

Julia Schmidt, BSc

**Towards the site-specific incorporation of  
 $N^G, N^G$ -dimethylarginine with mutated  
pyrrolysyl-tRNA synthetases into eGFP in *E. coli***

**Master Thesis**

to achieve the university degree of

Diplom-Ingenieur

Master degree programme: Biotechnology

Submitted to

**Graz University of Technology**

Institute of Molecular Biotechnology

Supervisor

Ao.Univ.-Prof. Mag.rer.nat. Dr.rer.nat Anton Glieder

Institute of Molecular Biotechnology

Dipl.-Ing. Dr. techn. Birgit Wiltschi

Austrian Center of Industrial Biotechnology

Graz, April 2021

## **Affidavit**

I declare that I have authored this thesis independently, that I have not used other than the declared sources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to UniGRAZonline and TUGRAZonline is identical to the present master thesis.

Date: \_\_\_\_\_

Signature: \_\_\_\_\_

## **Acknowledgements**

I express my special thanks to Dr. Birgit Wiltschi for giving me the opportunity to work on this project and supervising me. With her expertise in the field of synthetic biology she always provided advice and helped me to overcome obstacles. Through the many fruitful discussions we had, I succeeded in working precisely and logically consistent.

I want to thank Prof. Dr. Anton Glieder for officially supervising me through my master thesis. The participation in the project enabled me to expand my biotechnological knowledge and made me confident in carrying out many experiments.

My gratitude also goes to my colleagues. Dajana Kolanovic, MSc, Dipl. Ing. Alena Voit and Rajeev Pasupuleti, MSc helped whenever I needed advice or faced challenges. Their experience and knowledge greatly facilitated the practical work and the finding of new solutions. I want to thank my team members Bernhard Maierhofer, BSc and Markus Braun, BSc for such a pleasant team atmosphere and the numerous scientific discussions we had. It has been a real pleasure working with all the group members.

Finally, I want to express great thanks to my parents for supporting and encouraging me in each section of my studies and throughout the time writing this thesis.

# Table of Contents

Affidavit .....	2
Acknowledgements .....	3
Abbreviations .....	6
Kurzfassung.....	8
Abstract .....	9
1. Introduction .....	10
1.1 Amino acids and the standard genetic code.....	10
1.2 Translation of proteins and aminoacylation .....	10
1.3 Protein engineering using non-canonical amino acids .....	11
1.4 Expansion of the genetic code enables site-specific incorporation of ncAAs.....	12
1.5 Pyrrolysyl-tRNA synthetase/tRNA <sub>CUA</sub> <sup>Pyl</sup> as an orthogonal pair .....	12
1.6 Screening assays for orthogonal pairs .....	13
1.7 Site-specific incorporation of <i>N</i> <sup>G</sup> , <i>N</i> <sup>G</sup> -dimethylarginine into eGFP .....	14
2. Aim of the thesis .....	17
3. Materials.....	18
3.1 Chemicals and components .....	18
3.2 Stocks, solutions and buffers .....	18
3.3 Liquid media.....	21
3.4 Solid media .....	23
3.5 Marker .....	23
3.6 Kits.....	23
3.7 Bacterial strains .....	23
3.8 Enzymes.....	25
4. Methods.....	26
4.1 Plasmids.....	26
4.2 Preparation of competent cells .....	26

4.3	Agarose gel electrophoresis.....	27
4.4	Cloning.....	27
4.5	Transformation of <i>E. coli</i> Top10 F' and <i>E. coli</i> BL21(DE3) and transformation efficiency.....	28
4.6	Ligation.....	29
4.7	Restriction analysis with <i>EcoRI</i> and <i>HindIII</i> .....	29
4.8	Sequencing.....	29
4.9	Expression of <i>E. coli</i> BL21(DE3) and fluorescence assay.....	30
4.10	Protein expression analysis.....	31
4.11	SDS-polyacrylamide gel electrophoresis.....	31
4.12	Chloramphenicol growth assay.....	32
5.	Results and Discussion.....	34
5.1	Construction of the plasmid library.....	34
5.1.1	Preparation of plasmid backbone pT7x3-MmOP eGFP Y40am.....	34
5.1.2	Preparation of inserts.....	36
5.1.3	Ligation of backbone vectors and inserts.....	39
5.2	Fluorescence assay with ADMARS variants.....	42
5.2.1	Cultivation in 96-deep-well plates.....	43
5.2.2	Cell density and fluorescence assay.....	44
5.3	Fluorescence assay with SDLSL mutant.....	48
5.4	SDS-PAGE.....	51
5.5	Chloramphenicol growth assay.....	54
6.	Conclusion.....	61
7.	Outlook.....	64
	Supplementary figures.....	66
	List of figures, tables and equations.....	71
	References.....	73

