Engineering protein biosynthesis apparatus, advanced design and screening strategies for small and fluorinated substrates in orthogonal translation

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat)

submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Tuyet Mai Thi To

M. Sc. Chem

Berlin, May 2022

1st reviewer: Prof. Dr. Beate Koksch (Freie Universität Berlin)

2nd reviewer: Prof. Dr. Nediljko Budisa (University of Manitoba)

Date of thesis defense: 30th November 2022

COGITO, ERGO SUM.

René Descartes

Danksagung

An erster Stelle möchte ich meinen zwei Betreuern Frau Prof. Dr. Beate Koksch und Herrn Prof. Dr. Nediljko Budisa für die Aufnahme in ihre Arbeitsgruppen sowie die hervorragende und wissenschaftliche Betreuung bei der Durchführung meiner gesamten Arbeit danken. Ich danke euch sowohl für das spannende und anspruchsvolle Thema als auch für eurer Verständnis und Vertrauen in guten sowie schlechten Zeiten, wenn etwas nicht auf Anhieb funktioniert hat.

Außerdem möchte ich mich bei Herrn Prof. Dr. Thomas Friedrich bedanken, für die gute Kooperation bezüglich der prolin-basierten Untersuchung in verschiedenen fluoreszierenden Proteinen.

Für die Bereitstellung der fluorierten Aminosäure TfeGly und des Kristallisationsliganden 2AA-Abu danke ich Dr. Johann Moschner und Suvrat Chowdhary.

Dr. Christian Roth danke ich für die Unterstützung bei den Kristallisationsexperimenten und die fachlichen Anregungen zu meiner allgemeinen Arbeit.

Bei meinen Arbeitsgruppen AG Koksch, AG Biocat und den Aluminis bedanke ich mich für das angenehme Arbeitsklima und den produktiven Austausch von Informationen.

Der besondere Dank geht an Dr. Jan-Stefan Völler. Danke, dass du mich bei all meiner Arbeiten unterstützt und an mich geglaubt hast. Ohne dich hätte ich meinen Masterabschluss nicht geschafft und mich nicht in meiner Dissertation zurechtfinden können.

Dr. Tobias Baumann danke ich für seine stets bereite Hilfestellung. Immer, wenn mein Computer nicht mehr richtig funktioniert oder ich nach einer Lösung des Problems in meiner Arbeit suche, warst du der fachlich kompetente, an den ich mich wenden kann.

Bei meinem Langzeit-Bürogenosse Dr. Georg Johannes Freiherr von Sass bedanke ich mich dafür, dass er eine anstrengende Kollegin sechs Jahre lang in jeglicher Hinsicht erdulden konnte.

Ich bin der ”Gourmet Lunch-Gruppe & friends”, Dr. Isabella Tolle, Dr. Ying Ma, Dr. Federica Agostini, Dr. Jessica Hannah Nickling, Christin Treiber-Kleinke und unserem einzigartigen Mann der Gruppe, Dr. Georg Johannes Freiherr von Sass, sehr dankbar für die kulinarischen Highlights aus aller Welt. Danke für die netten Gespräche, die eine schöne Abwechslung für den Tag waren.

Weiterhin möchte ich mich bei Tam Dang vom AK Süßmuth für seine Hilfsbereitschaft und für die angenehmen Gespräche bedanken. Danke für die nette Begleitung seit dem Studium und bis zum “bitteren“ Ende der Arbeit.

Sehr herzlich bedanken möchte ich mich bei Dr. Allison Ann Berger und Dr. Jan-Stefan Völler für die schnelle Durchsicht dieser Arbeit.

Den größten Dank muss ich meiner Familie ausdrücken. Ihr wart sowohl meine seelische als auch mentale Stütze, die mich seit Beginn meines Lebens bedingungslos liebt. Danke an meine “kleine” Schwester, die mir bei der Formatierung der Arbeit geholfen hat.

Zuletzt bedanke ich mich vor allem bei meiner besseren Hälfte, Dennis, und seiner Mama, dass sie mich jederzeit unterstützt haben.

Parts of this work

Publications

Exner, M. P.; Kuenzl, T.; **To, T. M. T**; Ouyang, Z.; Schwagerus, S.; Hoesl, M. G.; Hackenberger, C. P. R.; Lensen, M.; Panke, S.; Nediljko, B.; Design of S-Allylcysteine in Situ Production and Incorporation Based on a Novel Pyrrolysyl-tRNA Synthetase Variant, *Chembiochem*, 18, 85-90, 2017. (DOI: https://doi.org/10.1002/cbic.201600537)

To, T. M. T; Kubyshkin, V.; Schmitt, F.-J.; Budisa, N.; Friedrich, T.; Residue-Specific Exchange of Proline by Proline Analogs in Fluorescent Proteins: How "Molecular Surgery" of the Backbone Affects Folding and Stability, *Journal of visualized experiments: JOVE*, 180, 2022. (DOI: https://doi.org/10.3791/63320)

Posters

Engineering of an isoleucyl-tRNA synthetase for incorporation of fluorinated amino acids, 8th Peptide Engineering Meeting (PEM8 2018), Berlin, Germany, 08.11.2018 – 10.11.2018

Engineering of an isoleucyl-tRNA synthetase for incorporation of fluorinated amino acids, 8th Chemical Protein Synthesis Meeting (CPS 2019), Berlin, Germany, 16.06.2018 – 19.06.2019

Presentations

Labeling of novel fluorescent proteins by fluorinated amino acids, GRK 1582/2 Kollegseminar, Berlin, Germany, 23.11.2016

Isoleucyl-tRNA synthetase engineering towards activation of fluorinated amino acids, GRK 1582/2 Workshop, Berlin, Germany, 29.03.2018

Further publication beyond this work

To, T. M. T; Völler, J.-S.; Biava, H.; Koksch, B.; Budisa, N.; Global substitution of hemeproteins with noncanonical amino acids in Escherichia coli with intact cofactor maturation machinery, *Enzyme and microbial technology*, **106**, 55-59, 2017.

Abstract

Protein engineering is a comprehensive toolbox for the chemical modification of enzymes in particular, and for the expansion of molecular functional diversity in general.[1] In recent decades, two different categories have become established for the engineering of proteins. These include the approach of directed evolution approaches on the one hand and the strategies of rational protein design on the other hand.[2–6] In particular, the use of noncanonical amino acids to introduce new functionalities has gained importance in the engineers' toolbox. These include isostructural analogues of canonical amino acids as well as molecules with reactivities that can provide sites for further protein modifications.

In this study, we have presented a strategy for manipulating the protein biosynthesis machinery towards the incorporation of noncognate fluorinated substrates. In general, fluorinated amino acids are not genetically encoded. These mainly synthetic building block are valuable for the design of particularly stable protein folds and for targeting highly specific protein-protein interactions. Fluorine is small and has a very low polarizability and the strongest inductive effect among the chemical elements found on earth.[7] Due to these unique stereoelectronic properties, fluorine substitution is advantageously used in protein and peptide design. In this context, the strategy of directed evolution was applied to construct isoleucyl-transfer ribonucleic acid synthetase libraries for the isoleucine AUA rare codon reassignment with small aliphatic fluorinated amino acids, such as L-trifluoroethylglycine, by random mutagenesis. A suitable screening plasmid containing a mutant of superfolder green fluorescent protein (sfGFP) as reporter protein and a modified isoleucine transfer ribonucleic acid ($tRNA^{Ile1}_{UAU}$) from *Escherichia coli* was produced to create an enhanced molecular adaptor level for gene expression.[8] However, the required selection strain could not be constructed by genome editing due to the complexity of essential gene modification.

In the second part of the study, different reporter proteins were used in advanced design with noncanonical amino acids for improvement of their biophysical, chemical, and biological properties. A robust alkene-tagged sfGFP variant was obtained, which is a valuable target in medicinal chemistry. In addition, the residue-specific incorporation of proline analogues into green fluorescent protein (GFP) derivates ─ enhanced green fluorescent protein (EGFP), NowGFP, and KillerOrange ─ enables the study of the role of prolines in the typical β-barrel structure organization.

Zusammenfassung

Protein-Engineering beschreibt ein Prozess zur chemischen Modifizierung von Enzymen bzw. im Allgemeinen zur Erweiterung der molekularen Funktionsvielfalt.[1] In den letzten Jahrzehnten haben sich zwei unterschiedliche Kategorien für das Engineering von Proteinen etabliert. Dazu gehören zum einen die Ansätze der gerichteten Evolution und zum anderen die Strategien des rationalen Proteindesigns.[2–6] Insbesondere die Verwendung nichtkanonischer Aminosäuren zur Einführung neuer Funktionalitäten hat im Werkzeugkasten (Toolbox) der Ingeneure an Bedeutung gewonnen. Dazu gehören isostrukturelle Analoga kanonischer Aminosäuren sowie Moleküle mit Reaktivität, die Stellen für weitere Proteinmodifikationen bieten können.

In dieser Arbeit haben wir eine Strategie zur Manipulation der Proteinbiosynthesemaschinerie vorgestellt, um nicht-kanonische fluorierte Substrate in Proteine einzubauen. Fluorierte Aminosäuren sind im Allgemeinen nicht genetisch kodiert. Diese hauptsächlich synthetische Bausteine sind wertvoll für das Design besonders stabiler Proteinfaltungen und für die gezielte Steuerung hochspezifischer Protein-Protein-Wechselwirkungen. Fluor ist klein, hat zu dem eine sehr geringe Polarisierbarkeit und die stärkste induktive Wirkung unter den chemischen Elementen, die auf der Erde vorkommen.[7] Aufgrund dieser einzigartigen stereoelektronischen Eigenschaften werden Fluorsubstitutionen beim Protein- und Peptiddesign vorteilhaft eingesetzt. In diesem Zusammenhang wurde die Strategie der gerichteten Evolution angewendet, um Isoleucyl-Transfer-Ribonukleinsäure-Synthetase-Bibliotheken für die seltene Isoleucin-AUA-Codon-Neuzuordnung mit kleinen aliphatischen fluorierten Aminosäuren, wie L-Trifluorethylglycin, durch Zufallsmutagenese zu konstruieren. Ein geeignetes Screening-Plasmid, das eine Mutante des grün fluoreszierenden Superfolder-Proteins (sfGFP) als Reporterprotein und einer modifizierten Isoleucin-Transfer-Ribonukleinsäure ($tRNA^{Ile1}_{UAU}$) von *Escherichia coli* enthielt, um eine erhöhte molekulare Adaptor-Level für die Genexpression zu schaffen.[8] Der benötigte Selektionsstamm konnte jedoch aufgrund der Komplexität der essentiellen Genmodifikation nicht durch Genomeditierung konstruiert werden.

Im zweiten Teil der Arbeit wurden verschiedene Reporterproteine verwendet, um ihre biophysikalischen, chemischen und biologischen Eigenschaften mittels fortgeschrittenem Design mit nichtkanonischen Aminosäuren zu verbessern. Es wurde eine robuste, mit Alken markierte sfGFP-Variante synthetisiert, die ein wertvolles Ziel in der medizinischen Chemie darstellt. Darüber hinaus ermöglicht der ortsspezifische Einbau von Prolin-Analoga in Derivate des grün fluoreszierenden Proteins (GFP) – verbessertes grün fluoreszierendes Protein (EGFP), NowGFP und KillerOrange – die Untersuchung der Rolle von Prolinen in der typischen β-Barrel-Strukturorganisation.

Symbols and abbreviations

A	adenine or alanine
Å	Ångström
AA	amino acids
AA-AMP	aminoacyl-adenylate
AARS	aminoacyl-tRNA synthetase
AA-tRNA	aminoacyl-tRNA
Abu	α-aminobutyric acid or L-homoalanine
Abu-2AA	(L-homoalanyl) amino-2‘-deoxyadenosine
AC	anticodon
Amp	ampicillin
AMP	adenosine monophosphate
AmpR	ampicillin resistance
AMP-PNP	adenosine-5‘-[(β,γ)imido] triphosphate
APS	ammonium persulfate
A-site	aminoacyl-site
ATP	adenosine triphosphate
BI	before induction
BIS-TRIS	Bis-(2-hydroxy-ethyl)-amino-tris(hydroxymethyl)-methane
bp	base pairs
BSA	bovine serum albumin
c	concentration
C	cytosine
cAA	canonical amino acid/s
CAGO	CRISPR/Cas9-assisted sgRNA-free one-step
calc.	calculated
Cas9	CRISPR-associated protein 9
CAT	chloramphenicol acetyltransferase
CD	circular dichroism
cfsfGFP	cysteine-free superfolder green fluorescent protein
chr	chromosome
Cm	chloramphenicol
CmR	chloramphenicol resistance
CP1	connective peptide 1
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats

CRO	chromophore
cryo-EM	cryo-electron microscopy
CuAAC	cooper-catalyzed 1,3-dipolar alkyne-azide cycloaddition
CV	column volume
°C	degree Celsius
D	dihydrouridine or aspartic acid
Da	Dalton
deg	degree
DfeGly	(2*S*)-2-amino-4,4-difluorobutanoic acid or L-difluoroethylglycine
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
dsDNA	double-stranded deoxyribonucleic acid
DTT	dithiothreitol
DYT	Double Yeast Tryptone
ε_M	molar extinction coefficient/s
E	elution
E. coli	*Escherichia coli*
EDTA	ethylene-diamine-tetraacetic acid
EF	elongation factor
EGFP	enhanced green fluorescent protein
em.	emission
E-site	exit-site
ESI	electron spray ionization
exc.	excitation
et al.	*et alii*
EtBr	ethidium bromide
fAA	fluorinated amino acid/s
FACS	fluorescence-activated cell sorting
ft	flow through
FRET	fluorescence resonance energy transfer
g	gram or standard gravity unit
G	guanine or glycine
GFP	green fluorescent protein
GTP	guanosine-5’-triphosphate
h	hour/s
HCl	hydrochloric acid

H_6-tag	polyhistidine tag (HHHHHH)
HIC	hydrophobic interaction chromatography
HPLC	high performance liquid chromatography
HTS	high throughput screen
I	isoleucine or after induction
IEX	ion exchange chromatography
Ile-AMS	5 ’-*N*-[*N*-(L-isoleucyl)sulfamoyl]adenosine
IMAC	immobilized metal anion chromatography
IPTG	isopropyl-β-D-1-thiogalactopyranoside
k	kilo
k_{hyd}	hydrolysis rate constant
Kan	kanamycin
KanR	kanamycin resistance
kb	kilobase
λ	wavelength
L	liter, lysidine or leucine
LB	Lysogeny Broth
LC-ESI MS	liquid chromatography - electrospray ionization mass spectrometry
lys	lysate
µ	micro
m	mass, milli or meter
M	marker or molar
MAGE	Multiplex Automated Genome Engineering
MALDI-TOF MS	matrix assisted laser desorption ionization-time of flight mass spectrometry
MES	2-(*N*-morpholino)ethanesulfonic acid
MfeGly	(2*S*)-2-amino-4-monofluorobutanoic acid or L-monofluoroethylglycine
$MgCl_2$	magnesium chloride
$Mg(OAc)_2$	magnesium acetate
min	minute/s
Mj	*Methanocaldococcus jannaschii*
Mm	*Methanosarcina mazei*
M. mazei	*Methanosarcina mazei*
M. mobile	*Mycoplasma mobile*
mRNA	messenger ribonucleic acid
M_W	molecular weight

n	nano
NaCl	sodium chloride
NaOH	sodium hydroxide
ncAA	noncanonical amino acid/s
NMM	new minimal medium
NMR	nuclear magnetic resonance spectroscopy
NPC	nascent polypeptide chain
NTP	nucleoside triphosphate
ø	diameter
OD	optical density
OD_{600}	optical density at λ = 600 nm
o-pair	orthogonal suppressor AARS:$tRNA_{sup}$ pair
PAM	Protospacer Adjacent Motif
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
pel	pellet
PMSF	phenylmethylsulfonyl fluoride
PP_i	pyrophosphate
ppm	parts per million
pre-crRNA	precursor CRISPR RNA
P-site	peptidyl-site
Pyl	pyrrolysine
RBS	ribosomal binding site
RF	release factor
RNA	ribonucleic acid
ROS	reactive oxygen species
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
rt	room temperature
s	second/s
Sac	*S*-allyl-L-cysteine
S. aureus	*Staphylococcus aureus*
Sc	*Saccharomyces cerevisiae*
SCS	stop codon suppression
SDS	sodium dodecyl sulfate

SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SeC	selenocysteine
SEC	size exclusion chromatography
sfGFP	superfolder green fluorescent protein
sgRNA	single guide RNA
SPI	selective pressure incorporation
SPPS	solid phase peptide synthesis
ssDNA	single-stranded deoxyribonucleic acid
sup	suppressor
θ	molar ellipticity
t	time
T	thymine or temperature
T_m	melting temperature
TB	Terrific Broth
TCEP	tris(2-carboxyethyl)phosphine
TEV	*tobacco etch virus* protease
TfeGly	(2*S*)-2-amino-4,4,4-trifluorobutanoic acid or L-trifluoroethylglycine
TfeGly-2AA	2‘-(L-trifluoroethylglycyl) amino-2‘-deoxyadenosine
TilS	lysidine-tRNA synthetase
TψC	thymidine
tracrRNA	trans-activating CRISPR RNA
TRIS	tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
$tRNA_{sup}$	suppressor tRNA
T. thermophilus	*Thermus thermophilus*
U	uracil
UV	ultraviolet
V	volt or volume
Vis	visible
v/v	volume per volume
w	wash
wt	wildtype
w/v	weight per volume
3D	three-dimensional

The single- or three-letter code were used for abbreviation of canonical amino acids. [*Eur. J. Biochem.* **1984**, *138*, 9–37].

Table of Contents

1. Introduction

1.1 The genetic code – How everything started.

The genetic code is defined as the biochemical basis of heredity encoded in terms of codons in the genetic materials, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), that ubiquitously determine the specific amino acid (AA) sequence of proteins in the three-domain system of life.[9,10] This definition was made possible the enormous amount of results from fundamental biological research since the early 19th century. Neither Berzelius, the creator of the term 'protein' in 1838 (Greek *proteios*, meaning primary), nor the discoverers of the building blocks of these 20 'macromolecular nitrogen-containing substances' ─ now known as AA, starting in 1810 with Wollaston and his isolation of 'cystic oxide', the oxidized dimeric form of cysteine ─ were aware of their contribution to deciphering the complex genetic information flow.[11–13] Although the first work about the biopolymer "nuclein", the former term of modern nucleic acid, was published in 1871 by Miescher, it took over 80 years to solve the structure of the polymeric macromolecule as a double helix by Watson and Crick in 1953.[14,15] Subsequently, the connection between the DNA, RNA and proteins was postulated as the central dogma of biology by Crick in 1958.[9] The major part of this theory implied DNA as the reliable storage device of the genetic information, which exactly predetermine the sequence of proteins. DNA passed the information by generation of its own transcript as RNA and following translation into proteins at the ribosome, supported by an adapter molecule – transfer RNA (tRNA) - with anticodon, which is recognized by a certain codon on the RNA. "Specific enzymes", presently known as aminoacyl-tRNA synthetases (AARS), attach a particular AA covalently to each tRNA via catalysis of a specific aminoacylation reaction. The dogma clarifies exclusively the unidirectional genetic flow of information from DNA to protein and allows no possible transfer from protein back to the nucleic acids. In 1961 the historic poly-U experiment by Nirenberg and Matthaei illustrated not only the first triplet base sequence (codon), UUU, corresponding to phenylalanine (Phe, F) in a protein, but also confirmed Crick's former postulate.[16] By 1966, 64 possible coding units had been identified and assigned to the different AA, thereby moved the humanity one step closer to deciphering the genetic code.[17]

1.1.1 Codon usage bias

The double helical DNA is composed of a pentose-phosphate scaffold linked to the nucleobases Adenine (A), Cytosine (C), Guanine (G), and Thymine (T), that defined, in fact,

the primary sequence of proteins. The two strands are connected to each other by hydrogen bonds between opposite C-G and A-T base pairs. The RNA is constructed similarly with the exception, that T is substituted with Uracil (U). Since these essential nucleic acids consist of four bases, in each case, there are 4×4×4 = 64 possible triplets (Tab. 1).[17]

Table 1 Genetic code in RNA format and codon usage bias in *E. coli* genes. % represents the average frequency this codon is used per 100 codons and the ratio shows the abundance of that codon relative to all of the codons for that particular AA. The start and stop codons for translation process are highlighted in green and red, respectively. Table modified from Maloy *et al*.[18]

		Codon	AA	%	Ratio	Codon	AA	%	Ratio	Codon	AA	%	Ratio	Codon	AA	%	Ratio		
first base	U	UUU	Phe	1.9	0.51	UCU	Ser	1.1	0.19	UAU	Tyr	1.6	0.53	UGU	Cys	0.4	0.43	U	third base
		UUC	Phe	1.8	0.49	UCC	Ser	1.0	0.17	UAC	Tyr	1.4	0.47	UGC	Cys	0.6	0.57	C	
		UUA	Leu	1.0	0.11	UCA	Ser	0.7	0.12	UAA	Stop	0.2	0.62	UGA	Stop	0.1	0.30	A	
		UUG	Leu	1.1	0.11	UCG	Ser	0.8	0.13	UAG	Stop	0.03	0.09	UGG	Trp	1.4	1.00	G	
	C	CUU	Leu	1.0	0.10	CCU	Pro	0.7	0.16	CAU	His	1.2	0.52	CGU	Arg	2.4	0.42	U	
		CUC	Leu	0.9	0.10	CCC	Pro	0.4	0.10	CAC	His	1.1	0.48	CGC	Arg	2.2	0.37	C	
		CUA	Leu	0.3	0.03	CCA	Pro	0.8	0.20	CAA	Gln	1.3	0.31	CGA	Arg	0.3	0.05	A	
		CUG	Leu	5.2	0.55	CCG	Pro	2.4	0.55	CAG	Asn	2.9	0.69	CGG	Arg	0.5	0.08	G	
	A	AUU	Ile	2.7	0.47	ACU	Thr	1.2	0.21	AAU	Asn	1.6	0.39	AGU	Ser	0.7	0.13	U	
		AUC	Ile	2.7	0.46	ACC	Thr	2.4	0.43	AAC	Asn	2.6	0.61	AGC	Ser	1.5	0.27	C	
		AUA	Ile	0.4	0.07	ACA	Thr	0.1	0.30	AAA	Lys	3.8	0.76	AGA	Arg	0.2	0.04	A	
		AUG	Met	2.6	1.00	ACG	Thr	1.3	0.23	AAG	Lys	1.2	0.24	AGG	Arg	0.2	0.03	G	
	G	GUU	Val	2.2	0.29	GCU	Ala	1.8	0.19	GAU	Asp	3.3	0.59	GGU	Gly	2.8	0.38	U	
		GUC	Val	1.4	0.20	GCC	Ala	2.3	0.25	GAC	Asp	2.3	0.41	GGC	Gly	3.0	0.40	C	
		GUA	Val	1.2	0.17	GCA	Ala	2.1	0.22	GAA	Glu	4.4	0.70	GGA	Gly	0.7	0.09	A	
		GUG	Val	2.4	0.34	GCG	Ala	3.2	0.34	GAG	Glu	1.9	0.30	GGG	Gly	0.9	0.13	G	
		U				C				A				G					
		second base																	

Three of them encode termination of the translation, the stop codons: UAA (*ochre*), UGA (*opal*) and UAG (*amber*), and the remaining 61 specify the 20 canonical amino acids (cAA, **Table 1**). Each codon specifies one cAA, meaning the genetic code is unambiguous, but except for Methionine (Met – AUG) and Tryptophan (Trp – UUG), all cAA can be assigned to more than one codon, resulting in the phenomenon of codon degeneracy.[19,20] Moreover, encoding codons are quantitatively proportional to the abundance of the corresponding cAA in the proteome, but the frequency of these synonymous codons are not identical, which is defined as the codon bias in organisms.[21] In this regard, the translation level can be fine-tuned by accurate codon usage, e.g. by accurate usage of different rare codon copy numbers, which are less abundant utilized for cAA assignment.[8,22–24] However, proteinogenic AA can also be

encoded by stop codons, such as in methanogenic archaea the 21st AA selenocysteine (SeC) by UGA and pyrrolysine (Pyl) by UAG, respectively.[25–27] Additionally, proteins are complex highly specific biomacromolecules and require multifaceted building blocks. Expanding the AA repertoire is indispensable and can be achieved in living organisms by different methods.[28] Though, the major class is post-translational modification, whereby the existing side chain of the AA is modified or a new chemical motive is introduced.[29] For instance, using the cofactor biotin in biotinylation targeting the ε-amino group of a lysine (K) residue at the active center converts the inactive apocarboxylase into a fully active holocarboxylase.[30,31] There are over 900 naturally occurring AA, including free AA and as intermediates in metabolism, demonstrating the importance of AA diversity in appropriate protein functions.[32,33] More detail is provided in the next Section.

1.1.2 Amino acids – fundamental building block of functional diversity

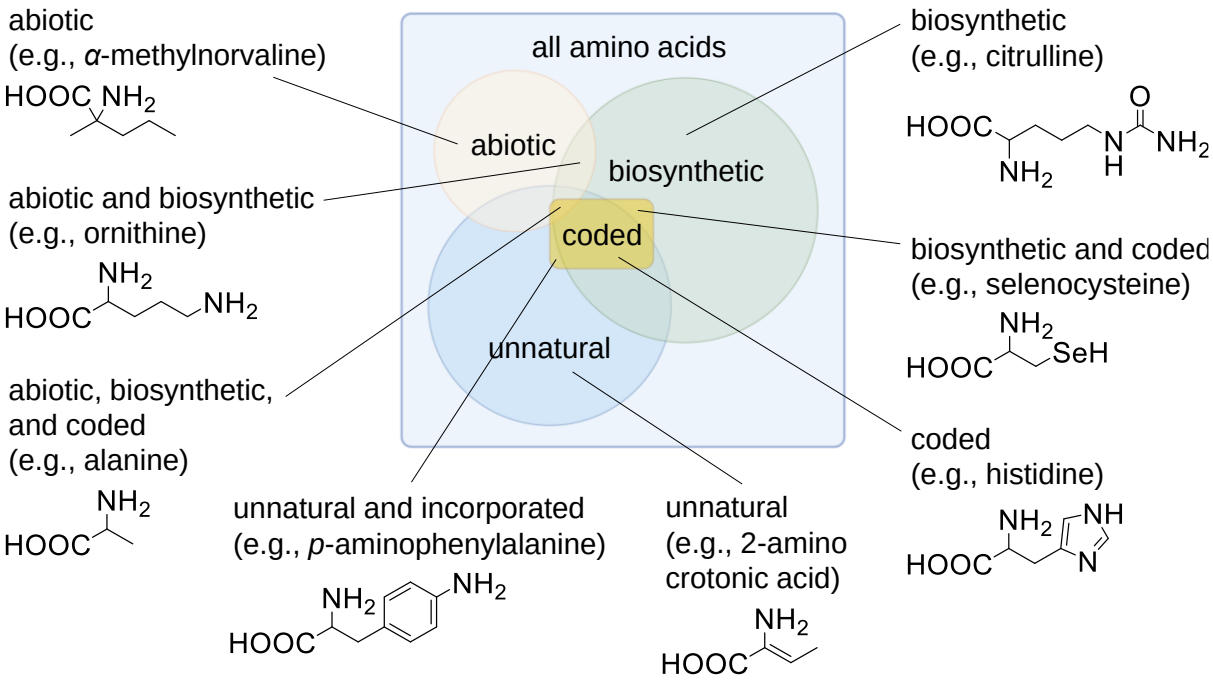


Figure 1 Categories of amino acids with examples represented in a Venn diagram. The three main groups include abiotic, that are probably synthetized before life evolved, biosynthetic derived from natural biosynthetic pathways, and unnatural AA. Of all the AA, genetically coded composed only a small number. AA, such as alanine, is assigned to an overlapping section – abiotic, biosynthetic, and coded – since it is a proteinogenic AA having an origin from abiotic synthesis. Figure adopted from Lu *et al*. (2006) and created with ChemDraw.[35]

The standard alphabet of proteins, shared by most organisms on earth, consists of 20 encoded L-α-amino acids (**Figure 2A**). Their general structure contains carboxylic acid (-COOH) and amino (-NH_2) functional groups at the central α-carbon atom, as well as a covalently bonded hydrogen atom and an either aliphatic or aromatic side chain (-R) (**Figure 2A**). According to different aspects several classifications exist. For example, in terms of nutrition, amino acids are categorized into three groups: essential, nonessential, and conditionally essential.[34] With respect to evolution of the AA alphabet, the origin of synthesis is given in three main categories: abiotic, biosynthetic, and engineered (**Figure 1**).[35]

Here, AA beyond the standard alphabet are considered. In the context of this study, classification based on genetic background is relevant. This corresponds to the two groups: natural encoded canonical amino acids (cAA) and added unnatural noncanonical amino acids (ncAA) (**Figure 2**). The standard 20 proteinogenic cAA are genetically encoded and serve as the fundamental basis of protein structure and function (**Section 1.1.1** and **1.3**). In contrast, a small proportion of the countless ncAA with unusual side chains, D-stereochemistry, or atypical backbone bonding (**Figure 2B**) are incorporated into proteins by engineered translation apparatuses (**Section 1.2.3**).[36,37] However, the interest in encoding ncAA, either found in nature or synthetically produced, is particularly extensive due to their broad applications in chemistry, medicine and biology.[38,39] Especially in drug discovery and development, introduction of chemical functionality into peptides and proteins via ncAA could improve their bioavailability, permeability, stability, and selectivity of the resulting product.[40–43] Newly functionalized biomolecules can be further applied for bioorthogonal reactions, among others to achieve complex *in situ* assembly at expected target site.[44] For example, the use of thiol-ene hydrogel in drug delivery studies is enabled by introduction of reactive thiol and alkene functional groups into protein-based therapeutics.[45]

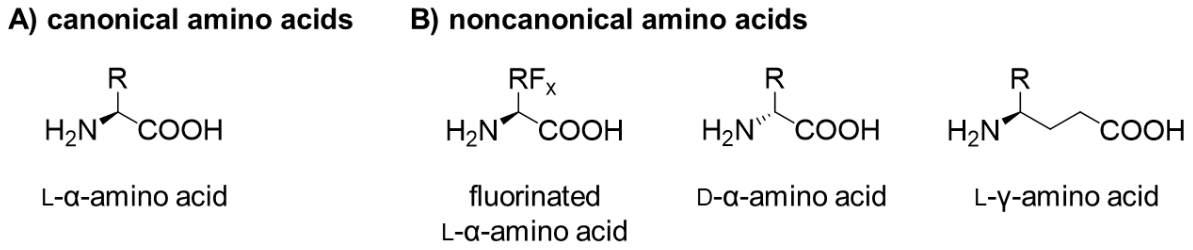


Figure 2 Simplified structural representation of canonical amino acids and examples of noncanonical amino acids.

In medicinal chemistry and chemical biology, fluorinated amino acids are the most utilized ncAA.[46,47] Not surprisingly, 15 % of 200 top-selling pharmaceuticals in 2020 contain at least one fluorine atom.[48] Due to the similar atomic size compared to hydrogen, fluorine displays a

perfect substituent regarding molecular volume and geometry.[49] In proteins, fluorine can interact with both polar and hydrophobic moieties, resulting in polar interactions with donors of hydrogen bond (e.g., backbone amino groups, polar side chains), hydrophobic interactions with lipophilic side chains, and orthogonal multipolar interactions, e.g., with backbone carbonyl groups, or sulphur containing side chain (Cys).[50] Besides, it possesses three non-bonding electron pairs, the highest electronegativity in the periodic table and very low polarizability.[51] These unique physico-chemical properties enable the strongest chemical bonds with carbon (C–F bond), adjustable hyper-conjugative and inductive electron withdrawing effect upon fluorination.[52] Thus, directed modification of peptides and proteins can be performed to fine-tune properties including conformation, folding, enzymatic activity, acidity/basicity, and stability (proteolytic, thermal as well as in organic solvent).[7,53–56] Moreover, fluorine is nearly unrepresented in natural biological environments, for which it can be further deployed for bioorthogonal labeling of biomolecules in order to investigate protein-protein interactions, protein structure and dynamics by highly accurate ^{19}F NMR analysis.[57–59]

1.2 Synthesis of proteins

1.2.1 *In vitro* strategies of protein synthesis

A fundamental approach of chemical peptide production is solid-phase peptide synthesis (SPPS), making peptides up to 50 AA in length accessible.[60] For “difficult” (e.g., containing much hydrophobic AA and bulky side chains) and longer sequences (≤ 100 AA) the use of microwave-assisted SPPS promotes significantly increased reaction times and a product with higher purity level.[61] To overcome the size limit of polypeptide chains (> 100 AA), shorter peptide fragments are prepared by SPPS and chemoselectively combined by chemical ligation, such as click reaction, Staudinger ligation, and native chemical ligation.[62,63] However, this stepwise production method has a negative impact on the overall yield, since isolation and purification of intermediates are required. Subsequent folding of the peptide to achieve an intact protein can occur as a significant drawback of this technique.

Another valuable approach is the cell-free gene expression system applying chemically aminoacylated dinucleotides with attached AA. Appropriate AA is enzymatically ligated with truncated tRNA bearing appropriate anticodon (**Section 1.2.2.1**) for *in vitro* translation.[64–66] This method allows rapid and inexpensive protein production by easy access to the cellular processes due to the simplified representation of the in-cell environment. The major drawbacks of this technique are the complex chemical synthesis of the compounds and their low stability. Therefore, large polypeptides and proteins are typically achieved through a

biological expression system using the cellular machinery, that will be explained in detail by the coming **Section 1.2.2**.

1.2.2 Protein biosynthesis

In order to encode the universal stored genetic information into functional proteins, organisms have to undergo a biological multi-step process, initiated with transcription of DNA into RNA, then followed by translation of RNA into protein. Various types of RNA having different functions facilitate the second part of protein biosynthesis: the ribosomal RNAs (rRNA), the major catalytic component of the ribosome, tRNA, the adapter molecule between codons and AA, and mRNA, as template for the biological protein synthesis.

As briefly outlined above, translation start involves two components of the ribosomal translation apparatus that are responsible for the accurate interpretation of the genetic code: AARS and tRNA (**Section 1.2.2.1** and **1.2.2.2**). AARS specifically selects its cognate AA from the cellular repertoire of cAA and loads it to the corresponding tRNA consuming chemical energy, adenosine triphosphate (ATP) (**Section 1.2.2.2**). This connection is clearly defined by the encoding nucleotide sequence of the mRNA. The charged aminoacyl-tRNA (AA-tRNA) is now transferred to the ribosome by the guanosine-5'-triphosphate hydrolase (GTPase) elongation factor Tu (EF-Tu) forming a ternary EF-Tu-AA-tRNA-GTP complex. The ribosome, a multifunctional ribonucleoprotein complex within all living organisms. In prokaryotic cells, it consists of a small 30S and a large 50S subunit with three tRNA binding sites: the aminoacyl-

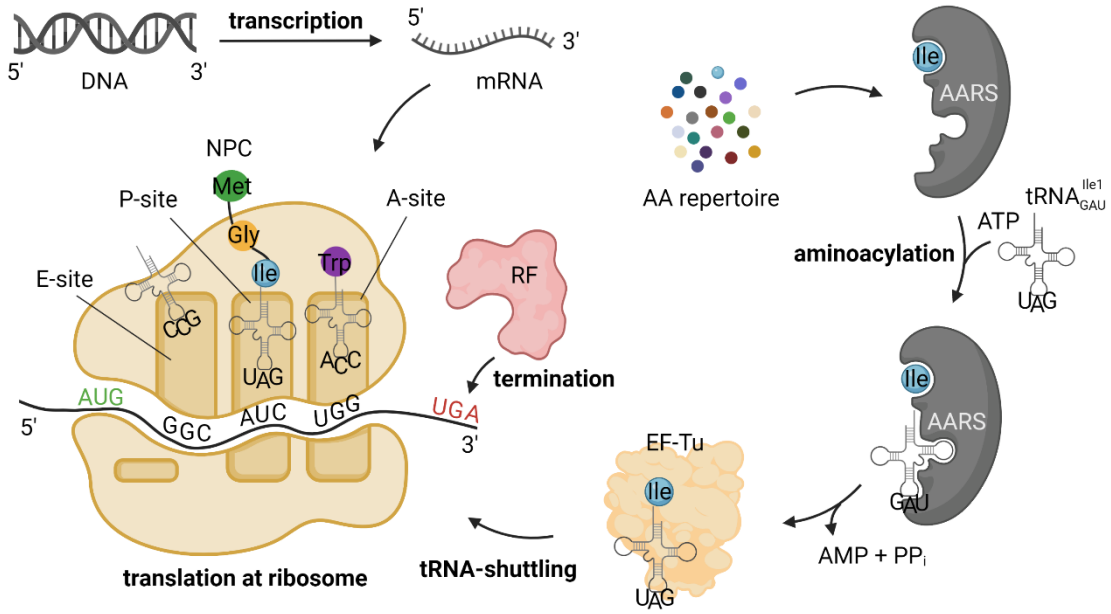


Figure 3 Schematic presentation of cellular protein translation system. Aminoacyl-site (A-site), peptidyl-site (P-site), exit-site (E-site), nascent polypeptide chain (NPC), release factor (RF), elongation factor Tu (EF-Tu). Figure created with BioRender.com.

site (A-site), the peptidyl-site (P-site) and the exit-site (E-site). An accurate codon-anticodon base pairing allows the AA-tRNA to release from the ternary complex and fully accommodate at the A-site. Afterwards, the peptidyl-tRNA in the P-site is transferred to the AA-tRNA in the A-site by a transpeptidation reaction of the polypeptide chain. The peptidyl-tRNA in the A-site is elongated and then moved to the P-site, thereby extends the nascent polypeptide chain (NPC) by one AA. The deacylated tRNA is transferred to the E-site and leaves the ribosome from there. Termination of the translation is initiated once a stop codon of the mRNA resides at the A-site. A release factor (RF) enters the A-site instead of a AA-tRNA and triggers the release of the NPC from the P-site.[20] A schematic outline of the ribosomal protein synthesis is summarized in **Figure 3**.

On average, the translation in *E. coli* proceeds at a maximal speed of 20 AA per second with an estimated error rate of 10^{-4}.[67] Therefore, different involving enzymes at each step implement proofreading mechanisms and enable a high fidelity process. One important step is the correct codon-anticodon pairing that is assumed for several further reactions.[68] First, the 16S-rRNA of the 30S subunit stabilizes the mRNA for the translation process by forming hydrogen bonds between its three conserved bases (G530, A1492, and A1493) and the codon-anticodon complex, only if the first two base pairs of the complex complemented perfectly, following the Watson-Crick geometry. Second, the correct pair enables the fully accommodation of AA-tRNA in the A-site of the ribosome and release of EF-Tu by GTP-hydrolysis. In case misacylated tRNA is formed, the initial binding to EF-Tu will be negatively affected and thus, reduces the amount of incorrectly charged tRNA. Furthermore, mischarging of a tRNA is avoided by the specificity of AARS (**Section 1.2.2.2**). Nevertheless, a small number of misacylated tRNA still arrives at the ribosome and non-cognate AA is introduced into the NPC. Exploiting this disturbance is a useful tool for genetic code engineering (**Section 1.2.3**)

1.2.2.1 Transfer RNAs

As previously mentioned in **Section 1.1.1**, tRNAs serve as molecular adaptors between the 61 sense codons and the 20 AA during translation. Depending on identity and organism, tRNAs consist of 72 - 95 ribonucleotides and modified nucleobases.[69] They adopt a cloverleaf secondary structure with five characteristic structural features: dihydrouridine (D) stem and loop, anticodon (AC) stem and loop, variable loop, thymidine (TψC) stem and loop and acceptor stem with 3' - CCA end (see **Figure 4A**).[70] Once a L-shaped tertiary structure is formed with 15 conserved and eight semi-conserved residues, the two active ends of tRNA become clearly visible: the anticodon loop and the acceptor stem (**Figure 4**). Therefore,

important identity elements are located in these areas (**Figure 4B**).[71]

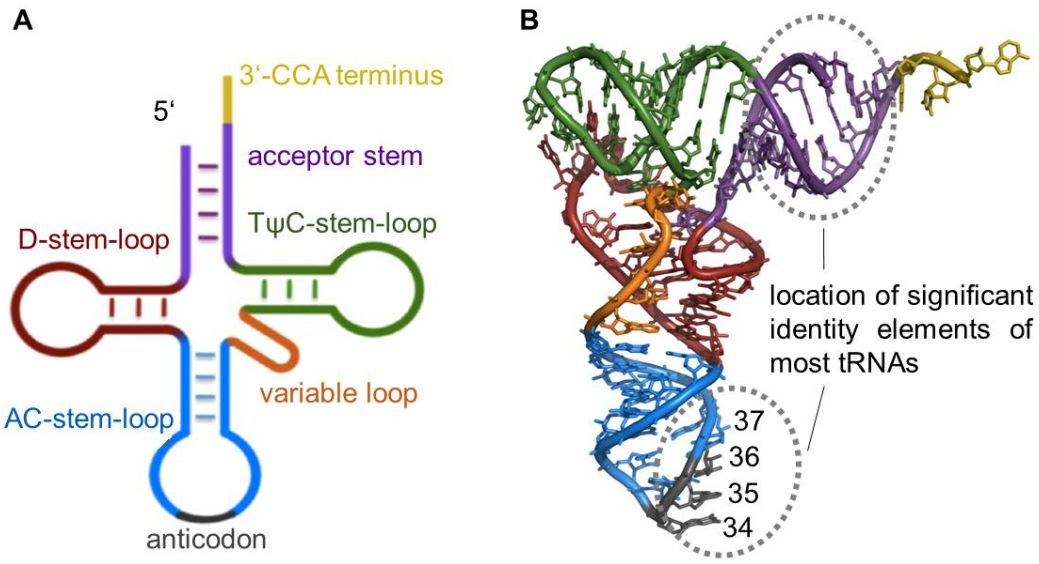


Figure 4 Schematic presentation of the cloverleaf secondary structure of tRNA (A) and its general tertiary structure (B). Figure modified from Shareghi *et al*. (2012).[70]

In general, tRNAs are transcribed from a single promoter into precursor by a RNA polymerase-mediated reaction. This tRNA gene product undergoes a maturation process followed by additional modifications to ensure the structural integrity and composition of the final adapter molecule. While 5' transcript processing is mainly performed by the ribonucleoprotein RNase P, the maturation at the 3' end requires a combination of endonucleolytic cleavage events facilitated by exoribonucleases, comprising RNases II, BN, D, PH, PNPase and T. This underlines the complexity and stringent necessity of the accurate functionalization of the AA attachment site, the 3' - CAA terminus.[72]

Unusual nucleotides found in tRNA molecules are produced by covalent modification of a normal nucleotide in the precursor tRNA chain, catalyzed by tRNA modification enzymes.[73] About 10 % of the tRNA sequence is chemically modified post-transcriptional, which enables both functional diversity as well as fine-tuned tRNA structure, which is critical for the stability, folding and decoding of tRNAs.[74] It is presumed that for instance, modifications outside of the anticodon regions destabilize the overall molecule structure. The highest abundance and variety of mutations are remarkably located at positions 34 and 37 of the AC-stem-loop. Both are associated with maintaining the translation accuracy as well as efficiency. Especially, position 34 represents the first base of the anticodon, also known as the wobble base. This position proves necessary in wobble base pairing with the third nucleobase of a codon on the mRNA, in order to facilitate expanded translational capacity.[75] In bacterial genomes there are only 46 tRNA genes involved in decoding the 61 sense codons.[76] Hence, by the non-Watson-

Crick base pairing a single tRNA sufficiently recognizes more than one synonymous codon of a specific AA. tRNAs aminoacylated with the same AA are termed as isoacceptors.

The cognate tRNA is accurately selected from the cellular pool by a particular AARS. The matured tRNA possesses important interaction and binding identity moieties including positive (determinants) as well as negative (anti-determinants) elements that induce appropriate aminoacylation and impede mischarging. These include nucleobases primarily found at the two distal ends of the adapter molecule: the acceptor stem, and the anticodon loop.[77] According to the fact that activated AA is ultimately transferred to the 3' - CAA terminus for charging, the important role of the acceptor stem in recognition is certain.[78] On the one hand, the discriminator nucleotide N73 and the first four base pairs of the AA accepting stem are required for efficient aminoacylation process in case they are correctly aligned in the cognate AARS's catalytic core.[79] On the other hand, the same residues can act as anti-determinants and hinder the tRNA charge by non-cognate AARS. Moreover, the anticodon loop is involved in identity recognition by most AARS. For example, position 37 only participates in the tRNA charging step of class I synthetases (**Section 1.2.2.2**). Another identity element is the anticodon, especially positions 35 and 36. They are in close contact with the AARS systems and essential for codon-anticodon base pairing. The wobble base 34 contributes to the property as an anti-determinant element, for instance modification of C34 to L34 of the *E. coli* $tRNA^{Ile2}_{CAU}$ prevents a misaminoacylation by non-cognate methionyl-tRNA synthetase (MetRS) (details in **Section 1.2.3.1**). Minor identity elements are found in other tRNA domains such as D-stem, variable loop and inside of the L-shaped tRNA, but less abundant.[80–84]

1.2.2.2 Aminoacyl-tRNA synthetases

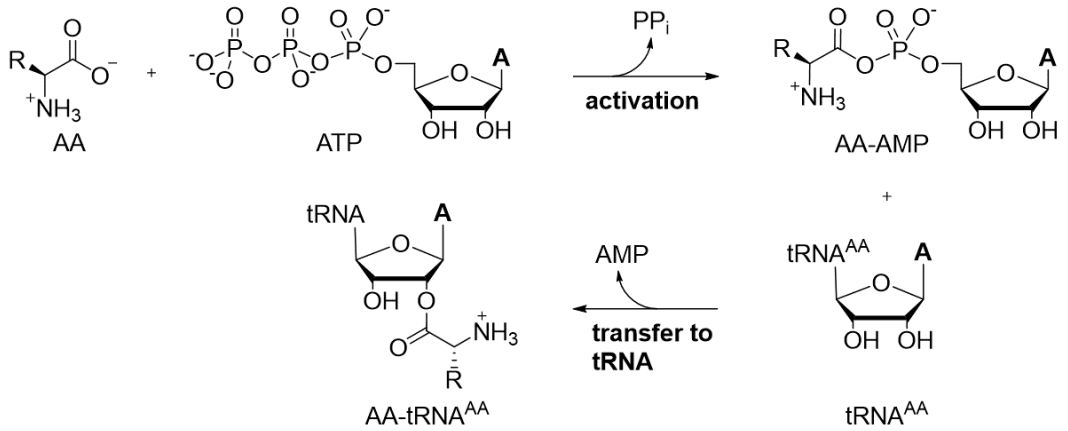


Figure 5 The tRNA aminoacylation reaction. Figure created with ChemDraw.

Having their similarity across organisms throughout all domains of life, AARS are involved in

a diverse range of cellular processes such as intron splicing, signaling pathways, human HIV type 1 viral assembly, and antiapoptotic interactions with apoptosis signal-regulating kinase 1.[85–88] However, the most important and critical role of this family of enzymes is in protein biosynthesis, in order to ensure accurate translation. These multitasking molecules ligate corresponding tRNAs with their cognate AA according to the genetic information given on a mRNA with both high cognate substrate recognition and stringent proofreading efficiency. The mentioned aminoacylation reaction occurs in two steps (**Figure 5**). First, AA and ATP align in the catalytic site of the AARS, thereby activate the AA by a condensation reaction forming a stable aminoacyl-adenylate intermediate (AA-AMP) bound to the enzyme (**Figure 5, activation**). Pyrophosphate (PP_i) is released from the active center as a side product. In the second step, the carbonyl carbon of the adenylate undergoes an esterification reaction with one of the hydroxyl groups of the tRNA's A76, resulting in AMP and an aminoacyl-tRNA (AA-$tRNA^{AA}$, **Figure 5**, **transfer to tRNA**), which is then used for protein translation at ribosomes (**Figure 3**).[89] The catalytic mechanism of the aminoacylation is universally conserved, whereas AARS possess extensive structural, and in certain cases, functional variety. There are 23 AARS known to date that can be classify into two evolutionary very distinct classes (I and II) with respect to the structural topologies of their catalytic sites (**Table 2**). Class I enzymes are mainly monomeric and feature a catalytic domain as nucleotide-binding Rossmann fold, including two highly conserved motifs KMSKS and HIGH (**Figure 6A**). These significant sequence elements enable ATP binding in an extended configuration and stabilization of the aminoacylation intermediate AA-AMP:AARS, respectively (**Figure 6A**). Besides, class I

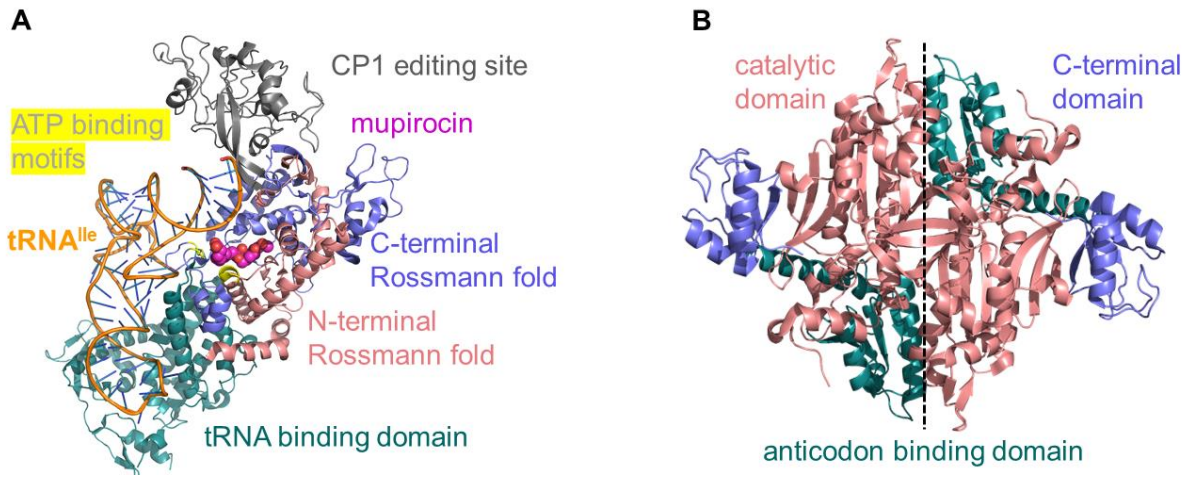


Figure 6 Crystal structure of the monomeric class I enzyme isolecyl-tRNA synthetase (IleRS) from *Staphylococcus aureus* complexed with $tRNA^{Ile}$ and mupirocin as an inhibitor at the active site (A), and the homodimeric class II enzyme prolyl-tRNA synthetase (ProRS) from *Thermus thermophilus* (B). ATP binding motifs, KMSKS and HIGH are shown in yellow. The connective peptide (CP1) is a highly conserved proofreading domain of class I enzymes. Structures are generated from PDB entries 1QU2, and 1H4S by PyMOL, respectively.

synthetases access the tRNA from the minor groove of the acceptor stem and undergo aminoacylation reaction at the 2'-OH of the tRNA's terminal adenosine. In contrast, the class II enzymes generally occur multimeric with the active domain composed of an antiparallel β-fold bearing seven β-sheets and three flanking helices (**Figure 6B**). The catalytic antiparallel β-fold of class II tRNA-ligases consists of three conserved residues, namely motif 1, 2, and 3 (**Figure 6B**). Motif 1 is important for dimerization, as motifs 2 and 3 are responsible for ATP and AA recognition. A bent configuration with the γ-phosphate folding back over the adenine ring characterizes the ATP binding within this class. Furthermore, class II synthetases access the acceptor stem via the major groove side and use the 3'-OH of the terminal adenosine for aminoacylation.[89] Both classes also exhibit pronounced differences according to the kinetics of their aminoacylation reaction. For class I AARS (except for IleRS and some GluRS) aminoacyl-tRNA release is the rate limiting step whereas it is the AA activation rate in case of class II. In addition, the anticodon loop of the tRNA is mainly used by class I AARS as identity element, which is not common for class II enzymes.

Table 2 Classification of AARS based on the structure and their particular properties. α, monomer; α_2, homodimer; α_4, homotetramer; $\alpha_2\beta_2$, heterotetramer.[89,90]

	Class I			**Class II**		
		Structure	Editing		Structure	Editing
Subclass a	MetRS	α_2	Yes	SerRS	α_2	Yes
	ValRS	α	Yes	ThrRS	α_2	Yes
	LeuRS	α	Yes	AlaRS	α_2	Yes
	IleRS	α	Yes	$GlyRS_a$	α_2	No
	CysRS	α, α_2	No	ProRS	α_2	Yes
	ArgRS	α_2	No	HisRS	α_2	No
Subclass b	GluRS	α	No	AspRS	α_2	No
	GlnRS	α	No	AsnRS	α_2	No
	LysRS-I	α	No	LysRS-II	α_2	Yes
Subclass c	TyrRS	α	No	$GlyRS_b$	$\alpha_2\beta_2$	No
	TrpRS	α_2	No	PheRS	$\alpha_2\beta_2$, α_2	Yes
				PylRS	α_2	No
				SepRS	α_4	No
Catalytic domain	Rossmann fold			Antiparallel β-fold		
tRNA binding site	Minor grove			Major grove		
Aminoacylation site	2'-OH			3'-OH		
Anticodon recognition	Most			Few		

Based on phylogenetic analysis, the differences in structural and mechanical properties as well as domain arrangement divide both AARS categories into three further sub-classes (a, b, c). Each of them is related to their AA substrates with similar chemical properties (**Table 2**). For example, synthetases of the subclass Ia tend to load AA that are aliphatic (Val, Leu, Ile) and thiolated (Met, Cys), respectively. Interestingly, there are similar substrate tendencies within subgroup and across the classes, such as polar and charged AA are considered as substrates of both subclass Ib enzymes (Gln, Glu, Lys) and IIb counterparts (Asn, Asp, Lys).[91] Among all three domains of life, it is expected to have 20 AARS for activation of all 20 cAA involved in protein biosynthesis. However, the set of encoded AARS differs for each organism. Especially bacterial and archaeal genomes lack up to three AARS genes, such as GlnRS, AsnRS, and CysRS.[92,93] Therefore, these organisms evolved indirect aminoacylation pathways to compensate the missing AARS. Another reaction using this indirect charging strategy is the aminoacylation of $tRNA^{SeC}$ since there is no gene encoding SecRS throughout living organisms. In order to synthesize the SeC-$tRNA^{SeC}$, cells first load Ser onto $tRNA^{SeC}$ involving endogenous SerRS. The resulting Ser-$tRNA^{SeC}$ is then modified to SeC-$tRNA^{SeC}$ by different enzymes upon the organism.[94,95]

To achieve high translational fidelity and acceptable quality control, AARS must be able to distinguish their cognate substrate from those structurally similar AA. As mentioned previously, identity elements of the cognate tRNA isoacceptors aid in proper recognition (**Section 1.2.2.1**). Another strategy is by AA discrimination mechanisms, such as exclusion of noncognate AA by size, and charge, respectively, or use of coordinated metal ions as cofactor to bind certain functional groups of the substrate.[96–98] Besides, there are evolved intramolecular and intermolecular editing activities that are proposed to be the second sieve after the aforementioned recognition mechanisms. A highly conserved proofreading domain, e. g. the connective peptide I (CP1, **Figure 6A**), is only present in class I enzymes and ensures the accurate discrimination of very similar aliphatic AA (Leu, Ile, Val) by the appropriate AARS.[99–101] In comparison, the editing domains of the class II AARS are more idiosyncratic.[102] All common known *in vivo* editing pathways are summarized in the following **Figure 7**. Depending on the editing occurring before or after the transfer of the AA to corresponding tRNA, editing can be categorized into two parts: pre- and post-transfer editing. The former can further be divided in two subclasses including tRNA-independent and –dependent reactions. The first part implies the selective tRNA-independent release of the AA-AMP intermediate from the AARS to the cytosol followed by a spontaneous hydrolysis of this labile phosphoester bond. The latter encloses the tRNA-dependent cleavage of the enzyme bound intermediate-tRNA complex, *AARS*:AA-AMP-$tRNA^{AA}$, into its individual components (**Figure 7A**). The described reactions take place both in the active site of the AARS and in an enzyme-independent manner. In the post-transfer editing part, the ester bond is typically hydrolyzed in

a domain apart from the activation center (**Figure 7B**).[103] For this purpose, the enzyme-bound AA-tRNA, *AARS*:AA-tRNA^{AA}, changes its conformation and rearranges the 3' terminus harboring the AA for translocation from the active site to the editing site, while the core of the tRNA remains bound to the enzyme.[104–106] Once the AA-tRNA^{AA} is released from AARS other components participate in the translation quality control machinery such as EF-Tu (**Section 1.2.2**) and *trans*-editing factors (**Figure 7B**).

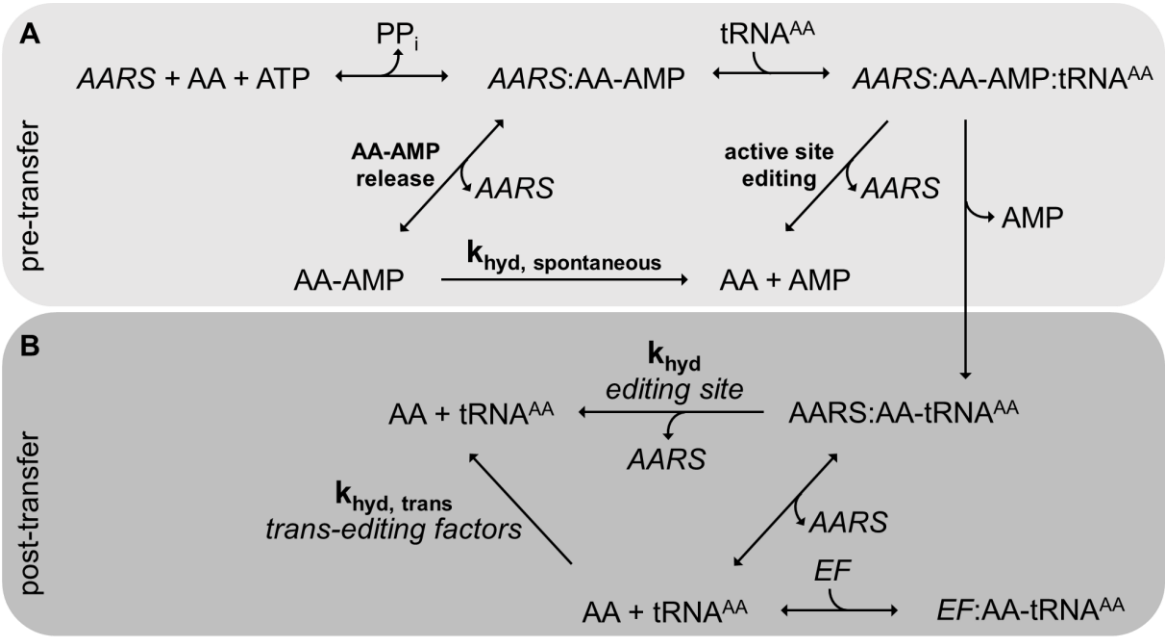


Figure 7 Schematic overview of editing pathways including A) pre-transfer and B) post-transfer parts. Editing events are represented in bold and involved enzymes in italics. k_{hyd} stands for hydrolysis reaction caused by different factors.

Overall, AARS are crucial enzymes for plenty biological processes, especially protein biosynthesis. These ligases of Crick's adaptor molecules perform their function with high accuracy. Since editing mechanisms could only evolve against naturally occurring AA, AARS also possess a significant substrate tolerance and permit the recognition and charging of structurally related moieties onto tRNAs. This phenomenon provides an important tool for synthetic biology, engineering AARS towards incorporation of ncAA with respect to expand the genetic code and therefore, increase the protein function diversity.

1.3 Protein engineering

Given by their tremendous importance in health, medicine, industry, and biotechnology, proteins have become a target for research with respect to increase their range of applicability.[107] Protein engineering is a procedure by which structure and function of the biomolecules are specifically modified for enhanced stability and/or improved enzymatic properties for further application, such as protein labelling. These powerful biotechnological tools include the design of protein models using molecular modeling techniques, site-specific mutagenesis of existing genes, and directed evolution methods for generation of new natural products.[108–112] Especially, the usage of ncAA in polypeptides facilitates the extension of their chemical functionalities (**Section 1.1.2**). Proteins containing such extended side chains can be deployed in subsequent analysis and bioorthogonal reactions to further investigate the structural, functional effects, as well as inter- and intramolecular interactions. For example, using an alkyne functionalized barstar with different galactose and lactose derivatives in cooper-catalyzed 1,3-dipolar alkyne-azide cycloaddition (CuAAC), so called 'click' reaction, Budisa and co-workers could provide artificially glycosylated protein for lectin binding studies.[113–115]

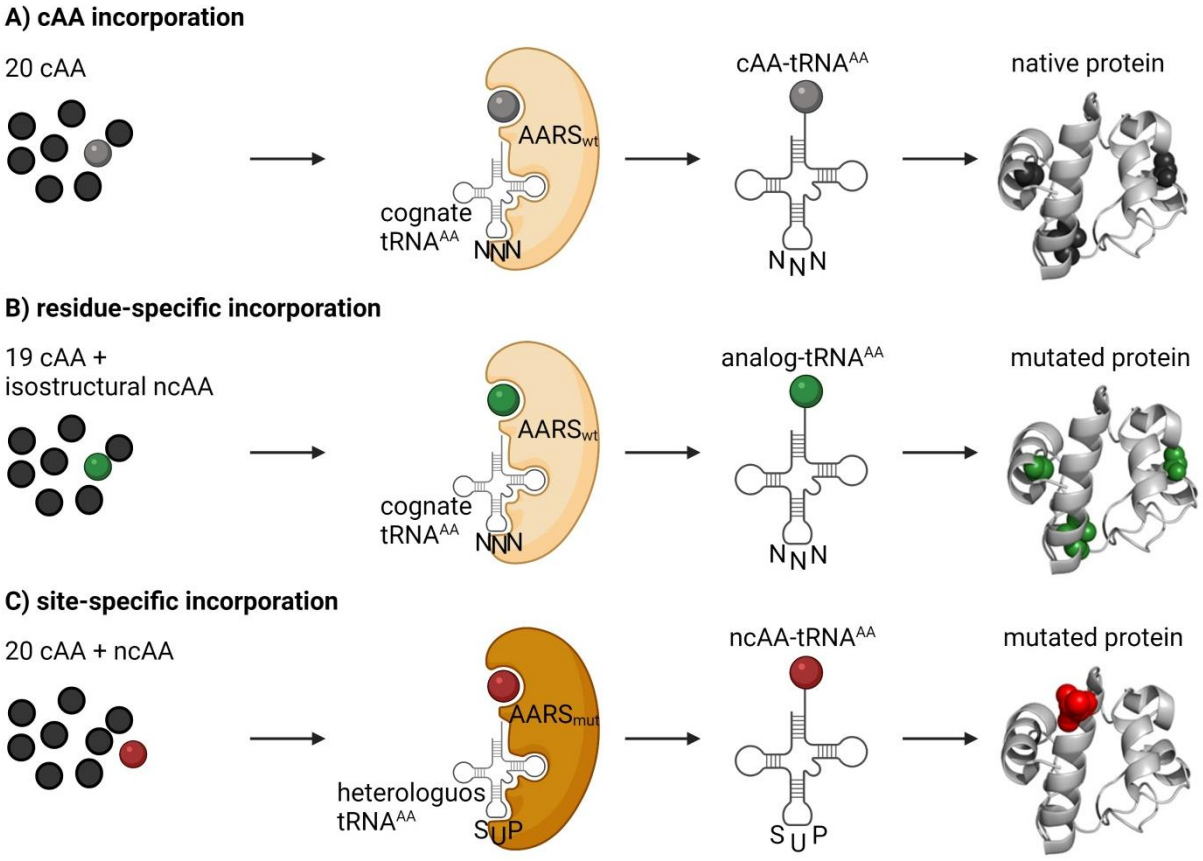


Figure 8 Schematic comparison of SPI and SCS with the native cAA incorporation. Figure created with BioRender.com and modified PDB entry 1HRC.

For the *in vivo* incorporation of the new conjugation moieties, coding units must be generated, either by redefining or by recoding the genetic code. That means on the one hand, cAA can be replaced by isostructural ncAA (genetic code engineering) and on the other hand, that ncAA can be added to the natural repertoire (genetic code expansion).[116,117] Therefore, two different approaches are applied: selective-pressure incorporation (**Section 1.2.3.1**) and stop-codon suppression (**Section 1.2.3.2**) method. A schematic comparison is shown in **Figure 8**.

