorophenylalanine residues. *Macromolecules* **1994,** *27* (19), 5471-5475.
33. Barton, W. A.; Tzvetkova-Robev, D.; Erdjument-Bromage, H.; Tempst, P.; Nikolov, D. B., Highly efficient selenomethionine labeling of recombinant proteins produced in mammalian cells. *Protein Sci* **2006,** *15* (8), 2008-2013.
34. Suchanek, M.; Radzikowska, A.; Thiele, C., Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nature Methods* **2005,** *2* (4), 261-268.
35. Hong, S. H.; Kwon, Y. C.; Jewett, M. C., Non-standard amino acid incorporation into proteins using Escherichia coli cell-free protein synthesis. *Front Chem* **2014,** *2*, 34.
36. Khambhati, K.; Bhattacharjee, G.; Gohil, N.; Braddick, D.; Kulkarni, V.; Singh, V., Exploring the Potential of Cell-Free Protein Synthesis for Extending the Abilities of Biological Systems. *Frontiers in Bioengineering and Biotechnology* **2019,** *7* (248).
37. Ceelen, L. M.; Decostere, A.; Ducatelle, R.; Haesebrouck, F., Cytolethal distending toxin generates cell death by inducing a bottleneck in the cell cycle. *Microbiol Res* **2006,** *161* (2), 109-120.
38. Orbán, E.; Proverbio, D.; Haberstock, S.; Dötsch, V.; Bernhard, F., Cell-free expression of G-protein-coupled receptors. *Methods Mol Biol* **2015,** *1261*, 171-95.
39. Worst, E. G.; Exner, M. P.; De Simone, A.; Schenkelberger, M.; Noireaux, V.; Budisa, N.; Ott, A., Cell-free expression with the toxic amino acid canavanine. *Bioorg Med Chem Lett* **2015,** *25* (17), 3658-60.
40. Carlson, E. D.; Gan, R.; Hodgman, C. E.; Jewett, M. C., Cell-free protein synthesis: applications come of age. *Biotechnol Adv* **2012,** *30* (5), 1185-1194.
41. Chin, J. W., Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annual Review of Biochemistry* **2014,** *83* (1), 379-408.
42. Venkat, S.; Chen, H.; Gan, Q.; Fan, C., The Application of Cell-Free Protein Synthesis in Genetic Code Expansion for Post-translational Modifications. *Front Pharmacol* **2019,** *10*, 248-248.
43. Chemla, Y.; Ozer, E.; Schlesinger, O.; Noireaux, V.; Alfonta, L., Genetically expanded cell-free protein synthesis using endogenous pyrrolysyl orthogonal translation system. *Biotechnol Bioeng* **2015,** *112* (8), 1663-72.

44. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the genetic code of Escherichia coli. *Science* **2001,** *292* (5516), 498-500.
45. Scheidler, C. M.; Vrabel, M.; Schneider, S., Genetic Code Expansion, Protein Expression, and Protein Functionalization in Bacillus subtilis. *ACS Synth Biol* **2020,** *9* (3), 486-493.
46. Wang, F.; Robbins, S.; Guo, J.; Shen, W.; Schultz, P. G., Genetic incorporation of unnatural amino acids into proteins in Mycobacterium tuberculosis. *PloS one* **2010,** *5* (2), e9354-e9354.
47. Lemke, E. A.; Summerer, D.; Geierstanger, B. H.; Brittain, S. M.; Schultz, P. G., Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat Chem Biol* **2007,** *3* (12), 769-72.
48. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G., An Expanded Eukaryotic Genetic Code. *Science* **2003,** *301* (5635), 964.
49. Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W., Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat Chem* **2012,** *4* (4), 298-304.
50. Greiss, S.; Chin, J. W., Expanding the Genetic Code of an Animal. *Journal of the American Chemical Society* **2011,** *133* (36), 14196-14199.
51. Li, F.; Zhang, H.; Sun, Y.; Pan, Y.; Zhou, J.; Wang, J., Expanding the Genetic Code for Photoclick Chemistry in E. coli, Mammalian Cells, and A. thaliana. *Angewandte Chemie International Edition* **2013,** *52* (37), 9700-9704.
52. Vargas-Rodriguez, O.; Sevostyanova, A.; Söll, D.; Crnković, A., Upgrading aminoacyl-tRNA synthetases for genetic code expansion. *Curr Opin Chem Biol* **2018,** *46*, 115-122.
53. Fath, M. J.; Kolter, R., ABC transporters: bacterial exporters. *Microbiological Reviews* **1993,** *57* (4), 995.
54. Yang, N. J.; Hinner, M. J., Getting across the cell membrane: an overview for small molecules, peptides, and proteins. *Methods in molecular biology (Clifton, N.J.)* **2015,** *1266*, 29-53.
55. Berg, J. M.; Tymoczko, J. L.; Stryer, L., *Membrankanäle und -pumpen - Stryer Biochemie*. 6 ed.; Springer Spektrum Akademischer Verlag: Heidelberg, 2010.
56. Luo, X.; Fu, G.; Wang, R. E.; Zhu, X.; Zambaldo, C.; Liu, R.; Liu, T.; Lyu, X.; Du, J.; Xuan, W.; Yao, A.; Reed, S. A.; Kang, M.; Zhang, Y.; Guo, H.; Huang, C.; Yang, P.-Y.; Wilson, I. A.; Schultz, P. G.; Wang, F., Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. *Nature chemical biology* **2017,** *13* (8), 845-849.
57. Payne, J. W.; Gilvarg, C., The role of the terminal carboxyl group on peptide transport in Escherichia coli. *J Biol Chem* **1968,** *243* (2), 335-40.
58. Verkamp, E.; Backman, V. M.; Björnsson, J. M.; Söll, D.; Eggertsson, G., The periplasmic dipeptide permease system transports 5-aminolevulinic acid in Escherichia coli. *J Bacteriol* **1993,** *175* (5), 1452-6.
59. Smith, M. W.; Tyreman, D. R.; Payne, G. M.; Marshall, N. J.; Payne, J. W., Substrate specificity of the periplasmic dipeptide-binding protein from Escherichia coli: experimental basis for the design of peptide prodrugs. *Microbiology* **1999,** *145 (Pt 10)*, 2891-901.
60. Ko, W.; Kumar, R.; Kim, S.; Lee, H. S., Construction of Bacterial Cells with an Active Transport System for Unnatural Amino Acids. *ACS Synth Biol* **2019,** *8* (5), 1195-1203.
61. Zhou, H.; Cheung, J. W.; Carpenter, T.; Jones, S. K.; Luong, N. H.; Tran, N. C.; Jacobs, S. E.; Galbada Liyanage, S. A.; Cropp, T. A.; Yin, J., Enhancing the incorporation of lysine derivatives into proteins with methylester forms of unnatural amino acids. *Bioorganic & Medicinal Chemistry Letters* **2020,** *30* (2), 126876.
62. Takimoto, J. K.; Xiang, Z.; Kang, J.-Y.; Wang, L., Esterification of an unnatural amino acid structurally deviating from canonical amino acids promotes its uptake and incorporation

into proteins in mammalian cells. *Chembiochem : a European journal of chemical biology* **2010,** *11* (16), 2268-2272.
63. Mehl, R. A.; Anderson, J. C.; Santoro, S. W.; Wang, L.; Martin, A. B.; King, D. S.; Horn, D. M.; Schultz, P. G., Generation of a Bacterium with a 21 Amino Acid Genetic Code. *Journal of the American Chemical Society* **2003,** *125* (4), 935-939.
64. 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.
65. Kirilov, K. T.; Golshani, A.; Ivanov, I. G., Termination Codons and Stop Codon Context in Bacteria and Mammalian Mitochondria. *Biotechnology & Biotechnological Equipment* **2013,** *27* (4), 4018-4025.
66. Xie, J.; Schultz, P. G., An expanding genetic code. *Methods* **2005,** *36* (3), 227-38.
67. Zhang, Z.; Alfonta, L.; Tian, F.; Bursulaya, B.; Uryu, S.; King, D. S.; Schultz, P. G., Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc Natl Acad Sci U S A* **2004,** *101* (24), 8882-7.
68. Anderson, J. C.; Schultz, P. G., Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. *Biochemistry* **2003,** *42* (32), 9598-608.
69. Köhrer, C.; Yoo, J. H.; Bennett, M.; Schaack, J.; RajBhandary, U. L., A possible approach to site-specific insertion of two different unnatural amino acids into proteins in mammalian cells via nonsense suppression. *Chem Biol* **2003,** *10* (11), 1095-102.
70. Wan, W.; Huang, Y.; Wang, Z.; Russell, W. K.; Pai, P. J.; Russell, D. H.; Liu, W. R., A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. *Angew Chem Int Ed Engl* **2010,** *49* (18), 3211-4.
71. 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.
72. Xiao, H.; Chatterjee, A.; Choi, S.-h.; Bajjuri, K. M.; Sinha, S. C.; Schultz, P. G., Genetic Incorporation of Multiple Unnatural Amino Acids into Proteins in Mammalian Cells. *Angewandte Chemie International Edition* **2013,** *52* (52), 14080-14083.
73. Italia, J. S.; Addy, P. S.; Erickson, S. B.; Peeler, J. C.; Weerapana, E.; Chatterjee, A., Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites. *J Am Chem Soc* **2019,** *141* (15), 6204-6212.
74. Zeng, Y.; Wang, W.; Liu, W. R., Towards reassigning the rare AGG codon in Escherichia coli. *Chembiochem* **2014,** *15* (12), 1750-4.
75. Mukai, T.; Yamaguchi, A.; Ohtake, K.; Takahashi, M.; Hayashi, A.; Iraha, F.; Kira, S.; Yanagisawa, T.; Yokoyama, S.; Hoshi, H.; Kobayashi, T.; Sakamoto, K., Reassignment of a rare sense codon to a non-canonical amino acid in Escherichia coli. *Nucleic acids research* **2015,** *43* (16), 8111-8122.
76. Kwon, I.; Kirshenbaum, K.; Tirrell, D. A., Breaking the Degeneracy of the Genetic Code. *Journal of the American Chemical Society* **2003,** *125* (25), 7512-7513.
77. Sheppard, K.; Yuan, J.; Hohn, M. J.; Jester, B.; Devine, K. M.; Söll, D., From one amino acid to another: tRNA-dependent amino acid biosynthesis. *Nucleic Acids Res* **2008,** *36* (6), 1813-25.
78. O'Donoghue, P.; Sheppard, K.; Nureki, O.; Söll, D., Rational design of an evolutionary precursor of glutaminyl-tRNA synthetase. *Proceedings of the National Academy of Sciences of the United States of America* **2011,** *108* (51), 20485-20490.
79. Hohsaka, T.; Kajihara, D.; Ashizuka, Y.; Murakami, H.; Sisido, M., Efficient Incorporation of Nonnatural Amino Acids with Large Aromatic Groups into Streptavidin in In Vitro Protein Synthesizing Systems. *Journal of the American Chemical Society* **1999,** *121* (1), 34-40.

80. Hohsaka, T.; Ashizuka, Y.; Murakami, H.; Sisido, M., Incorporation of Nonnatural Amino Acids into Streptavidin through In Vitro Frame-Shift Suppression. *Journal of the American Chemical Society* **1996,** *118* (40), 9778-9779.
81. 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), 441-4.
82. Wang, K.; Neumann, H.; Peak-Chew, S. Y.; Chin, J. W., Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat Biotechnol* **2007,** *25* (7), 770-7.
83. Li, L.; Degardin, M.; Lavergne, T.; Malyshev, D. A.; Dhami, K.; Ordoukhanian, P.; Romesberg, F. E., Natural-like replication of an unnatural base pair for the expansion of the genetic alphabet and biotechnology applications. *J Am Chem Soc* **2014,** *136* (3), 826-9.
84. Morris, S. E.; Feldman, A. W.; Romesberg, F. E., Synthetic Biology Parts for the Storage of Increased Genetic Information in Cells. *ACS Synth Biol* **2017,** *6* (10), 1834-1840.
85. Seo, Y. J.; Hwang, G. T.; Ordoukhanian, P.; Romesberg, F. E., Optimization of an unnatural base pair toward natural-like replication. *J Am Chem Soc* **2009,** *131* (9), 3246-52.
86. Feldman, A. W.; Romesberg, F. E., In Vivo Structure-Activity Relationships and Optimization of an Unnatural Base Pair for Replication in a Semi-Synthetic Organism. *J Am Chem Soc* **2017,** *139* (33), 11427-11433.
87. Ast, M.; Gruber, A.; Schmitz-Esser, S.; Neuhaus, H. E.; Kroth, P. G.; Horn, M.; Haferkamp, I., Diatom plastids depend on nucleotide import from the cytosol. *Proc Natl Acad Sci U S A* **2009,** *106* (9), 3621-6.
88. Feldman, A. W.; Fischer, E. C.; Ledbetter, M. P.; Liao, J. Y.; Chaput, J. C.; Romesberg, F. E., A Tool for the Import of Natural and Unnatural Nucleoside Triphosphates into Bacteria. *J Am Chem Soc* **2018,** *140* (4), 1447-1454.
89. Ledbetter, M. P.; Karadeema, R. J.; Romesberg, F. E., Reprograming the Replisome of a Semisynthetic Organism for the Expansion of the Genetic Alphabet. *J Am Chem Soc* **2018,** *140* (2), 758-765.
90. Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Corrêa, I. R., Jr.; Romesberg, F. E., A semi-synthetic organism with an expanded genetic alphabet. *Nature* **2014,** *509* (7500), 385-8.
91. Zhang, Y.; Lamb, B. M.; Feldman, A. W.; Zhou, A. X.; Lavergne, T.; Li, L.; Romesberg, F. E., A semisynthetic organism engineered for the stable expansion of the genetic alphabet. *Proc Natl Acad Sci U S A* **2017,** *114* (6), 1317-1322.
92. Zhang, Y.; Ptacin, J. L.; Fischer, E. C.; Aerni, H. R.; Caffaro, C. E.; San Jose, K.; Feldman, A. W.; Turner, C. R.; Romesberg, F. E., A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* **2017,** *551* (7682), 644-647.
93. 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. *Nature Chemical Biology* **2020**.
94. Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G., A New Functional Suppressor tRNA/Aminoacyl−tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins. *Journal of the American Chemical Society* **2000,** *122* (20), 5010-5011.
95. Steer, B. A.; Schimmel, P., Major anticodon-binding region missing from an archaebacterial tRNA synthetase. *J Biol Chem* **1999,** *274* (50), 35601-6.
96. Davis, L.; Chin, J. W., Designer proteins: applications of genetic code expansion in cell biology. *Nature Reviews Molecular Cell Biology* **2012,** *13* (3), 168-182.
97. Willis, J. C. W.; Chin, J. W., Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat Chem* **2018,** *10* (8), 831-837.

98. Meineke, B.; Heimgärtner, J.; Lafranchi, L.; Elsässer, S. J., Methanomethylophilus alvus Mx1201 Provides Basis for Mutual Orthogonal Pyrrolysyl tRNA/Aminoacyl-tRNA Synthetase Pairs in Mammalian Cells. *ACS Chem Biol* **2018,** *13* (11), 3087-3096.
99. Wan, W.; Huang, Y.; Wang, Z.; Russell, W. K.; Pai, P.-J.; Russell, D. H.; Liu, W. R., A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in Escherichia coli. *Angewandte Chemie International Edition* **2010,** *49* (18), 3211-3214.
100. 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), 3197-203.
101. Nozawa, K.; Ishitani, R.; Nureki, O., [Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality]. *Seikagaku* **2010,** *82* (7), 617-23.
102. 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. *The Journal of biological chemistry* **2012,** *287* (39), 32738-32746.
103. Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Nureki, O.; Yokoyama, S., Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon Methanosarcina mazei. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2006,** *62* (Pt 10), 1031-3.
104. 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.
105. 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. *Nature Chemical Biology* **2017,** *13* (12), 1261-1266.
106. Yanagisawa, T.; Sumida, T.; Ishii, R.; Yokoyama, S., A novel crystal form of pyrrolysyl-tRNA synthetase reveals the pre- and post-aminoacyl-tRNA synthesis conformational states of the adenylate and aminoacyl moieties and an asparagine residue in the catalytic site. *Acta Crystallogr D Biol Crystallogr* **2013,** *69* (Pt 1), 5-15.
107. Wan, W.; Tharp, J. M.; Liu, W. R., Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim Biophys Acta* **2014,** *1844* (6), 1059-1070.
108. Yu, Z.; Pan, Y.; Wang, Z.; Wang, J.; Lin, Q., Genetically encoded cyclopropene directs rapid, photoclick-chemistry-mediated protein labeling in mammalian cells. *Angew Chem Int Ed Engl* **2012,** *51* (42), 10600-4.
109. 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. *Journal of Molecular Biology* **2008,** *378* (3), 634-652.
110. Ambrogelly, A.; Gundllapalli, S.; Herring, S.; Polycarpo, C.; Frauer, C.; Söll, D., Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc Natl Acad Sci U S A* **2007,** *104* (9), 3141-6.
111. Elliott, T. S.; Townsley, F. M.; Bianco, A.; Ernst, R. J.; Sachdeva, A.; Elsässer, S. J.; Davis, L.; Lang, K.; Pisa, R.; Greiss, S.; Lilley, K. S.; Chin, J. W., Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. *Nature biotechnology* **2014,** *32* (5), 465-472.
112. Wang, K.; Sachdeva, A.; Cox, D. J.; Wilf, N. M.; Lang, K.; Wallace, S.; Mehl, R. A.; Chin, J. W., Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat Chem* **2014,** *6* (5), 393-403.
113. Tharp, J. M.; Ehnbom, A.; Liu, W. R., tRNAPyl: Structure, function, and applications. *RNA Biology* **2018,** *15* (4-5), 441-452.

114. Wang, L.; Schultz, P. G., A general approach for the generation of orthogonal tRNAs. *Chemistry & Biology* **2001,** *8* (9), 883-890.
115. Hanahan, D.; Jessee, J.; Bloom, F. R., [4] Plasmid transformation of Escherichia coli and other bacteria. In *Methods in Enzymology*, Academic Press: 1991; Vol. 204, pp 63-113.
116. Umehara, T.; Kim, J.; Lee, S.; Guo, L. T.; Söll, D.; Park, H. S., N-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo. *FEBS Lett* **2012,** *586* (6), 729-33.
117. Hohl, A.; Karan, R.; Akal, A.; Renn, D.; Liu, X.; Ghorpade, S.; Groll, M.; Rueping, M.; Eppinger, J., Engineering a Polyspecific Pyrrolysyl-tRNA Synthetase by a High Throughput FACS Screen. *Scientific Reports* **2019,** *9* (1), 11971.
118. Lin, A. E.; Lin, Q., Rapid Identification of Functional Pyrrolysyl-tRNA Synthetases via Fluorescence-Activated Cell Sorting. *Int J Mol Sci* **2018,** *20* (1), 29.
119. Miao, H.; Yu, C.; Yao, A.; Xuan, W., Rational design of a function-based selection method for genetically encoding acylated lysine derivatives. *Organic & Biomolecular Chemistry* **2019,** *17* (25), 6127-6130.
120. Bryson, D. I.; Fan, C.; Guo, L.-T.; Miller, C.; Söll, D.; Liu, D. R., Continuous directed evolution of aminoacyl-tRNA synthetases. *Nature Chemical Biology* **2017,** *13* (12), 1253-1260.
121. 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), e156.
122. Förster, C.; Limmer, S.; Ribeiro, S.; Hilgenfeld, R.; Sprinzl, M., Ternary complex between elongation factor Tu.GTP and Phe-tRNA(Phe). *Biochimie* **1993,** *75* (12), 1159-66.
123. Wang, J.; Kwiatkowski, M.; Forster, A. C., Kinetics of tRNA(Pyl) -mediated amber suppression in Escherichia coli translation reveals unexpected limiting steps and competing reactions. *Biotechnol Bioeng* **2016,** *113* (7), 1552-9.
124. Eargle, J.; Black, A. A.; Sethi, A.; Trabuco, L. G.; Luthey-Schulten, Z., Dynamics of Recognition between tRNA and elongation factor Tu. *J Mol Biol* **2008,** *377* (5), 1382-405.
125. Italia, J. S.; Latour, C.; Wrobel, C. J. J.; Chatterjee, A., Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes. *Cell Chem Biol* **2018,** *25* (10), 1304-1312.e5.
126. Italia, J. S.; Addy, P. S.; Wrobel, C. J.; Crawford, L. A.; Lajoie, M. J.; Zheng, Y.; Chatterjee, A., An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nat Chem Biol* **2017,** *13* (4), 446-450.
127. Isaacs, F. J.; Carr, P. A.; Wang, H. H.; Lajoie, M. J.; Sterling, B.; Kraal, L.; Tolonen, A. C.; Gianoulis, T. A.; Goodman, D. B.; Reppas, N. B.; Emig, C. J.; Bang, D.; Hwang, S. J.; Jewett, M. C.; Jacobson, J. M.; Church, G. M., Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **2011,** *333* (6040), 348-53.
128. 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* **2013,** *342* (6156), 357-60.
129. Ostrov, N.; Landon, M.; Guell, M.; Kuznetsov, G.; Teramoto, J.; Cervantes, N.; Zhou, M.; Singh, K.; Napolitano, M. G.; Moosburner, M.; Shrock, E.; Pruitt, B. W.; Conway, N.; Goodman, D. B.; Gardner, C. L.; Tyree, G.; Gonzales, A.; Wanner, B. L.; Norville, J. E.; Lajoie, M. J.; Church, G. M., Design, synthesis, and testing toward a 57-codon genome. *Science* **2016,** *353* (6301), 819-22.
130. Fredens, J.; Wang, K.; de la Torre, D.; Funke, L. F. H.; Robertson, W. E.; Christova, Y.; Chia, T.; Schmied, W. H.; Dunkelmann, D. L.; Beránek, V.; Uttamapinant, C.; Llamazares, A. G.; Elliott, T. S.; Chin, J. W., Total synthesis of Escherichia coli with a recoded genome. *Nature* **2019,** *569* (7757), 514-518.

131. Wang, K.; Fredens, J.; Brunner, S. F.; Kim, S. H.; Chia, T.; Chin, J. W., Defining synonymous codon compression schemes by genome recoding. *Nature* **2016,** *539* (7627), 59-64.
132. 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), 344-54.
133. Tuller, T.; Waldman, Y. Y.; Kupiec, M.; Ruppin, E., Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci U S A* **2010,** *107* (8), 3645-50.
134. Goodman, D. B.; Church, G. M.; Kosuri, S., Causes and effects of N-terminal codon bias in bacterial genes. *Science* **2013,** *342* (6157), 475-9.
135. Hutchison, C. A., 3rd; Chuang, R. Y.; Noskov, V. N.; Assad-Garcia, N.; Deerinck, T. J.; Ellisman, M. H.; Gill, J.; Kannan, K.; Karas, B. J.; Ma, L.; Pelletier, J. F.; Qi, Z. Q.; Richter, R. A.; Strychalski, E. A.; Sun, L.; Suzuki, Y.; Tsvetanova, B.; Wise, K. S.; Smith, H. O.; Glass, J. I.; Merryman, C.; Gibson, D. G.; Venter, J. C., Design and synthesis of a minimal bacterial genome. *Science* **2016,** *351* (6280), aad6253.
136. Lang, K.; Chin, J. W., Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chemical Reviews* **2014,** *114* (9), 4764-4806.
137. Saxon, E.; Bertozzi, C. R., Cell Surface Engineering by a Modified Staudinger Reaction. *Science* **2000,** *287* (5460), 2007.
138. Borrmann, A.; Milles, S.; Plass, T.; Dommerholt, J.; Verkade, J. M. M.; Wießler, M.; Schultz, C.; van Hest, J. C. M.; van Delft, F. L.; Lemke, E. A., Genetic Encoding of a Bicyclo[6.1.0]nonyne-Charged Amino Acid Enables Fast Cellular Protein Imaging by Metal-Free Ligation. *ChemBioChem* **2012,** *13* (14), 2094-2099.
139. Plass, T.; Milles, S.; Koehler, C.; Szymański, J.; Mueller, R.; Wießler, M.; Schultz, C.; Lemke, E. A., Amino Acids for Diels–Alder Reactions in Living Cells. *Angewandte Chemie International Edition* **2012,** *51* (17), 4166-4170.
140. Uttamapinant, C.; Howe, J. D.; Lang, K.; Beránek, V.; Davis, L.; Mahesh, M.; Barry, N. P.; Chin, J. W., Genetic Code Expansion Enables Live-Cell and Super-Resolution Imaging of Site-Specifically Labeled Cellular Proteins. *Journal of the American Chemical Society* **2015,** *137* (14), 4602-4605.
141. Mayer, S. V.; Murnauer, A.; von Wrisberg, M.-K.; Jokisch, M.-L.; Lang, K., Photo-induced and Rapid Labeling of Tetrazine-Bearing Proteins via Cyclopropenone-Caged Bicyclononynes. *Angewandte Chemie International Edition* **2019,** *58* (44), 15876-15882.
142. Tsai, Y.-H.; Essig, S.; James, J. R.; Lang, K.; Chin, J. W., Selective, rapid and optically switchable regulation of protein function in live mammalian cells. *Nature chemistry* **2015,** *7* (7), 554-561.
143. 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. *Journal of the American Chemical Society* **2008,** *130* (29), 9268-9281.
144. Weeks, C. L.; Polishchuk, A.; Getahun, Z.; Degrado, W. F.; Spiro, T. G., Investigation of an unnatural amino acid for use as a resonance Raman probe: Detection limits, solvent and temperature dependence of the νC≡N band of 4-cyanophenylalanine. *J Raman Spectrosc* **2008,** *39* (11), 1606-1613.
145. Wang, J.; Xie, J.; Schultz, P. G., A Genetically Encoded Fluorescent Amino Acid. *Journal of the American Chemical Society* **2006,** *128* (27), 8738-8739.
146. Chatterjee, A.; Guo, J.; Lee, H. S.; Schultz, P. G., A Genetically Encoded Fluorescent Probe in Mammalian Cells. *Journal of the American Chemical Society* **2013,** *135* (34), 12540-12543.

147. Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; Ge, Y.; Gunawardena, J.; Hendrickson, R. C.; Hergenrother, P. J.; Huber, C. G.; Ivanov, A. R.; Jensen, O. N.; Jewett, M. C.; Kelleher, N. L.; Kiessling, L. L.; Krogan, N. J.; Larsen, M. R.; Loo, J. A.; Ogorzalek Loo, R. R.; Lundberg, E.; MacCoss, M. J.; Mallick, P.; Mootha, V. K.; Mrksich, M.; Muir, T. W.; Patrie, S. M.; Pesavento, J. J.; Pitteri, S. J.; Rodriguez, H.; Saghatelian, A.; Sandoval, W.; Schlüter, H.; Sechi, S.; Slavoff, S. A.; Smith, L. M.; Snyder, M. P.; Thomas, P. M.; Uhlén, M.; Van Eyk, J. E.; Vidal, M.; Walt, D. R.; White, F. M.; Williams, E. R.; Wohlschlager, T.; Wysocki, V. H.; Yates, N. A.; Young, N. L.; Zhang, B., How many human proteoforms are there? *Nat Chem Biol* **2018,** *14* (3), 206-214.
148. Sadakierska-Chudy, A.; Filip, M., A comprehensive view of the epigenetic landscape. Part II: Histone post-translational modification, nucleosome level, and chromatin regulation by ncRNAs. *Neurotox Res* **2015,** *27* (2), 172-197.
149. Elsässer, S. J.; Ernst, R. J.; Walker, O. S.; Chin, J. W., Genetic code expansion in stable cell lines enables encoded chromatin modification. *Nature Methods* **2016,** *13* (2), 158-164.
150. Kim, C. H.; Kang, M.; Kim, H. J.; Chatterjee, A.; Schultz, P. G., Site-specific incorporation of ε-N-crotonyllysine into histones. *Angewandte Chemie (International ed. in English)* **2012,** *51* (29), 7246-7249.
151. Rehn, A.; Lawatscheck, J.; Jokisch, M.-L.; Mader, S. L.; Luo, Q.; Tippel, F.; Blank, B.; Richter, K.; Lang, K.; Kaila, V. R. I.; Buchner, J., A methylated lysine is a switch point for conformational communication in the chaperone Hsp90. *Nature Communications* **2020,** *11* (1), 1219.
152. Tarrant, M. K.; Cole, P. A., The chemical biology of protein phosphorylation. *Annu Rev Biochem* **2009,** *78*, 797-825.
153. Fan, C.; Ip, K.; Söll, D., Expanding the genetic code of Escherichia coli with phosphotyrosine. *FEBS letters* **2016,** *590* (17), 3040-3047.
154. Zhang, M. S.; Brunner, S. F.; Huguenin-Dezot, N.; Liang, A. D.; Schmied, W. H.; Rogerson, D. T.; Chin, J. W., Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nature methods* **2017,** *14* (7), 729-736.
155. Fottner, M.; Brunner, A. D.; Bittl, V.; Horn-Ghetko, D.; Jussupow, A.; Kaila, V. R. I.; Bremm, A.; Lang, K., Site-specific ubiquitylation and SUMOylation using genetic-code expansion and sortase. *Nat Chem Biol* **2019,** *15* (3), 276-284.
156. Ryan, D. P.; Matthews, J. M., Protein–protein interactions in human disease. *Current Opinion in Structural Biology* **2005,** *15* (4), 441-446.
157. 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. *Proceedings of the National Academy of Sciences* **2002,** *99* (17), 11020.
158. Chou, C.; Uprety, R.; Davis, L.; Chin, J. W.; Deiters, A., Genetically encoding an aliphatic diazirine for protein photocrosslinking. *Chemical Science* **2011,** *2* (3), 480-483.
159. Tippmann, E. M.; Liu, W.; Summerer, D.; Mack, A. V.; Schultz, P. G., A Genetically Encoded Diazirine Photocrosslinker in Escherichia coli. *ChemBioChem* **2007,** *8* (18), 2210-2214.
160. 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. *Journal of the American Chemical Society* **2002,** *124* (31), 9026-9027.
161. Murale, D. P.; Hong, S. C.; Haque, M. M.; Lee, J.-S., Photo-affinity labeling (PAL) in chemical proteomics: a handy tool to investigate protein-protein interactions (PPIs). *Proteome Science* **2017,** *15* (1), 14.
162. Pham, N. D.; Parker, R. B.; Kohler, J. J., Photocrosslinking approaches to interactome mapping. *Current opinion in chemical biology* **2013,** *17* (1), 90-101.