## Abbreviations

AA	amino acid
aaRS	aminoacyl-tRNA synthetase
ADMA	$N^G, N^G$ -dimethylarginine
ADMARS	ADMA-tRNA synthetase
Arg	arginine
BocK	$N_\alpha$ -( <i>tert.</i> -butoxycarbonyl)-L-lysine
bp	base pair
cAA	canonical amino acid
CAT	chloramphenicol acetyltransferase
cfu	colony forming units
Cm	chloramphenicol
ddH <sub>2</sub> O	double-distilled water
dH <sub>2</sub> O	distilled water
DNA	deoxyribonucleic acid
DWP	deep-well plate
<i>E. coli</i>	<i>Escherichia coli</i>
eGFP	enhanced green fluorescence protein
HPLC-MS	high performance liquid chromatography-mass spectrometry
IPTG	isopropyl- $\beta$ -D-thiogalactopyranoside
Kan	kanamycin
LB	Luria-Bertani medium
<i>Ma</i>	<i>Methanomethylophilus alvus</i>
<i>Mm</i>	<i>Methanosarcina mazei</i>
ncAA	non-canonical amino acid
ONC	overnight culture
OP, o-pair	orthogonal pair

Pyl	pyrrolysine
PylRS	pyrrolysyl-tRNA synthetase
rpm	revolutions per minute
WT	wild type
YE	yeast extract

## Kurzfassung

ADMA-tRNA Synthetasen (ADMARSen) sind stark mutierte Aminoacyl-tRNA Synthetasen, die synthetisch entwickelt wurden, um nicht-kanonische Aminosäuren wie  $N^G, N^G$ -Dimethylarginin (ADMA) an der Position eines Amber Stopcodons einzubauen. Das Design der ADMARSen basierte auf der Wildtyp-Struktur der Pyrrolysyl-tRNA Synthetase von *Methanosarcina mazei* (*MmPylRS*) und sollte die Substratbindungskapazitäten des Enzyms beträchtlich erweitern. Das orthogonale Paar bestand aus ADMARS und der zugehörigen  $tRNA_{CUA}^{Pyl}$  (ADMA-OP). Wir generierten eine Plasmid Library aus ADMARS Expressionsvektoren. Die Backbone Vektoren enthielten ein Amber mutiertes Grün fluoreszierendes Protein-Gen (eGFP) und die zugehörige  $tRNA_{CUA}^{Pyl}$ .

Die ADMA-OPs wurden verwendet um ADMA ortsspezifisch in das eGFP als Zielprotein einzubauen. Um ein orthogonales Mutanten-Paar, das ADMA einbaut, zu finden, führten wir einen eGFP Fluoreszenz-Assay durch. Das eGFP-Gen enthielt eine Amber-Mutation an einer zulässigen Stelle stromaufwärts des Fluorophors. eGFP Expression und Fluoreszenz waren daher vom ADMA-Einbau an der Amber-Position abhängig. Bedauerlicherweise zeigte der Assay, dass eGFP nicht exprimiert und ADMA nicht in die Amber Position eingebaut wurde.

Um zu überprüfen, ob die ADMARSen exprimiert wurden, führten wir eine SDS-PAGE durch. ADMARS hatte ein berechnetes Molekulargewicht von 50.7 kDa und wir erwarteten eine detektierbare Bande am Gel, weil die Expression von einem T7 Promotor kontrolliert wurde. Die ADMARSen wurden jedoch nicht im erwünschten Ausmaß exprimiert. Die nächsten Schritte wären eine Wiederholung der Expressionsanalyse der ADMARSen, eine sorgfältige Analyse der ADMA-Aufnahme und der intrazellulären Akkumulation sowie, wenn nötig, eine Überarbeitung des Designs des orthogonalen Paares.