1.3.1 Selective pressure incorporation method (SPI)

SPI is the simplest popular approach for genetic code engineering. For this purpose, a host strain auxotrophic for the cAA for substitution is cultivated in a chemically defined medium containing all 20 cAA until mid-log phase, at the point where the certain cAA is depleted from the growth medium. Then a bioisostere of the missing cAA is added and the target protein expression is induced. Under these cultivation conditions, the appropriate endogenous AARS with substrate promiscuity charges its cognate tRNA with the chemically non-canonical analogues that are similar in shape, size, and biophysical properties. The proteome-wide residue-specific incorporation of the isostructural ncAA is achieved by using this misacylated analogue-tRNA^{AA} (**Figure 8B**). This codon reassignment technique allows codon-specific modification of proteins and often initiates beneficial synergistic effects on folding, stability, or activity of the target.[118,119] For example, substitution of Trp by 4-aminotryptophan in barstar and Met by norleucine in the cytochrome P450 BM-3 heme domain generates proteins with predefined spectral properties and increased activities.[120,121] To improve the incorporation efficiency, an inducible expression system needs to be tightly regulated, for instance by using inducible strongly repressed promoters.[122] In addition, several engineering approaches can be deployed, e. g. by an increase in concentration of the catalyzing enzyme resulting from an overexpression of endogenous AARS or targeted mutation in AARS active sites.[123–125]

1.3.2 Stop codon suppression method (SCS)

In contrast to SPI, SCS facilitates the site-directed modification of proteins by expanding the proteinogenic AA repertoire by a 21st AA.[126] To encode the new AA in the translational machinery, one of the three stop (nonsense) codons is typically used for reassignment: TAG (*amber*), TGA (*opal*) and TAA (*ochre*). Furthermore, SCS requires an orthogonal engineered suppressor AARS:tRNA$_{sup}$ pair (o-pair), which has no cross-reactivity with endogenous translation components (**Figure 8C**).[127] It is realized, i.e. by (i) taking an heterologous

AARS:tRNA pair from a distant species for redesigning the binding pocket of the AARS in order to aminoacylate only the ncAA rather than the 20 cAA; (ii) modifying the tRNA towards $tRNA_{sup}$ that bears an anticodon complementary to the nonsense codon and must not be recognized by the endogenous AARS.[128] TyrRS:$tRNA_{CUA}^{Tyr}$ pair from *Methanocaldococcus jannaschii*, PylRS:$tRNA_{CUA}^{Pyl}$ pair from *Methanosarcina bakeri* and *Methanosarcina mazei* are frequently used for engineering orthogonal platforms, which are optimally suited for highly efficient TAG suppression in *E. coli* host systems.[129–131] Based on the available crystal structure of the enzyme, residues in the first shell of the active centers are identified that are supposed to interact with the appropriate ncAA.[132] Only these few positions are rationally chosen for randomization by site-saturation mutagenesis to generate libraries bearing high side chain diversity. Subsequently, the resulting AARS mutants are transformed into competent cells containing cognate suppressor $tRNA_{CUA}$ for *in vivo* enrichment of the desired mutants, usually in alternate rounds of positive and negative selections.[133] On an additional plasmid, reporter systems are provided for selection readout. For positive rounds, an essential gene for survival is supplied, e. g. chloramphenicol acetyltransferase (CAT) and a toxic gene for negative rounds, e. g. barnase. Both genes contain in-frame TAG codons at permissive sites to determine suppression efficiency. In positive selection rounds, all cAA and the ncAA of interest are provided to the growth medium containing the selection marker chloramphenicol. Cells expressing AARS mutants that suppress the in-frame stop codons of CAT merely survive, whereas inactive ones lead to cell death and are therefore eliminated from the library. Plasmids harboring the active AARS mutants are isolated from cells and transformed in fresh cells for the following negative selection round. In absence of the ncAA, orthogonal suppressor AARS are selected since AARS constructs aminoacylating endogenous tRNAs and endogenous AARS charging $tRNA_{CUA}$ with cAA enable the toxic protein biosynthesis causing cell death. Promising AARS mutants are isolated from survived cells and employed for further rounds of double-sieve selection. Resulted mutants are used in an additional screening procedure, often requiring a reporter with traceable spectroscopical properties like fluorescent proteins (**Section 1.2.3.3**). This supplementary step provides insights into the efficiency and quantity of enzyme activity as well as to eliminate further false positive mutants. The novel AA sequences of the active sites are obtained via sequencing analysis of the isolated plasmids. By aligning these outcomes, a substantial sequence convergence can be evaluated and then deployed for other ncAA. Modification of the translational apparatus with respect to minimize the competitive reactions by endogenous release factors (RF) is another important effort of engineering of the ribosomal translation system. For instance, the suppression efficiency is significantly enhanced by reduction of the binding affinity or even deletion of RF1, the competitor of the suppressor $tRNA_{CUA}$.[134,135]

Development of a host strain lacking TAG codons or evolving the orthogonal suppressor tRNAs also leads to improvement of the desired SCS activity.[136,137] Over 200 ncAA have been incorporated into recombinant proteins to date by using the randomized active sites of these AARS for nonsense codon reassignment.[138–141]

1.3.3 Sense codon reassignment

In order to create a synthetic organism with an expanded genetic code, sense codon reassignment is a promising strategy that combines the underlying concepts of both aforementioned approaches of protein engineering. A proteome-wide incorporation of ncAA into proteins is implemented in response to degenerated sense codons.[116,142,143] In general, rarely used or 'blank' codons are required to add novel AA to the proteinogenic AA repertoire. Since it is assumed that between 30-40 of the total 61 sense codons are essential for genetic encoding in an organism, over 20 remaining sense codons are dispensable as 'blank' codons in recoding.[144] Besides, suitable codons can be redefined by selective modification of the decoding pathways, in other words by complete liberation of the native functional information of the particular base triplets.[145,146] It is beneficial to choose codons encoding AA that are as little as possible involved in catalytic and metabolic functions to minimize undesirable impact by the residue-specific ncAA insertion. Chemically inert and apolar AA like Leu, Ile, and Val are reasonable choices for this purpose.[145] Further important requirement for reassignment is the introduction of an orthogonal AARS:tRNA pair that can outcompete the endogenous translation machinery in assigning a selected codon.[143] A designed tRNA bears a complementary anticodon for a designated codon. An engineered AARS exclusively activates and loads novel AA on to the corresponding tRNA. The resulting charged ncAA-tRNA is then used for ribosomal translation of proteins containing the desired novel side chain. Sense codon reassignment is a promising strategy and can be exploited for reprogramming cells with respect to expand the protein building block by incorporation of multiple non-natural chemical functionalities.[147] The following Section will address the major background for experimental implementation of this approach in depth.

1.3.4 Towards an efficient ncAA incorporation system

1.3.4.1 Codon for reassignment

The selection of a sense codon amenable to reassignment is crucial for incorporation of ncAA

into proteins. A promising candidate is the rare codon AUA that combines beneficial properties. Along with AUC and AUU, AUA decodes all Ile in bacteria, a chemically inert AA that participates in catalytic and metabolic functions just to a small extent. The two most frequent codons are translated by the high abundant $tRNA^{Ile1}_{GAU}$ containing the GAU anticodon via Watson-Crick base and wobble base pairing, respectively (**Figure 9**). In contrast, the rare AUA decoding process utilized the low abundant $tRNA^{Ile2}_{CAU}$ that bears the anticodon CAU.[145,148] In the first step, C34 of the $tRNA^{Ile2}_{CAU}$ is modified to lysidine (L, 2-lysyl-cytidine) in a two-step reaction catalyzed by the lysidine-tRNA synthetase (TilS, **Figure 10**).[149–151] Lysidine of the resulted $tRNA^{Ile2}_{LAU}$ solely pairs with the third base adenine of the AUA codon, based on the complete conversion of base-pairing specificity (**Figure 9**). Therefore, lysidinylation by TilS plays an important role in maintaining the accuracy of decoding process.[152] The lysidine formation enables the maturation of $tRNA^{Ile2}$ for the aminoacylation by IleRS.[153] Besides, it prevents misacylation by MetRS since the less abundant $tRNA^{Ile2}_{CAU}$ carries the same anticodon CAU like $tRNA^{Met}_{CAU}$ (**Figure 9**). In absence of the C34 modification AUA positions exclusively encode Met using the endogenous MetRS.[154,155] The rare codon AUA occurs at a frequency of about 0.4 % in *E. coli* (**Table 1**), making up 5797 positions of the whole genome.[156] Therefore, using this rare codon for ncAA reassignment would probably not disturb cellular functions.

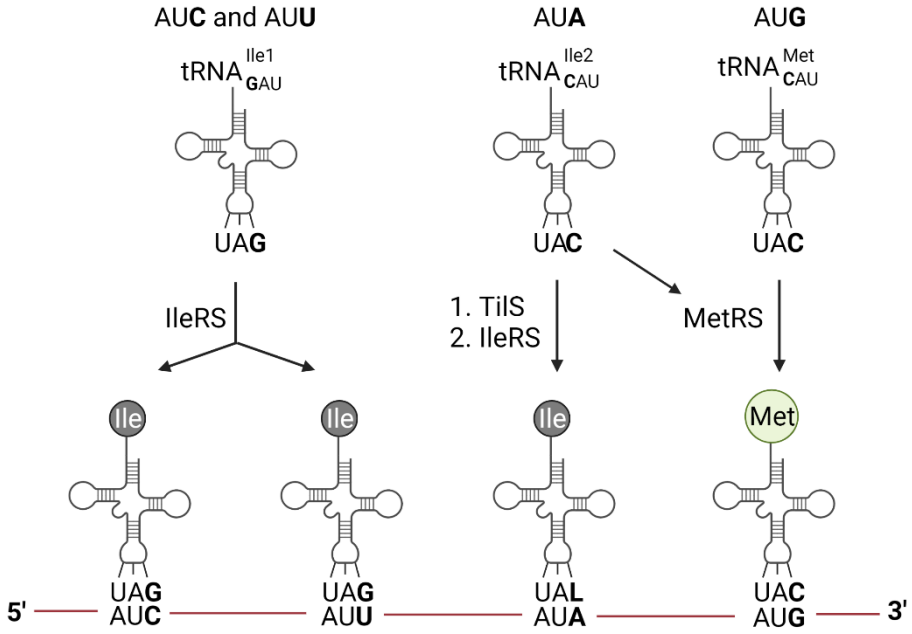


Figure 9 Schematic presentation of Ile decoding routes and potential misacylation by MetRS in *E. coli*. Figure created with BioRender.com.[145]

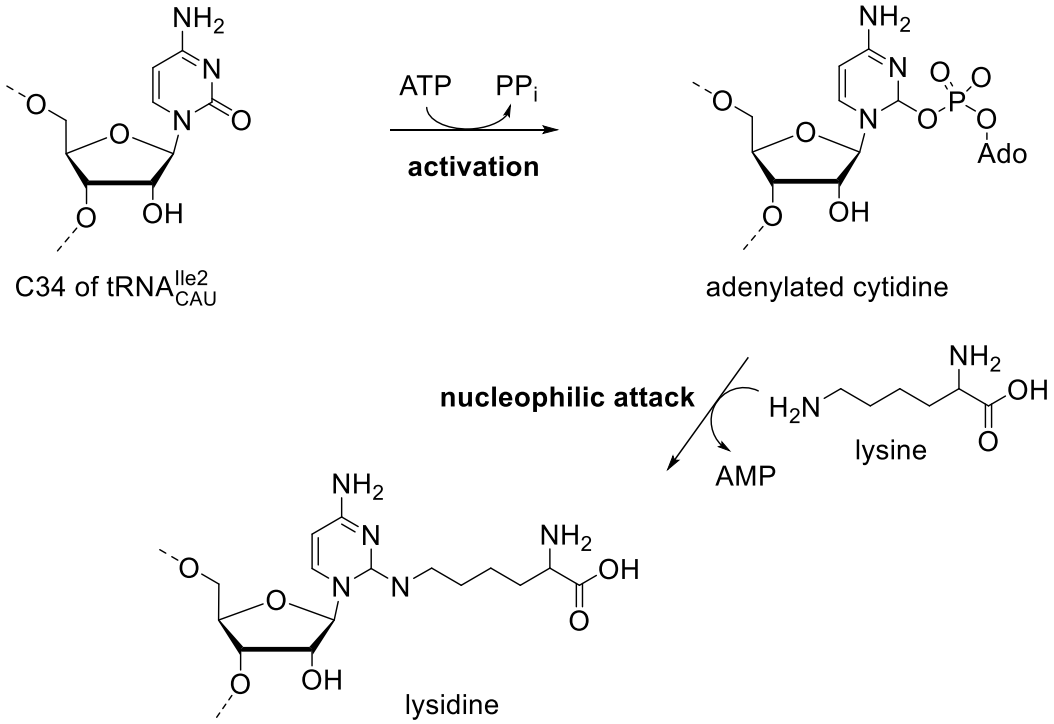


Figure 10 Two-step reaction mechanism of lysidine formation catalyzed by TilS. TilS can discriminate $tRNA^{Ile2}_{CAU}$ from $tRNA^{Met}_{CAU}$. Initially, this enzyme catalyzes the activation of C2 carbonyl group of C34 of only $tRNA^{Ile2}_{CAU}$ using ATP as a substrate. A nucleophilic attack of the adenylated C2 group by the ε-amino group of lysine leads to the release of AMP and formation of lysidine. Figure adopted from Suzuki *et al*. (2010) and created with ChemDraw.[150]

1.3.4.2 Redesign participants of the translational apparatus

As mentioned above, the ribosomal translation system is complex and involves multiple enzymes (**Section 1.2.2)**. The prediction of the ultimate AA sequence for efficient and specified catalysis activity is impossible to make. However, redesigning of key enzymes for incorporation of ncAA into proteins is one step towards solving the problem. The most used component for a systematic directed evolution is AARS.[157] In general, active sites of AARS are engineered based on their known crystal structure and homolog from other organisms, respectively. For example, by introducing a single mutation T415G at the active site of the phenylalanyl-tRNA synthetase (PheRS) from *Saccharomyces cerevisiae* (*Sc*) and using the modified *Sc*$tRNA^{Phe}_{AAA}$, 2-L-napthylalanine (2Nal) could be incorporate into proteins in response to UUU codon.[158] To further enhance the substrate specificity with respect to 2Nal substitute, a library originating from this mutant is created and promising mutants selected by a positive-negative screening system using the well-known green fluorescent protein (GFP) in fluorescence-activated cell sorting (FACS).[159] For this purpose, four residues within the substrate binding pocket of *Sc*PheRS are randomized. The resulted 10^6 membered library is

transformed into cells expressing GFP, with the particularity that 12 Phe are encoded by UUU. Assigning of cAA other than Phe at appropriate positions implement a 20-fold reduction of fluorescence. For a positive screening, the expression of GFP and *Sc*PheRS library are induced in presence of 2Nal and allow separation of mutants that efficiently activate Phe from those that charge other cAA and 2Nal. Dimly fluorescent cells are then grown for the negative screening in which expression of proteins is induced without 2Nal. This time strongly fluorescent cells are enriched by FACS, since it is expected that highly selective mutants towards 2Nal will not catalyze any cAA misincorporation except Phe at designated sites in GFP. Following a further round of positive screening, promising clones are selected for sequencing and AA activation analysis. Although an engineered *Sc*PheRS mutant suffers a significant loss in the activity, a remarkable enhancement in selectivity towards 2Nal could be achieved in comparison to starting *Sc*PheRS(T415G).

One further step towards ncAA incorporation is the selective modification of tRNA. In general, the AC-stem loop is the favorite target for this purpose. Especially position 34, the first base of the anticodon of the tRNA, is known to have one of the highest modification rates (**Section 1.2.2.1**). Besides, position 35 and 36 can also be rationally mutated with respect to ncAA incorporation.[160,161] For instance, by changing the anticodon CCA of *E. coli.* $tRNA^{Trp}_{CCA}$ to UCA the novel functioned adapter molecule $tRNA^{Trp}_{UCA}$ can be obtained. In combination with liberated *E. coli* TrpRS this suppressor $tRNA^{Trp}_{UCA}$ was efficiently used in TGA suppression for site-specific incorporation of several Trp analogues in HEK293T cells as well as *E. coli*.[161] Construction of a tRNA library is another opportunity to improve the ncAA charging activity. Based on the known crystal structure of the tRNA of interest, positions excluding identity elements are preferably chosen for randomization. When applying the adapter molecule in nonsense suppression, in which a general enrichment of a desired mutant is performed *in vivo* using alternate rounds of double-sieve selection.[133]

1.3.4.3 Fluorescent proteins – universal reporters

A useful tool for rapid evaluation of ncAA incorporation efficiency and fidelity during and after protein translation are reporter genes.[162] Usually, molecules owning fluorescent property are applied since the initial screening and quantitative characterization of the ncAA incorporation procedure are easy to observe.[163] By far the most common reporter is green fluorescent protein (GFP) and its engineered derivates.[164] The main advantage of these proteins is their robust structure once folded and the range of wavelength at which they can be tracked.[165] The discovery of GFP dates to 1961, as a by-product during the isolation of aequorin, a

luminescent substance, from the jellyfish *Aequorea victoria*.[166,167] Since then, this 238 AA protein gains in importance due to its versatile applicability in protein engineering, biophysical and biomedical studies.[168–175] GFP's unique β-barrel structure consists of eleven β-strands surrounding a single central α-helix that bears the chromophore (**Figure 11A**).[176] Both ends of the β-barrel scaffold are capped by short distorted helical segments to protect the central element from bulk solvent. The resulting tertiary structure equals a nearly perfect cylinder with a diameter of 24 Å and 42 Å in length (**Figure 11A**).

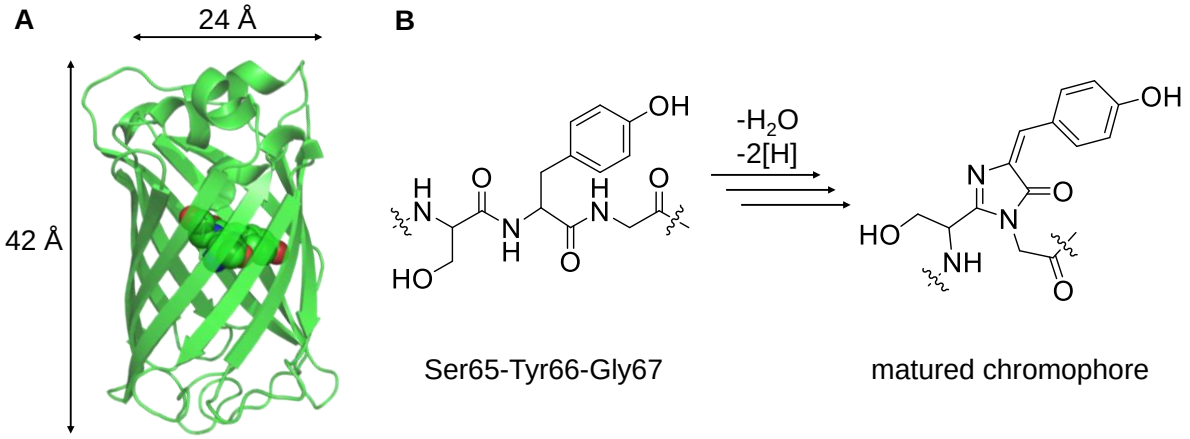
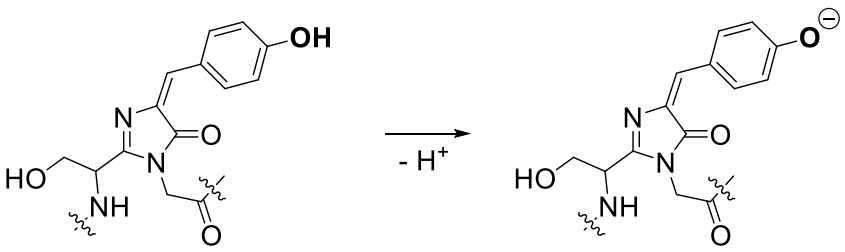


Figure 11 Structure of enhanced green fluorescent protein from *Aequorea victoria* (A) and the summarized auto-catalytic maturation of a tyrosine-based chromophore (B). Crystal structure modified from PDB entry 2Y0G and chemical equation created with ChemDraw.

The strong visible absorbance and fluorescence of GFP originates from the intrinsic chromophore in the core, a *para*-hydroxy-benzylidene-imidazolinone derivative, which is formed from an autocatalytic reaction of the backbone in three distinct kinetic steps (**Figure 11B**).[177,178] Starting from the tripeptide Ser65-Tyr66-Gly67 in an extended conformation, a covalent rearrangement occurs and promotes a cyclization reaction at the cost of one water molecule (dehydration). In the third and slowest phase of chromophore maturation the cyclized imidazole-5-on intermediate oxidizes, resulting in a visually fluorescent chromophore (**Figure 11B**). Once it is formed, the matured chromophore tightens the surrounding β-barrel scaffold by noncovalent interactions and therefore additionally stabilizes the overall protein structure.[179] Naturally occurring GFP absorbs light strongly at 395 nm and about two times weaker at 475 nm. The absorption maxima correlate with the two states of the chromophore: neutral and anionic (**Figure 12**).[179] The protonation states exist in equilibrium controlled by the internal hydrogen bonding network.[180] Excitation at particular wavelengths lead to green fluorescence emission with the maximum at 508 nm.



neutral chromophore anionic chromophore

Figure 12 Protonation states of the matured tyrosine-based chromophore in GFP. Chemical equation created with ChemDraw.[179]

The matured chromophore in its isolated form lacks fluorescence, pointing out the importance of tightly packed chromophore in the β-barrel scaffold with regard to the fluorescent property.[181,182] It also means that the chromophore in GFP is intrinsic to the primary structure. Based on this knowledge, the chromophore and its environment have been extensively engineered to develop a wide range of fluorescent proteins.[183] For instance, introducing two point mutations in the chromophore and its neighboring surrounding: Ser65T and F64L yield a variant with 35-fold improvement in brightness, increased solubility and stability as well as a faster post-translational oxidation reaction of chromophore maturation compared to the ancestral protein.[184] This mutant was named as enhanced green fluorescent protein (EGFP).[184,185] In order to obtain GFP variants with different colors for wider application, chromophore Tyr66 is substituted by histidine (His) and Trp, respectively.[186] By this direct modification of the core a blue shift in excitation and emission maximum could be observed. Novel GFP mutants with multiple colors were yielded and named according to their color, e.g. blue fluorescent protein (BFP) or cyan fluorescent protein (CFP).[177,187] Further mutations in the chromophore environment of these colored variants and substitution of cAA with ncAA lead to a great number of proteins with enhanced properties and distinct fluorescent profiles.[188–193]

The remarkable class of fluorescent proteins are utilized not only as reporters of gene expression but also as fusion proteins for mechanism studies or selectable markers for protein localization and folding as well as biosensors and probes for protein-protein interactions.[194–198] Experimental implementation is simple due to the wide availability of spectroscopical equipment in most laboratories. For high throughput screening, FACS and a microtiter plate reader, capable of recording fluorescence and absorption over a long-time range, are often deployed. Using GFP as a reporter enables *in vivo* monitoring of cell activities since its fluorescent property is not sensitive to the environment once the matured protein is formed.[183]

1.4 Protein structure

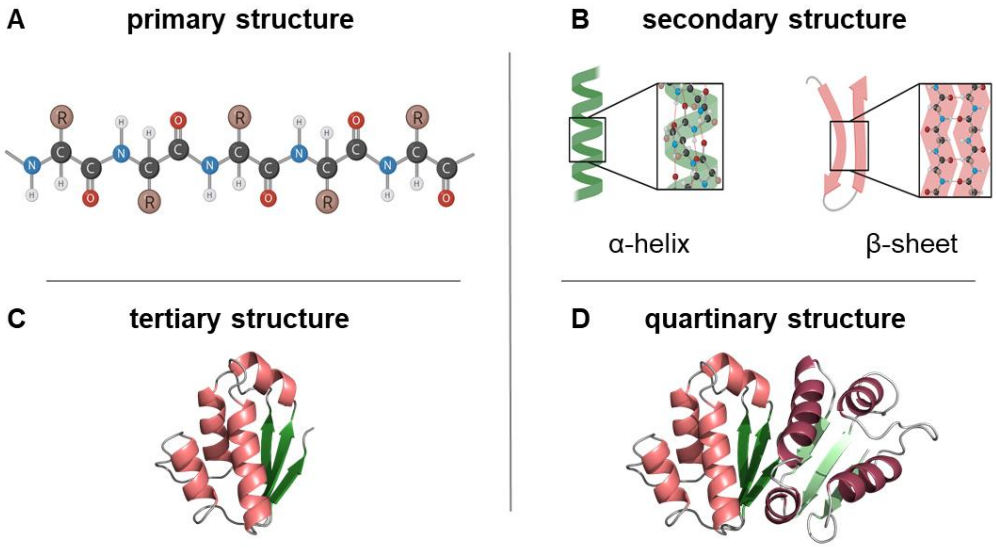


Figure 13 Types of protein structures: primary (A), secondary (B), tertiary (C), and quaternary structure (D). Figure created with BioRender.com using modified structure from PDB entry 1A19.

Protein function is determined by its structure which includes four distinct levels: primary, secondary, tertiary, and quaternary structure (**Figure 13**).[199] The linear sequence of AA known as primary structure is unique for each protein and given by their coding gene according to the central dogma of molecular biology (**Figure 13A**).[9] AA are linked together by the amide groups, so-called peptide bonds. The amino terminus (N-terminus) and carboxyl terminus (C-terminus) make up the two free ends of the polypeptide chain (**Figure 13A**). After synthesis of AA sequences at the ribosome (**Section 1.2.2**), extended chains are arranged into three-dimensional (3D) compact globular entities consisting of one or more individual protein molecules (subunits): tertiary and quaternary structures, respectively (**Figure 13C, D**). In addition, secondary structures, highly regular local sub-structures, form domains within these arrangements (**Figure 13B**). The most common types are α-helix and β-sheet (β-strand) that represent the key structural features of proteins and are stabilized by hydrogen bonding networks (**Figure 13B**).[200] The energetical driving force behind this folding process is the hydrophobic effect, which describes the tendency to minimize the number of hydrophobic residues exposed to solvent water due to decrease in solvation free energy.[201] The 3D structure of proteins therefore usually possesses a hydrophobic core and a charged hydrophilic surface.[202] Since there are several energetically favorable conformations, protein structures are dynamic and fluctuate between these similar folds. Besides hydrogen bonding (non-covalent interaction), covalent bonds (disulfide bridges), Van der Waals forces, ionic, and

hydrophobic interactions contribute as well to the stabilization of protein structure.[203,204] Only intact folded proteins can retain their functions. Protein misfolding potentially results in highly organized fibrillar aggregates that might cause diseases like Alzheimer or Parkinson.[205,206] Therefore, cells often recruit chaperones as an essential assistant for proper protein assembly and folding.[207]

To understand the mechanism of protein function, the native protein structure must be solved. Common methods include cryo-electron microscopy (cryo-EM), nuclear magnetic resonance spectroscopy (NMR), and X-ray crystallography.[208–211] Even though each technique has its own advantages and disadvantages, the latter is the most applied one for accurate creation of the atomic model. For this purpose, proteins are purified and crystallized in an aqueous environment using vapor diffusion, microbatch, microdialysis, and free-interface diffusion method, respectively.[212] Disordered soluble protein is unsaturated and is required to be converted to the crystalline ordered state in supersaturated region, at which the protein concentration is above its solubility limit (**Figure 14**). Here, a distinction is made between three sub regions: lower (metastable zone), intermediate (labile zone), and very high supersaturation (precipitation) (**Figure 14**). Ideally, a molecule forms a single crystal in case it first reaches the labile zone, where both nucleation and a slow growth process are promoted, and subsequently enters the metastable zone for proper crystal growth.[213] In practice, it needs

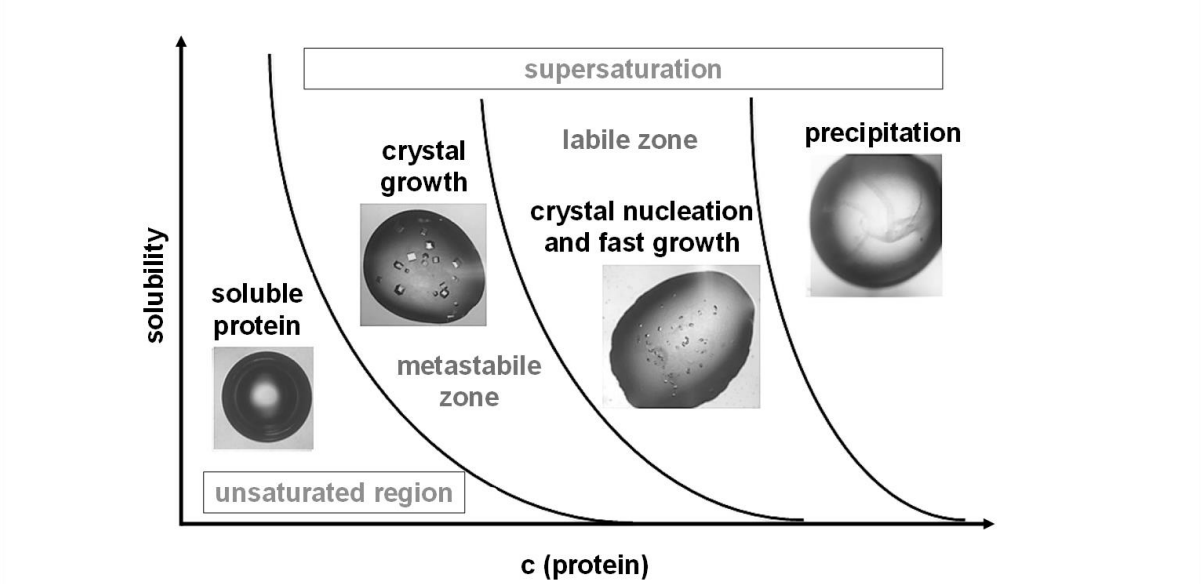


Figure 14 Schematic presentation of the stable states (liquid, crystalline, precipitation) of protein as a function of crystallization variables (solubility and concentration).
Figure adapted from Russo *et al*. (2013).[208]

many approaches with systematic variation of numerous parameters, for instance precipitant type, concentration, temperature, pH, and additives, to obtain the optimal condition. A variety of possible initial screening conditions rises experimentally that needs to be further optimized

and upscaled for high quality crystals. Single crystals in a range of 20-50 nm are sufficient for the following high-resolution X-ray analysis.[214] Protein crystals diffract the X-ray beam into a characteristic pattern, from which distribution of electrons within this biomolecule can be evaluated. The location of each atom is then determined by interpreting the resulting electron density map. The resulting detailed biological macromolecular crystal structure is published in Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB), the currently biggest archive of 3D structures of proteins, nucleic acids, and complex assemblies.[215,216]

2. Aim of this work

The major goal of this study is to evolve a ribosomal translation system towards reassignment of the Ile rare codon AUA to small aliphatic fluorinated amino acids. For this purpose, key participants in protein biosynthesis are selected: IleRS and $tRNA^{Ile}$ from *E. coli*. Engineering of these involved parties includes the crystallization of IleRS, the generation of IleRS libraries, and modification of the $tRNA^{Ile}$ anticodon. Furthermore, a suitable screening strategy for selection of the active IleRS mutant has to be developed. This requires, amongst other things, a 'blank' codon for reassignment, a convenient strain and appropriate reporter for detection as well as characterization of expression outcomes.

In addition, reporter proteins are investigated for enhancement of useful properties. Therefore, different ncAA were introduced into several β-barrel fluorescent protein variants. The useful tools of protein engineering SPI and SCS approaches are applied for incorporation, respectively. Resulted products are subsequently analyzed by spectroscopic methods and employed for further applications.

3. Results and discussion

3.1 Crystallization of IleRS

3.1.1 Production of substrates

(2*S*)-2-aminobutanoic acid (Abu) and its γ-fluorinated derivatives (2S)-2-amino-4-monofluorobutanoic acid (L-monofluoroethylglycine, MfeGly), (2S)-2-amino-4,4-difluorobutanoic acid (L-difluoroethylglycine, DfeGly), (2S)-2-amino-4,4,4-trifluorobutanoic acid (L-trifluoroethylglycine, TfeGly) were extensively investigated in the Koksch laboratory in the context of peptide and protein environment (**Figure 15**). They showed that the impact of fluorination on key properties of the resulting artificial biological polymers depends on the nature of fluorinated side chain as well as on the molecular interactions within the designed system.[55] To be more specific, polar character is introduced into otherwise hydrophobic AA side chain by small numbers of fluorine atoms.[217] Folding of modified biopolymers is altered due to the change in the secondary structure propensity of fluorinated aliphatic AA.[54,218] Furthermore, fluorine can cause a significant effect on the proteolytic stability of peptides and increase resistance to denaturation of proteins by organic solvents, respectively. [56,219,220]

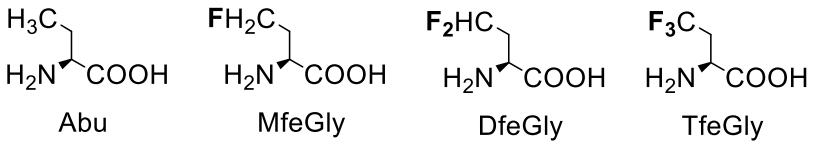


Figure 15 Chemical structures, IUPAC names and common synonyms of ncAA relevant for this study. (2*S*)-2-aminobutanoic acid (L-homoalanine, Abu) and three of its fluorinated derivatives (2*S*)-2-amino-4-monofluorobutanoic acid (L-monofluoroethylglycine, MfeGly), (2*S*)-2-amino-4,4-difluorobutanoic acid (L-difluoroethylglycine, DfeGly), (2*S*)-2-amino-4,4,4-trifluorobutanoic acid (L-trifluoroethylglycine, TfeGly).

Fluorinated analogues of Abu are activated by *E. coli* IleRS, whereby TfeGly has shown the best kinetic values (**Figure 15**).[221,222] To gain insights into the intra- and intermolecular interactions of small aliphatic ncAA within the active centers of *E. coli* IleRS, X-ray crystallographic experiments were conducted. The structural information could be used for rational enzyme engineering or for targeting specific residues in directed evolution. Moreover, the crystal structure would allow insights into the mode of activation and editing of the ncAA. In this study we concentrated on crystallization of *E. coli* IleRS in complex with TfeGly that was synthetized by colleagues from the Koksch laboratory.

For protein expression, an available pQE-80L plasmid carrying the synthetase gene was taken

from the plasmid database of the Budisa group (**Figure 16**). The protein bears a N-terminal His-tag (H_6) for purification followed by a *tobacco etch virus* protease (TEV) recognition site for subsequent removal of the tag (**Section 6.3.1, 6.3.5**).[223] Expression was successfully performed in *E*. *coli* strain BL21 (DE3) at 28 °C for 6 h, using 1.0 mM IPTG for induction (**Section 6.2.5**).

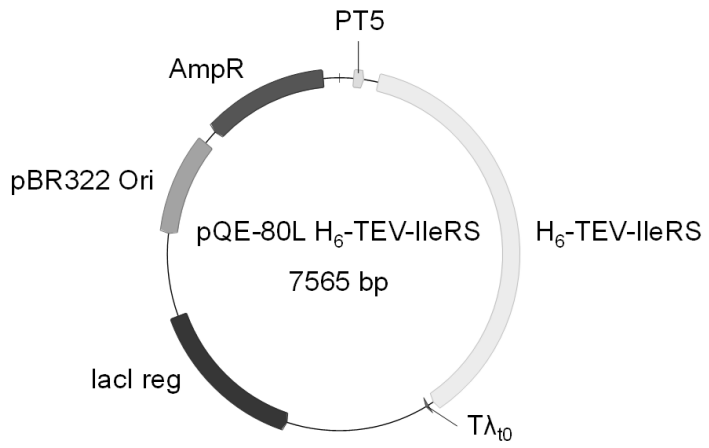


Figure 16 Map of expression plasmid pQE-80L H_6-TEV-IleRS. The bacterial expression vector pQE-80L carries the gene of interest (H_6-TEV-IleRS) under control of the inducible T5 promotor (PT5). Ampicillin resistance is used to select cells containing the desired plasmid. Further features of this vector are λ_{t0} terminator ($T\lambda_{t0}$) for enhanced plasmid and mRNA stability, *lacI* regulatory gene (lacI reg) for transcription regulation, and pBR322 origin of replication (pBR322 Ori), that initiates replication of plasmid in expression host.

Overexpression of target protein could be detected by SDS-PAGE gel analysis (**Figure 17A**). We first purified H_6-TEV-IleRS by immobilized metal anion chromatography (IMAC) according to **Section 6.3.1**. Eluted fraction was analyzed by LC-ESI MS to identify the molecular weight (**Figure 17D**). While the calculated molecular weight of H_6-TEV-IleRS ($M_{W,\ calc.}$) is 106.963 kDa, a peak maximum was found at 106.979 kDa. The positive mass shift of 16 Da can be assigned to a methionine oxidation within the full-length protein. Major band at expected size confirmed the presence of H_6-TEV-IleRS on the SDS-PAGE gel (**Figure 17B**). Unfortunately, there are two minor unassigned band in protein mixture that need to be removed prior to crystallization events. Hydrophobic interaction chromatography (HIC) was then applied in order to improve the protein purity (**Section 6.3.2**). Unwanted compounds were still present in the eluted fraction indicating a similar hydrophobicity compared to the desired product (**Figure 17B**). In a further separation attempt, ion exchange chromatography (IEX) was applied (**Section 6.3.3**). To analyze purified proteins more sensitively, the SDS-PAGE gel from the third purification step was silver stained (**Section 6.4.3**). Finally, we could separate impurities from the target protein based on their stronger affinity to the anion exchanger (**Figure 17C**).

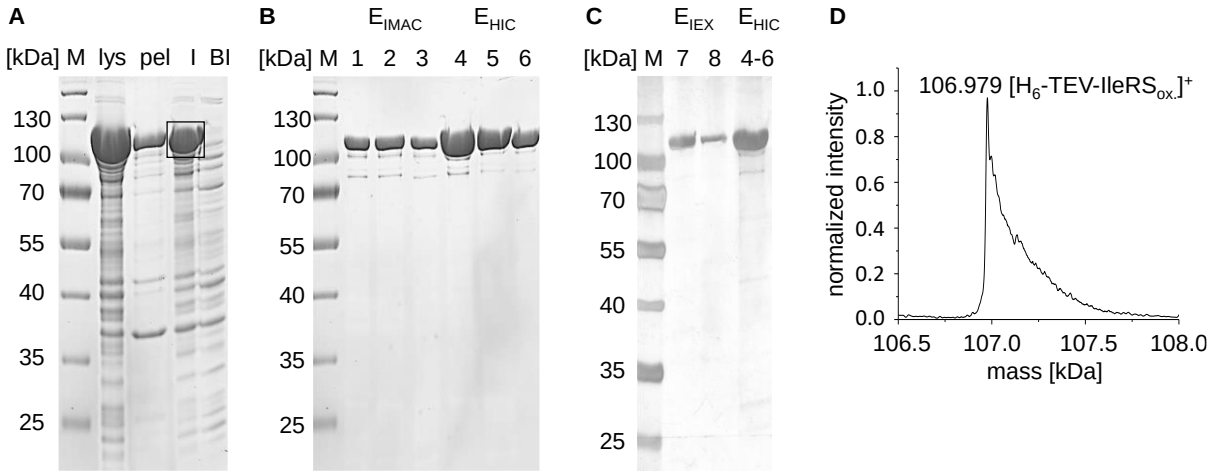


Figure 17 Analysis of expressed H_6-TEV-IleRS in *E. coli* BL21 (DE3) by 12 % SDS-PAGE gel (A-C) and LC-ESI MS (D). (**A**) Before (BI) and after induction (I) fractions from standard gene expression as well as insoluble (pel) and soluble (lys) samples after cell lysis are applied for analysis. Black frame indicates the overexpression of protein. The calculated molecular weight of H_6-TEV-IleRS ($M_{W, calc.}$) is 106.963 kDa. (**B**) Eluted fractions from immobilized metal anion chromatography (E_{IMAC}) are 1-3 and from hydrophobic interaction chromatography (E_{HIC}) 4-6. Impurities after purification are smaller bands below the prominent target product bands. (**C**) The purity of elutions from ion exchange chromatography purification (E_{IEX}) 7 and 8 are significantly increased in comparison to pooled $E_{HIC,4-6}$. SDS gels from (**A**) and (**B**) are Coomassie stained, whereas the gel shown in (**C**) is silver stained for higher detection sensitivity. The same protein marker (M) is used for (**A-C**). (**D**) Deconvoluted mass of purified H_6-TEV-IleRS ($M_{W, found}$) 106.979 kDa shows a difference of 16 Da, that can be assigned to one Methionine oxidation of H_6-TEV-IleRS.

However, the protein stability may suffer after three purification steps, that take at least a week. Since the crystallization process in general occupies a lot of time, we should provide a freshly prepared sample for improved success rate. We assumed that the side products are artefacts from expression process and optimized the expression conditions to reduce the purification steps. The parameters of choice are on the one hand the concentration of inducer and on the other hand the duration time of expression due to their simple implementation. In this part, isopropyl-β-D-1-thiogalactopyranoside (IPTG) is used for induction of the expression process, which is known to cause appreciable damage to the host cell.[224] Therefore, the concentration of IPTG should be effectively tuned for each protein production to alleviate this negative side effect. The second parameter is chosen for variation based on the assumption that slight bands ranging from approximately 80 to 100 kDa are degraded forms of the synthetase (**Figure 17B**). First, the impurities of the elution fractions possess similar hydrophobicity compared to the desired protein since it could not be separated from the mixture by HIC. It can be concluded that these are truncated products of the target protein. Second, the longer expressed recombinant proteins remain in the cellular environment, the higher the probability

of proteolysis. Although expression host *E. coli* BL21 (DE3) is deficient in the cytoplasmic Lon protease, others like ClpAP are still present and tend to cleave overexpressed protein at multiple sites.[225,226] Thus, we need to adjust the duration of target gene expression. 0.5 and 1 mM IPTG were tested for induction of protein biosynthesis. Cells were then harvested at distinct time points and used for analysis by SDS-PAGE. The results of optimization experiments show slight increase of H_6-TEV-IleRS yield when induced with higher concentration of IPTG (**Figure 18A**). However, the number of unwanted side products in the area of interest rises coincidentally (**Figure 18A**, black frame). We observed the same trend for the expression duration: side products increase with extended time. On this account we decided to use 0.5 mM IPTG and express protein in 4 h. After the purification by IMAC we obtained H_6-TEV-IleRS samples with negible side products (**Figure 18B**).

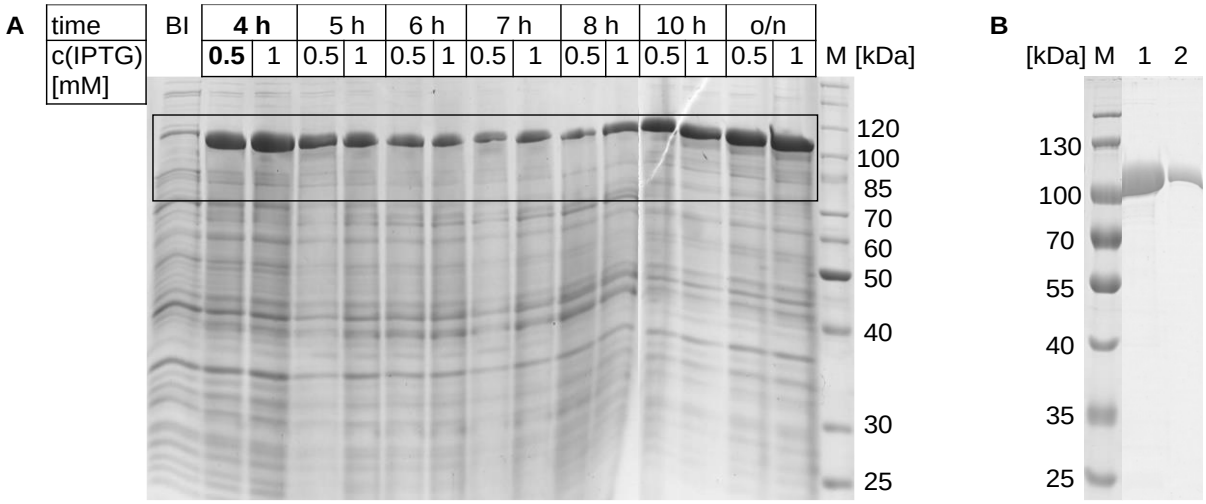


Figure 18 Fine tuning of expression conditions (A) and purity H6-TEV-IleRS expressed in *E. coli* BL21 (DE3) with optimized parameters (B). (**A**) Samples after each duration time were taken and normalized to OD_{600} = 1 prior to analysis. Black frame highlights the area of interest. Protein expression induced by 0.5 mM IPTG shows best result after 4 h (conditions shown in bold). For comparison, sample before induction (BI) is loaded to emphasize the overexpression at different time point (bold bands at ~107 kDa). (**B**) Two eluted fractions from optimized expression after IMAC (1 and 2). Both analyses were performed with 12 % Coomassie stained SDS-PAGE gels. Different markers (M) were used: unstained (**A**) and prestained protein ladder (**B**).

In the next step, we removed H_6-tag from the N-terminus of the synthetase by TEV protease according to **Section 6.3.5** prior to crystallization (**Figure 19**).Therefore, TEV protease was expressed using pET28a-H_6-TEV plasmid (**Figure 19A**). 6 mg purified TEV were yielded from 1 L cell culture (**Figure 19B**). H_6-tag was efficiently cleaved and removed from the reaction mixture by IMAC. Resulted product was analyzed by LC-ESI MS (**Figure 20**). For comparison, expressed synthetase prior to the TEV-cleavage was also submitted for mass analysis. Again,

the detected $M_{W,\,found}$ (H_6-TEV- IleRS) possesses +16 Da difference compared to the $M_{W,\,calc}$ (H_6-TEV-IleRS), while $M_{W,\,calc.}$ (IleRS) of 104.353 kDa after tag-removal is in accordance with its observed mass (104.355 kDa) within the error range of our instrument. Hence, we assumed that the N-terminal start Methionine is oxidized, which in turn did not impede the TEV protease recognition and allowed the complete cleavage of the purification tag.

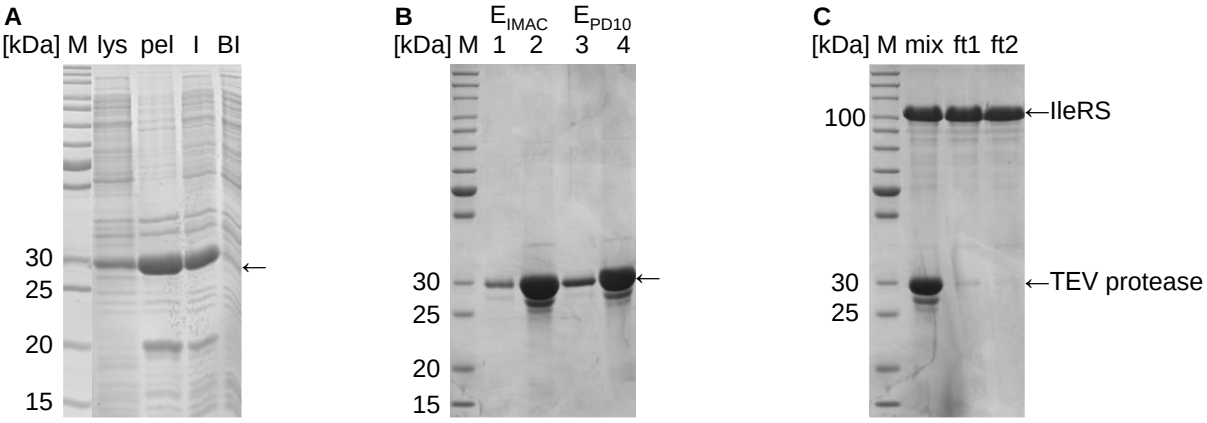
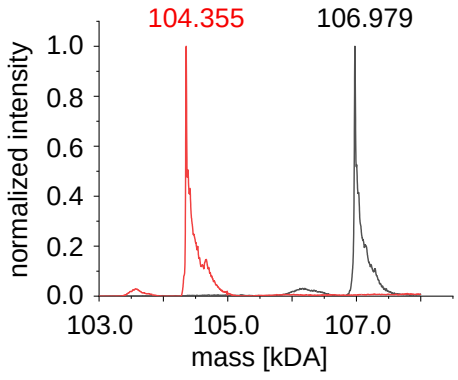


Figure 19 Analysis of H_6-TEV protease expression in *E. coli* BL21 (DE3) gold (A), purification (B) and overnight cleavage results (C) by 12 % SDS-PAGE. (**A**) Before (BI) and after induction (I) fractions from standard gene expression of TEV protease as well as insoluble (pel) and soluble (lys) samples after cell lysis are applied for analysis. $M_{W,\,calc.}$ (H_6-TEVprotease) is 28.705 kDa. Overexpression could be observed at expected size (arrow). (**B**) 1 and 2 are eluted fractions from IMAC (E_{IMAC}); 3 and 4 are isolated fractions after desalting column PD10. Protein mass was verified by LC-MS. Calculated size of target protein is marked with arrow. (**C**) H_6-tag cleavage from H_6-TEV-IleRS was performed overnight at 4 °C (**Section 6.3.5**). The incubated mixture (mix) as well as flowthrough fractions (ft1, ft2) were then loaded on SDS-PAGE gel. $M_{W,\,calc.}$ (H_6-TEV-IleRS) is 106.963 kDa. The same marker (M) was used for (**A**-**C**).



protein	$M_{W,\,calc.}$ [kDa]	$M_{W,\,found}$ [kDa]
H_6-TEV-IleRS	106.963	106.979
IleRS	104.353	104.355

– H_6-TEV-IleRS
– IleRS

Figure 20 LC-ESI MS spectra of purified H_6-TEV-IleRS and IleRS. While the deconvoluted mass found for IleRS 104.355 kDa corresponds to $M_{W,\,calc.}$ (IleRS) 104.353 kDa, observed mass of H_6-TEV-IleRS exhibits an addition of 16 Da.

Subsequent purification via size exclusion chromatography (SEC, **Section 6.3.4**) yields fractions containing monomeric IleRS that were pooled and concentrated. **Figure 21** shows the chromatogram and the corresponding result from SDS-PAGE analysis. Concentration of about 50 mg/mL could be determined by the Lambert Beer law (**Section 6.4.1.2**). Protein was directly used for crystallization. Residual substance was dispensed in 20 µL aliquots, flash frozen in liquid nitrogen and stored at -80 °C.

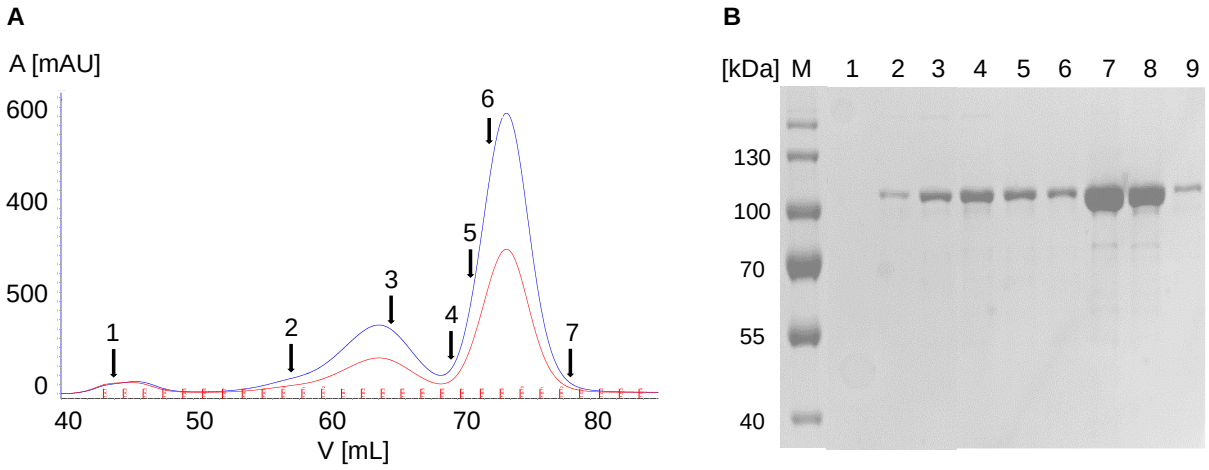


Figure 21 Analysis of IleRS purification by SEC. (**A**) Chromatogram of elution step shows several peaks implying different components within the sample. Blue curve represents the absorption detected at 280 nm (A_{280}) and red curve at 260 nm (A_{260}). The A_{260}/ A_{280} ratio of about 0.6 at the main peak indicates the presence of pure protein sample. The first maximum (1) can be associated to aggregation products, the second to oligomers (2), the third to dimers (3) and the fourth to monomers (4-7). (**B**) Corresponding 9 % SDS-PAGE gel of selected SEC elution fractions 1-9 are applied with marker (M). Observed bands at ~100 kDa in samples 2-9 confirm the former assignment ($M_{W, calc.}$ (IleRS) = 104.353 kDa). Since the samples were treated with SDS prior to loading, oligomeric state of proteins cannot be detected.

In summary, we succeeded to optimize the expression and purification of IleRS mutants for crystallization purpose. A complete protocol for synthesis of protein with high purity could be established and used for large scale production of the desired enzyme.

3.1.2 Screening and optimization

We first set up screening trials with TfeGly as a ligand, H_6-TEV-IleRS and IleRS, respectively. This will indicate whether the cleavage of N-terminal H_6-tag is required prior to crystallization process. From the entire 1344 screening conditions, no promising formation of crystal could be observed for H_6-TEV-IleRS, but five conditions with IleRS yield some crystalline materials that are good starting points for further refinement (**Figure 22**). Therefore, it is clear that the

N-terminal H_6-tag disturbs the crystallization process and needs to be cleaved.

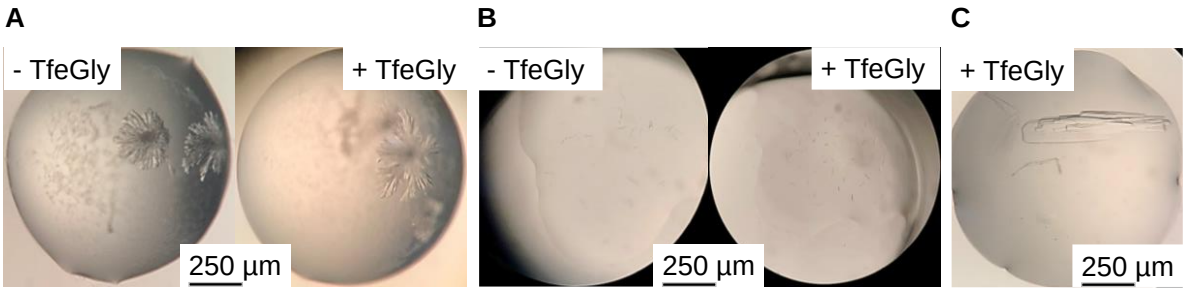


Figure 22 Droplets of promising conditions containing protein-like crystals in a 96-well sitting drop microplate. (**A**) Hampton Index: 0.1 M BIS-TRIS, pH 6.5, 0.2 M $MgCl_2$, 25 % w/v PEG 3350, (**B**) JBScreen JSGC++: 100 mM TRIS, pH 8.0, 40 % v/v 2-Methyl-2,4-pentanediol, (**C**) JBScreen Basic HTS (100 mM MES, pH 6.5, 200 mM $Mg(OAc)_2$, 20 % w/v PEG 8000). For each trial, c (IleRS) = 20 mg/mL and c (TfeGly) = 7.5 mM were used, and screening plates were stored at 18 °C. Yielded crystals were in the approximate dimensions of 50 µm (**A**, **B**) as well as about 600 µm (**C**). Pictures were taken after 4 weeks.

A diffraction test measurement revealed that the crystal obtained in JBScreen Basic HTS condition (IleRS_screen_1) showed protein-like diffraction pattern that was then selected for the rational optimization in order to produce diffraction-quality protein crystals. The concentration of protein and precipitant, as well as pH were varied to fine-tune the conditions (**Section 6.3.6.4**). Upon scale-up to a 24-well plate, expected high-quality crystals remained absent. A crucial reason of this failed attempt is the modified kinetics of the crystallization process compared to the small scale environment that is induced by several factors.[227] These include the tenfold increased volume of the sample droplet, changing the method from sitting drop to hanging drop vapor diffusion and different materials of the surface, where the crystallization proceeds (**Section 6.3.6**).

To reduce the aforementioned modifications in experimental performance during the optimization procedure, 48-well plate for sitting drop vapor diffusion method was applied with fivefold increased volume. In this context, the protein droplet is also in contact with a polystyrene surface as in the case of screening well. For the same protein concentration as used for initial trial (c (IleRS) = 20 mg/mL), crystals could only be reproduced at the identical composition of crystallization mixture (**Figure 23A**). Whereas applying less concentrated protein sample (c (IleRS) = 15 mg/ml) yielded considerable crystals with sharp edges at lower precipitant level, 17 % w/v PEG 8000 (**Figure 23B**). However, obtained crystals are intertwined and another screening method should be applied to increase the numbers of initial hits. Since lower concentration of protein resulted in more crystals, subsequent experiments were performed with c (IleRS) = 15 mg/mL.

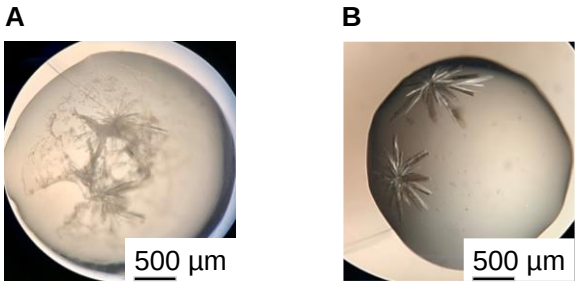


Figure 23 Experimental outcome from optimization of IleRS_screen_1 condition. (**A**) IleRS_opt_1 (100 mM MES, pH 6.5, 200 mM $Mg(OAc)_2$, 20 % w/v PEG 8000, c (IleRS) = 20 mg/mL), (**B**) IleRS_opt_2 (100 mM MES, pH 6.5, 200 mM $Mg(OAc)_2$, 17 % w/v PEG 8000, c (IleRS) = 15 mg/mL). For each trial c (TfeGly) = 7.5 mM were used, and screening plates were stored at 18 °C. Yielded crystals were in the approximate dimensions of 600 µm. Pictures were taken after 4 weeks.

Micro seeding was the next screening method we have used (**Section 6.3.6.3**). As indicated in the concentration-dependent phase diagram (**Figure 14**), nucleation is the essential start of crystallization (**Section 1.4**). Nevertheless, this first step in the labile zone can be evaded by directly introducing e.g., already-grown small crystals of the same protein into the metastable region for crystal growth. Therefore, crystals obtained from optimization trials were employed as ready-made nuclei (**Figure 23**). Five promising hits were then applied for further optimization in 48-well plates. Unfortunately, diffraction patterns later revealed that even nicely grown single crystals can be characteristically assigned to salt crystals. It is known, that seeds often dissolve in solutions and are protein-specific.[228,229] The assuming crystals of IleRS are possibly not suitable as the 'ultimate' nucleant in our case. Hence, further screening of seeds is needed which would outgrow the time frame of this study.

It should be noted that IleRS contains several active centers (**Figure 24**) and therefore, possesses intrinsic molecule's flexibility due to its potential conformational dynamic.[20,230] To increase the protein rigidity, we either supplied all possible ligands of IleRS binding sites or used one short independently folded domain of IleRS. Implementation of the former attempt is achieved by adding non-hydrolyzable substrates to crystallization mixture: adenosine-5'-[(β,γ)imido] triphosphate (AMP-PNP) as ATP substitution; TfeGly for the activation site; 2'-(L-trifluoroethylglycyl) amino-2'-deoxyadenosine (TfeGly-2AA) as charged tRNA analogue for the CP1 domain (**Figure 24**). Since TfeGly is not available in large excess, we first synthetized 2'-(L-homoalanyl) amino-2'-deoxyadenosine (Abu-2AA) as a nonfluorinated precursor of the charged tRNA analogue, to investigate the tendency of interactions within the editing domain. For following experiments, AMP-PNP, TfeGly and Abu-2AA were actually used to fill the binding pockets of IleRS.

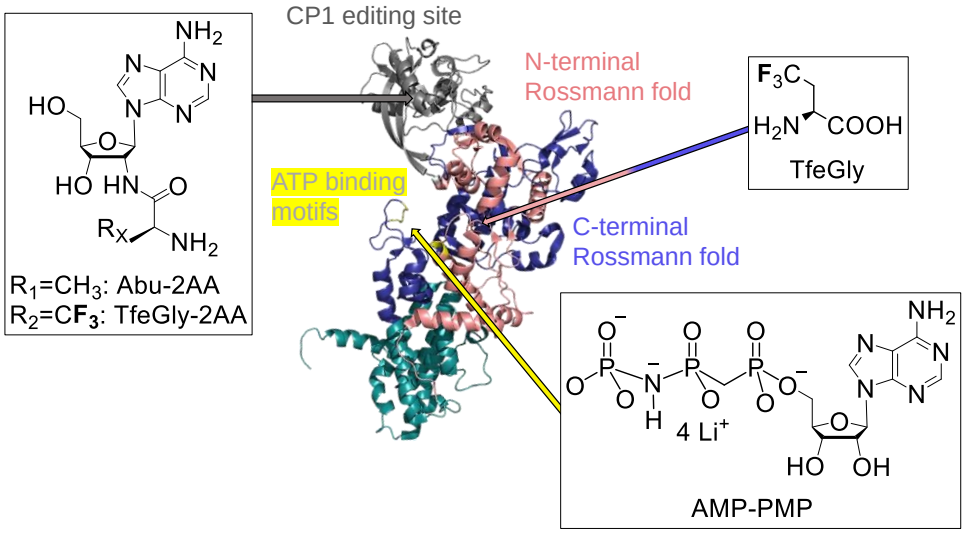


Figure 24 Crystal structure of the monomeric IleRS from *Thermus thermophilus* with corresponding non-hydrolyzable substrate analogues for appropriate binding pockets. Colored arrows indicate the destination site of particular ligands. Crystal structure is modified by PyMOL using PDB entry 1JZQ.

From 1344 screening trials, nice crystal growth could be observed for the same condition, in which IleRS crystallized with TfeGly, IleRS_screen_1 (**Figure 22C**). The new screen composition (17 mg/mL IleRS, 7.5 mM TfeGly, 1 mM Abu-2AA, 5 mM AMP-PNP in 100 mM MES, pH 6.5, 200 mM $Mg(OAc)_2$, 20 % w/v PEG 8000) is named IleRS_fullscreen_1 (**Figure 25**). We noticed that the growth took several weeks compared to the previous in Ile_screen_1, which we did not expect. In case the protein rigidity is increased by adding the substrates for the active sites, it should stabilize the structure and therefore, accelerate the growth process. From a different perspective, further additives decrease the final protein concentration in the equilibrated drop and in this way modify the kinetics of crystallization

Figure 25 Observation of crystal growth within 15 weeks. IleRS_fullscreen_1 sample is composed of 17 mg/mL IleRS, 7.5 mM TfeGly, 1 mM Abu-2AA, 5 mM AMP-PNP in 100 mM MES, pH 6.5, 200 mM $Mg(OAc)_2$, 20 % w/v PEG 8000. Although the crystal is intertwined, but there is an increase in size from approximately 250 µm to 500 µm. Due to the slower growth rate from 7th to 15th week compared to the starting halftime, we assumed that the crystal size reached its maximum.

process, which can lead to a lower growth rate of the crystal formation. However, we chose this condition for further refinement in 48-well plate. After several months, no high-quality single crystal could be observed. A diffraction test measurement was performed by Dr. Christian Roth (Max Planck Institute of Colloids and Interfaces, Potsdam, Germany). The result provided us a protein-specific pattern, but with lower resolution due to the small size of the crystals.

The failed attempts with the full-length protein led us to change the strategy to crystallize shorter independently folded domain of IleRS. Previous studies have shown that CP1 domain of *T. thermophilus* can be crystallized in complex with pre- and post-transfer editing substrate analogues.[231,232] Based on the alignment of *T. thermophilus* CP1 with *E. coli* CP1 the analogue sequence was determined for the *E. coli* equivalent and cloned into the pQE-80L H_6-TEV-IleRS plasmid to replace the full-length gene (**Figure 26**).

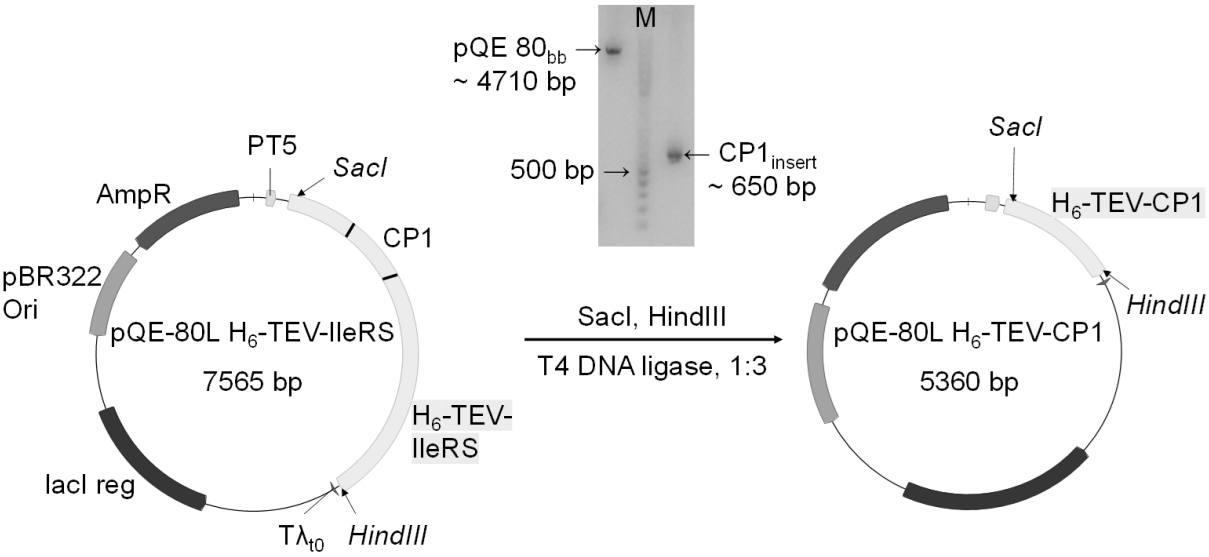


Figure 26 Schematic presentation of expression plasmid generation. pQE80 backbone ($pQE\ 80_{bb}$) is amplified between the restriction sites of *SacI* and *HindIII*, while the CP1 insert ($CP1_{insert}$) is obtained from amplification of the CP1 sequence extended by *SacI* and *HindIII* recognition sites. After restriction digestion and ligation of the purified fragments according to the protocol in **Section 6.1.3**, final plasmid pQE-80L H_6-TEV-CP1 could be isolated and verified by sequencing.