163. Nguyen, T.-A.; Cigler, M.; Lang, K., Expanding the Genetic Code to Study Protein–Protein Interactions. *Angewandte Chemie International Edition* **2018,** *57* (44), 14350-14361.
164. Preston, G. W.; Wilson, A. J., Photo-induced covalent cross-linking for the analysis of biomolecular interactions. *Chemical Society Reviews* **2013,** *42* (8), 3289-3301.
165. Tanaka, Y.; Bond, M. R.; Kohler, J. J., Photocrosslinkers illuminate interactions in living cells. *Molecular BioSystems* **2008,** *4* (6), 473-480.
166. Hino, N.; Oyama, M.; Sato, A.; Mukai, T.; Iraha, F.; Hayashi, A.; Kozuka-Hata, H.; Yamamoto, T.; Yokoyama, S.; Sakamoto, K., Genetic Incorporation of a Photo-Crosslinkable Amino Acid Reveals Novel Protein Complexes with GRB2 in Mammalian Cells. *Journal of Molecular Biology* **2011,** *406* (2), 343-353.
167. Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokoyama, S., Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat Methods* **2005,** *2* (3), 201-6.
168. Joiner, C. M.; Breen, M. E.; Clayton, J.; Mapp, A. K., A Bifunctional Amino Acid Enables Both Covalent Chemical Capture and Isolation of in Vivo Protein-Protein Interactions. *Chembiochem* **2017,** *18* (2), 181-184.
169. Hoffmann, J.-E.; Dziuba, D.; Stein, F.; Schultz, C., A Bifunctional Noncanonical Amino Acid: Synthesis, Expression, and Residue-Specific Proteome-wide Incorporation. *Biochemistry* **2018,** *57* (31), 4747-4752.
170. Long, M. J. C.; Poganik, J. R.; Aye, Y., On-Demand Targeting: Investigating Biology with Proximity-Directed Chemistry. *Journal of the American Chemical Society* **2016,** *138* (11), 3610-3622.
171. Chmura, A. J.; Orton, M. S.; Meares, C. F., Antibodies with infinite affinity. *Proceedings of the National Academy of Sciences* **2001,** *98* (15), 8480.
172. Cigler, M.; Müller, T. G.; Horn-Ghetko, D.; von Wrisberg, M.-K.; Fottner, M.; Goody, R. S.; Itzen, A.; Müller, M. P.; Lang, K., Proximity-Triggered Covalent Stabilization of Low-Affinity Protein Complexes In Vitro and In Vivo. *Angewandte Chemie International Edition* **2017,** *56* (49), 15737-15741.
173. Kobayashi, T.; Hoppmann, C.; Yang, B.; Wang, L., Using Protein-Confined Proximity To Determine Chemical Reactivity. *Journal of the American Chemical Society* **2016,** *138* (45), 14832-14835.
174. Furman, J. L.; Kang, M.; Choi, S.; Cao, Y.; Wold, E. D.; Sun, S. B.; Smider, V. V.; Schultz, P. G.; Kim, C. H., A Genetically Encoded aza-Michael Acceptor for Covalent Cross-Linking of Protein–Receptor Complexes. *Journal of the American Chemical Society* **2014,** *136* (23), 8411-8417.
175. Liu, J.; Li, S.; Aslam, N. A.; Zheng, F.; Yang, B.; Cheng, R.; Wang, N.; Rozovsky, S.; Wang, P. G.; Wang, Q.; Wang, L., Genetically Encoding Photocaged Quinone Methide to Multitarget Protein Residues Covalently in Vivo. *Journal of the American Chemical Society* **2019,** *141* (24), 9458-9462.
176. Wang, N.; Yang, B.; Fu, C.; Zhu, H.; Zheng, F.; Kobayashi, T.; Liu, J.; Li, S.; Ma, C.; Wang, P. G.; Wang, Q.; Wang, L., Genetically Encoding Fluorosulfate-l-tyrosine To React with Lysine, Histidine, and Tyrosine via SuFEx in Proteins in Vivo. *J Am Chem Soc* **2018,** *140* (15), 4995-4999.
177. Xuan, W.; Shao, S.; Schultz, P. G., Protein Crosslinking by Genetically Encoded Noncanonical Amino Acids with Reactive Aryl Carbamate Side Chains. *Angewandte Chemie (International ed. in English)* **2017,** *56* (18), 5096-5100.
178. Xuan, W.; Li, J.; Luo, X.; Schultz, P. G., Genetic Incorporation of a Reactive Isothiocyanate Group into Proteins. *Angew Chem Int Ed Engl* **2016,** *55* (34), 10065-8.
179. Huguenin-Dezot, N.; Alonzo, D. A.; Heberlig, G. W.; Mahesh, M.; Nguyen, D. P.; Dornan, M. H.; Boddy, C. N.; Schmeing, T. M.; Chin, J. W., Trapping biosynthetic acyl-

enzyme intermediates with encoded 2,3-diaminopropionic acid. *Nature* **2019,** *565* (7737), 112-117.
180. Courtney, T.; Deiters, A., Recent advances in the optical control of protein function through genetic code expansion. *Current Opinion in Chemical Biology* **2018,** *46*, 99-107.
181. 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 : a European journal of chemical biology* **2018,** *19* (20), 2178-2185.
182. 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. *Angewandte Chemie (International ed. in English)* **2014,** *53* (15), 3932-3936.
183. Deiters, A.; Groff, D.; Ryu, Y.; Xie, J.; Schultz, P. G., A Genetically Encoded Photocaged Tyrosine. *Angewandte Chemie International Edition* **2006,** *45* (17), 2728-2731.
184. Gautier, A.; Nguyen, D. P.; Lusic, H.; An, W.; Deiters, A.; Chin, J. W., Genetically encoded photocontrol of protein localization in mammalian cells. *J Am Chem Soc* **2010,** *132* (12), 4086-8.
185. 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. *Journal of the American Chemical Society* **2014,** *136* (6), 2240-2243.
186. Yamaguchi, A.; Iraha, F.; Ohtake, K.; Sakamoto, K., Pyrrolysyl-tRNA Synthetase with a Unique Architecture Enhances the Availability of Lysine Derivatives in Synthetic Genetic Codes. *Molecules* **2018,** *23* (10).
187. Beránek, V.; Willis, J. C. W.; Chin, J. W., An Evolved Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetase/tRNA Pair Is Highly Active and Orthogonal in Mammalian Cells. *Biochemistry* **2019,** *58* (5), 387-390.
188. 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*, 374146.
189. Seki, E.; Yanagisawa, T.; Kuratani, M.; Sakamoto, K.; Yokoyama, S., Fully Productive Cell-Free Genetic Code Expansion by Structure-Based Engineering of Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetase. *ACS Synth Biol* **2020**.
190. 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), 32738-46.
191. Madeira, F.; Park, Y. M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A. R. N.; Potter, S. C.; Finn, R. D.; Lopez, R., The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research* **2019,** *47* (W1), W636-W641.
192. Borrel, G.; Parisot, N.; Harris, H. M. B.; Peyretaillade, E.; Gaci, N.; Tottey, W.; Bardot, O.; Raymann, K.; Gribaldo, S.; Peyret, P.; O'Toole, P. W.; Brugère, J.-F., Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. *BMC Genomics* **2014,** *15* (1), 679.
193. Maeder, D. L.; Anderson, I.; Brettin, T. S.; Bruce, D. C.; Gilna, P.; Han, C. S.; Lapidus, A.; Metcalf, W. W.; Saunders, E.; Tapia, R.; Sowers, K. R., The Methanosarcina barkeri Genome: Comparative Analysis with Methanosarcina acetivorans and Methanosarcina mazei Reveals Extensive Rearrangement within Methanosarcinal Genomes. *Journal of Bacteriology* **2006,** *188* (22), 7922.
194. Assis das Graças, D.; Thiago Jucá Ramos, R.; Vieira Araújo, A. C.; Zahlouth, R.; Ribeiro Carneiro, A.; Souza Lopes, T.; Azevedo Baraúna, R.; Azevedo, V.; Cruz Schneider, M. P.; Pellizari, V. H.; Silva, A., Complete Genome of a Methanosarcina mazei Strain Isolated

from Sediment Samples from an Amazonian Flooded Area. *Genome Announc* **2013,** *1* (3), e00271-13.
195. Sherman, F.; Stewart, J. W.; Tsunasawa, S., Methionine or not methionine at the beginning of a protein. *Bioessays* **1985,** *3* (1), 27-31.
196. Murnauer, A. Design einer photoinduzierbaren Diels-Alder Cycloaddition. Master Thesis, Technische Universität München, Garching, 2018.
197. Mayer, S. V. Expanding the chemical biology toolbox: Site-specific incorporation of unnatural amino acids and bioorthogonal protein labeling to study structure and function of proteins. Dissertation, Technische Universität München, Garching, 2019.
198. Lecker, S. H.; Goldberg, A. L.; Mitch, W. E., Protein Degradation by the Ubiquitin–Proteasome Pathway in Normal and Disease States. *Journal of the American Society of Nephrology* **2006,** *17* (7), 1807.
199. Crespo-Yàñez, X.; Aguilar-Gurrieri, C.; Jacomin, A. C.; Journet, A.; Mortier, M.; Taillebourg, E.; Soleilhac, E.; Weissenhorn, W.; Fauvarque, M. O., CHMP1B is a target of USP8/UBPY regulated by ubiquitin during endocytosis. *PLoS Genet* **2018,** *14* (6), e1007456.
200. Teixeira, L. K.; Reed, S. I., Ubiquitin Ligases and Cell Cycle Control. *Annual Review of Biochemistry* **2013,** *82* (1), 387-414.
201. Komander, D.; Rape, M., The Ubiquitin Code. *Annual Review of Biochemistry* **2012,** *81* (1), 203-229.
202. Hershko, A.; Ciechanover, A., THE UBIQUITIN SYSTEM. *Annual Review of Biochemistry* **1998,** *67* (1), 425-479.
203. Kulathu, Y.; Komander, D., Atypical ubiquitylation — the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. *Nature Reviews Molecular Cell Biology* **2012,** *13* (8), 508-523.
204. Komander, D.; Clague, M. J.; Urbé, S., Breaking the chains: structure and function of the deubiquitinases. *Nature Reviews Molecular Cell Biology* **2009,** *10* (8), 550-563.
205. Jin, J.; Li, X.; Gygi, S. P.; Harper, J. W., Dual E1 activation systems for ubiquitin differentially regulate E2 enzyme charging. *Nature* **2007,** *447* (7148), 1135-8.
206. Pelzer, C.; Kassner, I.; Matentzoglu, K.; Singh, R. K.; Wollscheid, H. P.; Scheffner, M.; Schmidtke, G.; Groettrup, M., UBE1L2, a novel E1 enzyme specific for ubiquitin. *J Biol Chem* **2007,** *282* (32), 23010-4.
207. Handley, P. M.; Mueckler, M.; Siegel, N. R.; Ciechanover, A.; Schwartz, A. L., Molecular cloning, sequence, and tissue distribution of the human ubiquitin-activating enzyme E1. *Proc Natl Acad Sci U S A* **1991,** *88* (1), 258-62.
208. Michelle, C.; Vourc'h, P.; Mignon, L.; Andres, C. R., What was the set of ubiquitin and ubiquitin-like conjugating enzymes in the eukaryote common ancestor? *J Mol Evol* **2009,** *68* (6), 616-28.
209. Li, W.; Bengtson, M. H.; Ulbrich, A.; Matsuda, A.; Reddy, V. A.; Orth, A.; Chanda, S. K.; Batalov, S.; Joazeiro, C. A., Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling. *PLoS One* **2008,** *3* (1), e1487.
210. Tokgöz, Z.; Bohnsack, R. N.; Haas, A. L., Pleiotropic Effects of ATP·Mg2+ Binding in the Catalytic Cycle of Ubiquitin-activating Enzyme. *Journal of Biological Chemistry* **2006,** *281* (21), 14729-14737.
211. Schulman, B. A.; Wade Harper, J., Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nature Reviews Molecular Cell Biology* **2009,** *10* (5), 319-331.
212. Lee, I.; Schindelin, H., Structural Insights into E1-Catalyzed Ubiquitin Activation and Transfer to Conjugating Enzymes. *Cell* **2008,** *134* (2), 268-278.
213. Metzger, M. B.; Hristova, V. A.; Weissman, A. M., HECT and RING finger families of E3 ubiquitin ligases at a glance. *Journal of Cell Science* **2012,** *125* (3), 531.

214. Luo, J.; Liu, Q.; Morihiro, K.; Deiters, A., Small-molecule control of protein function through Staudinger reduction. *Nature Chemistry* **2016,** *8* (11), 1027-1034.
215. Kang, J.-Y.; Kawaguchi, D.; Wang, L., Genetically Encoding Unnatural Amino Acids in Neurons In Vitro and in the Embryonic Mouse Brain for Optical Control of Neuronal Proteins. *Methods in molecular biology (Clifton, N.J.)* **2018,** *1728*, 263-277.
216. Erickson, S. B.; Mukherjee, R.; Kelemen, R. E.; Wrobel, C. J. J.; Cao, X.; Chatterjee, A., Precise Photoremovable Perturbation of a Virus–Host Interaction. *Angewandte Chemie International Edition* **2017,** *56* (15), 4234-4237.
217. Engelke, H.; Chou, C.; Uprety, R.; Jess, P.; Deiters, A., Control of protein function through optochemical translocation. *ACS Synth Biol* **2014,** *3* (10), 731-6.
218. Bamford, C. H.; Norrish, R. G. W., 359. Primary photochemical reactions. Part VII. Photochemical decomposition of isovaleraldehyde and di-n-propyl ketone. *Journal of the Chemical Society (Resumed)* **1935,** (0), 1504-1511.
219. Piovesan, D.; Minervini, G.; Tosatto, S. C., The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res* **2016,** *44* (W1), W367-74.
220. Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S., Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem Biol* **2008,** *15* (11), 1187-97.
221. Chou, P. Y.; Fasman, G. D., Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol* **1978,** *47*, 45-148.
222. MacArthur, M. W.; Thornton, J. M., Influence of proline residues on protein conformation. *J Mol Biol* **1991,** *218* (2), 397-412.
223. Richardson, J. S.; Richardson, D. C., Principles and Patterns of Protein Conformation. In *Prediction of Protein Structure and the Principles of Protein Conformation*, Fasman, G. D., Ed. Springer US: Boston, MA, 1989; pp 1-98.
224. Weiss, M. S.; Jabs, A.; Hilgenfeld, R., Peptide bonds revisited. *Nature Structural Biology* **1998,** *5* (8), 676-676.
225. Plutner, H.; Cox, A. D.; Pind, S.; Khosravi-Far, R.; Bourne, J. R.; Schwaninger, R.; Der, C. J.; Balch, W. E., Rab1b regulates vesicular transport between the endoplasmic reticulum and successive Golgi compartments. *J Cell Biol* **1991,** *115* (1), 31-43.
226. Cherfils, J.; Zeghouf, M., Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol Rev* **2013,** *93* (1), 269-309.
227. Vetter, I. R.; Wittinghofer, A., The guanine nucleotide-binding switch in three dimensions. *Science* **2001,** *294* (5545), 1299-304.
228. Nuoffer, C.; Balch, W. E., GTPases: multifunctional molecular switches regulating vesicular traffic. *Annu Rev Biochem* **1994,** *63*, 949-90.
229. Isberg, R. R.; O'Connor, T. J.; Heidtman, M., The Legionella pneumophila replication vacuole: making a cosy niche inside host cells. *Nat Rev Microbiol* **2009,** *7* (1), 13-24.
230. Derré, I.; Isberg, R. R., Legionella pneumophila Replication Vacuole Formation Involves Rapid Recruitment of Proteins of the Early Secretory System. *Infection and Immunity* **2004,** *72* (5), 3048.
231. Murata, T.; Delprato, A.; Ingmundson, A.; Toomre, D. K.; Lambright, D. G.; Roy, C. R., The Legionella pneumophila effector protein DrrA is a Rab1 guanine nucleotide-exchange factor. *Nat Cell Biol* **2006,** *8* (9), 971-7.
232. Müller, M. P.; Peters, H.; Blümer, J.; Blankenfeldt, W.; Goody, R. S.; Itzen, A., The Legionella Effector Protein DrrA AMPylates the Membrane Traffic Regulator Rab1b. *Science* **2010,** *329* (5994), 946.
233. Brombacher, E.; Urwyler, S.; Ragaz, C.; Weber, S. S.; Kami, K.; Overduin, M.; Hilbi, H., Rab1 guanine nucleotide exchange factor SidM is a major phosphatidylinositol 4-

phosphate-binding effector protein of Legionella pneumophila. *J Biol Chem* **2009,** *284* (8), 4846-56.
234. Chen, X.-H.; Xiang, Z.; Hu, Y. S.; Lacey, V. K.; Cang, H.; Wang, L., Genetically encoding an electrophilic amino acid for protein stapling and covalent binding to native receptors. *ACS chemical biology* **2014,** *9* (9), 1956-1961.
235. Mihai Gazdag, E.; Streller, A.; Haneburger, I.; Hilbi, H.; Vetter, I. R.; Goody, R. S.; Itzen, A., Mechanism of Rab1b deactivation by the Legionella pneumophila GAP LepB. *EMBO Rep* **2013,** *14* (2), 199-205.
236. Schoebel, S.; Oesterlin, L. K.; Blankenfeldt, W.; Goody, R. S.; Itzen, A., RabGDI Displacement by DrrA from Legionella Is a Consequence of Its Guanine Nucleotide Exchange Activity. *Molecular Cell* **2009,** *36* (6), 1060-1072.
237. 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), 361-74.
238. AbouElfetouh, A.; Kuhn, M. L.; Hu, L. I.; Scholle, M. D.; Sorensen, D. J.; Sahu, A. K.; Becher, D.; Antelmann, H.; Mrksich, M.; Anderson, W. F.; Gibson, B. W.; Schilling, B.; Wolfe, A. J., The E. coli sirtuin CobB shows no preference for enzymatic and nonenzymatic lysine acetylation substrate sites. *Microbiologyopen* **2015,** *4* (1), 66-83.
239. Yu, D.; Ellis, H. M.; Lee, E. C.; Jenkins, N. A.; Copeland, N. G.; Court, D. L., An efficient recombination system for chromosome engineering in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **2000,** *97* (11), 5978-5983.
240. Datsenko, K. A.; Wanner, B. L., One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences* **2000,** *97* (12), 6640.
241. Wang, J.; Sarov, M.; Rientjes, J.; Hu, J.; Hollak, H.; Kranz, H.; Xie, Y.; Stewart, A. F.; Zhang, Y., An improved recombineering approach by adding RecA to λ Red recombination. *Molecular Biotechnology* **2006,** *32* (1), 43-53.
242. Heermann, R.; Zeppenfeld, T.; Jung, K., Simple generation of site-directed point mutations in the Escherichia coli chromosome using Red®/ET® Recombination. *Microbial Cell Factories* **2008,** *7* (1), 14.
243. Bayley, H., *Photogenerated Reagents in Biochemistry and Molecular Biology*. Elsevier: 2000; Vol. 12, p 8-24 & 112-137.
244. Mackinnon, A. L.; Taunton, J., Target Identification by Diazirine Photo-Cross-linking and Click Chemistry. *Curr Protoc Chem Biol* **2009,** *1*, 55-73.
245. Ai, H.-w.; 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.
246. Cigler, M. Genetically encoding unnatural amino acids: Novel tools for protein labelling and chemical stabilization of low-affinity protein complexes. Dissertation, Technische Universität München, Garching 2019.
247. Maleknia, S. D.; Downard, K. M., Radical approaches to probe protein structure, folding, and interactions by mass spectrometry. *Mass Spectrometry Reviews* **2001,** *20* (6), 388-401.
248. Kiselar, J. G.; Chance, M. R., Future directions of structural mass spectrometry using hydroxyl radical footprinting. *Journal of Mass Spectrometry* **2010,** *45* (12), 1373-1382.
249. Konermann, L.; Pan, J.; Liu, Y.-H., Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society Reviews* **2011,** *40* (3), 1224-1234.
250. Wang, L.; Chance, M. R., Protein Footprinting Comes of Age: Mass Spectrometry for Biophysical Structure Assessment. *Molecular & cellular proteomics : MCP* **2017,** *16* (5), 706-716.
251. Sinz, A., Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom Rev* **2006,** *25* (4), 663-82.

252. Rappsilber, J., The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J Struct Biol* **2011,** *173* (3), 530-40.
253. Leitner, A.; Walzthoeni, T.; Kahraman, A.; Herzog, F.; Rinner, O.; Beck, M.; Aebersold, R., Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics* **2010,** *9* (8), 1634-49.
254. Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G., High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proceedings of the National Academy of Sciences* **2000,** *97* (11), 5802.
255. Sinz, A.; Arlt, C.; Chorev, D.; Sharon, M., Chemical cross-linking and native mass spectrometry: A fruitful combination for structural biology. *Protein Sci* **2015,** *24* (8), 1193-1209.
256. Leitner, A.; Walzthoeni, T.; Aebersold, R., Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat Protoc* **2014,** *9* (1), 120-37.
257. Sinz, A., The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks. *Expert Rev Proteomics* **2014,** *11* (6), 733-43.
258. Yu, C.; Huang, L., Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Analytical Chemistry* **2018,** *90* (1), 144-165.
259. Kao, A.; Chiu, C. L.; Vellucci, D.; Yang, Y.; Patel, V. R.; Guan, S.; Randall, A.; Baldi, P.; Rychnovsky, S. D.; Huang, L., Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol Cell Proteomics* **2011,** *10* (1), M110.002212.
260. Müller, M. Q.; Dreiocker, F.; Ihling, C. H.; Schäfer, M.; Sinz, A., Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS. *Anal Chem* **2010,** *82* (16), 6958-68.
261. Iacobucci, C.; Götze, M.; Ihling, C. H.; Piotrowski, C.; Arlt, C.; Schäfer, M.; Hage, C.; Schmidt, R.; Sinz, A., A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions. *Nat Protoc* **2018,** *13* (12), 2864-2889.
262. Kaake, R. M.; Wang, X.; Burke, A.; Yu, C.; Kandur, W.; Yang, Y.; Novtisky, E. J.; Second, T.; Duan, J.; Kao, A.; Guan, S.; Vellucci, D.; Rychnovsky, S. D.; Huang, L., A New In Vivo Cross-linking Mass Spectrometry Platform to Define Protein-Protein Interactions in Living Cells. *Molecular & Cellular Proteomics* **2014**, mcp.M114.042630.
263. Mitchell Wells, J.; McLuckey, S. A., Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*, Academic Press: 2005; Vol. 402, pp 148-185.
264. Biemann, K., [25] Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. In *Methods in Enzymology*, Academic Press: 1990; Vol. 193, pp 455-479.
265. Müller, M. Q.; Zeiser, J. J.; Dreiocker, F.; Pich, A.; Schäfer, M.; Sinz, A., A universal matrix-assisted laser desorption/ionization cleavable cross-linker for protein structure analysis. *Rapid Communications in Mass Spectrometry* **2011,** *25* (1), 155-161.
266. Yang, Y.; Song, H.; He, D.; Zhang, S.; Dai, S.; Lin, S.; Meng, R.; Wang, C.; Chen, P. R., Genetically encoded protein photocrosslinker with a transferable mass spectrometry-identifiable label. *Nature Communications* **2016,** *7* (1), 12299.
267. Hermanson, G. T., Chapter 1 - Introduction to Bioconjugation. In *Bioconjugate Techniques (Third Edition)*, Hermanson, G. T., Ed. Academic Press: Boston, 2013; pp 1-125.
268. Shang, X.; Chen, Y.; Wang, N.; Niu, W.; Guo, J., Oxidation-induced generation of a mild electrophile for proximity-enhanced protein–protein crosslinking. *Chemical Communications* **2018,** *54* (33), 4172-4175.

269. Kusano, S.; Haruyama, T.; Ishiyama, S.; Hagihara, S.; Nagatsugi, F., Development of the crosslinking reactions to RNA triggered by oxidation. *Chemical Communications* **2014,** *50* (30), 3951-3954.
270. Weissbach, H.; Resnick, L.; Brot, N., Methionine sulfoxide reductases: history and cellular role in protecting against oxidative damage. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2005,** *1703* (2), 203-212.
271. Lin, Z.; Johnson, L. C.; Weissbach, H.; Brot, N.; Lively, M. O.; Lowther, W. T., Free methionine-(R)-sulfoxide reductase from Escherichia coli reveals a new GAF domain function. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** *104* (23), 9597-9602.
272. Ejiri, S.-i.; Weissbach, H.; Brot, N., The purification of methionine sulfoxide reductase from Escherichia coli. *Analytical Biochemistry* **1980,** *102* (2), 393-398.
273. Makukhin, N.; Havelka, V.; Poláchová, E.; Rampírová, P.; Tarallo, V.; Strisovsky, K.; Míšek, J., Resolving oxidative damage to methionine by an unexpected membrane-associated stereoselective reductase discovered using chiral fluorescent probes. *The FEBS Journal* **2019,** *286* (20), 4024-4035.
274. Ezraty, B.; Bos, J.; Barras, F.; Aussel, L., Methionine Sulfoxide Reduction and Assimilation in Escherichia coli: New Role for the Biotin Sulfoxide Reductase BisC. *Journal of Bacteriology* **2005,** *187* (1), 231.
275. Michel, A.; Agerer, F.; Hauck, C. R.; Herrmann, M.; Ullrich, J.; Hacker, J.; Ohlsen, K., Global regulatory impact of ClpP protease of Staphylococcus aureus on regulons involved in virulence, oxidative stress response, autolysis, and DNA repair. *J Bacteriol* **2006,** *188* (16), 5783-96.
276. Baker, T. A.; Sauer, R. T., ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim Biophys Acta* **2012,** *1823* (1), 15-28.
277. Sauer, R. T.; Bolon, D. N.; Burton, B. M.; Burton, R. E.; Flynn, J. M.; Grant, R. A.; Hersch, G. L.; Joshi, S. A.; Kenniston, J. A.; Levchenko, I.; Neher, S. B.; Oakes, E. S.; Siddiqui, S. M.; Wah, D. A.; Baker, T. A., Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* **2004,** *119* (1), 9-18.
278. Gottesman, S.; Roche, E.; Zhou, Y.; Sauer, R. T., The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **1998,** *12* (9), 1338-47.
279. Katayama-Fujimura, Y.; Gottesman, S.; Maurizi, M. R., A multiple-component, ATP-dependent protease from Escherichia coli. *J Biol Chem* **1987,** *262* (10), 4477-85.
280. Wong, P.; Houry, W. A., Chaperone networks in bacteria: analysis of protein homeostasis in minimal cells. *Journal of Structural Biology* **2004,** *146* (1), 79-89.
281. Maurizi, M. R.; Thompson, M. W.; Singh, S. K.; Kim, S. H., Endopeptidase Clp: ATP-dependent Clp protease from Escherichia coli. *Methods Enzymol* **1994,** *244*, 314-31.
282. Gottesman, S., Proteolysis in bacterial regulatory circuits. *Annu Rev Cell Dev Biol* **2003,** *19*, 565-87.
283. Frees, D.; Qazi, S. N. A.; Hill, P. J.; Ingmer, H., Alternative roles of ClpX and ClpP in Staphylococcus aureus stress tolerance and virulence. *Molecular Microbiology* **2003,** *48* (6), 1565-1578.
284. Frees, D.; Sørensen, K.; Ingmer, H., Global virulence regulation in Staphylococcus aureus: pinpointing the roles of ClpP and ClpX in the sar/agr regulatory network. *Infect Immun* **2005,** *73* (12), 8100-8.
285. Böttcher, T.; Sieber, S. A., Structurally refined beta-lactones as potent inhibitors of devastating bacterial virulence factors. *Chembiochem* **2009,** *10* (4), 663-6.
286. Yu, A. Y.; Houry, W. A., ClpP: a distinctive family of cylindrical energy-dependent serine proteases. *FEBS Lett* **2007,** *581* (19), 3749-57.

287. Wang, J.; Hartling, J. A.; Flanagan, J. M., The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. *Cell* **1997,** *91* (4), 447-56.
288. Joshi, S. A.; Hersch, G. L.; Baker, T. A.; Sauer, R. T., Communication between ClpX and ClpP during substrate processing and degradation. *Nat Struct Mol Biol* **2004,** *11* (5), 404-11.
289. Stahlhut, S. G.; Alqarzaee, A. A.; Jensen, C.; Fisker, N. S.; Pereira, A. R.; Pinho, M. G.; Thomas, V. C.; Frees, D., The ClpXP protease is dispensable for degradation of unfolded proteins in Staphylococcus aureus. *Scientific Reports* **2017,** *7* (1), 11739.
290. Kim, D. Y.; Kim, K. K., The structural basis for the activation and peptide recognition of bacterial ClpP. *J Mol Biol* **2008,** *379* (4), 760-71.
291. Kolygo, K.; Ranjan, N.; Kress, W.; Striebel, F.; Hollenstein, K.; Neelsen, K.; Steiner, M.; Summer, H.; Weber-Ban, E., Studying chaperone-proteases using a real-time approach based on FRET. *J Struct Biol* **2009,** *168* (2), 267-77.
292. Gribun, A.; Kimber, M. S.; Ching, R.; Sprangers, R.; Fiebig, K. M.; Houry, W. A., The ClpP double ring tetradecameric protease exhibits plastic ring-ring interactions, and the N termini of its subunits form flexible loops that are essential for ClpXP and ClpAP complex formation. *J Biol Chem* **2005,** *280* (16), 16185-96.
293. Lee, B. G.; Park, E. Y.; Lee, K. E.; Jeon, H.; Sung, K. H.; Paulsen, H.; Rübsamen-Schaeff, H.; Brötz-Oesterhelt, H.; Song, H. K., Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. *Nat Struct Mol Biol* **2010,** *17* (4), 471-8.
294. Geiger, S. R.; Böttcher, T.; Sieber, S. A.; Cramer, P., A Conformational Switch Underlies ClpP Protease Function. *Angewandte Chemie International Edition* **2011,** *50* (25), 5749-5752.
295. Szyk, A.; Maurizi, M. R., Crystal structure at 1.9Å of E. coli ClpP with a peptide covalently bound at the active site. *Journal of Structural Biology* **2006,** *156* (1), 165-174.
296. Gersch, M.; List, A.; Groll, M.; Sieber, S. A., Insights into structural network responsible for oligomerization and activity of bacterial virulence regulator caseinolytic protease P (ClpP) protein. *The Journal of biological chemistry* **2012,** *287* (12), 9484-9494.
297. Zeiler, E.; List, A.; Alte, F.; Gersch, M.; Wachtel, R.; Poreba, M.; Drag, M.; Groll, M.; Sieber, S. A., Structural and functional insights into caseinolytic proteases reveal an unprecedented regulation principle of their catalytic triad. *Proc Natl Acad Sci U S A* **2013,** *110* (28), 11302-7.
298. Kimber, M. S.; Yu, A. Y.; Borg, M.; Leung, E.; Chan, H. S.; Houry, W. A., Structural and theoretical studies indicate that the cylindrical protease ClpP samples extended and compact conformations. *Structure* **2010,** *18* (7), 798-808.
299. Vomacka, J.; Korotkov, V. S.; Bauer, B.; Weinandy, F.; Kunzmann, M. H.; Krysiak, J.; Baron, O.; Böttcher, T.; Lorenz-Baath, K.; Sieber, S. A., An Aromatic Hydroxyamide Attenuates Multiresistant Staphylococcus aureus Toxin Expression. *Chemistry – A European Journal* **2016,** *22* (5), 1622-1630.
300. Hackl, M. W.; Lakemeyer, M.; Dahmen, M.; Glaser, M.; Pahl, A.; Lorenz-Baath, K.; Menzel, T.; Sievers, S.; Böttcher, T.; Antes, I.; Waldmann, H.; Sieber, S. A., Phenyl Esters Are Potent Inhibitors of Caseinolytic Protease P and Reveal a Stereogenic Switch for Deoligomerization. *Journal of the American Chemical Society* **2015,** *137* (26), 8475-8483.
301. Hofbauer, B.; Vomacka, J.; Stahl, M.; Korotkov, V. S.; Jennings, M. C.; Wuest, W. M.; Sieber, S. A., Dual Inhibitor of Staphylococcus aureus Virulence and Biofilm Attenuates Expression of Major Toxins and Adhesins. *Biochemistry* **2018,** *57* (11), 1814-1820.
302. Radzicka, A.; Wolfenden, R., Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases. *Journal of the American Chemical Society* **1996,** *118* (26), 6105-6109.

303. Lan, Y.; Langlet-Bertin, B.; Abbate, V.; Vermeer, L. S.; Kong, X.; Sullivan, K. E.; Leborgne, C.; Scherman, D.; Hider, R. C.; Drake, A. F.; Bansal, S. S.; Kichler, A.; Mason, A. J., Incorporation of 2,3-Diaminopropionic Acid into Linear Cationic Amphipathic Peptides Produces pH-Sensitive Vectors. *ChemBioChem* **2010,** *11* (9), 1266-1272.
304. Purich, D. L., *Enzyme Kinetics: Catalysis and Control: A Reference of Theory and Best-Practice Methods*. Elsevier: 2010; Vol. 1, p 404; Table 7.6.
305. Grimsley, G. R.; Scholtz, J. M.; Pace, C. N., A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci* **2009,** *18* (1), 247-251.
306. Blow, D. M.; Birktoft, J. J.; Hartley, B. S., Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature* **1969,** *221* (5178), 337-40.
307. R. B. Silverman, M. W. H., *The Organic Chemistry of Drug Design and Drug Action*. Academic Press: 2014; Vol. 3, p 172; Chapter 4
308. Fetzer, C.; Korotkov, V. S.; Thänert, R.; Lee, K. M.; Neuenschwander, M.; von Kries, J. P.; Medina, E.; Sieber, S. A., A Chemical Disruptor of the ClpX Chaperone Complex Attenuates the Virulence of Multidrug-Resistant Staphylococcus aureus. *Angewandte Chemie International Edition* **2017,** *56* (49), 15746-15750.
309. Gersch, M.; Kolb, R.; Alte, F.; Groll, M.; Sieber, S. A., Disruption of oligomerization and dehydroalanine formation as mechanisms for ClpP protease inhibition. *J Am Chem Soc* **2014,** *136* (4), 1360-6.
310. 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), 546.
311. Gersch, M.; Gut, F.; Korotkov, V. S.; Lehmann, J.; Böttcher, T.; Rusch, M.; Hedberg, C.; Waldmann, H.; Klebe, G.; Sieber, S. A., The Mechanism of Caseinolytic Protease (ClpP) Inhibition. *Angewandte Chemie International Edition* **2013,** *52* (10), 3009-3014.
312. Gersch, M.; Famulla, K.; Dahmen, M.; Göbl, C.; Malik, I.; Richter, K.; Korotkov, V. S.; Sass, P.; Rübsamen-Schaeff, H.; Madl, T.; Brötz-Oesterhelt, H.; Sieber, S. A., AAA+ chaperones and acyldepsipeptides activate the ClpP protease via conformational control. *Nature Communications* **2015,** *6* (1), 6320.
313. Hammill, J. T.; Miyake-Stoner, S.; Hazen, J. L.; Jackson, J. C.; Mehl, R. A., Preparation of site-specifically labeled fluorinated proteins for 19F-NMR structural characterization. *Nat Protoc* **2007,** *2* (10), 2601-7.
314. Chiu, J.; Tillett, D.; Dawes, I. W.; March, P. E., Site-directed, Ligase-Independent Mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb. *J Microbiol Methods* **2008,** *73* (2), 195-8.
315. Zon, G.; Gallo, K. A.; Samson, C. J.; Shao, K. L.; Summers, M. F.; Byrd, R. A., Analytical studies of 'mixed sequence' oligodeoxyribonucleotides synthesized by competitive coupling of either methyl- or beta-cyanoethyl-N,N-diisopropylamino phosphoramidite reagents, including 2'-deoxyinosine. *Nucleic Acids Res* **1985,** *13* (22), 8181-96.
316. Kayushin, A.; Korosteleva, M.; Miroshnikov, A., Large-scale solid-phase preparation of 3'-unprotected trinucleotide phosphotriesters--precursors for synthesis of trinucleotide phosphoramidites. *Nucleosides Nucleotides Nucleic Acids* **2000,** *19* (10-12), 1967-76.
317. Kayushin, A. L.; Korosteleva, M. D.; Miroshnikov, A. I.; Kosch, W.; Zubov, D.; Piel, N., A convenient approach to the synthesis of trinucleotide phosphoramidites--synthons for the generation of oligonucleotide/peptide libraries. *Nucleic Acids Res* **1996,** *24* (19), 3748-55.
318. Mauriala, T.; Auriola, S.; Azhayev, A.; Kayushin, A.; Korosteleva, M.; Miroshnikov, A., HPLC electrospray mass spectrometric characterization of trimeric building blocks for oligonucleotide synthesis. *J Pharm Biomed Anal* **2004,** *34* (1), 199-206.
319. Yagodkin, A.; Azhayev, A.; Roivainen, J.; Antopolsky, M.; Kayushin, A.; Korosteleva, M.; Miroshnikov, A.; Randolph, J.; Mackie, H., Improved synthesis of

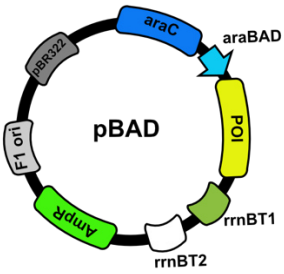
trinucleotide phosphoramidites and generation of randomized oligonucleotide libraries. *Nucleosides Nucleotides Nucleic Acids* **2007,** *26* (5), 473-97.
320. Neylon, C., Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res* **2004,** *32* (4), 1448-59.
321. Sondek, J.; Shortle, D., A general strategy for random insertion and substitution mutagenesis: substoichiometric coupling of trinucleotide phosphoramidites. *Proc Natl Acad Sci U S A* **1992,** *89* (8), 3581-5.
322. Cadwell, R. C.; Joyce, G. F., Randomization of genes by PCR mutagenesis. *PCR Methods Appl* **1992,** *2* (1), 28-33.
323. Wilson, D. S.; Keefe, A. D., Random mutagenesis by PCR. *Curr Protoc Mol Biol* **2001,** *Chapter 8*, Unit8.3.
324. Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M. Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P., A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* **2012,** *30* (10), 918-920.
325. Hoopmann, M. R.; Zelter, A.; Johnson, R. S.; Riffle, M.; MacCoss, M. J.; Davis, T. N.; Moritz, R. L., Kojak: efficient analysis of chemically cross-linked protein complexes. *Journal of Proteome Research* **2015,** *14* (5), 2190-2198.
326. Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **2008,** *26* (12), 1367-1372.
327. Hoopmann, M. R.; Mendoza, L.; Deutsch, E. W.; Shteynberg, D.; Moritz, R. L., An Open Data Format for Visualization and Analysis of Cross-Linked Mass Spectrometry Results. *Journal of the American Society for Mass Spectrometry* **2016,** *27* (11), 1728-1734.
328. Perez-Riverol, Y.; Csordas, A.; Bai, J. W.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Perez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaino, J. A., The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Research* **2019,** *47* (D1), D442-D450.
329. Kabsch, W., XDS. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 125-32.
330. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J Appl Crystallogr* **2007,** *40* (Pt 4), 658-674.
331. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **2011,** *67* (Pt 4), 235-42.
332. Kovalevskiy, O.; Nicholls, R. A.; Long, F.; Carlon, A.; Murshudov, G. N., Overview of refinement procedures within REFMAC5: utilizing data from different sources. *Acta Crystallogr D Struct Biol* **2018,** *74* (Pt 3), 215-227.
333. Miller, C. A.; Ingmer, H.; Cohen, S. N., Boundaries of the pSC101 minimal replicon are conditional. *Journal of bacteriology* **1995,** *177* (17), 4865-4871.
334. Lusetti, S. L.; Cox, M. M., The Bacterial RecA Protein and the Recombinational DNA Repair of Stalled Replication Forks. *Annual Review of Biochemistry* **2002,** *71* (1), 71-100.