## Abstract

ADMA-tRNA synthetases (ADMARSs) are highly mutated aminoacyl-tRNA synthetases designed to incorporate non-canonical amino acids like  $N^G, N^G$ -dimethylarginine (ADMA) at the position of an amber stop codon. The design of the ADMARSs was based on the structure of wild type pyrrolysyl-tRNA synthetase from *Methanosarcina mazei* (*MmPylRS*) and intended to largely expand the substrate binding capacities of the enzyme. The orthogonal pair consisted of ADMARS and the cognate tRNA<sub>CUA</sub><sup>Pyl</sup> (ADMA-OP). We generated a library of ADMARS variant expression vectors. The backbone vectors harbored an amber mutated enhanced green fluorescent protein (eGFP) gene and the cognate tRNA<sub>CUA</sub><sup>Pyl</sup>.

ADMA-OPs were employed to incorporate ADMA site-specifically into the eGFP as a target protein. To select a mutant o-pair which incorporates ADMA, we performed an eGFP fluorescence assay. The eGFP gene carried an amber mutation at a permissive site upstream of the fluorophore. Therefore, eGFP expression and fluorescence were dependent on the incorporation of ADMA at the amber position. Unfortunately, the assay indicated that full-length eGFP was not expressed and ADMA was not incorporated at the amber position.

To explore whether ADMARSs were expressed, we performed SDS-PAGE. ADMARSs had a calculated molecular weight of 50.7 kDa and we expected them to be detectable on the gel because their expression was controlled by a T7 promoter. However, ADMARSs were not expressed to the desired extent. Next steps would be the repetition of the expression analysis of ADMARSs, the thorough analysis of ADMA uptake and intracellular accumulation and if necessary, the redesign of the orthogonal pair.

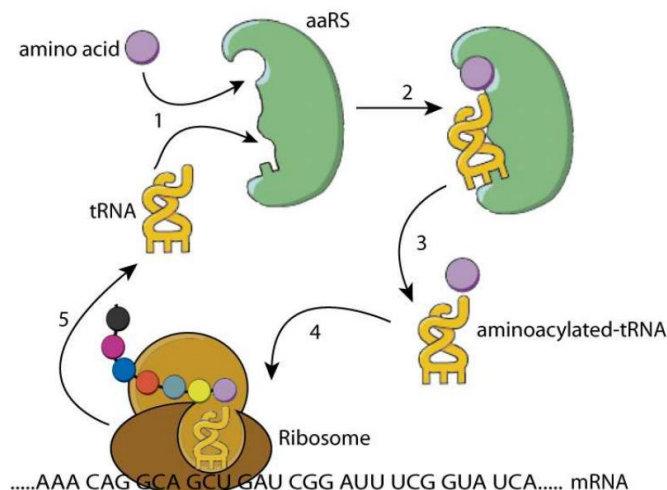
# 1. Introduction

## 1.1 Amino acids and the standard genetic code

Amino acids (AAs) are the structural units of peptides and proteins. In nature, 20 canonical amino acids (cAAs) are encoded by the standard genetic code and used for protein biosynthesis. The standard genetic code consists of 64 codons and each codon consists of three nucleobases. 61 of them are sense codons as they encode a cAA. Each cAA can be encoded by two to six codons. Methionine and tryptophan are the exception since they are only encoded by one codon each. The three remaining codons are “nonsense” or rather stop codons as they do not encode a cAA but terminate translation. These are called amber (UAG), ochre (UAA) and opal (UGA) codons. The information stored in triplet codons on DNA level is transcribed into messenger RNAs (mRNAs) which convey this information to the ribosome for translation. Not only mRNA is crucial for ribosomal translation, though. Transfer RNAs (tRNAs), aminoacyl-tRNA synthetases (aaRSs), the ribosome itself and protein factors such as elongation and release factors are of high importance<sup>1</sup>.

## 1.2 Translation of proteins and aminoacylation

Which AA is inserted at a certain position in the polypeptide chain depends on the pairing of the codon presented on the mRNA with the assigned aminoacyl-tRNA<sup>3</sup>. Aminoacyl-tRNA synthetases (aaRSs) are a large group of enzymes which catalyze the charging of a tRNA with the corresponding proteinogenic amino acid. This process is called aminoacylation of tRNAs<sup>2,3</sup> and is shown in Figure 1.



**Figure 1 Aminoacylation of tRNAs.** (1) Recognition of cognate AA and tRNA by the aaRS, (2) binding of the cognate AA and tRNA by the aaRS, (3) release of aminoacyl-tRNA, (4) transport of aminoacyl-tRNA to the ribosome for (5) protein translation and recycling of tRNA. Figure taken from Wang *et al.*<sup>7</sup>.