The editing domain was subsequently expressed and purified using the optimized protocol as described in the previous Section. We performed the SDS-PAGE analysis of H_6-TEV-CP1, which showed pure product at expected size of 24.524 kDa (**Figure 27**). 20 mg/mL concentrated protein was used in the following crystallization experiments. 1344 screening trials were conducted using Abu-2AA as a non-hydrolyzable substrate. Nonetheless, we were not able to observe promising starting points for refinements, neither with nor without Abu-

2AA. We assume that the N-terminal purification tag disturb the crystallization process and need to be cleaved. Furthermore, the CP1 domain from *E. coli* is probably not as stable as the one from *T. thermophilus* and degrades during the crystallization attempts.

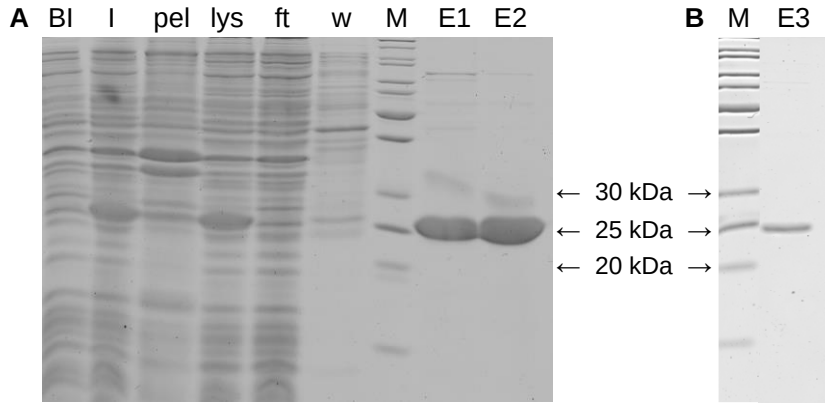


Figure 27 Analysis of H_6-TEV CP1 expression in E. coli BL21 (DE3) and purification by IMAC (A) and after SEC (B) on 12 % SDS-PAGE gel. (**A**) Before (BI) and after induction (I) fractions from standard gene expression of CP1 domain as well as insoluble (pel) and soluble (lys) samples after cell lysis are applied for expression analysis. Overexpression could be observed at expected size of approximately 24.5 kDa. From IMAC purification, flowthrough (ft) and wash (w) fractions were collected and analyzed with eluted fractions 1 and 2 (E1, E2). Enrichment of desired fragment could be obtained. (**B**) After SEC purification step, elution fraction 3 (E3) provides a sample in very high purity for crystallization purpose. The same marker (M) was used for both analyses.

Taken together, certain starting points for subsequent refinement are obtained from initial screens of full-length IleRS, but first attempts with CP1 domain remain unsuccessful. However, production efforts of a single crystal could not provide a crystal with suitable size for structure determination.

3.2 Reassignment of AUA codon to fluorinated Abu derivatives

Previous studies have shown that Abu and its fluorinated derivatives are activated by *E. coli* IleRS, but translation of these non-cognate substrates in place of Ile positions is not supported due to the proofreading activity of the synthetase.[233,234] However, applying e.g., the post-transfer editing defective mutant IleRS(Ala10) enables the accumulation of TfeGly-tRNA^{Ile} *in vitro*, and *in vivo* incorporation of TfeGly into the model peptide VW18, respectively. This translation was not selective, since replacement of Ile by other cAA like Leu, Val and Met could be detected via mass analysis, indicating the biosynthesis product as a protein mixture.[222] Nevertheless, it was the first reported ribosomal translation of small fluorinated ncAA in response to Ile codons. Starting from these results, we just have developed a new strategy to increase the selectivity of *E. coli* IleRS towards reassignment of Ile codon to TfeGly. The AUA rare codon was elected due to the advantages of AUA rare codon usage as blank codon for emancipation. As mentioned before (**Section 1.3.4.1**), the AUA rare codon decoding process in *E. coli* is catalyzed the lysidine-tRNA synthetase (TilS). By removing this essential enzyme from the Ile decoding pathway, the AUA rare codon becomes a 'blank codon' for reassignment by other AA. Our idea was to generate a strain that is able to regulate Ile decoding in the cell as circumstances require. Besides, the genomic *tilS* is deleted, meaning these cells would die in case alternative AUA decoding mechanism is absent. To prevent this, we supplied the cells a W905R mutant of *E. coli* IleRS that efficiently aminoacylated tRNA with UAU anticodon.[153] This implied the additional modification of $tRNA^{Ile1}_{GAU}$, and $tRNA^{Ile2}_{CAU}$ to $tRNA^{Ile1}_{UAU}$, and $tRNA^{Ile2}_{UAU}$, respectively. Furthermore, IleRS(W905R) libraries were created on the rescue plasmid that simultaneously functionalized for both screening and selection of the desired synthetase mutants. Thus, a screening plasmid encoding the library of synthetase, modified tRNA^{Ile}, and reporter protein was designed according to **Figure 28**. Detailed information about the strain and screening plasmid compositions will be given by the following Sections.

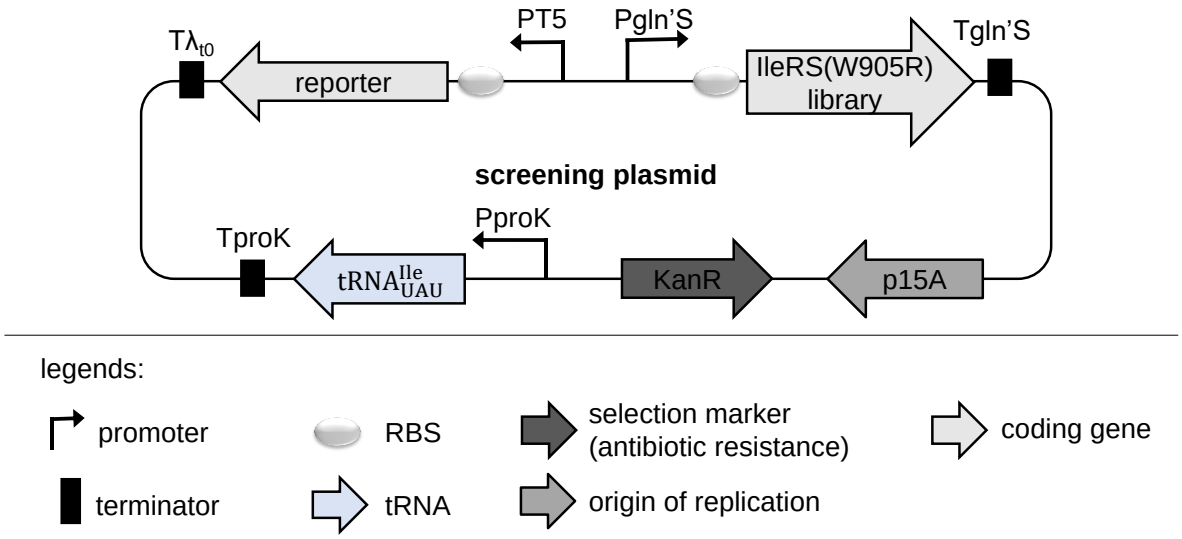


Figure 28 Graphic map of screening plasmid used for screening and selection of AARS mutants specific for the desired ncAA translation.

3.2.1 IleRS libraries

In general, enhancement of the catalytic efficiency of an enzyme is achieved by redesigning its binding pockets. In particular when involved residues of AARS form a relaxed overall hydrophobic cavity around the bound substrate instead of a complementary shape, tailoring AARS binding cleft is an essential step towards substrate recognition and specification. A comprehensive library is the best means for this purpose to cover necessary mutations. Since IleRS, a class Ia synthetase, possesses two distinct catalytic domains, both sites were considered for the library design. *T. thermophilus* IleRS possesses around 24 % sequence identity compared to *E. coli* IleRS, but approximately 90 % homology of the active centers, since these areas contain well conserved residues in the two bacteria species. Based on the published crystal structure of *T. thermophilus* homolog several positions in the first shell of the active and editing site of *E. coli* IleRS are chosen for randomization. To be more precise, residues are considered that face to the substrate side chain and are within 5 Å environment of the substrate (**Figure 29**). Most of those positions are known to have specific structural and functional impact on IleRS.[99,235–237] For instance, the hydrophobic cleft of activation site is formed by the residues P57, W569 and W534. Furthermore, T243 and D342 are essential elements of the editing pocket.[238] Mutation to IleRS(T243R, D342A) inactivates the deacylation ability of misacylated tRNA.[239]

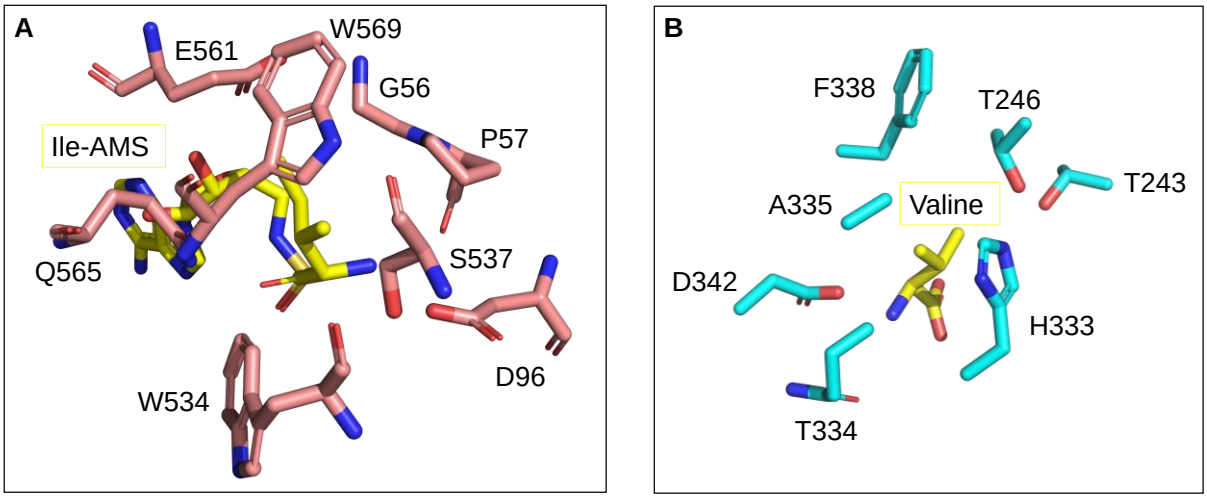


Figure 29 Residues facing the substrate side chain within 5 Å of *T. thermophilus* IleRS active (A) and editing site (B). Appropriate co-crystallized substrates are shown in yellow sticks: (**A**) non-hydrolyzable analogue of Ile-AMP 5'-*N*-[*N*-(L-isoleucyl)sulfamoyl]adenosine (Ile-AMS), (**B**) valine. Residues are annotated in single letter code with corresponding positions in *E. coli* IleRS. Structures are modified from PDB entry 1JZQ (**A**) and 1UE0 (**B**) by PyMOL.

Randomization of all relevant residues of the binding pockets would yield a library of about 20^{15} = 3.3 · 10^{19} members. Due to limited transformation efficiency in *E. coli* (10^8 - 10^9 recombinants), we decided to separate the two centers and generate smaller libraries to fully cover the possible IleRS mutants.[240,241] Taken from the ability of *Mycoplasma mobile* enzyme decoding the AUA codon without assistance of *tilS*, W905R mutation was inserted into *E. coli* IleRS to generate a mutant capable for recognition and charge of $tRNA^{Ile}_{UAU}$ (see **Section 3.2.2** for details).[153] For the libraries of the activation site IleRS(W905R, Ala10) sequence served as the template on the basis of two facts: 1) former studies revealed that Ala10 mutant of *E. coli* IleRS inactivated the editing function and enabled translation of noncognate substrates including TfeGly;[222] 2) mutations influencing binding of the substrate side chain during aminoacylation reaction did not affect the editing activity.[242] Inactivation of the editing function is therefore essential for translation of noncognate substrates in case the editing site is not changed coincidently. Based on these points, we focused on introduction of further mutations into the activation site of this construct. In this way, we tailored the synthetic subsite of the synthetase to explicitly aminoacylate fluorinated ncAA without any subsequent disturbance by the enzyme's proofreading function.[222] Since the competing cAA are limited to Ile, Leu, Val, and Met, we expected a reduction in size of the activation site would exclude cAA larger than TfeGly and therefore, lower the number of canonical competing substrates of the synthetase. Especially, G56 is one residue of the synthetic site that aroused our interest. This position has been shown to minimize the binding pocket due to its close proximity to the side chain of the

substrate (**Figure 30**).[243]

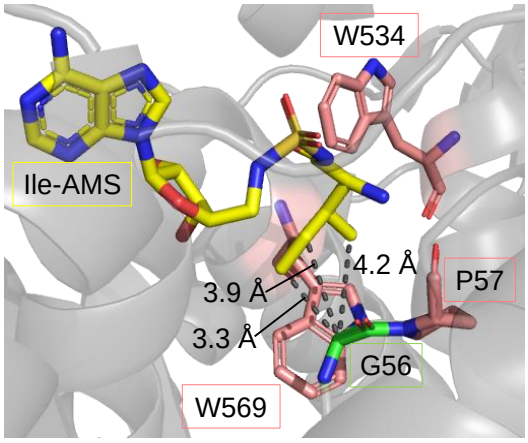


Figure 30 Location of G56 and the hydrophobic core relative to the substrate Ile-AMS within the activation site of *T. thermophilus* IleRS crystal structure. The distance between G56 (green) and the side chain of Ile-AMS (yellow) ranges from 3.3 Å to 4.2 Å. Residues are annotated in single letter code with corresponding positions in *E. coli* IleRS. Structure is modified from PDB entry 1JZQ.

We have introduced different mutations into *E. coli* IleRS (Ala10) mutant, in order to test the impact of this position on the translation efficiency for TfeGly. Gly is the smallest nonpolar and aliphatic cAA with neutral net charge. On the one hand, it is reasonable to replace this residue with the next larger cAA with the similar property, alanine (Ala, A), and on the other hand, with polar cAA, such as asparagine (Asn, N) as well as proline (Pro, P) having unique property among all cAA. Ala is a favored neutral substitution of Gly due to its similar small steric size. G56P substitution was shown to disable Ile and Val activation ability of IlleRS.[243] Asn replacement was chosen to investigate the impact of an local increased polarity of the active site on the intermolecular interactions with TfeGly. Via site-directed mutagenesis, the plasmid pQE 80L IleRS (Ala10) containing G56A, G56P, and G56N, respectively, was generated. Protein expression and purification was performed according to **Section 6.2.5** and **6.3.1**. Unfortunately, the activity assay (**Section 6.3.7**) revealed a decrease in activation efficiency of the mutants with narrowed synthetic site (**Figure 31**). Even substitution with the smallest change, G56A, resulted in 7.7-fold decrease in activity compared with IleRS (Ala10). The G56N mutant had almost lost the activation efficiency for TfeGly. Proline substitution provoked a 2.6-fold reduced activity. Having a closer look at the sequence, G56 is a part of a 22-membered loop (KKGKKTFILHD**G**PPYANGSIHI) that directly faces the catalytic core. Ala is known to be preferred in α-helices and Asn is not a typical member of any substructures.[244] Since Gly and Pro are frequently found in loop structures, substitution of Gly by Pro causes the smallest structural change.[245] The formation of loop is in turn important early in folding,

since it enables the exploration of energetically favorable interactions of the polypeptide chain and leads to chain compaction.[246] As the entire activation efficiency of IleRS (Ala10) mutants towards TfeGly is substantially affected upon reduction of the binding pocket size by G56 mutations, this position is removed from the list of residues that are considered for the synthetic site libraries.

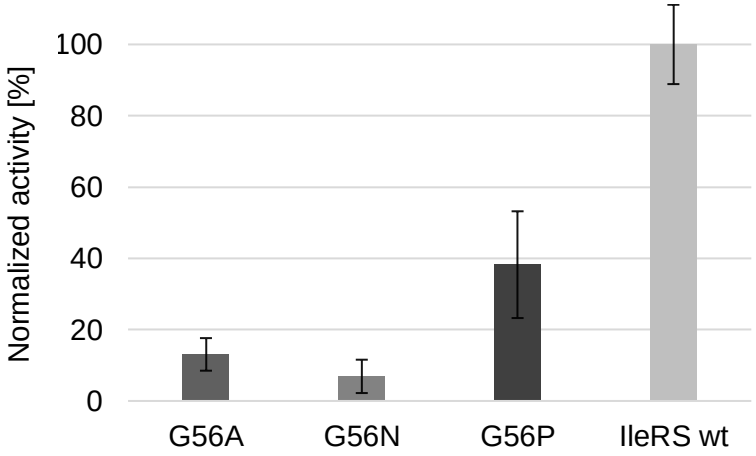


Figure 31 Malachite green activity assay. Activity of 1 µM *E. coli* IleRS (Ala10) mutants towards 8 mM TfeGly. This assay relies on the colorimetric quantification of the complex formed between free orthophosphate (PP_i) and Malachite green molybdate. The amount of green molybdophosphoric acid complexes is determined by spectroscopic measurement at 650 nm and is directly correlated with c (PP_i) in the reaction mixture. As PP_i is formed during AA activation step of AARS, the Malachite green can be used to quantify the amount of PP_i generated by IleRS (Ala10) mutant activity. c (PP_i) formed by initial IleRS (Ala10) (IleRS wt) was put on a level with 100 % of enzyme activity to normalize the activity of other IleRS (Ala10) mutants having smaller synthetic center: G56A, G56N, G56P. Values represent the average of five samples with the standard deviation as error bars.

Since the accurate information of the interactions of TfeGly within the active centers are not available, we have identified the homologue *T. thermophilus* IleRS residues in *E. coli* and randomly selected four positions for different libraries of each catalytic binding pocket (**Table 3**). $20^4 = 1.6 \cdot 10^5$ sequences resulted from a respective library, making it experimentally implementable. pQE-80L *Ec*IleRS was taken from the database and initially used as the template for insertion of W905R and/or Ala10 mutations. Five of six libraries were then successfully synthetized via site-saturation mutagenesis using NNK degenerate codons at selected positions. By any means, the IleRS library E2 could not be produced. Starting from these five libraries 15 different screening plasmids will be constructed which cover a great number of testing diversity.

Table 3 Relevant residues and randomly composed libraries of active (A1-A3) and editing site (E1-E3) of *E. coli* IleRS.

Randomized positions	Library description
P57, D96, S537, W569	A1
W534, E561, Q565, W569	A2
P57, D96, W534, E561	A3
T246, H333, A335, H338	E1
T243, H333, T334, D342	E2
T246, T334, A335, D342	E3

By sequencing of obtained complete library mixtures, we could survey the quality of the IleRS libraries (**Figure 32**). Library A1 could be randomized perfectly since all positions were occupied by NNK. The resulting library size is $1.6 \cdot 10^5$. For all other libraries, randomization of one or two positions were failed to be accomplished. Library A2 did not cover Pro, Thr, and Ala substitution at W569 residue causing a decrease of $2.4 \cdot 10^4$ mutants. Library A3 lacked Ile, Thr, Lys at P57 and Phe, Tyr, Cys at D96. In case of library E1, T246 was not mutated to five cAA: Cys, Trp, Arg, Glu, Gly. However, the overall size of these libraries is relatively equal to the expected one and missing substitutions are compensated by other cAA with similar properties. Only library E3 possesses the lowest number of members. Here, 2400 mutants could be synthetized, made up 1.5 % of target value. This is due to the nearly unchanged residues T246 and D342. T246P, D342E, and D342G could be obtained for this editing site library. One possible reason is the bias towards cytosine and guanine, which is often favored for annealing reactions due to the stronger interaction of GC base pairs.[247] As template DNA, only wildtype sequence is available at the beginning of a PCR reaction. Thus, the three formed hydrogen bonds of G and C are stronger, compared to the two hydrogen bonds of A and T, GC interaction preferably leads those primers containing many cytosines or guanines to anneal to the ACT codon of T246, and GAC of D342, respectively. As a consequence, C/G-rich templates (CCG, CCT for T246 and GAG, GGG, GGC for D342) are produced for the following PCR cycles. Nonetheless, the choice of residues for the libraries were random and the number of mutants in the six individual library pools is sufficiently enough to deem the constructed library satisfactory.

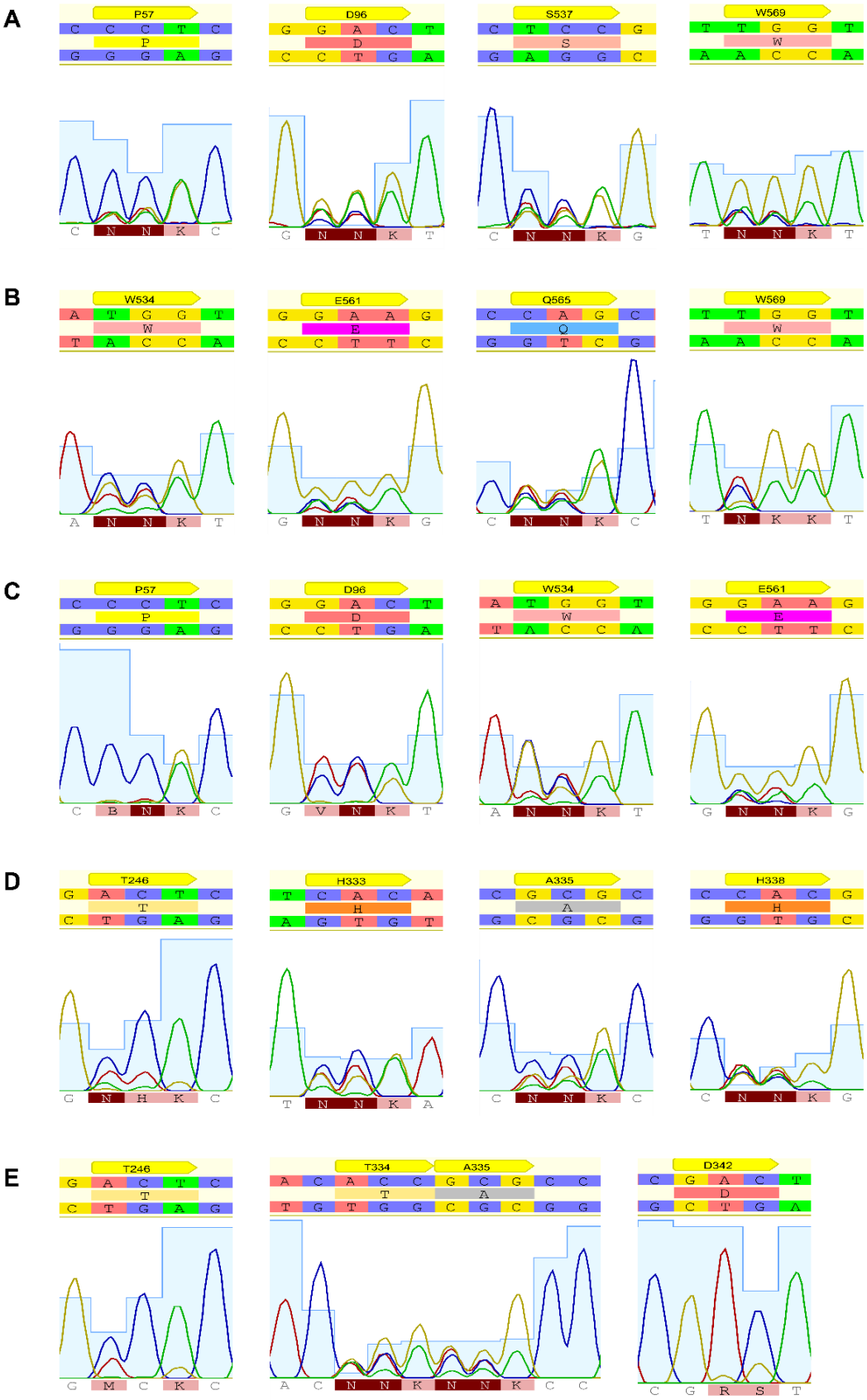


Figure 32 Sequencing results of the library mixture. Randomized positions of **A**) library A1 $1.6 \cdot 10^5$, **B**) library A2, **C**) library A3, **D**) library E1, **E**) library E3, are shown with sequencing chromatogram. The size of each library amount to A1 = $1.6 \cdot 10^5$, A2 = $1.36 \cdot 10^5$, A3 = $1.16 \cdot 10^5$, E1 = $1.2 \cdot 10^5$, and E3 = $2.4 \cdot 10^3$.

3.2.2 Modification of tRNA^{Ile}

E. coli exhibits five tRNA^{Ile}: ileT, ileU, ileV, ileX and ileY (**Figure 33**). The genes encoding ileT, ileU and ileV tRNA are identical. They own the anticodon GAU (**Figure 33**, yellow framed) that pairs with the two most frequent codons AUC and AUU. The remaining sequences of ileX and ileY tRNA bear the anticodon CAU (**Figure 33**, brown framed) and are involved in decoding the rare codon AUA by means of *tilS* (**Figure 9**). They differ only in two bases within the acceptor stem (**Figure 33**, green framed)

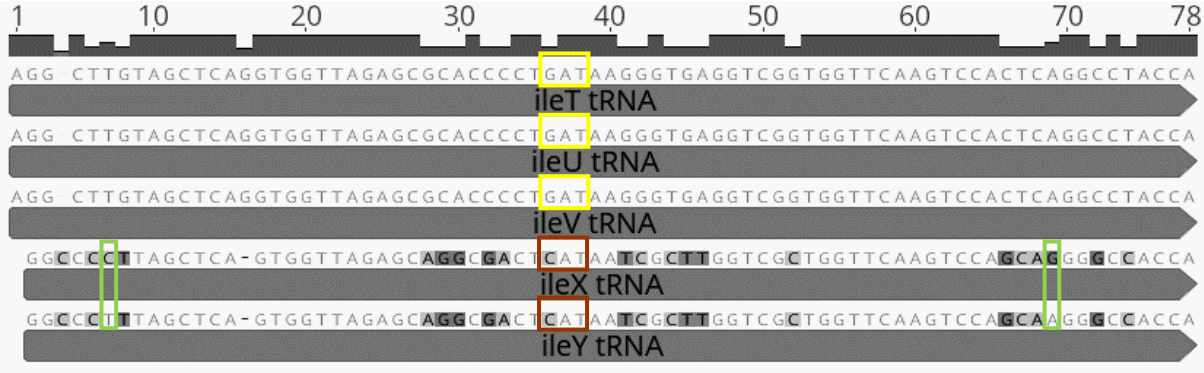


Figure 33 Sequences of tRNA^{Ile} in *E. coli*. Anticodons of ileT, ileU and ileV tRNA are highlighted by yellow frames and the one of ileX and ileY in brown. The difference in sequence between ileX and ileY are shown in green frames. Since all sequences are given in DNA format, uracil nucleobase (U) is denoted here as thymine (T). Alignment was performed by the Geneious prime software.

In this study, we amplified the inserts $tRNA^{Ile1}_{UAU}$ and $tRNA^{Ile2}_{UAU}$, from the database plasmids pNB26'2_EctRNA_TAT-EcIleRS, and pNB26'2_EctRNA_TAT2-EcIleRS, respectively. The nucleobase at position 34 of different tRNA^{Ile} was substituted with U (**Figure 34**). Both $tRNA^{Ile}_{UAU}$ did not exist in *E. coli* and cannot be charged by wildtype IleRS.[153] Only W905R mutant showed efficient recognition of these modified tRNAs, comparable with the efficiency towards native $tRNA^{Ile}_{CAU}$.[153] By supplying the modified $tRNA^{Ile}_{UAU}$:IleRS(W905R) pair, we provide the cells a new ability to decipher the Ile rare codon AUA that is independent from the *tilS* activity. Since we aimed to work with a *tilS*-knockout strain, the $tRNA^{Ile}_{UAU}$:IleRS(W905R) pair is the only opportunity for cells to survive. Templates used for amplification were the native gene sequences of $tRNA^{Ile1}_{GAU}$ (equals to ileT, ileU, ileV) and $tRNA^{Ile2}_{CAU}$ (equals to ileX). We aimed to test the different recognition ability of IleRS for both types of tRNA which is correlated with the frequency of codon usage.[248] Genes were successfully amplified via site-directed mutagenesis PCR. Results were verified by sequencing with corresponding primers (**Appendix**).

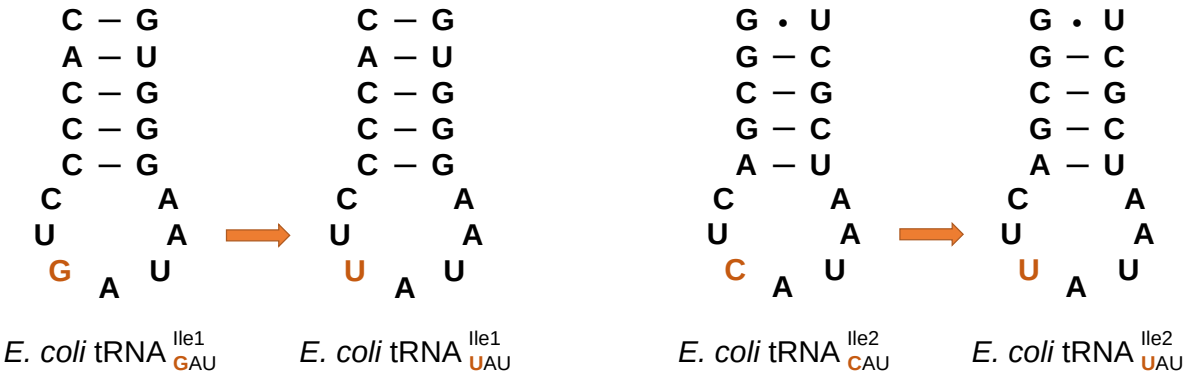


Figure 34 The anticodon stem-loop sequences of *E. coli* tRNA^{Ile} with highlighted purposed mutation at position 34.

3.2.3 Design of an expression reporter protein

Selection and detection of gene expression are essential to high throughput screening assays that for example enable analysis of protein translation efficiency. For nonsense codon reassignment, double-sieve selection is a commonly used approach for fast screening and enrichment of promising AARS mutants. However, such a standard genetic selector does not exist for evaluation of sense codon emancipation. Each newly reassigned system requires a suitable genetic selection strategy. In this work, our reliable selection is the cell survival. Our *tilS* knockout strain is only rescued by proper assignment of Ile rare codon with Ile during growth phase and later with TfeGly that is catalyzed by IleRS mutants from the generated library (the strain concept is supplied in **Section 3.2.5**). Nevertheless, this library might contain mutants that translate other cAA at AUA positions. The identification of reassigned AA would be difficult if there is no protein containing AUA which can be isolated and purified for analysis, e.g., by mass spectrometry. Therefore, we used a screening system as a modest way for the confirmation of reassigned AA and assessment of the incorporation efficiency. Two different reporter concepts were applied to readout the best IleRS mutant from the libraries determined by fluorescence (**Figure 35**). The reporter superfolder green fluorescent protein (sfGFP) is chosen due to its advantages in folding and stability compared to wildtype GFP.[189]

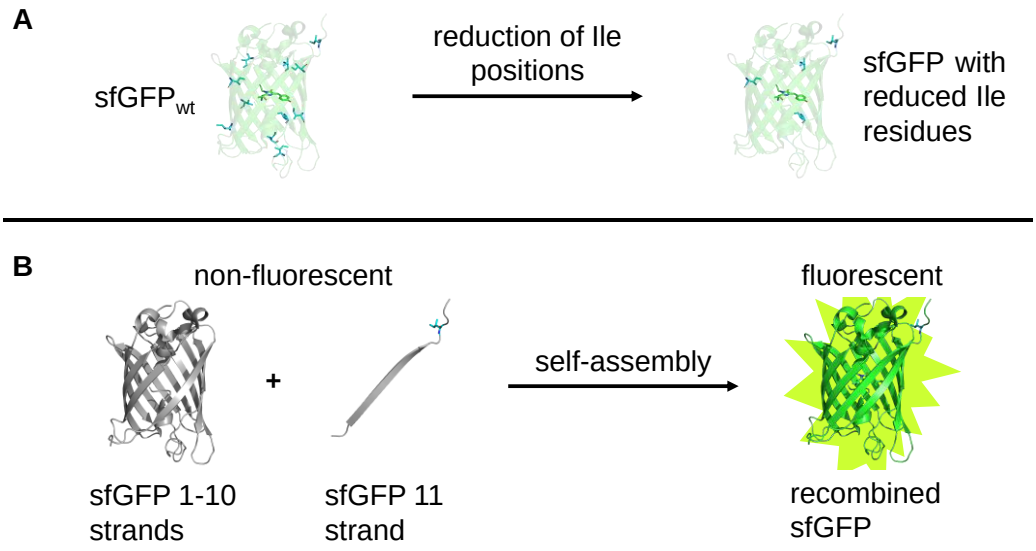


Figure 35 Two ways of reporter usage. Structures were modified from PDB entry 2B3P by PyMOL.

3.2.3.1 A modified sfGFP for screening

The full-length sfGFP was used in the first strategy (**Figure 35A**). It is known that higher number of positions, need to be substituted in one protein, can lower the translation efficiency. This probably leads to incomplete replacement of the cAA by its analogue which in turn resulted in an inhomogeneous expression product. In addition, multiple substitution would complicate the assessment of changes. The impact of each single mutation cannot be clearly identified and individually assigned. To avoid the limitation by the residue number, nine of eleven Ile positions in sfGFP were gradually substituted by other cAA (**Figure 36**). Mutations I14M, I47V, I123T, I128T, I136T, I152V, I161L, I167V, I188A were selected based on their presence in mutants of fluorescent proteins.[249–252] For example, I128T and I167V are two of seven mutations that were introduced into sfGFP fragment for enhanced protein solubility.[250] Since the fluorescent protein structures were not abolished by the selected mutations, we assumed that transfer of the respective mutations into our sfGFP sequence would not disturb the intact protein folding. Only I98 and I229 were excluded from mutagenesis (**Figure 36**, yellow). I98 is located close to the chromophore. Specifying this position by one cAA would take risk to lose the fluorescence ability. Therefore, this residue was used for randomization. I229 is described as one of the critical sites for the fluorescent formation of sfGFP and should not be substituted by any cAA.[253] Furthermore, we need to verify that these positions tolerate TfeGly incorporation, resulting in an fluorescent TfeGly-containing protein. Due to the similar size and physicochemical properties to Ile, TfeGly is a promising candidate for Ile substitution.

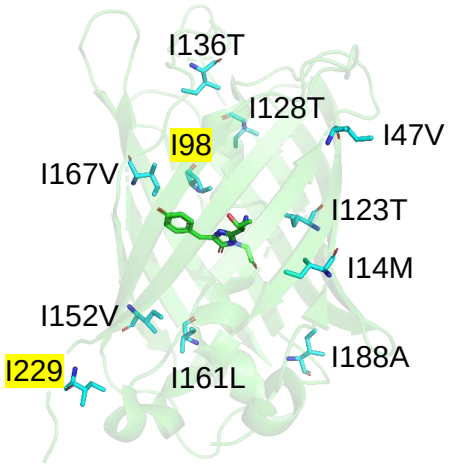


Figure 36 Isoleucine positions in sfGFP and their purposed mutations. The tyrosine-based chromophore, formed by the tripeptide T65-Y66-G67, is buried in the middle of the barrel structure (green stick). I98 and I229 are positions with no defined substitution (highlighted in yellow). Structure was modified from PDB entry 2B3P by PyMOL.

With this background, we first introduced the defined mutations into sfGFP one by one via site directed mutagenesis. pET28a+H_6-sfGFP-Strep was applied as the template DNA. Two most distant positions from the chromophore were chosen as starting points: I136 and I188. After sequence verification, modified sfGFP was transformed into an expression strain for a rapid check whether these mutations would affect the protein fluorescence. The results showed right fluorescent colonies upon irradiation by light at λ = 365 nm meaning these changes did not influence the spectroscopical property (**Figure 37**).

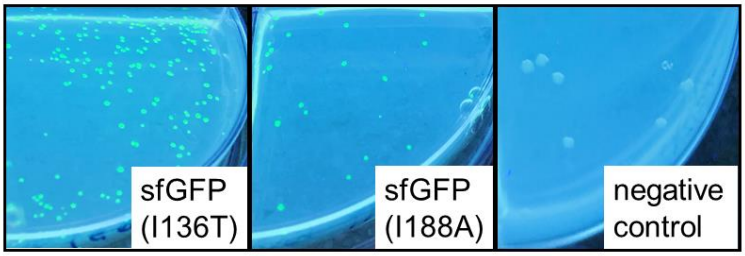


Figure 37 Rapid test of fluorescent behavior of cells containing sfGFP mutants. sfGFP bearing single mutation was introduced into BL21 (DE3) gold. LB agar plates were supplied with 1 mM IPTG for induction of gene expression. Cells containing sfGFP mutants exhibit fluorescence upon irradiation by λ = 365 nm (left and middle). *E. coli* BL21 (DE3) gold lacking the gene of sfGFP was used as the negative control (right). The numbers of colonies are irrelevant, since the number of plated cells were not normalized.

Following this work scheme, further mutations were consecutively introduced into sfGFP. In case that the last added mutation yielded nonfluorescent proteins, other residues were taken for mutagenesis. For example, introduction of I14M into sfGFP (I136T, I188A) resulted in properly folded protein, while further mutations lead to drastic loss in fluorescence (**Figure 38A**). Absorption of cell cultures at 600 nm (OD_{600}) was recorded, on the one hand, to monitor the impact of mutations on the cell fitness, on the other hand, to ensure that the fluorescence intensity has mainly originated from the protein property and is not a result from fluctuating cell density (**Figure 38B**). Since the fluorescence intensity decreased while OD_{600} values remained similar, we concluded that this mutation somehow induced an improper protein folding. Another explanation could be a quenching effect of the introduced mutation, which seems unlikely in our case. In order to test the opportunity for protein fluorescence conservation, we decided to keep the modification of I14 residue until the final round and randomize this position to cover all 19 cAA substitutions.

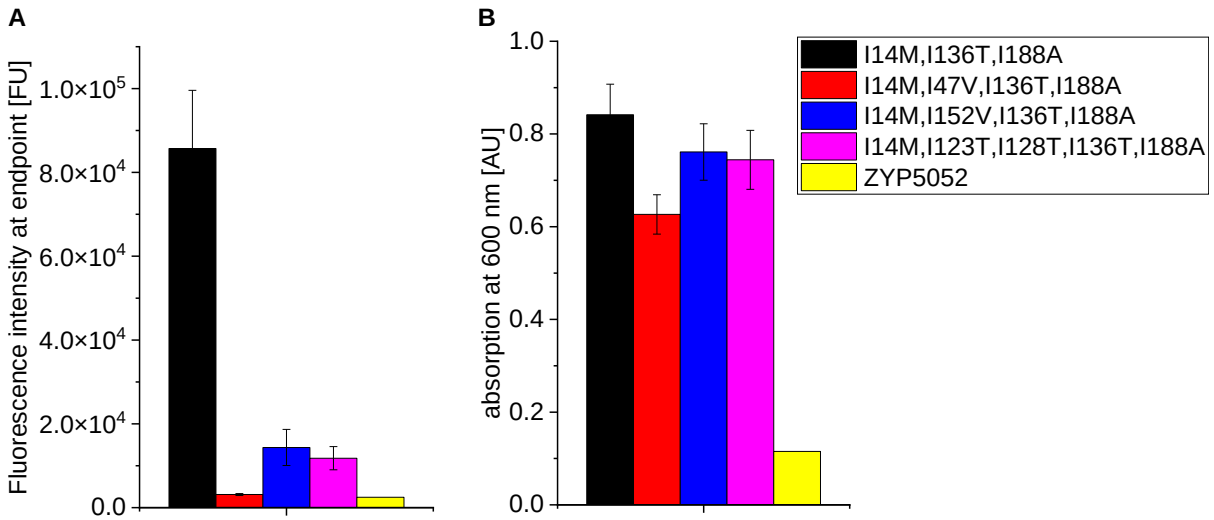


Figure 38 Spectroscopical characterization of I14M containing sfGFP mutants. (**A**) Fluorescence spectrum was monitored at T = 30 °C, $\lambda_{exc.}$ = 488 nm and $\lambda_{em.}$ = 510 nm for 24 h expression in ZYP 5052 medium. Intensities from endpoint measurement is displayed in the bar chart. (**B**) Absorption spectrum was determined after 24 h expression at 600 nm. Five clones of each mutant at OD_{600} ~ 0.2 was used for incubation start. Bar charts were prepared by the OriginPro tool.

Using NNK degenerate codons at selected positions I14, I98 and I229 randomization was successfully performed via site-saturation mutagenesis. There was no fluorescence detectable for those mutants in a rapid plate assay (data not shown). In accordance with our expectation, cAA substitution of I98 and I229 residue disturbed the chromophore formation, and the protein folding, respectively. I14 was in turn an essential residue for the barrel structure once the number of mutations is increased. It can therefore be concluded that each

modification we introduced into the sequence caused small changes in protein structure assembly. By the addition of a rising number of mutations, insignificant changes accumulated, resulting in misfolding of the protein. That is the case, for example, when the barrel structure could not be developed and/or the chromophore is not properly formed. Nevertheless, a quantitative fluorescence analysis of expressed sfGFP mutants was implemented to confirm the results from the rapid assay. For comparison, five clones were used for sfGFP constructs with defined mutations which could remain fluorescent. In case of randomization, 20 clones of each position were inoculated for expression analysis, resulting, for instance, in 40 clones for sfGFP bearing both I14NNK and I98NNK mutations. The reason for these numbers is that each randomized position can be occupied by 20 possible cAA. Therefore, at least 20 clones for one NNK position is required to cover the potential mutants. With two randomizing positions in one construct there are $20^2 = 400$ possibilities of AA compositions. Since the initial assay already showed no fluorescence on plate, we decided to pick 40 representative clones for analysis. The resulted fluorescence and absorption intensity at 600 nm of cell cultures at endpoint are shown in **Figure 39**. Absorption values of all tested clones remained similar, indicating an independency of fluorescence intensity from cell density (**Figure 39B**). Reduction of five Ile residues revealed the best result (**Figure 39A**, black column). Half a fluorescence intensity was decreased by the presence of two additional substitutions (**Figure 39A**, cyan column). Further introduction of mutations into sfGFP lead to nearly complete loss of the fluorescence (**Figure 39A**, yellow column). Randomization of selected positions could not gain a maintenance of the spectroscopical property (**Figure 39A**, blue, purple, brown column). As already mentioned, both Ile positions, 98 and 229, are critical residues for sfGFP and could not be replaced by other cAA. Alternatively, we only performed the measurement with 40 clones, which equaled to maximal 10 % of the possible mutants for two randomized positions within one construct. A promising construct may have been missed by this screening. However, the observed negative impact of increasing number of mutations on the fluorescence behavior would argue against this idea. Since our screening method requires an adequate fluorescence intensity for readout, we continued with the first two sfGFP mutants from this analysis: sfGFP(I123T, I128T, I136T, I152V, I188A) and sfGFP(I123T, I128T, I136T, I152V, I161L, I167V, I188A).

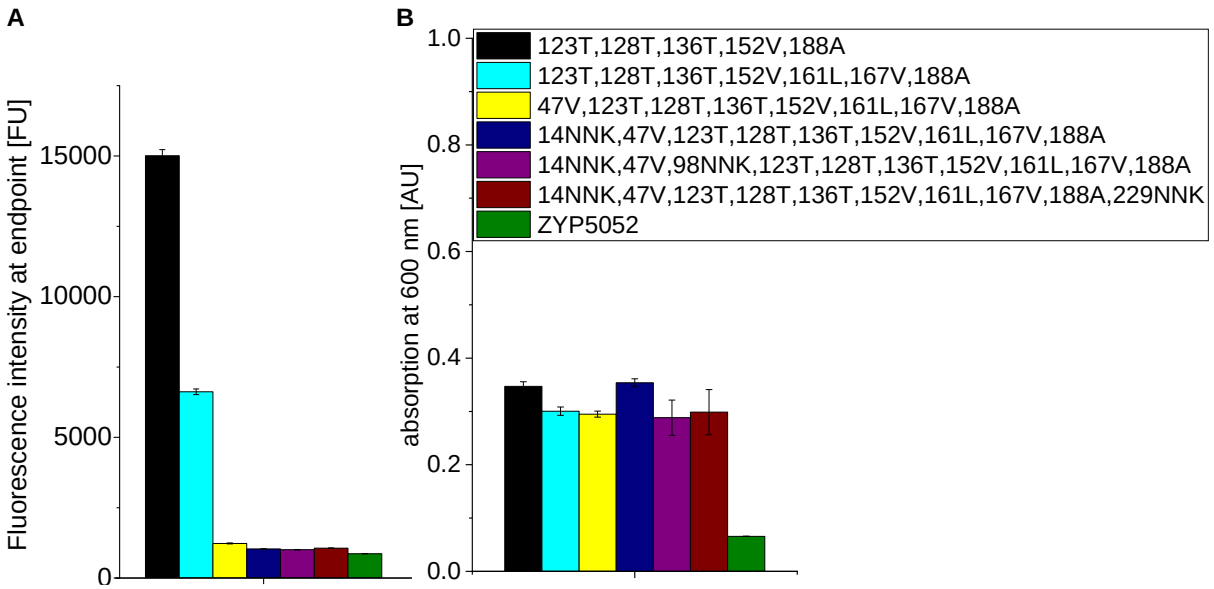


Figure 39 Spectroscopical characterization of sfGFP mutants. (**A**) Fluorescence spectrum was monitored at T = 30 °C, $\lambda_{exc.}$ = 488 nm and $\lambda_{em.}$ = 510 nm for 24 h expression in ZYP 5052 medium. Intensities from endpoint measurement is displayed in the bar chart. (**B**) Absorption spectrum was determined after 24 h expression at 600 nm. Different number of clones were set up for the sfGFP mutants: for each of the first three (black, cyan, and yellow) five clones, for the 4th (dark blue) 20 clones, and for the last two 40 clones each. OD_{600} ~ 0.2 was used for incubation start. Descriptions of mutations were shortened to the number of Ile position and replacing AA. Bar charts were prepared by the OriginPro tool.

Substitution of the remaining Ile frequent codons by AUA resulted in reporter protein containing six rare codons (sfGFP(I14AUA, I47AUA, I98AUA, I123T, I128T, I136T, I152V, I161AUA, I167AUA, I188A, I229AUA)) and four rare codons (sfGFP(I14AUA, I47AUA, I98AUA, I123T, I128T, I136T, I152V, I161L, I167V, I188A, I229AUA)), respectively. The sfGFP constructs are abbreviated as sfGFP(6xAUA) and sfGFP(4xAUA) in following sections.

3.2.3.2 Split sfGFP

The second screening reporter strategy recruited an engineered sfGFP as a split protein that is able to self-assemble without assistance of supplementary protein-protein interactions (**Figure 35B**).[250] The split construct is separated between the tenth and eleventh β-strand of the full-length sequence into two fragments: GFP1-10 (residues 1-214) and GFP11 (residues 215-230). The GFP11 is the short C-terminal peptide sequence and consists of 16 AA. GFP1-10 includes the tripeptide for the chromophore formation. This fragment remains nonfluorescent, since the conserved E222 residue for maturation is on the missing GFP11 part.[254] The two fragments reconstitutes to a fluorescent sfGFP after a complete chromophore

maturation reaction. This convenient split system is a popular tag for application like protein quantification, visualization of protein subcellular localization, single-molecule imaging, as well as *in vitro* protein complex assembly.[255–258]

Our experimental setup is divided in two parts. First, the N-terminal fragment GFP1-10 is expressed during the cell growth process, in which 20 cAA are present. Thus, all Ile residues encoded by abundant codons (AUC and AUU) on this fragment are properly assigned with Ile by the endogenous translation system. Second, the endogenous IleRS is inactivated, e. g., by temperature, and only IleRS mutants supplied by our library are able to decode AUA positions. TfeGly is added to the growth media and GFP11 expression is induced. In this way, the single AUA codon on the C-terminal part will be exclusively reassigned by TfeGly. Expression of each fragment is controlled by different promoters in order to enable time-shifted induction. The advantage of this method is that only one residue of the reporter protein will be substituted by a ncAA and therefore, provides a homogeneous product. Subsequential isolation and purification of the reconstituted sfGFP can be further characterized by mass analysis.

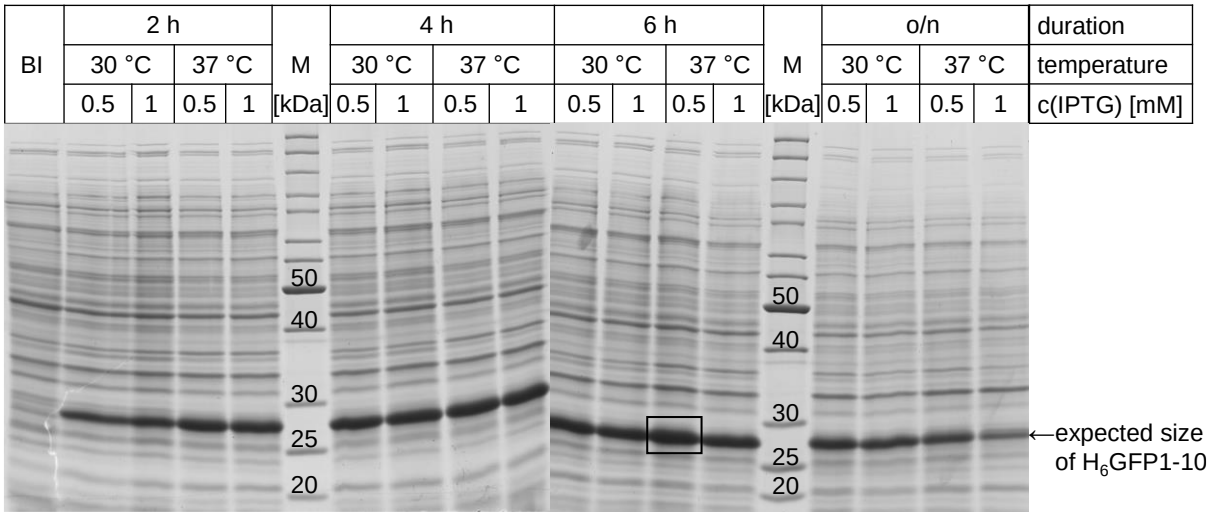


Figure 40 Determination of expression conditions of H_6-GFP1-10 in *E. coli* Dh10b on 15 % SDS-PAGE gels. Before induction fraction (BI) and protein expression samples were collected after several time points. Protein production was induced by 0.5 mM or 1 mM IPTG. Expression was performed at 30 °C, and 37 °C, respectively. Overexpression is observed for protein around expected size ($M_{W, calc.}$ (H_6-GFP1-10) = 26.5 kDa). Black frame highlights the best result of the overall test. Black arrow indicates the overexpressed H_6-GFP1-10 band. Unstained marker (M) was used for this analysis.

The plasmid pET28a+H_6-sfGFP-Strep was used as the starting point for generation of GFP1-10. Since the wildtype sequence of this N-terminal fragment is not soluble in aqueous solvents, seven mutations were introduced into this sequence, according to the literature, to increase its solubility by approximately 50 %: N39I, T105K, E111V, I128T, K166T, I167V and S205T.[250]

Substitutions were implemented by site directed mutagenesis. The eleventh strand of sfGFP was then deleted from the gene yielding a GFP1-10 with N-terminal H_6-tag that was verified by sequencing. Test expression was performed to determine the expression parameters. Subsequent analysis by SDS-PAGE revealed best result when expression was induced by 0.5 mM IPTG for 6 h at 37 °C (**Figure 40**, black frame). This outcome supports our described experimental project in which the GFP1-10 is expressed during the cell growth. Expression duration of 6 h can be realized because our strain is expected to grow in a longer time period. However, we decided to use 30 °C for protein production, since our screening strain should carry a genomic thermosensitive IleRS mutant that would lose its catalytic efficiency upon increasing temperature (see strain information in **Section 3.2.5**).

GFP11 was ordered as a short oligonucleotide fragment that contained AUA rare codon for encoding the I229 residue. It can be directly used as the template to clone on the screening plasmid using appropriate primers.

Both reporter systems were successfully generated on several plasmids. In the next Section, the composition of all relevant components to a selection and screening plasmid will be described in more detail.

3.2.4 Assembly of an ultimate selection and screening system

The selection and screening plasmid was constructed according to the map on **Figure 41**. To avoid lots of selection markers (antibiotics) in a host strain, which would increase the stress on cells, we composed the system on a single plasmid. The effect of antibiotic agents on cell fitness cannot be predicted for new strains. Hence, we chose ampicillin (Amp), and kanamycin (Kan) for selection of cells harboring our plasmid with appropriate resistance gene, respectively. To allow production of AARS library (plasmid IV) and corresponding modified $tRNA^{Ile}_{UAU}$ (plasmid I) from the starting point, both were cloned under control of constitutive promoters/terminators, glnS' (plasmid II) and proK, on a pPAB26 backbone (plasmid V, **Figure 41**). Reporter protein sfGFP bearing a N-terminal H_6-tag for purification and several AUA codons (H_6-sfGFP(xAUA)) was in turn controlled by the inducible strong promoter T5. All important features were amplified from prepared constructs using appropriate primers. Purified amplicons were assembled to the final screening system via golden-gate cloning (**Figure 41**).

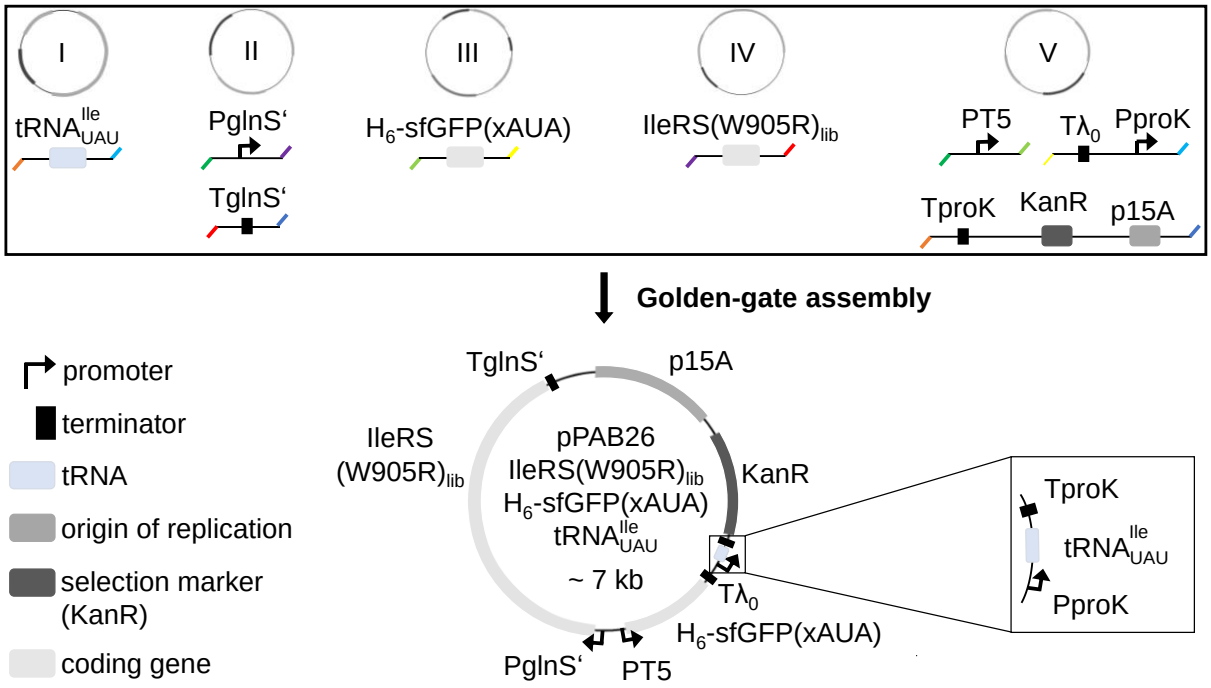


Figure 41 Assembly strategy of the screening plasmid. Relevant fragments were amplified from following plasmids: I) pNB26'2 *Ec*tRNA$_{UAU}$ *Ec*IleRS, II) pBU16-1GK *Mj*oNBY, III) pET-28a H$_6$-sfGFP-Strep, IV) pQE-80L *Ec*IleRS(W905R)$_{lib}$, V) pPAB26 CAT(Q98TAG, D181TAG) PT5 H$_6$-SUMO$_{opt}$-sfGFP(R2TAG)-Strep *Mj*tRNA$_{opt, CUA}$. Applied primers contained Eco31I cut sites and extended particular modules with overhangs that were unique for recombination. The color of the overhangs indicated the cohesion of each fragment.

In order to test the functionality of the rare codon screening concept, we first have synthetized the screening plasmid haboring the *E. coli* IleRS (W905R, Ala10) mutant and one copy of $tRNA^{Ile1}_{UAU}$. To compare the influence of the number of rare codons on the translation efficiency, sfGFP (6xAUA), sfGFP (4xAUA), and the mutant showing the best fluorescence intensity after reduction of Ile positions from previous Section, were also cloned on the test system. The obtained screening plasmid carried the gene for KanR that was changed to AmpR afterwards. All plasmids were transformed into *E. coli* Dh10b and Top10 cells to determine the production of protein which was correlated to the fluorescence intensity. *E. coli* Dh10b and Top10 are two most commonly suitable strains in screening libraries due to, inter alia, their great transformation efficiency (greater than 1 - 5 x 10^8 cfu/µg in case of chemically competent cells), and reduced recombination events (recA1).[259–261] Absorption at 600 nm and fluorescence of samples were detected after overnight expression. Similar absorbance of tested populations could be observed, and it can be concluded that our system has no influence on the cell fitness (data not shown). Therefore, we normalized the fluorescence values to the corresponding OD_{600} to analyze the fluorescence intensity relative to the number of cells in a sample.

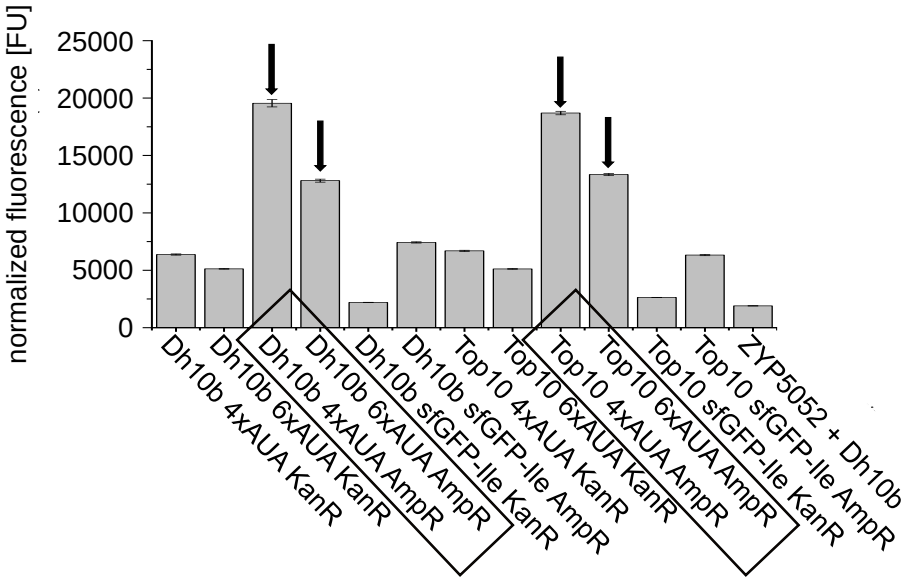


Figure 42 Fluorescence assay of different sfGFP mutants bearing AmpR or KanR in *E. coli* Dh10b and Top10. Bacterial cells were supplied with plasmids haboring kanamycin (KanR) or ampicillin resistance (AmpR), IleRS (W905R) mutant, modified $tRNA_{UAU}^{Ile}$, and particular sfGFP variants: sfGFP(6xAUA) = 6xAUA, sfGFP(4xAUA) = 4xAUA, sfGFP(I123T, I128T, I136T, I152V, I188A) = sfGFP-Ile. Arrows highlight the best experimental setups that will be the starting point for further operations. Fluorescence was normalized to corresponding OD_{600}. Five biological replicates were used for each measurement.

No significant difference could be observed between the appointed strains (**Figure 42**). This is not surprising since they are closely related and have the same genotypes. Interestingly, cells of one setup having the KanR produced less fluorescent protein, compared to the one with AmpR. It has been described by Liu *et al.* that supplementation of 50 mg/L and higher concentration of Kan decreases the transcription level of the gene encoding GFP by over 50 %, while the transcription of the KanR gene increases over 2.5-fold.[262] Although the corresponding experiments were conducted in other than that used in our study, the negative effect of Kan on gene expression level can be assigned to our case. 50 mg/L was an established concentration of Kan added to the growth medium in our lab. Owing to the current observation, we will lower the concentration of this aminoglycoside antibiotic to 25 mg/L in future application. One important determination is that the expression efficiency of reporter mutant without rare codon is ten-fold worse compared to the similar construct containing AUA codons at appropriate Ile residues (**Figure 42**). It is described that the translation rate of rare codons is predicted to be low which correlates with the less abundant cognate tRNAs.[21] As we increased the concentration of the relevant components by supplying the cells with the necessary *E. coli* $tRNA_{UAU}^{Ile1}$:IleRS(W905R) pair, the catalytic affinity to decode rare codon is simultaneously enhanced. Therefore, expression of mutants harboring the rare codon AUA for

Ile decoding is preferred over the one having the same mutations but encoded all Ile residues by frequent codons, AUC and AUU. Besides, sfGFP(4xAUA) exhibits higher fluorescence in relation to sfGFP(6xAUA) (**Figure 42**). To recap, sfGFP(4xAUA) equals to sfGFP(I123T, I128T, I136T, I152V, I161L, I167V, I188A) containing AUA rare codons at the four remaining Ile positions and sfGFP(6xAUA) corresponds to sfGFP(I123T, I128T, I136T, I152V, I188A) with the AUA rare codon encoding all of six remaining Ile residues. Previous results demonstrated about 50 % decrease in fluorescence of sfGFP(I123T, I128T, I136T, I152V, I161L, I167V, I188A) in comparison to the outcome of sfGFP(I123T, I128T, I136T, I152V, I188A) expressed in BL21 (DE3) gold (**Figure 39A**). A possible explanation for this contrary observation is that the higher number of rare codons within one gene probably constitute a ramp for expression and consequently decelerates the translation process. Moreover, the rare codon usage is a context dependent matter since the codon pair bias implies a selective distribution of nucleotides neighboring a particular codon.[263] It is assumed that the potential interactions of tRNAs in the A- and P-sites of the ribosome can affect the translation rate. Thus, substitution of AUU encoding I161 and I167 residues by AUA might cause unwanted interactions between the resulting tRNAs at the translation apparatus.

Based on these facts, we proceeded with the synthesis of the ultimate screening plasmids according to the described assembly strategy (**Figure 41**). Verified libraries were assembled with appropriate backbone setup. Since our test revealed a functioning system with the modified frequent $tRNA^{Ile1}_{UAU}$, we continued with this tRNA and did not integrate $tRNA^{Ile2}_{UAU}$ to our project. Due to the lack of time, we could only proceed with one reporter protein strategy. Incorporation of the split sfGFP method requires supplementary plasmids. Therefore, we exclusively consider the first reporter strategy in order to keep our screening system as simple as possible. For this purpose, both sfGFP(6xAUA) and sfGFP(4xAUA) constructs were used. sfGFP(I123T, I128T, I136T, I152V, I188A) was served as a negative control. With the same setup, only sfGFP containing AUA codon gain intact folded proteins. Before induction, Ile is depleted from growth medium and TfeGly is supplied. Only expression that efficiently incorporates TfeGly at AUA positions yields fluorescent protein. AmpR was applied as selection marker for all resulting plasmids. The following table summarizes the plasmids that were finally obtained in the context of screening for promising IleRS candidates from the libraries.

Table 4 Plasmid constructed for screening.