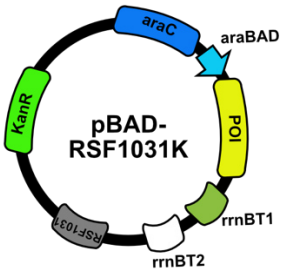
335. Poteete, A. R., What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. *FEMS Microbiol Lett* **2001,** *201* (1), 9-14.
336. Hoopmann, M. R.; Zelter, A.; Johnson, R. S.; Riffle, M.; MacCoss, M. J.; Davis, T. N.; Moritz, R. L., Kojak: efficient analysis of chemically cross-linked protein complexes. *J Proteome Res* **2015,** *14* (5), 2190-8.

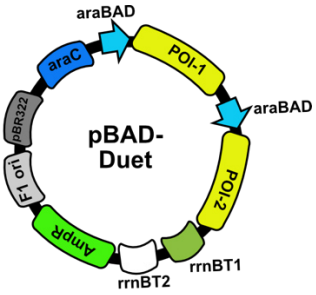
7. APPENDIX

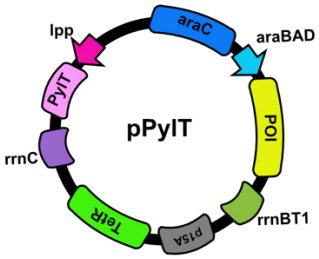
7.1. General

7.1.1. Plasmids maps & backbones & cloning

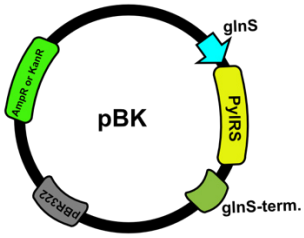




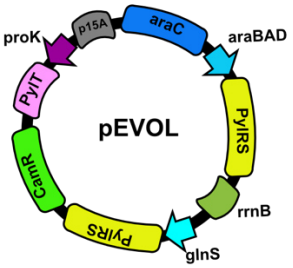




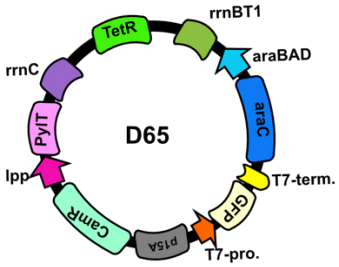
PylT (Ma or Mb)



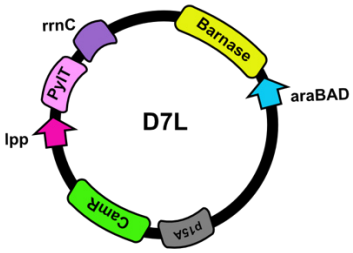
PylRS (Ma or Mb)



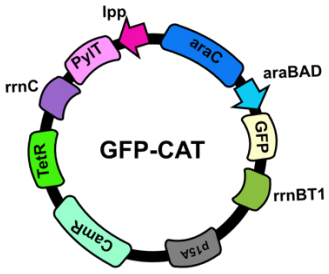
PylRS (Mm or Mb)



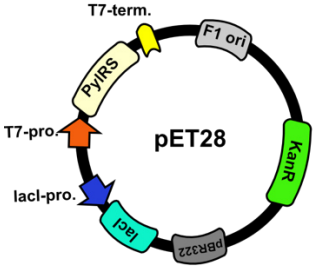
PylT (Ma or Mb)

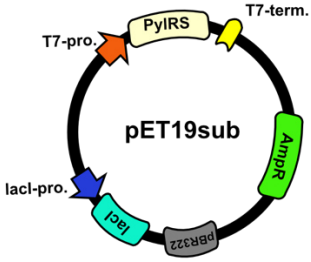


PylT (Ma or Mb)

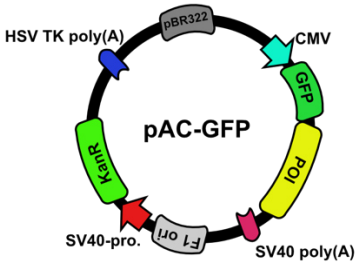


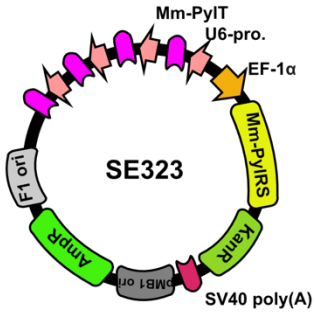
PylT (Ma or Mb)





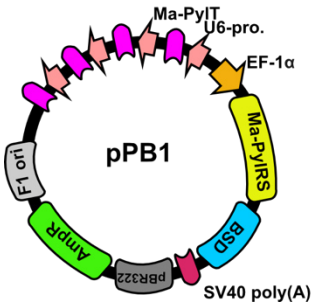
PylRS (Ma or Mb)

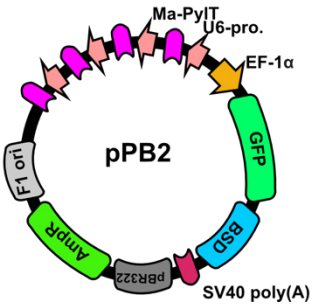



Mm-PylT
U6-pro.
EF-1α
SE323
Mm-PylRS
KanR
pMB1 ori
AmpR
F1 ori
SV40 poly(A)


Mm-PylT
U6-pro.
EF-1α
SE362
GFP
pBR322
AmpR
F1 ori
SV40 poly(A)


Mm-PylT
U6-pro.
EF-1α
SE348
mCherry
GFP
KanR
pUC ori
AmpR
F1 ori
SV40 poly(A)


Ma-PylT
U6-pro.
EF-1α
pPB1
Ma-PylRS
BSD
pBR322
AmpR
F1 ori
SV40 poly(A)


Ma-PylT
U6-pro.
EF-1α
pPB2
GFP
BSD
pBR322
AmpR
F1 ori
SV40 poly(A)

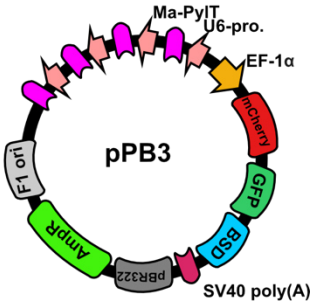

Ma-PylT
U6-pro.
EF-1α
pPB3
mCherry
GFP
BSD
pBR322
AmpR
F1 ori
SV40 poly(A)

Table 64: Cloning of plasmids

Code	Plasmid	Method	Template	Primer 5'→3' & Enzymes
MW-P6	pBAD_SaClpP-S98A-StrepII	Site-directed Mutagenesis	pBAD_SaClpPwt- StrepII	NEB170: TATGGCTGCAGCGATGGGATCAT NEB169: CCGATACAAATTGTTTGAACATC
MW-P8	pPylT_SaClpP-S98TAG-StrepII	Site-directed Mutagenesis	pPylT_SaClpP-S98TAG-His6	NEB202: cagttcgaaaaaTAAAGCTCGAGCGAAGCT NEB203: cgggtggctccaTTTTGTTTCAGGTACCATCAC
MW-P11	pPylT_SaClpPwt-His6	DNA Assembly	B: pPylT_sfGFP-V2P-N150TAG-His6	P186: CACCATCATCATCACCATTAAAGCTCGAGCGAAGC P187: GGTTAATTCCTCCTGTTAGCCCAAAAAACG
			I: pBAD_SaClpPwt- StrepII	P188: GGGCTAACAGGAGGAATTAACCATGAATTTAATTCCTACAGTTATTG P189: GCTTTAATGGTGATGATGATGGTGTTTTGTTTCAGGTACCATC
MW-P12	pPylT_SaClpP-S98A-His6	Site-directed Mutagenesis	pPylT_SaClpPwt-His6	NEB170: TATGGCTGCAGCGATGGGATCAT NEB169: CCGATACAAATTGTTTGAACATC
MW-P14	pPylT_SaClpP-S98TAG-His6	Site-directed Mutagenesis	pPylT_SaClpPwt-His6	NEB158: TATGGCTGCAtagATGGGATCATTC NEB159: CCGATACAAATTGTTTGAAC
MW65	pBAD_sfGFP-V2P-N150TAG-His6	Site-directed Mutagenesis	pBAD_sfGFP-N150TAG-His6	NEB81: TCACAACGTGtgcGATGGTAGCG NEB82: CGGATTTTAAAGTTCGCTTTG
MW72	pPylT_sfGFP-N150TAG-E173C-His6	Site-directed Mutagenesis	pPylT_sfGFP-N150TAG-His6	NEB81: TCACAACGTGtgcGATGGTAGCG NEB82: CGGATTTTAAAGTTCGCTTTG
MW79	pPylT_sfGFP-V2P-N150TAG-His6-MaPylT	Restriction Cloning	B: pPylT_sfGFP-N150TAG-His6 I: Gene String MaPylT	AclI + NsiI-HF
MW80	pPylT_sfGFP-V2P-N150TAG-E173C-His6	Site-directed Mutagenesis	pPylT_sfGFP-N150TAG-His6	NEB81: TCACAACGTGtgcGATGGTAGCG NEB82: CGGATTTTAAAGTTCGCTTTG
MW59	pBAD-RSF1031K_Gro7	DNA Assembly	B: pBAD-RSF1031K	P178: TGCGTTTCTACAAACTC P179: AGACAATTGACGGCTTG
			I: M2021_pGro7 (Itzen Lab TUM)	P180: GTCAAGCCGTCAATTGTCTTCGAAGAATAGTTACGGCTTATG P181: CAAAAGAGTTTGTAGAAACGCACCGTGTCTGAATCTTATACTAG
MW-E1	*Ma*-D65	Restriction Cloning	B: *Mb*-D65	P203: ggaaaGGCGCGCCggtgaagaagctcgacc P204: ggaaaCGTACGGCCCTGCACCATTATGTTC AscI + BsiWI-HF
			I: pPylT_sfGFP-V2P-N150TAG-His6-MaPylT	P201: ggaaaGGCGCGCCCAGATAAAATATTTCTAGATTTCAG P202: ggaaaCGTACGCACAGCTTGTCTGTAAG AscI + BsiWI-HF
MW-E2	*Ma*-D7L	Restriction Cloning	B: *Mb*-D7L	P205: ggaaaGGCGCGCCcgtttaagggcaccaataactg P206: ggaatggtgcatgcTTATC AscI + SphI
			I: pPylT_sfGFP-V2P-N150TAG-His6-MaPylT	P207: ggaaaGCATGCCCACTCCAAGAATTGGAG P208: ggaaaGGCGCGCCCACAGCTTGTCTGTAAGC AscI + SphI
MW-E3			B: *Mb*-PylT_sfGFP150TAG CAT	StuI + NsiI-HF

Code	Plasmid	Method	Template	Primer 5'→3' & Enzymes
	Ma-pPylT_sfGFP150TAG CAT	Restriction Cloning	I: pPylT_sfGFP-V2P-N150TAG-His6-MaPylT	P209: ggaaaAGGCCTCAGATAAAATATTTCTAGATTTCAGTG P210: ggaaaATGCATCACAGCTTGTCTGTAAGC StuI + NsiI-HF
MW-Ma1	pET28_His-Throm-Ma-PylRSwt	Restriction Cloning	B: pET28_His-Throm-Mm-PylRS(Y306G, Y384F, I405R)(185-454) (Schneider Lab; LMU)	NdeI + NotI-HF
			I: pBK_Ma-PylRSwt	P199: ggaaggCATATGACCGTGAAATATACCG P200: gtgcggccgcTTAATTGATTTTTGCACCATTCAG
MW-Ma2	pET28_His-Throm-Ma-PylRS-MF3 (Y126G)	Restriction Cloning	B: pET28_His-Throm-Mm-PylRS(Y306G, Y384F, I405R)(185-454) (Schneider Lab; LMU)	NdeI + NotI-HF
			I: pBK_Ma-PylRS-MF3 (Y126G)	P199: ggaaggCATATGACCGTGAAATATACCG P200: gtgcggccgcTTAATTGATTTTTGCACCATTCAG NdeI + NotI-HF
MW-Ma3	pET28_His-Throm-Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)	Restriction Cloning	B: pET28_His-Throm-Mm-PylRS(Y306G, Y384F, I405R)(185-454) (Schneider Lab; LMU)	NdeI + NotI-HF
			I: pBK_Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)	P199: ggaaggCATATGACCGTGAAATATACCG P200: gtgcggccgcTTAATTGATTTTTGCACCATTCAG NdeI + NotI HF
MW-Ma4	pET28_His-Throm-Ma-PylRS-MF18 (M129A, N166Q, V168S)	Restriction Cloning	B: pET28_His-Throm-Mm-PylRS(Y306G, Y384F, I405R)(185-454) (Schneider Lab; LMU)	NdeI + NotI-HF
			I pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)	P199: ggaaggCATATGACCGTGAAATATACCG P200: gtgcggccgcTTAATTGATTTTTGCACCATTCAG NdeI + NotI HF
MW-Ma5	pET28_His-Throm-Ma-PylRS-MF18 (M129A, N166Q, V168S)-H227I-Y228P	Restriction Cloning	B: pET28_His-Throm-Mm-PylRS(Y306G, Y384F, I405R)(185-454) (Schneider Lab; LMU)	NdeI + NotI-HF
			I: pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)-H227I-Y228P	P199: ggaaggCATATGACCGTGAAATATACCG P200: gtgcggccgcTTAATTGATTTTTGCACCATTCAG NdeI + NotI-HF
MW-Ma6	pBK_Ma-PylRSwt	Site-directed Mutagenesis	pBK_Ma-PylRS-MF3 (Y126G)	NEB192: ACCGAATCTGtatAGCGTTATGCGTGATCTGCG NEB193: GCCAGCATCGGACGCAGT

Code	Plasmid	Method	Template	Primer 5'→3' & Enzymes
MW-Ma9	pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)	Site-directed Mutagenesis	pBK_Ma-PylRSwt	NEB194: GTATAGCGTTgcgCGTGATCTGCG NEB195: AGATTCGGTGCCAGCATC NEB196: gtctGATATGGGTCCGCGTGGT NEB197: agctgCAGCATGGTAAATTCTTCCAGATG
MW-Ma12	pBK_Ma-PylRS-H227I-Y228P	Site-directed Mutagenesis	pBK_Ma-PylRSwt	NEB224: AGTTGGTCCGattccgCTGGATGCAGCAC NEB225: GCTGCGCTACAAACTTCT
MW-Ma13	pBK_Ma-PylRS-MF3 (Y126G)-H227I-Y228P	Site-directed Mutagenesis	pBK_Ma-PylRS-MF3 (Y126G)	NEB224: AGTTGGTCCGattccgCTGGATGCAGCAC NEB225: GCTGCGCTACAAACTTCT
MW-Ma14	pBK_Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)-H227I-Y228P	Site-directed Mutagenesis	pBK_Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)	NEB224: AGTTGGTCCGattccgCTGGATGCAGCAC NEB225: GCTGCGCTACAAACTTCT
MW-Ma15	pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)-H227I-Y228P	Site-directed Mutagenesis	pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)	NEB224: AGTTGGTCCGattccgCTGGATGCAGCAC NEB225: GCTGCGCTACAAACTTCT
MW-Ma16	pBK_Ma-PylRSwt-H227A	Site-directed Mutagenesis	pBK_Ma-PylRSwt	NEB212: AGTTGGTCCGgcgTATCTGGATGCAGCAC NEB213: GCTGCGCTACAAACTTCT
MW-Ma17	pBK_Ma-PylRS-MF3 (Y126G)-H227A	Site-directed Mutagenesis	pBK_Ma-PylRS-MF3 (Y126G)	NEB212: AGTTGGTCCGgcgTATCTGGATGCAGCAC NEB213: GCTGCGCTACAAACTTCT
MW-Ma18	pBK_Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)-H227A	Site-directed Mutagenesis	pBK_Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)	NEB212: AGTTGGTCCGgcgTATCTGGATGCAGCAC NEB213: GCTGCGCTACAAACTTCT
MW-Ma19	pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)-H227A	Site-directed Mutagenesis	pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)	NEB212: AGTTGGTCCGgcgTATCTGGATGCAGCAC NEB213: GCTGCGCTACAAACTTCT
MW-S45	pBK_Mb-PylRS-Y271C-N311Q-Y349F-V366C	Site-directed Mutagenesis	TM3-pBK_Mb-PylRS-Y349 (Lang Lab, TUM)	NEB160: CCCGACCCTGtgcAACTATCTGCGTAAAC NEB161: GCCAGCATCGGACGCAGG NEB162: CACCATGGTTcagTTTTGCCAAATG NEB163: AATTCTTCCAGGTGTTCTTTG NEB166: GAGCAGCGCGtgcGTGGGTCCGG NEB167: AGTTCCAGATCGCCATGC
MW-S8	pBK_Mb-MF18(A274G, N311Q, C313S)-A274G	Site-directed Mutagenesis	pBK_Mb-PylRS-L274A-N311Q-C313S (MF18)	NEB204: GTATAACTATggcCGTAAACTGGATCGTATTCTGCCGGG NEB205: AGGGTCGGGGCCAGCATC
MW-S9	pBK_Mb-MF18(L274A, N311Q, C313S)-M315G	Site-directed Mutagenesis	pBK_Mb-PylRS-L274A-N311Q-C313S (MF18)	NEB206: GTTTTCGCAAggcGGCAGCGGCTG NEB207: TGAACCATGGTGAATTCTTC
MW-S10	pBK_Mb-MF18(L274A, N311Q, C313S)-I378A	Site-directed Mutagenesis	pBK_Mb-PylRS-L274A-N311Q-C313S (MF18)	NEB208: TGAATGGGGCgcgGATAAACCGTGGATTG NEB209: CGATCCAGGCTAACCGGA
MW-S11	pBK_Mb-MF18(L274A, N311Q, C313S)-S313G	Site-directed Mutagenesis	pBK_Mb-PylRS-L274A-N311Q-C313S (MF18)	NEB216: GGTTCAGTTTggcCAAATGGGCAG NEB217: ATGGTGAATTCTTCCAGG
MW-S12	pBK_Mb-MF18(L274A, N311Q, C313S)-I378N	Site-directed Mutagenesis	pBK_Mb-PylRS-L274A-N311Q-C313S (MF18)	NEB218: TGAATGGGGCaacGATAAACCGTGG NEB219: CGATCCAGGCTAACCGGA

Code	Plasmid	Method	Template	Primer 5'→3' & Enzymes
MW-S13	pBK_Mb-MF18(L274A, N311Q, C313S)-I378Q	Site-directed Mutagenesis	pBK_Mb-PylRS-L274A-N311Q-C313S (MF18)	NEB220: TGAATGGGGCcagGATAAACCGTGGATTG NEB219: CGATCCAGGCTAACCGGA
MW-S14	pBK_Mb-MF18(L274A, N311Q, C313S)-M315C	Site-directed Mutagenesis	pBK_Mb-PylRS-L274A-N311Q-C313S (MF18)	NEB221: GTTTTCGCAAtgcGGCAGCGGCTG NEB222: TGAACCATGGTGAATTCTTC
MW-S15	pBK_Mb-MF18(L271A, N311Q, C313S)-A274G-M315G	Site-directed Mutagenesis	pBK_Mb-MF18(A274G, N311Q, C313S)-A274G	NEB206: GTTTTCGCAAggcGGCAGCGGCTG NEB207: TGAACCATGGTGAATTCTTC
MW-S16	pBK_Mb-MF18(L274A, N311Q, C313S)-A274G-I378A	Site-directed Mutagenesis	pBK_Mb-MF18(A274G, N311Q, C313S)-A274G	NEB208: TGAATGGGGCgcgGATAAACCGTGGATTG NEB209: CGATCCAGGCTAACCGGA
MW-S17	pBK_Mb-MF18(L274A, N311Q, C313S)-M315C-I378Q	Site-directed Mutagenesis	pBK_Mb-MF18(L274A, N311Q, C313S)-M315C	NEB220: TGAATGGGGCcagGATAAACCGTGGATTG NEB219: CGATCCAGGCTAACCGGA
MW-S18	pBK_Mb-MF18(L274A, N311Q, C313S)-M315C-I378N	Site-directed Mutagenesis	pBK_Mb-MF18(L274A, N311Q, C313S)-M315C	NEB218: TGAATGGGGCaacGATAAACCGTGG NEB219: CGATCCAGGCTAACCGGA
MW-S19	pBK_Mb-MF18(L274A, N311Q, C313S)-A274G-M315C-I378A	Site-directed Mutagenesis	pBK_Mb-MF18(L274A, N311Q, C313S)-A274G-I378A	NEB221: GTTTTCGCAAtgcGGCAGCGGCTG NEB222: TGAACCATGGTGAATTCTTC
MW-S20	pBK_Mb-MF18(L274A, N311Q, C313S)-A274G-M315G-I378A	Site-directed Mutagenesis	pBK_Mb-MF18(L271A, N311Q, C313S)-A274G-M315G	NEB208: TGAATGGGGCgcgGATAAACCGTGGATTG NEB209: CGATCCAGGCTAACCGGA
MW-S21	pBK_Mb-MF18(L274A, N311Q, C313S)-A274G-M315G-I287A-I378A	Site-directed Mutagenesis	pBK_Mb-MF18(L274A, N311Q, C313S)-A274G-M315G-I378A	NEB210: TCCGATCAAAgcgTTTGAAGTGGGCCCG NEB211: CCCGGCAGAATACGATCC
	pET301_Mm-PylRS	DNA Assembly	B_ pET301	P21: tcagcggtttttgtccataTAAGGGAGAGCGTCGAGAT P22: tctacgaacctgtaactgcaTTAGACGTCAGGTGGCACT
			I: pEVOL_Mm-PylRSwt-PylT (Lemke Lab, LMU)	PstI-HF + NdeI
	pET301_Mm-PylRS-Y306M-L309A-C3489A	Site-directed Mutagenesis	pET301_Mm-PylRS	NEBa: tatgcgCGCAAACTGGACCGTGCC NEBb: gttcatCAGATTTGGTGCCAGCATAGG NEBc: GCTGAACTTTgctCAAATGGGTTCAGGTTG NEBd: ATGGTAAACTCCTCCAGATG
	pEVOL_Mm-PylRS-Y306M-L309A-C348A-PylT (1x Mm-TEMPOH-I; 1x Mm-PylRS)	DNA Assembly	B: pEVOL_Mm-PylRSwt-PylT (Lemke Lab, LMU)	P17: gtcgaccatcatcatcatca P37: agatctaattcctcctgttagc
			I: pET301_Mm-PylRS-Y306M-L309A-C3489A	P35: gctaacaggaggaattagatctatggacaaaaaaccgctgaa P36: tgatgatgatgatggtcgacttacaggttcgtagagatcc
MW-S60	pEVOL_Mm-PylRS-Y306M-L309A-C348A-PylT (Mm-TEMPOH-I) (BrCnKRS)	DNA Assembly	B: pEVOL_Mm-PylRS-Y306M-L309A-C348A-PylT (1x Mm-TEMPOH-I; 1x Mm-PylRS)	P40: ctgcagtttcaaacgctaaa P41: atgggattcctcaaagcgta
			I: pET301_Mm-PylRS-Y306M-L309A-C3489A	P38: tacgctttgaggaatcccatatggacaaaaaaccgctgaa P39: tttagcgtttgaaactgcagttacaggttcgtagagatcc
	pBAD_Rab1b-Q67A	Restriction Cloning	pBad-His6-TEV-EBFP2 (Itzen Lab, UKE)	NdeI + HindIII-HF
MW-P41	pBAD_Rab1b-Q67A-His6	SLIM	pBAD_Rab1b-Q67A	P1: TAACTCGAGGGATCCGGG P2: gtgatgatggtggtgatgACCCATACGTTTTTTGATTT P3: CATCACCACCATCATCACTAACTCGAGGGATCCGGG P4: ACCCATACGTTTTTTGATTT

Code	Plasmid	Method	Template	Primer 5'→3' & Enzymes
MW-P42	pBAD_Rab1b-Q67A-R69TAG-His6	Site-directed Mutagenesis	pBAD_Rab1b-Q67A-His6	NEB1: TGGTGCAGAAtagTTCCGTACCATTACCAGC NEB2: GCGGTATCCCAGATCTGC
MW-P43	pBAD_Rab1b-Q67A-T72TAG-His6	Site-directed Mutagenesis	pBAD_Rab1b-Q67A-His6	NEB3: ACGTTTCCGTtagATTACCAGCTCCTATTATC NEB4: TCTGCACCAGCGGTATCC
MW-P46	pBAD_Rab1b-Q67A-D44TAG-His6	Site-directed Mutagenesis	pBAD_Rab1b-Q67A-His6	NEB99: CATTGGCGTCtagTTCAAAATCC NEB100: GTACTGATATAGGATTCGG
MW-P48	pBAD_Rab1b-Q67A-D31TAG-His6	Site-directed Mutagenesis	pBAD_Rab1b-Q67A-His6	NEB97: TTTTGCTGATtagACGTATACCGAATCCTATATC NEB98: CGCAGCAGCAGACAGCTT
	pBAD-DrrA$_{1-339}$	DNA Assembly	B: pBad-His6-TEV-EBFP2 (Itzen Lab, UKE)	P7: GAATTCGAAGCTTGGCTG P8: ATGTATATCTCCTTCTTAAAGTTAAAC
			I: pOPINE(n)EGFP_DrrA_AMPyD (1-339) (Itzen Lab, UKE)	P5:GAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGAGCATAATGGGGAGAATTA P6:GCCAAAACAGCCAAGCTTCGAATTCTTAACGTTGAACACCCAGCT
	pBAD-StrepII-DrrA$_{1-339}$	Site-directed Mutagenesis	pBAD-DrrA$_{1-339}$	P9:GGGTGCAGATGACGACGACAAGCATATGAGCATAATGGGGAGAATTAAAATG P10:TTTTCGAACTGCGGGTGGCTCCACATATGTATATCTCCTTCTTAAAGTTAAAC
MW-P54	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)	DNA Assembly	B: pBAD-RSF1031K	P13: TCAGAAGTGAAACGCCGTAG P14: TGGTGAATTCCTCCTGCTAG
			I: pBAD-StrepII-DrrA$_{1-339}$	P11: CTAGCAGGAGGAATTCACCAGAAATAATTTTGTTTAACTTTAAGAAG P12: CTACGGCGTTTCACTTCTGATTAACGTTGAACACCCAG
MW-P55	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)-D82C	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)	NEB53: TGATGCCTTTtgcGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC
MW-P56	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)-D82A	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)	NEB55: TGATGCCTTTgcgGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC
MW-P57	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)-D22C	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)	NEB45: TTTATACAGTtgcGAGCGAGACAAAC NEB46: CTACCAAATTGCTCTTCATTAAC
MW-P58	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)-D25C	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)	NEB47: TGATGAGCGAtgcAAACCATTACTATC NEB48: CTGTATAAACTACCAAATTGC
MW-P59	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)-R24C	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)	NEB49: CAGTGATGAGtgcGACAAACCATTAC NEB50: TATAAACTACCAAATTGCTC
MW-P93	pBAD-RSF1031K_StrepII-Ent-DrrA(1-647)-N451A-R453A-D480A-S483A	DNA Assembly	B: pBAD-RSF1031K-StrepII-DrrA1-339	P13: TCAGAAGTGAAACGCCGTAG P17: ATGCTTGTCGTCGTCATC
			I: pET19mod-DrrA(N451A;R453A;D480A;S483A)	P15:GAAAAGGGTGCAGATGACGACGACAAGCATATGAGCATAATGGGGAGAAT P16: GGCGCTACGGCGTTTCACTTCTGATCATTTTATCTTAATGGTTTGTC
MW-P94	pBAD-RSF1031K_StrepII-Ent-DrrA(1-647)-N451A-R453A-D480A-S483A-D82C	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-Ent-DrrA(1-647)-N451A-R453A-D480A-S483A	NEB53: TGATGCCTTTtgcGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC
	pBAD-RSF1031K_DrrA(16-352)	DNA Assembly	B: pBAD-RSF1031K-StrepII-DrrA$_{1-647}$	P20: TCAGAAGTGAAACGCCGTAG P21: TtgTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTGGTGAATTCC
			I: pBAD-RSF1031K-StrepII-DrrA$_{1-647}$	P18:GTTTAACTTTAAGAAGGAGATATACAAATGTTTGGTAGTTTATACAGTG P19:CTACGGCGTTTCACTTCTGATTACAACTTCTTGGCATTTTC

Code	Plasmid	Method	Template	Primer 5'→3' & Enzymes
MW-P60	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)	Site-directed Mutagenesis	pBAD-RSF1031K_DrrA(16-352)	P22:AAAGGAGAATCTTTATTTTCAGGGCTTTGGTAGTTTATACAGTGATGAGCGAG P23:TCGAACTGCGGGTGGCTCCAAGACGCCATTTGTATATCTCCTTCTTAAAGTTAA ACAAAATTATTTCTGGTGAATTC
MW-P62	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-D82C	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)	NEB53: TGATGCCTTTgcGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC
MW-P61	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-Q71C	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)	NEB59: TGGATTACGTtgcATTTACAAAGAAATGATTTATGATG NEB60: TTCTCCCACTCCTGCCAT
MW-P65	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-I123TAG	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)	NEB107: TGACGAAGATtagAAAAAAGTAAAACCGGTCTTTG NEB108: TTTTTAACAATAACAAAGGCATC
MW-P68	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-D131TAG	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)	NEB109: ACCGGTCTTTtagGCATTAAATAATTTATG NEB110: TTTACTTTTTTTATATCTTCGTCATTTTTAAC
MW-P70	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-Q138TAG	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)	NEB111: TAATTTATGTtagCGTATTTTTACTG NEB112: TTTAATGCGTCAAAGACC
MW-P75	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-E179TAG	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)	NEB113: AGCTGATGTTtagGCAACTGCAATG NEB114: ACCATCCCGTCTTTTAAC
MW-P76	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-A180TAG	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)	NEB115: TGATGTTGAAtagACTGCAATGTCAATTTTAACATC NEB116: GCTACCATCCCGTCTTTTAAC
MW-P77	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-S184TAG	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)	NEB117: AACTGCAATGtagATTTTAACATCAAAAC NEB118: GCTTCAACATCAGCTACC
MW-P78	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-K241TAG	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)	NEB119: AGTGAGTGTCtagACTCATATTATGC NEB120: ACTTCTGCACCCTCTTTAG
MW-P79	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)-E316TAG	Site-directed Mutagenesis	pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)	NEB121: AGAACATGATtagATGCCTTATTC NEB122: TTTTTGTGCTCAGCATGC
MW-P88	pBAD-RSF1031K_DrrA(16-533)-TEV-His6 (N451A-R453A-D480A-S483A)	DNA Assembly	B: pBAD-RSF1031K_StrepII-Ent-DrrA(1-647)-N451A-R453A-D480A-S483A	P28:GTTTAACTTTAAGAAGGAGATATAcaATGTTTGGTAGTTTATACAGTGATGAGCG P29:GCCCTGAAAATAAAGATTCTCAGCTTTATAAGAAAGATATTCGGATAAATTTTCACC
			I: pBAD-RSF1031K_StrepII-Ent-DrrA(1-647)-N451A-R453A-D480A-S483A	P30: CATTTGTATATCTCCTTCTTAAAG P31:GCTGAGAATCTTTATTTTCAGGGCGGCCATCACCACCATCATCACTGATCAGAAGTGAAACGC
MW-P89	pBAD-RSF1031K_DrrA(16-533)-D82C-TEV-His6 (N451A-R453A-D480A-S483A)	Site-directed Mutagenesis	pBAD-RSF1031K_DrrA(16-533)-TEV-His6 (N451A-R453A-D480A-S483A)	NEB53: TGATGCCTTTgcGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC
MW-P92	pBAD-RSF1031K_DrrA(16-533)-D82C-D512C-TEV-His6 (N451A-R453A-D480A-S483A)	Site-directed Mutagenesis	pBAD-RSF1031K_DrrA(16-533)-D82C-TEV-His6 (N451A-R453A-D480A-S483A)	NEB15: TAATGTTAATtgcGAAACATTAGAGAGTGTG NEB16: TTGCTTCTCTCAAGTTTTTG
MW-P100	pBAD-Duet_Rab1b-Q67A-His6-StrepII-Ent-DrrA(1-339)	DNA Assembly	B: pBAD-RSF1031K_StrepII-Ent-DrrA(1-339)	P26: GGCCCTCTAGATGCGGCCAAGAAACCAATTGTCCATATTG P27: ACAAAAGAGTTTGTAGAAACGCAA
			I: pBAD_Rab1b-Q67A-His6	P24: TTGCGTTTCTACAAACTCTTTTGT P25: GGCCGCATCTAGAGGGCC