As illustrated in Figure 1, the appropriate AA is activated by the aaRS under consumption of ATP. An aminoacyl adenylate is formed which interacts specifically with the acceptor stem and the anticodon of the corresponding tRNA. AMP is cleaved off and an aminoacyl-tRNA is formed. It interacts with a translation elongation factor under consumption of GTP. The elongation factor transports the aminoacyl-tRNA to the A-site of the ribosome. The tRNA anticodon interacts with the corresponding mRNA codon and translation can proceed<sup>3</sup>. Translation of mRNA into polypeptides continues until a stop codon is recognized in the A-site of the ribosome. Since there are no cognate tRNAs for stop codons, they are recognized by release factors mimicking tRNAs. This leads to the termination of ribosomal translation<sup>4</sup>.

### 1.3 Protein engineering using non-canonical amino acids

The number of naturally occurring AAs is many times higher than the 20 cAAs that are used for protein biosynthesis<sup>1</sup>. Besides the 20 cAAs, there are further AAs that can be incorporated into proteins<sup>5</sup>. Two naturally occurring non-canonical amino acids (ncAAs) are selenocysteine (Sec) and pyrrolysine (Pyl), the 21<sup>st</sup> and 22<sup>nd</sup> amino acid, respectively. Sec and Pyl can both be ribosomally translated into proteins in selected organisms<sup>1</sup>. Sec is found in mammalian proteins and Pyl in proteins necessary for methylamine utilization e.g., in *Methanosarcina mazei* or *Methanomethylophilus alvus*<sup>1,6</sup>. To encode ncAAs, an expansion of the genetic code is necessary since the 64 codons of the standard genetic code are already used for the 20 cAAs<sup>1</sup>. Natural expansion of the genetic code leads to the encoding of AAs that introduce novel properties and functionalities into proteins<sup>5</sup>.

## 1.4 Expansion of the genetic code enables site-specific incorporation of ncAAs

To code for ncAAs, sense codons can be reassigned. This approach is mainly used for the residue-specific incorporation of ncAAs. Alternatively, codons that do not encode an AA yet (e.g., stop codons) can be assigned for an ncAA. This approach usually leads to the site-specific incorporation of ncAAs<sup>7</sup>. Site-specific implies that the ncAA gets incorporated in response to a defined codon<sup>8</sup>. Sec and Pyl are encoded by in-frame stop codons, namely opal (UGA) and amber (UAG) stop codon, respectively<sup>1</sup>. Besides opal and amber, the ochre (UAA) stop codon can also be used for the incorporation of ncAAs<sup>9</sup>. Among the three stop codons, the amber codon is the least frequently occurring stop codon in most organisms<sup>7</sup>. Being the rarest codon, it is most frequently used as an in-frame stop codon for ncAA incorporation<sup>10</sup>. To encode an ncAA, the stop codon must not trigger but suppress translation termination. This phenomenon is called stop codon suppression (SCS). SCS requires a suppressor tRNA carrying an anticodon for the stop codon and an aaRS with a substrate specificity for the ncAA<sup>1</sup>. Assigning an in-frame stop codon to an ncAA leads to a competition with the release factor<sup>15</sup>, which can strongly reduce the SCS efficiency<sup>12</sup>. Usually, the release factor recognizes the stop codon and causes the translation termination by releasing the nascent polypeptide chain<sup>11</sup>. To overcome the release factors action, the charged suppressor tRNA must be present at a level that allows it to compete with the release factor for the amber codon. The cognate aaRS must specifically and efficiently recognize the ncAA<sup>12</sup>. Whether the ncAA incorporation is favored over the translation termination and how efficient the ncAA is incorporated also depends on the translational context. The translational context includes both mRNA and polypeptide environment<sup>13,14</sup>. The mRNA environment is defined as the upstream and downstream bases of the stop codon. The polypeptide environment is defined as the amino acids preceding and succeeding the ncAA<sup>15</sup>. The incorporation efficiency indicates to what extent the ncAA was inserted into the polypeptide<sup>7</sup>. Factors like polypeptide structure, hydrophobicity or hydrophilicity as well as polypeptide size influence to what extent the ncAA fits into the polypeptide.