Plasmid label	Abbreviation
pPAB26 IleRS(Ala10, W905R)A1 H_6-sfGFP(14AUA, 47V, 98AUA, 123T, 128T, 136T, 152V, 161AUA, 167V, 188A, 229AUA) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen A1 4xAUA
pPAB26 IleRS(Ala10, W905R)A1 H_6-sfGFP(14AUA, 47AUA, 98AUA, 123T, 128T, 136T, 152V, 161AUA, 167AUA, 188A, 229AUA) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen A1 6xAUA
pPAB26 IleRS(Ala10, W905R)A1 H_6-sfGFP(123T, 128T, 136T, 152V, 188A) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen A1 sfGFP
pPAB26 IleRS(Ala10, W905R)A2 H_6-sfGFP(14AUA, 47V, 98AUA, 123T, 128T, 136T, 152V, 161AUA, 167V, 188A, 229AUA) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen A2 4xAUA
pPAB26 IleRS(Ala10, W905R)A2 H_6-sfGFP(14AUA, 47AUA, 98AUA, 123T, 128T, 136T, 152V, 161AUA, 167AUA, 188A, 229AUA) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen A2 6xAUA
pPAB26 IleRS(Ala10, W905R)A2 H_6-sfGFP(123T, 128T, 136T, 152V, 188A) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen A2 sfGFP
pPAB26 IleRS(Ala10, W905R)A3 H_6-sfGFP(14AUA, 47V, 98AUA, 123T, 128T, 136T, 152V, 161AUA, 167V, 188A, 229AUA) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen A3 4xAUA
pPAB26 IleRS(Ala10, W905R)A3 H_6-sfGFP(14AUA, 47AUA, 98AUA, 123T, 128T, 136T, 152V, 161AUA, 167AUA, 188A, 229AUA) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen A3 6xAUA
pPAB26 IleRS(Ala10, W905R)A3 H_6-sfGFP(123T, 128T, 136T, 152V, 188A) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen A3 sfGFP
pPAB26 IleRS(W905R)E1 H_6-sfGFP(14AUA, 47V, 98AUA, 123T, 128T, 136T, 152V, 161AUA, 167V, 188A, 229AUA) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen E1 4xAUA
pPAB26 IleRS(W905R)E1 H_6-sfGFP(14AUA, 47AUA, 98AUA, 123T, 128T, 136T, 152V, 161AUA, 167AUA, 188A, 229AUA) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen E1 6xAUA
pPAB26 IleRS(W905R)E1 H_6-sfGFP(123T, 128T, 136T, 152V, 188A) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen E1 sfGFP
pPAB26 IleRS(W905R)E3 H_6-sfGFP(14AUA, 47V, 98AUA, 123T, 128T, 136T, 152V, 161AUA, 167V, 188A, 229AUA) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen E3 4xAUA

pPAB26 IleRS(W905R)E3 H_6-sfGFP(14AUA, 47AUA, 98AUA, 123T, 128T, 136T, 152V, 161AUA, 167AUA, 188A, 229AUA) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen E3 6xAUA
pPAB26 IleRS(W905R)E3 H_6-sfGFP(123T, 128T, 136T, 152V, 188A) $tRNA^{\mathbf{Ile1}}_{\mathbf{UAU}}$	pScreen E3 sfGFP

3.2.5 Generation of a selection strain

As mentioned before we aimed to generate a strain that can regulate Ile rare codon assignment upon requirement. This strain should be able to grow applying its endogenous IleRS and thus, all AUA codon is proteome-wide assigned with Ile. Thus, the essential cellular components are synthetized under nutrient-rich conditions with all of the 20 cAA to preserve their functionality. Once the strain reaches its mid-log phase, the activity of the genomic IleRS should be inactivated. Thenceforth, cells are exclusively grown by supplied IleRS mutants of a library. At the same time Ile is depleted from medium and expression of reporter protein containing AUA is induced in the presence of TfeGly. Ideally, only cells capable to efficiently incorporate the ncAA in response to the rare codons can produce fluorescent protein for selection. This assumes that TfeGly does not disturb protein folding and other 19 cAA substitutions will abolish the β-barrel structure. As our previous test revealed the loss of fluorescence by extreme reduction of Ile residues, it depends on the impact of TfeGly on protein fluorescence whether this concept would be a high-throughput screening assay. In this regard, we have no opportunity to test since there is no synthetase available that specifically translates TfeGly. Therefore, our strategy is to perform matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) analysis of the lysate of clones, showing different fluorescence intensity which was previously sorted e.g., by FACS.

The requirements of our strain can be accomplished by several steps (**Figure 43**). First of all, the strain needs an inducible control of the synthetase activity. The most straightforward solution is the regulation by temperature. A579P mutation in *ileS* (gene encoding IleRS) is found to cause a thermal instability at 42 °C.[264,265] Introduction of this mutation into *ileS* would allow the inactivation of IleRS upon transfer to the restrictive temperature. Furthermore, a W905R mutation is required to be inserted to *ileS* prior to the genomic deletion of *tilS*, in order to ensure the AUA codon deciphering of an *tilS* knockout strain for survival. Finally, the constructed strain must be auxotrophic for Ile, meaning it is unable to synthetize Ile due to disruption of the biosynthetic pathway. Thus, cells are driven toward incorporation of the cAA analogue upon cultivation in defined growth media containing the ncAA and in absence of the canonical counterpart. There is one important point to consider in the chronology of strain

alteration. By removing *tilS* from the modification pathway, the AUA codon becomes unassigned which is lethal for the cells. Hence, this deletion should be performed in the final step. A schematic workflow is represented in **Figure 43**.

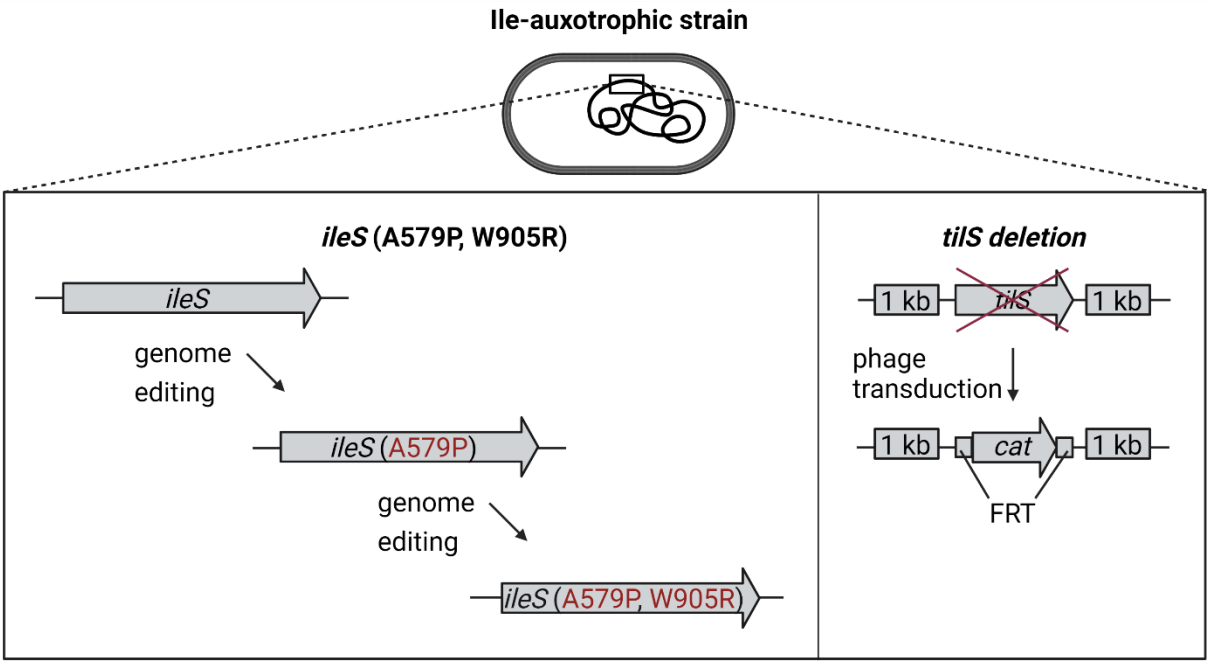


Figure 43 Requirements for the suitable selection strain and a schematic implementation.

First, we aimed to introduce the mutations into the genomic *ileS* of Ile-auxotroph strains. For this purpose, CAG18431 and ML6 were chosen based on their distinct advantageous properties. Both auxotrophic strains were generated by removal of different essential parts along the biosynthesis pathway of Ile (**Figure 44**).[266] CAG1841 is a MG1655 derivative (*E. coli* K12 strain) in which the gene for dihydroxy-acid dehydratase (*ilvD*, **Figure 44**) is replaced by the transposon Tn*10* encoding tetracycline resistance.[267,268] On the other hand, ML6 is derived from the *E. coli* C43(DE3) strain that is frequently used for overproduction of membrane proteins and toxic proteins.[269] The plasmid stability is significantly enhanced in this BL21 derivate and therefore allows the transformation of highly toxic genes into the expression strain.[270] Ile-auxotrophy is achieved by deletion of the genes encoding branched-chain amino acid aminotransferase (*ilvE*, **Figure 44**) and valine pyruvate aminotransferase (*avtA*) in C43(DE3).[271] *E. coli* K strain is commonly used for cloning and screening purpose while *E. coli* B strain is deployed for high yield expression. The suitability of those strains cannot be predicted in our case as we supplied a selection, screening, and expression system on one plasmid. Therefore, both of them are considered for the following experiments. Verification of Ile-auxotrophy in selected strains was already performed by colleagues of the Budisa group.

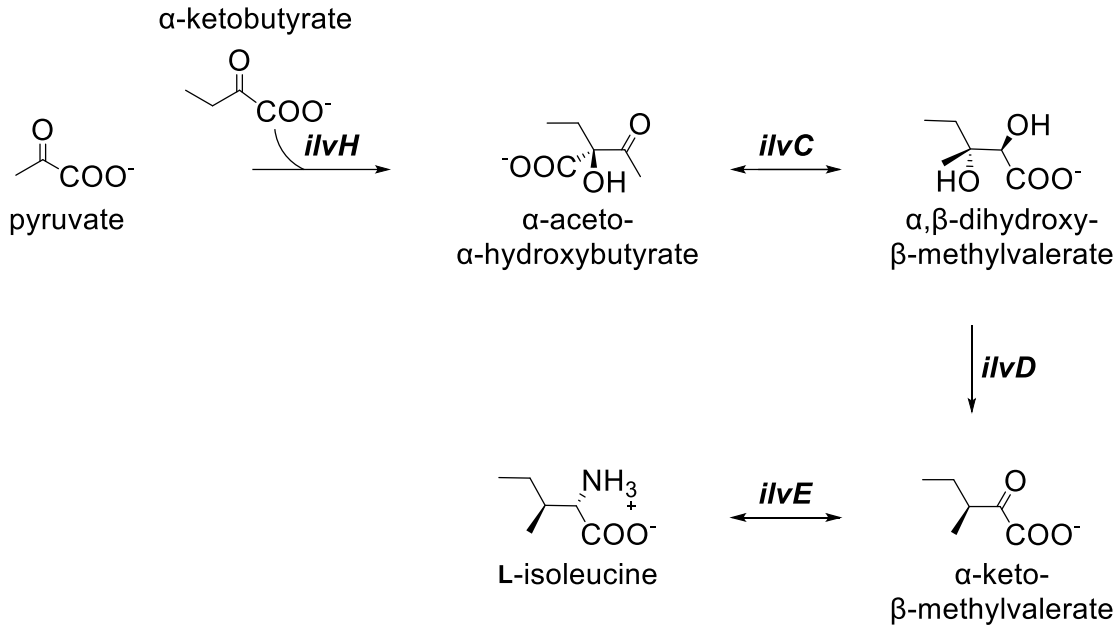


Figure 44 A section of the biosynthetic pathway of Ile in *E. coli*. The responsible genes for enzymes catalyzing each reaction step are acetolactate synthase (***ilvH***), ketol-acid reductoisomerase (***ilvC***), dihydroxy-acid dehydratase (***ilvD***), and branched-chain amino acid aminotransferase (***ilvE***).[266]

In order to prove the thermosensitivity of the IleRS(A579P) mutant, several tests were conducted prior to genomic editing. It should be noted that A579 is located on the 18-membered α-helix bearing residues in proximity of the cognate substrate at the activation site. Especially, one residue of the hydrophobic core, W569, is at the other end of this secondary structure element. A579 is the third to last residue of the α-helix and only 20.4 Å away from the substrate side chain. Here it is quite conceivable that this residue can have a huge impact on the activation affinity of the synthetase.

In this regard, H_6-TEV-IleRS, H_6-TEV-IleRS (A579P) as well as H_6-TEV-IleRS(Ala10) were expressed at 22 °C and purified (by IMAC) for comparison. Known from the crystallization part, we cleaved the N-terminal H_6-tag to avoid any disturbance, e.g., by the Met-oxidation of initiator Met. Products before and after TEV cleavage were analyzed by SDS-PAGE, revealed a pure sample for further experiments (**Figure 45A**). A loss of 2.1 kDa in mass could be observed in LC-MS analysis and thus verified the successful tag removal (data not shown). Using the malachite green assay (**Section 6.3.7**), the aminoacylation efficiency of the IleRS(A579P) mutant was determined after incubation at different temperature in order to check the thermosensitivity of this mutant (**Figure 45B**). The wildtype enzyme was applied as a standard of 100 % activity at 22 °C. In general, the IleRS(A579P) mutant activates Ile less

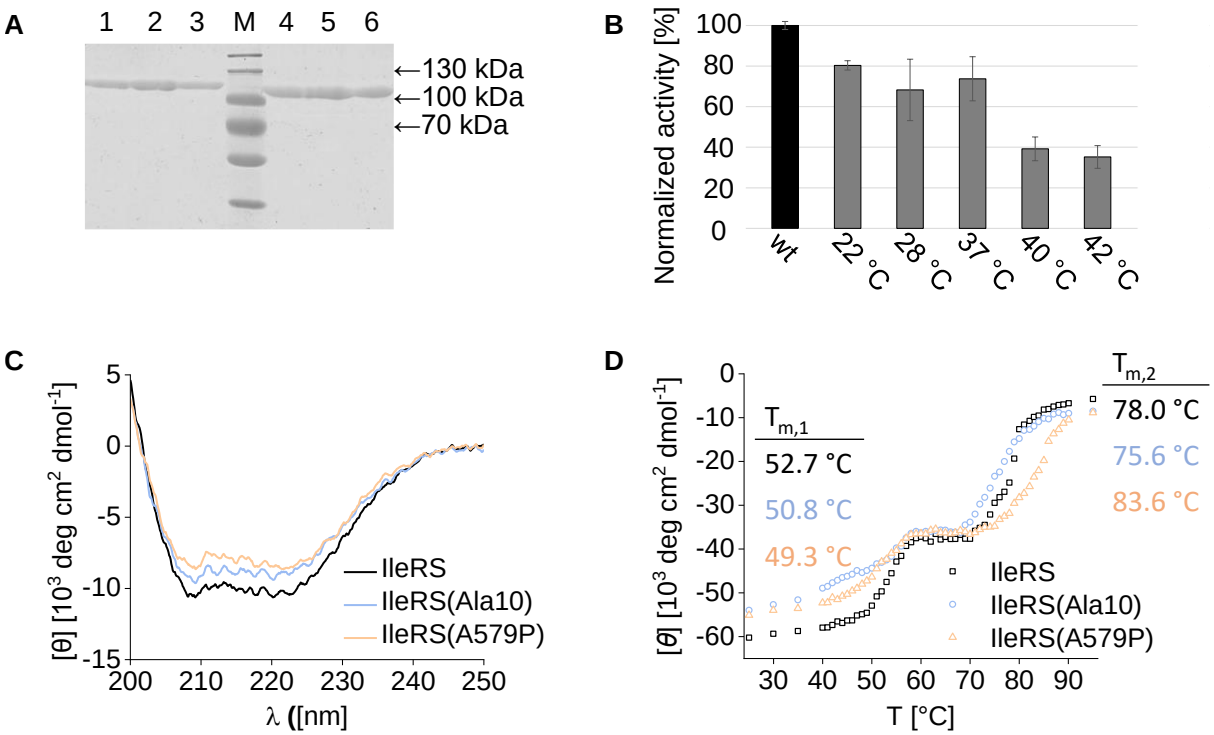


Figure 45 Property investigations of IleRS mutants by A) SDS-PAGE analysis, B) malachite green activity assay, C) spectroscopic CD measurement for determination of the secondary structure, and D) spectroscopic CD measurement for thermal stability analysis. A) Purified proteins were applied for the analysis on 12 % SDS gel. Expected size of H_6-TEV-IleRS (fraction 1) amount to 107.0 kDa, H_6-TEV-IleRS(Ala10) (fraction 2) 106.6 kDa, H_6-TEV-IleRS(A579P) (fraction 3) 107.0 kDa, and after cleavage IleRS (fraction 4) 104.4 kDa, IleRS(Ala10) (fraction 5) 104.0 kDa, IleRS(A579P) (fraction 6) 104.4 kDa. Corresponding size of marker (M) bands are shown with arrows. **B**) Malachite green assay of IleRS(A579P) (grey columns) expressed at 22 °C in comparison to analogously produced IleRS (wt, black column). IleRS(A579P) mutant was incubated at different temperature (22 °C, 28 °C, 37 °C, 40 °C, 42 °C) for 15 min prior to the application in colorimetric assay.10 µM enzyme was used with 5 µM Ile as the substrate. Values represent the average of five samples with the SD as error bars (blank subtracted). **C**) CD spectra of IleRS, IleRS(Ala10), and IleRS(A579P) at 20 °C. 1 µM of each enzyme was prepared in 25 mM phosphate buffer pH 8.0 containing 100 mM KCl. Curves are accumulated results of five measurements. **D**) Thermal denaturation curves of IleRS, IleRS(Ala10), and IleRS(A579P) recorded at 222 nm from 25 °C to 95 °C at a rate of 3 °C per min. The denaturation process is divided in two parts. For each separate phase, T_m values were calculated from a non-linear fit with a Boltzmann function applying the Levenberg-Marquardt algorithm. Analysis of data was processed in OriginPro program.

efficiently than IleRS. The treatment with temperature up to 37 °C did not significantly decrease the activity of the mutant. Incubation at 40 °C and higher led to 50 % of loss in activation efficiency. Even after a heat shock at 42 °C about 35 % of IleRS(A579P) was still active. There are two possible explanations for this observation. First, the A579P mutation is in principle able to induce a limitation in the enzyme's activity but cannot achieve a total

inactivation by temperature. Second, the mutant only loses its activity during the incubation at 42 °C and recovers after the transfer to non-restricted temperature. The latter is more likely since a wildtype enzyme is capable to refold after denaturation.[272] However, we will deplete Ile from the cultivation media at certain stage and supply TfeGly instead. Ile is therefore not available for the cells to incorporate into AUA codon as a competing reaction. Besides, the rare codon reassignment will be performed at 42 °C and enables a strong restriction to the catalytic function of A579P mutant. Hence, this residual activity of wildtype enzyme can be neglected.

According to **Section 6.4.7.**, CD measurements of the IleRS mutant were conducted in the next step to determine the changes in their secondary structures (**Figure 45C**). The results show a slight decrease in the α-helix content of the mutants in comparison to the wildtype. IleRS(A579P) possesses a greater loss in α-helix propensity among the two mutants. This is in agreement with our expectation since Pro is established as a potent breaker of secondary structures. As mentioned before, A579 is the third to last residue of an α-helix. It is most likely that this sub-structure is truncated by the substitution with Pro. However, the range of change is very small and might be a result of the measurement inaccuracy. Such a large macromolecule mainly consists of α-helices wherefore, a single mutation is expected to have minor impact on the overall secondary structure content of the protein.

In addition, thermal stability of IleRS mutants was investigated by CD (**Figure 45D**). The change of molar ellipticity ([θ]) at 222 nm was monitored as a function of the temperature. Resulted curves showed the denaturation of all constructs in two phases. It is in accordance with the folding hypothesis of large, single-chain proteins composed of multiple domains.[273] This states that folding is proceeded in two steps: first, domains fold independently and rapidly; and then associate in a subsequent slower step. The rate determining part typically governs the acquisition of enzyme activity. In this sense, the first denaturation stage implies the inactivation of IleRS catalytic function while the second phase is in the mean of entire structure destruction. We assumed that the association of N- and C-terminal Rossman fold is disrupted by the temperature and hence, leads to the loss of enzyme activity. As expected, A579P mutant is instable and exhibits the lowest melting temperature ($T_{m,1}$ (IleRS(A579P)) = 49.3 °C) in the first stage, followed by the post-transfer editing defective Ala10 mutant ($T_{m,1}$ (IleRS(Ala10)) = 50.8 °C). Wildtype enzyme is inactivated at 52.7 °C. However, in the second step of denaturation process, IleRS(A579P) showed the best resistance to temperature. The $T_{m,2}$ (IleRS(A579P) is 83.6 °C, 5.6 °C higher than from the wildtype. Complete denaturation of Ile(Ala10) is achieved by the lowest temperature, $T_{m,2}$ (IleRS(Ala10)) of 75.6 °C. By introducing ten Ala residues, several interactions within the editing site of IleRS are disrupted and thus, alters the CP1 scaffold. As each single domain of IleRS plays a role in structure stabilization,

Ala10 substitution might cause this significant contribution to the protein denaturation at higher temperature.

To conclude, the A579P offers a good opportunity for regulation of IleRS activity by temperature. It is important to incubate the cells at a maximum of 37 °C for growth, then increase to 42 °C for the enzyme inactivation and retain the restricted temperature for reassignment.

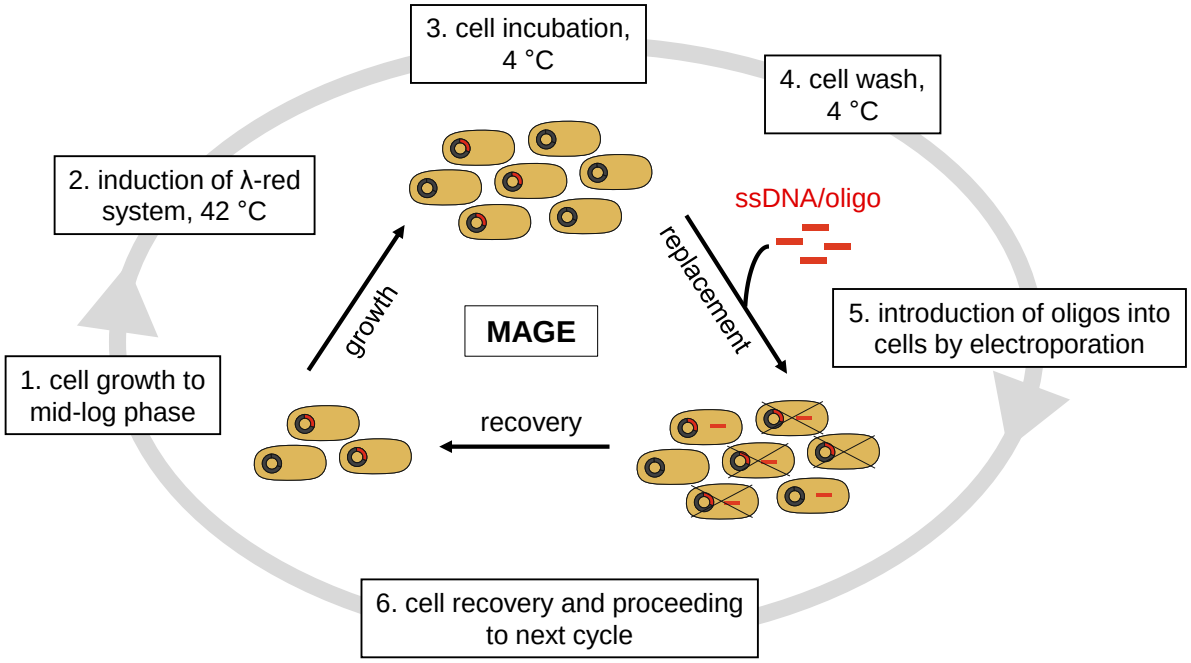


Figure 46 MAGE cycling workflow according to Wang *et al*.[274]

The insert of A579P mutation was attempted by modified MAGE (Multiplex Automated Genome Engineering), a high-thoughput genome editing method developed to simulaneously introduce several chromosomal changes in to a single cell or across a cell population.[274] This cyclical and scalable engineering approach represents an opportunity to construct both highly modifed genetic components and investigate large sequence landscapes by directed evolution in a semirational manner. To this end, MAGE iteratively applies single-stranded DNA (ssDNA) or oligo in a recombination-based genetic engineering (recombineering) protocol. The process is mediated by β protein of the bacteriophage λ-Red system, which consists of the three genes: *γ*, *β*, and *exo*.[275] Heterologous expression revealed corresponding gene products, Gam, Beta, and Exo. Gam inhibits the host RecBCD exonuclease V activity to prevent degradation of introduced foreign linear double-stranded DNA fragments (dsDNA). The single-stranded DNA (ssDNA) binding protein Beta promotes recombination via direction and hybridization of the linear fragment to its genomic complement, the lagging strand at the replication fork. In this

way, the target gene region is mutated by mismatch with the supplied ssDNA. Exo is a 5' to 3' exonuclease that processes linear dsDNA, leaving 3' overhangs as substrates for subsequent recombinogenesis. In MAGE, this protein is not relevant as only ssDNA is used. The general scheme of MAGE process is to put cell population in continual cycles of recombineering in order to enrich the populations carrying oligo-mediated allelic replacements (**Figure 46**). Genotyping or phenotyping are used to analyze subpopulations from MAGE cycles. Wang and coworkers developed an integrated prototype device that automates the MAGE process for a rapid and reliable continuous cellular programming.[274] For our study, such an instrument is not available. At this point, we intended to perform MAGE cycles manually.

We have designed an MAGE-oligo harboring the A579P mutation for recombineering that consists of 90 bp since the highest incorporation efficiency for ssDNA in the 70–90 bp range.[276] The codon decoding 579P is located in the middle of the sequence and hence to satisfy the minimum binding size for Beta (~30 bp).[277] According to the **Section 6.2.4.1**, five MAGE cycles were performed in CAG18431, and ML6, respectively. The λ-Red system was supplied on the auxillary plasmid pKD46 or pORTMAGE. pKD46 carries the λ-Red operon from bacteriophage λ as described above. Expression of the necessary genes is tightly regulated by pAraB promoter to prevent unwanted recombination under noninducing conditions. The temperature-sensitive origin of replication repA101ts was used to eliminate this plasmid from the strain after usage. pORTMAGE also harbors genes for the λ-Red recombinase enzymes as pKD46, and additionally, the dominant-negative mutator allele of the *E. coli* methyl-directed mismatch repair (MMR) protein MutL(E32K). As an essential part of the MutHLS complex responsible for MMR, MutL acts mainly to recruit the MutH endonuclease to the DNA damage sites. MutL(E32K) is a heat-sensitive mutant of MutL, which is inactivated upon higher temperature. Moreover, this mutant cannot be complemented by the native protein. In the absense of MutL, the MMR machinery is inactivated and hence enables the transient suppression of mismatch repair during oligonucleotide integration by MAGE process. Expression of both the λ-Red system and MutL(E32K) is controlled by the cI857 temperature-sensitive λ repressor which is removed at higher temperature. In other words, at nonpermissive temperatures (30-34 °C) the native MMR system is funtioning properly and the mutation rate of the host cells is limitted to wildtype level to avoid background mutations. By increasing the temperature to 42 °C the λ-Red recombinase enzymes as well as MutL(E32K) are expressed and therefore, the mutation rate is significantly increased. This allows the insertion of mutation by MAGE process without disturbance of the endogenous MMR system.

For the insertion of the A7579P mutation into *ileS* both plasmids were seperately used for λ-Red system supplementation. Different oligo concentrations were applied for transformation (see **Section 6.2.4.1**). After five MAGE cycles, 96 clones of each trial setup were selected for analysis. Single colonies were resuspended in LB media and spotted in the same amount on

two different LB-agar plates to duplicate the colony. One plate was incubated at 30 °C and the other at 42 °C. Cell colonies exhibiting hugh differences in growth behavior were picked for sequencing to verify the presense of the A579P mutation in the genome. It is expected that cells harboring the A579P mutation have nearly inactivated IleRS at 42 °C and thus struggle to survive. The resulting LB-agar plates after overnight incubation are shown in **Figure 47**.

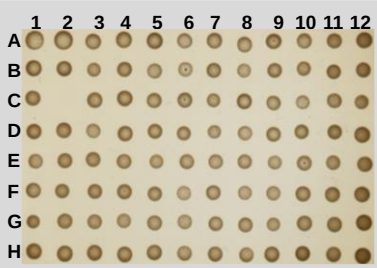
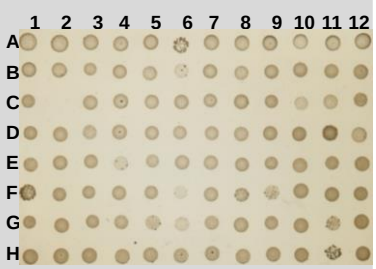
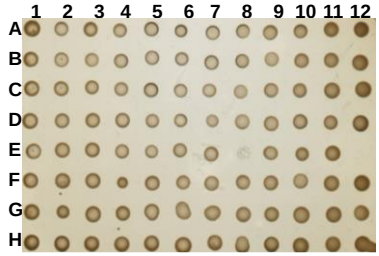
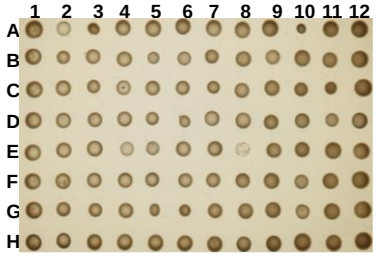
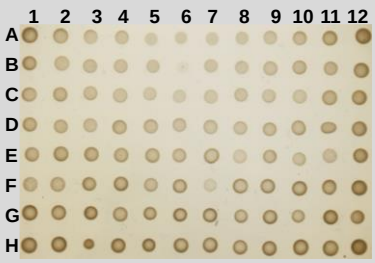
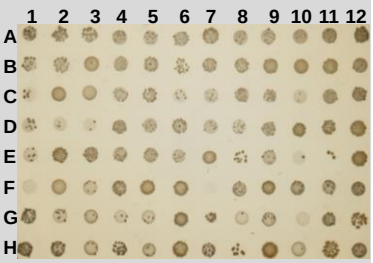
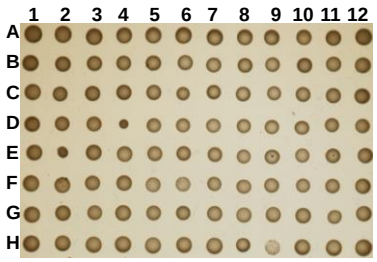
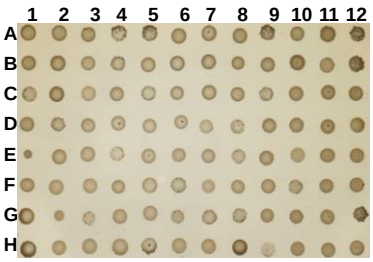
Strain	Supplemented plasmid	30 °C	42 °C
CAG18431	pKD46		
	pORTMAGE		
ML6	pKD46		
	pORTMAGE		

Figure 47 Results after five MAGE-cycles. The designed oligo and pKD46 or pORTMAGE were transformed into CAG18431, and ML6, respectively. Homogeneous spots indicate proper cell growth while isolated colonies are the results of a disruption in cell fitness.

Contrary to our expectations, pORTMAGE containing the ability for inactivation of endogenous MMR system could not enhance the mutation rate. Both strains showed higher amount of cells growing seperately when MAGE was assisted by the pkD46 plasmid. ML6 possesses a better

yield in mutated cells compared to CAG18431. This is obvious since ML6, an *E. coli* B strain derivate, is a commonly used expression strain. Therefore, the essential λ-Red recombinase enzymes are expressed with higher yield and are more concentrated for recombineering events. As a C43 (DE3) based strain, it is not surprising that ML6 stabilizes the auxiliary plasmid better than CAG18431 and is able to retain the plasmid for target gene expression. Almost all of the colonies produced by pKD46 in ML6 showed significant difference in growth behavior at the appointed temperature (**Figure 47**, part 3).

However, the sequencing results of promising colonies showed the presence of the A579 residue. No mutated *ileS* could be observed. We assumed that the introduced mismatch at position 579 first causes the irregular growth on the plates. By incubation in the liquid medium, meaning after several cell divisions, cells somehow repair this mutation to avoid the thermosensitivity.

We further applied a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/ CRISPR-associated protein 9 (Cas9) based approach for mutation of *ileS*. The CRISPR/Cas9 system is adapted from an evolved RNA-mediated adaptive defense system of bacteria and archaea, which protects them against invading viral DNA.[278] After an infection recovery, small pieces of the viruses' DNA is installed into the CRISPR array as protospacers. This kind of library supports the cells to a rapid response upon a reinfection with the same or closely related pathogen. To be more precise, the sequence information of CRISPR array is used to produce a large precursor CRISPR RNA (pre-crRNA) that is specifically recognized and targeted by Cas enzymes. These nucleases process the pre-crRNA to yield smaller mature crRNAs that complementary to the protospacer sequences of invasive genetic elements. Together with a trans-activating CRISPR RNA (tracrRNA) the crRNA forms a crRNA:tracrRNA duplex to guide Cas proteins to the targets for the cleavage of protospacer sequences of the invading virus. By now, numerous CRISPR/Cas systems containing different classes of Cas enzymes have been developed for addition, deletion, or modification of DNA sequences.[279] Cas9 is an endonuclease that can be most commonly found in this type of genome editing systems. In 2012 the Charpentier and Doudna labs could simplify the system by the combination of tracrRNA and crRNA into a single guide RNA (sgRNA).[280] Cas9 was shown to be equally effective towards binding of this synthetic sequence as to the crRNA:tracrRNA duplex. However, each target requires at least the design of a new sgRNA that can be the bottle neck of the experiment. Besides, off-target effects of the CRISPR/Cas9 system limit the targetable sequence space.[281] To overcome these restrictions, Zhao and coworkers developed the CRISPR/Cas9-assisted sgRNA-free one-step genome editing strategy (CAGO, **Figure 48**).[282] They were able to evolve a universal CRISPR/Cas9 recognition sequence, named N20PAM. N20PAM is supplied on an editing cassette containing homologous sequences of destination site and a marker for selection. Once the cassette is introduced into the target locus of the

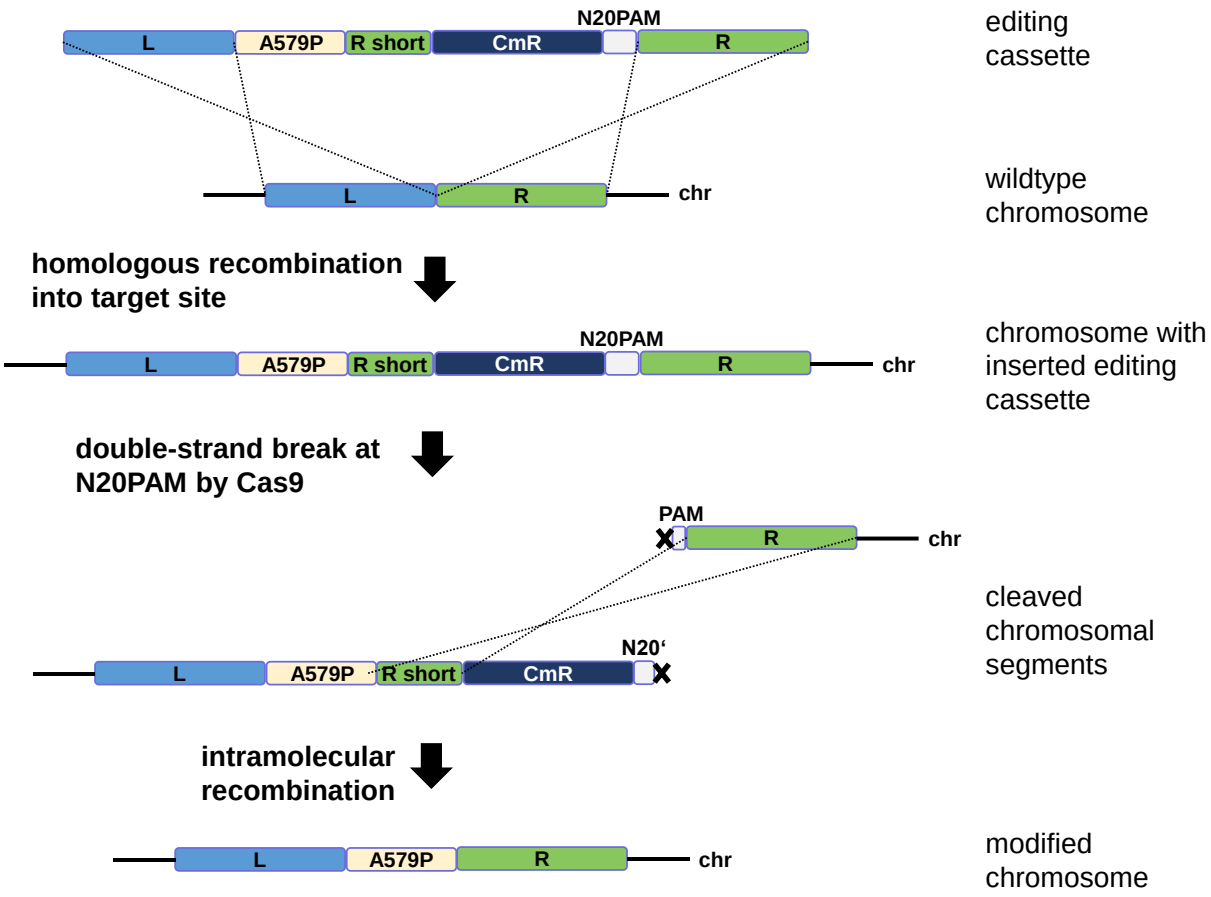


Figure 48 Schematic presentation of CAGO genome editing principle used in this study. Homologous fragments left (L) and right (R) from the chromosome insertion site equals to the sequence of *ileS* upstream (L) and downstream (R) of the A579 residue. R_short is the first 40 bp of the right homologous fragment containing the A579P mutation that is separately highlighted in front of the R-short fragment (beige). Chloramphenicol resistance was used for selection (CmR). Figure is adapted from Zhao *et al*.[282]

genome by homologous recombination, Cas9 recognizes and breaks the double-strand at N20PAM. Subsequent intrachromosomal recombination completes the scarless genome editing event. Genes encoding all of the necessary enzymes for this strategy are supplied on the pCAGO plasmid (**Figure 49**). These include the inducible λ-Red system (described in MAGE), Cas9 and sgRNA with the designed N20 sequence for specific targeting of N20M recognition site.

In this study, we have chosen *E. coli* MDS 42, Dh10b, Top10 for the A570P mutation of the *ileS*. Since the previous MAGE approach failed to edit the genome, we speculate that the genetic knockout in the Ile synthetic route might impair the mutation rate of *ileS* even though they originate from two distinct biological pathways. Hence, we decided to use strains that are more suitable for genome engineering. MDS42 is known to have a greatly reduced genome, with 704 deleted nonessential genes, leading to high electroporation efficiency and accurate propagation of recombinant genes and plasmids that were unstable in other strains.

Interestingly, this strain is free of mutation generating insertion sequence (IS) elements, therefore prevents insertion of knockout fragment into other positions in the genome.[283] Dh10b and Top10 are used to test the functionality of the screening system (**Section 3.2.4**).

The editing cassette was generated by golden gate assembly of four fragments: the left (L, 502 bp), and right homologous region of A579 residue in *ileS* (R, 503 bp), the first 70 bp of the right homologous fragment containing the A579P mutation (A579P, R-short) and the antibiotic resistance cassette with N20PAM recognition site (CmR, N20PAM, 1080 bp) (**Figure 48**, see **Section 6.2.4.2** for assembly details). The assembled editing product (~ 2.2 kb) was successfully verified by sequencing analysis. Cells containing the pCAGO plasmid were induced for expression of the λ-red system by IPTG at the early log phase (OD_{600} = 0.15 - 0.20). Cells were made electrocompetent and the editing cassette is introduced into these bacteria by electroporation. After 2 h recovery, cells were plated on LB-agar plate with Cm and Amp for selection of target clones. **Figure 49** visualizes the entire CAGO genome editing strategy. (For more experimental details see **Section 6.2.4.2**.)

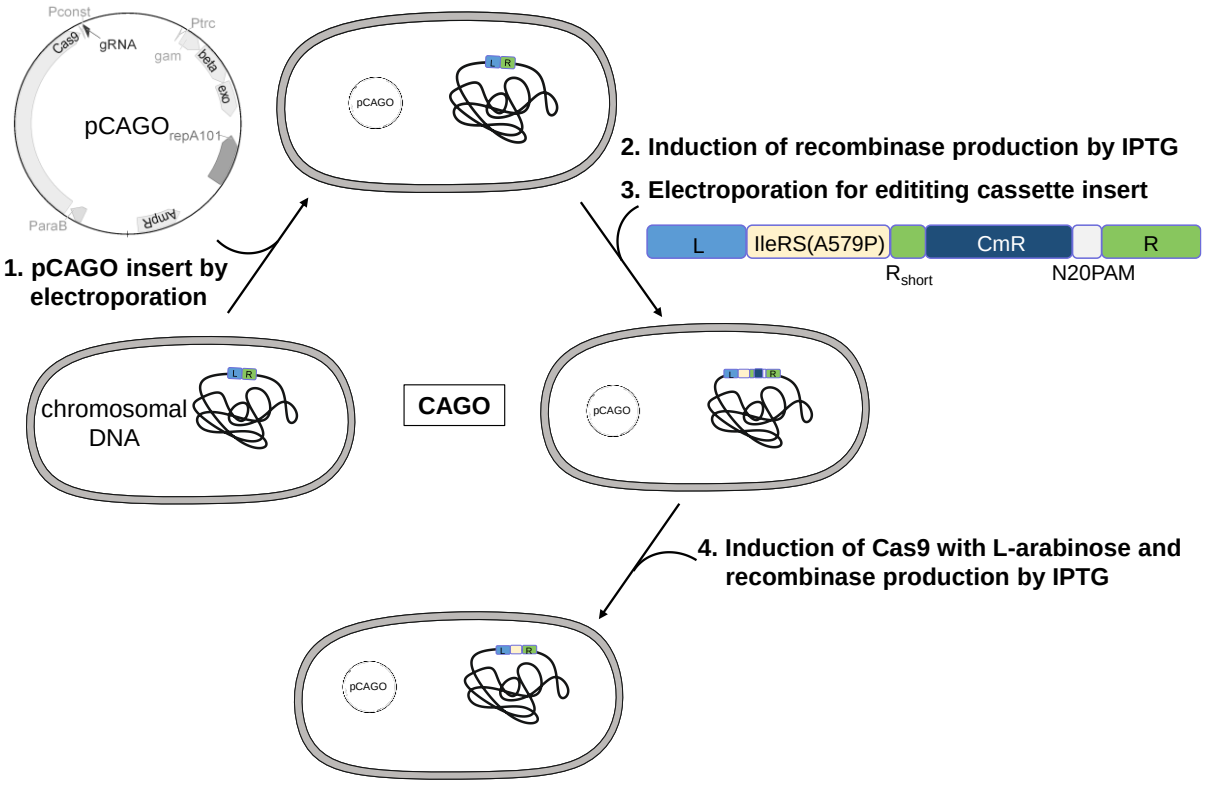


Figure 49 Workflow of CAGO genome editing technique according to Zhao *et al*.[282]

400 ng of the editing cassette was initially used for introduction into the target strain (**Section 6.2.4.2**). However, no clones could be obtained upon selection with chloramphenicol (Cm). This indicates that homologues recombination did not work, and the strain did not exhibit the

genomic CmR. We have further attempted to vary the inducer (IPTG, 0.1 mM, 0.2 mM, 0.3 mM, 0.5 mM, 1 mM) concentration for the expression of the λ-red system, which catalyzes the recombination event. Different amounts of the editing cassette were used, 800 ng, 1000 ng, and 1500 ng, in order to enhance the substrate availability. Again, no cell colonies could survive upon selection with Cm. To further enhance the homologous recombination efficiency, we extend the length of the flanking sequences from 500 bp to 1000 bp, since the efficiency is positively correlated with the length of the homologous sequences.[284] The CAGO technique was then deployed with different concentrations of IPTG (0.1 mM, 0.2 mM, 0.3 mM, 0.5 mM, 1 mM) and distinct amount of the assembled editing cassette (500 ng, 1000 ng, 1500 ng). Despite all efforts we were not able to introduce the editing cassette into the selected strain. After we have recapped the entire CAGO strategy, we have figured out the reason of the failure. By insertion of the editing cassette into the target site of *ileS*, the essential gene for survival is interrupted by the sequence of CmR and N20PAM. Therefore, no clone can be detected since the Ile translation of cells containing the editing cassette is disabled.

We have reassembled a new 3200 bp editing cassette, which consists of a 502 bp left homologous arm (L, base position 1268 – 1770 of *ileS*), a 1083 bp fragment within *ileS* including the mutation (A579P, base position 1771 – 2853 of *ileS* and stop codon), a short right homologous arm (R_{short}, the first 45 bp of R), a selection marker resistance gene (CmR, 1037 bp) with a specific 23 bp N20PAM, and a 510 bp right homologous arm (R, sequence downstream of *ileS*). By this newly designed cassette, *ileS* is not interrupted during the entire CAGO event and cells are capable to survive. Sequencing analysis verified the correct editing cassette sequence including the A579P mutation. The CAGO strategy was subsequently started analogous to the previous attempt. Finally, we were able to insert the editing cassette into the genome, but in bad yield. Only two clones of *E. coli* Top10 cells were obtained from the selection plate containing Cm and Amp. Since Cm was used for selection of cells harboring the editing cassette and Amp for selection of cells containing the pCAGO plasmid, these two clones are deemed to carry the editing cassette with CmR in their chromosomal DNAs. The presence of the introduced cassette was verified genomically by colony PCR using different combination of corresponding primers binding outside and inside of the cassette. It should be noted that these primers bind exclusively at the target sites and have no other complementary sequence in the entire genome. In absence of the editing cassette, no fragment can be produced. **Figure 50A** shows the scheme of amplification and appropriate expected size of the fragments, F1 and F2. The PCR products of both clones revealed the roughly corresponding bands (**Figure 50B**). F1 and F2 were sent for sequencing in order to phenotypically verify the presence of A579P mutation in *ileS* as well as the editing cassette itself. However, the fragments were not sequenceable. This might be due to the low concentration of the submitted DNA.

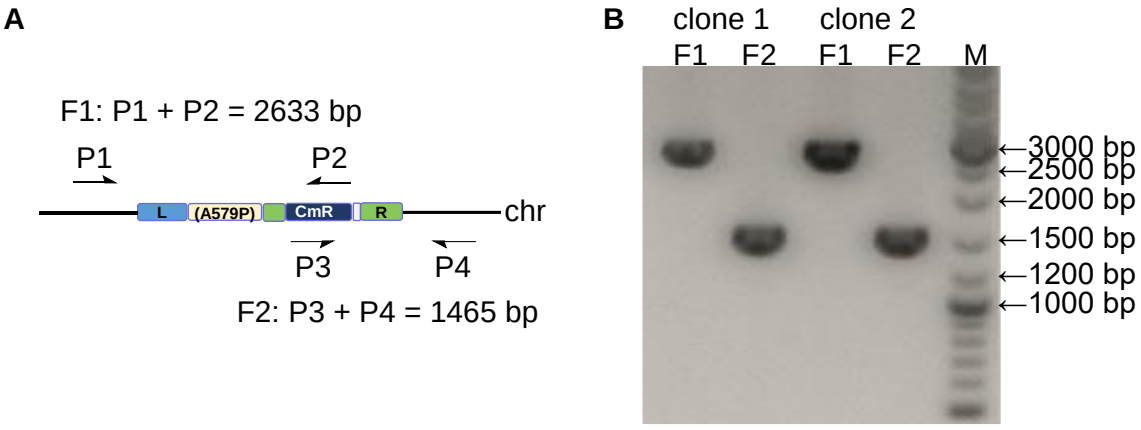


Figure 50 Genomic verification of the editing cassette in *E. coli* Top10. A) Fragments F1 and F2 are produced by the amplification of the target locus using primer pairs P1 + P2, and P3 + P4, respectively. P1 is a forward primer bound upstream of the left homologous arm. P2 is a reverse primer and binds within the CmR cassette. The forward primer P3 is complementary to a short CmR sequence. P4 is a reverse primer used for amplification from outside downstream of the cassette. **B**) Agarose gel electrophorese analysis of DNA fragments, F1 and F2, amplified from two obtained Top10 clones from the homologous recombination event. Particular expected size of the fragments correspond with the calculated values.

To save time, we have proceeded with the last step of the CAGO protocol, in which the selection marker CmR and the CRISPR/Cas9 recognition sequence N20PAM are removed from the target locus by intramolecular recombination (**Figure 49**, step 4). For this purpose, expression of Cas9 and the λ-Red recombinase enzymes were induced. Strain selection was performed on LB-agar supplied with Amp. To exclude the false negative clones still containing the CmR cassette, we innoculated the same single clone in LB media supplied with Amp or Amp and Cm. The desired strain would survive explicitly in LB Amp. Six clones proved to lose the CmR were then submitted for sequencing analysis of the A579P mutation region. Surprisingly, all of them showed wildtype sequence as observed after the MAGE attempts. It is known that editing of essential genes often cause inexplicable phenomena and represent on occasion an insuperable barrier for genome engineering. The reason for this failure remains a mystery and was not further investigated in this study.

Taken together, the strain construction could not be continued since insertion of the thermosensitivity by A579P mutation remains unsuccessful.

3.3 Advanced design of reporter protein

Protein design is a strategy of protein engineering in order to specifically adapt protein properties by target mutation and thus, to gain novel catalytic functions or/and structural improvements.[285] It represents a significant step toward fundamental understanding of origins of the catalysis and protein folding. Strategic introduction of mutations is implemented either at protein level by peptide synthesis or at DNA level by site-directed mutagenesis, and *de novo* gene synthesis, respectively. Rational protein design approaches require the AA sequence predictions by redesigning known sequence and structure of homologous native protein or by the computational *de novo* design, completely from scratch based on paradigm. For proteins of lengths greater than 100 residues permutation of existing gene-encoded sequence by site-directed mutagenesis is the favored approach. It allows low-cost production of rationally designed proteins by equipment, which is available in every gene technical laboratory. Codon manipulations are performed *in vivo* with high success rate. However, the classical site-directed mutagenesis is limited to the standard genetic code repertoire that encode the 20 cAA. Furthermore, replacement with structural similar cAA is not always possible, as for Trp, Met, His, or Pro, which often have key contributions to protein structures and functions.[39] For instance, substitution of the N-terminal Met by any other 19 cAA would lead to the loss of expression of the entire protein since this Met is essential as the start signal for protein translation. In order to fine-tune the protein scaffold, subtle changes is preferably introduced by 'atomic mutations' at the level of single atoms or atom groups.[286] These non-disruptive isosteric modifications enable the detailed investigation of interactions within the internal protein architecture and resulting properties. Potential candidates of such isosteric substituents of the cAA are ncAA that possess close structural similarity to their analogue and therefore, have been widely used in other forms of protein engineering, such as directed evolution (**Section 1.3**).[287] By incorporation of ncAA into proteins the limitations of classical site-directed mutagenesis can be overcome and thus, it is a suitable tool for an advanced protein design strategy.[288]

In this study, we intended to redesign reporter proteins with ncAA to enrich the diversity of these highly versatile biomolecules. cAA analogues were selected mainly based on their known impact and application in other fluorescent proteins (**Figure 51**).

S-allyl-L-cysteine (Sac, **Figure 51A**) is a small naturally occurring sulfur-containing ncAA derived from aged garlic extract and possesses a broad range of advantageous biological activities.[289–294] In peptide synthesis, the allyl side chain has been deployed for protection of the thiol functional group of cysteine (Cys, C, **Figure 51A**).[295] On the one hand, it prevents Cys among others from the irreversible oxidation and on the other hand, it functioned as a reactant in bioorthogonal reactions for protein modifications such as photo-induced thiol-ene

coupling.[296,297] For the reasons stated, Sac has been described as a chemically versatile opportunity for protein-labeling in biology and chemistry.[298]

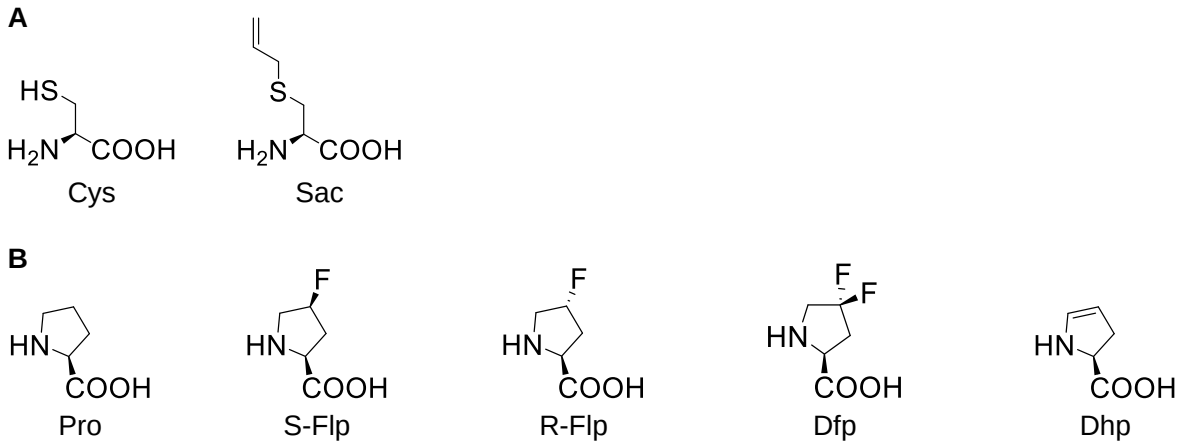


Figure 51 Chemical structures, IUPAC names and common synonyms of selected cAA and their isosteric analogues. A) 2-amino-3-sulfhydrylpropanoic acid (cysteine, Cys), 2-amino-3-(prop-2-en-1-ylsulphanyl)propanoic acid (*S*-allylcysteine, Sac), **B**) pyrrolidine-2-carboxylic acid (proline, Pro), (2*S*,4*S*)-4-fluoropyrrolidine-2-carboxylic acid ((4*S*)-4-fluoroproline, S-Flp), (2*S*,4*R*)-4-fluoropyrrolidine-2-carboxylic acid ((4*R*)-4-fluoroproline, R-Flp), (2*S*)-4,4-difluoro-pyrrolidine-2-carboxylic acid (4,4-difluoroproline, Dfp), (2*S*)-2,5-dihydro-1H-pyrrole-2-carboxylic acid (3,4-dehydroproline, Dhp).

Proline (Pro, P) is a unique proteinogenic α-amino acid in biological systems, or in chemical terms, a secondary amine. The side chain exhibits a distinctive cyclic structure that confer Pro with a remarkable conformational rigidity in comparison to the other 19 cAA. In proteins, Pro residues are frequently found in turn and loop structures or at the end of an α-helix and β-sheet.[246,299] This is due to the lack of a hydrogen on the α-amino group that disables the function as a hydrogen bond donor for stabilization of the secondary sub-structures. Moreover, the planar amide bond to Pro residue possesses a partial double bond character, which can exist in both *cis* and *trans* conformations (**Figure 52A**).[300,301] Due to the less steric repulsion of the *trans* isomer's amide hydrogen to the preceding C^{α} atom, most peptide bond preferably adopts the *trans* conformation. However, the slow cis-trans isomerization is in equilibrium and plays an important role in the rate determining step of the protein backbone folding process.[302,303] The second characteristic conformational transition of Pro comprises the fast ring puckering process (**Figure 52B**).[304,305] For each of an *cis* and *trans* isomer there are two puckered states of the pyrrolidine ring that relieve the structure from a strained planar conformation. Depending on the substituent at C^{γ}, either exo or endo ring pucker is favored. For example, (4*S*)-4-fluoroproline (4S-Flp) stabilizes the *cis* isomer and strongly prefers the C^{γ}-endo-pucker, while (4*R*)-4-fluoroproline (4R-Flp) favors the *trans* isomer and the C^{γ}-exo-

pucker.[306] Thus, together with other factors Pro ring puckering states is of great importance in organization of the backbone and interpretation of the structure stability.[307,308]

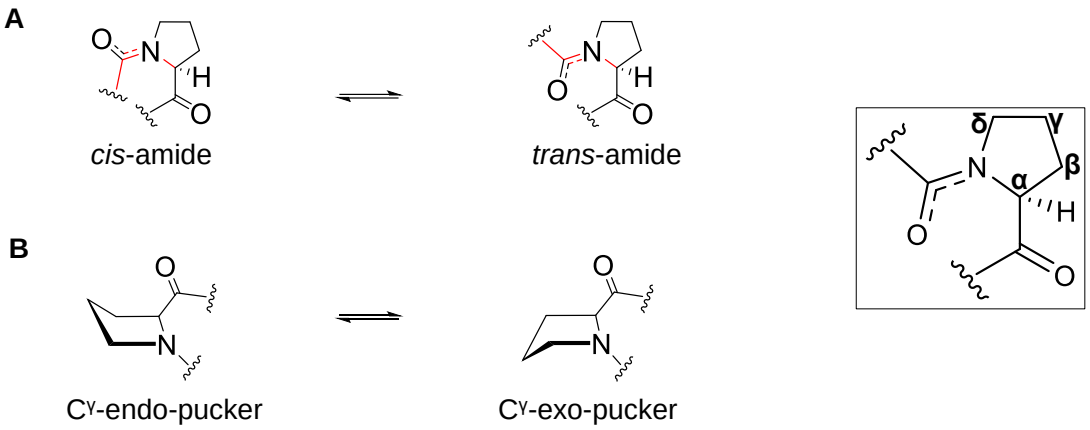


Figure 52 Characteristic conformational transitions of Pro residue. A) The structure formulas of two existing isomers of Pro in a peptide bond. Red lines indicate bonds that are considered for classification. **B**) The general structure of Pro ring puckering states include four conformations: *cis* isomer with C^{γ}-endo-pucker or *cis* isomer with C^{γ}-exo-pucker and *trans* isomer with C^{γ}-endo-pucker or *trans* isomer with C^{γ}-exo-pucker.

Over the last two decades, investigations of the Pro impact in the context of peptide and protein backbone have gained interest in many research groups. Especially, the usage of the isostructural analogues for replacement of Pro has become increasingly popular in modification of the biopolymer backbones. Here, we chose 4S-Flp, 4R-Flp, (4,4-difluoro-proline (Dfp), and 3,4-dehydroproline (Dhp) to manipulate the folding properties of resulting proteins upon the characteristic conformational transitions of substituents: 4S-Flp favors the C^{γ}-endo-pucker, 4R-Flp stabilizes the C^{γ}-exo-pucker, and Dfp has no ring pucker preference (**Figure 51B**).[306] Dhp possesses a flat ring structure, in which the ring pucker state is abolished due to double bond between C^{γ} and C^{δ} atom.[309]
Different fluorescent proteins are selected for distinct aims of the investigation. Experimental setups and results are described in the following two sections.

3.3.1 A protein scaffold for bioorthogonal reactions and selective labelling

In this part of the study, we have designed a reporter protein as a scaffold that is suitable for bioorthogonal reactions and selective labelling. For this purpose, a cysteine-free sfGFP (cfsfGFP) mutant carrying Sac in place of lysine (Lys, K) at position two was generated due to several reasons (**Figure 53**). First, K2 is an optimal basis for a scaffold that is deployed in

intermolecular reactions with respect to the solvent-exposed site. The introduced alken-tag at this position will minimally interfere with the protein innate functions, but allows chemical modification for further functionalization e.g., by photoinduced thiol-ene reaction. Second, sfGFP contains two Cys residues harboring the nucleophilic thiol group that are known to be easily oxidized irreversibly to e.g., sulfuric or sulfonic acid. Moreover, formation of disulfide bonds of a Cys residue in the β-barrel might abolish the tightly packed structure and therefore, lead to the loss of fluorescent. Since we have planned to perform bioorthogonal reactions in a redox environment, these residues would disturb the entire experiment. Therefore, both Cys residues were replaced by another cAA that was determined to retain most of the parental fluorescence.[310] C48S and C70M mutations are introduced into sfGFP(K2TAG) via site-directed mutagenesis resulting in the pET28a+ cfsfGFP(K2TAG)-H_6 plasmid for expression.

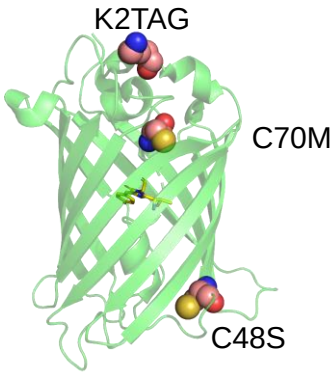


Figure 53 Selected positions for mutagenesis in sfGFP. sfGFP crystal structure illustrates the solvent-exposed K2 which was seleted for introduction of amber codon (TAG) for Sac incorporation via SCS. Both cysteine residues are proposed to replace by other cAA: C48S and C70M. The tyrosine-based chromophore is shown in yellow. Figure prepared based on PDB entry 2B3P using PyMOL.

Introduction of Sac into the TAG site is implemented by SCS using the evolved orthogonal pair derived from *Methanosarcina mazei* (*M. mazei*), $tRNA^{Sac}_{CUA}$:SacRS. The synthetase harbors the essential mutations in the *M. mazei* PylRS, C348W and W417S, that minimize the binding pocket of the enzyme and hence enable the translation of the small substrate, Sac.[139] The expression host *E. coli* BL21 AI was used for protein production. Simultaneously, we applied the genomically recoded *E. coli* C321.ΔA.exp (DE3), which was designed for enhanced suppression efficiency.[136] However, a test expression revealed a better production yield by BL21 AI compared to C321.ΔA.exp (DE3) (data not shown). It is not surprisingly that expression in a B strain results in quantitatively and qualitatively better outcomes than in a K12 derived strain usually used for cloning. Therefore, we proceeded with the BL21 AI host. Purification of expressed proteins by IMAC yielded approximately 7.2 mg per liter expression

culture. Analysis by SDS-PAGE and subsequently by LC-ESI MS verified the presence of target product (**Figure 54**). The incorporation of Sac is indicated by a positive mass shift of 40 Da at 27 753.6 Da that corresponds to the calculated value of 27 753.0 Da (**Figure 54B**). A small peak at 27 772.6 Da can be assigned to cfsfGFP(2Sac) with unmatured chromophore cfsfGFP(Sac)$_{pre}$. Longer incubation of the target protein in oxidative environment could decrease the undesired side product. Nonetheless, the purity of target sample is sufficiently for subsequent experiments.

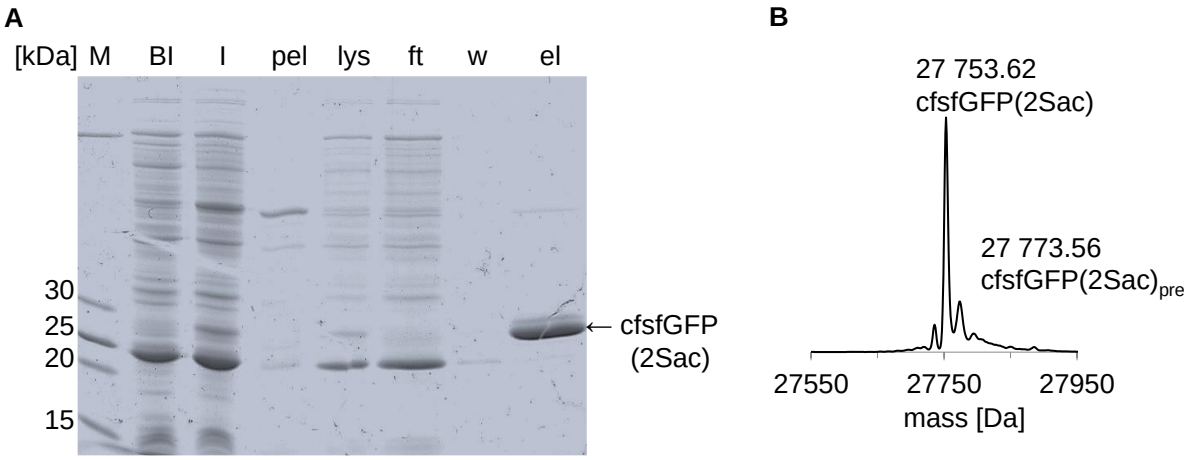


Figure 54 Analysis of expressed cfsfGFP(2Sac)-H$_6$ in C321.ΔA.exp (DE3) by SDS-PAGE (A) and by LC-ESI MS. A) Before (BI) and after induction (I) fractions from standard heterologous gene expression of cfsfGFP(2Sac) as well as insoluble (pel) and soluble (lys) samples after cell lysis are applied for expression analysis on 12 % SDS-PAGE gel. Overexpression could be observed at expected size of approximately 27.8 kDa (arrow). From IMAC purification, flowthrough (ft) and wash (w) fractions were collected and analyzed with eluted fraction (el). Enrichment of desired fragment could be obtained in high purity. Unstained marker (M) was used with appropriate fragment size (left lane). **B**) Deconvoluted mass spectrum of purified cfsfGFP variant shows the presence of cfsfGFP(2Sac) in the main peak, while the minor peak indicates the variant containing unmatured chromophore, cfsfGFP(2Sac)$_{pre}$. Calculated MW for matured cfsfGFP(2Sac): 27 753.0 Da and unmatured cfsfGFP(2Sac)$_{pre}$: 27 773.0 Da.

Next cfsfGFP(2Sac) was tested for its immobilizing capacities on novel thiol-functionalized hydrogels. The use of this polymer gains interests in encapsulation and delivery of active agents e.g., antimicrobial peptides, for biomedical applications.[311] 10 - 20 µg of characterized cfsfGFP(R2Sac) was used for coupling on thiolated 8PEG hydrogel (8PEG-SH) by thiol-ene conjugation (**Section 6.5.1**).[312–314] Initiation of the radical photochemical reaction was achieved by irradiation at λ = 365 nm in presence of photo-initiator for 10 min resulting in cfsfGFP(2Sac)-S-8PEG hydrogel. The success of the coupling reaction can be clearly verified by the presence of fluorescence of conjugated gel (**Figure 55A**). In comparison, no

fluorescence could be detected for gels mixed with cfsfGFP(2Sac) that was not capable to undergo the thiol-ene reaction due to the missing photoinduction. The successful conjugation of Sac-containing cfsfGFP variant on thiolated hydrogel represents a smart opportunity to visualize e.g., the drug delivery, *in vivo*.