Code	Plasmid	Method	Template	Primer 5'→3' & Enzymes
MW-P101	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-His6-StrepII-Ent-DrrA(1-339)	NEB1: TGGTGCAGAAtagTTCCGTACCATTACCAGC NEB2: GCGGTATCCCAGATCTGC
MW-P102	pBAD-Duet_Rab1b-Q67A-T72TAG-His6-StrepII-Ent-DrrA(1-339)	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-His6-StrepII-Ent-DrrA(1-339)	NEB3: ACGTTTCCGTtagATTACCAGCTCCTATTATC NEB4: TCTGCACCAGCGGTATCC
MW-P104	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)-D82A	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)	NEB55: TGATGCCTTTgcgGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC
MW-P105	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)-D82C	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)	NEB53: TGATGCCTTTgcGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC
MW-P106	pBAD-Duet_Rab1b-Q67A-T72TAG-His6-StrepII-Ent-DrrA(1-339)-D82C	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-T72TAG-His6-StrepII-Ent-DrrA(1-339)	NEB53: TGATGCCTTTgcGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC
MW-P107	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)-E38A	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)	NEB57: TGAGCGGTTGGGGATAGT NEB58: GAAAAAATTTgcgGAATATCAGAACAAATTGGCTAATCTAAG
MW-P108	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)-E38C	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)	NEB56:GAAAAAATTTgcGAATATCAGAACAAATTGGCTAATCTAAGTAAAATAATTC NEB57: TGAGCGGTTGGGGATAGT
MW-P109	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)-D165C	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)	NEB27: TACGCCTGATtgcTTGTTTGGTATGTTAAAAG NEB28: CCAATAAGACGGCTTGGATTG
MW-P110	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)-D172C	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)	NEB29: TATGTTAAAAtgcGGGATGGTAGCTG NEB30: CCAAACAAATCATCAGGC
MW-P111	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)-D165A-D172A	SLIM	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-339)	SLIM13: GCGTTGTTTGGTATGTTAAAAGCGGGGATGGTAGCTGATGTTG SLIM14: ATCAGGCGTACCAATAAGACG SLIM15: CGCTTTTAACATACCAAACAACGCATCAGGCGTACCAATAAG SLIM16: GGGATGGTAGCTGATGTTGAAG
MW-P123	pBAD-Duet_Rab1b-Q67A-His6-DrrA(16-352)	DNA Assembly	B: pBAD-Duet_Rab1b-Q67A-His6-StrepII-Ent-DrrA(1-339)	P132: TCAGAAGTGAAACGCCGTAG P165: TtgTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTGGTGAATTCC
			I: pBAD-RSF1031K_StrepII-Ent-DrrA(1-647)-N451A-R453A-D480A-S483A	P163:GTTTAACTTTAAGAAGGAGATATAcaATGTTTGGTAGTTTATACAGTG P164: CTACGGCGTTTCACTTCTGATTACAACTTCTTGGCATTTTC
MW-P130	pBAD-Duet_Rab1b-Q67A-D31TAG-His6-DrrA(16-352)	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-His6-DrrA(16-352)	NEB97: TTTTGCTGATtagACGTATACCGAATCCTATATC NEB98: CGCAGCAGCAGACAGCTT
MW-P131	pBAD-Duet_Rab1b-Q67A-D31TAG-His6-DrrA(16-352)-D131A	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-D31TAG-His6-DrrA(16-352)	NEB155: ACCGGTCTTTgcgGCATTAAATAATTTATG NEB26:
MW-P132	pBAD-Duet_Rab1b-Q67A-D31TAG-His6-DrrA(16-352)-E58A	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-D31TAG-His6-DrrA(16-352)	NEB156: TGAGGGCAATgcgGTATCTCCATG NEB157: TTTTCTCGAATTATTTTACTTAGATTAG
MW-P133	pBAD-Duet_Rab1b-Q67A-D31TAG-His6-DrrA(16-352)-E280A	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-D31TAG-His6-DrrA(16-352)	NEB153: TTTATTACGTgcgAAAAATTTATTGCCCG NEB154: CGTATAGTTCCTGGTGCC
MW-P124	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-DrrA(16-352)	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-His6-DrrA(16-352)	NEB1: TGGTGCAGAAtagTTCCGTACCATTACCAGC NEB2: GCGGTATCCCAGATCTGC
MW-P125	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-DrrA(16-352)-D82C	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-DrrA(16-352)	NEB53: TGATGCCTTTgcGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC

Code	Plasmid	Method	Template	Primer 5'→3' & Enzymes
MW-P144	pBAD-Duet_Rab1b-Q67A-His6-StrepII-TEV-DrrA(16-352)	DNA Assembly	B: pBAD-RSF1031K_StrepII-TEV-DrrA(16-352)	P26: GGCCCTCTAGATGCGGCCAAGAAACCAATTGTCCATATTG P27: ACAAAAGAGTTTGTAGAAACGCAA
			I: pBAD_Rab1b-Q67A-His6	P24: TTGCGTTTCTACAAACTCTTTTGT P25: GGCCGCATCTAGAGGGCC
MW-P145	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-His6-StrepII-TEV-DrrA(16-352)	NEB1: TGGTGCAGAAtagTTCCGTACCATTACCAGC NEB2: GCGGTATCCCAGATCTGC
MW-P150	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82C	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)	NEB53: TGATGCCTTTtgcGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC
MW-P151	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)	NEB55: TGATGCCTTTgcgGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC
MW-P154	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A-Q71C	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A	NEB59: TGGATTACGTtgcATTTACAAAGAAATGATTTATGATG NEB60: TTCTCCCACTCCTGCCAT
MW-P155	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A-Y78C	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A	NEB236: AGAAATGATTtgcGATGCCTTTGCG NEB237: TTGTAAATTTGACGTAATCC
MW-P156	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A-E75C	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A	NEB232: AATTTACAAAtgcATGATTTATGATGCCTTTGC NEB233: TGACGTAATCCATTCTCC
MW-P157	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A-D79C	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A	NEB230: AATGATTTATtgcGCCTTTGCGGC NEB231: TCTTTGTAAATTTGACGTAATC
MW-P158	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A-E87C	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A	NEB228: GCTTGGTGTCtgcATGCCGAAAGATATG NEB229: GCCGCAAAGGCATCATAA
MW-P159	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A-M88C	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-TEV-DrrA(16-352)-D82A	NEB234: TGGTGTCGAGtgcCCGAAAGATATGGAAGTCCATTTTGC NEB235: AGCGCCGCAAAGGCATCA
MW-P164	pBAD-Duet_Rab1b-Q67A-His6-StrepII-Ent-DrrA(1-647)(N451A-R453A-D480A-S483A)	DNA Assembly	B: pBAD-RSF1031K_StrepII-Ent-DrrA(1-647)-N451A-R453A-D480A-S483A	P26: GGCCCTCTAGATGCGGCCAAGAAACCAATTGTCCATATTG P27: ACAAAAGAGTTTGTAGAAACGCAA
			I: pBAD_Rab1b-Q67A-His6	P24: TTGCGTTTCTACAAACTCTTTTGT P25: GGCCGCATCTAGAGGGCC
MW-P165	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-647)(N451A-R453A-D480A-S483A)	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-His6-StrepII-Ent-DrrA(1-647)(N451A-R453A-D480A-S483A)	NEB1: TGGTGCAGAAtagTTCCGTACCATTACCAGC NEB2: GCGGTATCCCAGATCTGC
MW-P166	pBAD-Duet_Rab1b-Q67A-T72TAG-His6-StrepII-Ent-DrrA(1-647)(N451A-R453A-D480A-S483A)	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-His6-StrepII-Ent-DrrA(1-647)(N451A-R453A-D480A-S483A)	NEB3: ACGTTTCCGTtagATTACCAGCTCCTATTATC NEB4: TCTGCACCAGCGGTATCC

Code	Plasmid	Method	Template	Primer 5'→3' & Enzymes
MW-P167	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-647)-D82C (N451A-R453A-D480A-S483A)	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-StrepII-Ent-DrrA(1-647)(N451A-R453A-D480A-S483A)	NEB53: TGATGCCTTTtgcGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC
	pBAD-Duet_Rab1b-Q67A- His6-DrrA(16-533)-TEV-His6 (N451A-R453A-D480A-S483A)	DNA Assembly	B: pBAD-Duet_Rab1b-Q67A-His6-StrepII-Ent-DrrA(1-647)(N451A-R453A-D480A-S483A)	P150: CATTTGTATATCTCCTTCTTAAAG P173: gctgagaatctttattttcagggcggccatcaccaccatcatcacTGATCAGAAGTGAAACGC
			I: pBAD-RSF1031K_StrepII-Ent-DrrA(1-647)-N451A-R453A-D480A-S483A	P171:gtttaaCTTTAAGAAGGAGATATAcaAATGTTTGGTAGTTTATACAGTGATGAGCG P172: gccctgaaaataaagattctcagcTTTATAAGAAAGATATTCGGATAAATTTTCACC
MW-P160	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-DrrA(16-533)-TEV-His6 (N451A-R453A-D480A-S483A)	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A- His6-DrrA(16-533)-TEV-His6 (N451A-R453A-D480A-S483A)	NEB1: TGGTGCAGAAtagTTCCGTACCATTACCAGC NEB2: GCGGTATCCCAGATCTGC
MW-P161	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-DrrA(16-533)-D82C-TEV-His6 (N451A-R453A-D480A-S483A)	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-DrrA(16-533)-TEV-His6 (N451A-R453A-D480A-S483A)	NEB53: TGATGCCTTTtgcGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC
MW-P162	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-DrrA(16-533)-D82C (N451A-R453A-D480A-S483A)	Site-directed Mutagenesis	pBAD-Duet_Rab1b-Q67A-R69TAG-His6-DrrA(16-533)-D82C-TEV-His6 (N451A-R453A-D480A-S483A)	NEB85: TGATCAGAAGTGAAACGC NEB86: TTTATAAGAAAGATATTCGGATAAATTTTC
MM19	SE348_Rab1b-Q67A-R69TAG-His6_4xMmPylT	Restriction Cloning	B: SE348_mCherry-TAG-GFP-4xMmPylT	NheI-HF + BamHI-HF
			I: pBAD-Rab1b-Q67A-R69TAG-His6	P231: aattatGGATCCCTCGAGTTAgtg P232: CTAGAGCTAGCACCATGCCGGAATATGACTATC NheI-HF + BamHI-HF
MM27	pAcGFP-DrrA(8-533)-D110A-D112A-D82A	Site-directed Mutagenesis	pAcGFP-DrrA(8-533)-D110A-D112A	NEB55: TGATGCCTTTgcgGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC
MM33	pAcGFP-DrrA(8-533)-D110A-D112A-D82C	Site-directed Mutagenesis	pAcGFP-DrrA(8-533)-D110A-D112A	NEB53: TGATGCCTTTtgcGCGCTTGGTG NEB54: TAAATCATTTCTTTGTAAATTTGACGTAATC
MM35	pAcGFP-DrrA(8-533)-R70A-Q71A-K74A-D82A	Site-directed Mutagenesis	pAcGFP-DrrA(8-533)-R70A-Q71A-K74A	NEB55: TGATGCCTTTgcgGCGCTTGGTG NEB278: TAAATCATTTCAGCGTAAATAGCAG
MM36	pAcGFP-DrrA(8-533)-R70A-Q71A-K74A-D82C	Site-directed Mutagenesis	pAcGFP-DrrA(8-533)-R70A-Q71A-K74A	NEB53: TGATGCCTTTtgcGCGCTTGGTG NEB278: TAAATCATTTCAGCGTAAATAGCAG
MM32	pPB3_mCherry-TAG-GFP_4xMaPylT	Restriction Cloning	B: pPB2_sfGFP-His6_4xMaPylT I: SE348_mCherry-TAG-GFP-4xMmPylT	BsiWI-HF + AsiSI
MM28	pET19sub_Ma-PylRSwt	Restriction Cloning	B: pET19sub_Mm-PylRSwt	P221: GGCAAGCTTCTTAAGTTTAAACGC P222: ggaaaCCATGGTGAGGATCCGGCTGCTAACAAAG NheI-HF + NcoI-HF
			I: pPB1_Ma-PylRSwt_4xMaPylT	NheI-HF + NcoI-HF

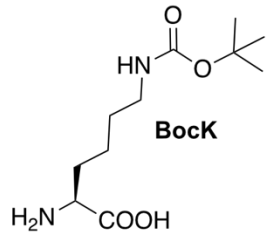
7.1.2. Substrate spectra of aaRSs

Table 65 Overview of the substrate spectra of different PylRS mutants

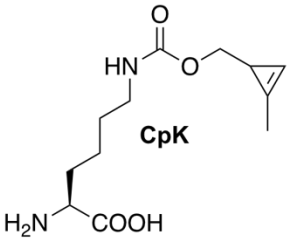
Code	Plasmid	Accepted uAA
MW-S26	pBK_Mb-PylRSwt (D4)	BocK, CpK
MW-S34 Mb-TEMPOH-I	pBK_Mb-PylRS-Y271M-L274A-C313A (Mb-TEMPOH-I) (BrCnKRS)	BrC6K. BrC7K. BrCo6K, BrC07K, BrCN7K, TM-439*, DiaziK
Mb-MF3	pBK_Mb_PylRS-Y271G-C313V (MF3)	BCNK, mTetK, TetK
Mb-MF9	pBK_Mb-PylRS-L266M-L270I-Y271F-L274A-C313F (MF9)	GGK, AcK
Mb-MF18	pBK_Mb-PylRS-L274A-N311Q-C313S (MF18)	AzGGK
MW-S45 pDAPRS	pBK_Mb-PylRS-Y271C-N311Q-Y349F-V366C	pDAP
MW-S59	pEVOL_Mm-PylRSwt-PylT	BocK, CpK
MW-S60 Mm-TEMPOH-I	pEVOL_Mm-PylRS-Y306M-L309A-C348A-PylT (Mm-TEMPOH-I) (BrCnKRS)	BrC6K. BrC7K. BrCo6K, BrC07K, BrCN7K, TM-439, DiaziK
MW-Ma6	pBK_Ma-PylRSwt	BocK, CpK, DiaziK, BrC6K. BrC7K. BrCo6K, BrC07K, BrCN7K*
Ma-MF3	pBK_Ma-PylRS-MF3 (Y126G)	BCNK
Ma-MF9	pBK_Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)	AcK
MW-Ma12	pBK_Ma-PylRS-H227I-Y228P	BocK, CpK, DiaziK, BrC6K, BrC7K, BrCO6K, BrCO7K, BrCN7K*, TM-439*
MW-Ma13	pBK_Ma-PylRS-MF3 (Y126G)-H227I-Y228P	BCNK
MW-Ma14	pBK_Ma-PylRS-MF9 (L121M, L125I, Y126F, M129A, V168F)-H227I-Y228P	AcK
MW-Ma15	pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)-H227I-Y228P	AzGGK
MW-Ma15	pBK_Ma-PylRS-MF18 (M129A, N166Q, V168S)-H227I-Y228P	AzGGK

* The aaRS seems to accept the uAA but no correct mass of the incorporated uAA into the protein was detected

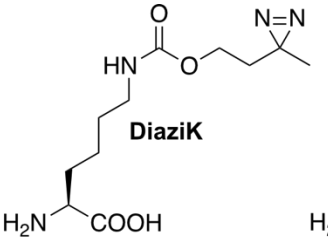
7.1.3. Structures of unnatural amino acids



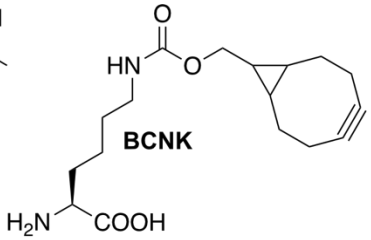
Molecular Weight: 246,31



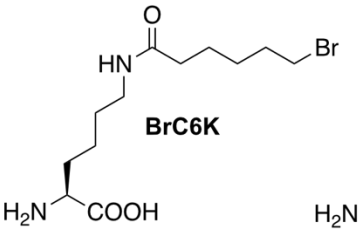
Molecular Weight: 256,30



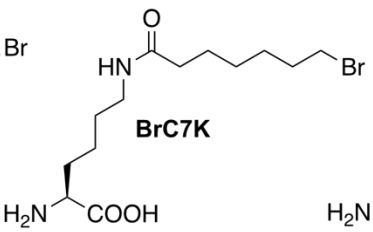
Molecular Weight: 272,31



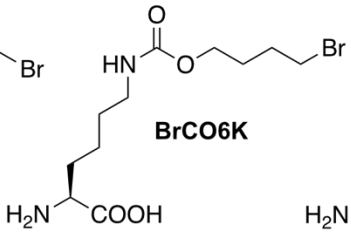
Molecular Weight: 322,41



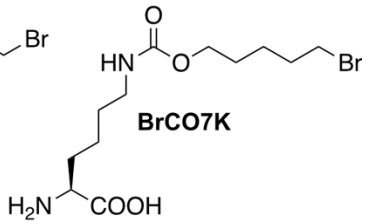
Molecular Weight: 323,23



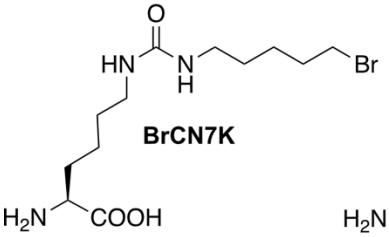
Molecular Weight: 337,26



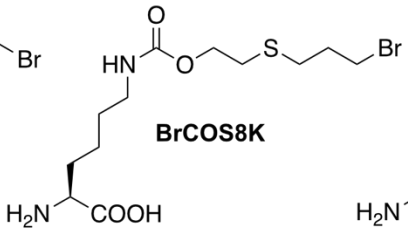
Molecular Weight: 325,20



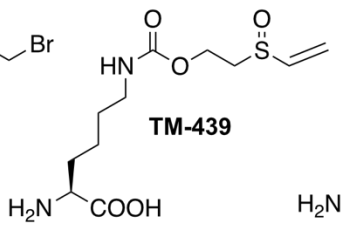
Molecular Weight: 339,23



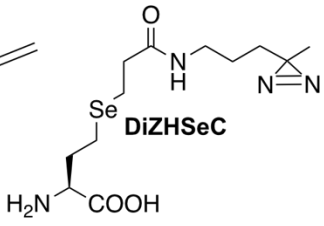
Molecular Weight: 338,25



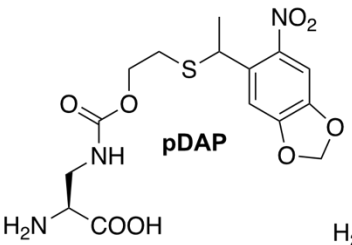
Molecular Weight: 371,29



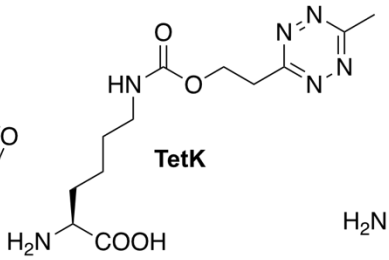
Molecular Weight: 292,35



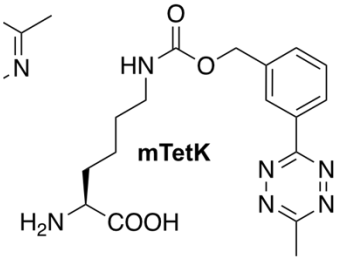
Molecular Weight: 349,30



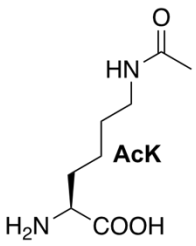
Molecular Weight: 401,39



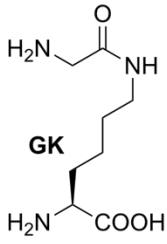
Molecular Weight: 312,33



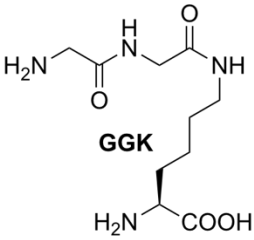
Molecular Weight: 374,40



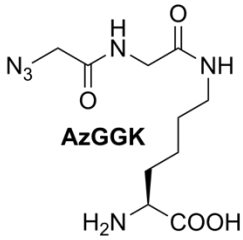
Molecular Weight: 188,23



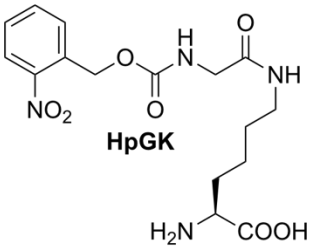
Molecular Weight: 203,24



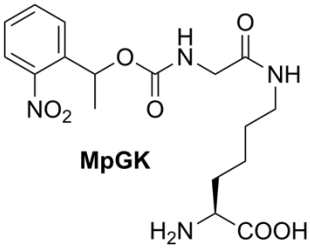
Molecular Weight: 260,29



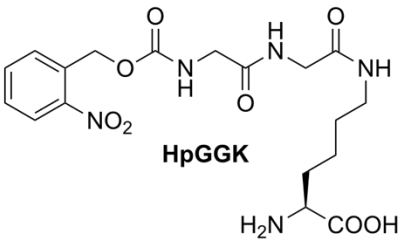
Molecular Weight: 286,29



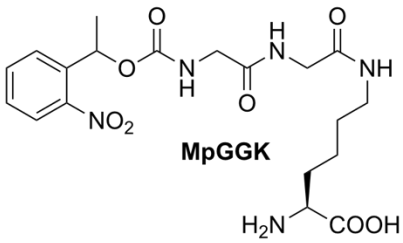
Molecular Weight: 382,37



Molecular Weight: 396,40



Molecular Weight: 439,43



Molecular Weight: 453,45

7.1.4. Molecular weights of proteins

Table 66 Overview of protein masses and extinction coefficients

Protein	Average mass [Da]	Monoisotopic mass [Da]	ε [M^{-1} cm^{-1}]
sfGFP-6xHis	27827.32	27809.92	18910
sfGFP-V2P-6xHis	27825.31	27807.91	18910
*Sa*ClpP-StrepII	22553.72	22539.59	14440
*Sa*ClpP-S98A-StrepII	22537.72	22523.59	14440
*Sa*ClpP-6xHis	22336.42	22322.45	14440
*Sa*ClpP-S98A-6xHis	22320.42	22306.46	14440
Rab1b_Q67A-6xHis	20301.00	20288.30	24410
Rab1b-6xHis	20429.13	20416.34	24410
StrepII-EntKin-DrrA$_{1-339}$	40719.37	40693.38	32890
StrepII-TEV-DrrA$_{16-352}$	40650.35	40624.56	34380
DrrA$_{16-352}$	38600.13	38575.62	27390
DrrA$_{16-533}$-TEV-His	60212.63	60174.51	32830
StrepII-EntKin-DrrA$_{1-647}$	75258.74	75211.22	62800
6xHis-Throm-*Mm*-C-term-PylRS-MF18	33322.06	33300.98	24410
6xHis-Throm-*Ma*-PylRS-MF18-H227I-Y228P	32821.23	32800.41	24410
Ma-PylRS-MF18-H227I-Y228P	30939.17	30919.54	24410
Mm-TEMPOH-I	50815.30	50783.30	-
Mm-PylRS	50921.42	50889.34	-
GroES	10386.95	10380.58	-
GroEL	57328.85	57292.69	-

7.1.5. Sources of crystal structures of proteins

Table 67 Overview about protein crystal structures obtained from the PDB

PDB code	Protein
2q7e	*Mm* pyrrolysyl-tRNA synthetase (catalytic domain) bound to an ATP analogue
6ezd	Pyrrolysyl-tRNA synthetase from *Canditatus Methanomethylophilus alvus* (MmaPylRS)
2e3c	Crystal structure of the catalytic domain of pyrrolysyl-tRNA synthetase (*Mm*)
5ud5	Crystal structure of the tRNA binding domain of Pyrrolysyl-tRNA synthetase bound to tRNA(Pyl) (*Mm*)
6jp2	Crystal structure of pyrrolysyl-tRNA synthetase from *Methanomethylophilus alvus*
2zin	Crystal structure of the catalytic domain of pyrrolysyl-tRNA synthetase in complex with BocLys and an ATP analogue (*Mm*)
3nku	Crystal structure of the N-terminal domain of DrrA/SidM from *Legionella pneumophila*
3nkv	Crystal structure of Rab1b covalently modified with AMP at Y77
3v5e	Crystal structure of ClpP from *Staphylococcus aureus* in the active, extended conformation
3v5i	The crystal structure of the mutant ClpP S98A (*Staphylococcus aureus*) (compact state)
3qwd	Crystal structure of ClpP from *Staphylococcus aureus* (compressed state)
2mws	Structure of the complex of ubiquitin and the ubiquitin-like (UBL) domain of Ddi1
2b3q	Crystal structure of a well-folded variant of green fluorescent protein

7.2. Chapter I

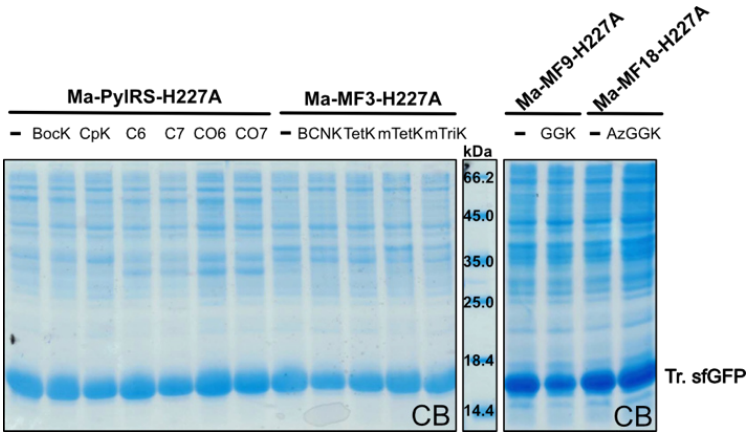


Figure 75: Amber suppression in sfGFP-V2P-N150TAG-His6 with *Ma*-PylRS mutants bearing a H227A mutation using different uAAs. SDS gels showing the cell lysates of the test expressions with different *Ma*-PylRS variants bearing H227A mutations. Expression of sfGFP-V2P-N150TAG-His6 using *Ma*-PylRS-H227A and *Ma*-MF3-H227A were performed in 2x YT medium. Expressions of sfGFP-V2P-N150TAG-His6 using *Ma*-MF9-H227A and *Ma*-MF18-H227A were done in AI medium without phenylalanine (AI -Phe). (CB = Coomassie brilliant blue; Met = methionine; Phe = phenylalanine; Tr. = truncated; "-" = no uAA added; C6 = BrC6K; C7 = BrC7K; CO6 = BrCO6K; CO7 = BrCO7K)

7.3. Chapter II

Table 68 Data collection and refinement statistics of the solved crystal structures *Mm*-MF18-CTD and *Ma*-MF18IP

	Mm-MF18-CTD	*Ma*-MF18IP
Wavelength	1.00	0.976
Resolution range	52.53 - 1.75 (1.81 - 1.75)	45.96 - 1.5 (1.55 - 1.5)
Space group	P 64	I 4
Unit cell	105.1 105.1 71.4 90 90 120	110.1 110.1 113.9 90 90 90
Total reflections	922,519 (88,681)	1,452,903 (123,913)
Unique reflections	45,233 (4,446)	107,620 (10,298)
Multiplicity	20.4 (19.7)	13.5 (12.0)
Completeness (%)	99.55 (98.69)	99.52 (95.46)
Mean I/sigma(I)	47.28 (1.11)	19.73 (1.30)
Wilson B-factor	31.96	22.04
R-merge	0.269 (3.155)	0.083 (1.764)
R-pim	0.069 (0.726)	0.023 (0.522)
CC1/2	0.985 (0.546)	1 (0.542)
CC*	0.996 (0.841)	1 (0.839)
Reflections used in refinement	45,054 (4,446)	107,610 (10,296)
Reflections used for R-free	2,250 (215)	5,381 (515)
R-work	0.202 (0.342)	0.220 (0.388)
R-free	0.217 (0.361)	0.232 (0.377)
CC (work)	0.925 (0.686)	0.956 (0.639)
CC (free)	0.941 (0.583)	0.953 (0.661)
Number of non-hydrogen atoms	2,330	4,808
macromolecules	2,176	4,327
ligands	33	80
solvent	121	401
Protein residues	262	540
RMS (bonds)	0.013	0.016
RMS (angles)	1.64	1.74
Ramachandran favored (%)	98.06	99.23
Ramachandran allowed (%)	1.94	0.77
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	2.92	1.05
Clashscore	0.68	3.54
Average B-factor	28.71	24.42
macromolecules	27.96	23.68
ligands	28.55	23.43
solvent	42.30	32.56
Number of TLS groups	1	

Statistics for the highest-resolution shell are shown in parentheses.

7.3.1. *M. mazei* synthetases

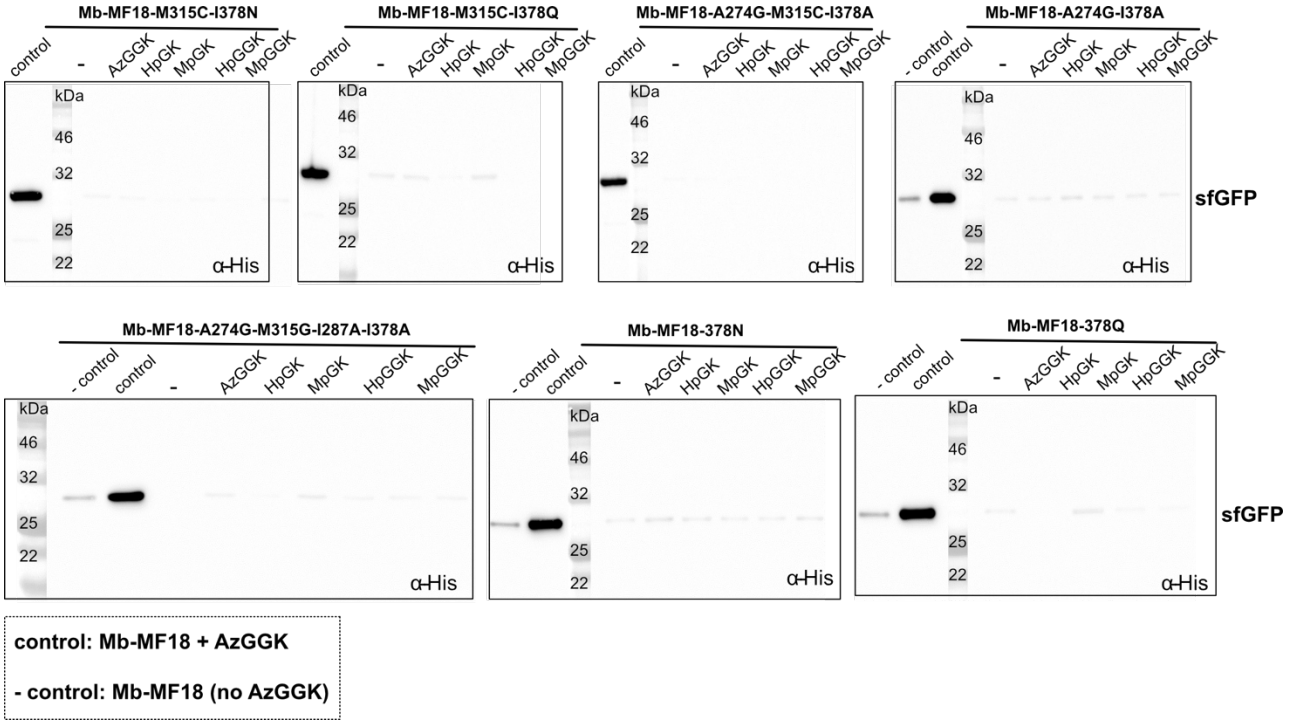


Figure 76: Western blots (α-His) of *E. coli* cell lysates obtained from test expressions for the incorporation of AzGGK derivatives into sfGFP-V2P-N150TAG-His6 using *Mb*-MF18 mutants. As positive control for successful sfGFP production served a cell lysate from an expression using AzGGK and *Mb*-MF18. As negative control (-control) an expression without uAA addition was used.