## 1.5 Pyrrolysyl-tRNA synthetase/tRNA<sub>CUA</sub><sup>Pyl</sup> as an orthogonal pair

AaRSs can be divided into Class I and Class II synthetases based on the structure of their active site. In class I synthetases, the active site contains a Rossmann dinucleotide-binding domain. The active site of class II synthetases contains an antiparallel beta-fold<sup>3</sup>. Evolutionary events

caused a certain incompatibility of aaRSs and tRNAs among species. For instance, an archaeal aaRS does not necessarily charge a bacterial tRNA with the corresponding amino acid. This builds the foundation for orthogonal aaRS/tRNA pairs (o-pairs) which can be used for the site-specific incorporation of ncAAs<sup>16</sup>. The best-studied example for such an o-pair is pyrrolysyl-tRNA synthetase/tRNA<sub>CUA</sub><sup>Pyl</sup> (PylRS/tRNA<sub>CUA</sub><sup>Pyl</sup>)<sup>17</sup>. Pyrrolysyl-tRNA synthetase (PylRS) is a class II synthetase<sup>17</sup> that has a natural specificity for Pyl and mediates the specific charging of tRNA<sub>CUA</sub><sup>Pyl</sup> with Pyl<sup>6</sup>. PylRS shows a low selectivity towards the tRNA anticodon but an outstandingly high substrate side chain promiscuity. PylRS recognizes also other ncAAs that have a similar structure to Pyl such as *N*<sub>α</sub>-(*tert*-butoxycarbonyl)-L-lysine (BocK). This high substrate tolerance is peculiar for an aaRS and makes PylRS an ideal mediator for site-specific incorporation of ncAAs<sup>17</sup>.

## 1.6 Screening assays for orthogonal pairs

To find a functional mutant of a PylRS/tRNA<sub>CUA</sub><sup>Pyl</sup> o-pair specific for an ncAA, a suitable selection or screening is necessary. A two-step selection procedure is often performed. It consists of a positive and a negative selection step constituting one selection round<sup>18</sup>. The concept behind this selection procedure is the linkage of the development of a desired phenotype with the suppression of a stop codon by an ncAA. If the stop codon is suppressed, the phenotype will be developed. If the stop codon is not suppressed, the phenotype will not be developed. Cells showing the desired phenotype in the positive selection are analyzed further in the negative selection. In general, two to three rounds of selection are performed<sup>19</sup>.

The positive-negative selection is based on an antibiotic resistance gene with mutated in-frame amber stop codons. It starts with a positive selection step. Cells transformed with the o-pair are grown in the presence of a drug or antibiotic, cAAs and the ncAA of interest. Cells survive when they express the antibiotic resistance gene which contains an amber codon at a permissive site. This is achieved by stop codon suppression and incorporation of an AA, either a cAA or the ncAA. Cells die when the o-pair is incapable of stop codon suppression and no AA is incorporated. At this stage mutants that incorporate both cAAs and ncAA are selected. The surviving cells from the first step are co-transformed with a toxic gene including an amber codon at a permissive site. The cells are grown in absence of the ncAA. Since the incorporation of a cAA at the amber position leads to the expression of the toxic protein, only those cells will survive which do not incorporate cAAs at the amber position. In the negative selection step, mutants that incorporate cAAs are sorted out<sup>20</sup>. The described methodology requires a co-

transformation of the plasmid with a toxic gene after positive selection, which is very time-consuming especially with many mutants. Therefore, dual positive/negative selection vectors harboring an antibiotic resistance gene are often used for both positive and negative selection<sup>19</sup>. Melançon *et al.*<sup>19</sup> used a *CAT*-uracil phosphoribosyltransferase (*UPRT*) fusion gene as a dual positive/negative selection marker. An amber stop codon is present at a permissive site in the *CAT* portion of the fusion gene. *UPRT* converts 5-fluorouracil to 5-fluoro-dUMP to inhibit cell death caused by thymidylate synthase. The positive selection step enriches mutants suppressing the amber codon in presence of both chloramphenicol (Cm) and the ncAA. The negative selection step sorts out mutants suppressing the amber codon and resisting cell death in the absence of the ncAA but the presence of 5-fluorouracil.