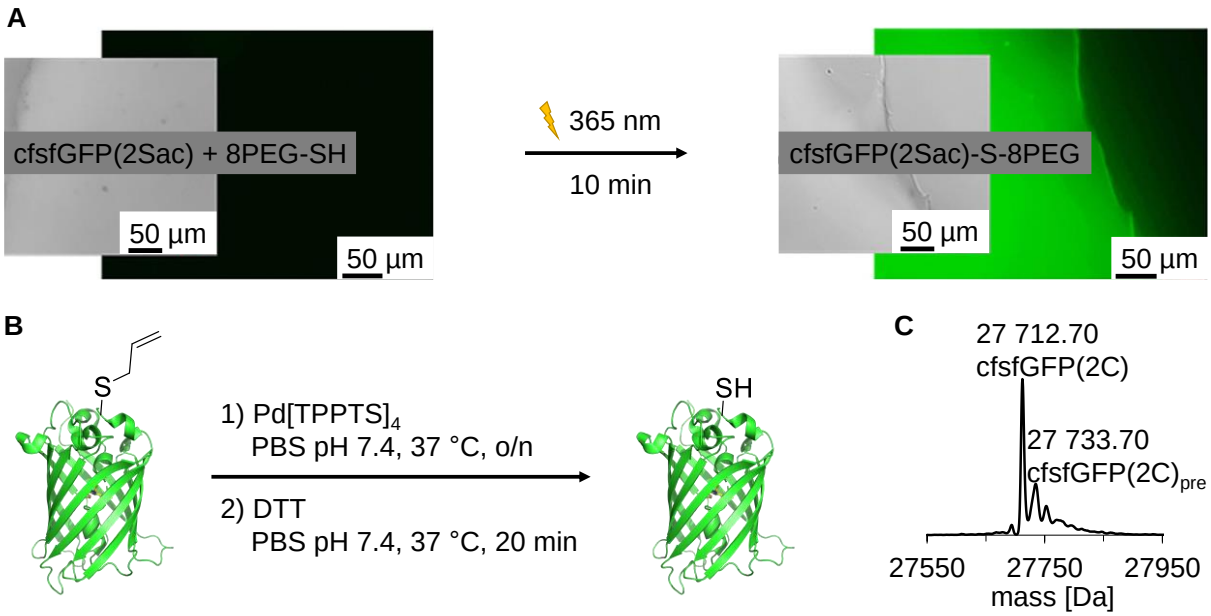


Figure 55 Application of cfsfGFP(2Sac) in bioorthogonal reaction (A) and chemical modification (B). A) Conjugation of cfsfGFP(2Sac) on 8PEG-SH gel was achieved by the photoinduced thiol-ene reaction. Hydrogel prior to the coupling remains dark, while immobilized cfsfGFP(2Sac)-S-8PEG gel shows bright fluorescence when analyzed by fluorescence microscopy. **B**) Deprotection of the Sac allyl group was performed in aqueous buffered solution at physiological pH, catalyzed by the $Pd[TPPTS]_4$ complex. **C**) LC-MS analysis of the sample after deprotection confirms the success of the reaction. Calculated MW for matured cfsfGFP(2C): 27 713.0 Da and unmatured $cfsfGFP(2C)_{pre}$: 27 733.0 Da.

We went further and proved the feasibility of the Sac allyl moiety for protection of Cys side chain of a protein. In collaboration with Sergej Schwagerus (FMP Berlin, AG Hackenberger), deprotection was performed in physiological environment using an $Pd[TPPTS]_4$ complex (**Figure 55B, Section 6.5.2**). LC-MS analysis verified the correct mass shift of -40 Da after the allyl cleavage, yielding a sfGFP mutant containing a Cys residue at the second position (27 712.7 Da, **Figure 55C**). Again, a small amount of the mutant harboring unmatured chromophore is detected as it was already present in the sample prior to deprotection.
Nevertheless, we could show that cfsfGFP(2Sac) is a robust designed protein scaffold capable to applied for controlled chemical modifications as it remains intact after photoinduced thiol-ene conjugation and Pd-catalyzed deprotection.

3.3.2 A proline-based design of fluorescent proteins

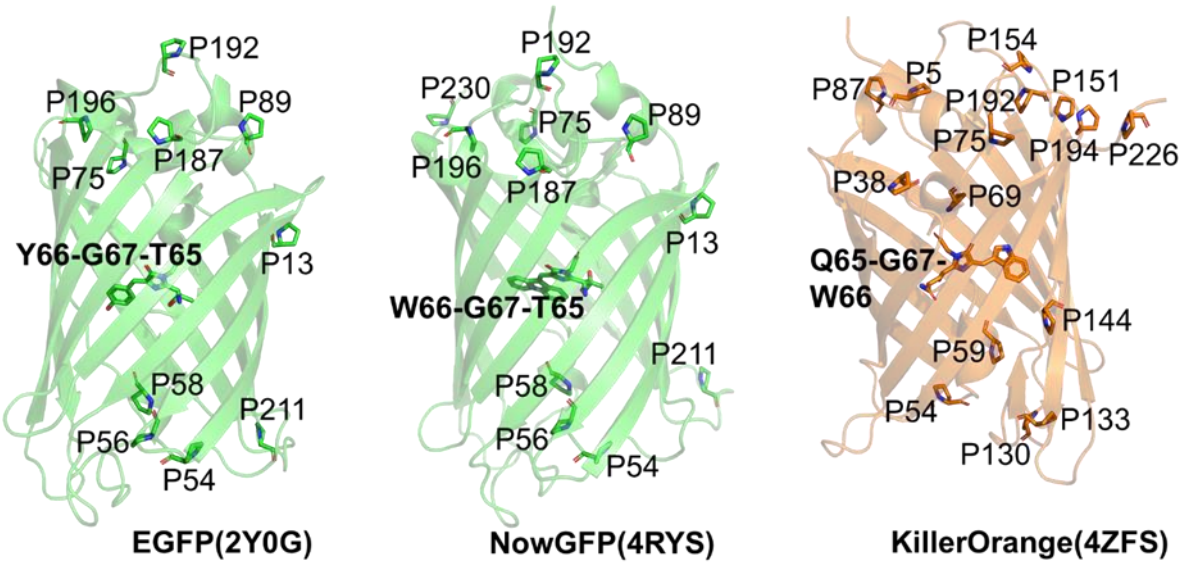


Figure 56 Fluorescent proteins of interest. The characteristic β-barrel structure of three different variants of fluorescent proteins are presented: EGFP, NowGFP, KillerOrange. Ribbon colors represent the color of fluorescence emission of each variant. Proline residues (one-letter code) are highlighted as sticks with the annotation of corresponding positions. Chromophores are shown with the initial tripeptide composition in bold. All structures were generated by PyMOL using appropriate PDB entries (in brackets).

As mentioned above, we aimed to use built-in chemical modifications in context of fine-tuning the advantageous properties of fluorescent proteins. This is achieved by substitution of cAA bearing essential structural and functional roles in a biopolymer with their similar ncAA counterparts. Pro is shown to have unique character, that is often indispensable for protein structures. However, Budisa and coworkers demonstrates that by design of EGFP using fluorinated Pro analogues folding and stability of the enzyme could be improved in comparison to the wildtype.[193] Inspired by these outcomes, we have transmitted the idea of "molecular surgery" to engineer novel fluorescent proteins towards enhanced properties. For this purpose, the designated set of Pro analogues (**Figure 51B**), described in **Section 3.3**, were applied to modify the structure of NowGFP and KillerOrange via SPI (**Section 1.3.1**). NowGFP is a variant of *A. victoria* GFP obtained by multiple mutagenesis of preceding fluorescent proteins.[190,315] It exhibits an anionic tryptophan-based chromophore and a long fluorescence lifetime with a high quantum yield. As a descendant of EGFP, NowGFP also contains the 10 conserved Pro residues like EGFP and an additional eleventh Pro at position 230 (**Figure 56**). Both GFP constructs are often used in cell biology and biophysical applications.[316–318] In contrast, KillerOrange is a chromoprotein anm2CP derivative found in the hydrozoan genus

Anthoathecata.[319,320] It possesses an equally tryptophan-based chromophore and 15 proline residues (**Figure 56**). As the name indicates, KillerOrange is an orange-emitting protein photosensitizer that generates reactive oxygen species (ROS) upon illumination by certain wavelength. Genetically encoded photosensitizers have been developed for light-induced production of ROS at desired locations *in vitro* and *in vivo*.[321] This promising optogenetic tool enables a controlled elimination of specific cell populations, target protein inactivation or DNA damage.[322–325] As a positive control, selected Pro analogues were also incorporated in EGFP for comparison.

All Pro residues of the model proteins were replaced by the aforementioned set of Pro analogues via SPI (**Section 6.2.5.2**). Proline-auxotrophic E. coli K12 strain JM83 was used as an expression host.[326] Fluorescent proteins were cloned with an H_6-tag for subsequent purification by IMAC. Cell pellets of the wildtype protein and variants bearing Pro-to-S-Flp/Dhp substitution showed the typical bright color of the particular autofluorescent intact chromophore, while variants with R-Flp and Dfp remained colorless. The loss of fluorescence is rather due to the lack of chromophore formation or misfolding that ends up as unfolded protein in inclusion bodies (**Figure 57A**). The presence of protein containing R-Flp (**Figure 57B-D**) and Dfp (data not shown) were confirmed by SDS-PAGE analysis. In case of R-Flp, the C^{γ}-exo-pucker conformation is favored, which is not preferred by the wildtype structures and in turn leads to protein misfolding.[55] Introduction of Dfp, which exhibits no apparent ring puckering preference, produced a dysfunctional protein due to the diminished velocity of the cis-trans isomerization at the Pro residue.[306] It should be noted that extraction of proteins from the inclusion bodies is possible via *in vitro* refolding techniques.[327] However, corresponding procedures are time consuming and often lead to poor protein yield. In contrast, native proteins, S-Flp and Dhp bearing variants are mainly in the soluble fractions (**Figure 57B-D**). As a C^{γ}-endo-pucker, S-Flp improved the stability of the barrel structure due to the increased total C^{γ}-endo ring puckering proportion of the entire biomolecule (7/10 Pro residues of the wildtype protein possesses C^{γ}-endo-pucker, after Pro-to-S-Flp replacement: 9/10). In the scope of this work, only soluble proteins were isolated by IMAC, resulted in 20-30 mg/L expression culture for EGFP, 60-80 mg/L for NowGFP and KillerOrange. The yields of wild type and mutated proteins were comparably similar. Subsequent analysis of enriched samples by LC-ESI MS proved the presence of desired isolates with high purity (**Figure 58**). Each Pro substitution by S-Flp resulted in a mass increase of 18 Da, while Dhp replacement produced a negative mass shift of 2 Da. Interestingly, detected mass of NowGFP is in total about 15 Da higher than the calculated value, which probably due to the oxidation of a Met residue (+16 Da, **Figure 58B**). In comparison, both Pro replacements by S-Flp and Dhp provided target proteins having the expected mass. Hence, incorporation of Pro analogues somehow prevents the unwanted side reaction.

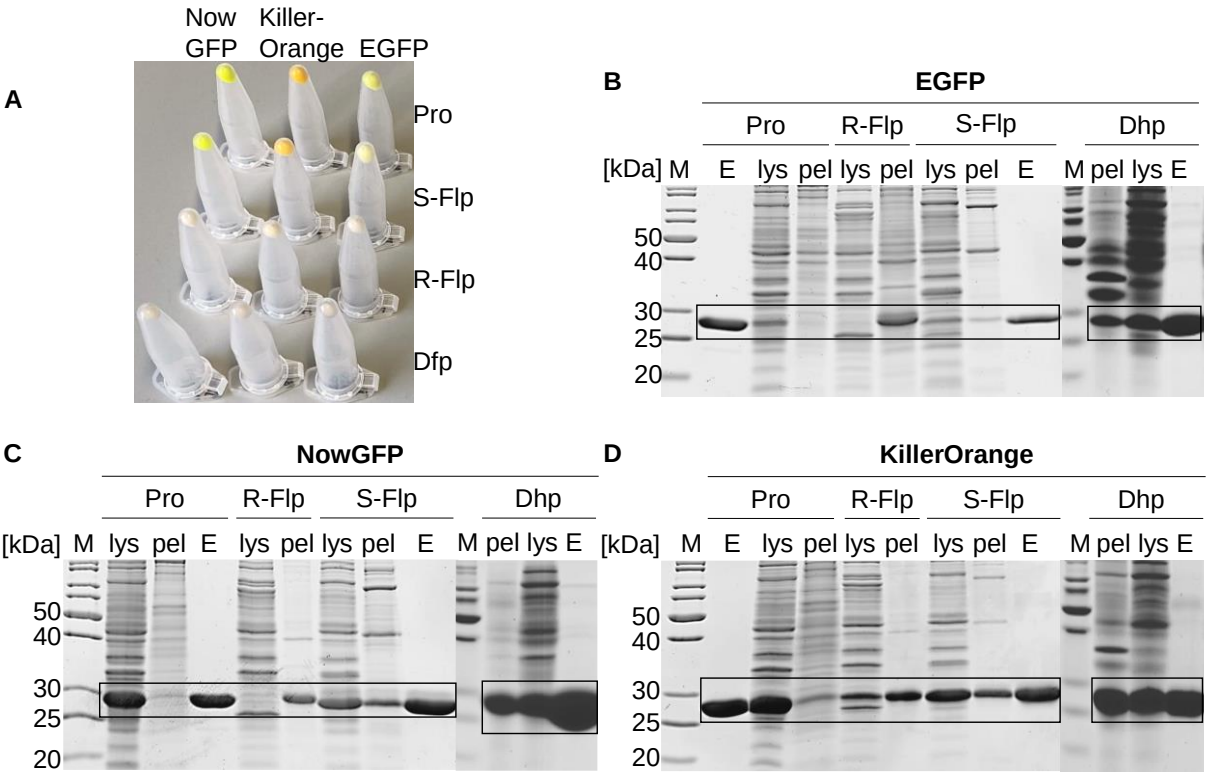


Figure 57 Analysis of fluorescent protein expression in *E.* coli JM83 by SDS-PAGE. A) 1 mL of cell pellets from expression culture having normalized OD_{600} of 2. Soluble (lys) and insoluble fractions (pel) of **B**) EGFP, **C**) NowGFP and **D**) KillerOrange variants were loaded on 15 % acrylamide gel to compare with the eluted fractions (E) from IMAC of soluble proteins. Expected regions of particular protein are highlighted (black frame). In fluorescent protein variants, all Pro residues were substituted by R-Flp, S-Flp and Dhp. The same marker (M) was used for the entire analysis.

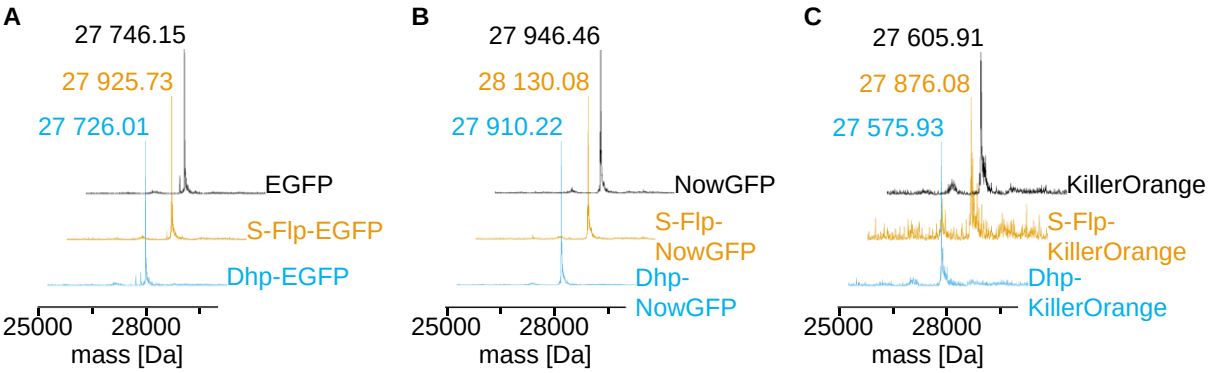


Figure 58 LC-ESI MS spectra of purified EGFP (A), NowGFP (B) and Killer-Orange variants (C). The main mass peak of deconvoluted LC-ESI MS spectra are labeled. All wildtype proteins are shown in black, S-Flp containing variants in orange and Dhp bearing contructs in cyan. **A**) $M_{W, calc.}$ (EGFP) = 27 745.33 Da, $M_{W, calc.}$ (S-Flp-EGFP) = 27 925.33 Da, $M_{W, calc.}$ (Dhp-EGFP) = 27 725.33 Da. **B**) $M_{W, calc.}$ (NowGFP) = 27 931.50 Da, $M_{W, calc.}$ (S-Flp-NowGFP) = 28 129.50 Da, $M_{W, calc.}$ (Dhp- NowGFP) = 27 909.50 Da. $M_{W, calc.}$ (KillerOrange) = 27 606.09 Da, $M_{W, calc.}$ (S-Flp- KillerOrange) = 27 876.09 Da, $M_{W, calc.}$ (Dhp- KillerOrange) = 27 576.09 Da.

To further characterize the changes in the context of ncAA incorporation induced spectroscopic properties, light absorbance and emission measurements were implemented (**Figure 59**). A maximum at 280 nm is observed in the UV-Vis absorption spectra, which is typical for aromatic residues Tyr and Trp. The chromophore of the different fluorescent protein derivates absorbed at distinct wavelengths. For EGFP variants, the maximum was found at 488 nm, while NowGFP constructs possessed the strongest absorbance at 493 nm (**Figure 59A,B**). Both variants show higher absorbance in presence of S-Flp. This is due to the effect of fluorination that probably enhances the Trp/Tyr absorption band. There are many Pro residues in the proximity of the aromatic residues, both in the primary sequence and in the tertiary structure. Especially at the proline-rich pentapeptide P54-V55-P56-W57-P58 (PVPWP motif), which is conserved in EGFP and NowGFP, fluorination of these buried residues is known to increase their hydrophobicity and thus stabilize the folded protein.[193] However, higher absorption values can be caused by an increased amount of improperly folded protein or immature chromophore. Since the concentration determination was performed by quantifying the absorbance characteristics of both, matured and immature chromophores, higher absorption at 280 nm in relation to the chromophores absorption wavelength is probably resulted from the normalization of the spectra (**Figure 59A,B**). This assumption is approved by a significant decreased ratio of chromophore to the tryptophan- and tyrosine-specific absorbance (ε_M(CRO)/ε_M(Tyr+Trp) = 0.96) compared to a higher ratio (1.57) of the wildtype protein S-Flp-bearing EGFP (**Table 5**). For KillerOrange, the two separate maxima were detected according to the two potential configurations and charged states of its chromophore (**Figure 59C**). The maximum at 455 nm is likely corresponded to an immature CFP-like chromophore, while the peak around 510 nm is associated with a mature mHoney-Dew-like tryptophan-based chromophore.[328] The characteristic double-peak signature for neutral tryptophan-based chromophores can be observed for all KillerOrange variants. Pro-to-S-Flp substitution of this photosensitizer positively influenced the contribution of the different chromophore states. The maturation reaction is promoted by the introduction of fluorine in the protein structure. In contrast, incorporation of Dhp had no impact on the absorbance spectra of EGFP or NowGFP but induced a slight improvement of the chromophore maturation of KillerOrange.

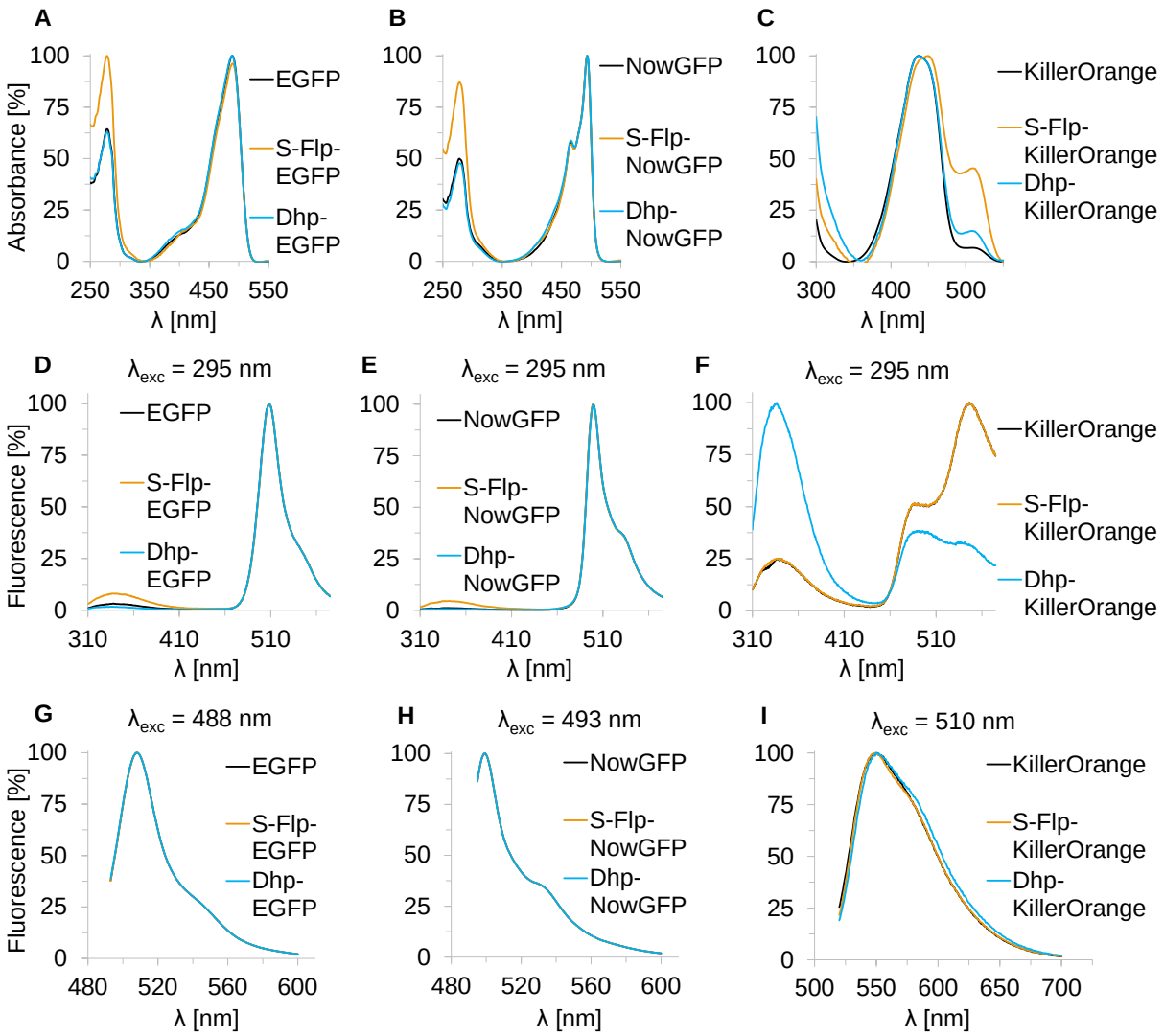


Figure 59 Absorption and fluorescence emission analysis of fluorescent protein variants.

Table 5 Molar extinction coefficients (ε_M) of EGFP variants at appropriate wavelengths. The calculation of ε_M (in $M^{-1} \cdot cm^{-1}$) values is based on the recorded UV-Vis absorption spectra of selected EGFP variants with known protein concentrations. Respective wavelength of 488 nm correlates to the chromophore-specific absorbance wavelength, whereas 280 nm indicates the maximum absorbance, resulted from aromatic residues, Tyr and Trp.

λ[nm]	ε_M[M^{-1} cm^{-1}] (EGFP)	ε_M[M^{-1} cm^{-1}] (S-Flp-EGFP)	ε_M[M^{-1} cm^{-1}] (Dhp-EGFP)
488 (≡ CRO)	31657 (± 1341)	22950 (± 290)	27800 (± 542)
280 (≡ Tyr+Trp)	20116 (± 172)	23800 (± 715)	17300 (± 554)

In addition, we have analyzed the fluorescence emission spectra of selected variants using corresponding chromophore-associated absorbance wavelength of appropriate protein variant for excitation (**Figure 59G-I**). These spectra remained unchanged meaning the Pro analogue incorporation had no influence in the physical and chemical vicinity of the chromophore. Similar results could be observed for the fluorescence emission intensities of EGFP and NowGFP variants, when excited at the tryptophan-characteristic absorbance maximum of 295 nm. In contrast, KillerOrange showed remarkable differences upon excitation at 295 nm. This experiment monitors the fluorescence resonance energy transfer (FRET) or direct excitonic coupling reaction between the Trp aromatic side chain and the mature chromophore, having a distance shorter than 25 Å. The strong chromophore-based fluorescence intensity of EGFP and NowGFP variants are predominant towards their weak tryptophan-specific emission, when excited at 295 nm (**Figure 59D,E**). However, the S-Flp-containing variants possessed a slightly larger tryptophan-based emission, which arose from a countless participation of the misfolded protein containing Trp and an immature chromophore. For KillerOrange, the tryptophan-based fluorescence intensity was significantly increased. This is likely due to the absence of fluorescence quenching, which is induced by the mechanism of excitation energy transfer or exciton coupling. The wildtype protein and the S-Flp-bearing variant showed similar Trp emission profile, while Pro-to-Dhp substitution triggered a drastic decrease in chromophore emission, which is probably caused by minor structural effects (**Figure 59F**).

For comparison of folding properties, more precisely the formation of the chromophore, unfolding/renaturation experiments were implemented. Therefore, fluorescence emission intensity measurements were recorded for proteins in their native state, immediately after a chemical denaturation, and after 24 h of the refolding process (**Section 6.4.6**). The refolding efficiency was estimated by the ratio of native proteins to the corresponding renatured samples. Purified proteins were denatured by boiling in 8 M urea and subsequently refolded at room temperature by 100-fold dilution into the same buffer without urea, but 5 mM DTT. Over 30 min refolding kinetics were monitored by fluorescence at 295 nm, and 488 nm, respectively. EGFP variants showed fast refolding velocity at both wavelength (**Figure 60**). The recovery of the chromophore fluorescence intensity is two time slower than the recovery of tryptophan-specific emission (**Figure 60**). The respective period of time needed for emission recovery was 1500 s in case of the chromophore (**Figure 60B**) and 750 s for the tryptophan-based emission (**Figure 60A**). Pro-to-Dhp substitution in turn did not influence the protein refolding kinetic. In contrast, no refolding could be observed for NowGFP and KillerOrange variants after denaturation, leading to the absence of chromophore fluorescence signal (data not shown). EGFP differs from the parent GFP in three positions, compared to NowGFP with 27 mutations in total. In fact, the vast of mutations may destroy the ability of self-assembly and

chromophore formation after denaturation. KillerOrange is a photosensitizer, derived from the hydrozoan chromoprotein variant KillerRed.[321] Although it possess the robust β-barrel structure of the homolog GFP, the refolding property lacks completely.

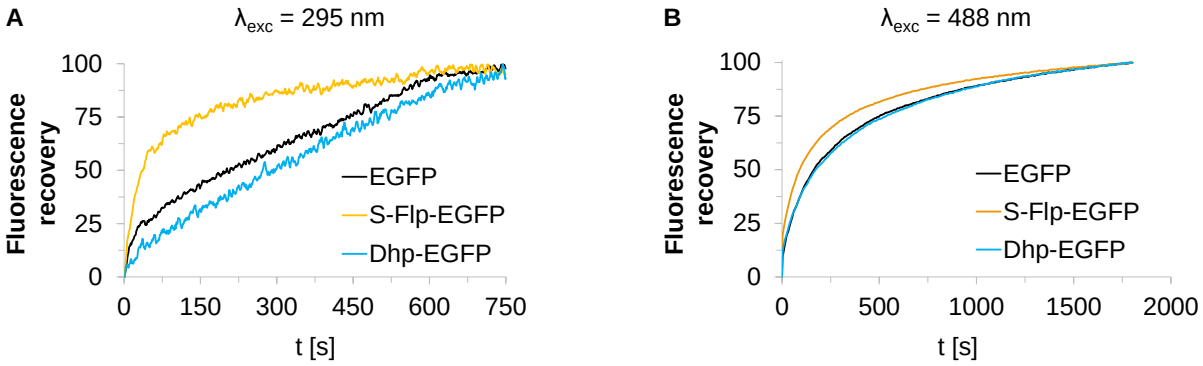


Figure 60 The protein refolding and chromophore maturation process of EGFP variants, monitored by fluorescence emission upon excitation at 295 nm (A), and 488 nm (B). **A**) The tryptophan-specific fluorescence emission at 330 nm was tracked upon excitation with ultraviolet light ($\lambda_{exc.}$ = 295 nm). (**B**) Monitoring of the fluorescence amplitude development at the chromophore emission 509 nm upon excitation with green light ($\lambda_{exc.}$ = 488 nm). The time-dependent fluorescence recovery were normalized to unity (100%) according to the amplitude at the endpoint of the monitoring timescale. Black curves represent the spectra of the wildtype protein, while S-Flp-containing variant is shown in orange curves, and Dhp-bearing proteins in blue curves, respectively.

Excited at the tryptophan-specific wavelength, all of the EGFP variants showed partial fluorescence emission recovery (**Figure 61A-C**). A slightly better yield of refolded protein could be obtained with S-Flp-EGFP. Surprisingly, EGFP wildtype and Dhp containing variant possess larger tryptophan-specific fluorescence, but red-shifted, indicates that the structural environment of the single W57 residue have changed after reassembly (**Figure 61A,C**). It should be noted that monitoring of the denaturation-refolding process of one protein by tryptophan-specific emission represents the emission behavior of immediate vicinity of Trp. Denaturation will not lead to total loss of the Trp-fluorescence, but only a shift of the emission maximum. Therefore, quantitative evaluation of recovered protein amount is not possible since each environment reveals a different emission pattern and intensity. Upon excitation by 488 nm, the recovery of chromophore fluorescence was monitored. S-Flp-EGFP recovered more than 91 % of its initial fluorescence, whereas EGFP and Dhp-EGFP only reached 43 %, and 53 %, respectively (**Figure 61D-F**). These results are in agreement with literature data.[193]

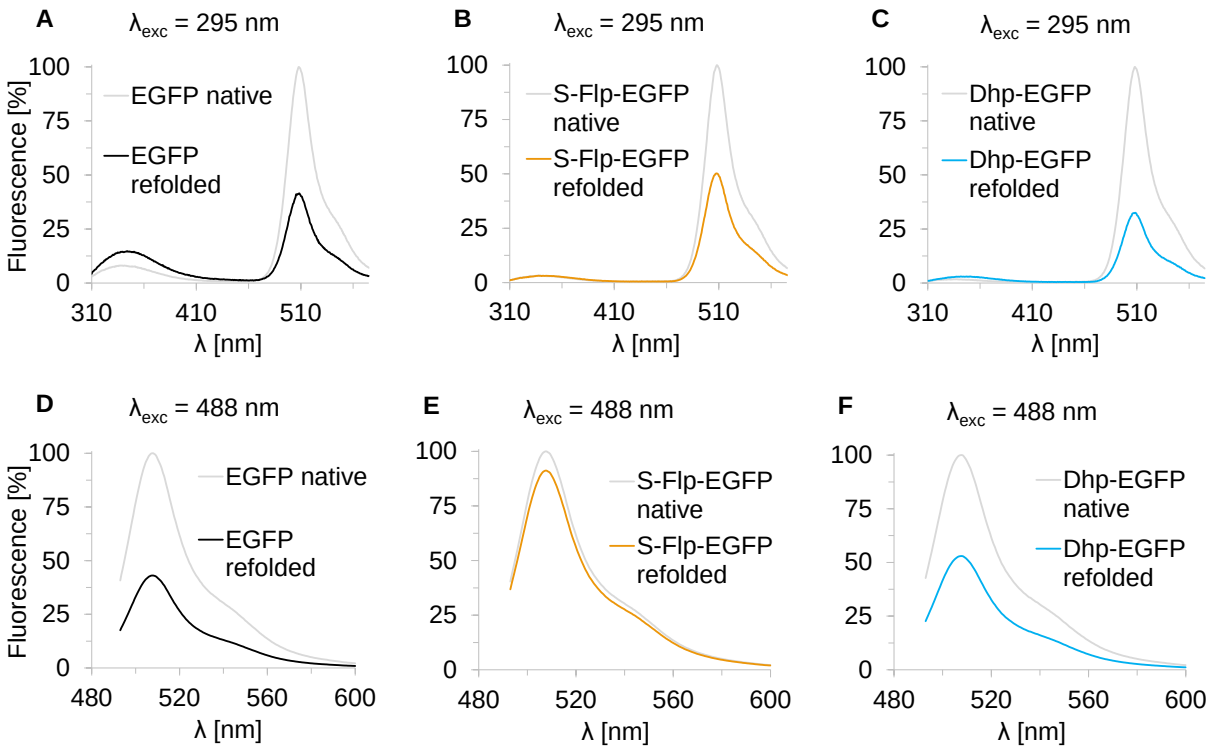


Figure 61 Fluorescence emission spectra of refolded EGFP variants upon excitation at 295 nm (A-C), and 488 nm (D-F). Normalized fluorescence emission spectra of 0.3 µM solutions of fluorescent protein variants in the native state (grey curves) and after denaturation and subsequent refolding. **A**, **D**) EGFP; **B**, **E**) S-Flp-EGFP; **C**, **F**) Dhp-EGFP. Refolded samples of EGFP correspond to black curve, S-Flp-EGFP to orange and Dhp-EGFP to cyan, respectively. The fluorescence intensity of each refolded protein variant are normalized to the maximum fluorescence of the appropriate native state.

Taken together, by intensive investigation we were able to determine the diverse impact, introduced by different Pro analogues. We have successfully showed that subtle changes of C^{γ} position at the level of single atoms and even by the substitution with an isomer is sufficient enough e.g., to abolish the protein structure. Pro replacement to R-Flp and Dfp of all fluorescent variant decreased the protein solubility massive. An improvement in refolding velocity could be obtained by Pro-to-S-Flp replacement in EGFP, making this advanced design more advantageous than substitution with cAA, that might end up in a dysfunctional protein. Despite our expectation, refolding experiment could not be performed for NowGFP and KillerOrange variants due to their total loss of structure recovery ability.

4. Conclusion and outlook

In a major part of this study, protein engineering was used for a design process to produce peptide and protein structures containing noncognate fluorinated substrates in *E. coli*.
In earlier research by a previous colleague, ribosomal translation of the fluorinated AA TfeGly in response to Ile codon was achieved by using an IleRS mutant (IleRS(Ala10)) that had a defect in post-transfer editing.[222] However, Val, Met, and Leu were also found to be canonical substitutions for Ile. Therefore, the selectivity of the synthetase needs to be improved. Our aim was to design a directed evolution strategy, using the wildtype IleRS enzyme of *E. coli*.
In this context, crystallization attempts of IleRS were conducted after successful synthesis of the necessary substrates. An initial screening with the full-length *E. coli* synthetase and ligands (TfeGly, Abu-2AA, AMP-PMP) yielded promising starting points, which, however, could not be optimized for the production of diffraction-quality protein crystals. We found that the crystallization process is disturbed by the N-terminal H_6-tag that definitely has to be removed prior to crystallization experiments. A side benefit of the tag-cleavage is the removal of the start-methionine, which is prone to be oxidized after the isolation and purification from the cell lysate. Also, the shorter CP1 domain could not be crystallized. It should be noted that the H_6-tag was not removed from the editing domain after purification, which needs to be executed in the future experiments. Alternatively, IleRS enzymes from homologous organisms, such as *T. thermophilus* and *S. aureus*, which have already been successfully crystallized, can be used in complex with TfeGly and the non-hydrolyzable charged tRNA-analogue, TfeGly-2AA, for crystallization. It is very likely that the crystals will be obtained with such an established set-up.
Since the structural elucidation of *E. coli* IleRS failed, we changed our strategy and constructed IleRS libraries based on the active site environments of homologous *T. thermophilus* IleRS. For this purpose, residues within 5 Å of the synthetic and the editing pockets facing the substrate side chain were selected to construct three IleRS of each catalytic center. Although the expected library size could not be perfectly achieved, the number of IleRS mutants is sufficient for selection.
In the next step, sfGFP reporter protein harboring reduced number of Ile positions were constructed by converting the common Ile codons AUC and AUU to the rare AUA codon by a silent mutation, resulting in sfGFP(6xAUA) and sfGFP(4xAUA). We could show that the maintenance of the fluorescent ability of sfGFP dependents on the combination of the Ile-substitutions. The two full-length reporter protein variants that exhibit sufficient fluorescence intensity can be used to study the impact of rare codon number on expression efficiency. Alternatively, a split sfGFP system was cloned to compensate for the disadvantages of the first reporter construct. It is not clear whether the replacement of Ile with TfeGly in

sfGFP(6xAUA) and sfGFP(4xAUA) would lead to complete loss of fluorescence. In the split concept, sfGFP1-10 fragment is expressed in presence of Ile and sfGFP11 fragment containing one AUA codon in presence of TfeGly. Self-assembly of these fragments would yield sfGFP that has a substitution from Ile to TfeGly. The likelihood is very high that replacement of the single Ile229 residue with a fAA with similar size and physicochemical properties (e.g., TfeGly) to Ile would maintain the emission behavior of the protein. This split system was generated as an alternative reporter system and was not further investigated in this study.

In the later phase of the study, we showed that the screening plasmid containing the necessary regulators for expression was fully assembled. The final test of the screening system in selected *E. coli* displayed the functionality of the system *in vivo*. Antibiotic resistance affected the expression of the fluorescent protein, which is consistent with data from the literature.[262] Moreover, our sfGFP mutants bearing rare codons provided better fluorescence in our screening system than the sfGFP construct having abundant codons for Ile decoding. This result indicates that our system is suitable for AUA codon reassignment.

Despite all attempts to generate a selection strain, we were not able to introduce relevant mutations (A579P, W905R) into *ileS* by different genome editing approaches. We assumed that the mutation in this gene, which is essential for survival, would be repaired immediately afterwards, resulting in a wildtype enzyme. This issue can be solved by a different strategy. First, the cells are supplied with a copy of *ileS* containing the desired mutations on a plasmid. The endogenous copy of *ileS* is then removed from the genome, resulting in an *ileS*-knockout strain. Cell survival is explicitly enabled by the supplied IleRS(A579P, W505R) mutant.

The effects of ncAA incorporation into reporter proteins were analyzed after the “atomic mutations” were introduced into selected proteins. We have shown that cfsfGFP(2Sac) is suitable for chemical modification by thiol-ene coupling and can act as a cleavable protecting group for cysteine in the redox-environment. This background can be used for further applications, e.g., drug delivery, by using the alkene group as a tag for other coupling.

For EGFP, NowGFP and KillerOrange, the proline-based study showed that subtle changes in the side chain configuration of Pro were sufficient to maintain (Pro-to-S-Flp/Dhp substitution) or disturb (Pro-to-R-Flp/Dfp) the formation of the β-barrel structure. R-Flp showed an unfavorable C^{γ}-exo conformation in the original protein structures, while *trans*-to-*cis* isomerization of Dfp is the rate-limiting step in protein folding. Both effects probably triggered the protein misfolding in all tested fluorescent protein derivates. In turn, the replacement of Pro by S-Flp/Dhp promoted protein maturation. Moreover, S-Flp exhibited faster recovery of fluorescence after denaturation of the EGFP variant, which was not observed for NowGFP and KillerOrange variants. However, we have shown that ‘molecular surgery’ could be easily applied to several β-barrel constructs as a useful tool of protein engineering.

5. Materials

5.1 Technical equipment

Balances	
Mettler PE 3600 Deltarange	Mettler Toledo (Gießen, Germany)
TE 1502S	Sartorius (Göttingen, Germany)
Centrifuges	
Avanti J-26 XP	Beckman Coulter (Krefeld, Germany)
Centrifuge 5418 R	Eppendorf AG (Hamburg, Germany)
Centrifuge 5810 R	Eppendorf AG (Hamburg, Germany)
MiniSpin	Eppendorf AG (Hamburg, Germany)
Cooling centrifuge	Eppendorf AG (Hamburg, Germany)
Table centrifuge	VWR International GmbH (Darmstadt, Germany)
Gel electrophoresis	
Electrophoresis unit	Hoefer Scientific Instruments (Holliston, MA, USA)
Horizontal agarose gel system	Factory of Max-Planck Institute for Biochemistry (Martinsried, Germany)
Vertical SDS-gel system	Factory of Max-Planck Institute for Biochemistry (Martinsried, Germany)
Incubators, mixers and shakers	
Ecotron	Infors HT (Einsbach, Germany)
Incubator series B, KB	Binder (Tuttlingen, Germany)
Multitron	Infors HT (Einsbach, Germany)
Liquid Chromatography	
Äkta pure	GE Healthcare Life Sciences (München, Germany)
Äkta purifier	GE Healthcare Life Sciences (München, Germany)
Peristaltic pump P1	Pharmacia Biotech (now: GE Healthcare Life Sciences, München, Germany)
Mass spectrometry	
Agilent 6530 Accurate-Mass Q-TOF	Agilent (Santa Clara, CA, USA)

Spectroscopy

BioPhotometer plus	Eppendorf AG (Hamburg, Germany)
CD spectrometer J-815	Jasco Deutschland (Groß-Umstadt, Germany)
Fluorescence spectrometer LS 55	Perkin Elmer (Rodgau, Germany)
Microplate reader Infinite M200	Perkin Elmer (Rodgau, Germany)
Ultrospec 6300 pro	Amersham Biosciences (now: GE Healthcare Lifes Science, München, Germany)

Thermocyclers

Mastercycler Gradient	Eppendorf AG (Hamburg, Germany)
Peqstar 2x Gradient	Peqlab (Erlangen, Germany)

Thermomixer

Mixing Block MB-102	Bioer Technology (Binjiang, China)
Thermomixer 5437	Eppendorf AG (Hamburg, Germany)
Thermomixer compact	Eppendorf AG (Hamburg, Germany)

Miscellaneous

Gel-documentation system Felix 2050	Biostep (Jahnsdorf, Germany)
Ice machine Scotsman AF 80	Scotsman (Vernon Hills, IL, USA)
IKA Combimag RET	IKA (Staufen, Germany)
MicroPulser™	Bio-Rad Laboratories (München)
Microwave KOR-6305	Daewoo (Butzbach, Germany)
Microwave	SHARP (Osaka, Japan)
Orbital shaker Rotamax 120	Heidolph (Schwabach, Germany)
pH-Meter S20-SevenEasy™	Mettler Toledo (Gießen, Germany)
Potentiostat Autolab PGSTAT12	Metrohm Autolab B.V. (Utrecht, NL)
Power supply Consort EV261 and E143	Sigma-Aldrich (now: Merck KGAA, Darmstadt, Germany)
Power supply Power Pack P25 T	
Scanner ViewPix 700	Biometra (Jena, Germany)
Scintillation counter Wallac 1409	Biostep (Jahnsdorf, Germany)
Sonopuls HD 3200	Wallac (Freiburg, Germany)
Sonotrodes MS72, KE76	Bandelin (Berlin, Germany)
Vortex Genie™	Bandelin (Berlin, Germany)
Water bath VWB 12	Bender & Hobein AG (Zürich, Switzerland)
	VWR International GmbH (Darmstadt, Germany)

5.2 Extendable materials

Centrifugal concentrators	
Amicon® Ultra 0.5 mL centrifugal filters (3 kDa, 10 kDa, 30 kDa)	Merck KGAA, Darmstadt, Germany
Centrifugal devices, Microsep™ Advance (3 kDa, 10 kDa, 30 kDa)	Pall GmbH (Dreieich, Germany)
Centrifugal devices, Macrosep™ Advance (3 kDa, 10 kDa, 30 kDa)	Pall GmbH (Dreieich, Germany)
Columns:	
HiLoad™ 16/600 Superdex™ 200 pg	GE Healthcare Life Sciences (München, Germany)
HisTrap™ HP	GE Healthcare Life Sciences (München, Germany)
HiTrap™ Q Sepharose	GE Healthcare Life Sciences (München, Germany)
HiTrap™ Phenyl FF (Low Sub)	GE Healthcare Life Sciences (München, Germany)
PD-10 Desalting Columns	GE Healthcare Life Sciences (München, Germany)
Dialysis membranes:	
Spectra/por molecularporous membrane (3 kDa, 10 kDa, 30 kDa)	Spectrum Laboratories, Rancho Dominguez CA, USA
Electroporation cuvettes:	
Electroporation cuvettes (1 mm)	VWR International GmbH (Darmstadt, Germany)
Filter paper:	
Whatman 3MM	Whatman International Ltd, Maidstone Kent, UK
Sterile filters:	
Steritop™ 500 mL (ø 0.22 µm)	Merck KGAA, Darmstadt, Germany
Syringe filter ROTILABO® (ø 0.22 µm, 0.45 µm)	Carl Roth (Karlsruhe, Germany)
Crystallization accessories:	
MRC 2 lens crystallization plate	MRC Laboratory of Molecular Biology, Cambridge, UK

MRC MAXI 48 well plate	MRC Laboratory of Molecular Biology, Cambridge, UK
ComboPlate™ 24-well protein crystallization plate	Greiner Bio-One International, Frickenhausen, Germany

5.3 Chemicals

All standard chemicals were purchased from Alfa Aesar by Thermo Fisher Scientific (Waltham, USA), Carl Roth GmbH (Karlsruhe, Germany), Sigma Aldrich (now: Merck KGAA (Darmstadt, Germany), or VWR International GmbH (Darmstadt, Germany) unless otherwise specified. Following fAA and crystallization ligand were obtained from the members of the Koksch group: (2*S*)-4,4,4-trifluoroethylglycine (TfeGly, Dr. Johann Moschner), and 2'-((2*S*)-2-aminobutanoic acid) amino-2'-deoxyadenosine (Abu-2AA, Dr. Johann Moschner and Suvrat Chowdhary).

5.4 Media and supplements

All components of media and supplements were purchased from Carl Roth GmbH (Karlsruhe, Germany) or Sigma-Aldrich (now: Merck KGAA, Darmstadt, Germany).
For bacterial growth, fermentation, and protein expression under non-limiting conditions Lysogeny Broth (LB), Double Yeast Tryptone broth (DYT), and Terrific Broth (TB) medium were used, respectively. The liquid medium was autoclaved for 20 min at 121 °C and 1.5 bars. For agar plates, 1.5 % agar was added to the LB medium prior to sterilization. For selective pressure incorporation of ncAA, *E. coli* cells were grown in New Minimal Medium (NMM).[329] Stock solutions of salts for NMM were autoclaved, and the remaining sterile filtered, respectively. Auto-induction medium ZYP-5052 was used for expression of wildtype protein as well as for Stop Codon Suppression (SCS) experiments.[330] Stock solutions for ZYP-5052 were sterilized by filtration (ø 0.22 µm), except ZY-stock, which was autoclaved. Super Optimal Broth (SOB) medium and Super Optimal broth with Catabolite repression (SOC) medium was used for recovery of chemical- or electrocompetent cells after transformation.[331]
All media used in this work are listed in **Table 6.**

Table 6 List of media and their compositions.

Name	Composition per liter
LB	10 g tryptone 5 g yeast extract 10 g NaCl
DYT	16 g tryptone 10 g yeast extract 5 g NaCl
TB	12 g tryptone 24 g yeast ectract 5 g glycerol 17 mM KH_2PO_4 72 mM K_2HPO_4
SOB	20 g tryptone 5 g yeast ectract 0.6 g NaCl 0.2 g KCl 10 mM $MgSO_4$ 10 mM $MgCl_2$
SOC	SOB 20 mM glucose
ZYP5052	10 g tryptone 5 g yeast extract 50 mM KH_2PO_4 50 mM Na_2HPO_4 25 mM $(NH_4)_2SO_4$ 2 mM $MgSO_4$ 0.5 g glucose 5 g glycerol 2 g α-lactose 0.2 mL 1000x ZYP5052-trace elements

NMM	22.5 mM KH_2PO_4 50 mM K_2HPO_4 7.5 mM $(NH_4)_2SO_4$ 8.5 mM NaCl 1 mM $MgSO_4$ 20 mM glucose 1 mg $CaCl_2$ 1 mg $FeCl_2$ 10 mg biotin 10 mg thiamine 50 mg cAA 1 mL 1000x NMM-trace elements

(1000x NMM-trace elements: 10 µg/mL (Cu^{2+}, Zn^{2+}, Mn^{2+}, $MoOH^{2+}$); 1000x ZYP5052-trace elements: 50 mM Fe^{3+}, 20 mM Ca^{2+}, 10 mM (Mn^{2+}, Zn^{2+}), 2 mM (Co^{2+}, B^{3+}, Cu^{2+}, Ni^{2+}, Mo^{6+}, Se^{4+})

All media supplements used during this study were sterilized by filtration (ø 0.22 µm) and are outlined in **Table 7**.

Table 7 List of supplements and their stock concentrations.

Name	**Stock concentration**
Ampicillin	100 mg/mL
Arabinose	20 % w/v
Chloramphenicol	37 mg/mL
Glucose	20 % w/v
Kanamycin	50 mg/mL
Streptomycin/Spectinomycin	50 mg/mL /100 mg/mL
Tetracycline	20 mg/mL
fAA	100 µM

5.5 Buffers and solution

All buffers were prepared with dH_2O, and ddH_2O, respectively. For protein purification using a column, the solution was additionally filtered (ø 0.22 µm) to remove solid particles. The pH values were adjusted after combining all components and cooling the solutions to 4 °C. The table below gives an overview of all used buffers, and solutions including their compositions.

Table 8 List of used buffers and their compositions in their particular applications.

Application	Name	Composition
Agarose Gel Electrophoresis	6x DNA loading dye	0.25 % bromphenol blue 0.25 % xylencyanole 30 % glycerol
	50x TAE buffer	2 M TRIS 2M acetic acid 10 % v/v 0.5 M EDTA pH 8.0
Colony PCR	TE buffer	10 mM TRIS pH 7.4 1 mM EDTA pH 8.0
Bradford protein assay	Roti™-Quant	5x concentrated
Protein purification by IMAC	Lysis buffer	50 mM TRIS-HCl pH 7.4 - 8.0 150 mM NaCl 1 mM DTT 10 % v/v glycerol
	Binding buffer	50 mM TRIS-HCl pH 7.4 - 8.0 150 mM NaCl 1 mM DTT 10 % v/v glycerol 10 mM imidazole

	Wash buffer	50 mM TRIS-HCl pH 7.4 - 8.0 500 mM NaCl 1 mM DTT 10 % v/v glycerol 20 - 40 mM imidazole
	Elution buffer	50 mM TRIS-HCl pH 7.4 - 8.0 150 mM NaCl 1 mM DTT 10 % v/v glycerol 250 - 500 mM imidazole
	Dialysis buffer	50 mM TRIS-HCl pH 8.0 150 mM NaCl 1 mM DTT 10 % v/v glycerol
	Storage buffer	50 mM TRIS-HCl pH 8.0 150 mM NaCl 1 mM DTT 25 % v/v glycerol
Protein purification by HIC	Binding/wash buffer	25 mM HEPES pH 7.5 1.7 M $(NH_4)_2SO_4$ 10 % v/v glycerol
	Elution buffer	25 mM HEPES pH 7.5 10 % v/v glycerol
	Dialysis/storage buffer	50 mM TRIS-HCl pH 8.0 150 mM NaCl 1 mM DTT 10 % v/v glycerol
Protein purification by IEX	Binding buffer	50 mM TRIS-HCl pH 8.0 5 mM NaCl 10 % v/v glycerol

	Wash buffer	50 mM TRIS-HCl pH 8.0 10 mM NaCl 10 % v/v glycerol
	Elution buffer	50 mM TRIS-HCl pH 8.0 1 M NaCl 10 % v/v glycerol
	Dialysis/storage buffer	50 mM TRIS-HCl pH 8.0 150 mM NaCl 1 mM DTT 10 % v/v glycerol
IleRS purification by IMAC	Binding buffer	50 mM TRIS-HCl pH 8.0 150 mM KCl 1 mM DTT 10 % v/v glycerol 10 mM imidazole
	Wash buffer	50 mM TRIS-HCl pH 8.0 500 mM KCl 1 mM DTT 10 % v/v glycerol 25 mM imidazole
	Elution buffer	50 mM TRIS-HCl pH 8.0 150 mM KCl 1 mM DTT 10 % v/v glycerol 250 mM imidazole
	Dialysis buffer	50 mM TRIS-HCl pH 8.0 150 mM KCl 1 mM DTT 10 % v/v glycerol

	Storage buffer	50 mM TRIS-HCl pH 8.0 150 mM KCl 2 mM TCEP 10 % v/v glycerol
IleRS purification by Gelfiltration	SEC buffer	20 mM BICIN-KOH pH 8.0 150 mM KCl 10 % v/v glycerol 2 mM TCEP
SDS-PAGE	5x SDS loading dye	80 mM TRIS pH 6.8 10 % SDS 0.2 % w/v bromophenol blue 12.5 % v/v glycerol 4 % v/v mercaptoethanol
	10x SDS running buffer	31 g/L TRIS 10 g/L SDS 144 g/L glycine
	Coomassie staining solution	1 g Coomassie Brilliant Bue R-250 500 mL ethanol 100 mL glacial acetic acid add ddH_2O to V_{final} = 1 L
	Separation gel	380 mM TRIS-HCl pH 8.8 12 -15 % acrylamide (37.5:1) 0.1 % SDS 0.05 % APS 0.05 % TEMED
	Stacking gel	125 mM TRIS-HCl pH 6.8 5 % acrylamide (37.5:1) 0.1 % SDS 0.05 % APS 0.17 % TEMED
Silver staining of	50 % MeOH	50 % v/v methanol

protein gels	5 % MeOH	5 % v/v methanol
	32 µM DTT	1.6 µL 1 M dithiothreitol add ddH_2O to V_{final} = 50 mL
	Silver solution	0.1% w/v $AgNO_3$
	Developing solution	3% w/v Na_2CO_3 0.02 % w/v formaldehyde
	Stop reagent	solid citric acid (monohydrate)
TEV cleavage	TEV protease reaction buffer	50 mM TRIS-HCl pH 8.0 150 mM NaCl 1 mM EDTA 5 mM DTT 10 % w/v glycerol

5.6 Kits

Name	**Supplier**
GeneJET™ Gel Extraction Kit	Thermo Fisher Scientific (Waltham, USA)
GeneJET™ PCR Purification Kit	Thermo Fisher Scientific (Waltham, USA)
GeneJET™ Plasmid Mini-prep Kit	Thermo Fisher Scientific (Waltham, USA)
GeneJET™ Plasmid Midi-prep Kit	Thermo Fisher Scientific (Waltham, USA)
Phosphate assay kit	Abcam (Cambridge, England)

5.7 Synthetic oligonucleotides

All oligonucleotides, purchased from Sigma Aldrich (now: Merck KGAA, Darmstadt, Germany), are resuspended in nH_2O to a final concentration of 100 µM and stored at 20 °C. The working concentration for PCR is 10 µM and 100 µM for MAGE, and CRISPR-Cas9, respectively. Oligonucleotides with the length up to 40 bp are ordered in desalted quality, whereas longer ones in HPLC- or PAGE-purified form to exclude truncated products.

5.8 Strains

Table 9 Strains used in this study.

Strain name	Genotype	Source	Application
E. coli BL21 AI	*E. coli* B/r C> F^- *ompT hsdSB (rB-mB-) gal dcm araB::T7RNAP*-tet*A*	Budisa group	Gene expression
E. coli BL21 (DE3)	*E. coli* B C> F^- *ompT gal dcm+ lon hsdSB(r_B^- m_B^-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	Budisa group[332]	Gene expression
E. coli BL21 Gold (DE3)	*E. coli* B C> F^- *ompT hsdSB(r_B^- m_B^-) dcm+ TetR gal λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) endA Hte*	Budisa group	Gene expression
E. coli BL21 (DE3)	*E. coli* B C> F^- *ompT gal dcm+ lon hsdSB(r_B^- m_B^-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) ΔproA*	Budisa group	Gene expression
E. coli C321.ΔA.exp (DE3)	*E. coli* MG1655 Δ(*ybh*B-*bio*AB)::*zeoR* Δ*prfA* λ(DE3)	Budisa group	Gene expression
E. coli CAG18341	*E. coli* K12 C> F^- *λ- ilvG- rfb-50 rph-1 nadA3052::Tn10 (KanR)1 rph-1 ilvD*	Budisa group Coli Genetic Stock Center (CGSC)[267]	Gene expression, cloning
E. coli ML6	*E. coli* B C> F^- *ompT gal* $hsdS_B$ (rB^- mB^-) *dcm lon λ*(DE3 *[lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) cyo::kan ilvE avtA*	Budisa group[267]	Gene expression

E. coli Dh10b	*E. coli* K12 C> F- *mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 rpsL150(Str^R) araD139 Δ(ara, leu)7697 galU galK λ- nupG*	Budisa group Invitrogen[261]	Gene expression, cloning, plasmid preparation
E. coli JM83	*E. coli* K12 C> F- *ara Δ(lac-proAB) [Φ80dlac Δ(lacZ)M15] rpsL(StrR) thi*	Budisa group	Gene expression
E. coli MDS^{TM}42	*E. coli* K12 MG1655 *C> multiple deletions of insertion elements*	Budisa group Scarab Genomics[283]	Gene expression, cloning
E. coli NEB10-beta	*E. coli* K12 C> F- *Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15e14-ϕ80dlacZ ΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	Budisa group	Gene expression, cloning, plasmid preparation
E. coli Omnimax	*E. coli* K12 C> F- *{proAB+ lacIq lacZΔM15 Tn10(TetR) Δ(ccdAB)} mcrA Δ(mrr-hsdRMSmcrBC) φ80 (lacZ)ΔM15 Δ(lacZYA-argF) U169 endA1 recA1 supE44 thi-1 gyrA96 relA1 tonA panD*	Budisa group	Cloning, plasmid preparation
E. coli TOP10	E. coli K12 C> F- *mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZ*ΔM15 ΔlacX74 recA1 araD139 Δ(ara leu) 7697 galU galK rpsL (StrR) *endA1 nupG*	Budisa group[260]	Gene expression, cloning, plasmid preparation

5.9 Enzymes and biomolecular reagents

Thermo Fisher Scientific (Waltham, USA), New England Biolabs (Frankfurt am Main, Germany), and Carl Roth (Karlsruhe, Germany) are the suppliers of all used enzymes and biomolecular reagents, listed in **Table 10**.

Table 10 Enzymes and biomolecular reagents used in this study.

Name	**Supplier**
DNase	Carl Roth (Karlsruhe, Germany)
FastDigest restriction enzymes	Thermo Fisher Scientific (Waltham, USA)
High Fidelity restriction enzymes	New England Biolabs (Frankfurt am Main, Germany)
Lysozyme	Carl Roth (Karlsruhe, Germany)
Phusion High-Fidelity DNA Polymerase	New England Biolabs (Frankfurt am Main, Germany) Thermo Fisher Scientific (Waltham, USA)
Pyrophosphatase	New England Biolabs (Frankfurt am Main, Germany)
Q5® High-Fidelity DNA Polymerase	New England Biolabs (Frankfurt am Main, Germany)
restriction enzymes	New England Biolabs (Frankfurt am Main, Germany) Thermo Fisher Scientific (Waltham, USA)
RNase	Carl Roth (Karlsruhe, Germany)
T4 DNA Ligase	Thermo Fisher Scientific (Waltham, USA)
Taq DNA Polymerase	Budisa group
TEV protease	This study

5.10 Plasmids

Table 11 Plasmids used in this study

Name	promoter	Resistance marker	Source/Reference
pQE80L H_6-TEV-IleRS	T5	AmpR	Budisa group
pQE80L H_6-TEV-IleRS (Ala10)	T5	AmpR	This study
pQE80L H_6-TEV-IleRS(A579P)	T5	AmpR	This study
pQE80L H_6-TEV-IleRS(G56A)	T5	AmpR	A. J. Mamoud (supervised by Dr. J.-S. Völler and me)
pQE80L H_6-TEV-IleRS(G56N)	T5	AmpR	A. J. Mamoud (supervised by Dr. J.-S. Völler and me)
pQE80L H_6-TEV-IleRS(G56P)	T5	AmpR	A. J. Mamoud (supervised by Dr. J.-S. Völler and me)
pQE80L H_6-TEV-CP1	T5	AmpR	This study
pQE80L H_6-TEV-IleRS(W905R)	T5	AmpR	This study
pET28a+ H_6-TEV protease	T7	KanR	Dobbek group
pNB26'2_EctRNA_UAU-EcIleRS	pLac	KanR	Budisa group
pNB26'2_EctRNA_UAU_2-EcIleRS	pLac	KanR	Budisa group
pET28a+ sfGFP-H_6	T7	KanR	Budisa group
pET28a+H_6-sfGFP(I136T)-Strep	T7	KanR	This study
pET28a+H_6-sfGFP(I188A)-Strep	T7	KanR	This study
pET28a+H_6-sfGFP(I136T,I188A)-Strep	T7	KanR	This study
pET28a+H_6-sfGFP(I14M,I136T,I188A)-Strep	T7	KanR	This study
pET28a+H_6-sfGFP(I14M,I47V,I136T, I188A)-Strep	T7	KanR	This study
pET28a+H_6-sfGFP(I14M,I152V,I136T, I188A)-Strep	T7	KanR	This study

pET28a+H_6-sfGFP(I14M,I123T,I128T, I136T,I188A)-Strep	T7	KanR	This study
pET28a+H_6-sfGFP(I123T,I128T,I136T, I152V,I188A)-Strep	T7	KanR	This study
pET28a+H_6-sfGFP(I123T,I128T,I136T, I152V,I161L,I167V,I188A)-Strep	T7	KanR	This study
pET28a+H_6-sfGFP(I47V,I123T,I128T, I136T,I152V,I161L,I167V,I188A)-Strep	T7	KanR	This study
pET28a+H_6-sfGFP(I14NNK,I47V, I123T,I128T,I136T,I152V,I161L,I167V, I188A)-Strep	T7	KanR	This study
pET28a+H_6-sfGFP(I14NNK,I47V, I123T,I128T,I136T,I152V,I161L,I167V, I188A)-Strep	T7	KanR	This study
pET28a+H_6-sfGFP(I14NNK,I47V, I98NNK,I123T,I128T,I136T,I152V,I161L, I167V,I188A)-Strep	T7	KanR	This study
pET28a+H_6-sfGFP(I14NNK,I47V, I123T,I128T,I136T,I152V,I161L,I167V, I188A,I229NNK)-Strep	T7	KanR	This study
pET28a+ H_6-GFP1-10	T7	KanR	This study
pBU16_1GK_*Mj o*NBY	glnS'	AmpR	Budisa group
pPAB26 CAT(Q98TAG,D181TAG) PT5 His-$SUMO_{opt}$-sfGFP(R2TAG)-Strep *Mj* tRNA(opt, CUA)	T5, proK	KanR	Budisa group
pKD46	pBAD	AmpR	Budisa group
pORTMAGE-2	pL	AmpR	Addgene #72677
pCAGO	pTrc, pBAD	AmpR	GeneScript[282]
pJZ_Ptrp_*Mm*PylTS(C348W,W417S)	pTrp	CmR	Budisa group
pJZ_Ptrp_*Mb*PylTS(N311M,C313Q, W382N,V366G,R85H) = PCC1RS	pTrp	CmR	This study
pET28a cfsfGFP(2TAG)-H_6	T5	KanR	This study
pQE80L EGFP-H_6	T5	AmpR	Budisa group
pQE80L H_6-NowGFP(H7Q)	T5	AmpR	This study
pQE80L H_6-KillerOrange	T5	AmpR	This study

6 Methods

6.1 Molecular biological methods

6.1.1 Plasmid DNA preparation

Relative to the amount of expected plasmid DNA from *E. coli* cells, either GeneJet Plasmid Mini-prep Kit (Thermo Fisher Scientific, Waltham, USA) was used for small-scale isolation of plasmid DNA, or GeneJET Plasmid Midi-prep Kit (Thermo Fisher Scientific, Waltham, USA) for medium-scale purpose. According to the manufacturer's instructions, 5 mL and 50 mL cell culture were applied, respectively. The eluted plasmid DNA was stored at -20 °C until further use.

6.1.2 Polymerase chain reaction (PCR)

The amplification of DNA sequences for site directed-mutagenesis, cloning, and insert verification was performed by PCR in LabCycler thermocycler.

6.1.2.1 Standard PCR

Primers used in standard PCR were manually designed by means of ApE and Geneious software, respectively. The particular reactions were performed with either Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, Waltham, USA and New England Biolabs, Frankfurt am Main, Germany) or Q5® High-Fidelity DNA polymerase (New England Biolabs, Frankfurt am Main) utilizing subsequent protocol (**Table 12**) and hot-start program to prevent primer dimerization (**Table 13**)

Table 12 Composition of a 50 µL standard PCR mixture.

compound	**amount**
5x HF buffer	10.0 µL
5x GC Enhancer (optional)	10.0 µL
10 µM dNTPs	1.0 µL
10 µM primer forward	2.5 µL
10 µM primer reverse	2.5 µL

template	10 - 50 ng
DMSO (optional)	2.5 – 10 %
polymerase	0.5 µL
nH_2O	add to 50 µL

Table 13 Hot-start PCR program

PCR step	temperature	time	cycles
initial denaturation	98 °C	3 min	1
denaturation	98 °C	10 s	30 - 35
annealing	54 – 72 °C	20 s	
extension	72 °C	22.5 s/kb	
final extension	72 °C	5 – 10 min	1
end	12 °C	hold	1

For PCR with several side products, two-step PCR was also used to increase specific product yield (**Table 14**).