7.3.2. *M. alvus* synthetases

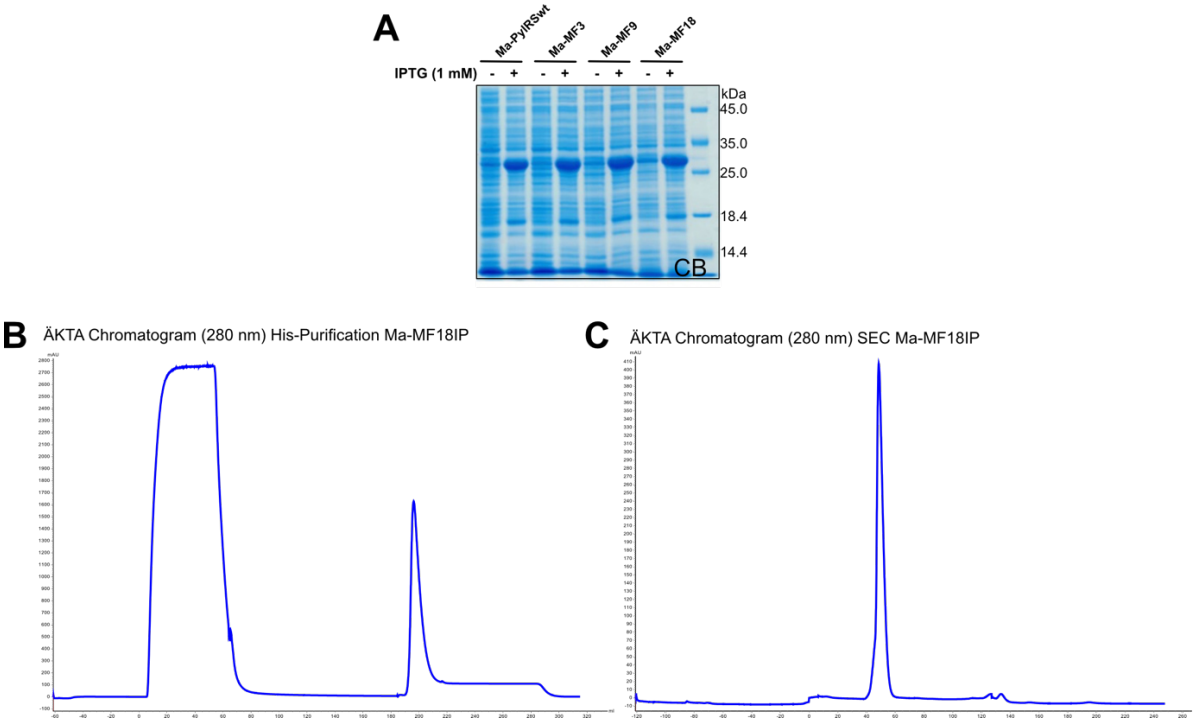


Figure 77: Expression of *Ma*-PylRS variants and purification of *Ma*-MF18-H227I-Y228P for crystallization experiments. A: SDS gel *of E. coli* Rosetta (DE3) cell lysates obtained from the induced (+ IPTG) and uninduced (- IPTG) expression of *Ma*-PylRS variants. B: ÄKTA chromatograms (280 nm) of the His-purification step of *Ma*-MF18-H227I-Y228P (*Ma*-MF18IP) for later crystallization experiments. C: ÄKTA chromatograms (280 nm) of the SEC-purification step of *Ma*-MF18IP for later crystallization experiments. (CB = Coomassie brilliant blue; IPTG = Isopropyl-β-D-thiogalactopyranoside; SEC = size exclusion chromatography)

7.4. Chapter III

7.4.1. Stabilization of Rab1b-DrrA complexes

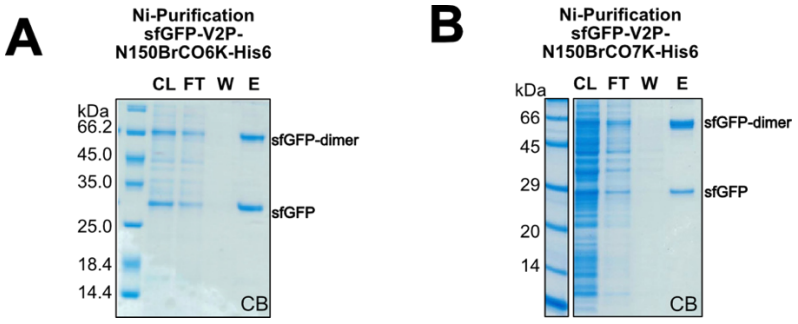


Figure 78: SDS gels of Ni-purifications of amber suppressions with sfGFP-V2P-N150TAG-His6 and *Mb*-TEMPOH-I using either BrCO6K or BrCO7K in *E. coli*. A: Purification of sfGFP-V2P-N150BrCO6K-His6. B: sfGFP-V2P-N150BrCO7K-His6. (CB = Coomassie brilliant blue, CL = Cleared lysate, FT = Flow through, W = Wash, E = Elution)

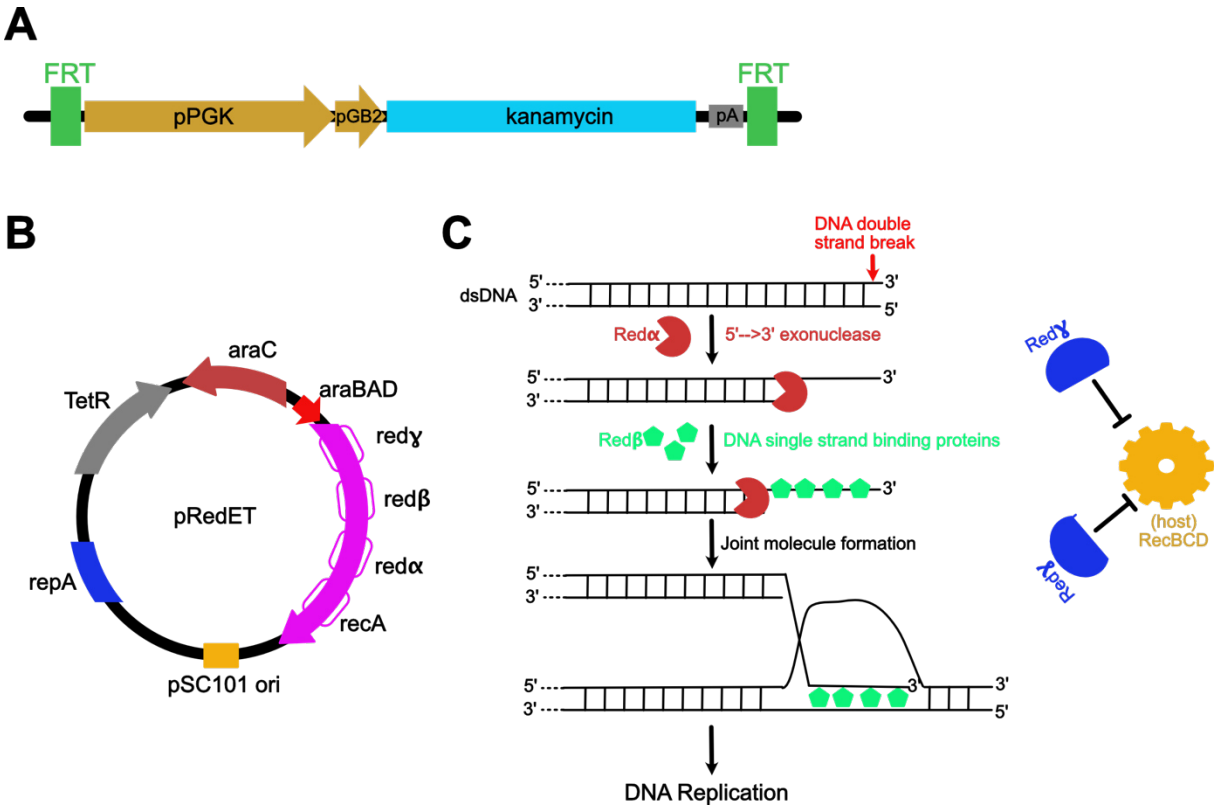


Figure 79: Necessary components for *E. coli* gene deletion (from Quick & Easy Gene Deletion Kit, Gene Bridges GmbH). A: Scheme of the Kan^R-cassette used for homologous recombination with the E. coli target gene. B: Scheme of the pRedET plasmid encoding all necessary proteins for RedET-mediated homologous recombination (Redα, β, γ) regulated through an araBAD promotor. The replication and propagation of pRedET is regulated by the thermo-sensitive pSC101 ori and required RepA.[333] RecA is an *E. coli* protein involved in DNA repair mechanisms.[241, 334] C: Principle of RedET-mediated homologous recombination. Redα recognize dsDNA breaks and removes with its exonuclease activity bases from the 5'-end. Redβ binds to new derived 3'-DNA single strand. Joint molecule formation can now occur and is resolved during replication leading to partial DNA replacement. Redγ inhibits *E. coli* RecBCD which is involved in DNA repair.[239, 242, 335] (Kan^R = kanamycin resistance, pPGK= eukaryotic pPGK-promotor, pGB2 = prokaryotic pGK2-promotor, pA = polyadenylation sequence, dsDNA = double stranded DNA; genes are written in small letters and corresponding proteins are written with a capital)

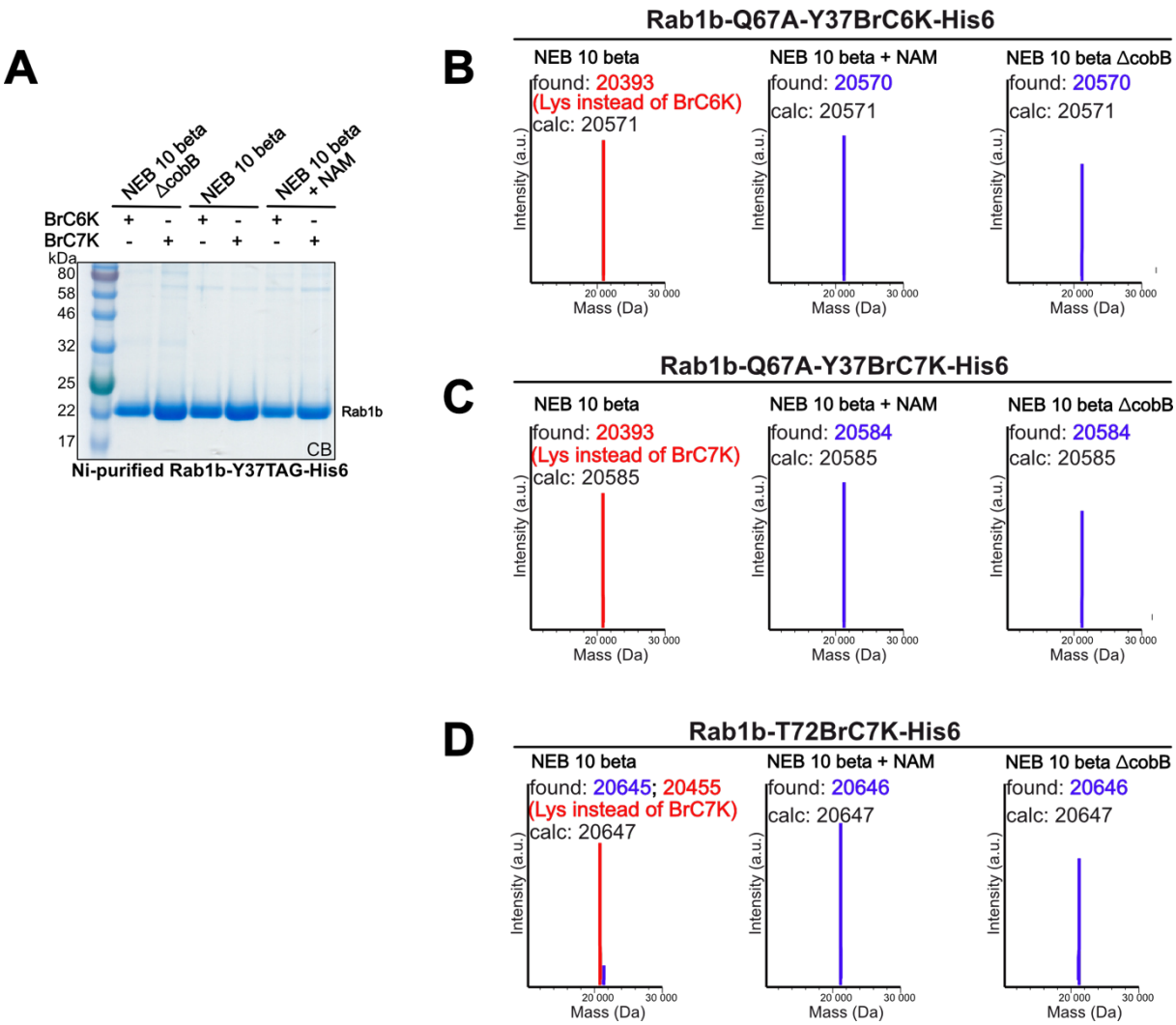


Figure 80: Amber suppression of Rab1b-Y37TAG-His6 and Rab1b-T72TAG-His6 with BrC6K and BrC7K using different expression conditions and *E. coli* strains. A: SDS gel of Ni-purified Rab1b-Y37BrC6K-His6 and Rab1b-Y37BrC7K-His6 obtained from expression with either *E. coli* NEB 10-beta (+/- NAM) or *E. coli* NEB 10-beta ΔcobB. B: ESI-MS data (average masses) of purified Rab1b-Y37BrC6K-His6 obtained from expression with either *E. coli* NEB 10-beta (+/- NAM) or *E. coli* NEB 10-beta ΔcobB. C: ESI-MS data (average masses) of purified Rab1b-Y37BrC7K-His6 obtained from expression with either *E. coli* NEB 10-beta (+/- NAM) or *E. coli* NEB 10-beta ΔcobB. D: ESI-MS data (average masses) of purified Rab1b-T72BrC7K-His6 obtained from expression with either *E. coli* NEB 10-beta (+/- NAM) or *E. coli* NEB 10-beta ΔcobB. CB = Coomassie brilliant blue, NAM = nicotinamide)

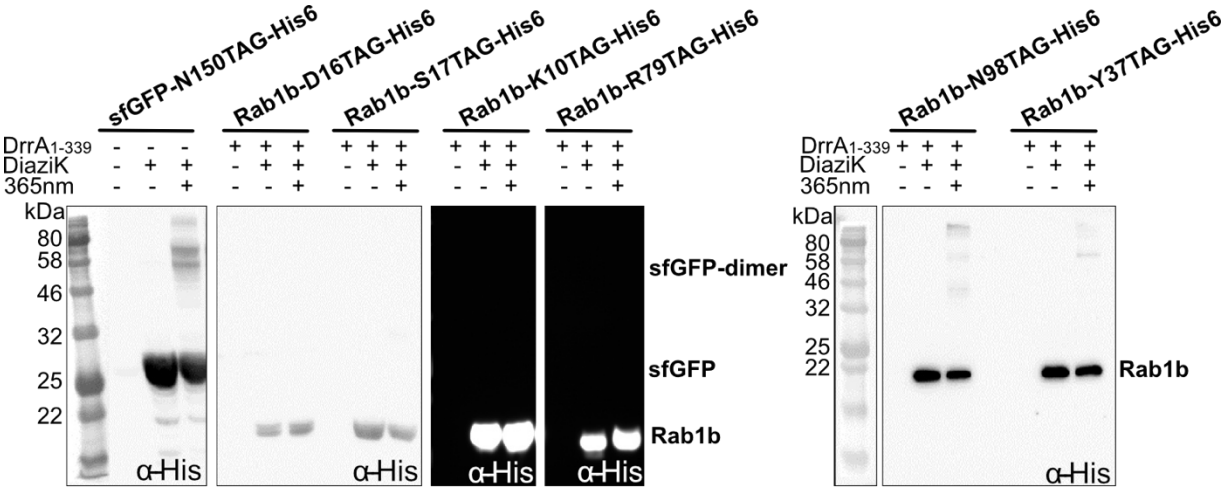


Figure 81: Identification of suitable positions in Rab1b for crosslink formation to DrrA-ATase using a screen with DiaziK. A: Western blot (α-His) of cell lysates from test expressions with triple-transformed (Rab1b-TAG-His6 variant, DrrA$_{1-339}$, *Mb*-TEMPOH-I) *E. coli* NEB 10-beta using DiaziK. Samples were irradiated with UV light (365 nm) for 10 min to induce crosslink formation. (365 nm = UV light at 365 nm for 10 min).

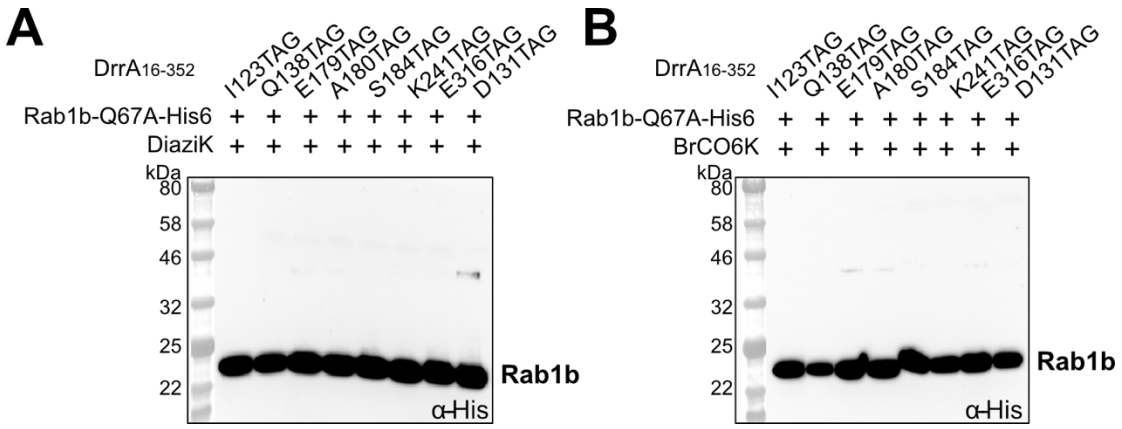


Figure 82: Identification of suitable positions in DrrA$_{16-352}$ for crosslink formation to Rab1b-Q67A-His6. A: Western blot (α-His) of cell lysates from test expressions with triple-transformed (Rab1b-Q67A-His6, DrrA$_{16-352}$-TAG variant, *Mb*-TEMPOH-I) *E. coli* NEB 10-beta using DiaziK. Samples were irradiated with UV light (365 nm) for 10 min to induce crosslink formation. (365 nm = UV light at 365 nm for 10 min). B: Western blot (α-His) of cell lysates from test expressions with triple-transformed (Rab1b-Q67A-His6, DrrA$_{16-352}$-TAG variant, *Mb*-TEMPOH-I) *E. coli* NEB 10-beta using BrCO6K.

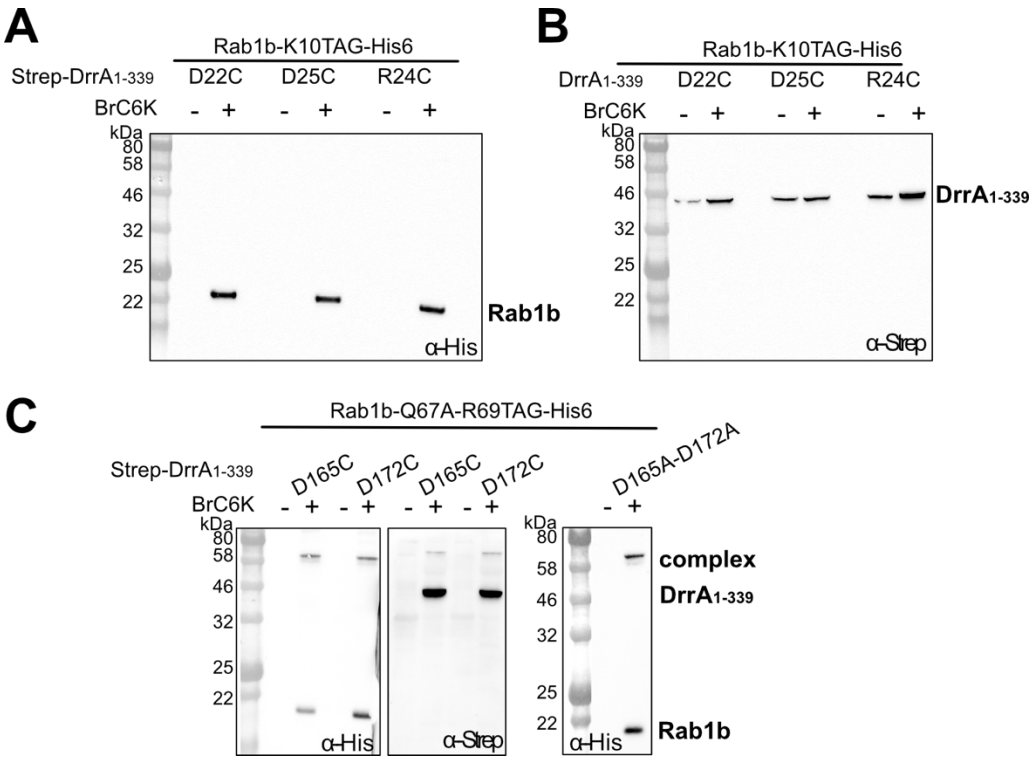


Figure 83: Validation of further Rab1b-Q67A-TAG-His6 and DrrA$_{1-339}$-cysteine mutants for crosslink formation using BrC6K. A: Western blot (α-His) of cell lysates from test expressions with *E. coli* NEB 10-beta using Rab1b-K10TAG-His6 combined with different DrrA$_{1-339}$ cysteine mutants. B: Corresponding western blot (α-Strep) to the experiment from A. C: Western blot (α-His and α-Strep) of cell lysates from test expressions with *E. coli* NEB 10-beta using Rab1b-Q67A-R69TAG-His6 combined with different DrrA$_{1-339}$ cysteine mutants or the corresponding double alanine mutant.

Scan	PreMass	Ch	kojak_score	delta_score	ppm_error	Peptide
116	992.5188	3	0.1950	0.1950	-2.804	R.NSAYKVVK.E
135	1387.5739	4	0.0800	0.0600	8.7119	K.AQYEEM[147]AKC[160]SR.A
961	3185.7184	8	0.3150	0.3050	1.4114	K.AIRLRAISALDRTFISQIPVDLEANEM[147]K.E
1158	2357.3506	8	0.9700	0.1450	8.8373	K.KMLILEKR.M
1968	2926.3432	8	0.1550	0.0550	-0.3601	K.K[304]ILVFENEMETHKNIEDEYKNK.F
2033	8360.1617	8	0.0300	0.0300	-5.695	R.ENLIGDGDIDAAC[160]ADGEGNNDTDAQDPNHNYPC[160]NDC[160]RDGVDDRDY[492]DY[492]QAPNY[
2051	4881.9233	4	0.4750	0.0050	2.6312	R.C[160]YC[160]QYSGYVNAC[160]EQSC[160]PNSNC[160]ENVDEC[160]SRGDSSLR.F
2185	2097.8911	7	0.3100	0.0350	-6.589	K.EPGK.R
2389	1282.7426	4	1.1150	0.0300	-5.396	S.HSSK.G
2390	1293.6940	4	2.5100	1.2600	0.1087	K.RFELHAEHKK.E
2391	1445.8575	4	0.9450	0.0050	9.8831	K.IKSFSGKK.G
2392	1641.9054	5	0.6300	0.1300	8.9575	K.KLLVFENEAATRR.N
2393	1533.8806	4	1.5150	0.3000	-4.387	R.HGSNPIR.R
2394	1659.8470	4	0.9600	0.0350	-6.559	R.GDC[160]ALKDAR.A
2395	3624.5779	4	0.7450	0.0050	7.2318	K.C[160]SQSENPSEC[160]QDKGEEELEK.Y
2396	1816.8179	2	0.7850	0.4300	-8.901	K.C[160]AC[160]SHR.E
2397	4011.7152	4	0.2000	0.1350	9.9200	K.SGGLSY[492]LSRSSFGGGY[492]DGGYRRSQSVSVLAQR.G
2399	1641.9056	4	0.5500	0.0900	8.8482	K.KLLVFENEAATRR.N
2400	3705.6085	4	0.2250	0.0050	9.7167	R.TSQSLEIELQSHLSMK[304]ESLEHTLEETK[304]AR.Y
2401	2883.4728	4	0.3000	0.3000	-5.602	R.YTAIEVDLALK[304]VNMLAQVDRLLR.A
2402	3789.5600	4	0.4300	0.0050	0.8993	K.C[160]SQSENPSEC[160]QDKGEEELEK.Y
2403	1674.8215	4	0.5250	0.0050	8.5587	K.DGGDAR.K
2404	3195.4587	4	0.3600	0.1500	-3.476	K.YILQGVTSWGLGC[160]ARPNK[304]PGVY[492]VR.V
2406	1392.7453	4	0.8350	0.7000	9.4813	R.GAKRR.S
2408	2265.1378	4	1.2550	0.8900	2.2376	R.FAADVPSPADKWR.E
2412	1502.7504	4	0.4700	0.2200	-3.188	K.KAANAK[304]KDAVSPK.M
2413	1903.0165	4	0.2750	0.0250	5.7648	R.DGKVISSREHVQSRPL.-
2414	1448.6252	4	0.5750	0.4800	-0.7333	K.KEHDEMPYSDAK.A
2421	4970.9518	4	0.5300	0.0050	-2.104	K.GFLVRVTAIFIDLALK.V

Figure 84: MS/MS analysis and identification of the crosslinked peptide in the Rab1b-Q67A-R69BrC6K-His6-DrrA$_{1\text{-}339}$ complex. Best hit indicates a crosslink to D82 in DrrA$_{1\text{-}339}$. The peptide was identified and validated using the software Kojak[327, 336].

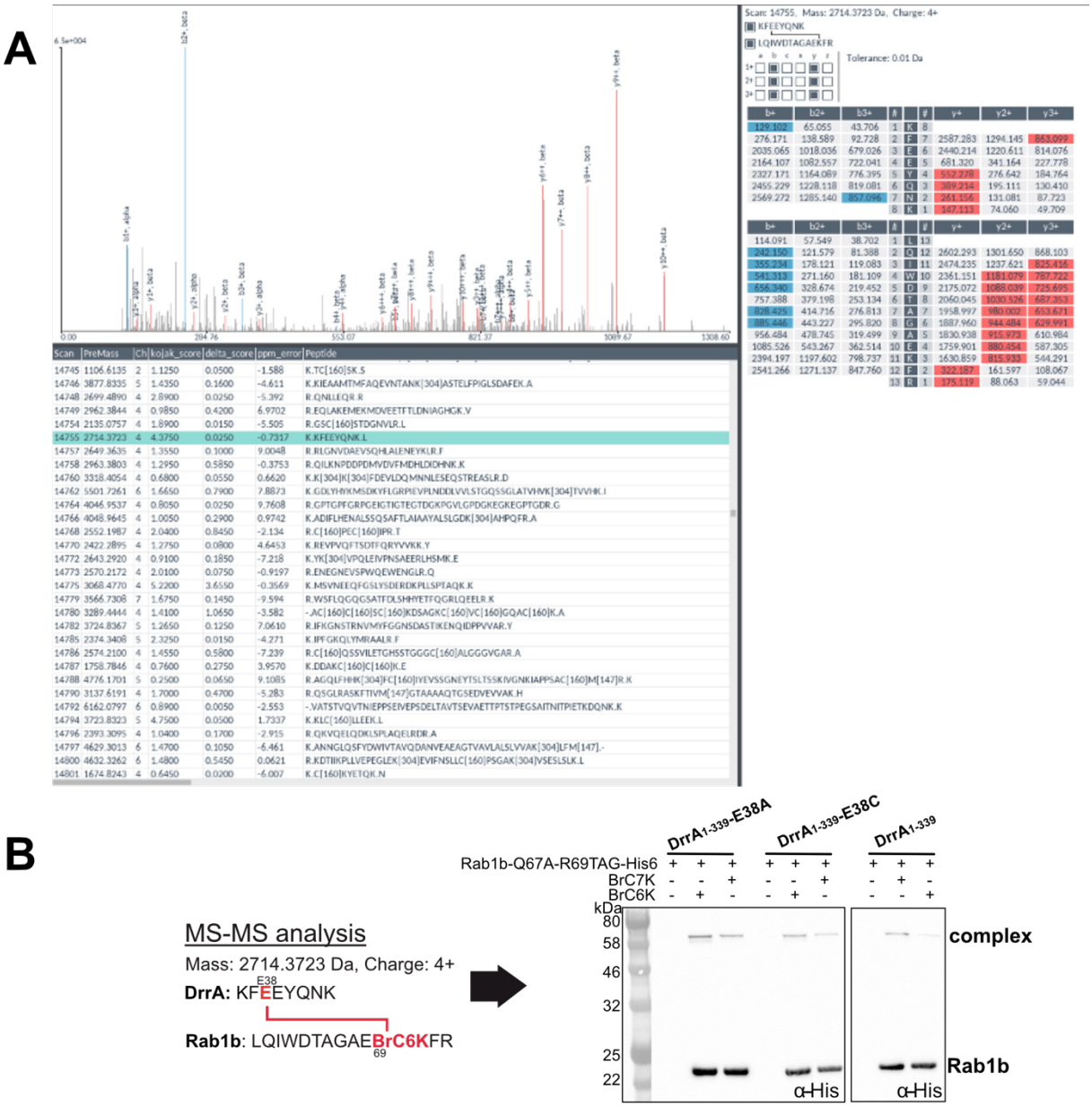


Figure 85: A: MS/MS analysis and identification of the crosslinked peptide in the Rab1b-Q67A-R69BrC6K-His6-DrrA$_{1-339}$ complex. A possible second hit indicates a crosslink to E38 in DrrA$_{1-339}$. The peptide was identified and validated using the software Kojak[327, 336]. B: MS/MS analysis of the identified crosslinked peptide Rab1b-Q67A-R69CrC6K-His6 and DrrA$_{1-339}$ shows a potential crosslink to E38 in DrrA$_{1-339}$. Crosslink is schematically indicated in red. Western blot (α-His) of cell lysates from test expressions using BrC6K or BrC7K in co-transformed *E. coli* NEB 10-beta with pBAD-Duet_Rab1b-Q67A-R69TAG-His6_Strep-DrrA$_{1-339}$-E38X and pEVOL-*Mm*-TEMPOH-I-PylT. (X = alanine or cysteine or aspartate)

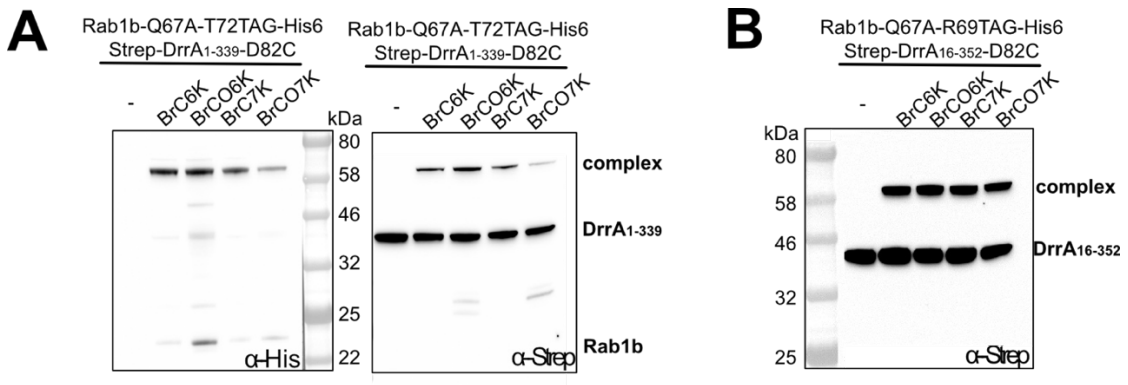


Figure 86: Crosslink formation between Rab1b-Q67A-XXXTAG-His6 and Strep-DrrA$_{1-339}$-D82C using different uAAs (BrC6K, BrC7K, BrCO6K, BrCO7K). A: Western blot (α-His) and western blot (α-Strep) of cell lysates from test expressions with co-transformed (pBAD-Duet_Rab1b-Q67A-T72TAG-His6_Strep-DrrA$_{1-339}$-D82C and pEVOL- *Mm*-TEMPOH-I-PylT) *E. coli* NEB 10-beta using different uAAs. B: Western blot (α-Strep) of cell lysates from test expressions with co-transformed (pBAD-Duet_Rab1b-Q67A-R69TAG-His6_Strep-DrrA$_{1-339}$-D82C and pEVOL-*Mm*-TEMPOH-I-PylT) *E. coli* NEB 10-beta using different uAAs.

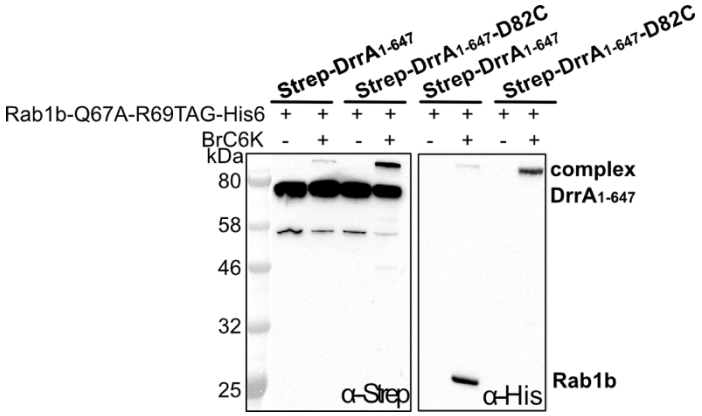


Figure 87: Crosslink formation between Rab1b-Q67A-R69TAG-His6 and Strep-DrrA$_{1-647}$-D82X using BrC6K. Western blot (α-Strep) and western blot (α-His) of cell lysates from test expressions with co-transformed (pBAD-Duet_Rab1b-Q67A-R69TAG-His6_Strep-DrrA$_{1-647}$-D82C and pEVOL- *Mm*-TEMPOH-I-PylT) *E. coli* NEB 10-beta using BrC6K. (X = aspartate or cysteine)

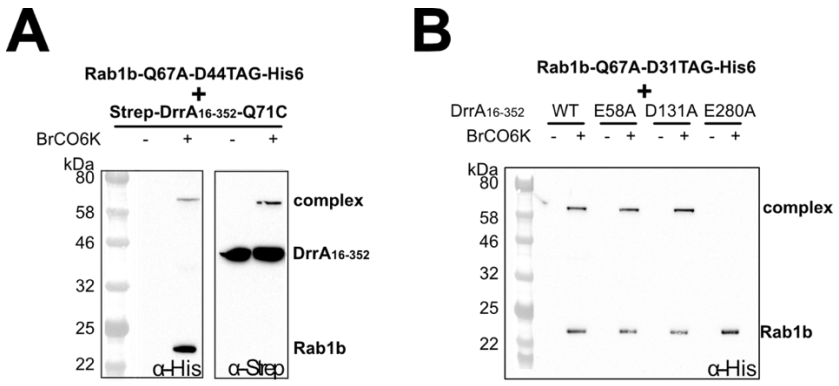


Figure 88: Validation of the Rab1b-DrrA interface. A: Western blot (α-His and α-Strep) of cell lysates of amber suppression using Rab1b-Q67A-D44BrCO6K-His6 and Strep-DrrA$_{16-352}$-Q71C. B: Western blot (α-His) of cell lysates of amber suppression using Rab1b-Q67A-D31BrCO6K-His6 in combination with different DrrA$_{16-352}$-alanine mutants which were supposed to be possible crosslink candidates and to identify the respective crosslinking partner.