Table 14 Two-step PCR program

PCR step	temperature	time	cycles
initial denaturation	98 °C	1 min	1
denaturation	98 °C	10 s	5 – 10
annealing	54 – 72 °C	20 s	
extension	72 °C	22.5 s/kb	
denaturation	98 °C	10 s	25 – 30
annealing/extension	72 °C	22.5 s/kb + 20 s	
final extension	72 °C	5 – 10 min	1
end	12 °C	hold	1

Parental DNA template was destroyed by subsequent digestion with the Type IIM restriction enzyme DpnI (Thermo Fisher Scientific, Waltham, USA) for 3 h at 37 °C or overnight at room temperature. The PCR products were analyzed for correct size and purity on agarose gels (**Section 6.4.2**). To get rid of salts and other reagents the PCR product was purified with the GeneJet PCR Purification Kit (Thermo Fisher Scientific, Waltham, USA) and eluted with nH_2O, e.g., prior to electroporation or restriction digest. In order to isolate a certain fragment from a

product mixture, a preparative gel electrophoresis was run to extract the desired DNA from the gel, which was then purified with geneJet Gel Extraction Kit (Thermo Fisher Scientific, Waltham, USA). The product was eluted with nH_2O, frequently, to avoid salts in the extracted sample, which can interfere in subsequent reactions.

6.1.2.2 Site-directed mutagenesis

Particular mutations were introduced into a DNA sequence using the Megawhop cloning procedure, the Golden Gate Assembly strategy or QuikChange Site-Directed Mutagenesis procedure (Agilent Technologies, Waldbronn, Germany) (**Section 6.1.3.3**).[333–335] Primers were designed with the help of the Geneious software, the online tool NEB Golden Gate Assembly Tool, and according to the QuikChange instruction manual, respectively. The composition of the PCR, the programs and the subsequent steps were described in previous **Section 6.1.2.1**. 2 - 5 µl of purified PCR product was transformed into 50 µL chemically competent or 50 µL electrocompetent *E. coli* cells (**Section 6.2**). Plasmid DNA was isolated from at least three individual clones and the successful mutagenesis was verified by sequencing (Microsynth Seqlab GmbH, Göttingen, Germany).

6.1.2.3 Colony PCR

Colony PCR was carried out in order to identify the presence of both certain chromosomal and plasmid DNA sequence in intact *E. coli* cells. Each single colony was resuspended in 10 µL TE-buffer and used as template in a colony PCR performed with following composition (**Table 15**) and protocol (**Table 16**).

Table 15 Composition of a 20 µL standard colony PCR mixture

compound	**amount**
5x Dream Taq polymerase buffer green	2.0 µL
10 µM dNTPs	0.5 µL
10 µM primer forward	0.5 µL
10 µM primer reverse	0.5 µL
template	1.0 µL
Taq DNA polymerase	1.0 µL
nH_2O	14.5 µL

Table 16 Colony PCR program

PCR step	temperature	time	cycles
initial denaturation	95 °C	3 min	1
denaturation	95 °C	30 s	30 - 35
annealing	54 – 72 °C	45 s	
extension	72 °C	1 min/kb	
final extension	72 °C	10 - 20 min	1
end	12 °C	hold	1

6.1.3 Cloning

6.1.3.1 Restriction digest

In order to analyze the plasmid or to prepare it for further cloning procedures purified PCR products were digested with restriction enzymes according to manufacturer's instructions. Depending on analytical or preparative purpose 200 ng and 1 - 5 µg of DNA was used, respectively. After 1 - 2 h of incubation 6x DNA loading dye was added to each sample and loaded on an 1 % agarose gel for analysis. For further cloning procedures the digested DNA fragments were directly spin-column or agarose gel purified with the GeneJet Gel Extraction Kit (Thermo Fisher Scientific, Waltham, USA). The eluted DNA concentration could be increased by minimized elution volume.

6.1.3.2 Ligation

The purified DNA fragments including backbone and inserts were ligated by T4 DNA ligase. Given by the property of the fragments, sticky-end, and blunt-end ligation was performed using the composition in **Table 17** and **18**, respectively. The volume of insert DNA was determined independently by following equation:

$$V_{insert} = \frac{1}{c_{insert}} \cdot x \cdot m_{backbone} \cdot \frac{s_{insert}}{s_{backbone}}$$

V = volume in µL
m = mass in ng
x = size in bp
s = factor of insert molar excess
c = concentration in $\frac{ng}{\mu L}$

Table 17 Reaction mixture for sticky-end ligation.

compound	amount
10x T4 DNA ligase buffer	2 µL
backbone DNA	30 - 200 ng
insert DNA	3:1 to 5:1 molar ratio over backbone
T4 DNA ligase	1 µL
nH_2O	add to 20 µL

Table 18 Reaction mixture for blunt-end ligation.

compound	amount
10x T4 DNA ligase buffer	2 µL
backbone DNA	30 - 200 ng
insert DNA	3x to 5x molar excess over backbone
50 % PEG 4000 solution	2 µL
T4 DNA ligase	1 µL
nH_2O	add to 20 µL

Typically, ligation was carried out for10 - 50 min at room temperature, 1 h at 22 °C, or overnight at 16 °C. 2 - 5 µL of the reaction mixture was transformed into 50 µL chemically competent, 2.5 µL for 50 µL electrocompetent *E. coli* cells, whereas the ligation product was desalted by spin-column purification or dialysis (MF-Millipore Membrane, 0.025 µm pore size, ø mm) prior to transformation.

6.1.3.3 Golden Gate Assembly

The Golden Gate Assembly strategy was applied for efficient seamless cloning of screening plasmids and synthetase libraries. In the first step, the class IIS restriction endonuclease BsaI (New England Biolabs, Frankfurt am Main, Germany) or Eco31I FD (Thermo Fisher Scientific, Waltham, USA) recognized asymmetric DNA sequences and cleaved outside of their recognition site, after 1 h incubation at 37 °C. The digested DNA fragment having sticky ends with 4 bp overhang offered a precise template for following ligation with T4 DNA ligation (**Section 6.1.3.2**). Alternatively, digestion and ligation were carried out using one-pot protocol listed below (**Table 19**).

Table 19 Program of one-pot assembly.

assembly step	temperature	time	cycles
digestion	37 °C	3 min	25 - 30
ligation	16 °C	4 min	
heat inactivation	80 °C	5 min	1

For library cloning, each backbone and insert were amplified in eight PCR mixtures (á 20 µL), in order to increase the randomization rate (see **Section 3.2.1**). 1 µg of purified backbone was then applied with a fivefold molar excess of pure inserts and the ligation preparation was split into four tubes, each 20 µL total volume. Ligation mix was spin-column purified and transformed twice into electrocompetent *E. coli* cells to enhance the transformation efficiency. The assembly product was verified by sequencing of isolated plasmid DNA.

6.2 Microbiological methods

6.2.1 Cultivation and storage of *E. coli* cells

According to the usage, *E. coli* strains were cultivated in 2 - 20 mL liquid and solid media (**Section 5.4**) supplemented with appropriate selection antibiotics or other supplements like ncAA (**Sections 5.4**, **6.2.5.2** and **6.2.5.3**). The cultures were incubated in 12 mL cultivation tubes or flasks with different sizes at 30 °C or 37 °C and 200 - 250 rpm. Bacterial clones were kept on agar plates for a maximum of six weeks at 4 °C. For long-term storage *E. coli* strains were stored in liquid media supplemented with 25 % w/v glycerol at -80 °C.

6.2.2 Preparation of competent *E. coli* cells

In order to prepare right competent cells for high transformation efficiency, it was crucial to pre-cool every reaction tube (50 mL-falcon and 1.5 mL Eppendorf tube) and solution on ice, which were applied in the entire preparation procedure. An *E. coli* strain was grown in LB with respective antibiotics overnight at 37 °C, 220 rpm and used for inoculation of the main culture on the next day.

6.2.2.1 Chemically competent *E. coli* cells[241]

200 mL LB media or DYT media with appropriate antibiotics were inoculated with 2 mL of an overnight culture (1:100). Cells were incubated at 37 °C and 220 rpm, until they reached the OD_{600} between 0.2 - 0.5. Culture was then transferred to 50 mL-falcons and chilled on ice for 20 min. Cells were harvested by centrifugation (3000 g, 4 °C, 7 min) and washed in 45 mL 75 mM $CaCl_2$. After they have chilled on ice for 30 min, cells were harvested (3000 g, 4 °C, 10 min) and resuspended in 2 mL 75 mM $CaCl_2$ and 25 % w/v glycerol, resulting in 100-fold concentration. Finally, competent cells were aliquoted in 1.5 mL microcentrifuge tubes (100 µL each) and stored at -80 °C until use.

6.2.2.2 Electrocompetent *E. coli* cells[240]

2 mL of an overnight culture was incubated in 200 mL LB media or DYT media with required antibiotics (1:100) at 37 °C and 220 rpm, until it reached the OD_{600} between 0.2 - 0.5. Cells were immediately transferred to 50mL-falcons and kept on ice for 20 min. Afterwards, they were harvested (3000 g, 4 °C, 7 min) and washed three times with 20 mL 10 % w/v glycerol (3000 g, 4 °C, 10 min). Electrocompetent *E. coli* cells were then resuspended in 1 mL ice-cold 10 % w/v glycerol, aliquoted in 1.5 mL microcentrifuge tubes (50 µL) and stored at -80 °C until transformation.

6.2.3 Transformation of competent *E. coli* cells

Each aliquot of frozen or freshly prepared cells were kept on ice for 5 min and then supplied with up to a maximum of 50 ng plasmid for standard transformation. In other cases, like ligation product and library transformation, 5 to 10 % of the competent cell's volume was used for higher efficiency. Again, it was necessary to keep every step on ice and use pre-cooled reaction tubes.

6.2.3.1 Transformation of chemically competent *E. coli* cells

20 – 50 ng plasmid DNA was added to one 50 µL aliquot of chemically competent cells and incubated on ice for 20 – 30 min. Cells were heat shocked at 42 °C for 30 – 60 s and immediately put on ice to rest for 2 min. 950 µL SOB or SOC media were added to the reaction

tube and cells could then recover at 37 °C for 1 h or 30 °C for 2 h. Finally, they were plated on LB-agar plates containing 1 % w/v glucose as well as appropriate selection marker and incubated overnight depending on approach at 30 °C or 37 °C or kept at room temperature for 72 h.

6.2.3.2 Electroporation

10 – 50 ng plasmid DNA was pre-mixed in 50 µL electrocompetent cells and transferred to the cold electroporation cuvette (1 mm gap width). Cells were electroporated in an electroporator (BioRad Gene Pulser Xcell) by applying an electrical pulse of 1.8 kV for approximately 5 ms and immediately resuspended in 950 µL SOB or SOC media. Cells were then recovered at 37 °C for 1 h or 30 °C for 2 h. Subsequently, electroporated cells were plated on LB-agar (1 %) with particular supplements and incubated upon approach at 30 °C or 37 °C or kept at room temperature for 72 h.

6.2.4 Genome engineering

6.2.4.1 Multiplex Automated Genome Engineering (MAGE)

In order to introduce the thermosensitive mutation A579P in IleRS of convenient strains CAG18431 and ML6, MAGE was applied, though not automated but manually.[274,276] Using the Geneious software, a 91 bp, HPLC-grade single stranded oligo was designed according to the protocol of Wang *et al*, which carries the desired mutation A579P and mediates the allelic replacement.[276] This oligo was integrated into the genome of the selected strains using only Beta component of the lambda red (λ-red) recombination system, which was supplied on the plasmids pKD46 and pORTMAGE, respectively (**Section 5.10**).[275,336] These two plasmids differed in the regulation of the Beta gene by various promoters: on the one hand expression of the recombinase is under control of the araB promoter on pKD46, that need 1 % w/v L-arabinose as inducer, and on the other hand the temperature inducible (42 °C) expression system pORTMAGE is based on the phage lambda pL promoter regulation.
Strains carrying pKD46 or pORTMAGE were grown in 10 mL LB medium with 100 µg/mL Amp at 30 °C and 220 rpm until they reached the mid-log phase at OD_{600} ~ 0.20 to 0.35. The expression of the recombinase enzyme was induced according to their promoter. At OD_{600} ~ 0.4 to 0.7 cells were chilled on ice for 20 min to prevent degradation. Subsequently, they were made electrocompetent by washing three times with ice-cold sterile water at 4 °C (3000 g,

7 min) and resuspended in 100 µL water. 1 µL 100 µM oligo (MAGEIleRStsA579P) was mixed with 50 µL freshly prepared cells and applied to the electroporation cuvette (1 mm gap width). After electroporation at 1.8 kV for ~ 5 ms approximately 95 % of cells were killed, therefore recovery in 5 mL TB medium at 30 °C and 220 rpm for 1 h was indispensable. 5 mL LB medium with appropriate selection marker were added to the mixture and cells were grown to mid-log phase (OD_{600} ~ 0.20 to 0.35) for the subsequent cycle. The schematic summary of the MAGE cycling procedures are shown in **Figure 46** (**Section 3.2.5**). Three to five cycles of this genome engineering method were achieved for each mutation purpose. 1:5000 diluted cells were plated on LB-Amp_{100} agar plates and incubated at 30 °C overnight. To check the success of the mutation, single clones were duplicated on two LB-Amp_{100} agar plates. One was incubated at 30 °C and the other one at 42 °C overnight. The strain carrying the thermosensitive mutation of IleRS is capable to grow on plate, incubated at 30 °C and dies at the higher temperature, 42 °C. Plates were scanned and evaluated individually. Promising clones were selected to amplify the genome fragment containing the mutation A579P. The PCR products were spin-column purified and sent for sequencing.

6.2.4.2 CRISPR/Cas9

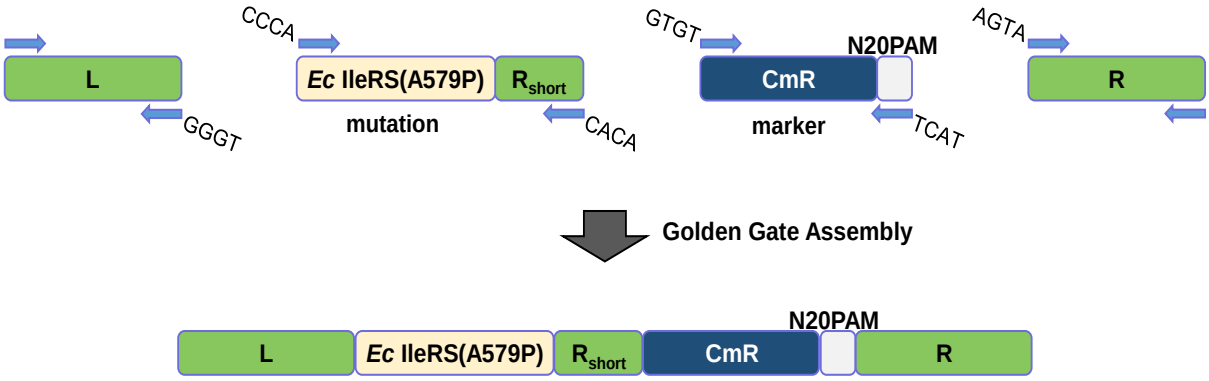


Figure 62 Assembly of editing cassette. Homologous regions (*green*), upstream (left arm, L) and downstream (shorter and full-length right arm, R_{short} and R) of the A579P mutation (highlighted in beige) were amplified from the editing strain using appropriate primers (*blue arrows*, **Appendix**). Selection marker cassette (*blue*, CmR) and the specific CRISPR/ Cas9 recognition site (*grey*, N20PAM) were obtained with amplification of the plasmid pKD46 (**Section 5.10**) employing corresponding primers (*blue arrows*). Amplified products carrying the optimized 4 bp type IIS restriction linkers (complement sequences on top and bottom of *blue arrows*) were applied in Golden Gate Assembly to gain the editing cassette (**Section 6.1.3.3**). Direction and accuracy of the specific digestion and ligation of the fragments were given by the four sticky base pairs.

The introduction of the thermosensitive mutation A579P in genomic IleRS gene was carried

out by a CPRISPR/cas9-assisted gRNA-free one-step (CAGO) genome editing technique, whereby both necessary enzymatic systems, λ-red and CRISPR/Cas9, were supplied on one plasmid, namely pCAGO (**Figure 49**).[279,282,337–339] The λ-red system (Exo, Beta, and Gam), controlled by the trc promoter (Ptrc), was responsible for the insertion of the editing cassette and for the intramolecular homologous recombination. The double-strand break at target sequence and the subsequent removal of the editing cassette for a scarless genome editing were achieved by the CRISPR/Cas9 system, which was under control of pBAD promoter (ParaB). With the CAGO technique, no specific gRNA was needed due to a universal N20PAM sequence (Protospacer Adjacent Motif), supplied on an editing cassette (**Figure 48**), which was generated by Golden Gate Assembly (**Section 6.1.3.3**). The editing tool is consisted of a left homologous arm (L), a fragment including the mutation (*Ec* IleRS(A579P)), a short right homologous arm (R_{short}, 45 bp), a selection marker (Chloramphenicol resistance gene, CmR, 1037 bp) with a specific 23 bp N20PAM, and a right homologous arm (R). The length of the homologous fragments varies from 500 to 1000 bp. Homologous regions were essential for the λ-red recombineering procedures, whereas the N20PAM was necessary for the recognition and accurate direction of the CRISPR/Cas9 system.

In the first step, the plasmid pCAGO was transformed in *E. coli* cells, MDS42, Dh10b and Top10, by electroporation. One clone of each cells was allowed to grow in 5 mL LB media and 100 µg/mL Amp at 30 °C until OD_{600} ~ 0.2 was reached (**Figure 49**). Then, expression of the λ-red recombineering system was induced with 0.1 - 1 mM IPTG. At OD_{600} ~ 0.6 cells were washed three times with ice-cold 10 % w/v glycerol and resuspended in 50 µL wash solution. In the next step, 400 - 1500 ng editing cassettes (several compositions of editing cassette are given in **Section 3.2.5**) were inserted in the electrocompetent cells by electroporation in a cuvette (1 mm gap width) at 1.8 kV for 5 - 6 ms. Cells were immediately resuspended in 1 mL LB media and recovered for 2 - 12 h at 22 °C. Subsequently, cells were plated on LB-agar with 100 µg/mL Amp and 30 µg/mL Cm to select cells carrying the editing cassette. Survived clones were amplified by PCR with primers, one binds outside and one inside of the editing cassette, to verify the presence of this fragment in the appropriate genome. Amplified product with expected fragment size was send for sequencing. Correct clones were transferred into 5 mL LB media with 100 µg/mL Amp, 1 mM IPTG and 0.2 - 2.0 % w/v L - arabinose and incubated over night to induce the λ-red recombinases and the CRISPR/Cas9 system. 100 µL cells were then plated on LB-agar with Amp.

Six clones were incubated in LB media with Amp, and Amp in combination with Cm, respectively. Cultivated cells were analyzed by colony PCR (**Section 6.1.2.3**) with primers that bind upstream and downstream of the homology arms. Clones exclusively survived in LB media with Amp and expected PCR products were send for DNA sequencing (**Section 6.4.8**) to verify the thermosensitive mutation A579P.

6.2.5 Gene expression in *E. coli*

For a standard recombinant protein expression, the plasmid carrying the target gene under control of inducible promoter was transformed into a suitable *E. coli* expression strain (**Section 5.8**, **Table 9**). Single colony was then inoculated in 10 - 20 mL LB with appropriate antibiotic and grown overnight at 37 °C. A part of this preculture was used for 500 - 1000 mL main culture (ratio 1:100). The cells were grown at 37 °C, 220 rpm until OD_{600} reached 0.5 - 1.0. According to the type of promoter, gene expression was induced by addition of 0.1 - 1.0 mM IPTG, and 0.2 - 2.0 % w/v L – arabinose, respectively. Proteins were expressed at 22 - 37 °C, 220 rpm for 4 - 12 h, according to protein properties. Cell culture was harvested by centrifugation at 4 °C for 20 min at 5000 g. The pellet was stored at -80 °C until further usage. Culture samples prior and after induction were normalized to OD_{600} = 1.0 per 1 mL cell culture and taken for SDS-PAGE analysis.

6.2.5.1 Test expression

In order to evaluate solubility of proteins *in vivo*, small scale expressions were performed.
An overnight culture was used to inoculate 5 - 20 mL LB medium with appropriate antibiotic at a ratio 1:100. Cells were grown at 37 °C, 220 rpm until OD_{600} ~ 0.5. Gene expressions were induced by addition of 0.5 mM IPTG and 1 % w/v L – arabinose, respectively. Expression cultures were shaken for 4 h at 37 °C, 220 rpm. Cells were sedimented at 4 °C for 10 min at 8000 g. The harvested pellets were stored at -80 °C until further usage. For analysis via SDS-PAGE, samples of non-induced and induced cultures were normalized to OD_{600} = 1.0 per 1 mL cell culture.

6.2.5.2 Selective-pressure incorporation (SPI)

Auxotrophic *E. coli* strains were used for SPI method that incorporated ncAA analogues residue-specifically at cAA positions missing in the AA biosynthesis of the strain. As part of this work Pro of different fluorescent proteins were exchanged by their isostructural analogues. Strains used for these purposes are listed with appropriate application in the table below.

Table 20 Summary of auxotrophic strains used in SPI methods. For detailed genotype description and plasmid information see **Section 5.8** and **5.10**.

strain	auxotrophy	target protein	promoter/inducer	incorporated ncAA
JM83	Pro	EGFP NowGFP KillerOrange	PT5/IPTG	(4S)-FPro, (4R)-FPro, 4,4-F_2Pro, 3,4 DhPro
BL21 ΔproA				

A plasmid encoding the target gene under control of T5 promoter was transformed into competent cells. A single colony was grown in 10 mL LB with antibiotic and 1 % glucose at 37 °C, 220 rpm overnight. 0.5 - 1.0 L NMM containing 50 mg/L of all cAA, and appropriate antibiotic were inoculated 1:200 with the preculture. Cells were then cultured at 37 °C, 220 rpm. At OD_{600} ~ 1.0 cells were washed with 45 mL ice-cold NMM without cAA (10 min, 4 °C, 3500 g) and resuspended in 5 mL NMM at room temperature. Subsequently, washed cells were added to 0.5 - 1.0 L freshly prepared NMM with antibiotic, 0.05 - 1.0 mM ncAA and all cAA except Pro, depending on AA that should be exchanged. Cultures were shaken at 37 °C for 40 min to deplete residual intracellular concentration of cAA for substitution. Target gene expression was induced with 0.5 - 1.0 mM IPTG and performed at 30 °C for another 6-14 h. Finally, the expression culture was harvested by centrifugation at 4 °C, 8000 g for 20 min. Pellet was washed in 15 mL lysis buffer and stored at -80 °C. Normalized samples of non-induced and induced cultures were taken for SDS-PAGE analysis.

6.2.5.3 Stop codon suppression (SCS)

The site-specific incorporation of *S*-allylcysteine (Sac) at amber stop codon positions the cysteine-free superfolder GFP mutant (cfsfGFP) was performed in BL21 AI or *E. coli* C321.ΔA.exp (DE3), bearing orthogonal pairs via SCS (**Table 21**).

Table 21 Summary of SCS setup.[139]

Target protein	promoter/ inducer	AARS used for ncAA incorporation	in-frame stop codon position	incorporated ncAA
cfsfGFP	PT5/IPTG	*Mm* SacRS(C348W,W417S)	2	Sac
		Mb PCC1RS		

Single colonies of *E. coli* carrying the plasmids encoding cfsfGFP(2TAG) and the orthogonal pair *Mm* SacRS (C348W, W417S): $tRNA_{CUA}^{Pyl}$, and *Mb* PCC1RS: $tRNA_{CUA}^{Pyl}$ (**Section 5.10**), respectively, was inoculated in 5 mL DYT media with Cm37, Amp100 and 1 % glucose overnight at 37 °C, 220 rpm. The preculture was then added 1:1000 to a 0.5 - 1.0 L DYT media containing appropriate antibiotics and 0.2 % glucose. Cells were grown to OD_{600} ~ 1.0, then supplied with 5 mM Sac and 1 mM IPTG. Protein expression was performed at 37 °C, 220 rpm for 4 h. Cell culture was harvested by centrifugation and washed with 15 mL lysis buffer. Finally, pellet was freezed at -80 °C. Wild type cfsfGFP was expressed using the protocol from **Section 6.2.5**. Non-induced and induced samples were taken for SDS-PAGE analysis, after normalized to OD_{600} = 1 per 1 mL culture.

6.2.6 Cell lysis

6.2.6.1 Standard cell lysis

Expressed cells were thawed on ice and resuspended in lysis buffer using the ratio of 1 g wet cells in 5 mL buffer (**Section 5.5**). 1.5 mg lysozyme, 135 mg DNase, 135 mg RNase, 3 mM $MgCl_2$ and 1 mM PMSF were added to each gram of cell mass. Cell suspension was then incubated on ice for 1 h and mixed occasionally by inverting 4 - 6 times. Cells were lyzed until the solution became clear by a pressure-based homogenizer (Microfluidizer M-110L) or via sonification (Sonopuls HD3200) twice for 6 min each with 45 % amplitude at an interval of 2 s on and 5 s off. The cell lysate was separated after centrifugation at 4 °C and 18 000 g for 1 h and filtered (ø 0.45 µm), subsequently. Filtrate was used for immediate purification or frozen at -80 °C using liquid nitrogen. Cell debris was resuspended in the same amount of nH_2O. 80 µL of lysate (soluble fraction) and disrupted cell pellet (insoluble fraction) were kept for SDS-PAGE analysis.

6.2.6.2 Cell lysis using B-PER® Bacterial Protein Extraction Reagent

Another effective method to obtain the soluble fraction from cells was the chemically lysis using B-PER® Bacterial Protein Extraction Reagent, according to the manufacturers' protocol. This technique was beneficial in case of high number of samples and saving time.

6.3 Biochemical methods

6.3.1 Immobilized metal anion chromatography (IMAC)

For purification of all H_6-tagged proteins IMAC was used as the first convenient method. 1 mL and 5 mL HisTrap™ columns (Ni-NTA column from GE Healthcare Life Sciences, München, Germany) was attached to a peristaltic pump for manual handling and connected to ÄKTA FPLC system for automated process. All buffers used for this purification technique are listed in **Section 5.5**, **Table 8**.

The Ni-NTA column was equilibrated with 5 column volume (CV) IMAC binding buffer prior sample loading. The filtered lysate of the cell lysis was then loaded twice to increase the number of binding proteins. Unbound compounds were removed by a wash step with 30 CV appropriate IMAC wash buffer. The desired product was eluted with IMAC elution buffer by step (5 CV each step) or linear gradient (15 CV in total) of increasing imidazole concentration. Relevant fractions were pooled and dialyzed three times against dialysis buffer to get rid of imidazole and salt excess. For a fast buffer exchange, PD-10 desalting column (GE Healthcare Life Sciences, München, Germany) was applied according to the manufacturers' protocol. Depending on further usage, dialysis buffer was equal to the binding buffer of following purification method or storage buffer of the purified protein. At each step of the purification procedure, an 80 µL sample was taken for SDS-PAGE analysis.

6.3.2 Hydrophobic interaction chromatography (HIC)

In case the initial purification by IMAC yielded impure protein sample, HIC can be performed to separate proteins based on their different hydrophobicity. Therefore, samples already adjusted to HIC binding buffer were loaded onto an equilibrated 1 mL HiTrap™ Phenyl FF (Low Sub) column (GE Healthcare Life Sciences, München, Germany) and washed with 20 CV HIC wash buffer. The target protein was eluted by applying a linear gradient of ammonium sulphate concentration from 1.7 M to 0 M (15 CV in total). Fractions containing desired proteins were pooled and dialyzed three times against dialysis buffer that depended on the subsequent application. At each step of purification, an 80 µL sample was saved for analysis by SDS-PAGE. All detailed buffer compositions are listed in **Section 5.5**, **Table 8**.

6.3.3 Ion exchange chromatography (IEX)

To get rid of impurity in protein sample after initial purification by IMAC or HIC, ion exchange chromatography (IEX) was applied. Eluates were dialyzed against the IEX binding buffer and loaded onto an equilibrated anion exchange column, 1 mL or 5 mL HiTrap™ Q Sepharose (GE Healthcare Life Sciences, München, Germany). After a 15 CV wash step using IEX wash buffer, desired protein was eluted from the column with linear gradient of NaCl concentration from 0.1 M to 1.0 M (15 CV in total). Relevant fractions were pooled and dialyzed three times against dialysis buffer. For the SDS-PAGE analysis 80 µL of each purification step were taken. Detailed information of buffer compositions can be found in **Section 5.5**, **Table 8**.

6.3.4 Size exclusion chromatography (SEC)

Prior to crystallization experiments, IleRS was additionally purified by SEC, a high-resolution gel filtration technique to separate molecules according to their size, meaning the discrimination between monomer, oligomer, and aggregated forms of the target protein. HiLoad™ 16/600 Superdex™ 200 pg (GE Healthcare Life Sciences, München, Germany) attached to ÄKTA FPLC system was used in this context. The product from TEV cleavage was loaded on the equilibrated column and eluted isocratically with the SEC buffer (1.3 CV). Based on the absorption at λ = 280 mm (A_{280nm}) eluates were selected, pooled, and concentrated to 1-2 mL with centrifugal devices (**Section 5.2**). SDS-PAGE analysis was carried out with samples collected at each fractionation peak to ensure the correct protein size. Remaining purified samples were aliquoted (25 µL per portion), immediately frozen with liquid nitrogen and stored at -80 °C until further usage. All described steps of SEC were performed in the labs of Dr. Christian Roth (Max Planck Institute of Colloids and Interfaces, Potsdam, Germany).

6.3.5 Digestion with TEV protease

The N-terminal H_6-tag of H_6-TEV-IleRS was cleaved by TEV protease that was heterologously expressed from pET28a+H_6-TEV protease plasmid in *E. coli*. The purified protease digested specifically at the TEV recognition site between H_6-tag and IleRS sequence.
For this purpose, H_6-TEV-IleRS from IMAC purification was transferred into TEV reaction buffer (**Section 5.5**, **Table 8**). At 4 °C, the overnight cleavage was carried out in 1:1 ratio (TEV protease: H_6-TEV-IleRS) on an orbital shaker (Rotamax 120, Heidolph, Schwabach, Germany) with gentle shaking. Obtained reaction mixture was loaded on Ni-NTA column to

separate target protein from the cleaved H_6-tag and TEV protease. IleRS was then collected in the flow through and concentrated to 2 - 3 mL prior purification by SEC (**Section 6.3.4**).

6.3.6 Crystallization of AARS

The gene encoding IleRS mutants were cloned on an expression plasmid and transformed into *E. coli* BL21 (DE3) gold (**Section 5.8**, **5.10**). Enzymes were expressed and purified by IMAC using the N-terminal H_6-tag first, followed by SEC according to protocols in **Section 6.3.1** and **6.3.4**. In case of H_6-tag cleavage, TEV-protease was added to the desalted eluted protein sample from IMAC (**Section 6.3.5**). AARS mutants were crystallized in collaboration with Dr. Christian Roth (Max Planck Institute of Colloids and Interfaces, Potsdam, Germany).

6.3.6.1 Method

The vapor diffusion technique was performed in sitting-drop (for screening) and hanging-drop (for optimization) format for the IleRS crystallization trials (**Figure 63**).[340]

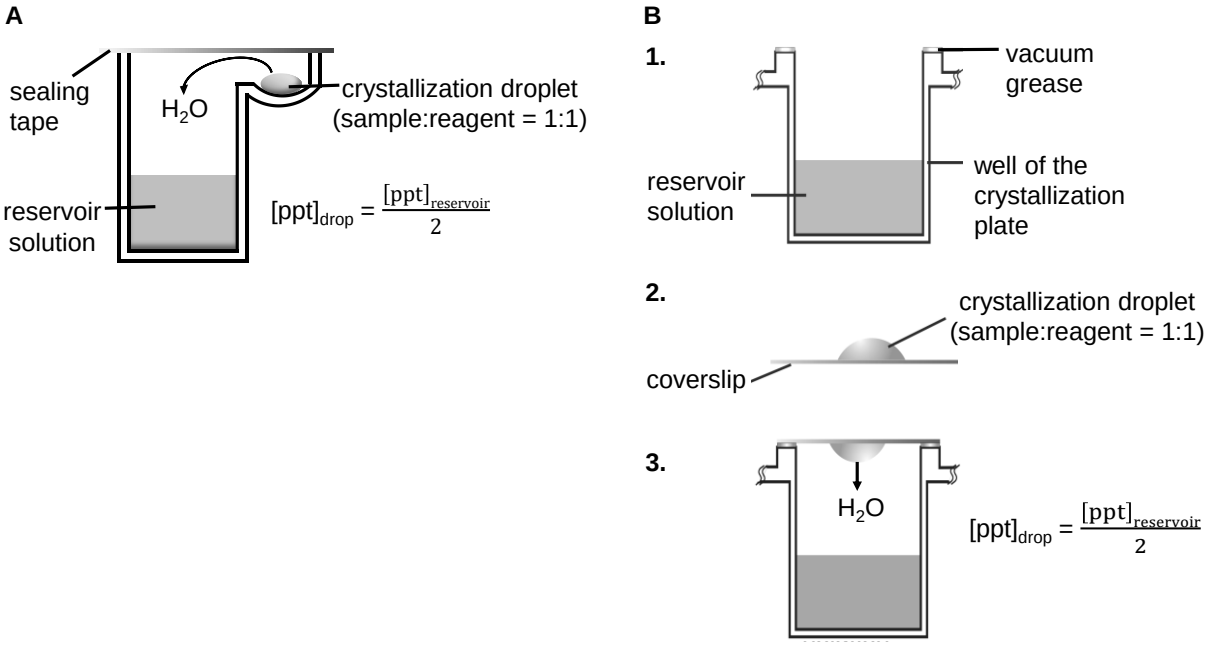


Figure 63 Schematic principle of crystallization method: sitting-drop (A) and hanging-drop vapor diffusion (B). A) The protein solution is in a sitting droplet with 1:1 ratio of the reservoir solution containing crystallizing agents at 2-fold higher concentrations than in the drop. In this way, equilibration of the separated solutions is induced by vapor diffusion. **B**) The same principle is applied to this form of vapor diffusion method with the difference that crystallization droplet is placed on a siliconized coverslip and hang over the reservoir.

A final protein concentration of 20 mg/mL was incubated with 20-fold molar excess of (*S*)-TfeGly (~8 mM) in SEC buffer (**Section 5.5**). The protein-ligand solution was then mixed 1:1 with appropriate precipitants in particular crystallization plates. In later experiments 5 mM AMP-PNP and 1 mM Abu-2AA were added to the mixture in order to stabilize the IleRS structure during crystal formation.

6.3.6.2 Screening

For preliminary identification of crystallization conditions of IleRS was performed by sparse matrix screening. The 2-lens crystallization plate in the 96 well plate format (MRC laboratory of molecular biology, Cambridge, UK) was used for the sitting-drop vapor diffusion system (**Figure 63A**) in combination of seven commercially available crystallization kits, containing 672 different conditions: JBScreen JCSG++, JBScreen Basic, JBScreen PACT++ 1-4, JBScreen PEG-Salts, JBScreen Classic I, JBScreen Classic II, JBScreen Wizard (Jena Bioscience, Jena, Germany), and Index HT (Hampton Research, CA, USA). Two sample droplets were set up for each plate by an Oryx pipetting robot system (Douglas Instruments Ltd., Berkshire, UK), resulting in 1344 screening conditions per plate. Droplet one is composed of 0.1 µL 20 mg/mL IleRS solution containing 7.5 mM TfeGly and 0.1 µL reservoir solution, while droplet two lacked the ncAA. 60 µl of the screening solution was filled into the reservoir chamber. The plate was finally sealed with a plastic film to avoid evaporation and stored at 19°C for drop-reservoir equilibration. After preparation, plates were immediately inspected under a light microscope. Plate observation was repeated every week until no further change in crystal size. Scoring protocol was prepared for conditions and changes in the crystallization drops. For illustrative purposes, pictures were taken from the promising variants. Prior to a diffraction test, crystals were harvested using a nylon fiber loop, shock-frozen and stored in liquid nitrogen.

6.3.6.3 Seeding

The micro seeding method was used as an alternative strategy to obtain protein crystals. Already-grown small crystals of IleRS with 7.5 mM TfeGly in 100 mM MES, pH 6.5, 200 mM $Mg(OAc)_2$, 20 % w/v PEG 8000 (IleRS_screen_1, **Figure 22**, **Section 3.1.2**) were crushed with a small glass tool under observation by the light microscope. The micro crystals were then transferred to a microcentrifuge tube containing glass beads in the 40 µL solution of the previous condition. The mixture was vortexed for 3 min and subsequently cooled on ice until

further use. This seed stock was used in 1:1:1 ratio of the droplet (seed stock: protein sample: reservoir) for screening trials. Remaining seed stock was shock-frozen by liquid nitrogen and stored at -80 °C for further usage.

6.3.6.4 Optimization

Initial attempts for fine screening of crystallization conditions were performed in a 24-well protein crystallization plate, based on the hanging drop vapor diffusion method (**Figure 63B**). Therefore, screening conditions yielding promising crystals were prepared with varied pH and concentration of precipitants. In case of the IleRS_screen_1 condition, pH 6.2, 6.6 and 7.2 were adjusted in solution containing 16, 18, 20, and 22 % w/v PEG 8000, respectively. Each plate was used for crystallization of IleRS with TfeGly (12 wells) and without TfeGly. 600 µl of screening solution was added to the reservoir. The crystallization droplet was prepared with 1 µL protein (with or without the ligand) and 1 µL of reservoir solution. The mixture was pipetted on a siliconized glass plate and placed over the reservoir with the drop facing downwards. All plates were inspected immediately after preparation and stored at 19 °C. Droplet observation was performed weekly, and changes were noted in a scoring protocol.
Optimization experiments were carried out in 48-well protein crystallization plate using the sitting drop vapor diffusion method (**Figure 63A**). Particular screening conditions resulting promising hits were fine-tuned by varying pH value, salt, concentration of protein and concentration of precipitants. For this purpose, 0.5 µL protein (with or without the ligand) was mixed with 5 µL of reservoir solution and pipet to each well with corresponding reservoir. The plate was then sealed with a plastic film to avoid evaporation and stored at 19°C for drop-reservoir equilibration. After preparation, plates were immediately inspected under a light microscope. Plate observation was repeated every week until no further change in crystal size. Scoring protocol was prepared for conditions and changes in the crystallization drops. For illustrative purposes, pictures were taken from the promising variants. Prior to a diffraction test, crystals were harvested using a nylon fiber loop, shock-frozen and stored in liquid nitrogen.

6.3.7 Malachite green assay

The aminoacylation activity of certain AARS was determined by a malachite green assay. Basically, in the first step AARS activates the AA ATP-dependently, under formation of an aminoacyl-adenylate along with pyrophosphate (PP_i). Therefore, evaluation of the resulting

amount of PP_i provided the catalytic efficiency of the AARS in this initial step. In this context, the released PP_i was measured after addition of pyrophosphatase (PPiase) that decomposed PP_i into inorganic phosphate ions (P_i). The colorimetric phosphate assay kit (Abcam, Cambridge, England) was used, which contained a formulation of malachite green and ammonium molybdate. In reaction with phosphate ions a chromogenic complex was formed and detectable by absorption measurement at λ = 650 nm. A schematic presentation of the phosphate assay is given in the following **Figure 64**.

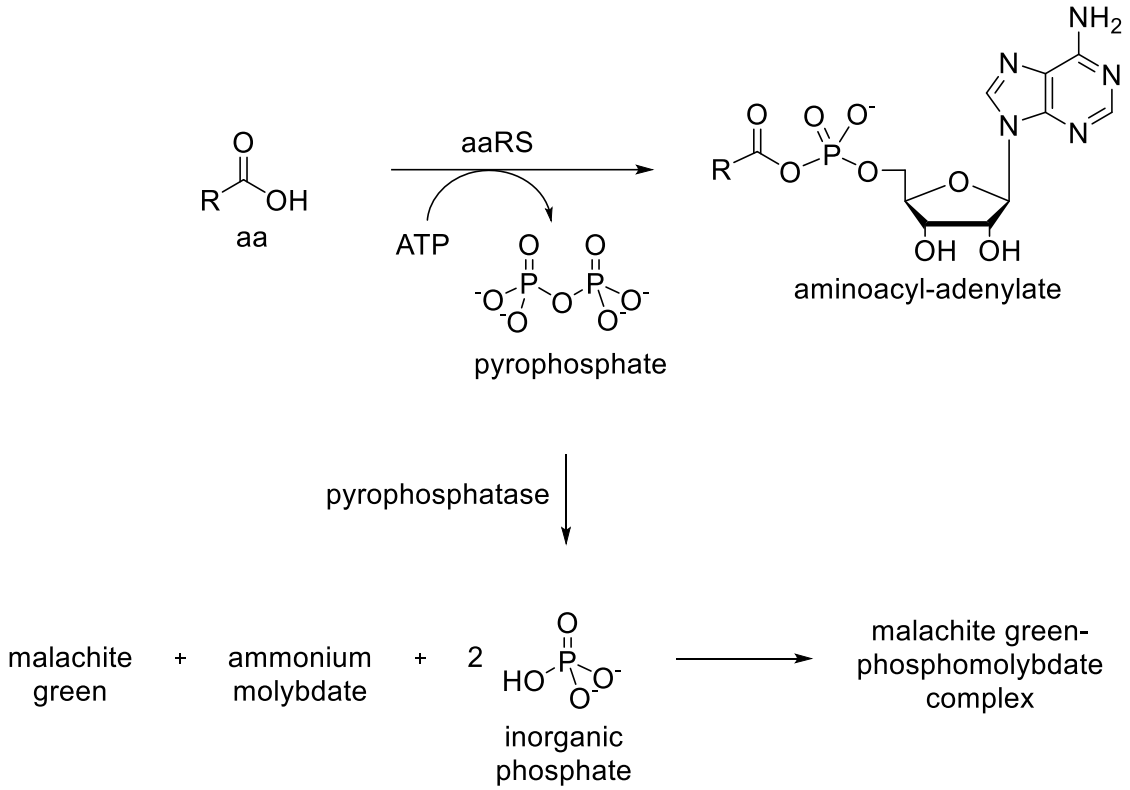


Figure 64 Aminoacylation-malachite green coupled assay according to the manufacturer's description.[341]

The thermostable PPiase from *Thermococcus litoralis* was purchased from New England Biolabs (Frankfurt am Main, Germany). Tested AARS were stored in 100 mM TRIS (pH 7.5 at rt), 100 mM KCl, 10 % w/v glycerol, 1 mM DTT to final concentrations of 1 µM, and 5 µM, respectively. All stock solutions used for this assay were freshly prepared. **Table 22** summarizes the composition of the reaction mixture.

Table 22 Reaction set up for malachite green activity assay.

compound	stock concentration	final concentration	volume in sample
TRIS pH 7.5 (rt)	1 M	100 mM	10 µL

$MgCl_2$	50 mM	5 mM	10 µL
ATP	100 mM	1 mM	1 µL
BSA	10 mg/mL	0.025 mg/mL	0.25 µL
PPiase	2000 u/mL	1 u/mL	1 µL
ncAA	40 mM	8 mM	20 µL
AARS	25 µM	1 or 5 µM	4 or 20 µL
nH_2O	-	-	fill up to 85 µL

A master mix without ncAA was prepared for multiple measurements. The reaction mixture was split into 96-well plate, 65 µL per well, and appropriate ncAA was then added to the wells. Finally, 15 µL of assay reagent (included in phosphate assay kit) was pipetted to each reaction. The plate was mixed and incubated at rt for 30 min protected from light. The output was measured at λ = 650 nm on the Microplate reader Infinite M200 (Perkin Elmer, Rodgau, Germany). A standard curve was prepared using phosphate standard solution (included in phosphate assay kit) according to the manufacturers' instruction. As a control, samples with cognate substrate and in absence of any AA was performed. All measurements were set up in duplicate.

6.4 Analytical methods

6.4.1 Concentration determination

6.4.1.1 DNA

DNA concentrations were determined in the elution buffer by absorbance measurements at 260 nm. The sample purity was assessed by the ratio of absorbance at 260 nm and 280 nm. DNA solution is deemed "pure" at a 260/280 value of 1.8. Lower ratios indicate the presence of contaminations, such as protein or phenol. Higher 260/280 values cannot be assigned to an issue. Another indication for impurity is illustrated by the 260/230 ratio in the range of 2.0-2.2. Low values might result from substances like ethanol from plasmid elution buffer, while higher ratios imply the inappropriateness of the blank solution.[342] Based on the consumption that at 260 nm the absorption value (A_{260}) of 1 equals to one unit of 50 µg/mL (U). DNA concentration (c (DNA)) was calculated using following formula:

$$c\,(DNA) = A_{260} \cdot U$$

6.4.1.2 Protein

6.4.1.2.1 Absorbance

For the determination of protein concentration, the theoretical molar extinction coefficient at 280 nm ($\varepsilon_{M,\ 280}$ in $M^{-1}\ cm^{-1}$) was calculated using the known AA sequence applied to the web-based software, ProtParam (Expasy Proteomics server, http://web.expasy.org/protparam). Protein absorbance at 280 nm (A_{280}) was measured and inserted together with the molecular weight (M_W in g/mol) and a cuvette pathlength (d) of 1 cm in the Lambert-Beer equation to obtain the protein concentration (c (protein) in mg/mL):

$$c\ (protein) = \frac{A_{280}}{\varepsilon_{M,280} \cdot d} \cdot M_W$$

Protein solutions having a 260/280 absorbance ratio of 0.6 are deemed pure. Higher values indicate contamination by e.g., DNA.

The concentration of fluorescent protein was additionally determined by the absorption wavelength of each chromophore (**Table 23**). Thus, the exact number of mature construct within the sample can be identified. The molar extinction coefficient at certain absorption maximum was taken from the published studies. Analogous to the previous determination by A_{280}, chromophore concentration was determined based on Lambert-Beer equation. The following table summarizes the values necessary for the calculation of protein concentration.

Table 23 Fluorescent proteins with corresponding characteristic values.

Protein	Wavelength (nm)	ε_M [$M^{-1}\ cm^{-1}$]	M_W [Da]
sfGFP	485	83 300[189]	29 190
EGFP	488	55 000[343]	27 745
NowGFP	493	53 600[190]	27 932
KillerOrange	514	22 600[320]	27 606

6.4.1.2.2 Bradford assay

The Bradford method was also applied for protein concentration determination. In an acidic environment, the dye Coomassie Brilliant Blue - G250 binds to the cationic, apolar hydrophobic side chain of proteins and forms a Coomassie G250-protein complex. As a consequence, the reddish-brown cationic dye undergoes a spectral shift to the blue anionic complex-form, which results in the absorption maximum change from 470 nm to 610 nm. At

595 nm the greatest difference between these two states is observed, making it the optimal wavelength for absorption measurements of the stable complex.[344] As the amount of formed complex is directly correlated with the protein concentration, this dye can be used as a rapid and sensitive method for protein quantification.

Five times concentrated Bradford reagent solution (Roti-Quant, **Table 8**), containing the G250 dye, was diluted 1:5 with ddH_2O and filtered (ø 0.22 µm) to remove solid particles. For the standard curve, 10 µL of each sample of the bovine serum albumin (BSA) dilution series (0, 0.10, 0.20, 0.50, 0.75, 1.00, 1.50 mg/mL) were prepared and added to 990 µL of the filtered Bradford solution. The mixture was incubated for 10 minutes in the dark to allow complete formation of the dye-protein complex. Subsequently, the specific absorption at 595 nm was measured by BioPhotometer plus (**Section 5.1**). The same procedures were implemented with protein sample of interest. The concentration was then determined by using the BSA standard curve.

6.4.2 Agarose gel electrophoresis

Double stranded DNA fragments of different sizes were separated by agarose gel electrophoresis. In general, 30-50 mL of an 0.5-1.5 % agarose gel were prepared in TAE buffer (**Section 5.5**), whereby separation of larger fragments, such as plasmid backbones (> 3 kb) in higher concentrated agarose gel provided better outcomes. For visualization of preparative gels, 0.5 µg/mL of the DNA-intercalating dye ethidiumbromide was added, while analytical gels were stained by 0.1-0.4 µL GelRed® Nucleic Acid Gel Stain (Biotium, Freemont, USA). Samples were mixed with 6x DNA loading buffer (Thermo Scientific, Darmstadt, Germany) prior to loading. The molecular size of analyzed fragments were estimated using the GeneRuler Mix Ladder (Thermo Scientific, Darmstadt, Germany). Gels were run at 80-100V for 30-90 min upon application. The separated DNAs were then visualized by illumination with UV light. Evaluation of the gels was performed by the gel documentation system Felix 2050 (Biostep, Burkhardsdorf, Germany). DNA fragments were extracted from preparative gel and purified by the GeneJET^{TM} Gel Extraction Kit for further usage.

6.4.3 Polyacrylamide gel electrophoresis

The analysis of protein expression level was performed by the polyacrylamide gel electrophoresis (PAGE). Soluble and insoluble fractions of protein expression as well as purified proteins were separated by SDS-PAGE using 10-15 % acrylamide gels with

PageRuler prestained or unstained Protein Ladder (Thermo Fisher Scientific, Waltham, USA) as a standard. The samples were initially mixed with 5x SDS loading dye (**Section 5.5**) and boiled at 95 °C for 10 min. Electrophoresis was performed in a vertical SDS-gel system (Factory of the Max-Planck Institute for Biochemistry, Martinsried, Germany) at room temperature between 90 V (stacking gel) and 120 V (running gel). Protein bands were finally visualized by 1 % w/v Coomassie Brilliant Blue R250 or silver staining.

6.4.4 Liquid chromatography-electrospray ionization mass spectroscopy (LC-ESI MS)

After external calibration, intact mass measurements of purified fluorescent proteins were performed by LC-ESI MS on an Agilent 6530 quadrupole time of flight (QTOF) instrument coupled with an Agilent 1260 HPLC system (Agilent technologies, Santa Clara, USA). At the system flow rate 0.3 mL/min 10 µL of 0.1 mg/mL concentrated sample was loaded on the attached C5 column, 3 micron (Supelco analytical, Sigma-Aldrich, Merck KGaA, Darmstadt, Germany). A gradient from 10 % to 80 % acetonitrile was run over 20 minutes. The obtained mass spectra were deconvoluted with maximum entropy by the Agilent MassHunter Qualitative Analysis software.

6.4.5 Ultraviolet-visible (UV-Vis) spectroscopy

UV-Vis absorption spectra of the intact fluorescent proteins in PBS buffer were recorded at 22 °C on a LAMBDA 950 UV/Vis/NIR spectrophotometer (Pelkin Elmer, Rodgau, Germany). Absorption data were monitored in the wavelength range of 250-600 nm using a 10 mm micro quartz cuvette with a pathlength of 1 cm (Hellma). The obtained data was analyzed by the UV-WinLab software and subsequently plotted by the OriginPro program.

6.4.6 Fluorescence spectroscopy

6.4.6.1 Fluorescence emission of protein variants

Fluorescence emission spectra of the intact fluorescent proteins in PBS buffer were recorded at 22 °C on a fluorescence spectrometer LS 55 (Pelkin Elmer, Rodgau, Germany). Protein concentrations were determined by both Bradford assay (**Section 6.4.1.2.2**) and calculation from the UV-absorption values with known molar extinction coefficient (ε_M) for particular protein (**Section 6.4.1.2.1**, **Table 24**). Excitation wavelengths applied for fluorescence

measurements using excitation/emission slits of 5.0 nm and scan speed of 100 nm min^{-1} are listed below.

Table 24 Applied excitation wavelengths in fluorescence measurements.

Protein variants	$\lambda_{excitation}$
EGFP	488 nm
NowGFP	493 nm
KillerOrange	510 nm

All fluorescence measurements were performed in Hellma™ Far UV quartz fluorescence cuvettes (10 mm pathlength) with 0.3 µM concentrated proteins. Spectra were recorded from 5 nm above the excitation wavelength to 600 nm in case of EGFP, NowGFP, and up to 700 nm for KillerOrange variants. For tryptophan-based emission determination 295 nm was used for excitation.

6.4.6.2 Denaturation and refolding of protein variants

30 µM of each purified fluorescent protein variant was denatured in PBS containing 8 M urea and 5 mM DTT for 5 min at 95 °C. Denatured sample was diluted 100-fold with PBS and 5 mM DTT to induce renaturation, yielding 0.3 µM final protein concentration. 200 µL of protein sample was transferred to a Hellma™ Far UV quartz fluorescence cuvette (10 mm pathlength) and the refolding kinetic of proteins was immediately monitored at $\lambda_{em.}$ = 330 nm when excited at $\lambda_{exc.}$ = 295 nm and $\lambda_{em.}$ = 509 nm by excitation at $\lambda_{exc.}$ = 488 nm. Over 30 min refolding process was followed by the 'Timedrive' option of the fluorescence spectrometer LS 55 (Pelkin Elmer, Rodgau, Germany) using an interval of 3 sec and a slit of 2.5 nm. The obtained data were subsequently analyzed by the OriginPro program.
After 24 h incubation of refolded samples, fluorescence emission spectra were collected using the same condition as for native proteins in **Section 6.4.6.1**.

6.4.7 Circular dichroism (CD) spectroscopy

In this work, far-UV CD spectra of IleRS mutants were recorded at the Jasco J-815 spectropolarimeter equipped with a Peltier-type FDCD accessory (model PTC 423S/15, Easton, USA) to allow temperature control during the measurement. 1 µM of each protein sample was prepared in 25 mM phosphate buffer (pH 8.0), containing 100 KCl. Data were

collected at 20 °C using 1 mm Hellma^{TM} 110-QS quartz cuvette and applying following parameters: 1 nm bandwidth, 0.1 nm intervals and a scan speed of 100 nm min^{-1}. Baselines of buffer were recorded by using the same conditions and subtracted from the presented data. Each reported CD spectra represents the average of four measurements.
Thermal unfolding experiments of IleRS mutants were conducted in the same cuvette, equipped with a stopper to avoid evaporation. The CD signal of 5 µM concentrated sample in previously used buffer was recorded at 222 nm in a temperature range of 20-98 °C with a rate of 3 °C per min. The obtained CD data were processed by the software Spectra Manger Version 2 from Jasco and subsequently analyzed by the OriginPro program.

6.5 Bioorthogonal reactions

6.5.1 Thiol-ene coupling reaction

10 – 20 µg of purified cfsfGFP(2Sac) was mixed with the reaction solution containing photo initiator. This mixture was added to a freshly prepared thiolated 8PEG hydrogel (8PEG-SH) which was covered with a piece of aluminum foil and incubated for 10 min in total darkness. Irradiation with UV light (366 nm) for 10 min enabled the radical photochemical reaction. The conjugated gel was subsequently washed with TRIS buffer for 30 min and prepared for the analysis by fluorescent microscopy. Optical images were captured by Zhaofei Ouyang (Lensen group) using a Carl Zeiss fluorescence microscope (Gottingen, Germany). For fluorescence microscopy, the Axio Observer Z1 was applied. Pictures were taken by an AxioCam MRm digital camera and analyzed by the AxioVisionV4.8.1 software package (Carl Zeiss, Goettingen, Germany).

6.5.2 Cleavage of the allyl group

Deprotection of the allylic handle from the thiol functionality was catalyzed by a water-soluble Pd^0 complex. In PBS buffer at physiological pH, cfsfGFP(2Sac) was incubated with 100 eq. of $Pd[TPPTS]_4$ overnight at 37 °C to allow complete cleavage. The reaction was quenched by adding an excess of DTT and incubated at 37 °C for 20 min, in order to remove the metal from the deprotected protein. Protein sample was finally analyzed by LC-ESI MS to verify the success of the deallylation. The deprotection reaction was performed by Sergej Schwagerus (FMP Berlin, AG Hackenberger).

7. References

1. Zhao, H. (ed.). *Protein Engineering. Tools and applications.* 1st ed. (Wiley-VCH, Weinheim,).
2. Li, C., Zhang, R., Wang, J., Wilson, L. M. & Yan, Y. Protein Engineering for Improving and Diversifying Natural Product Biosynthesis. *Trends in biotechnology* **38,** 729–744; 10.1016/j.tibtech.2019.12.008 (2020).
3. Johannes, T. W. & Zhao, H. Directed evolution of enzymes and biosynthetic pathways. *Current opinion in microbiology* **9,** 261–267; 10.1016/j.mib.2006.03.003 (2006).
4. Eijsink, V. G. H., Gåseidnes, S., Borchert, T. V. & van den Burg, B. Directed evolution of enzyme stability. *Biomolecular engineering* **22,** 21–30; 10.1016/j.bioeng.2004.12.003 (2005).
5. To, K. K. W. & Cho, W. C. S. An overview of rational design of mRNA-based therapeutics and vaccines. *Expert opinion on drug discovery* **16,** 1307–1317; 10.1080/17460441.2021.1935859 (2021).
6. Woolfson, D. N. A Brief History of De Novo Protein Design: Minimal, Rational, and Computational. *Journal of Molecular Biology* **433,** 167160; 10.1016/j.jmb.2021.167160 (2021).
7. Salwiczek, M., Nyakatura, E. K., Gerling, U. I. M., Ye, S. & Koksch, B. Fluorinated amino acids: compatibility with native protein structures and effects on protein-protein interactions. *Chemical Society reviews* **41,** 2135–2171; 10.1039/c1cs15241f (2012).
8. Lipinszki, Z. *et al.* Enhancing the Translational Capacity of E. coli by Resolving the Codon Bias. *ACS synthetic biology,* 2656–2664; 10.1021/acssynbio.8b00332 (2018).
9. Crick, F. H. C. On protein synthesis. *Symposia of the Society for Experimental Biology,* 138–163 (1958).
10. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America,* 4576–4579; 10.1073/pnas.87.12.4576 (1990).
11. Erdmann, V. A. & Barciszewski, J. 2011: 50th anniversary of the discovery of the genetic code. *Angewandte Chemie (International ed. in English),* 9546–9552; 10.1002/anie.201103895 (2011).
12. Tanford, C., Reynolds, J. *Nature's robots: A history of proteins* (Oxford University Press, Oxford, 2001).
13. Hubert Bradford. Vickery, Carl L. A. Schmidt. The History of the Discovery of the Amino Acids. *Chemical Reviews,* 169–318 (1931).
14. Miescher, F. Über die chemische Zusammensetzung der Eiterzellen. *Medicinisch-chemische Untersuchungen,* 441–460 (1871).
15. Watson, J. D. and Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature,* 737–738 (1953).
16. Marshall W. Nirenberg and J. Heinrich Matthaei. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides, 1588–1602 (1961).
17. Nirenberg, M. *et al.* The RNA code and protein synthesis. *Cold Spring Harbor symposia on quantitative biology,* 11–24; 10.1101/sqb.1966.031.01.008 (1966).
18. Maloy, S. R. & Stewart, V. J. and Taylor, R. K. *Genetic analysis of pathogenic bacteria. A laboratory manual* (Cold Spring Harbor Laboratory Press, New York, 1996).
19. Jones Jr., W. O. and Nirenberg, M. W. Degeneracy in the amino acid code. *Biochimica et biophysica acta,* 400–406 (1966).
20. Berg, J. M., Tymoczko, J. L. & Stryer, L. *Biochemistry.* 7th ed. (Palgrave Macmillan, Basingstoke, 2011).
21. Quax, T. E. F., Claassens, N. J., Söll, D. & van der Oost, J. Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular cell,* 149–161; 10.1016/j.molcel.2015.05.035 (2015).
22. Wang, Y. *et al.* An Engineered Rare Codon Device for Optimization of Metabolic Pathways. *Scientific Reports,* 1–11 (2016).

23. Guimaraes, J. C. *et al.* A rare codon-based translational program of cell proliferation. *Genome Biology*; 10.1186/s13059-020-1943-5 (2020).

24. Angov, E. Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnology journal,* 650–659; 10.1002/biot.201000332 (2011).

25. A. Böck, K. Forchhammer, J. Heider, W. Leinfelder, G. Sawers, B. Veprek and F. Zinoni. Selenocysteine: the 21st amino acid. *Molecular Microbiology,* 515–520 (1991).

26. Hao, B. *et al.* Reactivity and chemical synthesis of L-pyrrolysine- the 22(nd) genetically encoded amino acid. *Chemistry & biology,* 1317–1324; 10.1016/j.chembiol.2004.07.011 (2004).

27. Rother, M. & Krzycki, J. A. Selenocysteine, pyrrolysine, and the unique energy metabolism of methanogenic archaea. *Archaea (Vancouver, B.C.)*; 10.1155/2010/453642 (2010).

28. Wiltschi, B. & Budisa, N. Natural history and experimental evolution of the genetic code. *Applied microbiology and biotechnology* **74,** 739–753; 10.1007/s00253-006-0823-6 (2007).

29. Voet, D., Voet, J. G. & Pratt, C. W. *Fundamentals of biochemistry. Life at the molecular level* (Wiley, Hoboken, NJ, 2016).

30. Tong, L. Structure and function of biotin-dependent carboxylases. *Cellular and molecular life sciences : CMLS* **70,** 863–891; 10.1007/s00018-012-1096-0 (2013).

31. Beckett, D. Specificity and selectivity in post-translational biotin addition. *Biochemical Society transactions,* 1577–1591; 10.1042/BST20180425 (2018).

32. Lu, Y. and Freeland, S. On the evolution of the standard amino-acid alphabet. *Genome Biology,* 102 (2006).

33. Gutiérrez-Preciado, A. & Romero, H. and Peimbert, M. An Evolutionary Perspective on Amino Acids. *Nature Education,* 29 (2010).

34. Lopez, M. J. & Mohiuddin, S. S. *Biochemistry, Essential Amino Acids* (StatPearls Publishing, Treasure Island (FL), 2021).

35. Lu, Y. & Freeland, S. On the evolution of the standard amino-acid alphabet. *Genome Biology* **7** (2006).

36. Almhjell, P. J., Boville, C. E. & Arnold, F. H. Engineering enzymes for noncanonical amino acid synthesis. *Chemical Society reviews* **47,** 8980–8997; 10.1039/c8cs00665b (2018).

37. Link, A. J., Mock, M. L. & Tirrell, D. A. Non-canonical amino acids in protein engineering. *Current opinion in biotechnology* **14,** 603–609; 10.1016/j.copbio.2003.10.011 (2003).

38. Jäckel, C. & Koksch, B. Fluorine in Peptide Design and Protein Engineering. *Eur. J. Org. Chem.* **2005,** 4483–4503; 10.1002/ejoc.200500205 (2005).

39. Agostini, F. *et al.* Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angewandte Chemie (International ed. in English)* **56,** 9680–9703; 10.1002/anie.201610129 (2017).

40. Inokuma, T. Synthesis of Non-canonical Amino Acids and Peptide Containing Them for Establishment of the Template for Drug Discovery. *Chemical & pharmaceutical bulletin* **69,** 303–313; 10.1248/cpb.c21-00031 (2021).

41. Luo, X., Wang, T.-S. A., Zhang, Y., Wang, F. & Schultz, P. G. Stabilizing Protein Motifs with a Genetically Encoded Metal-Ion Chelator. *Cell chemical biology* **23,** 1098–1102; 10.1016/j.chembiol.2016.08.007 (2016).

42. Rezhdo, A., Islam, M., Huang, M. & van Deventer, J. A. Future prospects for noncanonical amino acids in biological therapeutics. *Current opinion in biotechnology* **60,** 168–178; 10.1016/j.copbio.2019.02.020 (2019).

43. Ding, Y. *et al.* Impact of non-proteinogenic amino acids in the discovery and development of peptide therapeutics. *Amino acids* **52,** 1207–1226; 10.1007/s00726-020-02890-9 (2020).

44. Devaraj, N. K. The Future of Bioorthogonal Chemistry. *ACS central science* **4,** 952–959; 10.1021/acscentsci.8b00251 (2018).

45. Kharkar, P. M., Rehmann, M. S., Skeens, K. M., Maverakis, E. & Kloxin, A. M. Thiol-ene click hydrogels for therapeutic delivery. *ACS biomaterials science & engineering* **2,** 165–179; 10.1021/acsbiomaterials.5b00420 (2016).

46. Ojima, I. (ed.). *Fluorine in Medicinal Chemistry and Chemical Biology* (Wiley-Blackwell, Chichester, 2009).

47. Shah, P. & Westwell, A. D. The role of fluorine in medicinal chemistry. *Journal of enzyme inhibition and medicinal chemistry* **22,** 527–540; 10.1080/14756360701425014 (2007).

48. McGrath, N. A., Brichacek, M. & Njardarson, J. T. A Graphical Journey of Innovative Organic Architectures That Have Improved Our Lives. *J. Chem. Educ.* **87,** 1348–1349; 10.1021/ed1003806 (2010).

49. Bondi, A. van der Waals Volumes and Radii. *The Journal of physical chemistry* **68,** 441–451; 10.1021/j100785a001 (1964).

50. Bauer, M. R. *et al.* Harnessing Fluorine-Sulfur Contacts and Multipolar Interactions for the Design of p53 Mutant Y220C Rescue Drugs. *ACS chemical biology* **11,** 2265–2274; 10.1021/acschembio.6b00315 (2016).

51. Smart, B. E. Fluorine substituent effects (on bioactivity). *Journal of Fluorine Chemistry* **109,** 3–11; 10.1016/S0022-1139(01)00375-X (2001).

52. Robalo, J. R., Huhmann, S., Koksch, B. & Vila Verde, A. The Multiple Origins of the Hydrophobicity of Fluorinated Apolar Amino Acids. *Chem* **3,** 881–897; 10.1016/j.chempr.2017.09.012 (2017).

53. Marsh, E. N. G. Fluorinated proteins: from design and synthesis to structure and stability. *Accounts of Chemical Research* **47,** 2878–2886; 10.1021/ar500125m (2014).

54. Jäckel, C., Salwiczek, M. & Koksch, B. Fluorine in a native protein environment--How the spatial demand and polarity of fluoroalkyl groups affect protein folding. *Angewandte Chemie (International ed. in English)* **45,** 4198–4203; 10.1002/anie.200504387 (2006).

55. Berger, A. A., Völler, J.-S., Budisa, N. & Koksch, B. Deciphering the Fluorine Code-The Many Hats Fluorine Wears in a Protein Environment. *Accounts of Chemical Research* **50,** 2093–2103; 10.1021/acs.accounts.7b00226 (2017).

56. Huhmann, S. & Koksch, B. Fine-Tuning the Proteolytic Stability of Peptides with Fluorinated Amino Acids. *Eur. J. Org. Chem.* **2018,** 3667–3679; 10.1002/ejoc.201800803 (2018).

57. Li, D. *et al.* Protein-protein interaction analysis in crude bacterial lysates using combinational method of 19F site-specific incorporation and 19F NMR. *Protein & cell* **8,** 149–154; 10.1007/s13238-016-0336-8 (2017).

58. Danielson, M. A. & Falke, J. J. Use of 19F NMR to probe protein structure and conformational changes. *Annual review of biophysics and biomolecular structure* **25,** 163–195; 10.1146/annurev.bb.25.060196.001115 (1996).

59. Buchholz, C. R. & Pomerantz, W. C. K. 19F NMR viewed through two different lenses: ligand-observed and protein-observed 19F NMR applications for fragment-based drug discovery. *RSC chemical biology* **2,** 1312–1330; 10.1039/d1cb00085c (2021).

60. Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. Merrifield, R. B. (1963). Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. Journal of the American Chemical Society, 85(14), 2149–2154. doi:10.1021/ja00897a025 . *Journal of the American Chemical Society,* 2149–2154; 10.1021/ja00897a025 (1963).

61. Collins, J. M., Singh, S. K. & Vanier, G. S. Microwave technology for solid phase peptide synthesis. It's not just for difficult peptides. *Chimica Oggi - Chemistry Today,* 26–29 (2012).

62. Varela, Y. F., Vanegas Murcia, M. & Patarroyo, M. E. Synthetic Evaluation of Standard and Microwave-Assisted Solid Phase Peptide Synthesis of a Long Chimeric Peptide Derived from Four Plasmodium falciparum Proteins. *Molecules (Basel, Switzerland)* **23**; 10.3390/molecules23112877 (2018).

63. Kulkarni, S. S., Sayers, J., Premdjee, B. & Payne, R. J. Rapid and efficient protein synthesis through expansion of the native chemical ligation concept. *Nat Rev Chem* **2**; 10.1038/s41570-018-0122 (2018).

64. Garenne, D. *et al.* Cell-free gene expression. *Nat Rev Methods Primers,* 1–18; 10.1038/s43586-021-00046-x (2021).

65. Robertson, S. A., Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. and Schultz, P. G. The use of 5'-phospho-2-deoxyribocytidylylriboadenosine as a facile route to chemical aminoacylation of tRNA. *Nucleic acids research,* 9649–9660 (1989).

66. Hecht, S. M., Alford, B. L., Kuroda, Y. and Kitano, S. "Chemical aminoacylation" of tRNA's. *The Journal of biological chemistry,* 4517–4520 (1978).

67. Ogle, J. M. & Ramakrishnan, V. Structural insights into translational fidelity. *Annual review of biochemistry,* 129–177; 10.1146/annurev.biochem.74.061903.155440 (2005).

68. Gromadski, K. B., Daviter, T. & Rodnina, M. V. A uniform response to mismatches in codon-anticodon complexes ensures ribosomal fidelity. *Molecular cell* **21,** 369–377; 10.1016/j.molcel.2005.12.018 (2006).

69. Giegé, R. and Frugier, M. Transfer RNA Structure and Identity. Available at https://www.ncbi.nlm.nih.gov/books/NBK6236/ (2000-2013).

70. Shareghi, P., Wang, Y., Malmberg, R. & Cai, L. Simultaneous prediction of RNA secondary structure and helix coaxial stacking. *BMC genomics,* 1-11; 10.1186/1471-2164-13-S3-S7 (2012).

71. Dirheimer, G., Keith, G., Dumas, P. & Westhof, E. Primary, Secondary, and Tertiary Structures of tRNAs. In *tRNA: Structure, Biosynthesis, and Function,* edited by D. Söll & U. L. RajBhandary (ASM Press, Washington, DC, USA, 1995), pp. 93–126.

72. Shepherd, J. & Ibba, M. Bacterial transfer RNAs. *FEMS microbiology reviews,* 280–300; 10.1093/femsre/fuv004 (2015).

73. Boccaletto, P. *et al.* MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic acids research,* D303-D307; 10.1093/nar/gkx1030 (2018).

74. Alberts, B. *Molecular biology of the cell.* 4th ed. (Garland Science, New York, 2002).

75. Crick, F. H. C. Codon-anticodon pairing: The wobble hypothesis. *Journal of Molecular Biology,* 548–555 (1966).

76. Chan, P. P. & Lowe, T. M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic acids research,* D184-9; 10.1093/nar/gkv1309 (2016).

77. Giegé, R., Sissler, M. & Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic acids research,* 5017–5035; 10.1093/nar/26.22.5017 (1998).

78. Dickson, L. A. and Schimmel, P. R. Structure of Transfer RNA-Aminoacyl Transfer RNA Synthetase Complexes Investgated By Nuclease Digestion. *Archives of Biochemistry and Biophysics,* 638–645 (1975).

79. Schimmel, P. R. Approaches to understanding the mechanism of specific protein-transfer RNA interactions. *Accounts of Chemical Research,* 411–418 (1977).

80. Rogers, M. J., Adachi, T. & Inokuchi, H. and Söll, D. Functional communication in the recognition of tRNA by Escherichia coli glutaminyl-tRNA synthetase. *Proceedings of the National Academy of Sciences of the United States of America,* 291–295 (1994).

81. Borel, F., Vincent, C. & Leberman, R. and Härtlein, M. Seryl-tRNA synthetase from Escherichia coli: implication of its N-terminal domain in aminoacylation activity and specificity., 2963–2969 (1994).

82. Shepherd, J. & Ibba, M. Relaxed substrate specificity leads to extensive tRNA mischarging by Streptococcus pneumoniae class I and class II aminoacyl-tRNA synthetases. *mBio,* e01656-14; 10.1128/mBio.01656-14 (2014).

83. Hale, S. P., Auld, D. S., Schmidt, E. & Schimmel, P. Discrete determinants in transfer RNA for editing and aminoacylation. *Science (New York, N.Y.),* 1250–1252; 10.1126/science.276.5316.1250 (1997).

84. Dudock, B., DiPeri, C., Scileppi, K. & Reszelbach, R. The yeast phenylalanyl-transfer RNA synthetase recognition site: the region adjacent to the dihydrouridine loop. *Proceedings of the National Academy of Sciences of the United States of America,* 681–684; 10.1073/pnas.68.3.681 (1971).

85. Ko, Y. G. *et al.* Glutamine-dependent antiapoptotic interaction of human glutaminyl-tRNA synthetase with apoptosis signal-regulating kinase 1. *The Journal of biological chemistry,* 6030–6036; 10.1074/jbc.M006189200 (2001).

86. Javanbakht, H. *et al.* The interaction between HIV-1 Gag and human lysyl-tRNA synthetase during viral assembly. *The Journal of biological chemistry,* 27644–27651; 10.1074/jbc.M301840200 (2003).

87. Park, H., Park, S. G., Kim, J., Ko, Y.-G. & Kim, S. Signaling pathways for TNF production induced by human aminoacyl-tRNA synthetase-associating factor, p43. *Cytokine,* 148–153; 10.1006/cyto.2002.1992 (2002).

88. Akins, R. A. and Lambowitz A. M. A protein required for splicing group I introns in Neurospora mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof. *Cell,* 331–345 (1987).

89. Rubio Gomez, M. A. and Ibba, M. Aminoacyl-tRNA synthetases. *RNA,* 910–936; 10.1261/rna.071720.119 (2020).

90. Mirande, M. The Aminoacyl-tRNA Synthetase Complex. *Sub-cellular biochemistry,* 505–522; 10.1007/978-3-319-46503-6_18 (2017).

91. Burbaum, J. J. & Schimmel, P. Structural relationships and the classification of aminoacyl-tRNA synthetases. *Journal of Biological Chemistry,* 16965–16968; 10.1016/S0021-9258(19)47323-7 (1991).

92. Tumbula, D. *et al.* Archaeal Aminoacyl-tRNA Synthesis: Diversity Replaces Dogma. *Genetics,* 1269–1276; 10.1093/genetics/152.4.1269 (1999).

93. Bult, C. J. *et al.* Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science,* 1058–1073 (1996).

94. Allmang, C. & Krol, A. Selenoprotein synthesis: UGA does not end the story. *Biochimie,* 1561–1571; 10.1016/j.biochi.2006.04.015 (2006).

95. Forchhammer, K., Boesmiller, K. & Böck, A. The function of selenocysteine synthase and SELB in the synthesis and incorporation of selenocysteine. *Biochimie,* 1481–1486; 10.1016/0300-9084(91)90181-Y (1991).

96. Zhang, C.-M., Christian, T., Newberry, K. J., Perona, J. J. & Hou, Y.-M. Zinc-mediated Amino Acid Discrimination in Cysteinyl-tRNA Synthetase. *Journal of Molecular Biology,* 911–917; 10.1016/S0022-2836(03)00241-9 (2003).

97. Sankaranarayanan, R. *et al.* Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase. *Nature structural biology,* 461–465 (2000).

98. Valencia-Sánchez, M. I. *et al.* Structural Insights into the Polyphyletic Origins of Glycyl tRNA Synthetases. *The Journal of biological chemistry,* 14430–14446; 10.1074/jbc.M116.730382 (2016).

99. Fukai, S. *et al.* Structural Basis for Double-Sieve Discrimination of L-Valine from L-Isoleucine and L-Threonine by the Complex of tRNAVal and Valyl-tRNA Synthetase. *Cell,* 793–803; 10.1016/S0092-8674(00)00182-3 (2000).

100. Lincecum, T. L. *et al.* Structural and Mechanistic Basis of Pre- and Posttransfer Editing by Leucyl-tRNA Synthetase. *Molecular cell,* 951–963; 10.1016/S1097-2765(03)00098-4 (2003).

101. Fersht, A. R. Editing mechanisms in protein synthesis. Rejection of valine by the isoleucyl-tRNA synthetase. *Biochemistry,* 1025–1030 (1977).

102. Martinis, S. A. & Boniecki, M. T. The balance between pre- and post-transfer editing in tRNA synthetases. *FEBS letters,* 455–459; 10.1016/j.febslet.2009.11.071 (2010).

103. Tawfik, D. S. & Gruic-Sovulj, I. How evolution shapes enzyme selectivity - lessons from aminoacyl-tRNA synthetases and other amino acid utilizing enzymes. *The FEBS journal,* 1284–1305; 10.1111/febs.15199 (2020).

104. Dock-Bregeon, A.-C. *et al.* Transfer RNA–Mediated Editing in Threonyl-tRNA Synthetase. *Cell,* 877–884; 10.1016/s0092-8674(00)00191-4 (2000).

105. Palencia, A. *et al.* Structural dynamics of the aminoacylation and proofreading functional cycle of bacterial leucyl-tRNA synthetase. *Nature structural & molecular biology,* 677–684; 10.1038/nsmb.2317 (2012).

106. Silvian, L. F., Wang, J. & Steitz, T. A. Insights into editing from an ile-tRNA synthetase structure with tRNAile and mupirocin. *Science (New York, N.Y.),* 1074–1077 (1999).

107. Foo, J. L., Ching, C. B., Chang, M. W. & Leong, S. S. J. The imminent role of protein engineering in synthetic biology. *Biotechnology advances,* 541–549; 10.1016/j.biotechadv.2011.09.008 (2012).

108. Chen, Z. & Zeng, A.-P. Protein engineering approaches to chemical biotechnology. *Current opinion in biotechnology,* 198–205; 10.1016/j.copbio.2016.07.007 (2016).

109. Glasscock, C. J., Lucks, J. B. & DeLisa, M. P. Engineered Protein Machines: Emergent Tools for Synthetic Biology. *Cell chemical biology,* 45–56; 10.1016/j.chembiol.2015.12.004 (2016).

110. Li, C., Zhang, R., Wang, J., Wilson, L. M. & Yan, Y. Protein Engineering for Improving and Diversifying Natural Product Biosynthesis. *Trends in biotechnology,* 729–744; 10.1016/j.tibtech.2019.12.008 (2020).

111. Mau Goh, K., Poh Hong, G., @ Pearly Ng, N. H. C., Kian Piaw, C. & Raja Abdul Rahman, R. N. Z. Trends and Tips in Protein Engineering, A Review. *Jurnal Teknologi,* 21–31; 10.11113/jt.v59.1574 (2013).

112. Budisa, N. Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire. *Angewandte Chemie (International ed. in English),* 6426–6463; 10.1002/anie.200300646 (2004).

113. Artner, L. M. *et al.* Site-selective modification of proteins for the synthesis of structurally defined multivalent scaffolds. *Chemical communications (Cambridge, England)* **48,** 522–524; 10.1039/c1cc16039g (2012).

114. Mamidyala, S. K. & Finn, M. G. In situ click chemistry: probing the binding landscapes of biological molecules. *Chemical Society reviews* **39,** 1252–1261; 10.1039/b901969n (2010).

115. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* **40,** 2004–2021; 10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5 (2001).

116. Budisa, N. *Engineering the genetic code. Expanding the amino acid repertoire for the design of novel proteins* (Wiley-VCH, Weinheim, 2006).

117. Benner, S. A. Expanding the genetic lexicon: Incorporating non-standard amino acids into proteins by ribosome-based synthesis. *Trends in biotechnology,* 158–163; 10.1016/0167-7799(94)90076-0 (1994).

118. Hoesl, M. G. *et al.* Lipase Congeners Designed by Genetic Code Engineering. *ChemCatChem* **3,** 213–221; 10.1002/cctc.201000253 (2011).

119. Moroder, L. & Budisa, N. Synthetic biology of protein folding. *Chemphyschem : a European journal of chemical physics and physical chemistry* **11,** 1181–1187; 10.1002/cphc.201000035 (2010).

120. Cirino, P. C., Tang, Y., Takahashi, K., Tirrell, D. A. & Arnold, F. H. Global incorporation of norleucine in place of methionine in cytochrome P450 BM-3 heme domain increases peroxygenase activity. *Biotechnology and bioengineering* **83,** 729–734; 10.1002/bit.10718 (2003).

121. Budisa, N. *et al.* Global Replacement of Tryptophan with Aminotryptophans Generates Non-Invasive Protein-Based Optical pH Sensors. *Angew. Chem. Int. Ed.* **41,** 4066–4069 (2002).

122. Hösl, M. G. Expanding the Toolkit of Protein Engineering. Towards Multiple Simultaneous In Vivo Incorporation ofNoncanonical Amino Acids. PhD. Technischen Universität München, 2011.

123. Sharma, N., Furter, R., Kast, P. & Tirrell, D. A. Efficient introduction of aryl bromide functionality into proteins in vivo. *FEBS letters,* 37–40; 10.1016/S0014-5793(00)01120-0 (2000).

124. Döring, V. *et al.* Enlarging the amino acid set of Escherichia coli by infiltration of the valine coding pathway. *Science (New York, N.Y.),* 501–504; 10.1126/science.1057718 (2001).

125. Kiick, K. L., van Hest, J. C. M. & Tirrell, D. A. Expanding the Scope of Protein Biosynthesis by Altering the Methionyl-tRNA Synthetase Activity of a Bacterial Expression Host. *Angewandte Chemie (International ed. in English),* 2148–2152; 10.1002/1521-3773(20000616)39:12<2148::AID-ANIE2148>3.0.CO;2-7 (2000).

126. Smolskaya, S. & Andreev, Y. A. Site-Specific Incorporation of Unnatural Amino Acids into Escherichia coli Recombinant Protein: Methodology Development and Recent Achievement. *Biomolecules*; 10.3390/biom9070255 (2019).

127. Xie, J. and Schultz, P. G. An expanding genetic code. *Methods,* 227–238; 10.1016/j.ymeth.2005.04.010 (2005).

128. Maranhao, A. C. & Ellington, A. D. Evolving Orthogonal Suppressor tRNAs To Incorporate Modified Amino Acids. *ACS synthetic biology,* 108–119; 10.1021/acssynbio.6b00145 (2017).

129. Nozawa, K. *et al.* Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. *Nature,* 1163–1167; 10.1038/nature07611 (2009).

130. Blight, S. K. *et al.* Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature,* 333–335; 10.1038/nature02895 (2004).

131. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science (New York, N.Y.),* 498–500; 10.1126/science.1060077 (2001).

132. Hadd, A. & Perona, J. J. Recoding aminoacyl-tRNA synthetases for synthetic biology by rational protein-RNA engineering. *ACS chemical biology* **9,** 2761–2766; 10.1021/cb5006596 (2014).

133. Wang, L. & Schultz, P. G. A general approach for the generation of orthogonal tRNAs. *Chemistry & biology* **8,** 883–890 (2001).

134. Johnson, D. B. F. *et al.* RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nature chemical biology* **7,** 779–786; 10.1038/nchembio.657 (2011).

135. Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature biotechnology* **25,** 770–777; 10.1038/nbt1314 (2007).

136. Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science (New York, N.Y.)* **342,** 357–360; 10.1126/science.1241459 (2013).

137. Wang, N., Ju, T., Niu, W. & Guo, J. Fine-tuning interaction between aminoacyl-tRNA synthetase and tRNA for efficient synthesis of proteins containing unnatural amino acids. *ACS synthetic biology* **4,** 207–212; 10.1021/sb500195w (2015).

138. Willis, J. C. W. & Chin, J. W. Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nature chemistry,* 831–837; 10.1038/s41557-018-0052-5 (2018).

139. Exner, M. P. *et al.* Design of S-Allylcysteine in Situ Production and Incorporation Based on a Novel Pyrrolysyl-tRNA Synthetase Variant. *Chembiochem : a European journal of chemical biology,* 85–90; 10.1002/cbic.201600537 (2017).

140. Hauf, M. *et al.* Photoactivatable Mussel-Based Underwater Adhesive Proteins by an Expanded Genetic Code. *Chembiochem : a European journal of chemical biology,* 1819–1823; 10.1002/cbic.201700327 (2017).

141. Vargas-Rodriguez, O., Sevostyanova, A., Söll, D. & Crnković, A. Upgrading aminoacyl-tRNA synthetases for genetic code expansion. *Current opinion in chemical biology* **46,** 115–122; 10.1016/j.cbpa.2018.07.014 (2018).

142. Link, A. J. & Tirrell, D. A. Reassignment of sense codons in vivo. *Methods (San Diego, Calif.)* **36,** 291–298; 10.1016/j.ymeth.2005.04.005 (2005).

143. Biddle, W., Schmitt, M. A. & Fisk, J. D. Evaluating Sense Codon Reassignment with a Simple Fluorescence Screen. *Biochemistry* **54,** 7355–7364; 10.1021/acs.biochem.5b00870 (2015).

144. Krishnakumar, R. *et al.* Transfer RNA misidentification scrambles sense codon recoding. *Chembiochem : a European journal of chemical biology* **14,** 1967–1972; 10.1002/cbic.201300444 (2013).

145. Bohlke, N. & Budisa, N. Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: rare isoleucine codon AUA as a target for genetic code expansion. *FEMS microbiology letters* **351,** 133–144; 10.1111/1574-6968.12371 (2014).

146. Acevedo-Rocha, C. G. & Budisa, N. Xenomicrobiology: a roadmap for genetic code engineering. *Microbial biotechnology* **9,** 666–676; 10.1111/1751-7915.12398 (2016).

147. Robertson, W. E. *et al.* Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science (New York, N.Y.)* **372,** 1057–1062; 10.1126/science.abg3029 (2021).

148. Harada, F. & Nishimura, S. Purification and characterization of AUA specific isoleucine transfer ribonucleic acid from Escherichia coli B. *Biochemistry* **13,** 300–307 (1974).

149. Ikeuchi, Y. *et al.* Molecular mechanism of lysidine synthesis that determines tRNA identity and codon recognition. *Molecular cell* **19,** 235–246; 10.1016/j.molcel.2005.06.007 (2005).

150. Suzuki, T. & Miyauchi, K. Discovery and characterization of tRNAIle lysidine synthetase (TilS). *FEBS letters* **584,** 272–277; 10.1016/j.febslet.2009.11.085 (2010).

151. Soma, A. *et al.* An RNA-Modifying Enzyme that Governs Both the Codon and Amino Acid Specificities of Isoleucine tRNA. *Molecular cell* **12,** 689–698 (2003).

152. Nakanishi, K. *et al.* Structural basis for translational fidelity ensured by transfer RNA lysidine synthetase. *Nature* **461,** 1144–1148; 10.1038/nature08474 (2009).

153. Taniguchi, T. *et al.* Decoding system for the AUA codon by tRNAIle with the UAU anticodon in Mycoplasma mobile. *Nucleic acids research* **41,** 2621–2631; 10.1093/nar/gks1344 (2013).

154. Silva, F. J., Belda, E. & Talens, S. E. Differential annotation of tRNA genes with anticodon CAT in bacterial genomes. *Nucleic acids research* **34,** 6015–6022; 10.1093/nar/gkl739 (2006).

155. Muramatsu, T. *et al.* Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature* **336,** 179–181 (1988).

156. Lajoie, M. J. *et al.* Probing the limits of genetic recoding in essential genes. *Science (New York, N.Y.)* **342,** 361–363; 10.1126/science.1241460 (2013).

157. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annual review of biochemistry* **79,** 413–444; 10.1146/annurev.biochem.052308.105824 (2010).

158. Kwon, I., Kirshenbaum, K. & Tirrell, D. A. Breaking the degeneracy of the genetic code. *Journal of the American Chemical Society* **125,** 7512–7513; 10.1021/ja0350076 (2003).

159. Kwon, I. & Lim, S. in. Tailoring the substrate specificity of yeast phenylalanyl-tRNA synthetase toward a phenylalanine analog using multiple-site-specific incorporation. *ACS synthetic biology* **4,** 634–643; 10.1021/sb500309r (2015).

160. Summerer, D. *et al.* Genetically Encoded Fluorescent Amino Acid. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 9785–9789; 10.1021/ja062666k.s001 (2006).

161. Italia, J. S. *et al.* An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nature chemical biology* **13,** 446–450; 10.1038/nchembio.2312 (2017).

162. Mayer-Kuckuk, P., Menon, L. G., Blasberg, R. G., Bertino, J. R. & Banerjee, D. Role of reporter gene imaging in molecular and cellular biology. *Biological chemistry* **385,** 353–361; 10.1515/BC.2004.039 (2004).

163. Serganova, I. & Blasberg, R. G. Molecular Imaging with Reporter Genes: Has Its Promise Been Delivered? *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* **60,** 1665–1681; 10.2967/jnumed.118.220004 (2019).

164. Zimmer, M. Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chemical Reviews* **102,** 759–781; 10.1021/cr010142r (2002).

165. Pakhomov, A. A. & Martynov, V. I. GFP family: structural insights into spectral tuning. *Chemistry & biology* **15,** 755–764; 10.1016/j.chembiol.2008.07.009 (2008).

166. Brooks, S. The discovery of aequorin and green fluorescent protein. *J Microsc* **217,** 1–2; 10.1111/j.0022-2720.2005.01446.x (2005).

167. SHIMOMURA, O., JOHNSON, F. H. & SAIGA, Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *Journal of cellular and comparative physiology* **59,** 223–239; 10.1002/jcp.1030590302 (1962).

168. Akbar, M. & Kim, H. Y. Green fluorescent protein tagging: A novel tool in biomedical research. *Indian Journal of Biotechnology* **4,** 466–470 (2005).

169. Born, J. & Pfeifer, F. Improved GFP Variants to Study Gene Expression in Haloarchaea. *Frontiers in microbiology* **10,** 1200; 10.3389/fmicb.2019.01200 (2019).

170. Hanson, M. R. & Köhler, R. H. GFP imaging: methodology and application to investigate cellular compartmentation in plants. *Journal of Experimental Botany* **52,** 529–539; 10.1093/jxb/52.356.529 (2001).

171. Kobayashi, T. *et al.* Engineering a novel multifunctional green fluorescent protein tag for a wide variety of protein research. *PloS one* **3,** e3822; 10.1371/journal.pone.0003822 (2008).

172. Kumar, A. & Pal, D. GREEN FLUORESCENT PROTEIN AND THEIR APPLICATIONS IN ADVANCE RESEARCH. *jreas* **01,** 42–46; 10.46565/jreas.2016.v01i01.007 (2016).

173. Misteli, T. & Spector, D. L. Applications of the green fluorescent protein in cell biology and biotechnology. *Nature biotechnology* **15,** 961–964 (1997).

174. Sakamoto, S., Shoyama, Y., Tanaka, H. & Morimoto, S. Application of Green Fluorescent Protein in Immunoassays. *ABB* **05,** 557–563; 10.4236/abb.2014.56065 (2014).

175. Zhang, Y. *et al.* Identification of genes expressed in C. elegans touch receptor neurons. *Nature* **418,** 331–335; 10.1038/nature00891 (2002).

176. Ormo, M. *et al.* Crystal Structure of the Aequorea victoria Green Fluorescent Protein. *Science* **273,** 1392–1395 (1996).

177. Heim, R., Prasher, D. C. & Tsien, R. Y. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **91,** 12501–12504 (1994).

178. Reid, B. G. & Flynn, G. C. Chromophore Formation in Green Fluorescent Protein. *Biochemistry* **36,** 6786–6791 (1997).

179. Craggs, T. D. Green fluorescent protein: structure, folding and chromophore maturation. *Chemical Society reviews* **38,** 2865–2875; 10.1039/b903641p (2009).

180. Yang, L. *et al.* Role of Hydrogen Bonding in Green Fluorescent Protein-like Chromophore Emission. *Scientific Reports* **9,** 11640; 10.1038/s41598-019-47660-0 (2019).

181. Jain, R. K. & Ranganathan, R. Local complexity of amino acid interactions in a protein core. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 111–116 (2004).

182. McCapra, F., Razavi, Z. & Neary, A. P. The fluorescence of the chromophore of the green fluorescent protein of Aequorea and Renilla. *Journal of the Chemical Society, Chemical Communications* **12,** 790–791 (1988).

183. Stepanenko, O. V., Stepanenko, O. V., Kuznetsova, I. M., Verkhusha, V. V. & Turoverov, K. K. Beta-barrel scaffold of fluorescent proteins: folding, stability and role in chromophore formation. *International review of cell and molecular biology* **302,** 221–278; 10.1016/B978-0-12-407699-0.00004-2 (2013).

184. Zhang, G., Gurtu, V. & Kain, S. R. An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. *Biochemical and biophysical research communications* **227,** 707–711; 10.1006/bbrc.1996.1573 (1996).

185. Cubitt, A. B., Woollenweber, L. A. & Heim, R. Understanding structure-function relationships in the Aequorea victoria green fluorescent protein. *Methods in Cell Biology* **58,** 19–30 (1999).

186. Heim, R. & Tsien, R. Y. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current Biology* **6,** 178–182; 10.1016/S0960-9822(02)00450-5 (1996).

187. Tsien, R. Y. The green fluorescent protein. *Annual review of biochemistry* **67,** 509–544 (1998).

188. *A fluorescent protein database*. Available at https://www.fpbase.org/ ,

189. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology* **24,** 79–88; 10.1038/nbt1172 (2006).

190. Pletnev, V. Z. *et al.* Structure of the green fluorescent protein NowGFP with an anionic tryptophan-based chromophore. *Acta crystallographica. Section D, Biological crystallography* **71,** 1699–1707; 10.1107/S1399004715010159 (2015).

191. Shcherbakova, D. M., Hink, M. A., Joosen, L., Gadella, T. W. J. & Verkhusha, V. V. An orange fluorescent protein with a large Stokes shift for single-excitation multicolor FCCS and FRET imaging. *Journal of the American Chemical Society* **134,** 7913–7923; 10.1021/ja3018972 (2012).

192. Hyun Bae, J. *et al.* Expansion of the Genetic Code Enables Design of a Novel “Gold” Class of Green Fluorescent Proteins. *Journal of Molecular Biology* **328,** 1071–1081; 10.1016/S0022-2836(03)00364-4 (2003).

193. Steiner, T. *et al.* Synthetic biology of proteins: tuning GFPs folding and stability with fluoroproline. *PloS one* **3,** e1680; 10.1371/journal.pone.0001680 (2008).

194. Magliery, T. J. *et al.* Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *Journal of the American Chemical Society* **127,** 146–157; 10.1021/ja046699g (2005).

195. Kostyuk, A. I. *et al.* In Vivo Imaging with Genetically Encoded Redox Biosensors. *International journal of molecular sciences* **21**; 10.3390/ijms21218164 (2020).

196. Tanz, S. K., Castleden, I., Small, I. D. & Millar, A. H. Fluorescent protein tagging as a tool to define the subcellular distribution of proteins in plants. *Frontiers in plant science* **4,** 214; 10.3389/fpls.2013.00214 (2013).

197. Tebo, A. G. & Gautier, A. A split fluorescent reporter with rapid and reversible complementation. *Nature communications* **10,** 2822; 10.1038/s41467-019-10855-0 (2019).

198. Vasiljevic, S. *et al.* Green fluorescent protein as a reporter of prion protein folding. *Virology journal* **3,** 59; 10.1186/1743-422X-3-59 (2006).

199. Linderstrom-Lang, K. U. *Protein and enzymesLane medical lectures. Lane medical lectures, 1951* (Stanford University Press, Stanford, CA, 1952).

200. Eisenberg, D. The discovery of the α-helix and β-sheet, the principal structural features of proteins. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 11207–11210 (2003).

201. Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **437,** 640–647; 10.1038/nature04162 (2005).

202. Tanford, C. How protein chemists learned about the hydrophobic factor. *Protein Science* **6,** 1358–1366 (1997).

203. Baldwin, R. L. Dynamic hydration shell restores Kauzmann's 1959 explanation of how the hydrophobic factor drives protein folding. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 13052–13056; 10.1073/pnas.1414556111 (2014).

204. *Essentials of Medical Biochemistry.* 2nd ed. (Elsevier, 2015).

205. Fabrizio Chiti and Christopher M. Dobson. Protein Misfolding, Functional Amyloid, and Human Disease.

206. Bhagavan, N. V. & Ha, C.-E. Three-Dimensional Structure of Proteins and Disorders of Protein Misfolding. In *Essentials of Medical Biochemistry.* 2nd ed. (Elsevier2015), pp. 31–51.

207. Ellis, R. J. Molecular chaperones: assisting assembly in addition to folding. *Trends in biochemical sciences* **31,** 395–401; 10.1016/j.tibs.2006.05.001 (2006).

208. Russo Krauss, I., Merlino, A., Vergara, A. & Sica, F. An overview of biological macromolecule crystallization. *International journal of molecular sciences* **14,** 11643–11691; 10.3390/ijms140611643 (2013).

209. Hu, Y. *et al.* NMR-Based Methods for Protein Analysis. *Analytical chemistry* **93,** 1866–1879; 10.1021/acs.analchem.0c03830 (2021).

210. Clore, G. M. & Gronenborn, A. M. Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy. *Critical reviews in biochemistry and molecular biology* **24,** 479–564; 10.3109/10409238909086962 (1989).

211. Jonic, S. & Vénien-Bryan, C. Protein structure determination by electron cryo-microscopy. *Current opinion in pharmacology* **9,** 636–642; 10.1016/j.coph.2009.04.006 (2009).

212. Blundell, T. L. & Johnson, L. N. *Protein crystallography.* 7th ed. (Acad. Press, New York u.a., 1976).

213. McPherson, A. Chapter 1 Introduction to the Crystallization of Biological Macromolecules. In *Current Topics in Membranes, Volume 63. Chapter 1 Introduction to the Crystallization of Biological Macromolecules* (Elsevier2009), Vol. 63, pp. 5–23.

214. McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta crystallographica. Section F, Structural biology communications* **70,** 2–20; 10.1107/S2053230X13033141 (2014).

215. Berman, H. M. Westbrook, J. *et al.* The protein data bank. *Nucleic acids research* **28,** 235–242 (2000).

216. *RCSB PDB*. Available at https://www.rcsb.org/ ,

217. Hoffmann, W. *et al.* An Intrinsic Hydrophobicity Scale for Amino Acids and Its Application to Fluorinated Compounds. *Angewandte Chemie (International ed. in English)* **58,** 8216–8220; 10.1002/anie.201813954 (2019).

218. Gerling, U. I. M. *et al.* Fluorinated amino acids in amyloid formation: a symphony of size, hydrophobicity and α-helix propensity. *Chem. Sci.* **5,** 819–830; 10.1039/C3SC52932K (2014).

219. Budisa, N. *et al.* Efforts towards the design of 'teflon' proteins: in vivo translation with trifluorinated leucine and methionine analogues. *Chemistry & biodiversity* **1,** 1465–1475; 10.1002/cbdv.200490107 (2004).

220. Asante, V., Mortier, J., Schlüter, H. & Koksch, B. Impact of fluorination on proteolytic stability of peptides in human blood plasma. *Bioorganic & medicinal chemistry* **21,** 3542–3546; 10.1016/j.bmc.2013.03.051 (2013).

221. Jakubowski, H. Translational Accuracy of Aminoacyl-tRNA Synthetases: Implications for Atherosclerosis. *The Journal of Nutrition* **131,** 2983S–2987S (2001).

222. Völler, J.-S. Activation, proofreading and translation of aliphatic fluorinated amino acids by class Ia AARSs. PhD thesis. Freie Universität Berlin, 2016.

223. Carrington, J. C. & Dougherty, W. G. A viral cleavage site cassette: Identification of amino acid sequences required for tobacco etch virus polyprotein processing. *Proceedings of the National Academy of Sciences of the United States of America* **85,** 3391–3395 (1988).

224. Dvorak, P. *et al.* Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21(DE3) carrying a synthetic metabolic pathway. *Microbial cell factories* **14,** 201; 10.1186/s12934-015-0393-3 (2015).

225. Rozkov, A. Control of proteolysis of recombinant proteins in Escherichia coli. KTH Royal Institute of Technology, 2001.

226. Gottesman, S. Proteases and their targets in Escherichia coli. *Annual review of genetics* **30,** 465–506 (1996).

227. Chayen, N. E. & Saridakis, E. Protein crystallization: from purified protein to diffraction-quality crystal. *Nature methods* **5,** 147–153; 10.1038/nmeth.f.203 (2008).

228. D'Arcy, A., Mac Sweeney, A. & Haber, A. Using natural seeding material to generate nucleation in protein crystallization experiments. *Acta crystallographica. Section D, Biological crystallography* **59,** 1343–1346; 10.1107/s0907444903009430 (2003).

229. Bergfors, T. Seeds to crystals. *Journal of Structural Biology* **142,** 66–76; 10.1016/S1047-8477(03)00039-X (2003).

230. Buck, M. Crystallography. *Structure* **11,** 735–736; 10.1016/S0969-2126(03)00130-8 (2003).

231. Fukunaga, R. & Yokoyama, S. Crystallization and preliminary X-ray crystallographic study of the editing domain of Thermus thermophilus isoleucyl-tRNA synthetase complexed with pre- and post-transfer editing-substrate analogues. *Acta crystallographica. Section D, Biological crystallography* **60,** 1900–1902; 10.1107/S0907444904019511 (2004).

232. Fukunaga, R., Fukai, S., Ishitani, R., Nureki, O. & Yokoyama, S. Crystal structures of the CP1 domain from Thermus thermophilus isoleucyl-tRNA synthetase and its complex with L-valine. *The Journal of biological chemistry* **279,** 8396–8402; 10.1074/jbc.M312830200 (2004).

233. Völler, J.-S. *et al.* Discovery and Investigation of Natural Editing Function against Artificial Amino Acids in Protein Translation. *ACS central science* **3,** 73–80; 10.1021/acscentsci.6b00339 (2017).

234. Zivkovic, I., Moschner, J., Koksch, B. & Gruic-Sovulj, I. Mechanism of discrimination of isoleucyl-tRNA synthetase against nonproteinogenic α-aminobutyrate and its fluorinated analogues. *The FEBS journal* **287,** 800–813; 10.1111/febs.15053 (2020).

235. Bishop, A. C., Nomanbhoy, T. K. & Schimmel, P. Blocking site-to-site translocation of a misactivated amino acid by mutation of a class I tRNA synthetase. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 585–590; 10.1073/pnas.012611299 (2002).

236. Nakama, T., Nureki, O. & Yokoyama, S. Structural basis for the recognition of isoleucyl-adenylate and an antibiotic, mupirocin, by isoleucyl-tRNA synthetase. *The Journal of biological chemistry* **276,** 47387–47393; 10.1074/jbc.M109089200 (2001).

237. Schmidt, E. and Schimmel, P. Dominant Lethality by Expression of a Catalytically Inactive Class I tRNA Synthetase. *Proceedings of the National Academy of Sciences of the United States of America* **90,** 6919–6923 (1993).

238. Fukunaga, R. & Yokoyama, S. Structural basis for substrate recognition by the editing domain of isoleucyl-tRNA synthetase. *Journal of Molecular Biology* **359,** 901–912; 10.1016/j.jmb.2006.04.025 (2006).

239. Dulic, M., Cvetesic, N., Perona, J. J. & Gruic-Sovulj, I. Partitioning of tRNA-dependent editing between pre- and post-transfer pathways in class I aminoacyl-tRNA synthetases. *The Journal of biological chemistry* **285,** 23799–23809; 10.1074/jbc.M110.133553 (2010).

240. Dower, W. J., Miller, J. F. & Ragsdale, C. W. High efficiency transformation of E. coli by high voltage electroporation. *Nucleic acids research,* 6127–6145; 10.1093/nar/16.13.6127 (1988).

241. Dagert, M. and Ehrlich, S. D. Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. *Gene,* 23–28 (1979).

242. Jakubowski, H. Proofreading in trans by an aminoacyl-tRNA synthetase: a model for single site editing by isoleucyl-tRNA synthetase. *Nucleic acids research* **24,** 2505–2510; 10.1093/nar/24.13.2505 (1996).

243. Schmidt, E. & Schimmel, P. Mutational isolation of a sieve for editing in a transfer RNA synthetase. *Science* **264,** 265–267 (1994).

244. Muñoz, V. & Serrano, L. Helix design, prediction and stability. *Current opinion in biotechnology* **6,** 382–386; 10.1016/0958-1669(95)80066-2 (1995).

245. Richardson, J. S. The Anatomy and Taxonomy of Protein Structure. In *Advances in Protein Chemistry Volume 34* (Elsevier1981), Vol. 34, pp. 167–339.

246. Krieger, F., Möglich, A. & Kiefhaber, T. Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *Journal of the American Chemical Society* **127,** 3346–3352; 10.1021/ja042798i (2005).

247. Analytical, M. C. A. N. PCR - the polymerase chain reaction. *Analytical methods : advancing methods and applications* **6,** 333–336; 10.1039/c3ay90101g (2013).

248. Del Tito, B. J. *et al.* Effects of a minor isoleucyl tRNA on heterologous protein translation in Escherichia coli. *Journal of bacteriology* **177,** 7086–7091; 10.1128/jb.177.24.7086-7091.1995 (1995).

249. Nguyen, A. W. & Daugherty, P. S. Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nature biotechnology* **23,** 355–360; 10.1038/nbt1066 (2005).

250. Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nature biotechnology* **23,** 102–107; 10.1038/nbt1044 (2005).

251. Arai, Y. *et al.* Spontaneously Blinking Fluorescent Protein for Simple Single Laser Super-Resolution Live Cell Imaging. *ACS chemical biology* **13,** 1938–1943; 10.1021/acschembio.8b00200 (2018).

252. Heim, R. & Tsien, R. Y. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current Biology* **6,** 178–182; 10.1016/S0960-9822(02)00450-5 (1996).

253. Zhou, J., Lin, J., Zhou, C., Deng, X. & Xia, B. An improved bimolecular fluorescence complementation tool based on superfolder green fluorescent protein. *Acta biochimica et biophysica Sinica* **43,** 239–244; 10.1093/abbs/gmq128 (2011).

254. Barondeau, D. P., Putnam, C. D., Kassmann, C. J., Tainer, J. A. & Getzoff, E. D. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 12111–12116 (2003).

255. Kamiyama, D. *et al.* Versatile protein tagging in cells with split fluorescent protein. *Nature communications* **7,** 11046; 10.1038/ncomms11046 (2016).

256. Kim, Y. E., Kim, Y., Kim, J. A., Kim, H. M. & Jung, Y. Green fluorescent protein nanopolygons as monodisperse supramolecular assemblies of functional proteins with defined valency. *Nature communications* **6,** 7134; 10.1038/ncomms8134 (2015).

257. Pinaud, F. & Dahan, M. Targeting and imaging single biomolecules in living cells by complementation-activated light microscopy with split-fluorescent proteins. *Proceedings of the National Academy of Sciences of the United States of America* **108,** E201-10; 10.1073/pnas.1101929108 (2011).

258. van Engelenburg, S. B. & Palmer, A. E. Imaging type-III secretion reveals dynamics and spatial segregation of Salmonella effectors. *Nature methods* **7,** 325–330; 10.1038/nmeth.1437 (2010).

259. Dutra, B. E., Sutera, V. A. & Lovett, S. T. RecA-independent recombination is efficient but limited by exonucleases. *Proceedings of the National Academy of Sciences of the United States of America,* 216–221; 10.1073/pnas.0608293104 (2007).

260. invitrogen by life technologies. One Shot TOP10 Competent Cells 280126 Rev A.0. Available at https://assets.thermofisher.com/TFS-Assets/LSG/manuals/oneshottop10_man.pdf (2013).

261. Durfee, T. *et al.* The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. *Journal of bacteriology,* 2597–2606; 10.1128/JB.01695-07 (2008).

262. Liu, X. *et al.* The effects of kanamycin concentration on gene transcription levels in Escherichia coli. *3 Biotech* **10,** 93; 10.1007/s13205-020-2100-2 (2020).

263. Buchan, J. R., Aucott, L. S. & Stansfield, I. tRNA properties help shape codon pair preferences in open reading frames. *Nucleic acids research* **34,** 1015–1027; 10.1093/nar/gkj488 (2006).

264. Yanagisawa, T., Lee, J. T., Wu, H. C. & Kawakami, M. Relationship of protein structure of isoleucyl-tRNA synthetase with pseudomonic acid resistance of Escherichia coli. A proposed mode of action of pseudomonic acid as an inhibitor of isoleucyl-tRNA synthetase. *Journal of Biological Chemistry* **269,** 24304–24309; 10.1016/S0021-9258(19)51082-1 (1994).

265. Isaksson, L. A., Sköld, S. E., Skjöldebrand, J. & Takata, R. A procedure for isolation of spontaneous mutants with temperature sensitive of RNA and/or protein. *Molecular & general genetics : MGG* **156,** 233–237; 10.1007/BF00267177 (1977).

266. Miyanoiri, Y. *et al.* Highly efficient residue-selective labeling with isotope-labeled Ile, Leu, and Val using a new auxotrophic E. coli strain. *Journal of biomolecular NMR* **65,** 109–119; 10.1007/s10858-016-0042-0 (2016).

267. Singer, M. *et al.* A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of Escherichia coli. *Microbiological reviews,* 1–24; 10.1128/mr.53.1.1-24.1989 (1989).

268. Nichols, B. P., Shafiq, O. & Meiners, V. Sequence analysis of Tn10 insertion sites in a collection of Escherichia coli strains used for genetic mapping and strain construction. *Journal of bacteriology* **180,** 6408–6411; 10.1128/JB.180.23.6408-6411.1998 (1998).

269. Miroux, B. & Walker, J. E. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology* **260,** 289–298; 10.1006/jmbi.1996.0399 (1996).

270. Dumon-Seignovert, L., Cariot, G. & Vuillard, L. The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein expression and purification* **37,** 203–206; 10.1016/j.pep.2004.04.025 (2004).

271. Lin, M. T. *et al.* A rapid and robust method for selective isotope labeling of proteins. *Methods (San Diego, Calif.)* **55,** 370–378; 10.1016/j.ymeth.2011.08.019 (2011).

272. Michaels, J.-E. A., Shiba, K. & Miller, W. Autonomous folding of a C-terminal inhibitory fragment of Escherichia coli isoleucine-tRNA synthetase. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1433,** 103–109; 10.1016/S0167-4838(99)00153-3 (1999).

273. Creighton, T. E. (ed.). *Protein folding* (W. H. Freeman and Company, New York, 1992).

274. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature,* 894–898; 10.1038/nature08187 (2009).

275. 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 of the United States of America,* 6640–6645; 10.1073/pnas.120163297 (2000).

276. Wang, H. H. & Church, G. M. Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Methods in enzymology,* 409–426; 10.1016/B978-0-12-385120-8.00018-8 (2011).

277. Erler, A. *et al.* Conformational adaptability of Redbeta during DNA annealing and implications for its structural relationship with Rad52. *Journal of Molecular Biology* **391,** 586–598; 10.1016/j.jmb.2009.06.030 (2009).

278. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)* **315,** 1709–1712; 10.1126/science.1138140 (2007).

279. Mougiakos, I., Bosma, E. F., Vos, W. M. de, van Kranenburg, R. & van der Oost, J. Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. *Trends in biotechnology,* 575–587; 10.1016/j.tibtech.2016.02.004 (2016).

280. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337,** 816–821 (2012).

281. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research* **24,** 132–141; 10.1101/gr.162339.113 (2014).

282. Zhao, D. *et al.* CRISPR/Cas9-assisted gRNA-free one-step genome editing with no sequence limitations and improved targeting efficiency. *Scientific Reports,* 16624; 10.1038/s41598-017-16998-8 (2017).

283. Pósfai, G. *et al.* Emergent properties of reduced-genome Escherichia coli. *Science (New York, N.Y.),* 1044–1046; 10.1126/science.1126439 (2006).

284. Fujitani, Y., Yamamoto, K. & Kobayashi, I. Dependence of Frequency of Homologous Recombination on the Homology Length. *Genetics* **141,** 797; 10.1093/genetics/141.2.797 (1995).

285. Richardson, J. S. & Richardson, D. C. The de novo design of protein structures. *Trends in biochemical sciences* **14,** 304–309 (1989).

286. Budisa, N. *et al.* Atomic mutations in annexin V Thermodynamic studies of isomorphous protein variants. *European journal of biochemistry* **253,** 1–9 (1998).

287. Link, A. J., Mock, M. L. & Tirrell, D. A. Non-canonical amino acids in protein engineering. *Current opinion in biotechnology* **14,** 603–609; 10.1016/j.copbio.2003.10.011 (2003).

288. Minks, C., Alefelder, S., Moroder, L., Huber, R. & Budisa, N. Towards New Protein Engineering: In Vivo Building and Folding of Protein Shuttles for Drug Delivery and Targeting by the Selective Pressure Incorporation (SPI) Method. *Tetrahedron* **56,** 9431–9442; 10.1016/S0040-4020(00)00827-9 (2000).

289. Takemura, S. *et al.* S-Allyl cysteine improves nonalcoholic fatty liver disease in type 2 diabetes Otsuka Long-Evans Tokushima Fatty rats via regulation of hepatic lipogenesis and glucose metabolism. *Journal of Clinical Biochemistry and Nutrition* **53,** 94–101 (2013).

290. Omar, S. H. & Al-Wabel, N. A. Organosulfur compounds and possible mechanism of garlic in cancer. *Saudi pharmaceutical journal : SPJ : the official publication of the Saudi Pharmaceutical Society* **18,** 51–58; 10.1016/j.jsps.2009.12.007 (2010).

291. Borlinghaus, J., Albrecht, F., Gruhlke, M. C. H., Nwachukwu, I. D. & Slusarenko, A. J. Allicin: chemistry and biological properties. *Molecules (Basel, Switzerland)* **19,** 12591–12618; 10.3390/molecules190812591 (2014).

292. Colín-González, A. L. *et al.* The antioxidant mechanisms underlying the aged garlic extract- and S-allylcysteine-induced protection. *Oxidative medicine and cellular longevity* **2012,** 907162; 10.1155/2012/907162 (2012).

293. Hatono, S., Jimenez, A. & Wargovich, M. J. Chemopreventive effect of S-allylcysteine and its relationship to the detoxification enzyme glutathione S-transferase. *Carcinogenesis* **17,** 1041–1044 (1996).

294. Sumiyoshi, H. & Wargovich, M. J. Chemoprevention of 1,2-dimethylhydrazine-induced colon cancer in mice by naturally occurring organosulfur compounds. *Cancer research* **50,** 5084–5087 (1990).

295. Isidro-Llobet, A., Alvarez, M. & Albericio, F. Amino acid-protecting groups. *Chemical Reviews* **109,** 2455–2504; 10.1021/cr800323s (2009).

296. Boutureira, O. & Bernardes, G. J. L. Advances in chemical protein modification. *Chemical Reviews* **115,** 2174–2195; 10.1021/cr500399p (2015).

297. Torres-Kolbus, J., Chou, C., Liu, J. & Deiters, A. Synthesis of non-linear protein dimers through a genetically encoded Thiol-ene reaction. *PloS one* **9,** e105467; 10.1371/journal.pone.0105467 (2014).

298. Chalker, J. M., Bernardes, G. J. L., Lin, Y. A. & Davis, B. G. Chemical modification of proteins at cysteine: opportunities in chemistry and biology. *Chemistry, an Asian journal* **4,** 630–640; 10.1002/asia.200800427 (2009).

299. Morgan, A. A. & Rubenstein, E. Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome. *PloS one* **8,** e53785; 10.1371/journal.pone.0053785 (2013).

300. Ramachandran, G. N. & Sasisekharan, V. Conformation of polypeptides and proteins. *Advances in protein chemistry* **23,** 283–438; 10.1016/s0065-3233(08)60402-7 (1968).

301. MacArthur, M. W. & Thornton, J. M. Influence of proline residues on protein conformation. *Journal of Molecular Biology* **218,** 397–412; 10.1016/0022-2836(91)90721-H (1991).

302. Wedemeyer, W. J., Welker, E. & Scheraga, H. A. Proline cis-trans isomerization and protein folding. *Biochemistry* **41,** 14637–14644; 10.1021/bi020574b (2002).

303. Brandts, J. F., Halvorson, H. R. & Brennan, M. Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* **14,** 4953–4963; 10.1021/bi00693a026 (1975).

304. Di Wu. The puckering free-energy surface of proline. *AIP Advances* **3,** 32141; 10.1063/1.4799082 (2013).

305. Milner-White, E., Bell, L. H. & Maccallum, P. H. Pyrrolidine ring puckering in cis and trans-proline residues in proteins and polypeptides. *Journal of Molecular Biology* **228,** 725–734; 10.1016/0022-2836(92)90859-I (1992).

306. Kubyshkin, V., Davis, R. & Budisa, N. Biochemistry of fluoroprolines: the prospect of making fluorine a bioelement. *Beilstein journal of organic chemistry* **17,** 439–460; 10.3762/bjoc.17.40 (2021).

307. Schumacher, M., Mizuno, K. & Bächinger, H. P. The crystal structure of the collagen-like polypeptide (glycyl-4(R)-hydroxyprolyl-4(R)-hydroxyprolyl)9 at 1.55 A resolution shows up-puckering of the proline ring in the Xaa position. *The Journal of biological chemistry* **280,** 20397–20403; 10.1074/jbc.M501453200 (2005).

308. Ho, B. K., Coutsias, E. A., Seok, C. & Dill, K. A. The flexibility in the proline ring couples to the protein backbone. *Protein science : a publication of the Protein Society* **14,** 1011–1018; 10.1110/ps.041156905 (2005).

309. Kubyshkin, V. & Budisa, N. cis-trans-Amide isomerism of the 3,4-dehydroproline residue, the 'unpuckered' proline. *Beilstein journal of organic chemistry* **12,** 589–593; 10.3762/bjoc.12.57 (2016).

310. Suzuki, T. *et al.* Development of cysteine-free fluorescent proteins for the oxidative environment. *PloS one* **7,** e37551; 10.1371/journal.pone.0037551 (2012).

311. Caló, E. & Khutoryanskiy, V. V. Biomedical applications of hydrogels: A review of patents and commercial products. *European Polymer Journal* **65,** 252–267; 10.1016/j.eurpolymj.2014.11.024 (2015).

312. Yesildag, C., Ouyang, Z., Zhang, Z. & Lensen, M. C. Micro-Patterning of PEG-Based Hydrogels With Gold Nanoparticles Using a Reactive Micro-Contact-Printing Approach. *Frontiers in chemistry* **6,** 667; 10.3389/fchem.2018.00667 (2018).

313. Nojoumi, S., Ma, Y., Schwagerus, S., Hackenberger, C. P. R. & Budisa, N. In-Cell Synthesis of Bioorthogonal Alkene Tag S-Allyl-Homocysteine and Its Coupling with Reprogrammed Translation. *International journal of molecular sciences* **20**; 10.3390/ijms20092299 (2019).

314. Exner, M. P. *et al.* Incorporation of Amino Acids with Long-Chain Terminal Olefins into Proteins. *Molecules (Basel, Switzerland)* **21,** 287; 10.3390/molecules21030287 (2016).

315. Sarkisyan, K. S. *et al.* Green fluorescent protein with anionic tryptophan-based chromophore and long fluorescence lifetime. *Biophysical journal* **109,** 380–389; 10.1016/j.bpj.2015.06.018 (2015).

316. Guohong Zhang, Vanessa Gurtu & Steven R. Kain. An Enhanced Green Fluorescent Protein Allows Sensitive Detection of Gene Transfer in Mammalian Cells. *Biochemical and biophysical research communications* **227,** 707–711; 10.1006/bbrc.1996.1573 (1996).

317. Vámosi, G. *et al.* EGFP oligomers as natural fluorescence and hydrodynamic standards. *Scientific Reports* **6,** 33022; 10.1038/srep33022 (2016).

318. George Abraham, Bobin AND Sarkisyan, Karen S. AND Mishin, Alexander S. AND Santala, Ville AND Tkachenko, Nikolai V. AND Karp, Matti. Fluorescent Protein Based FRET Pairs with Improved Dynamic Range for Fluorescence Lifetime Measurements. *PloS one* **10,** 1–15; 10.1371/journal.pone.0134436 (2015).

319. Pletneva, N. V. *et al.* Crystal Structure of Phototoxic Orange Fluorescent Proteins with a Tryptophan-Based Chromophore. *PloS one* **10,** e0145740; 10.1371/journal.pone.0145740 (2015).

320. Sarkisyan, K. S. *et al.* KillerOrange, a Genetically Encoded Photosensitizer Activated by Blue and Green Light. *PloS one* **10,** e0145287; 10.1371/journal.pone.0145287 (2015).

321. Bulina, M. E. *et al.* A genetically encoded photosensitizer. *Nature biotechnology* **24,** 95–99; 10.1038/nbt1175 (2006).

322. Lin, J. Y. *et al.* Optogenetic inhibition of synaptic release with chromophore-assisted light inactivation (CALI). *Neuron* **79,** 241–253; 10.1016/j.neuron.2013.05.022 (2013).

323. Ryumina, A. P. *et al.* Flavoprotein miniSOG as a genetically encoded photosensitizer for cancer cells. *Biochimica et biophysica acta* **1830,** 5059–5067; 10.1016/j.bbagen.2013.07.015 (2013).

324. Serebrovskaya, E. O. *et al.* Light-induced blockage of cell division with a chromatin-targeted phototoxic fluorescent protein. *The Biochemical journal* **435,** 65–71; 10.1042/BJ20101217 (2011).

325. Williams, D. C. *et al.* Rapid and permanent neuronal inactivation in vivo via subcellular generation of reactive oxygen with the use of KillerRed. *Cell reports* **5,** 553–563; 10.1016/j.celrep.2013.09.023 (2013).

326. Larregola, M., Moore, S. & Budisa, N. Congeneric bio-adhesive mussel foot proteins designed by modified prolines revealed a chiral bias in unnatural translation. *Biochemical and biophysical research communications* **421,** 646–650; 10.1016/j.bbrc.2012.04.031 (2012).

327. Yamaguchi, H. & Miyazaki, M. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules* **4,** 235–251; 10.3390/biom4010235 (2014).

328. Ai, H., Shaner, N. C., Cheng, Z., Tsien, R. Y. & Campbell, R. E. Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. *Biochemistry* **46,** 5904–5910; 10.1021/bi700199g (2007).

329. Budisa, N. *et al.* High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. *European journal of biochemistry,* 788–796; 10.1111/j.1432-1033.1995.tb20622.x (1995).

330. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein expression and purification,* 207–234; 10.1016/j.pep.2005.01.016 (2005).

331. Hanahan, D. Studies on Transformation of Escherichia coli with Plasmids. *Journal of Molecular Biology,* 577–580 (1983).

332. Studier, F. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology,* 113–130; 10.1016/0022-2836(86)90385-2 (1986).

333. Agilent Technologies. Manual: QuikChange Site-Directed Mutagenesis Kit. Available at https://www.chem-agilent.com/pdf/strata/200518.pdf.

334. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PloS one* **3,** e3647; 10.1371/journal.pone.0003647 (2008).

335. Miyazaki, K. MEGAWHOP cloning: a method of creating random mutagenesis libraries via megaprimer PCR of whole plasmids. *Methods in enzymology,* 399–406; 10.1016/B978-0-12-385120-8.00017-6 (2011).

336. Nyerges, Á. *et al.* A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proceedings of the National Academy of Sciences of the United States of America,* 2502–2507; 10.1073/pnas.1520040113 (2016).

337. Bayat, H., Modarressi, M. H. & Rahimpour, A. The Conspicuity of CRISPR-Cpf1 System as a Significant Breakthrough in Genome Editing. *Current microbiology,* 107–115; 10.1007/s00284-017-1406-8 (2018).

338. Graham, D. B. & Root, D. E. Resources for the design of CRISPR gene editing experiments. *Genome Biology,* 260; 10.1186/s13059-015-0823-x (2015).

339. Yao, R. *et al.* CRISPR-Cas9/Cas12a biotechnology and application in bacteria. *Synthetic and systems biotechnology,* 135–149; 10.1016/j.synbio.2018.09.004 (2018).

340. McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta crystallographica. Section F, Structural biology communications,* 2–20; 10.1107/S2053230X13033141 (2014).

341. abcam. ab65622 Phosphate Assay Kit (Colorimetric). Available at https://www.abcam.com/ps/products/65/ab65622/documents/Phosphate-Assay-protocol-book-v6f-ab65622%20(website).pdf (2021).

342. Conklin, D. T042-NanoDrop-Spectrophotometers-Nucleic-Acid-Purity-Ratios.

343. McRae, S. R., Brown, C. L. & Bushell, G. R. Rapid purification of EGFP, EYFP, and ECFP with high yield and purity. *Protein expression and purification* **41,** 121–127; 10.1016/j.pep.2004.12.030 (2005).

344. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* **72,** 248–254; 10.1006/abio.1976.9999 (1976).

8. Appendix

8.1 ÄKTA FPLC chromatogram

Representative chromatograms from purification of H_6-TEV-IleRS are shown in following figures that were mentioned in the main part of the thesis. These include the initial H_6-TEV-IleRS three-step purification as well as a one-step purification after the optimization of expression conditions. For TEV protease and fluorescent proteins, purification was performed using a HisTrap column attached to a peristaltic pump. Therefore, no chromatogram could be obtained.

Wash buffer composition for the IMAC purification of H_6-TEV-IleRS was changed from 50 mM TRIS-HCl, pH 8.0, **150 mM KCl**, 1 mM DTT, 10 mM imidazole, 10 % v/v glycerol to 50 mM TRIS-HCl, pH 8.0, **500 mM KCl**, 1 mM DTT, 10 mM imidazole, 10 % v/v glycerol, in order to efficiently get rid of contaminations.

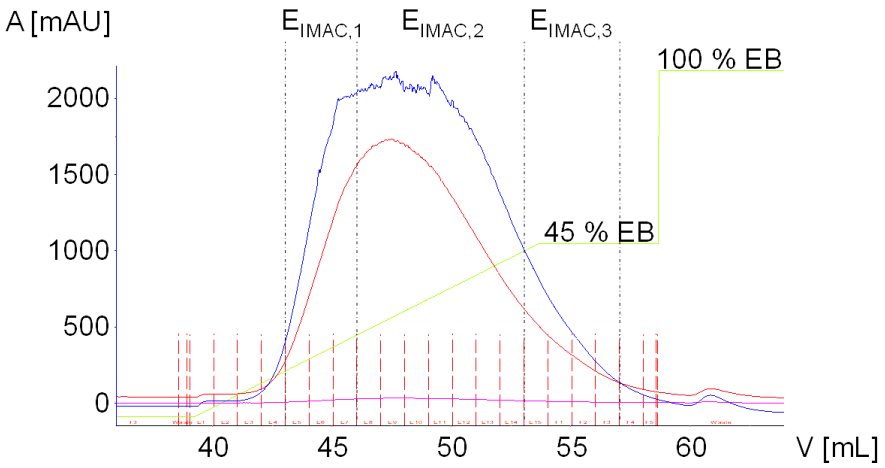


Figure 65 Analysis of H_6-TEV-IleRS purification by IMAC using a linear gradient for elution. Eluted fractions $E_{IMAC,1}$, $E_{IMAC,2}$ and $E_{IMAC,3}$ containing the target product were analyzed by SDS-PAGE. The corresponding result is shown in **Figure 17B**. A linear gradient from 0 to 50 % of elution buffer (EB: 50 mM TRIS-HCl, pH 8.0, 150 mM KCl, 1 mM DTT, 250 mM imidazole, 10 % v/v glycerol) resulted in product elution in a broad peak. Hence, further purification will apply step elution to obtain concentrated samples. Blue curve represents the absorption detected at 280 nm (A_{280}) and red curve at 260 nm (A_{260}). The A_{260}/ A_{280} ratio of about 0.9 of the main product indicates the presence of contaminations within this protein sample, which is in accordance with the SDS-PAGE results. Additionally, absorption at 320 nm was measured to monitor the background absorption, which is derived from e.g., precipitate. 50 mM TRIS-HCl, pH 8.0, 150 mM KCl, 1 mM DTT, 10 mM imidazole, 10 % v/v glycerol were used for equilibration of the column and 50 mM TRIS-HCl, pH 8.0, 150 mM KCl, 1 mM DTT, 25 mM imidazole, 10 % v/v glycerol were applied as wash buffer.

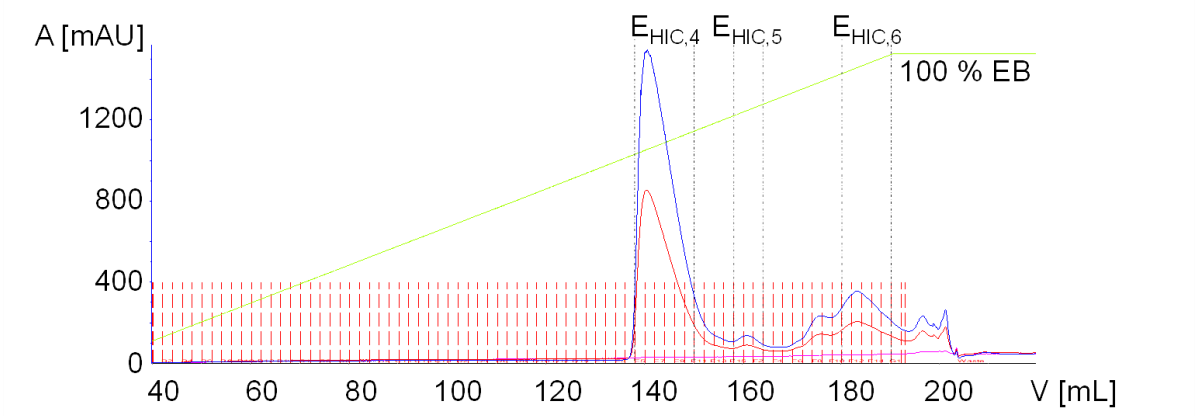


Figure 66 Analysis of H_6-TEV-IleRS purification by HIC using a linear gradient for elution. Eluted fractions $E_{HIC,4}$, $E_{HIC,5}$ and $E_{HIC,6}$ were applied for SDS-PAGE analysis. The corresponding result shows the desired product in all three fractions (**Figure 17B**). The unusual fractionation pattern may arise out of different protein mixture compositions. A linear gradient from 0 to 100 % of elution buffer was applied for elution (EB: 25 mM HEPES, pH 7.5, 10 % v/v glycerol). Blue curve represents the absorption detected at 280 nm (A_{280}) and red curve at 260 nm (A_{260}). However, the A_{260}/ A_{280} ratio of about 0.6 of the main product indicates a protein sample with improved purity, also observed on the SDS-PAGE gel (**Figure 17B**). Additionally, absorption at 320 nm was measured to monitor the background absorption, which is derived from e.g., precipitate. 25 mM HEPES, pH 7.5, 1.7 M $(NH_4)_2SO_4$, 10 % v/v glycerol were used as binding/wash buffer.