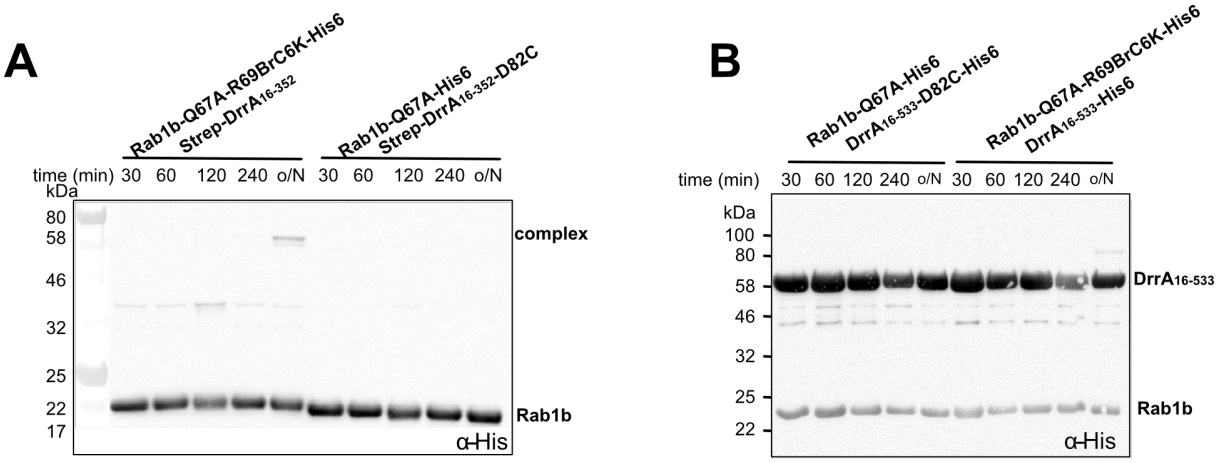


Figure 89: *In vitro* crosslinking experiments with Rab1b-Q67A-His variants and DrrA variants. A: Western blot (α-His) of the time dependent crosslink formation of samples containing purified Rab1b-Q67A-R69BrC6K-His6 and DrrA$_{16-352}$ or Rab1b-Q67A-His6 and DrrA$_{16-352}$-D82C. B: Western blot (α-His) of the time dependent crosslink formation of samples containing purified Rab1b-Q67A-His6 and DrrA$_{16-533}$-D82C or Rab1b-Q67A-R69BrC6K-His6 and DrrA$_{16-533}$. (o/N = overnight)

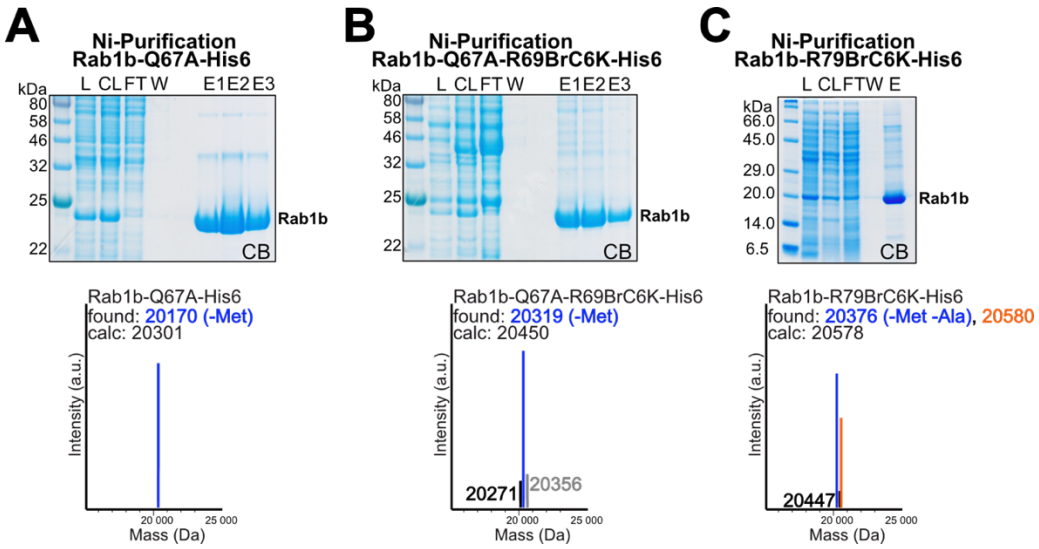


Figure 90: Ni-Purifications of Rab1b variants. A: SDS gels of the purification of Rab1b-Q67A-His6 and ESI-MS data (average masses) of purified Rab1b-Q67A-His6. B: SDS gels of the purification of Rab1b-Q67A-R69BrC6K-His6 and ESI-MS data (average masses) of purified Rab1b-Q67A-R69BrC6K-His6. C: SDS gels of the purification of Rab1b-R79BrC6K-His6 and ESI-MS data (average masses) of purified Rab1b-R79BrC6K-His6. (CB = Coomassie brilliant blue, L = Lysate; CL = Cleared lysate, FT = Flow-through, W = Wash, E1-3 = Elution 1-3)

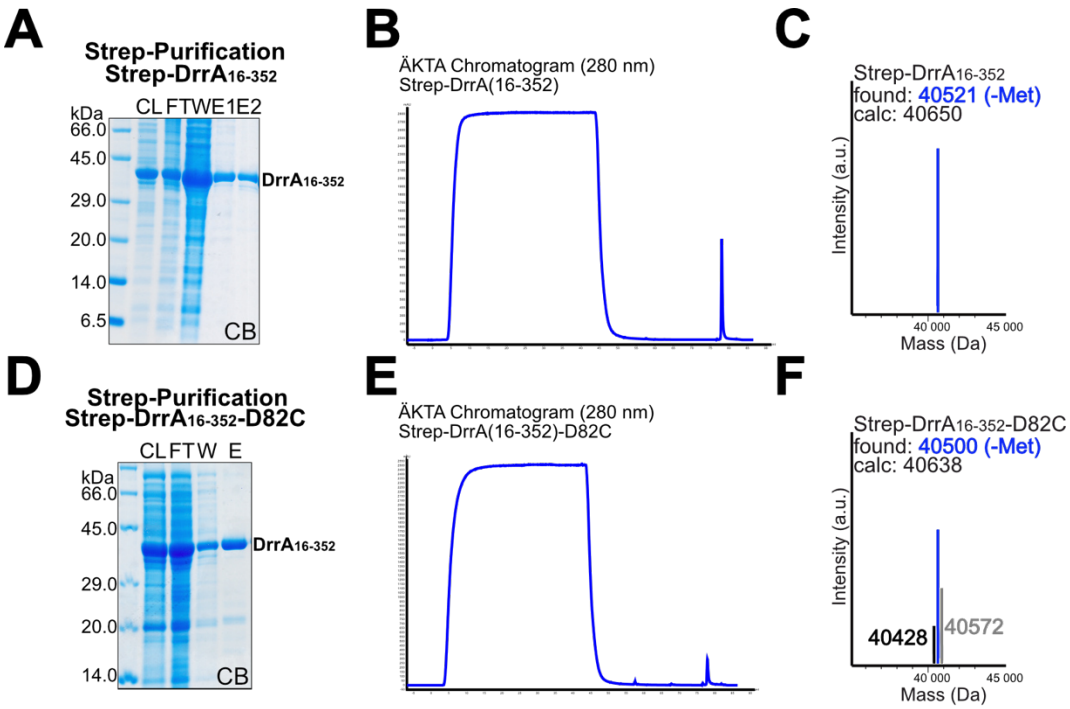


Figure 91: Strep-Purifications of Strep-DrrA16-352 variants. A: SDS gels of the purification of Strep-DrrA16-352. B: ÄKTA chromatogram (280 nm) of the Strep-purification of Strep-DrrA16-352. C: ESI-MS data (average masses) of purified Strep-DrrA16-352. D: SDS gels of the purification of Strep-DrrA16-352-D82C. E: ÄKTA chromatogram (280 nm) of the Strep-purification of Strep-DrrA16-352-D82C. F: ESI-MS data (average masses) of purified Strep-DrrA16-352-D82C. F: ÄKTA chromatogram (280 nm) of the Strep-purification of Strep-DrrA16-352. (CB = Coomassie brilliant blue, L = Lysate; CL = Cleared lysate, FT = Flow-through, W = Wash, E1-2 = Elution 1-2)

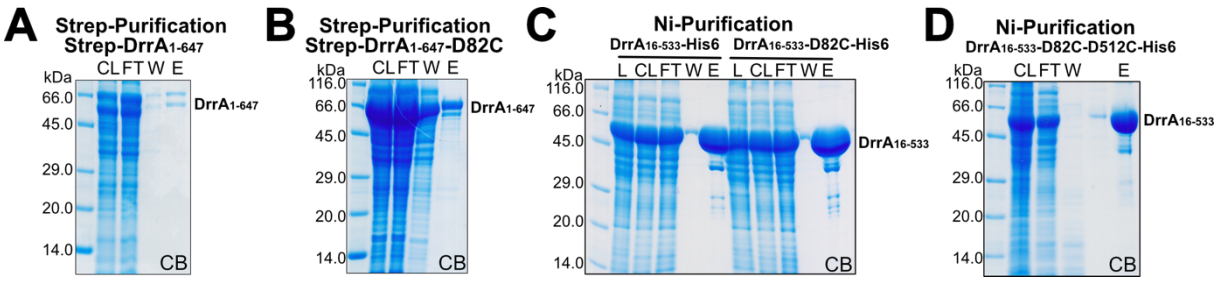


Figure 92: Purifications of Strep-DrrA1-647 variants and DrrA16-533-His6 variants. A: SDS gels of the Strep-purification of Strep-DrrA1-647. B: SDS gels of the Strep-purification of Strep-DrrA1-647-D82C. C: SDS gels of the Ni-purification of DrrA16-533-His6 and DrrA16-533-D82C-His6. D: SDS gels of the Ni-purification of DrrA16-533-His6 and DrrA16-533-D82C-D512C-His6. (CB = Coomassie brilliant blue, L = Lysate; CL = Cleared lysate, FT = Flow-through, W = Wash, E = Elution)

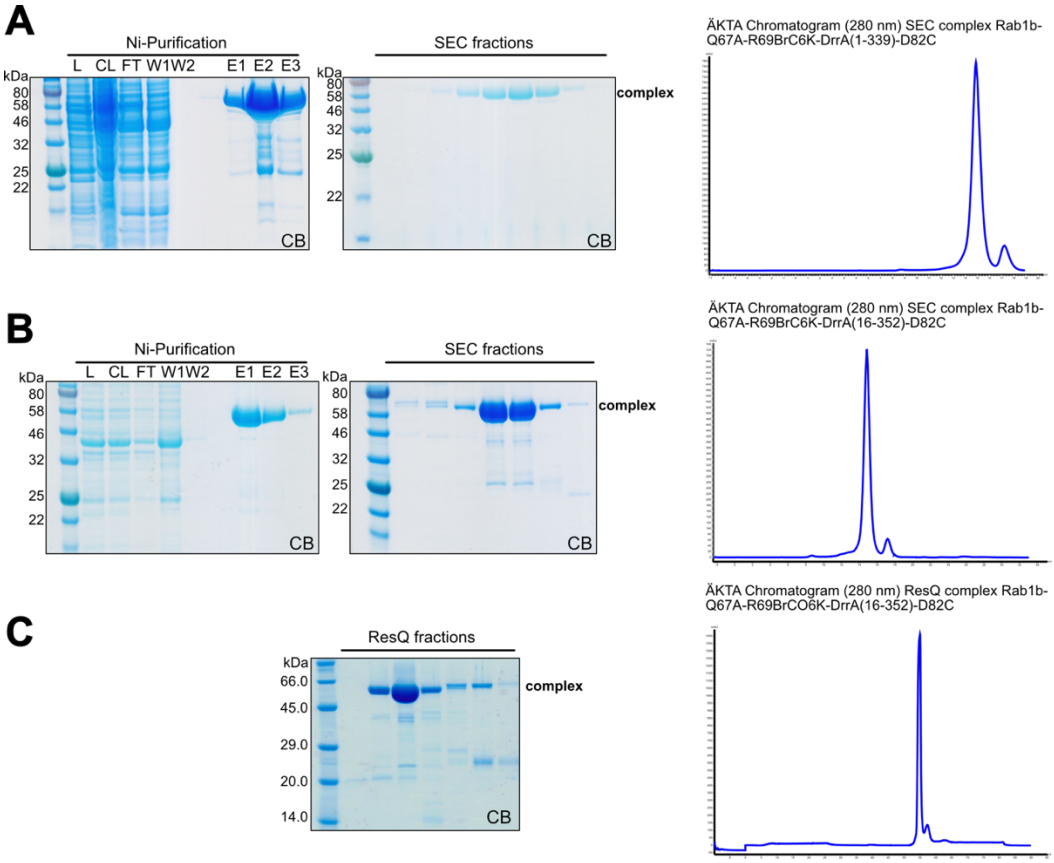


Figure 93: Purification of the Rab1b-Q67A-R69uAA-His6-DrrA-ATase-D82C complexes. A: SDS gel of the Ni-Purification and SEC fractions and ÄKTA chromatogram (280 nm) of Rab1b-Q67A-R69BrC6K-His6-DrrA$_{1-339}$-D82C complex. B: SDS gel of the Ni-Purification and SEC fractions and ÄKTA chromatogram (280 nm) of Rab1b-Q67A-R69BrC6K-His6-DrrA$_{16-352}$-D82C complex. C: SDS gel of the Ni-Purification and ResQ fractions and ÄKTA chromatogram (280 nm) of Rab1b-Q67A-R69BrCO6K-His6-DrrA$_{16-352}$-D82C complex. (CB = Coomassie brilliant blue, L = Lysate; CL = Cleared lysate, FT = Flow-through, W1/W2 = Wash 1/Wash2, E1-3 = Elution 1-3)

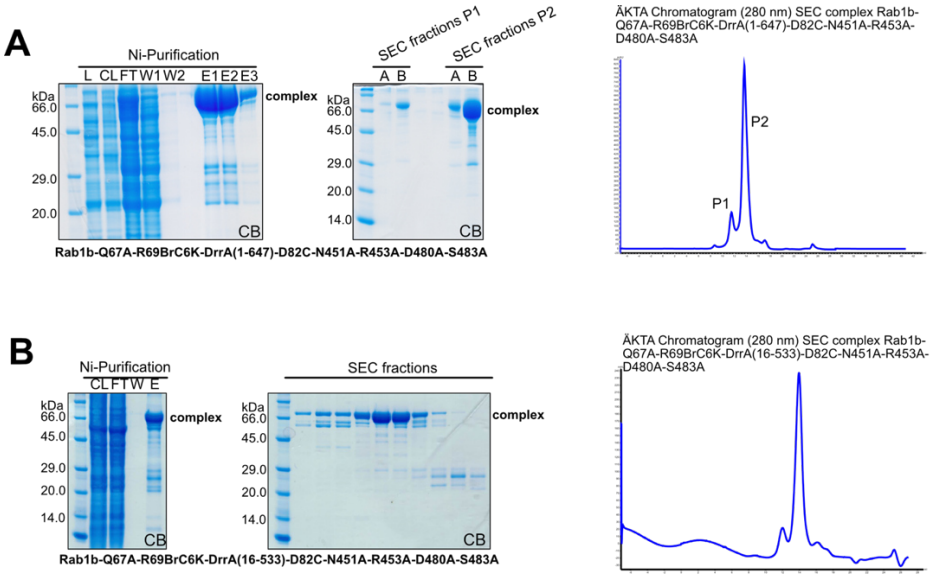


Figure 94: Purification of the Rab1b-Q67A-R69uAA-His6-DrrA-ATase-GEF-D82C complexes. A: SDS gel of the Ni-Purification and SEC fractions and ÄKTA chromatogram (280 nm) of Rab1b-Q67A-R69BrC6K-His6-DrrA$_{1-647}$-D82C complex. The ÄKTA chromatogram (280 nm) shows two peaks (P1 & P2). B: SDS gel of the Ni-Purification and SEC fractions and ÄKTA chromatogram (280 nm) of Rab1b-Q67A-R69BrC6K-His6-DrrA$_{16-533}$-D82C complex. (CB = Coomassie brilliant blue, L = Lysate; CL = Cleared lysate, FT = Flow-through, W1/W2 = Wash 1/Wash2, E1-3 = Elution 1-3, P1 = SEC peak 1, P2 = SEC peak 2, A = fraction A of P1 or P2, B = fraction B of P1 or P2)

7.4.2. MS^2-cleavable uAAs for protein crosslinking

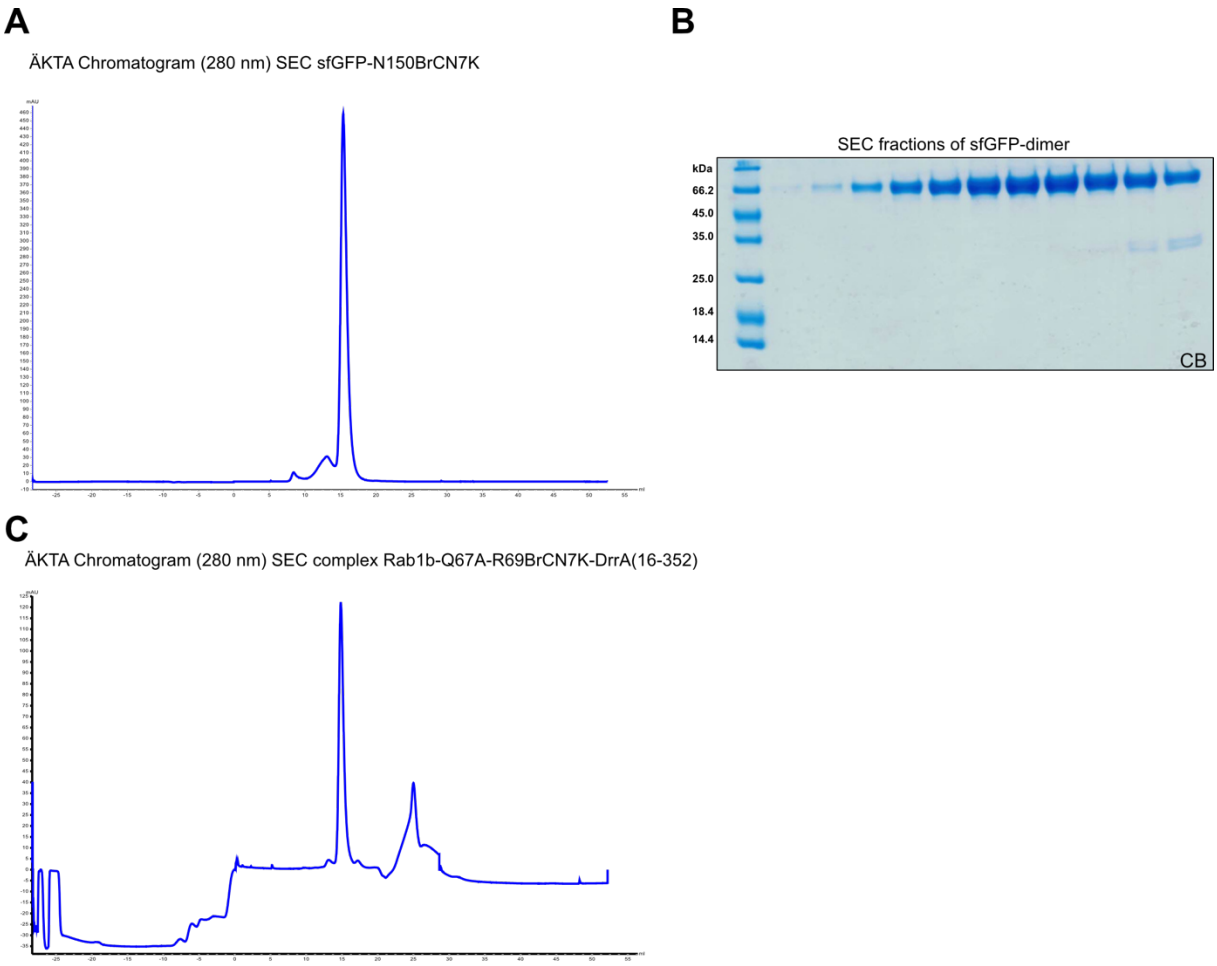


Figure 95: Purification of protein complexes crosslinked via BrCN7K. A: ÄKTA chromatogram (280 nm) of the SEC of the sfGFP-dimer. B: SDS gel of the fractions obtained from SEC of the sfGFP-dimer. C: ÄKTA chromatogram (280 nm) of the SEC of the Rab1b-Q67A-R69BrCN7K-His6-DrrA$_{16\text{-}352}$-D82C complex. (CB = Coomassie brilliant blue)

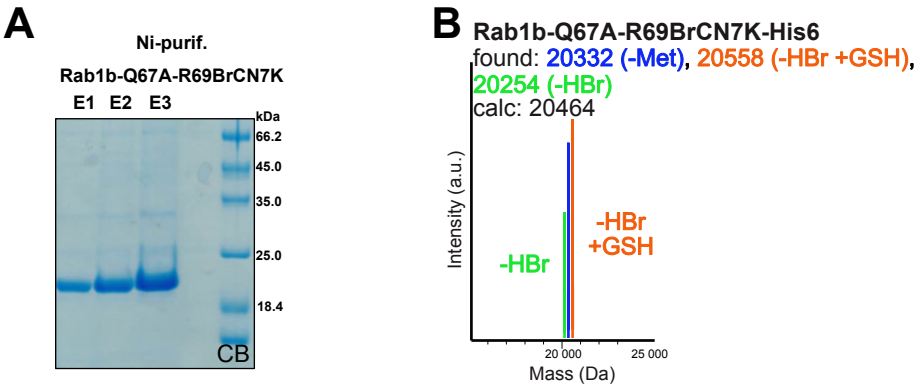


Figure 96: Ni-purification of Rab1b-Q67A-R69BrCN7K-His6 from *E. coli*. A: SDS gel of the elution fractions (E1 + E2 + E3) of the Ni-purification of Rab1b-Q67A-R69BrCN7K-His6 from *E. coli*. B: ESI-MS data (averaged masses) of purified Rab1b-Q67A-R69BrCN7K-His6 derived from B. (CB = Coomassie brilliant blue, GSH = glutathione)

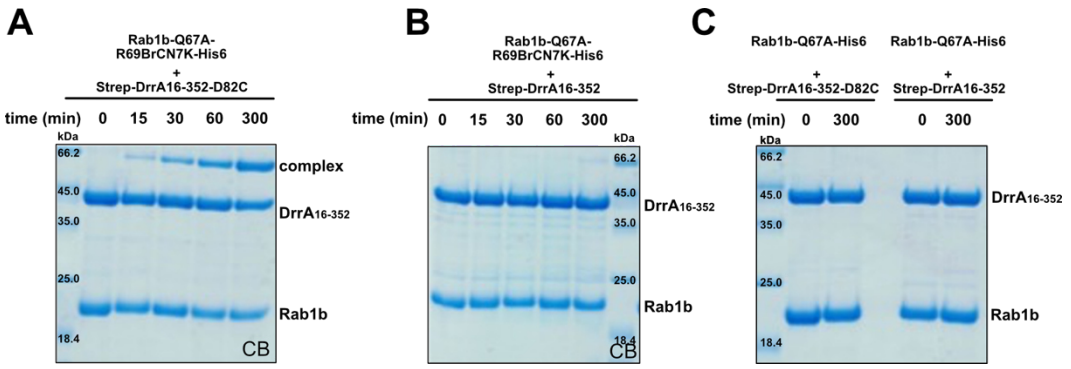
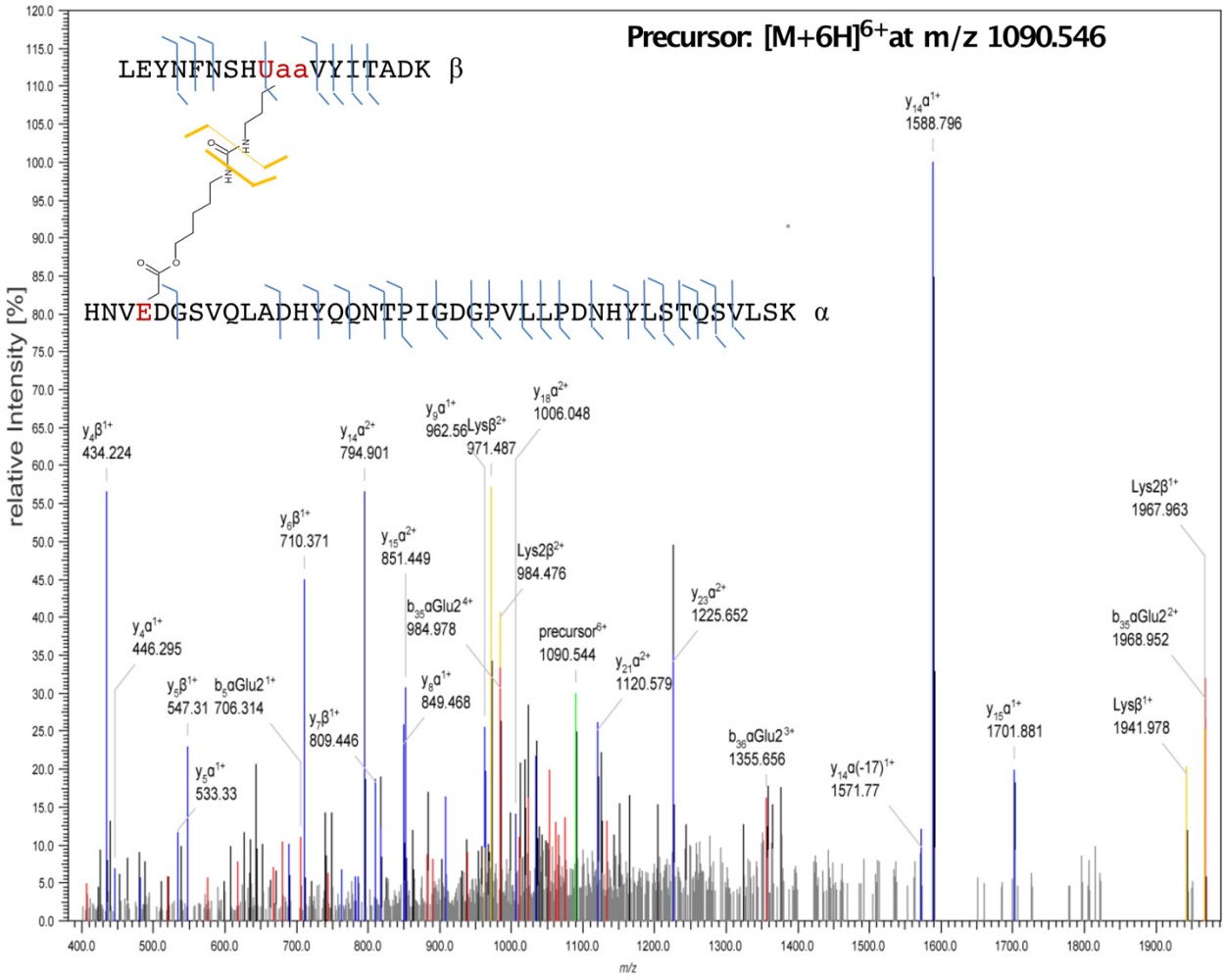


Figure 97: SDS gels of the time-dependent *in vitro* crosslinking of different combinations of purified Rab1b-Q67A-His6 variants and Strep-DrrA$_{16-352}$ variants. (CB = Coomassie brilliant blue)



Nr	MeroX Score	Pepscore 1	Pepscore 2	m/z	charge	M+H+	M+H+ calc.	deviation [ppi	peptide 1	peptide 2	link position 1	link position 2
1	66	263	29	973.493	7	6.808.406	6.807.408	146.61	[IRHNVxDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSK]	[LEYNFNSHUVYITADK]	x6	U9
2	66	284	35	1.135.579	6	6.808.436	6.807.408	151.03	[IRHNVxDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSK]	[LEYNFNSHUVYITADK]	x6	U9
3	59	104	11	934.896	7	6.538.228	6.538.223	0.81	[HNVxDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSK]	[LEYNFNSHUVYITADK]	x4	U9
4	50	131	40	1133.57	6	6.796.384	6.794.376	295.43	[HNVxDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSK]	[LEYNFNSHUVYITADKQK]	x4	U9
5	46	286	41	1.090.883	6	6.540.259	6.538.223	311.37	[HNVxDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSK]	[LEYNFNSHUVYITADK]	x4	U9

Figure 98: MS^2 analysis of the purified and digested sfGFP-dimer (sfGFP-N150BrCN7K-His6-dimer). MS^2 fragment spectrum of the sfGFP-dimer obtained from the data evaluation using MeroX (version 2.0.1.4). Cleavage products of the urea linker are colored in yellow. The y-ions (derived from C-terminal peptide fragments) are labeled in blue and b-ions (derived from N-terminal peptide fragments) are labeled in red which belong to the cleavage of the peptide backbone (the lowered number indicates the number of residues encompassed in the fragment). The table below listed the identified crosslinked peptide products obtained by MS^2 analysis of the sfGFP-dimer.

Nr	MeroX Score	Pepscore 1	Pepscore 2	m/z	charge	M+H+	M+H+ calc.	deviation [ppm]	peptide 1	protein 1	peptide 2	protein 2	link position 1	link position 2
1	121	147	69	849.096	+6	5089.541	5089.533	1.44	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
2	120	85	74	866.179	+4	3461.692	3461.694	-0.56	[EMIYDAFCALGVEMPK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
3	120	194	82	849.096	+6	5089.54	5089.533	1.23	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
4	120	149	59	849.096	+6	5089.541	5089.533	1.44	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
5	120	130	74	849.096	+6	5089.54	5089.533	1.3	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
6	119	182	97	849.096	+6	5089.54	5089.533	1.23	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
7	118	186	87	849.095	+6	5089.533	5089.533	-0.14	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
8	116	162	23	792.059	+6	4747.315	4747.307	1.78	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
9	116	147	30	792.059	+6	4747.317	4747.307	2.24	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
10	115	233	131	951.735	+4	3803.916	3803.921	-1.19	[TIKLQIWDTAGAEKFR]	>Rab1b	[EMIYDAFCALGVEMPK]	>DrrA	K14	C8
11	115	225	47	792.057	+6	4747.306	4747.307	-0.07	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
12	115	111	47	1018.713	+5	5089.536	5089.533	0.5	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
13	115	182	38	950.268	+5	4747.308	4747.307	0.37	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
14	115	171	30	792.059	+6	4747.318	4747.307	2.32	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
15	115	130	32	1018.712	+5	5089.532	5089.533	-0.22	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
16	114	172	37	950.271	+5	4747.327	4747.307	4.23	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
17	113	181	28	1268.643	+3	3803.914	3803.921	-1.75	[TIKLQIWDTAGAEKFR]	>Rab1b	[EMIYDAFCALGVEMPK]	>DrrA	K14	C8
18	113	136	41	792.057	+6	4747.307	4747.307	0.16	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
19	113	147	57	1018.716	+5	5089.553	5089.533	3.86	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
20	112	209	42	950.266	+5	4747.3	4747.307	-1.49	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
21	112	53	139	1273.139	+4	5089.534	5089.533	0.08	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
22	112	169	33	950.269	+5	4747.317	4747.307	2.23	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
23	111	145	55	1018.713	+5	5089.534	5089.533	0.14	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
24	111	145	48	1018.716	+5	5089.552	5089.533	3.62	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
25	111	133	53	792.058	+6	4747.312	4747.307	1.16	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
26	110	125	28	792.059	+6	4747.315	4747.307	1.78	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
27	110	58	114	1273.137	+4	5089.526	5089.533	-1.55	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
28	110	107	38	849.095	+6	5089.536	5089.533	0.43	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
29	108	61	120	1273.14	+4	5089.54	5089.533	1.23	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
30	108	115	43	1018.714	+5	5089.542	5089.533	1.64	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
31	108	162	71	727.94	+7	5089.539	5089.533	1.16	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
32	106	161	74	849.096	+6	5089.543	5089.533	1.8	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
33	106	128	50	950.269	+5	4747.317	4747.307	2.17	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
34	103	42	44	866.179	+4	3461.694	3461.694	-0.14	[EMIYDAFCALGVEMPK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
35	100	43	55	866.18	+4	3461.7	3461.694	1.63	[EMIYDAFCALGVEMPK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
36	98	143	54	727.94	+7	5089.538	5089.533	1.0	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
37	96	113	57	727.942	+7	5089.548	5089.533	2.93	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[TIKLQIWDTAGAEKFR]	>Rab1b	C8	K14
38	96	81	22	792.06	+6	4747.323	4747.307	3.32	[EMIYDAFCALGVEMPKDMEVHFAGSLAK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11
39	94	35	37	866.18	+4	3461.697	3461.694	0.78	[EMIYDAFCALGVEMPK]	>DrrA	[LQIWDTAGAEKFR]	>Rab1b	C8	K11

Figure 99: Table of the identified crosslinked peptide products obtained by MS^2 analysis of Rab1b-Q67A-R69BrCN7K-DrrA$_{16-352}$-D82C complex.