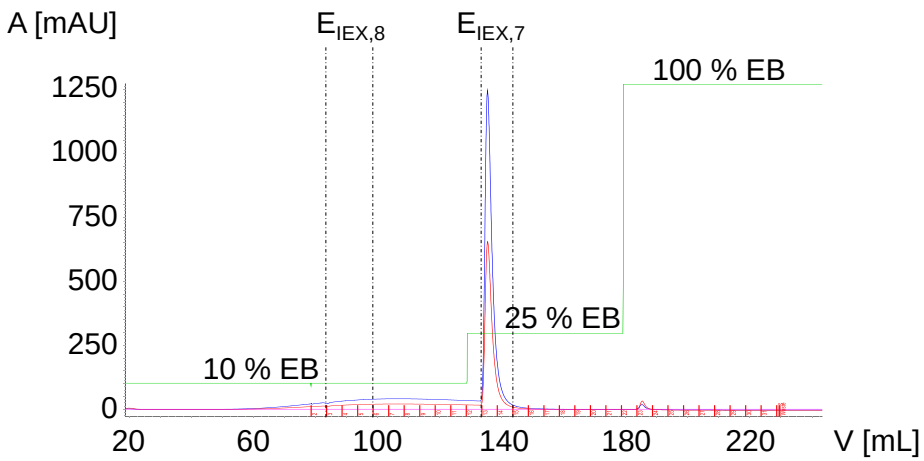


Figure 67 Analysis of H_6-TEV-IleRS purification by IEX using step elution. Eluted fractions $E_{IEX,7}$ and $E_{IEX,8}$ were analyzed by SDS-PAGE. The corresponding result is shown in **Figure 17C**. Therefore, it can be concluded that a small amount of target protein already eluted at 10 % elution buffer (EB: 50 mM TRIS-HCl, pH 8.0, 1 M NaCl, 10 % v/v glycerol), while the main part of the desired product is obtained with 25 % EB. Blue curve represents the absorption detected at 280 nm (A_{280}) and red curve at 260 nm (A_{260}). The A_{260}/ A_{280} ratio of about 0.6 at the main peak indicates a protein sample with high purity. Additionally, absorption at 320 nm was measured to monitor the background absorption, which is derived from e.g., precipitate. The percentage of EB was determined by previous purifications using the linear gradient of 0 to 50 % EB (data not shown). 50 mM TRIS-HCl, pH 8.0, 5 mM NaCl, 10 % v/v glycerol were used as binding buffer and 50 mM TRIS-HCl, pH 8.0, 10 mM NaCl, 10 % v/v glycerol for the wash step.

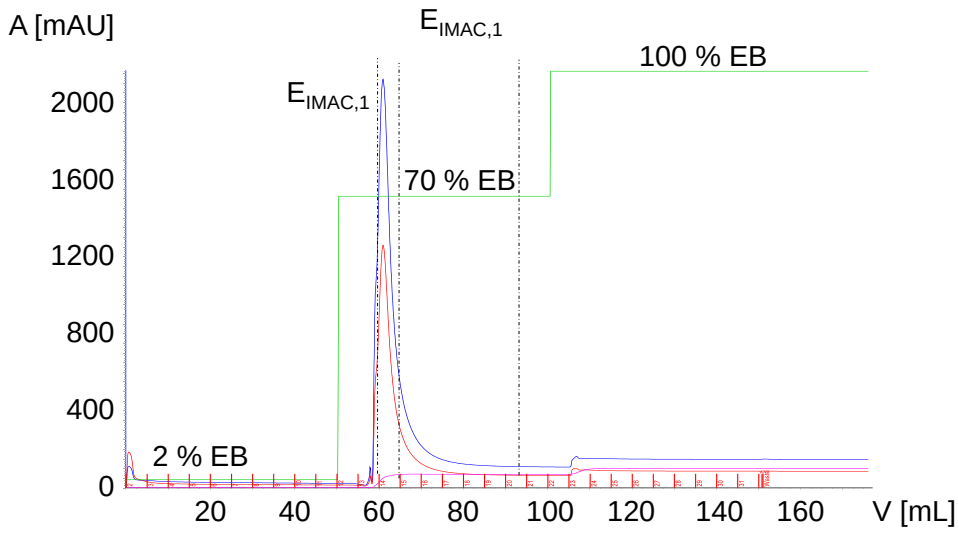


Figure 68 Purification of H_6-TEV-IleRS from optimized expression conditions using step elution. Eluted fractions $E_{IMAC,1}$ and $E_{IMAC,2}$ were analyzed by SDS-PAGE. The corresponding result is shown in **Figure 18B**. In the first step of elution, 2 % of elution buffer (EB: 50 mM TRIS-HCl, pH 8.0, 150 mM KCl, 1 mM DTT, 250 mM imidazole, 10 % v/v glycerol) was used to eliminate residual contaminations. In the second step, target protein was eluted using 70 % EB. Blue curve represents the absorption detected at 280 nm (A_{280}) and red curve at 260 nm (A_{260}). The A_{260}/ A_{280} ratio of about 0.6 of the main product indicates a protein sample with high purity, which is in accordance with the SDS-PAGE results. Additionally, absorption at 320 nm was measured to monitor the background absorption, which is derived from e.g., precipitate. 50 mM TRIS-HCl, pH 8.0, 150 mM KCl, 1 mM DTT, 10 mM imidazole, 10 % v/v glycerol were used for equilibration of the column and 50 mM TRIS-HCl, pH 8.0, 500 mM KCl, 1 mM DTT, 25 mM imidazole, 10 % v/v glycerol were applied as wash buffer.

8.2 Protein primary structures

H_6-TEV-IleRS

MRGS**HHHHHH**GSACEL*ENLYFQ*GMSDYKSSLNLPETGFPMRGDLAKREPGMLARWTDD
DLYGIIRAAKKGKKTFILHDGPPYANGSIHIGHSVNKILKDIIVKSKGLSGYDSPYVPGWDCHG
LPIELKVEQEYGKPGEKFTAAEFRAKCREYAATQVDGQRKDFIRLGVLGDWSHPYLTMDFK
TEANIIRALGKIIGNGHLHKGAKPVHWCVDCRSALAEAEVEYYDKTSPSIDVAFQAVDQDAL
KAKFAVSNVNGPISLVIWTTTPWTLPANRAISIAPDFDYALVQIDGQAVILAKDLVESVMQRIG
VTDYTILSTVKGAELELLRFTHPFMGFDVPAILGDHVTLDAGTGAVHTAPGHGPDDYVIGQK
YGLETANPVGPDGTYLPGTYPTLDGVNVFKANDIVVALLQEKGALLHVEKMQHSYPCCWR
HKTPIIFRATPQWFVSMDQKGLRAQSLKEIKGVQWIPDWGQARIESMVANRPDWCISRQRT
WGVPMSLFVHKDTEELHPRTLELMEEVAKRVEVNGIQAWWDLDAKEILGDEADQYVKVPD
TLDVWFDSGSTHSSVVDVRPEFAGHAADMYLEGSDQHRGWFMSSLMISTAMKGKAPYRQ
VLTHGFTVDGQGRKMSKSIGNTVSPQDVMNKLGADILRLWVASTDYTGEMAVSDEILKRAA
DSYRRIRNTARFLLANLNGFDPAKDMVKPEEMVVLDRWAVGCAKAAQEDILKAYEAYDFHE
VVQRLMRFCSVEMGSFYLDIIKDRQYTAKADSVARRSCQTALYHIAEALVRWMAPILSFTAD
EVWGYLPGEREKYVFTGEWYEGLFGLADSEAMNDAFWDELLKVRGEVNKVIEQARADKK
VGGSLEAAVTLYAEPELAAKLTALGDELRFVLLTSGATVADYNDAPADAQQSEVLKGLKVAL
SKAEGEKCPRCWHYTQDVGKVAEHAEICGRCVSNVAGDGEKRKFA

H_6-tag, *TEV recognition site*

H_6-TEV-IleRS(A579P)

MRGS**HHHHHH**GSACEL*ENLYFQ*GMSDYKSSLNLPETGFPMRGDLAKREPGMLARWTDD
DLYGIIRAAKKGKKTFILHDGPPYANGSIHIGHSVNKILKDIIVKSKGLSGYDSPYVPGWDCHG
LPIELKVEQEYGKPGEKFTAAEFRAKCREYAATQVDGQRKDFIRLGVLGDWSHPYLTMDFK
TEANIIRALGKIIGNGHLHKGAKPVHWCVDCRSALAEAEVEYYDKTSPSIDVAFQAVDQDAL
KAKFAVSNVNGPISLVIWTTTPWTLPANRAISIAPDFDYALVQIDGQAVILAKDLVESVMQRIG
VTDYTILSTVKGAELELLRFTHPFMGFDVPAILGDHVTLDAGTGAVHTAPGHGPDDYVIGQK
YGLETANPVGPDGTYLPGTYPTLDGVNVFKANDIVVALLQEKGALLHVEKMQHSYPCCWR
HKTPIIFRATPQWFVSMDQKGLRAQSLKEIKGVQWIPDWGQARIESMVANRPDWCISRQRT
WGVPMSLFVHKDTEELHPRTLELMEEVAKRVEVNGIQAWWDLDAKEILGDEADQYVKVPD
TLDVWFDSGSTHSSVVDVRPEFAGHAADMYLEGSDQHRGWFMSSLMIST**P**MKGKAPYRQ
VLTHGFTVDGQGRKMSKSIGNTVSPQDVMNKLGADILRLWVASTDYTGEMAVSDEILKRAA
DSYRRIRNTARFLLANLNGFDPAKDMVKPEEMVVLDRWAVGCAKAAQEDILKAYEAYDFHE
VVQRLMRFCSVEMGSFYLDIIKDRQYTAKADSVARRSCQTALYHIAEALVRWMAPILSFTAD
EVWGYLPGEREKYVFTGEWYEGLFGLADSEAMNDAFWDELLKVRGEVNKVIEQARADKK

VGGSLEAAVTLYAEPELAAKLTALGDELRFVLLTSGATVADYNDAPADAQQSEVLKGLKVAL
SKAEGEKCPRCWHYTQDVGKVAEHAEICGRCVSNVAGDGEKRKFA

H_6-tag, *TEV recognition site*, **A579P mutation**

H_6-TEV-IleRS(W905R)

MRGS**HHHHHH**GSACEL***ENLYFQ***SDYKSSLNLPETGFPMRGDLAKREPGMLARWTDDDLY
GIIRAAKKGKKTFILHDGPPYANGSIHIGHSVNKILKDIIVKSKGLSGYDSPYVPGWDCHGLPI
ELKVEQEYGKPGEKFTAAEFRAKCREYAATQVDGQRKDFIRLGVLGDWSHPYLTMDFKTE
ANIIRALGKIIGNGHLHKGAKPVHWCVDCRSALAEAEVEYYDKTSPSIDVAFQAVDQDALKA
KFAVSNVNGPISLVIWTTTPWTLPANRAISIAPDFDYALVQIDGQAVILAKDLVESVMQRIGVT
DYTILSTVKGAELELLRFTHPFMGFDVPAILGDHVTLDAGTGAVHTAPGHGPDDYVIGQKYG
LETANPVGPDGTYLPGTYPTLDGVNVFKANDIVVALLQEKGALLHVEKMQHSYPCCWRHK
TPIIFRATPQWFVSMDQKGLRAQSLKEIKGVQWIPDWGQARIESMVANRPDWCISRQRTW
GVPMSLFVHKDTEELHPRTLELMEEVAKRVEVNGIQAWWDLDAKEILGDEADQYVKVPDTL
DVWFDSGSTHSSVVDVRPEFAGHAADMYLEGSDQHRGWFMSSLMISTAMKGKAPYRQVL
THGFTVDGQGRKMSKSIGNTVSPQDVMNKLGADILRLWVASTDYTGEMAVSDEILKRAAD
SYRRIRNTARFLLANLNGFDPAKDMVKPEEMVVLDRWAVGCAKAAQEDILKAYEAYDFHEV
VQRLMRFCSVEMGSFYLDIIKDRQYTAKADSVARRSCQTALYHIAEALVRWMAPILSFTADE
VWGYLPGEREKYVFTGEWYEGLFGLADSEAMNDAFWDELLKVRGEVNKVIEQARADKKV
GGSLEAAVTLYAEPELAAKLTALGDELRFVLLTSGATVADYNDAPADAQQSEVLKGLKVALS
KAEGEKPRC**R**HYTQDVGKVAEHAEICGRCVSNVAGDGEKRKFA

H_6-tag, *TEV recognition site*, **W905R mutation**

H_6-TEV-IleRS(Ala10)

MRGS**HHHHHH**GSACEL*ENLYFQ*SDYKSSLNLPETGFPMRGDLAKREPGMLARWTDDDLY
GIIRAAKKGKKTFILHDGPPYANGSIHIGHSVNKILKDIIVKSKGLSGYDSPYVPGWDCHGLPI
ELKVEQEYGKPGEKFTAAEFRAKCREYAATQVDGQRKDFIRLGVLGDWSHPYLTMDFKTE
ANIIRALGKIIGNGHLHKGAKPVHWCVDCRSALAEAEVEYYDKTSPSIDVAFQAVDQDALKA
KFAVSNVNGPISLVIW**AAAAAAAAAA**RAISIAPDFDYALVQIDGQAVILAKDLVESVMQRIGV
TDYTILSTVKGAELELLRFTHPFMGFDVPAILGDHVTLDAGTGAVHTAPGHGPDDYVIGQKY
GLETANPVGPDGTYLPGTYPTLDGVNVFKANDIVVALLQEKGALLHVEKMQHSYPCCWRH
KTPIIFRATPQWFVSMDQKGLRAQSLKEIKGVQWIPDWGQARIESMVANRPDWCISRQRT
WGVPMSLFVHKDTEELHPRTLELMEEVAKRVEVNGIQAWWDLDAKEILGDEADQYVKVPD
TLDVWFDSGSTHSSVVDVRPEFAGHAADMYLEGSDQHRGWFMSSLMISTAMKGKAPYRQ
VLTHGFTVDGQGRKMSKSIGNTVSPQDVMNKLGADILRLWVASTDYTGEMAVSDEILKRAA

DSYRRIRNTARFLLANLNGFDPAKDMVKPEEMVVLDRWAVGCAKAAQEDILKAYEAYDFHE
VVQRLMRFCSVEMGSFYLDIIKDRQYTAKADSVARRSCQTALYHIAEALVRWMAPILSFTAD
EVWGYLPGEREKYVFTGEWYEGLFGLADSEAMNDAFWDELLKVRGEVNKVIEQARADKK
VGGSLEAAVTLYAEPELAAKLTALGDELRFVLLTSGATVADYNDAPADAQQSEVLKGLKVAL
SKAEGEKPRCWHYTQDVGKVAEHAEICGRCVSNVAGDGEKRKFA

H_6-tag, *TEV recognition site*, **Ala10 mutation**

H_6-TEV-CP1

MRGS**HHHHHH**GSACEL*ENLYFQ*GMYYDKTSPSIDVAFQAVDQDALKAKFAVSNVNGPISL
VIWTTTPWTLPANRAISIAPDFDYALVQIDGQAVILAKDLVESVMQRIGVTDYTILSTVKGAEL
ELLRFTHPFMGFDVPAILGDHVTLDAGTGAVHTAPGHGPDDYVIGQKYGLETANPVGPDGT
YLPGTYPTLDGVNVFKANDIVVALLQEKGALLHVEKMQHSY

H_6-tag, *TEV recognition site*

H_6-TEV protease

MG**HHHHHH**GESLFKGPRDYNPISSTICHLTNESDGHTTSLYGIGFGPFIITNKHLFRRNNGT
LVVQSLHGVFKVKNTTTLQQHLIDGRDMIIIRMPKDFPPFPQKLKFREPQREERICLVTTNFQ
TKSMSSMVSDTSCTFPSGDGIFWKHWIQTKDGQCGSPLVSTRDGFIVGIHSASNFTNTNNY
FTSVPKNFMELLTNQEAQQWVSGWRLNADSVLWGGHKVFMVKPEEPFQPVKEATQLMN
RRRRR

H_6-tag

H_6-sfGFP(I123T, I128T, I136T, I152V, I188A)

MGSS**HHHHHH**MRKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTT
GKLPVPWPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRA
EVKFEGDTLVNR**T**ELKG**T**DFKEDGN**T**LGHKLEYNFNSHNVY**V**TADKQKNGIKANFKIRHNV
EDGSVQLADHYQQNTP**A**GDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITH
GMDELYKSA*WSHPQFEK*

H_6-tag, **I123T, I128T, I136T, I152V, I188A mutations** (in chronological order), *Strep-tag*

H_6-sfGFP(I123T, I128T, I136T, I152V, I161L, I167V, I188A)

MGSS**HHHHHH**MRKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTT
GKLPVPWPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRA
EVKFEGDTLVNR**T**ELKG**T**DFKEDGN**T**LGHKLEYNFNSHNVY**V**TADKQKNG**L**KANFK**V**RHN
VEDGSVQLADHYQQNTP**A**GDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGIT
HGMDELYKSA*WSHPQFEK*

H_6-tag, **I123T, I128T, I136T, I152V, I161L, I167V, I188A mutations** (in chronological order), *Strep-tag*

H_6-sfGFP(N39I, T105K, E111V, I128T, K166T, I167V, S205T) = GFP1-10

MGSS**HHHHHH**MRKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDAT**I**GKLTLKFICTTG
KLPVPWPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDG**K**YKTRA**V**
VKFEGDTLVNRIELKG**T**DFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANF**TV**RHNVED
GSVQLADHYQQNTPIGDGPVLLPDNHYLSTQ**T**VLSKDPNEK

H_6-tag, **N39I, T105K, E111V, I128T, K166T, I167V, S205T mutations** (in chronological order)

cfsfGFP(R2TAG)-H_6

M*****KGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFI**S**TTGKLPVPWPTLVT
TLTYGVQ**M**FARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRI
ELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQN
TPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGS**HHHHHH**

*** in-frame TAG stop codon**, **C48S and C70M mutations** for cysteine-free construct (in chronological order), **H_6-tag**

EGFP-H_6

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL
VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV
NRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQ
QNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK**HHHHHH**

H_6-tag

H_6-NowGFP

MRGS**HHQHHH**GSVSKGEKLFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKMSLKFIC
TTGKLPVPWPTLKTTLTWGMQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT
RAEVKFEGDTLVNRIELKGVDFKEDGNILGHKLEYNAISGNANITADKQKNGIKAYFTIRHDV
EDGSVLLADHYQQNTPIGDGPVLLPDNHYLSTQSKQSKDPNEKRDHMVLLEFVTAAGIPLG
ADELYK

H_6-tag containing a Q mutation, which does not disturb protein purification and properties.

H_6-KillerOrange

MRGS**HHHHHH**GSECGPALFQSDMTFKIFIDGEVNGQKFTIVADGSSKFPHGDFNVHAVCE
TGKLPMSWKPICHLIQWGEPFFARYPDGISHFAQECFPEGLSIDRTVRFENDGTMTSHHTY
ELSDTCVVSRITVNCDGFQPDGPIMRDQLVDILPSETHMFPHGPNAVRQLAFIGFTTADGGL
MMGHLDSKMTFNGSRAIEIPGPHFVTIITKQMRDTSDKRDHVCQREVAHAHSVPRITSAIGS
DQD

H_6-tag

8.3 Nucleotide sequences

The nucleotides sequences are given in DNA format. Therefore, U is replaced by T.

$tRNA^{Ile1}_{UAU}$

AGGCTTGTAGCTCAGGTGGTTAGAGCGCACCCCT**TAT**AAGGGTGAGGTCGGTGGTTCA
AGTCCACTCAGGCCTACCA

UAU anticodon

$tRNA^{Ile2}_{UAU}$

GGCCCCTTAGCTCAGTGGTTAGAGCAGGCGACT**TAT**AATCGCTTGGTCGCTGGTTCAA
GTCCAGCAGGGGCCACCA

UAU anticodon

H_6-sfGFP(I14AUA, I47AUA, I98AUA, I123T, I128T, I136T, I152V, I161AUA, I167AUA, I188A, I229AUA) = H_6-sfGFP(6xAUA)

ATGGGCAGCAGC**CATCATCATCATCATCAC**ATGCGTAAAGGCGAAGAGCTGTTCACTG
GTGTCGTCCCT**ATA**CTGGTGGAACTGGATGGTGATGTCAACGGTCATAAGTTTTCCGTG
CGTGGCGAGGGTGAAGGTGACGCAACTAATGGTAAACTGACGCTGAAGTTC**ATA**TGTA
CTACTGGTAAACTGCCGGTACCTTGGCCGACTCTGGTAACGACGCTGACTTATGGTGTT
CAGTGCTTTGCTCGTTATCCGGACCATATGAAGCAGCATGACTTCTTCAAGTCCGCCAT
GCCGGAAGGCTATGTGCAGGAACGCACG**ATA**TCCTTTAAGGATGACGGCACGTACAAA
ACGCGTGCGGAAGTGAAATTTGAAGGCGATACCCTGGTAAACCGC**ACC**GAGCTGAAAG
GC**ACC**GACTTTAAAGAAGACGGCAAT**ACC**CTGGGCCATAAGCTGGAATACAATTTTAAC
AGCCACAATGTTTAC**GTG**ACCGCCGATAAACAAAAAAATGGC**ATA**AAAGCGAATTTTAA
A**ATA**CGCCACAACGTGGAGGATGGCAGCGTGCAGCTGGCTGATCACTACCAGCAAAAC
ACTCCA**GCC**GGTGATGGTCCTGTTCTGCTGCCAGACAATCACTATCTGAGCACGCAAA
GCGTTCTGTCTAAAGATCCGAACGAGAAACGCGATCATATGGTTCTGCTGGAGTTCGTA
ACCGCAGCGGGC**ATA**ACGCATGGTATGGATGAACTGTACAAAAGCGCTTGGAGCCACC
CGCAGTTCGAAAAATAA

H_6-tag, **I14AUA, I47AUA, I98AUA, I123T, I128T, I136T, I152V, I161AUA, I167AUA, I188A, I229AUA mutations** (in chronological order) , *Strep-tag*

H_6-sfGFP(I14AUA, I47AUA, I98AUA, I123T, I128T, I136T, I152V, I161L, I167V, I188A, I229AUA) = H_6-sfGFP(4xAUA)

ATGGGCAGCAGC**CATCATCATCATCATCAC**ATGCGTAAAGGCGAAGAGCTGTTCACTG
GTGTCGTCCCT**ATA**CTGGTGGAACTGGATGGTGATGTCAACGGTCATAAGTTTTCCGTG
CGTGGCGAGGGTGAAGGTGACGCAACTAATGGTAAACTGACGCTGAAGTTC**ATA**TGTA
CTACTGGTAAACTGCCGGTACCTTGGCCGACTCTGGTAACGACGCTGACTTATGGTGTT
CAGTGCTTTGCTCGTTATCCGGACCATATGAAGCAGCATGACTTCTTCAAGTCCGCCAT
GCCGGAAGGCTATGTGCAGGAACGCACG**ATA**TCCTTTAAGGATGACGGCACGTACAAA
ACGCGTGCGGAAGTGAAATTTGAAGGCGATACCCTGGTAAACCGC**ACC**GAGCTGAAAG
GC**ACC**GACTTTAAAGAAGACGGCAAT**ACC**CTGGGCCATAAGCTGGAATACAATTTTAAC
AGCCACAATGTTTAC**GTG**ACCGCCGATAAACAAAAAAATGGC**CTG**AAAGCGAATTTTAA
A**GTG**CGCCACAACGTGGAGGATGGCAGCGTGCAGCTGGCTGATCACTACCAGCAAAA
CACTCCA**GCC**GGTGATGGTCCTGTTCTGCTGCCAGACAATCACTATCTGAGCACGCAA
AGCGTTCTGTCTAAAGATCCGAACGAGAAACGCGATCATATGGTTCTGCTGGAGTTCGT
AACCGCAGCGGGC**ATA**ACGCATGGTATGGATGAACTGTACAAAAGCGCTTGGAGCCAC
CCGCAGTTCGAAAAATAA

H_6-tag, **I14AUA, I47AUA, I98AUA, I123T, I128T, I136T, I152V, I161L, I167V, I188A, I229AUA mutations** (in chronological order) , *Strep-tag*

8.4 Primers

Table 25 Primers and their applications.

application	Name	Sequence
Cloning of CP1 domain on pQE-80L plasmid	MT_CP1_HindIII_r	TAATTAAGCTTTTAATAGCTGTGCTGCATTTTCTCAAC
	MT_CP1_SacI_f	ATATGAGCTCGAGAATCTTTATTTTCAGGGCATGTATTACGACAAAACTTCTCCGTCC
	MT_CP1(225-426)_f	GTATTACGACAAAACTTCTCCGTCC
	MT_CP1(225-426)_r	ATAGCTGTGCTGCATTTTCTCAAC
	MT_IleRS_HindIII_f	AAGTTTGCCTGAAAGCTTAATTAGC
	MT_IleRS_SacI_r	CATGCCCTGAAAATAAAGATTCTCG
	MT_IleRS_start seq_rev	GCCCAGCGCGCGGATGATGTTGGC
reduction of Ile residues in sfGFP	MT_sfGFP(I136T)	CAATACCCTGGGCCATAAGCTGGAATAC
	MT_sfGFP(I136T)_r	GTATTCCAGCTTATGGCCCAGGGTATTG
	MT_sfGFP(I188A)	CACTCCAGCCGGTGATGGTCCTGTTCTGC
	MT_sfGFP(I188A)_r	GCAGAACAGGACCATCACCGGCTGGAGTG
	MT_I14M_BsaI_f	TATAGGTCTCATATGCTGGTGGAACTGGATGGTGATGTCAAC
	MT_I14M_BsaI_r	ATATGGTCTCACATAGGGACGACACCAGTGAACAGC
	MT_I47V_BsaI_f_new	TATAGGTCTCACGGCTGTACTACTGGTAAACTGCCGGTACC
	MT_I47V_BsaI_r	TATAGGTCTCAAGCCGAACTTCAGCGTCAGTTTACCATTAGTTGC
	MT_I123T_I128T_BsaI_r	ATTAGGTCTCACTGTGCCTTTCAGCTCGGTGCGGTTTACCAGGGTATCGC
	MT_I128T_BsaI_f	TATGGTCTCTACAGACTTTAAAGAAGACGGCAATACCCTGG
	MT_I152V_BsaI_f	TTAAGGTCTCACGTTCCGCCGATAAACAAAAAAATGGCCTG
	MT_I152V_BsaI_r	TATTGGTCTCTAACGTAAACATTGTGGCTGTTAAAATTGTATTCCAG
	MT_I152V_BsaI_f_new	TTAAGGTCTCACGTTACCGCCGATAAACAAAAAAATGGCCTGAAAG
	MT_161L_167V_f	ATATGGTCTCTCCTGAAAGCGAATTTTAAAGTGCGCCACAACGTGGAGGATGGC
	MT_161L_167V_r	AATTGGTCTCACAGGCCATTTTTTTGTTTATCGGCGGTCACG
	MT_I14NNK_f	TATAGGTCTCATNNKCTGGTGGAACTGGATGGTGATGTC
	MT_I14NNK_r	ATATGGTCTCAMNNAGGGACGACACCAGTGAACAGC
	MT_I98NNK_f	TATAGGTCTCAGNNKTCCTTTAAGGATGACGGCACGTAC
	MT_I98NNK_r	ATATGGTCTCAMNNCGTGCGTTCCTGCACATAGCC
	MT_I229NNK_f	TATAGGTCTCANNKACGCATGGTATGGATGAACTGTACAAAAG
	MT_I229NNK_r	ATATGGTCTCATMNNGCCCGCTGCGGTTACGAACTC
silent mutation of Ile abundant codons to rare AUA codon	MT_sfGFP(6xAUA) f1	AATTGGTCTCCGGCATAACGCATGGTATGGATGAACTGTACAAAAG
	MT_sfGF(6xAUA) f2	AATAGGTCTCACTGGTGGAACTGGATGGTGATGTCAAC
	MT_sfGF(6xAUA) f3	TAATGGTCTCATGTACTACTGGTAAACTGCCGGTACCTTG
	MT_sfGFP(6xAUA) f4	ATTAGGTCTCCCTTTAAGGATGACGGCACGTACAAAACG
	MT_sfGF(6xAUA) f5	TATAGGTCTCAAGCGAATTTTAAAATACGCCACAACGTGGAGGATG

	MT_sfGF(6xAUA) r1	ATTAGGTCTCACCAGTATAGGGACGACACCAGTGAACAGC
	MT_sfGF(6xAUA) r2	TTAAGGTCTCGTACATATGAACTTCAGCGTCAGTTTACCATTAGTTC
	MT_sfGFP(6xAUA) r3	TAATGGTCTCTAAAGGATATCGTGCGTTCCTGCACATAG
	MT_sfGF(6xAUA) r4	TAATGGTCTCTCGCTTTTATGCCATTTTTTTGTTTATCGGCGGTAAC
	MT_sfGFP(6xAUA) r5	TAATGGTCTCGTGCCCGCTGCGGTTACG
	MT_sfGF(4xAUA) f1	ATAAGGTCTCGGGCATAACGCATGGTATGGATGAACTGTACAAAAG
	MT_sfGF(4xAUA) f2	ATTAGGTCTCCTGGTGGAACTGGATGGTGATGTC
	MT_sfGF(4xAUA) f3	ATTAGGTCTCGTGTACTACTGGTAAACTGCCGGTACCTTG
	MT_sfGF(4xAUA) f4	ATTAGGTCTCATCCTTTAAGGATGACGGCACGTACAAAAC
	MT_sfGF(4xAUA) f5	ATTAGGTCTCAAGCGAATTTTAAAGTTCGCCACAACGTGGAGGATGG
	MT_sfGF(4xAUA) r1	ATTAGGTCTCCACCAGTATAGGGACGACACCAGTGAACAG
	MT_sfGF(4xAUA) r2	ATTAGGTCTCGTACAAACGAACTTCAGCGTCAGTTTACCATTAG
	MT_sfGF(4xAUA) r3	ATTAGGTCTCAAGGATATCGTGCGTTCCTGCACATAGC
	MT_sfGF(4xAUA) r4	ATTAGGTCTCTCGCTTTTATGCCATTTTTTTGTTTATCGGCGGTAAC
	MT_sfGF(4xAUA) r5	ATAAGGTCTCATGCCCGCTGCGGTTACGAAC
	MTsfGFP123_128AUA_Bsa_f	ATTAGGTCTCTCATAGACTTTAAAGAAGACGGCAATATACTGG
	MTsfGFP123_128AUA_Bsa_r	TAATGGTCTCTTATGCCTTTCAGCTCTATGCGGTTTACCAGGGTATCGC
	MTsfGFP136AUA_f	GACTTTAAAGAAGACGGCAATATACTG
	MTsfGFP136AUA_f1	GCAATATACTGGGCCATAAGCTGGAATACAA
	MTsfGFP136AUA_r	CAGTATATTGCCGTCTTCTTTAAAGTC
	MTsfGFP136AUA_r1	TTGTATTCCAGCTTATGGCCCAGTATATTGC
	MTsfGFP152AUA_Bsa_f	AATAGGTCTCACATAACCGCCGATAAACAAAAAAATGGC
	MTsfGFP152AUA_Bsa_r	TAATGGTCTCTATATGTAAACATTGTGGCTGTTAAAATTGTATTC
	MTsfGFP188AUA_Bsa_f	TTAAGGTCTCAATAGGTGATGGTCCTGTTCTGCTGC
	MTsfGFP188AUA_Bsa_r	ATTAGGTCTCACTATCGGAGTGTTTTGCTGGTAGTGATC
cloning of GFP1-10 on expression plasmid	MT_GFP1-10(AAT39ATT)_f	GTGACGCAACTATTGGTAAACTG
	MT_GFP1-10(ACG105AAG)_r	CGCGTTTTGTACTTGCCGTC
	MT_GFP1-10(GAA111GTC)_f	CGCGTGCGGTCGTGAAATTTG
	MT_GFP1-10(ATT128ACC)_r	CTTTAAAGTCGGTGCCTTTC
	MT_GFP1-10(AAA166ACC)_f	GAATTTTACCGTTCGCCACAACGTGGAG
	MT_GFP1-10(AGC205ACC)_r	GATCTTTAGACAGAACGGTTTGCG
	MT_GFP1-10(GAA111GTC)new_f	TACAAAACGCGTGCGGTCGTGAAATTTGAAGGCGATAC

	MT_GFP1-10(ATT128ACC)new_r	GCCGTCTTCTTTAAAGTCGGTGCCTTTCAGCTCAATGC
	MT_GFP1-10_f	TAATCTGCAGGAGATCCGGC
	MT_GFP1-10_r	CTGTCTAAAGATCCGAACGAGAAA
	MT_GFP11	ctaaccATGgCGTGACCACATGGTCCTTCATGAGTATGTAAATGCTGCTGGGATTACAGGTGGCGGCAAATTCctgcaggtcg
generation of IleRS libraries	IleRSAla10fw	CGCGCAATCTCTATTGC
	IleRSAla10rev	CCAGATCACCAGCGAGATTG
	Ala10_f	GCGGCTGCGGCGGCGGCGGCGGCGGCGGCG
	Ala10_r	CGCCGCCGCCGCCGCCGCCGCCGCAGCCGC
	IleRS fw seq 1	TACACTGGTGCGTTGACTGC
	IleRS fw seq 2	TGAAAGAGATCAAAGGCGTG
	Lac_sec_fwd	GACTCTCTTCCGGGCGCTAT
	MT_EcIleRS(W905R)_f	GTGAGAAGTGCCCACGCTGCCGTCACTAC
	MT_EcIleRS(W905R)_r	GTAGTGACGGCAGCGTGGGCACTTCTCAC
	MT_IleRS(W905R)seq_f	GTTGCAGACTATAACGACGCACC
	MT_active1_f1	AAATGGTCTCTTGGTNNKTTCATGTCTTCCCTAATGATCTCC
	MT_active1_f2	TAATGGTCTCTCCTTATGCGAATGGCAGCATTC
	MT_active1_f3	TAATGGTCTCTCTGGNNKTGCCACGGTCTGCC
	MT_active1_f4	TTAAGGTCTCTGGATCTACCCACTCTTCTGTTGTTGACG
	MT_active1_r1	AATTGGTCTCGAAGGMNNGCCATCATGCAGAATGAAGG
	MT_active1_r2	TATAGGTCTCACCAGCCAGGCACATACGGC
	MT_active1_r3	TAATGGTCTCTATCCMNNGTCAAACCATACGTCGAGAG
	MT_active1_r4	TTAAGGTCTCTACCACGGTGCTGGTCAGAAC
	MT_active2_f1	AAATGGTCTCTTCTGNNKGGTTCTGACNNKCACCGTGG
	MT_active2_f2	AAATGGTCTCATTTGACTCCGGATCTACCCACTCTTC
	MT_active2_r1	TATTGGTCTCACAAAMNNTACGTCGAGAGTATCCGGC
	MT_active2_r2	TAAAGGTCTCTCAGATACATGTCCGCTGCGTG
	MT_active3_f1	ATTAGGTCTCATCTGNNKGGTTCTGACCAGCACCG
	MT_active3_f2	ATATGGTCTCACCTTATGCGAATGGCAGCATTCATATTGG
	MT_active3_f3	ATTAGGTCTCATGCCACGGTCTGCCGATCG
	MT_active3_f4	AATTGGTCTCATTTGACTCCGGATCTACCCACTCTTCTG
	MT_active3_r1	AATTGGTCTCTAAGGMNNGCCATCATGCAGAATGAAGG
	MT_active3_r2	TAATGGTCTCTGGCAMNNCCAGCCAGGCACATAC
	MT_active3_r3	AATTGGTCTCTCAAAMNNTACGTCGAGAGTATCCGGCAC
	MT_active3_r4	TATAGGTCTCTCAGATACATGTCCGCTGCGTGAC
	MT_editing1_f1	ATTAGGTCTCTCCTGGCNNKGGCCCGGACG
	MT_editing1_f2	ATTAGGTCTCTCTGCCTGCGAACCGCGCAATC
	MT_editing1_r1	ATTAGGTCTCAGCAGMNNCCACGGCGTAGTGG
	MT_editing1_r2	TTAAGGTCTCACAGGMNNGGTMNNAACGGCACCGGTAC
	MT_editing2_f1	AGAGTGGGTCTCTGGACNNKTATGTGATCGGTCAGAAATACGGCC
	MT_editing2_f2	ATATGGTCTCTNNKCCGTGGACTCTGCCTGCGAACC
	MT_editing2_r1	AATTGGGTCTCCGMNNAGTGGTCCAGATCACCAGCGAG
	MT_editing2_r2	ATATGGTCTCAGTCCGGGCCGTGGCCAG
	MT_editing3_f1	AATTGGTCTCCCCTGGCCACGGCCCGGACNNKTATG
	MT_editing3_f2	AATTGGTCTCTGCCTGCGAACCGCGCAATC
	MT_editing3_r1	TTAAGGTCTCTAGGCAGMNNCCACGGCGTAGTG

	MT_editing3_r2	TTAAGGTCTCACAGGMNNMNNGTGAACGGCACCGGTACC
	MT_active2_r3	TTAAGGTCTCTACCACGGTGMNNGTCAGAACC
	MT_editing2n_f1	AGAGTGGGTCTCTCCTGGCCACGGCCCGGACNNKTATGTG
	MT_editing2n_r2	TTAAGGTCTCACAGGMNNGGTMNNAACGGCACCGGTACC
	MT_editing2new_r2	TTAAGGTCTCACAGGCGCMNNMNNAACGGCACCGGTACC
	MT_active2_r4	TTAAGGTCTCATGAAMMNACCACGGTGMNNGTCAGAACC
	MT_active2_f3	TAATGGTCTCTTTCATGTCTTCCCTAATGATCTCCACC
	MT_IRS(P57AAT)_f	GATGGCAATCCTTATGCGAATGGCAGCATTC
	MT_IRS(P57AAT)_r	GAATGCTGCCATTCGCATAAGGATTGCCATC
assembly of screening plasmids	MT1_pTB226_BsaI_f	ATATGGTCTCGGATTATTTTTCGAACTGCGGGTGGC
	MT1_pTB226_BsaI_r	TATAGGTCTCGCAGCAGCCATCATCATCATCATC
	MT2_pTB226_BsaI_f	ATATGGTCTCCGATTATTTTTCGAACTGCGGGTGGC
	MT2_pTB226_BsaI_r	ATATGGTCTCCATACCATGGGCAGCAGCCATCAT
	MT_pET28cl_f	TTAAGGTCTCAACGGTTCCTGGCCTTTTGCTG
	MT_pET28cl_r	TTAAGGTCTCACCGTAAAAAGGCCGCGTTGC
	MT_ID868AmpR_f	TAATGGTCTCCAATGAGCATTCAGCATTTTCGTGTGGCG
	MT_ID868AmpR_r	ATTAGGTCTCCGTTTTACCAATGTTTAATCAGGCTCGCGCCAATTTC
	MT_screenbb_f	ATATGGTCTCGAAACACCCCTTGTATTACTGTTTATGTAAGCAG
	MT_screenbb_r	TTAAGGTCTCGCATTAATTAATTGGACAAGGGTCCTTTTCCGC
	MT1_ID367_BsaI_f	ATTAGGTCTCCATGAGTGACTATAAATCAAGCCTG
	MT1_ID367_BsaI_r	ATAAGGTCTCCAGTCAGGCAAACTTACGTTTTTCACC
	MT1_ID690_BsaI_f	TATAGGTCTCCTAGGCCTGAGTGGACTTGAACC
	MT1_ID690_BsaI_r	ATATGGTCTCCCTTGTAGCTCAGGTGGTTAGAGC
	MT1_ID868_BsaI_f1	ATATGGTCTCCAACATCAATCATCCCCATAATCCTTGTTAG
	MT1_ID868_BsaI_f2	TATAGGTCTCGGACTGCAGTGATCATCTGACCGTAC
	MT1_ID868_BsaI_r1	TATAGGTCTCCTCATATGGGATTCCTCAAAGCGTAAAC
	MT1_ID868_BsaI_r2	TTAAGGTCTCGGGCAAAGCAGAAAAAACGCCGCTG
	MT1_pTB226_BsaI_f	ATATGGTCTCGGATTATTTTTCGAACTGCGGGTGGC
	MT1_pTB226_BsaI_r	TATAGGTCTCGCAGCAGCCATCATCATCATCATC
	MT1_pTB1172_BsaI_f1	TATAGGTCTCCTGCCCTAGACAGCTGGGCG
	MT1_pTB1172_BsaI_f2	TATAGGTCTCGCAAGCCTGGCGATCGCAATGCGGGG
	MT1_pTB1172_BsaI_f3	ATAAGGTCTCCGCTGCCCATGGTATATCTCCTTCTTATTTTTTTGAA
	MT1_pTB1172_BsaI_r1	TTAAGGTCTCCCCTACCAGTTTAAACAATTCGAAAAGCCTGC
	MT1_pTB1172_BsaI_r2	ATTAGGTCTCGAATCTGCAGGAGATCCGGCTGCTAAC
	MT1_pTB1172_BsaI_r3	ATATGGTCTCCTGTTTTCGCCAGATCCCGAAAAGTG
	MT2_ID367_BsaI_f	TATAGGTCTCCCCCATATGAGTGACTATAAATCAAGCCTG
	MT2_ID367_BsaI_r	TATAGGTCTCCTCAGGCAAACTTACGTTTTTCACC
	MT2_ID691_BsaI_f*	ATTAGGTCTCTGGTGGCCCCTGCTGGACTTG
	MT2_ID691_BsaI_r	ATATGGTCTCCTCGCCGGCCCCTTAGCTCAGTGG

	MT2_ID868_BsaI_f1	TATAGGTCTCGACATCAATCATCCCCATAATCCTTGTTAG
	MT2_ID868_BsaI_f2	TTAAGGTCTCGCTGACTGCAGTGATCATCTGACCGTAC
	MT2_ID868_BsaI_r1	TTAAGGTCTCGTGGGATTCCTCAAAGCGTAAAC
	MT2_ID868_BsaI_r2	TATAGGTCTCGGCAAAGCAGAAAAAACGCCGCTG
	MT2_pTB226_BsaI_f	ATATGGTCTCCGATTATTTTTCGAACTGCGGGTGGC
	MT2_pTB226_BsaI_r	ATATGGTCTCCATACCATGGGCAGCAGCCATCAT
	MT2_pTB1172_BsaI_f1	ATATGGTCTCCGCTTTGCCCTAGACAGCTGGGCG
	MT2_pTB1172_BsaI_f2	TATAGGTCTCGGCGATCGCAATGCGGGG
	MT2_pTB1172_BsaI_f3	TATAGGTCTCCGTATATCTCCTTCTTATTTTTTTGAATTCTG
	MT2_pTB1172_BsaI_r1	AGAGGGTCTCCCACCAGTTTAAACAATTCGAAAAGCCTGC
	MT2_pTB1172_BsaI_r2	ATATGGTCTCGATCTGCAGGAGATCCGGCT
	MT2_pTB1172_BsaI_r3	ATTAGGTCTCCATGTTTTCGCCAGATCCCGAA
sequencing of screening plasmids	MT_p15Aend_seq_f	CGAAAAGTGCCACCTGCTAGG
	MT_KanRend_seq_f	AGCTCGTTTCACGCTGAATATGG
	MT_ID364-BsaI_f	GATGGTCTCGAATGATCATCTGACCGTACGTTTC
	MT_ID364-BsaI_r	GATGGTCTCGTGGGCGCCTTTTTCGAACTGC
	pTAC_for	GAGCGGATAACAATTTCACACAGG
	GlnsTermbase_rev	AAAGCAGAAAAAACGCCGCTG
	MaH_54_ONBYRS_V263G_Rev	CTATTAACTGTCAAATCTCCACCAAATTTTTCTGGCC
	o_TB160_MjTyrRS_f	GCGAGGAAGAGTTAAGAGAGG
	sfGFP-SA_f	TGGATGAACTGTACAAAAGCG
	IleRSrevSeq	GAGTGACCAATATGAATGCTGC
	ileRS fw seq 4	TGGCGGAACACGCAGAAATC
	MaH_55_pBAB_tRNA_Seq	ATCCATGTTGCTGTTCAGACG
	o_TB66_sfGFP-StrepII_r	GGCTTTGTTAGCAGCCGGATCTCCTGCAGATTATTTTTCGAACTGCGGGTGGCTCCAAGCGCTTTTGTACAGTTCATCCATACCATGCG
	T5 Promoter	CCCGAAAAGTGCCACCTG
	GlnS prom forw	CATCAATCATCCCCATAATC
	any_pMec_glnSTer_rev	CTCAATATATTGCAGAGATCATG
insert of mutation A579P into genomic *ileS*	MT_IleRS(GCG579CCA)_f	GATCTCCACCCCAATGAAAGGC
	MT_IleRS(A579P)seq_r	CCATTTCACCGGTGTAGTCGG
	MT_IleRS(GCG579CCA)_r	GCC TTT CAT TGG GGT GGA GAT C
	MT_CRISPRs_IRS(579P)BsaI_f	TTAAGGTCTCACCCAATGAAGGGTAAAGCGCCGTATCG
	MT_CRISPRs_IRS_BsaI_r	TATTGGTCTCTACACAAGCCGTGGGTCAGTACCTGACG
	MT_CRISPRr_IRS_r	CTACGACGCGCCACACTG
	MT_CRISPRr_IRS_BasI_f	ATTAGGTCTCAAGTAATGAAGGGTAAAGCGCCGTATCG

	MT_CRISPRl_IRS(A579P)_BsaI_r	AATAGGTCTCATGGGGTGGAGATCATTAGGGAAGACATGAAC
	MT_CRISPRl_IRS_f	GATCATCTTCCGCGCGACG
	MT_CRISPR_seqIleR	CTCTTACGTGCCGATCAACGTC
	MT_CRISPR_seqCmR	CATCTGCCGACATGGAAGCCATC
	MT_CmRN20PAM_BsaI_r	TATAGGTCTCATACTTCGGTTCGATGGACTAACATGGGAATTAGCCATGGTCCATATG
	IleRSrevSeq2	CTTCTCACCTTCGGCTTTAC
	MT_CRISPRm_IRS_BsaI_r	TATAGGTCTCACTCATCAGGCAAACTTACGTTTTTCACCG
	MT_CRISPRr_IRS_BsaI_f	ATTAGGTCTCAAGTATGAGTCAATCGATCTGTTCAACAGGG
	MT_CRISPRr_IRS_r	CATCCGGCAGGGTTTATTATTGTTTTTTC
	MT_CRISPRr_IRS_r2	CCAATTTCTGGCTCGCCTGC
	MT_CRISPRs_IRS_BsaI_f	ATTAGGTCTCATGAGTCAATCGATCTGTTCAACAGGG
	MT_CRISPRs_IRS_BsaI_r	TTAAGGTCTCTACACCCAGCCACAGCCAGCGTAG
	MT_CRISPR seq_f1	CGTTGCAGACTATAACGACGCAC
	MT_CRISPR seq_r2	GTG GGC ACT TCT CAC CTT CGG C
	pKD46_AraC_For1	CAAATAGGGGTTCCGCGCAC
	MT_pKD3CmR_BsaI_f	AATAGGTCTCTGTGTAGGCTGGAGCTGCTTCG
	MT_pKD3CmR_BsaI_r	AATAGGTCTCTTCACCATGGGAATTAGCCATGGTCCATATG
	MT_cfsfGFP(C48S)_f	GTT CAT CTC TAC TAC TGG TAA ACT GC
cloning of cysteine-free sfGFP	MT_cfsfGFP(C70M)_r	CGG ATA ACG AGC AAA CAT CTG AAC

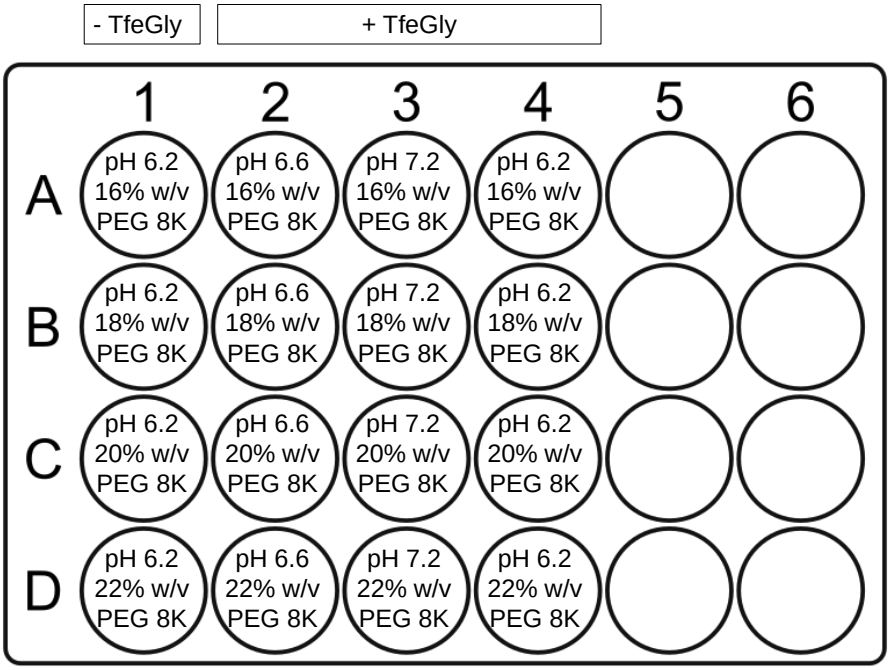
8.5 Crystallization conditions for optimization

Crystallization conditions for screening is according to the supplier's information.
For optimization experiments, pH value, salts and concentration of precipitants were varied to fine-tune the crystallization conditions as followed.

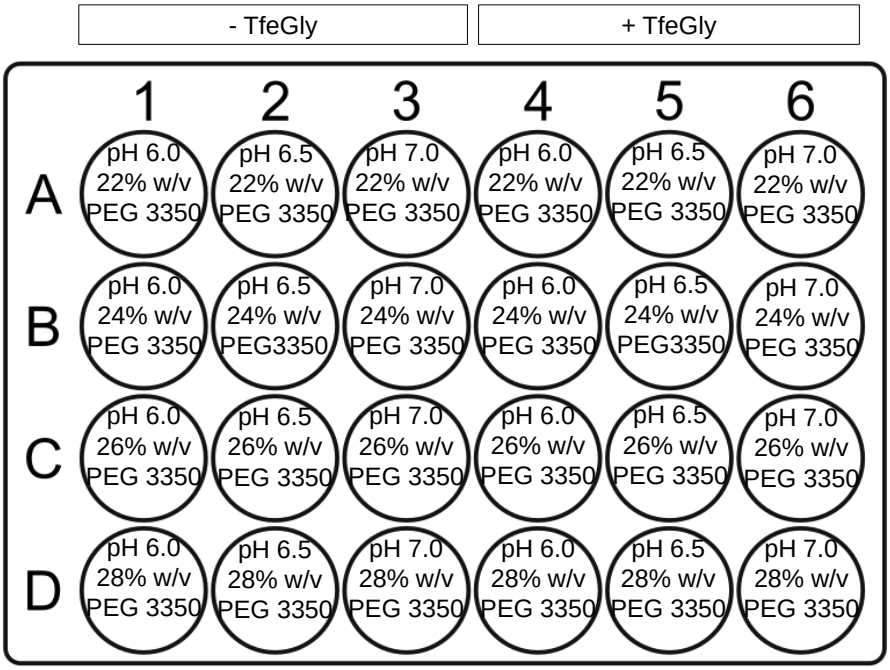
Optimization in 24-well plates

Optimization experiments were conducted based on the promising hits from screening. The letter and number denote the position of the condition from the screening in a 96-well. Composition of original condition used for fine-tuning is on the left, while the changes are shown in the appropriate wells.

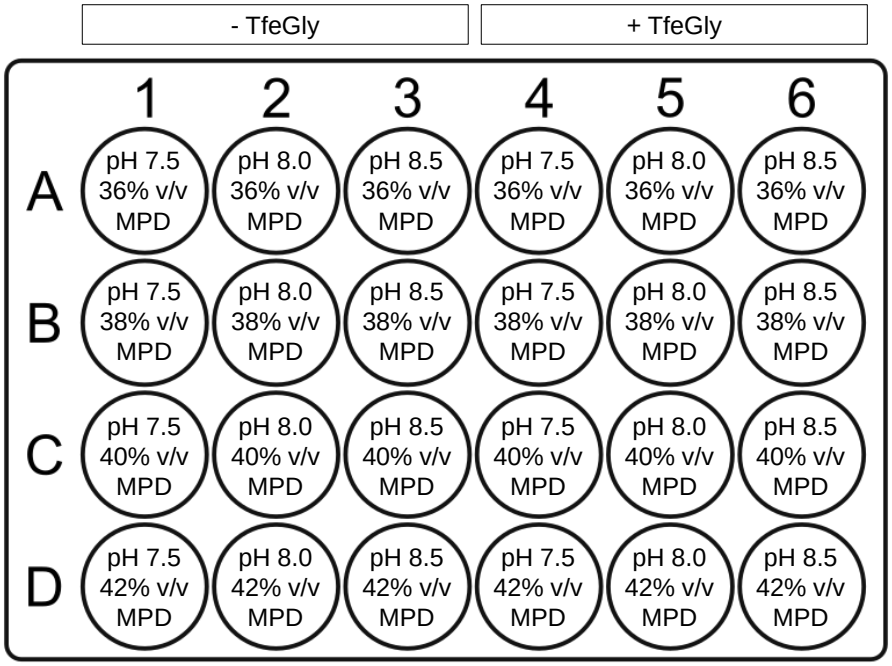
JBScreen Basic HTS_D9
20 mg/mL IleRS
± 8 mM TfeGly
100 mM MES, pH 6.5
200 mM $Mg(CH_3CO_2)_2$
20 % w/v PEG 8K



Hampton Index_ G11
20 mg/mL IleRS
± 8 mM TfeGly
100 mM BIS-TRIS, pH 6.5
200 mM $MgCl_2$
25 % w/v Polyethylene glycol 3350



JBScreen JCSG++_D8
20 mg/mL IleRS
± 8 mM TfeGly
100 mM TRIS, pH 8.0
40 % v/v 2-Methyl-2,4-pentanediol



Optimization in 48-well plates

Screening conditions resulting nice crystals are abbreviated in the first cell of each optimization plate as followed: Basic = JBScreen Basic, PACT = JBScreen PACT++ 1-4, and JCSG = JBScreen JCSG++ (Jena Bioscience, Jena, Germany). The letter and number denote the position of the condition from the screening in a 96-well. Changes are highlighted in bold.

Basic D9	15 mg/mL IleRS + 8 mM TfeGly			20 mg/mL IleRS + 8 mM TfeGly		
	1	2	3	4	5	6
A	0.1 M MES, pH **6.0** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **17 % w/v** PEG 8K
B	0.1 M MES, pH **6.0** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **18 % w/v** PEG 8K
C	0.1 M MES, pH **6.0** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **19 % w/v** PEG 8K
D	0.1 M MES, pH **6.0** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **20 % w/v** PEG 8K
E	0.1 M MES, pH **6.0** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **21 % w/v** PEG 8K
F	0.1 M MES, pH **6.0** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **22 % w/v** PEG 8K
G	0.1 M MES, pH **6.0** 0.2 M MgOAc **23 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **23 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **23 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **23 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **23 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **23 % w/v** PEG 8K
H	0.1 M MES, pH **6.0** 0.2 M MgOAc **24 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **24 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **24 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **24 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **24 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **24 % w/v** PEG 8K

JCSG C9	15 mg/mL IleRS			15 mg/mL IleRS + 8 mM TfeGly		
	1	2	3	4	5	6
A	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **20 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **20 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **20 %** w/v PG	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **20 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **20 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **20 %** w/v PG
B	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **21 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **21 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **21 %** w/v PG	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **21 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **21 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **21 %** w/v PG
C	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **22 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **22 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **22 %** w/v PG	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **22 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **22 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **22 %** w/v PG
D	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **23 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **23 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **23 %** w/v PG	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **23 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **23 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **23 %** w/v PG
E	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **24 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **24 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **24 %** w/v PG	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **24 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **24 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **24 %** w/v PG
F	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **25 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **25 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **25 %** w/v PG	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **25 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **25 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **25 %** w/v PG
G	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **26 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **26 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **26 %** w/v PG	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **26 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **26 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **26 %** w/v PG
H	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **27 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **27 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **27 %** w/v PG	0.1 M NaKP, pH **5.8** 10 % v/v glycerol **27 %** w/v PG	0.1 M NaKP, pH **6.2** 10 % v/v glycerol **27 %** w/v PG	0.1 M NaKP, pH **6.6** 10 % v/v glycerol **27 %** w/v PG

PACT D10	15 mg/mL IleRS			15 mg/mL IleRS + 8 mM TfeGly		
	1	2	3	4	5	6
A	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **19 %** w/v PEG 6k
B	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **20 %** w/v PEG 6k
C	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **21 %** w/v PEG 6k
D	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **22 %** w/v PEG 6k
E	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **23 %** w/v PEG 6k
F	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **24 %** w/v PEG 6k
G	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **25 %** w/v PEG 6k
H	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **26 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **26 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **26 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MgCl_2$ **26 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MgCl_2$ **26 %** w/v PEG 6k	0.1 M TRIS, pH **8.5** 200 mM $MgCl_2$ **26 %** w/v PEG 6k

PACT D10	15 mg/mL IleRS			15 mg/mL IleRS + 8 mM TfeGly		
	1	2	3	4	5	6
A	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **18 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MnCl_2$ **18 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **18 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **18 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MnCl_2$ **18 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **18 %** w/v PEG 6k
B	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MnCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MnCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **19 %** w/v PEG 6k
C	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MnCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MnCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **20 %** w/v PEG 6k
D	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MnCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MnCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **21 %** w/v PEG 6k
E	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MnCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MnCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **22 %** w/v PEG 6k
F	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MnCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MnCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **23 %** w/v PEG 6k
G	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **75** 200 mM $MnCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **75** 200 mM $MnCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **24 %** w/v PEG 6k
H	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MnCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $MnCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $MnCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $MnCl_2$ **25 %** w/v PEG 6k

PACT D10	15 mg/mL IleRS			15 mg/mL IleRS + 8 mM TfeGly		
	1	2	3	4	5	6
A	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **18 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $CoCl_2$ **18 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **18 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **18 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $CoCl_2$ **18 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **18 %** w/v PEG 6k
B	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $CoCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $CoCl_2$ **19 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **19 %** w/v PEG 6k
C	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $CoCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $CoCl_2$ **20 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **20 %** w/v PEG 6k
D	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $CoCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $CoCl_2$ **21 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **21 %** w/v PEG 6k
E	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $CoCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $CoCl_2$ **22 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **22 %** w/v PEG 6k
F	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $CoCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $CoCl_2$ **23 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **23 %** w/v PEG 6k
G	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **75** 200 mM $CoCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **75** 200 mM $CoCl_2$ **24 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **24 %** w/v PEG 6k
H	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $CoCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **7.0** 200 mM $CoCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **7.5** 200 mM $CoCl_2$ **25 %** w/v PEG 6k	0.1 M TRIS, pH **8.0** 200 mM $CoCl_2$ **25 %** w/v PEG 6k

PACT B11	15 mg/mL IleRS			15 mg/mL IleRS + 8 mM TfeGly		
	1	2	3	4	5	6
A	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **18 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **18 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **18 %** w/v PEG 6k	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **18 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **18 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **18 %** w/v PEG 6k
B	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **19 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **19 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **19 %** w/v PEG 6k	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **19 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **19 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **19 %** w/v PEG 6k
C	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **20 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **20 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **20 %** w/v PEG 6k	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **20 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **20 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **20 %** w/v PEG 6k
D	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **21 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **21 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **21 %** w/v PEG 6k	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **21 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **21 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **21 %** w/v PEG 6k
E	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **22 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **22 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **22 %** w/v PEG 6k	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **22 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **22 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **22 %** w/v PEG 6k
F	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **23 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **23 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **23 %** w/v PEG 6k	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **23 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **23 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **23 %** w/v PEG 6k
G	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **24 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **24 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **24 %** w/v PEG 6k	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **24 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **24 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **24 %** w/v PEG 6k
H	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **25 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **25 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **25 %** w/v PEG 6k	0.1 M MES, pH **5.5** 200 mM $CaCl_2$ **25 %** w/v PEG 6k	0.1 M MES, pH **6.0** 200 mM $CaCl_2$ **25 %** w/v PEG 6k	0.1 M MES, pH **6.5** 200 mM $CaCl_2$ **25 %** w/v PEG 6k

Basic D9	20 mg/ml H6-TEV-IleRS + 8 mM TfeGly			20 mg/ml H6-TEV-IleRS + 8 mM TfeGly + 5 mM AMP-PNP + 1 mM Abu-2AA		
	1	2	3	4	5	6
A	0.1 M MES, pH **6.0** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **17 % w/v** PEG 8K
B	0.1 M MES, pH **6.0** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **18 % w/v** PEG 8K
C	0.1 M MES, pH **6.0** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **19 % w/v** PEG 8K
D	0.1 M MES, pH **6.0** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **20 % w/v** PEG 8K
E	0.1 M MES, pH **6.0** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **21 % w/v** PEG 8K
F	0.1 M MES, pH **6.0** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **22 % w/v** PEG 8K
G	0.1 M MES, pH **6.0** 0.2 M MgOAc **23 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **23 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **23 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **23 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **23 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **23 % w/v** PEG 8K
H	0.1 M MES, pH **6.0** 0.2 M MgOAc **24 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **24 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **24 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **24 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **24 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **24 % w/v** PEG 8K

PACT C10	20 mg/ml H6-TEV-IleRS + 8 mM TfeGly + 5 mM AMP-PNP			20 mg/ml H6-TEV-IleRS + 8 mM TfeGly + 5 mM AMP-PNP + 1 mM Abu-2AA		
	1	**2**	**3**	**4**	**5**	**6**
A	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **15 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **15 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **15 % w/v** PEG 6K	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **15 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **15 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **15 % w/v** PEG 6K
B	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **16 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **16 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **16 % w/v** PEG 6K	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **16 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **16 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **16 % w/v** PEG 6K
C	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **17 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **17 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **17 % w/v** PEG 6K	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **17 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **17 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **17 % w/v** PEG 6K
D	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **18 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **18 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **18 % w/v** PEG 6K	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **18 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **18 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **18 % w/v** PEG 6K
E	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **19 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **19 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **19 % w/v** PEG 6K	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **19 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **19 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **19 % w/v** PEG 6K
F	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **20 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **20 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **20 % w/v** PEG 6K	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **20 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **20 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **20 % w/v** PEG 6K
G	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **21 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **21 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **21 % w/v** PEG 6K	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **21 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **21 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **21 % w/v** PEG 6K
H	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **22 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **22 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **22 % w/v** PEG 6K	0.1 M HEPES, pH **6.8** 0.2 M $MgCl_2$ **22 % w/v** PEG 6K	0.1 M HEPES, pH **7.0** 0.2 M $MgCl_2$ **22 % w/v** PEG 6K	0.1 M HEPES, pH **7.4** 0.2 M $MgCl_2$ **22 % w/v** PEG 6K

Basic D9	20 mg/ml IleRS + 8 mM TfeGly + 5 mM AMP-PNP + 1 mM Abu-2AA					
	1	**2**	**3**	**4**	**5**	**6**
A	0.1 M MES, pH **6.5** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **15 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **15 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **15 % w/v** PEG 8K	0.1 M HEPES, pH **7.5** 0.2 M MgOAc **15 % w/v** PEG 8K	0.1 M HEPES, pH **8.0** 0.2 M MgOAc **15 % w/v** PEG 8K
B	0.1 M MES, pH **5.5** 0.2 M MgOAc **16 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **16 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **16 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **16 % w/v** PEG 8K	0.1 M HEPES, pH **7.5** 0.2 M MgOAc **16 % w/v** PEG 8K	0.1 M HEPES, pH **8.0** 0.2 M MgOAc **16 % w/v** PEG 8K
C	0.1 M MES, pH **5.5** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M HEPES, pH **7.5** 0.2 M MgOAc **17 % w/v** PEG 8K	0.1 M HEPES, pH **8.0** 0.2 M MgOAc **17 % w/v** PEG 8K
D	0.1 M MES, pH **5.5** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M HEPES, pH **7.5** 0.2 M MgOAc **18 % w/v** PEG 8K	0.1 M HEPES, pH **8.0** 0.2 M MgOAc **18 % w/v** PEG 8K
E	0.1 M MES, pH **5.5** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M HEPES, pH **7.5** 0.2 M MgOAc **19 % w/v** PEG 8K	0.1 M HEPES, pH **8.0** 0.2 M MgOAc **19 % w/v** PEG 8K
F	0.1 M MES, pH **5.5** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M HEPES, pH **7.5** 0.2 M MgOAc **20 % w/v** PEG 8K	0.1 M HEPES, pH **8.0** 0.2 M MgOAc **20 % w/v** PEG 8K
G	0.1 M MES, pH **5.5** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M HEPES, pH **7.5** 0.2 M MgOAc **21 % w/v** PEG 8K	0.1 M HEPES, pH **8.0** 0.2 M MgOAc **21 % w/v** PEG 8K
H	0.1 M MES, pH **5.5** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M MES, pH **6.0** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M MES, pH **6.5** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M MES, pH **7.0** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M HEPES, pH **7.5** 0.2 M MgOAc **22 % w/v** PEG 8K	0.1 M HEPES, pH **8.0** 0.2 M MgOAc **22 % w/v** PEG 8K

Statutory Declaration

I hereby declare that I have written the present thesis independently, without enlisting any external assistance, and only using the specified aids.

Berlin, May 2022

Tuyet Mai Thi To