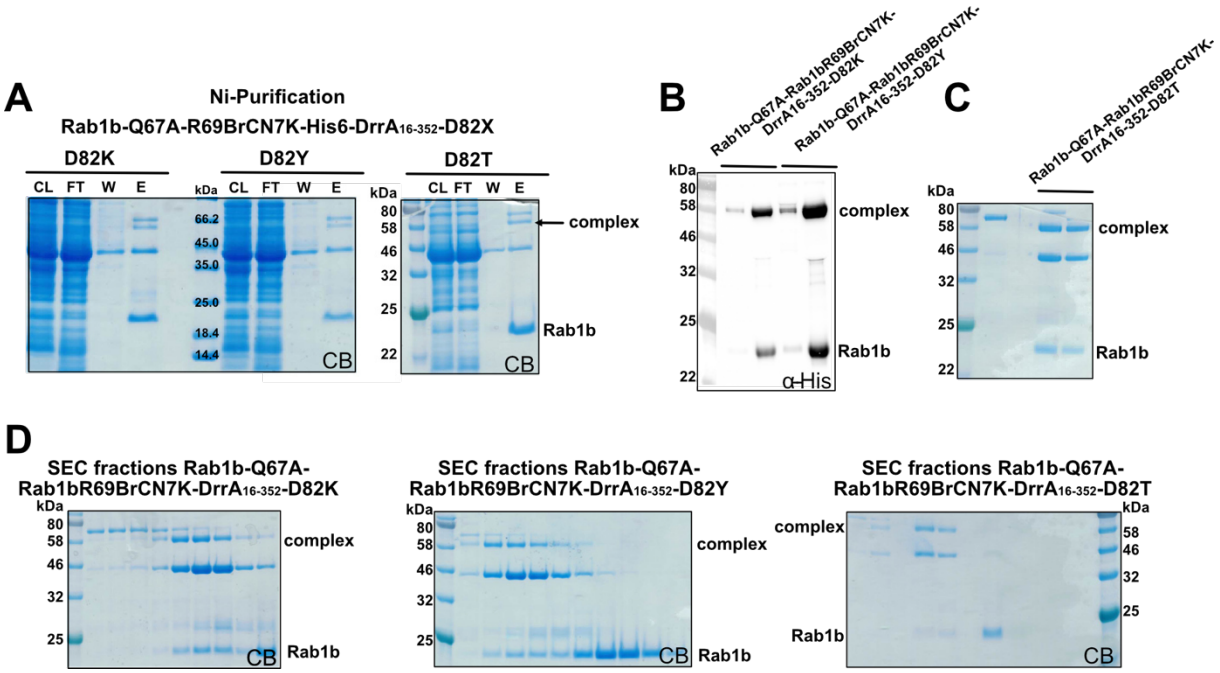
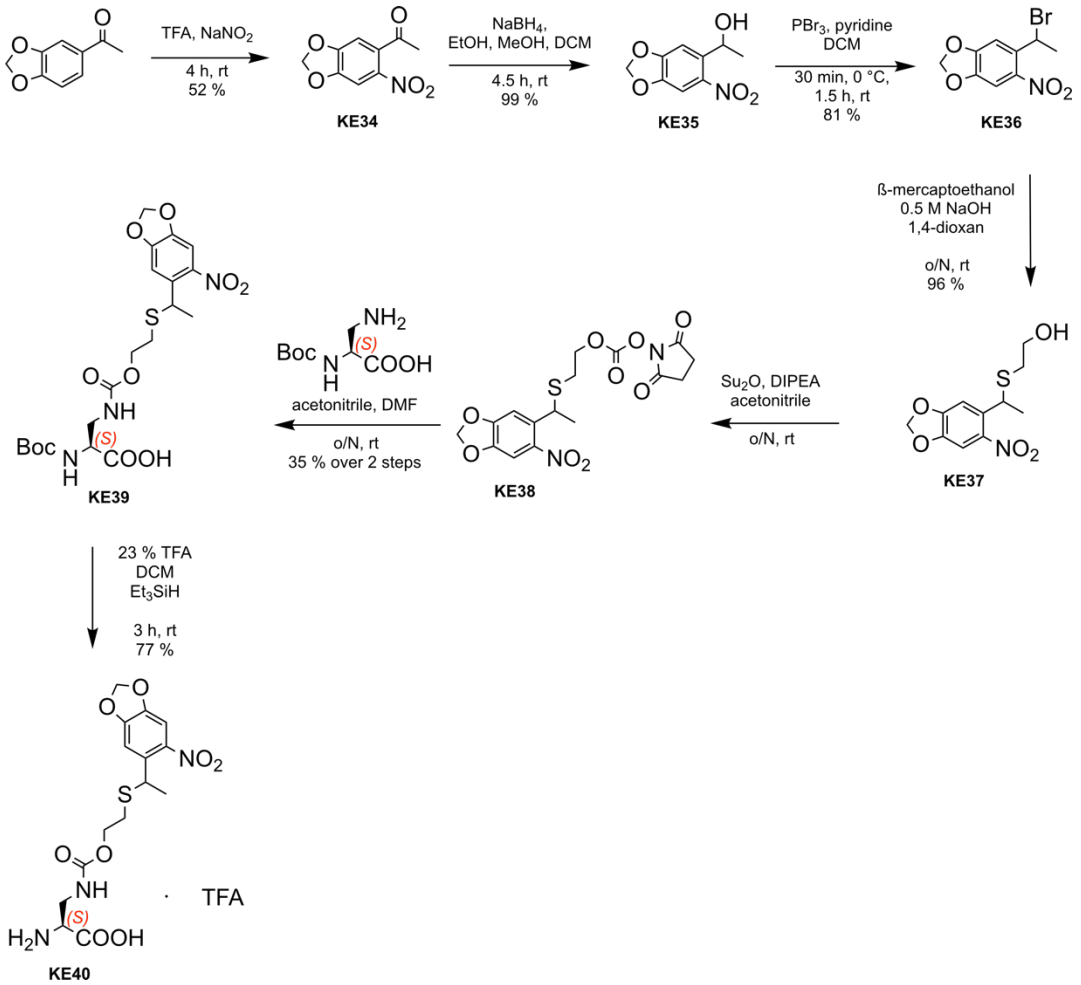


Figure 100: Purification of different Rab1b-Q67A-R69BrCN7K-DrrA$_{16-352}$-D82X complexes. A: SDS gels of the Ni-purifications of different Rab1b-Q67A-R69BrCN7K-DrrA$_{16-352}$-D82X complexes. B: Western blot (α-His) of the elution fractions obtained from Ni-purification of Rab1b-Q67A-R69BrCN7K-DrrA$_{16-352}$-D82K and Rab1b-Q67A-R69BrCN7K-DrrA$_{16-352}$-D82Y complexes. C: SDS gels of the elution fractions obtained from Ni-purification of Rab1b-Q67A-R69BrCN7K-DrrA$_{16-352}$-D82T. D: SDS gels of SEC fractions of different Rab1b-Q67A-R69BrCN7K-DrrA$_{16-352}$-D82X complexes. (X = K, lysine; Y, tyrosine; T, threonine)

7.4.3. Trapping *Sa*ClpP with interactors



Scheme 17 Synthesis of pDAP. This work was done by the collaboration partner, Konstantin Eckl (PhD student), of the group of Prof. Sieber (TUM)

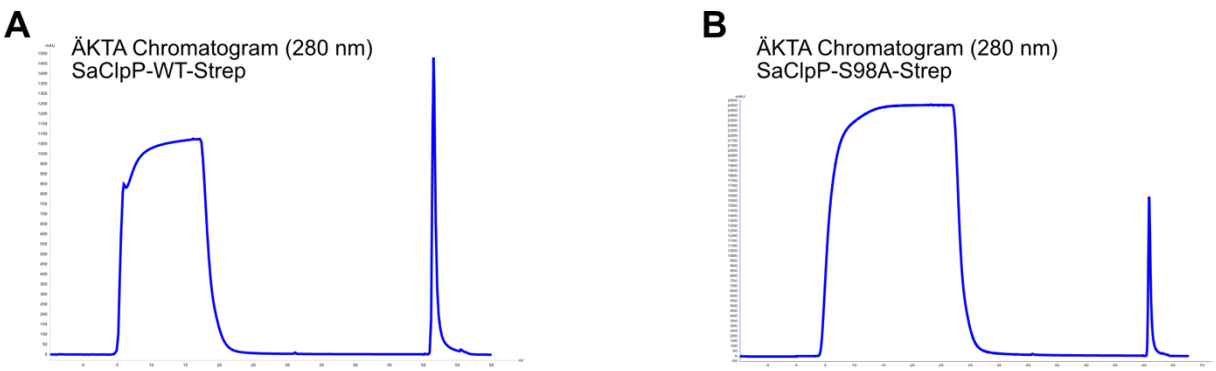
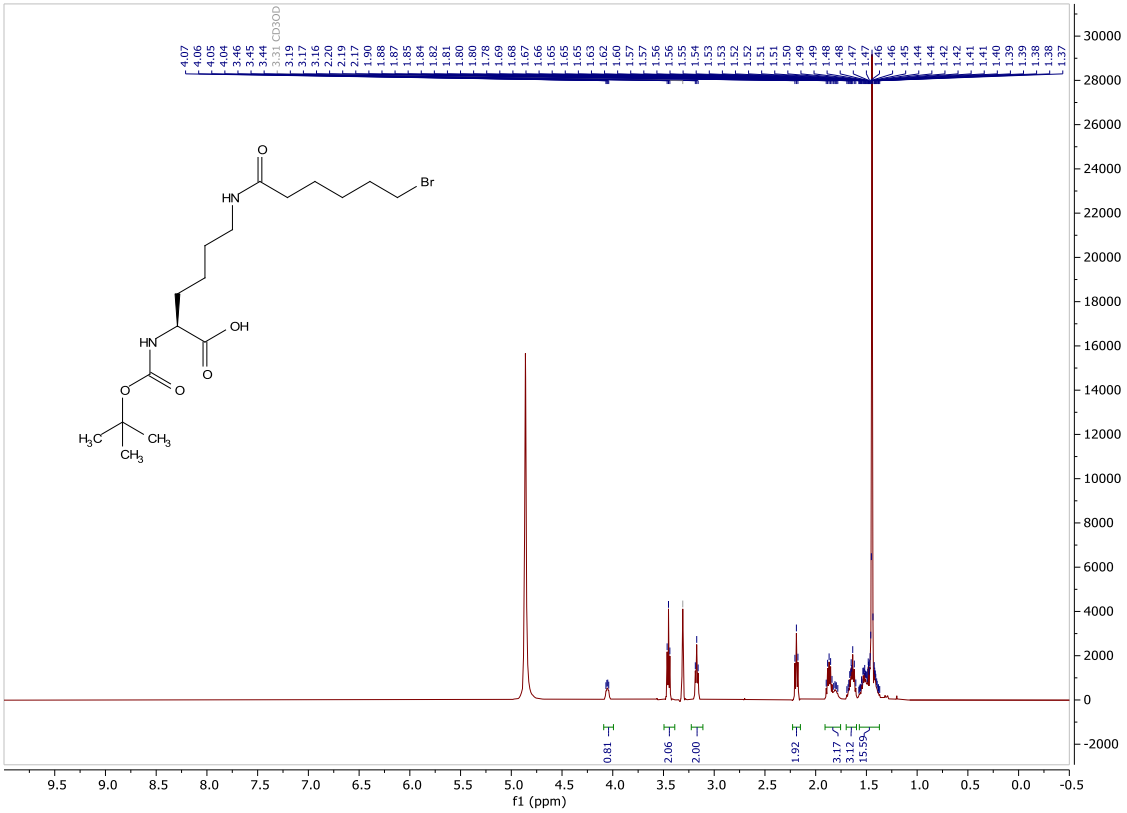


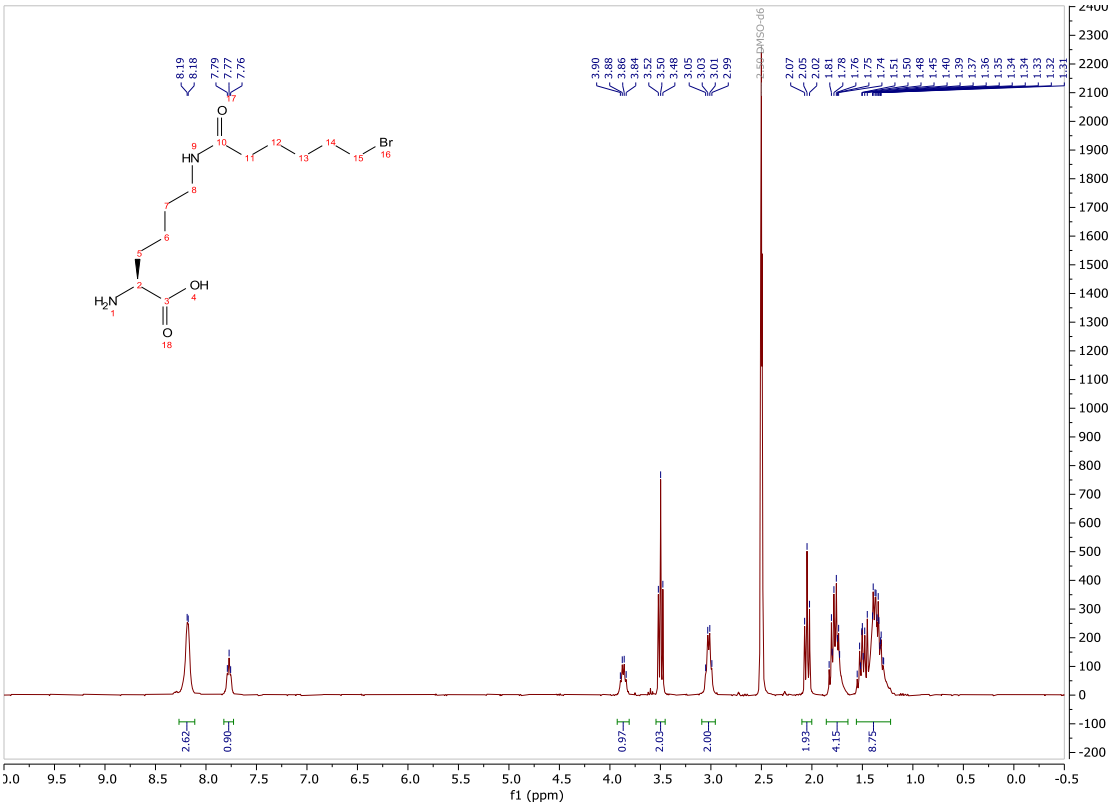
Figure 101: ÄKTA chromatograms (280 nm) of the Strep-purification of A: *Sa*ClpP-Strep and B: *Sa*ClpP-S98A-Strep using ÄKTA Pure.

7.5. NMR spectra

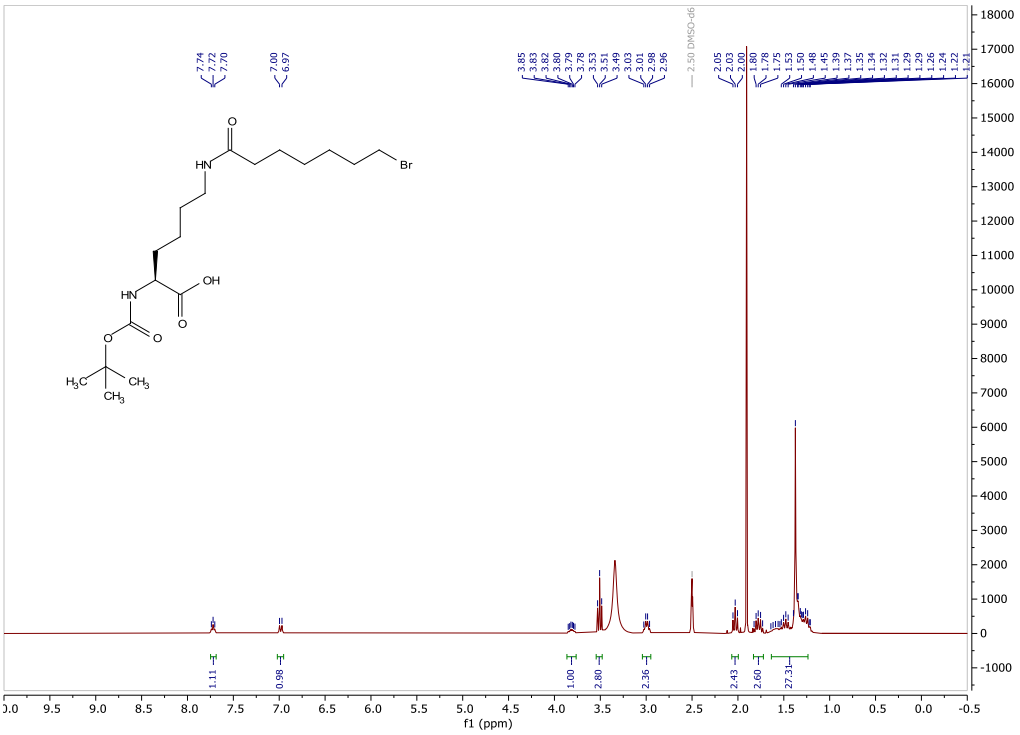
^{1}H-NMR: Boc-BrC6K



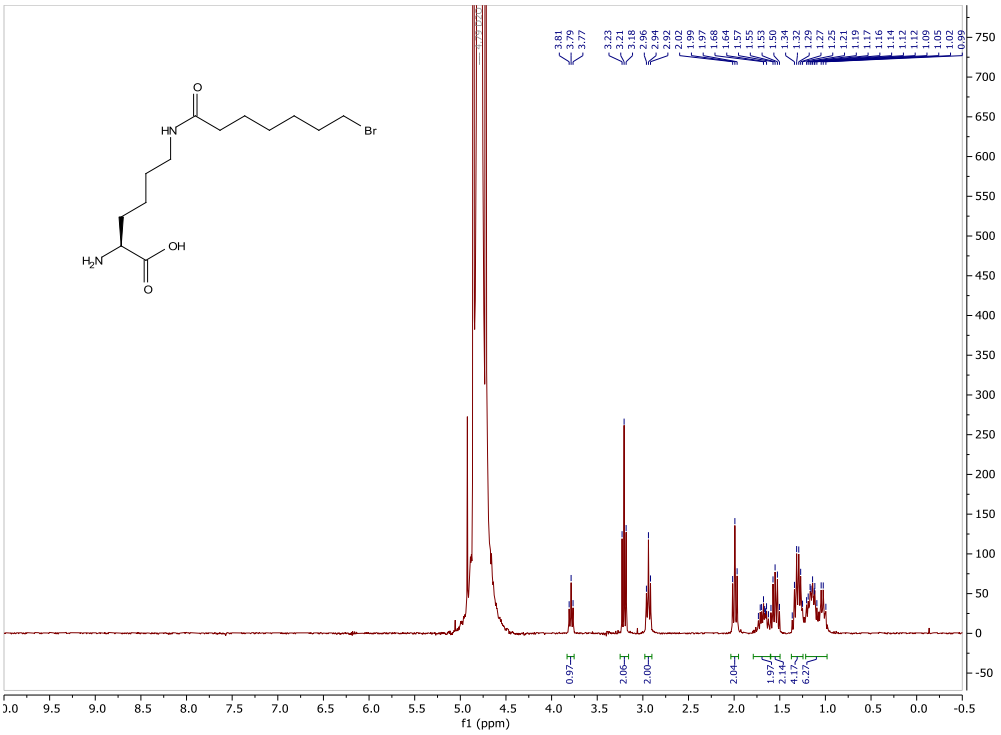
^{1}H-NMR: BrC6K



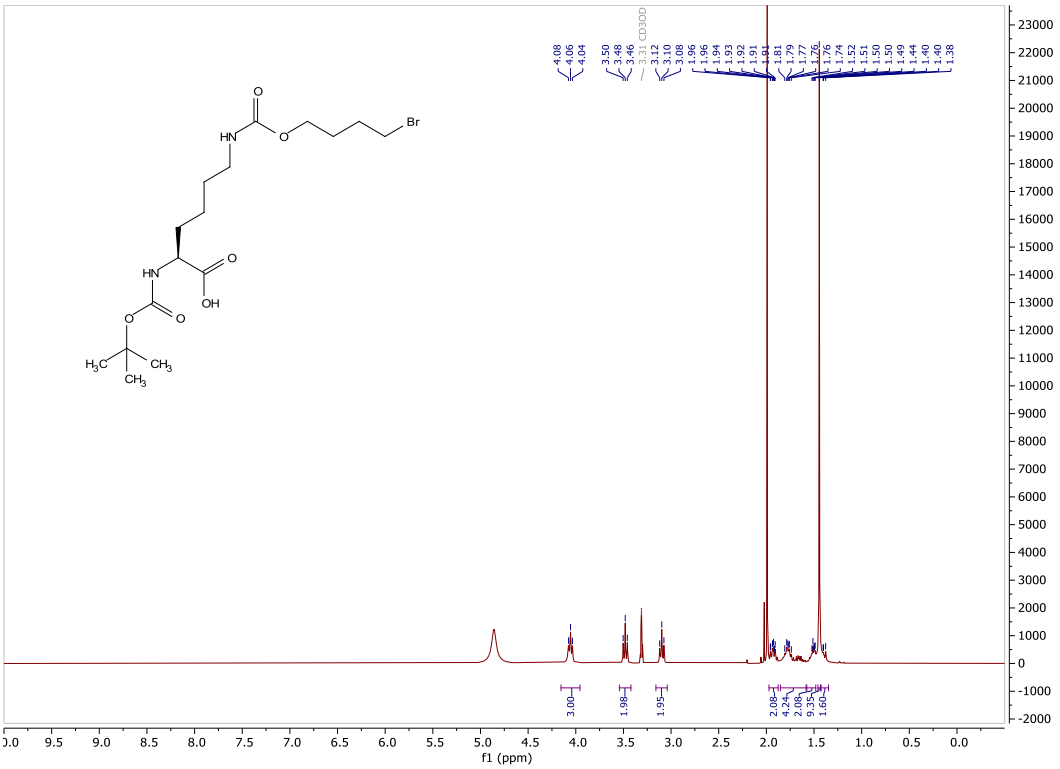
^{1}H-NMR: Boc-BrC7K



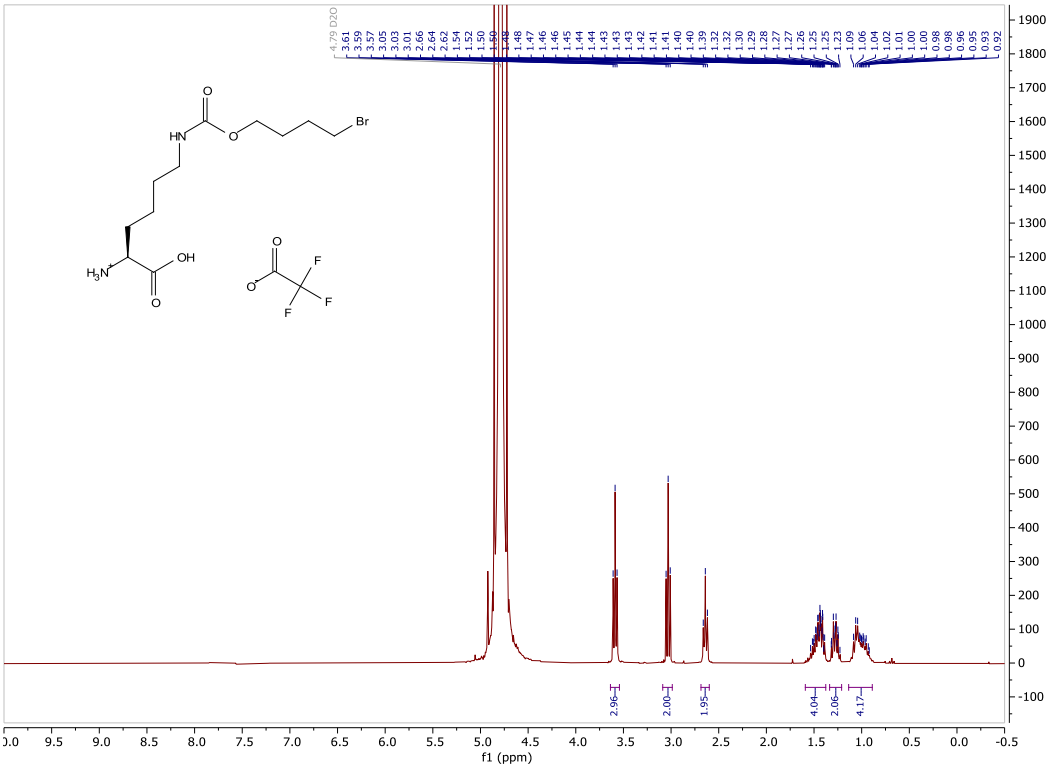
^{1}H-NMR: BrC7K



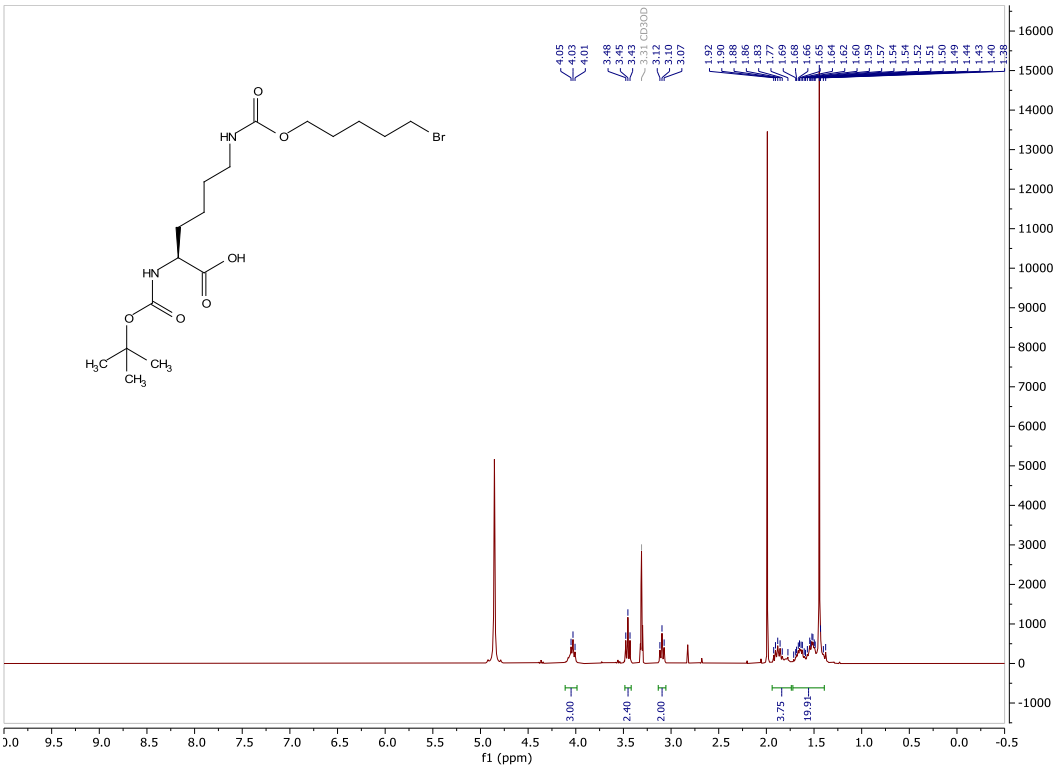
^{1}H-NMR: Boc-BrCO6K



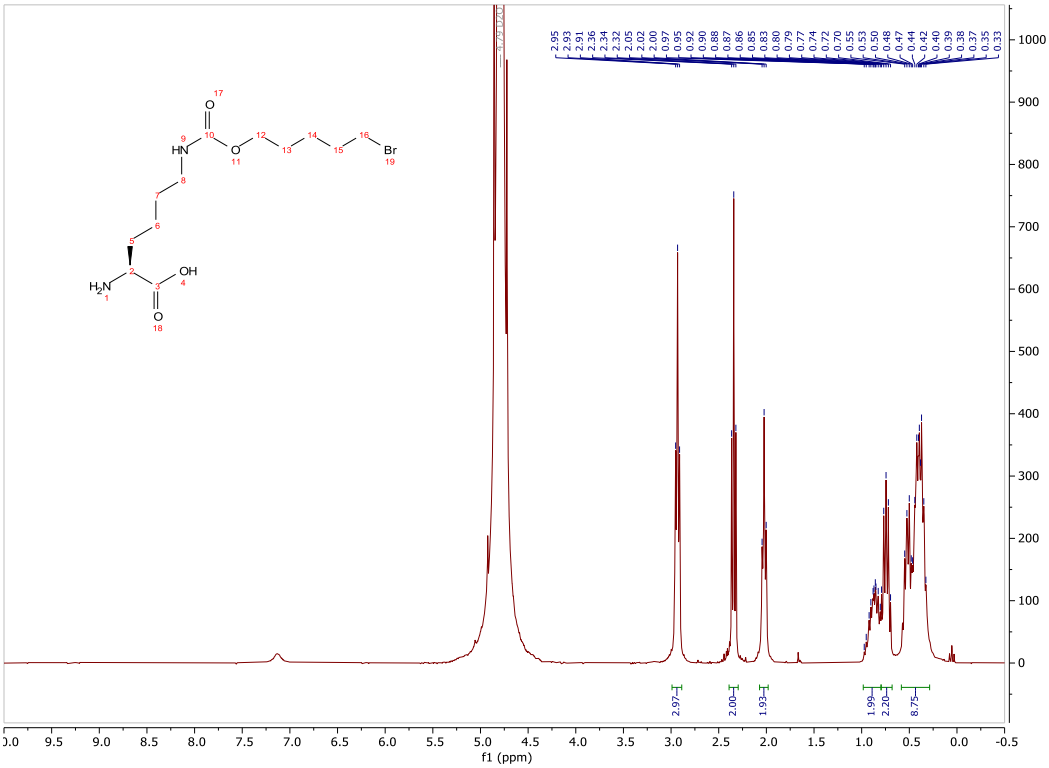
^{1}H-NMR: BrCO6K



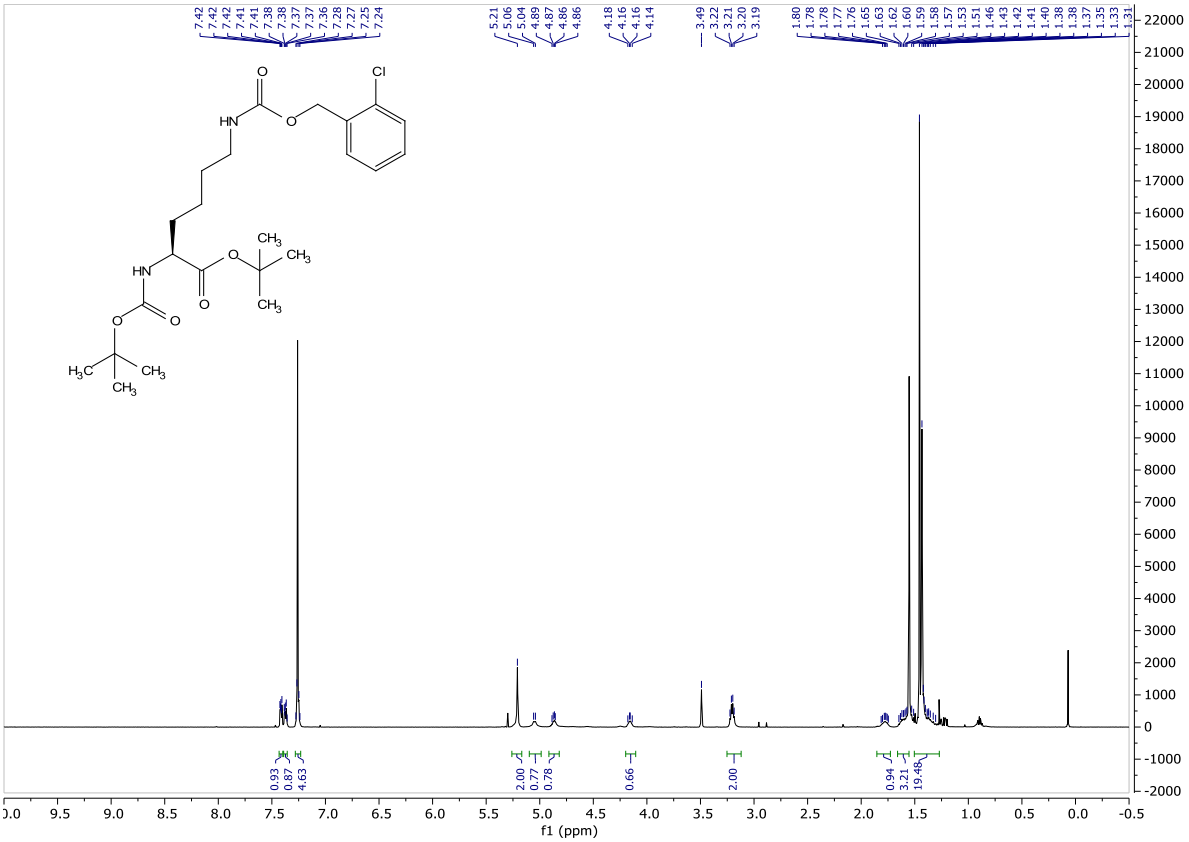
^{1}H-NMR: Boc-BrCO7K



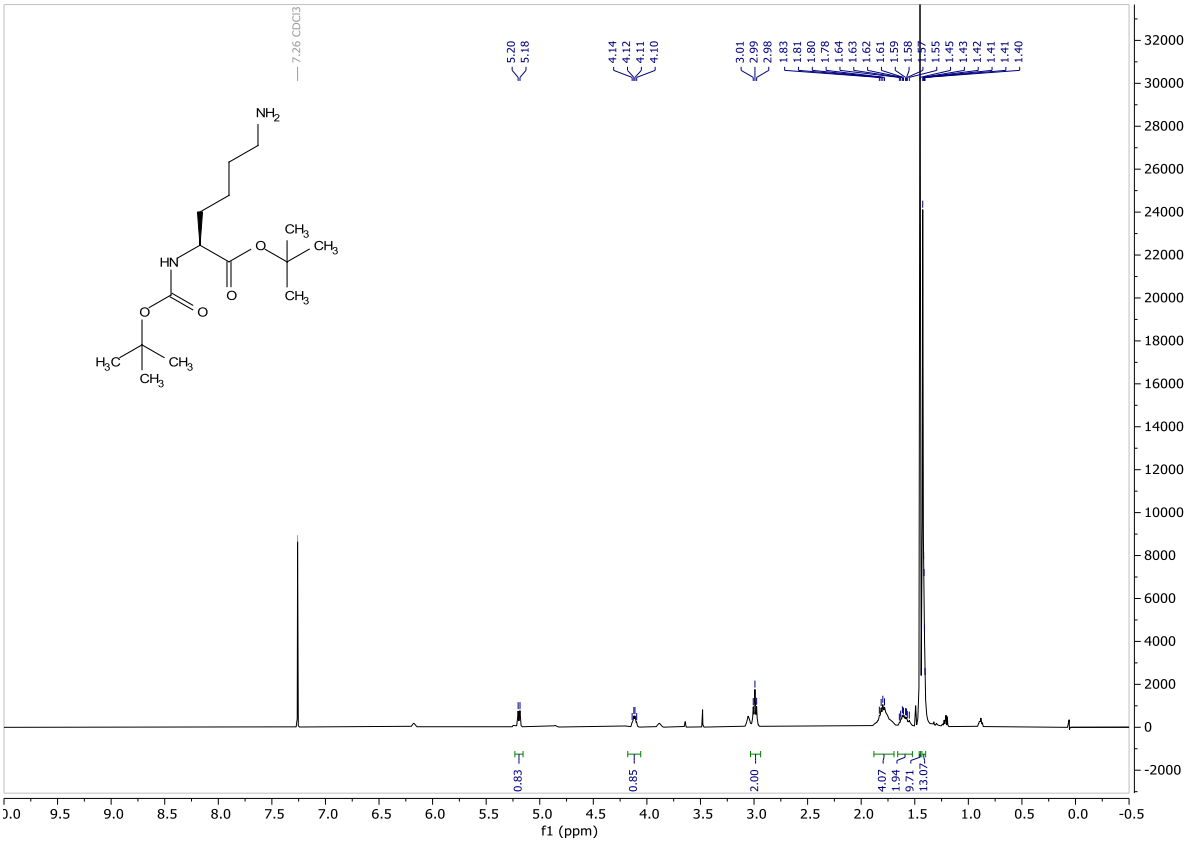
^{1}H-NMR: BrCO7K



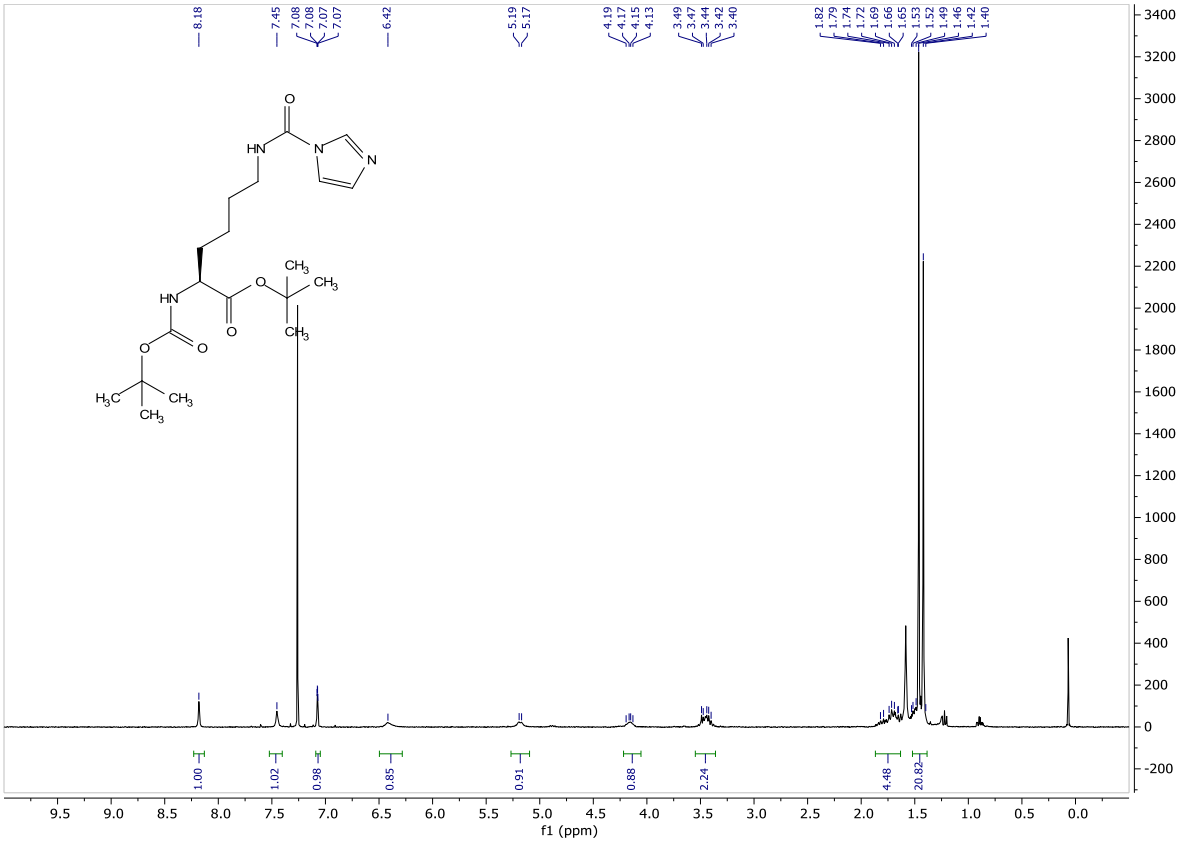
[1]H-NMR: *tert*-butyl esterified Boc-Lys(2-Cl-Cbz)-OH



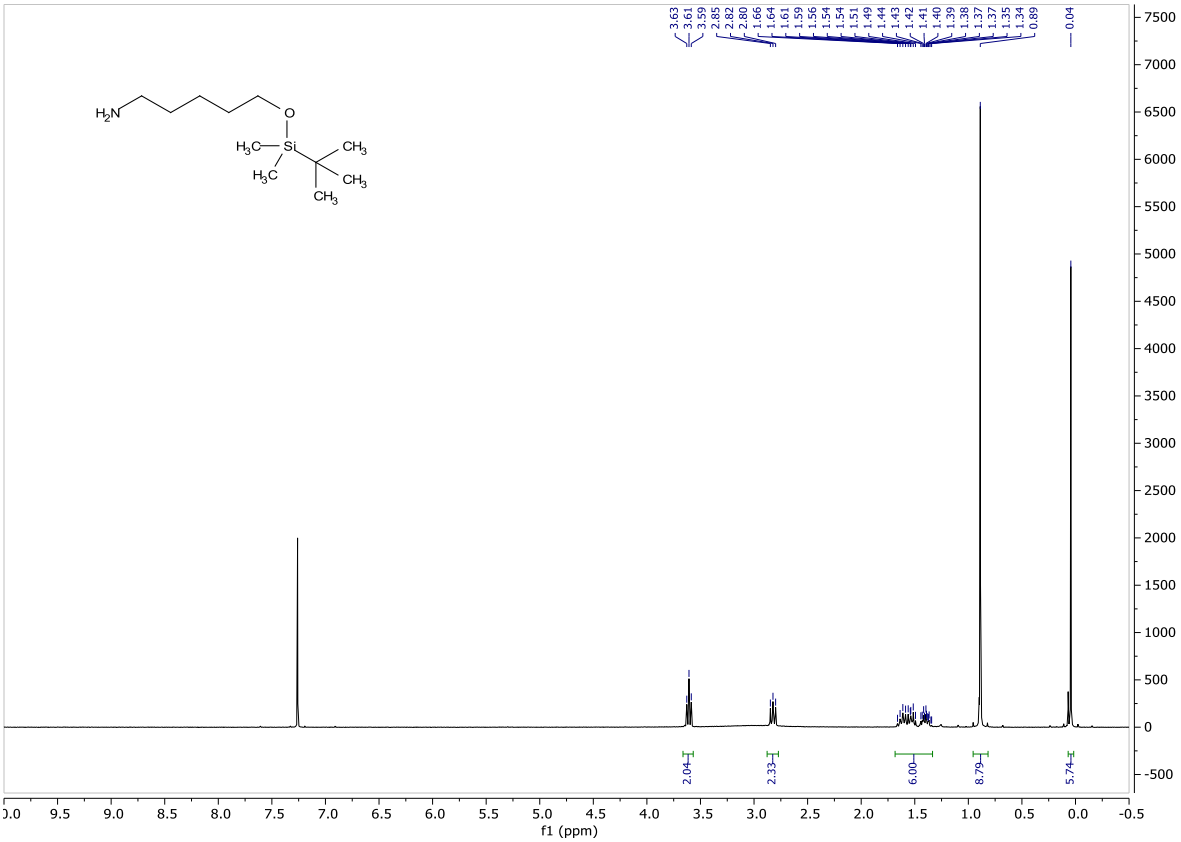
[1]H-NMR: *tert*-butyl esterified Boc-Lys-OH



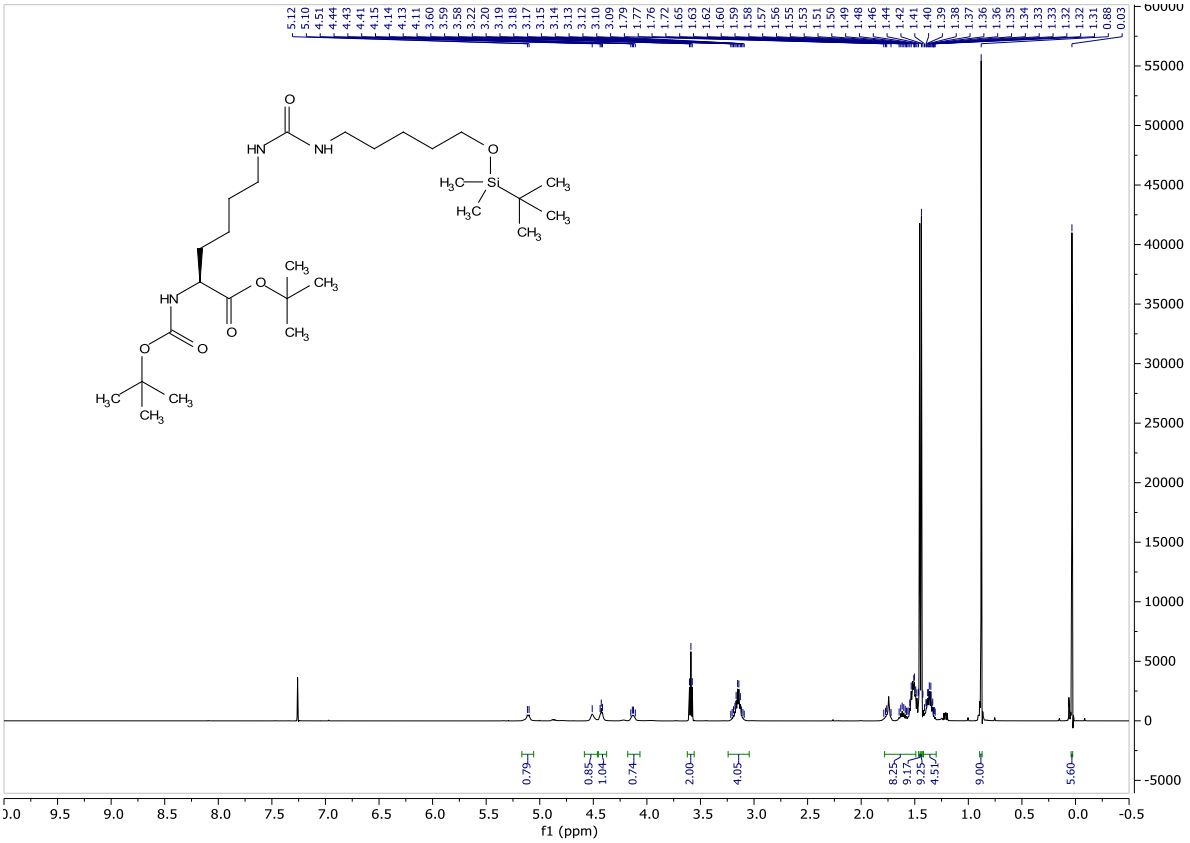
^{1}H-NMR: *tert*-butyl N^{2}-(*tert*-butoxycarbonyl)-N^{6}-(1*H*-pyrrole-1-carbonyl)-*L*-lysinate



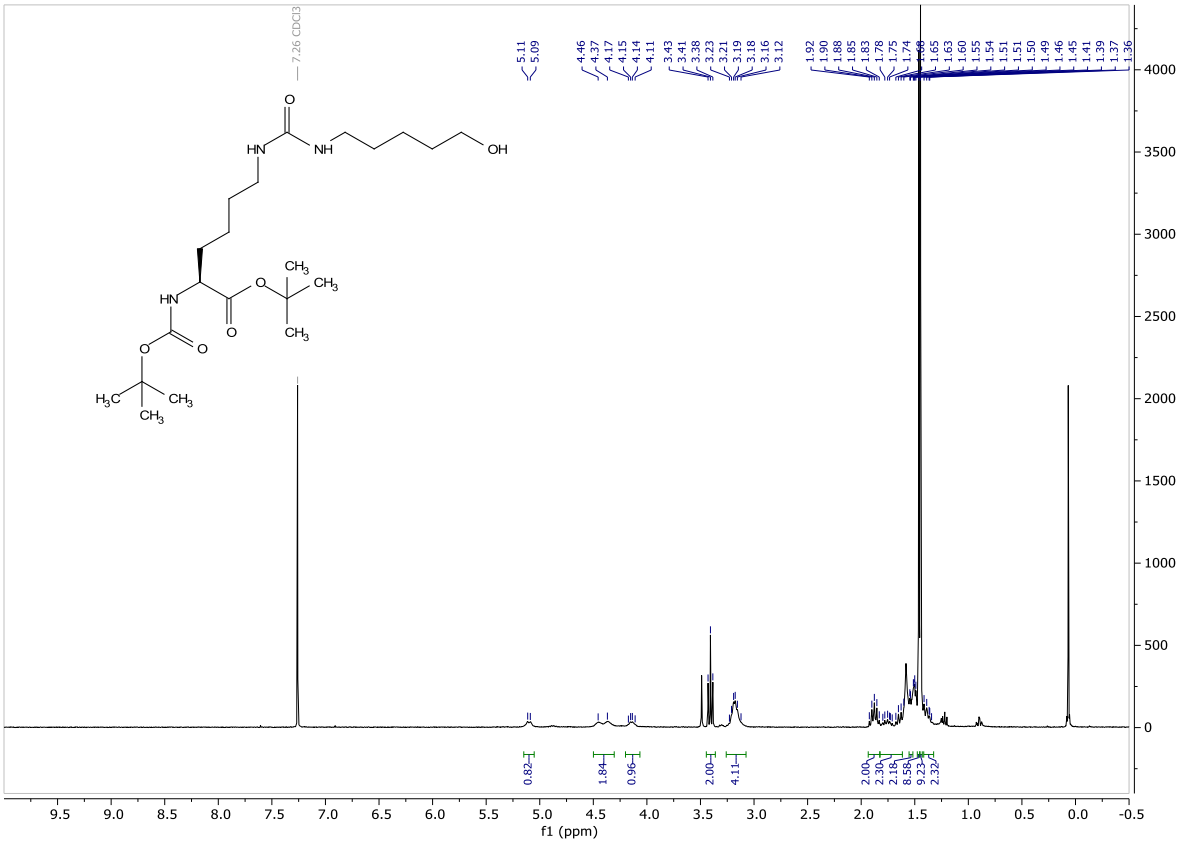
^{1}H-NMR: TBDMS protected 5-amino-1-pentanol



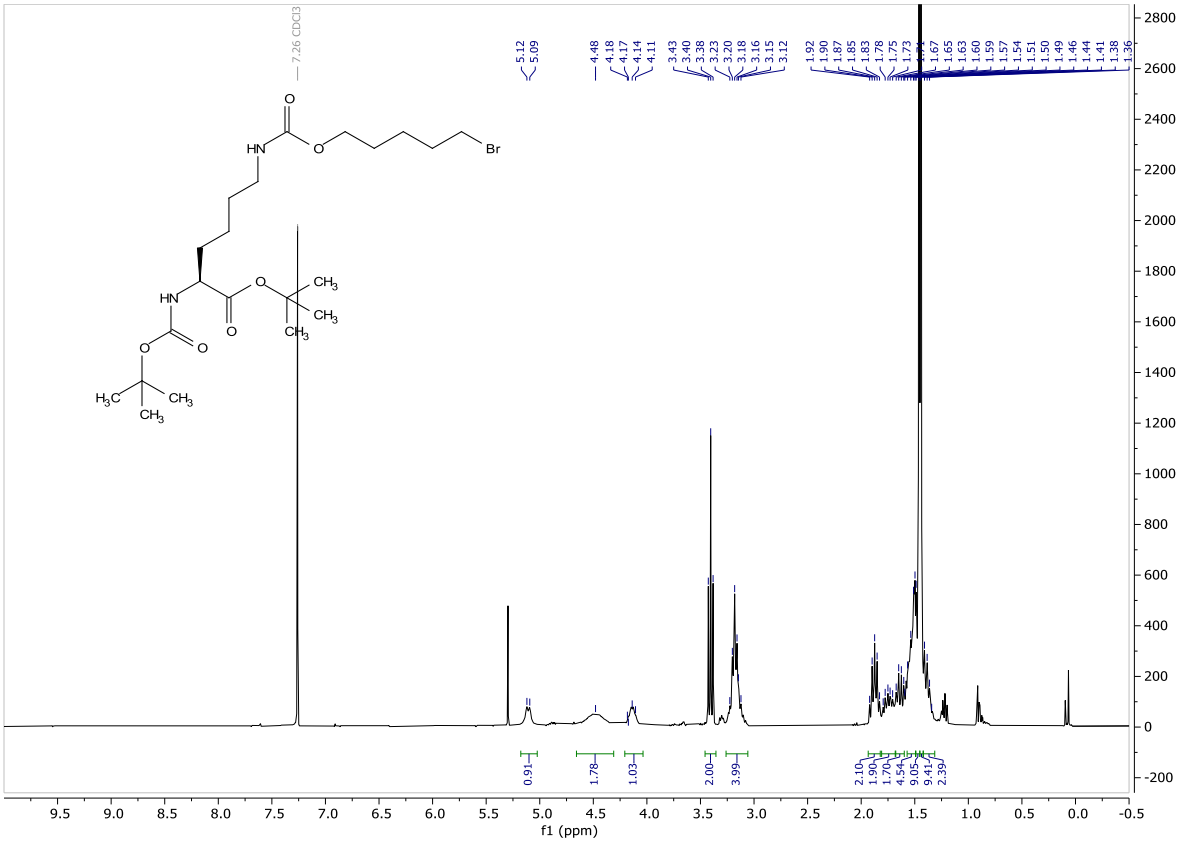
^{1}H-NMR: *tert*-butyl N^{2}-(*tert*-butoxycarbonyl)-N^{6}-((5-((*tert*-butyldimethylsilyl)oxy)pentyl)carbamo-yl)-*L*-lysinate



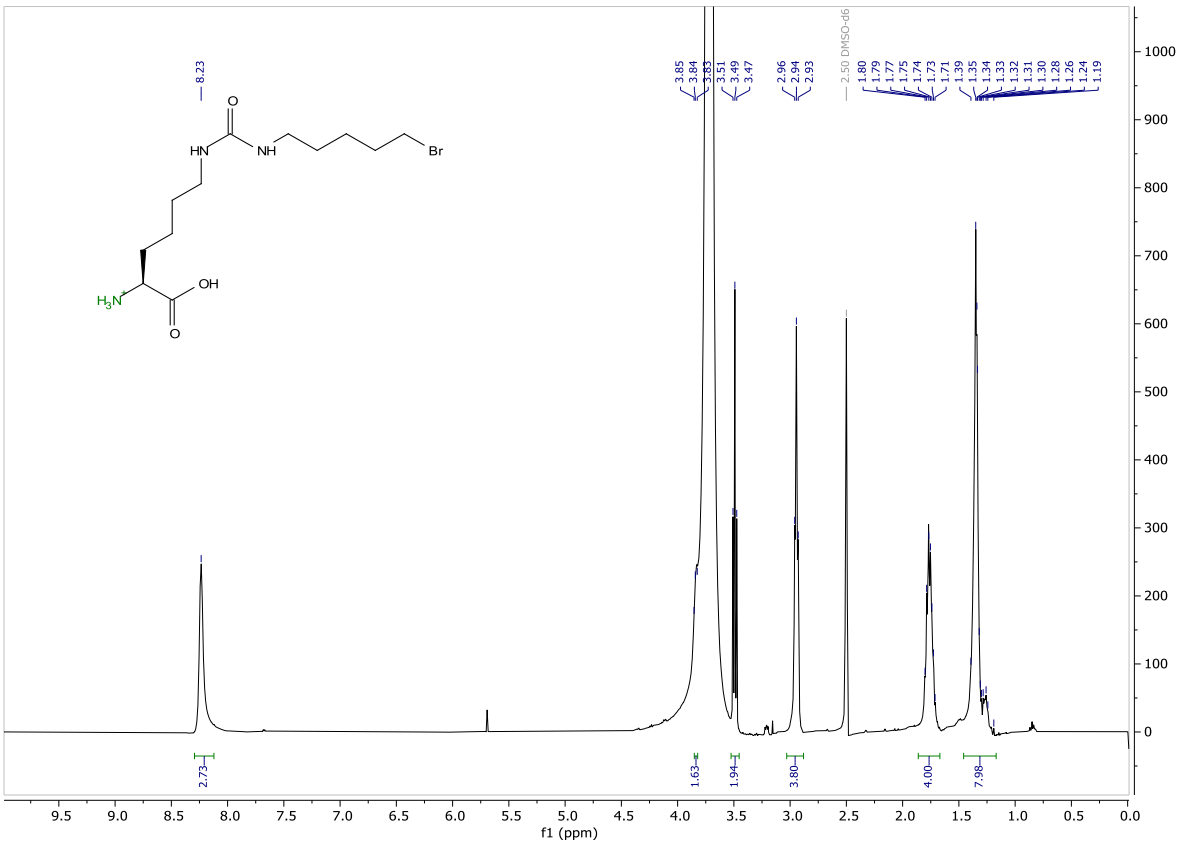
^{1}H-NMR: *tert*-butyl N^{2}-(*tert*-butoxycarbonyl)-N^{6}-(5-hydroxypentyl)carbamoyl)-*L*-lysinate



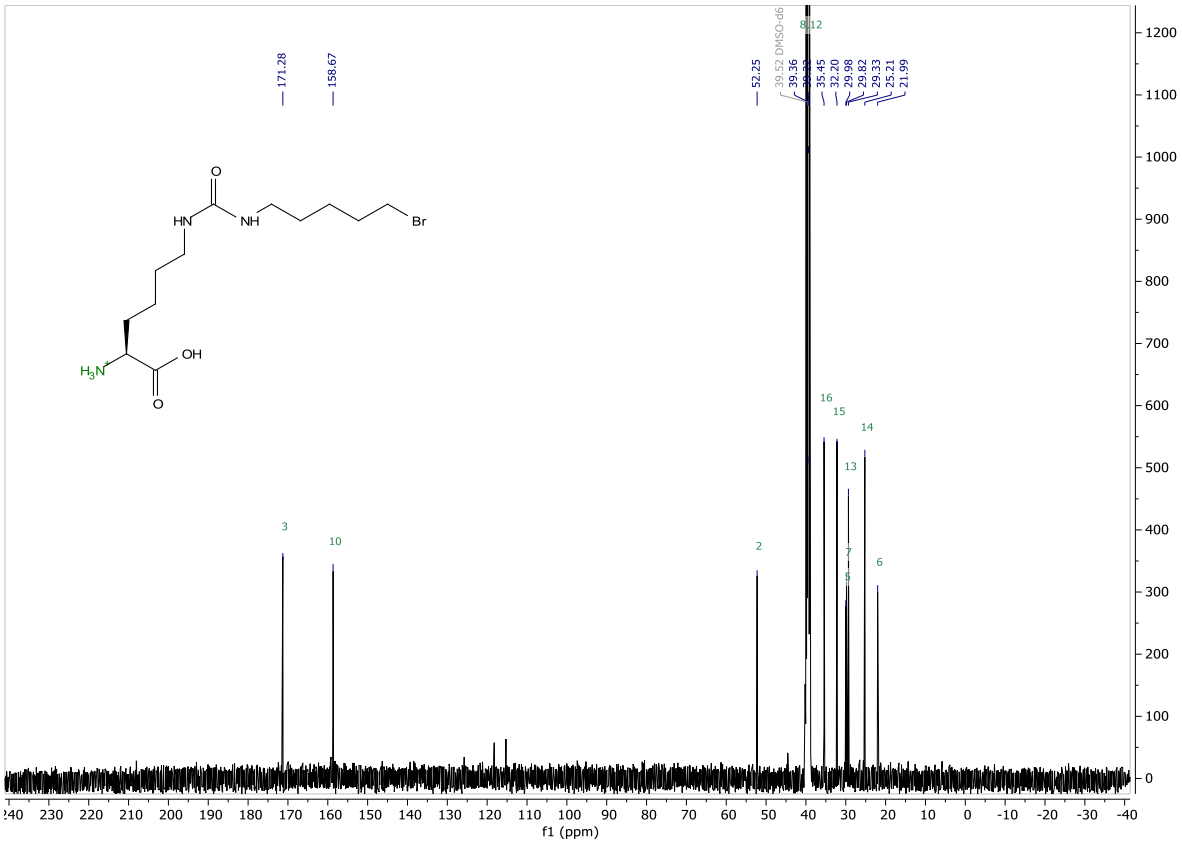
^{1}H-NMR: protected N^{6}-((5-bromopentyl)carbamoyl)-*L*-lysine (BrCN7K)



^{1}H-NMR: BrCN7K



^{13}C-NMR: BrCN7K



EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich eidesstattlich, dass ich die vorliegende Arbeit selbstständig angefertigt und keine anderen als die angegebenen Quellen oder Hilfsmittel verwendet habe. Alle in dieser Arbeit sinngemäß oder wortwörtlich übernommenen Stellen habe ich gekennzeichnet. Diese Arbeit wurde für keinen anderen akademischen Grad eingereicht wie angegeben.

....

Ort, Datum Unterschrift

ACKNOWLEDGEMENT - DANKSAGUNG

Es gibt noch so viel zu sagen, denn in den vier Jahren der Doktorarbeit ist unglaublich viel passiert. Tolle, erfolgreiche Momente wurden begleitet von so mancher Enttäuschung und Rückschlag. Das wäre, ohne eine Reihe von vielen besonderen Menschen, nicht zu schaffen gewesen, die mich während dieser Zeit begleitet und unterstützt haben und denen ich nun endlich dafür von ganzem Herzen danken möchte.

Beginnen möchte ich mit dir Kathrin. Vielen Dank, dass du mich im April 2016 in deine Gruppe aufgenommen hast –ohne dass es ein Aprilscherz war. Du hast den Weg zu dieser, meiner Doktorarbeit bereitet, indem du mir die Möglichkeit gegeben hast in deiner Gruppe forschen, lernen und sich entwickeln zu dürfen. Es war prägend mitzuerleben, wie eine junge Gruppe wächst und sich wandelt. In allen Lagen hattest du ein offenes Ohr, gabst Unterstützung und konstruktive Kritik. Danke für alles! Ich werde die Zeit nie vergessen und oft vermissen.
Als nächstes zu den einzigartigen Kollegen, die gleichzeitig auch zu Freunden geworden sind. Ihr seid das Herz der Arbeitsgruppe und habt den Arbeitsalltag immer zu einer wahren Freude gemacht - nicht nur wegen Cake o'clock und den KLangBang-Aktionen. Es war toll zu sehen, dass obwohl wir alle unterschiedlich sind, in Summe ein super Team gebildet haben, dass durch jeden einzelnen mit seinen Fähigkeiten bereichert wurde.

Liebste Susanne! Beim ersten Kennenlernen an meinen Bewerbungstag war ich schwer beeindruckt von den wunderschönen, pinken Tetrazinen, die überall in deinen Kolben, Säulen etc. verteilt waren. Aber schon nach ganz kurzer Zeit (wahrscheinlich ab dem ersten Arbeitstag^^) war ich beeindruckt von dir! Du bist eine ganz unglaublich bemerkenswerte Persönlichkeit und in vieler Hinsicht für mich ein Vorbild. Ich habe so viel von dir gelernt, arbeitstechnisch (du und Marko musstet mich ja als Neuling erstmal auf Kurs bringen) und auch menschlich und möchte nichts missen. Du bist uns bleibst mein liebster Mittagspausen-, Kaffeetrink-, Plätzchenback-, Chemie-, Spieleabend-, Konzert- und Ausgeh-Partner! Ich bin unendlich dankbar, dass sich unsere Wege gekreuzt haben und von da an (und noch weiter) gemeinsam verlaufen, sodass wir schon die ein oder andere Lebenslage zusammen durchgestanden haben. Man kann sich immer auf dich verlassen – durch Dick und Dünn – und unglaublich viel Spaß haben (aber besser ohne Zündquellen und entflammbare Gegenstände ^^) - eine wahre Freundin! DANKE!
Ach Marko! Was soll man über unsere Beziehung sagen?! Toni würde sie vielleicht auf All-You-Yan-Eat Running Sushi reduzieren, aber was Toni sagt ist ja eh egal^^ (Sorry Toni, nicht so gemeint). Aber Spaß bei Seite (auch wenn man den mit dir immer hat), ich möchte dir von ganzem Herzen danke sagen für alles, was wir in den letzten vier Jahren zusammen erlebt haben. Danke für die Hilfe beim Start in unserer Gruppe, danke für die ganzen chemischen Arbeiten, danke für die U-Boot-Projekte, danke für die Stunden „Spaß an Chemie mit Marko", danke für die Candy-Shop-Plünderungen und Süßigkeiten-Special-Editions, danke für die Gespräche, danke für die gemeinsamen Erlebnisse, danke für die gemeinsamen Feierabendbiere/Aperol Spritz, danke für die Kaffeepausen beim Griechen, danke für die Aufmunterungen und Aufmerksamkeiten etc... Legendär und nie vergessen bleiben unsere Fujikaiten-Besuche oder Vapiano im ZOB^^. Auch deine Wohnortsverlagerung schützt dich

nicht vor einem regelmäßigen Revival. Du bist eine Bereicherung für mein Leben und hast immer einen festen Platz in meinem Herzen.
Lieber Toni, natürlich ist nicht egal was du sagst! Du bist ein Hauptgewinn für uns in der Gruppe und das nicht nur, weil du chemisch so einiges auf dem Kasten hast, sondern auch weil du ein toller Mensch bist. Nach Marko bist du mein Chemie-Partner 2.0 geworden und hast die Marko-Lücke bestmöglich ausgefüllt – ebenbürtig! Ich bin dir unendlich dankbar für deine ganze Arbeit und Mühe und hoffe ich kann mich irgendwann revangieren. Mir ist es noch nicht gelungen dich davon zu überzeugen, dass Bio nicht nur Abfall ist, aber die Hoffnung lebt weiter. Hut ab, wie du dich für die Arbeitsgruppe einsetzt und gemeinsame Aktionen initiierst und organisierst. Das Teamfeeling lebt zum Großteil durch dein Engagement. Danke für die Weißwurstfrühstücke und die Krapfen-Runden und DICH, ich werde alles vermissen!
Kristina: was kann man sagen - eine tolle Chemikerin, starke Persönlichkeit, würdiger Trinkpartner, SFB1035-Retreat-Buddy und stets aufmerksamer Zuhörer. Wir haben so einiges erlebt, Tränen, Lachen – alles war dabei. Angespannte Situationen gab es viele, aber was zählt ist, dass wir uns aufeinander verlassen konnten und immer jedes Problem mit vereinten Kräften bewältigt haben. Ohne dich, deine Organisation und Hilfe würde die Chemie-Seite bei weitem nicht so gut laufen. Dein Gespür für Top-Playlists haben den Laboralltag und auch Feierlichkeiten bereichert und die Stimmung exorbitant gehoben – ich hoffe auf viele weitere musikalische Highlights.
Jetzt zu dir Vera: ich hatte die Hoffnung schon aufgegeben, Gesellschaft und in der Bio-Ecke des 6ten Stocks zu bekommen, aber dann kamst zum Glück du. Keiner ist ein besserer HEK-Zell-Flüsterer wie du und beim Thema Lab-Buddies macht dir so schnell keiner was vor. Deine kreative Gestaltung der Laborbuch-Einbände hat mich oft zum Lachen gebracht, danke für die schöne Zeit.
Anh, Max, und Marie-Lena, es war mir eine Ehre euch beim TUM-Campus-Lauf anzufeuern! Ich werde mich daran erinnern, wie ich morgens das türkise Fahrrad von MJ im Gang stehen gesehen habe und mir gleichzeitig dachte: „Oh man, sie ist so diszipliniert und sportlich, und das schon am Morgen. - krass!“. Aber nicht nur da bist du motiviert, nein, du hast dich auch um die (bitter nötige) Einrichtung der KLangLab-Website gekümmert und unsere Excel-Listen gehegt und gepflegt. Vielen Dank auch für die vielen Male am Mikroskop. Du bist über das IMPRES zu uns gekommen und seitdem weiß ich, dass dort nicht nur „komische“ Leute sind, sondern richtig prima Menschen! Anh, du Biochemiker mit etwas weniger Bio, ich hätte es nicht für möglich gehalten, dass ein Biochemiker so ein großes Wissen in Synthese haben kann, aber du hast mich eines Besseren belehrt. Unvergesslich sind auch deine „Asia-Boxen“ mit weiß Gott was für Lebensmitteln drin und deine Kuchen (besonderes der Karottenkuchen!), die waren einfach unfassbar gut. Danke, für die unterhaltsamen und interessanten U-Bahn-Fahrten zwischen Münchner Freiheit-Garching, die gefüllt mit allerhand (auch kritischen) Themen waren, über die ich vielleicht bis dato noch gar nicht nachgedacht hatte. Finally, Max, Master des Ubiquitin: Chapeau, dass du wirklich den Campus-Lauf durchgezogen (hinbekommen) hast! Aber „Hinbekommen“ tust du sowieso eigentlich alles. Noch nie habe ich einen Wissenschaftler wie dich getroffen. Kein „Bier vor vier“ (oder ähnliches) kann dein Wissen, deinen Eifer und deine Ideen bremsen. Du hast mir oft mit Ratschlägen und beim Brainstorming beiseite gestanden, vielen Dank!
Vielen Dank auch an unsere guten Seelen, die Auszubildenden (Tanja, Steffi) und TA Caro. Danke, Tanja und Steffi, dass ihr das Vertrauen hattet mir einen Teil eurer Ausbildung zu

überlassen. Das hat mir viel Freude gemacht und auch mich viel dazulernen lassen. Ihr alle habt so viel Unterstützung im Laboralltag geleistet – DANKE.

Danke auch an die anderen Arbeitsgruppen der großen Chemie-Fakultätsfamilie. Insbesondere an Astrid, die Mutter der Biochemie-Familie, du leistest so viel, bist stark, bist Problemlöser, bist Seelentröster und Mittagessen-Versorger. Danke auch an Bettina, die Managerin für alle NEB-Angelegenheiten.
Liebe Sabine, du und deine Gruppe ihr seid nicht wegzudenken. Immer hilfsbereit, geduldig und freundlich. Es war mir eine Ehre, die ersten Schritte im Bereich Proteinkristallisation mit dir gehen zu dürfen. Ich freue mich auf alles was kommt!
Ein herzliches Dankeschön gilt auch der Sieber-Gruppe (und ihren ehemaligen Mitgliedern). Matthias, ohne deine Unterstützung beim Thema MS/MS hätte sich einiges niemals so entwickelt, wie es jetzt gekommen ist. Wie du es meisterst Beruf und Familie unter einen Hut zu bekommen ist grandios. Du, mit deinem außergewöhnlichen Wissen, nicht nur im Bereich Proteomics, deinem Engagement und deiner tollen Persönlichkeit, bist eine unschätzbare Bereicherung für jedes Team. Liebe Anja, vielen Dank auch dir. Dafür, dass du immer ein toller Zuhörer warst, für dein positives Wesen, uns allerhand Proben gemessen und uns beraten hast. Und auch danke an Konstantin, es war toll dich als Praktikant in der Klang-Gruppe kennenlernen zu dürfen und später gemeinsam mit dir an *Sa*ClpP-pDAP zu arbeiten.
Und natürlich möchte ich auch ein Danke an die gesamte Gruppe von Tobias Gulder richten, mit dir wir eine lange Zeit das Labor geteilt und uns gegenseitig unterstützt haben.

Im Leben gibt es aber (zum Glück) nicht nur die Arbeitswelt, sondern auch das Privatleben mit Familie und Freunden.
Mama und Papa, jetzt ist es geschafft! Ein „Danke“ reicht gar nicht aus, um all das zu würdigen, was ihr für mich macht und tut seit 29 Jahren. Auch wenn ihr immer sagt: „wir verstehen leider sowieso nicht wirklich, was du da machst…“, habt ihr mehr dazu beigetragen mein Studium und die Promotion zu schaffen, als ihr glaubt. Egal, was ich machen wollte, ich hatte eure Unterstützung. Ihr haltet einem den Rücken frei und gebt mir ein Zuhause! Ich liebe euch!
Matthias, liebstes Bruderherz, danke, dass du immer da bist – Blut ist dicker als Wasser - Forever and ever and ever ^^!
Christoph, Irmgard und Alfons: auf euch kann ich immer zählen, egal in welcher Lebenslage. Auch hier habt mich immer bestärkt, gefördert und unterstützt, indem was ich getan habe. Danke!
Annika, Martin und mein süßer Theo, ihr seid auch ein Münchner Ersatz-Zuhause geworden. Angefangen beim der Waschmaschinen-Benutzung, der Fahrrad-Reparatur, Bohrarbeiten bis hin zum Kochen und Reden. Theo, du und dein Lachen haben noch immer jeden schlechten Tag zu einem guten gemacht. Ich weiß ich kann immer zu euch kommen. Herzlichen Dank!

Zuletzt möchte ich noch ein Dankschön an Carina und Fabi richten. Ihr zwei begleitet mich nun schon von Beginn des Studiums an auf meinem Weg. Carina: beste Praktikumspartnerin, Meister-Köchin, Word-Specialist (ohne dein wiederholtes geduldiges Zutun, wäre ich an den Gliederungen und Formatierungen verzweifelt) und natürlich außergewöhnliche Freundin. Fabi: Freund und absolutes Super-Brain, aber dabei so bescheiden und stets darauf aus, dein Wissen zu teilen, andere teilhaben zu lassen und zu helfen. Bleibt beide einfach genau wie ihr seid – bei euch zu sein ist immer ein Stück wie Zuhause. Danke für alles!