Another assay that can be used to analyze mutants of a PylRS/tRNA<sub>CUA</sub><sup>Pyl</sup> o-pair is the green fluorescent protein (GFP)-based screening. Expression of GFP is used for both positive and negative selection. An antibiotic resistance gene and the GFP gene depend on amber suppression for their expression. In the positive selection step, cells are grown in presence of an antibiotic and the ncAA. Fluorescing cells are enriched and processed in the next step, the negative selection<sup>21</sup>. Cells are grown in absence of both antibiotic and ncAA. GFP is expressed when a mutant o-pair is present that incorporates any cAA. GFP is not expressed when the o-pair specifically incorporates the ncAA. These nonfluorescent cells are analyzed in the next screening round. Again, two to three rounds of screening are performed<sup>22</sup>.

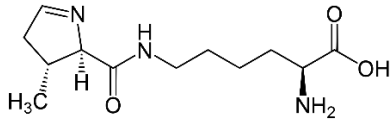
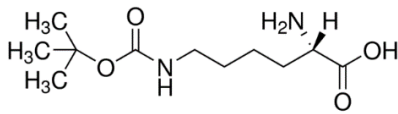
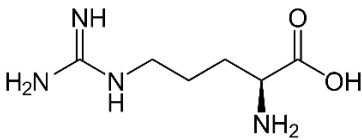
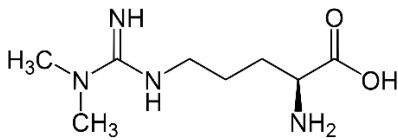
## 1.7 Site-specific incorporation of $N^G, N^G$ -dimethylarginine into eGFP

We aimed to exploit the high substrate side chain promiscuity of PylRS to incorporate  $N^G, N^G$ -dimethylarginine (ADMA) site-specifically into enhanced GFP (eGFP) as a target protein. We employed strongly mutated pyrrolysyl-tRNA synthetases from *Methanosarcina mazei* (*MmPylRSs*). For analysis of the mutants, we followed an adapted GFP-based screening assay<sup>37</sup>. Our plasmid library included an eGFP gene carrying an amber mutation at a permissive site upstream of the fluorophore. Therefore, eGFP expression and fluorescence were dependent on the incorporation of ADMA at the amber position. eGFP wild type was processed as a control. We only performed a positive selection by supplementing the media with ADMA. Monitoring the increase of eGFP fluorescence over time allowed us to identify mutants incorporating ADMA at the amber position within the eGFP gene.

In addition to ADMA, the amino acids BocK and arginine (Arg) were employed. BocK is a commercially available surrogate for pyrrolysine<sup>23</sup> and was used as a model substrate for

*MmPylRS*. Arg was used to verify that it is not accepted by the strongly mutated *MmPylRS*s. The structures of Pyl, BocK, Arg and ADMA are shown in Table 1.

**Table 1 Amino acids employed in this thesis.** Name, abbreviation, structure and molecular weight in [g/mol] are given for each AA.

Amino acid	Abbreviation	Structure	Molecular weight [g/mol]
pyrrolysine	Pyl		255.31
<i>N</i> <sub>α</sub> -( <i>tert</i> -butoxycarbonyl)-L-lysine	BocK		246.31
L-arginine	Arg		174.2
<i>N</i> <sup>G</sup> , <i>N</i> <sup>G</sup> -dimethylarginine	ADMA		202.25

Pyl and BocK are lysine-derivatives<sup>17</sup>. In Pyl, a pyrroline residue is attached to the *N*<sup>6</sup> (Table 1, first row) and in BocK, a *tert*-butyloxycarbonyl protecting group is attached to the *N*<sup>1</sup> (second row). ADMA is an arginine-derivative with two methyl-groups at the guanidine residue (last row).

Arginine methylation is a natural posttranslational modification of mammalian proteins and is catalyzed by a number of protein arginine methyltransferases. It is involved in biological processes like cell growth, proliferation, differentiation and transformation<sup>24</sup>. Arginine methylation also occurs in RNA-binding proteins (RBPs) which interact with RNA and assist in the formation of the ribonucleoprotein complex<sup>24,25</sup>. Arginine can be methylated once (mono-methylarginine) or twice (symmetrical and asymmetrical dimethylarginine)<sup>24</sup>. This thesis was part of a BioTechMed-Graz Flagship project which proposed that protein arginine methylation plays a critical role in the modulation of nuclear import, solubility and phase separation of

RBPs<sup>26</sup>. To study the methylation of RBPs, site-specific incorporation of methylated arginine into selected proteins would be of great value. Therefore, we aimed to incorporate ADMA as an ncAA site-specifically into eGFP as a target protein.



## 2. Aim of the thesis

The aim of this thesis was to discover a mutated orthogonal pair capable of site-specifically incorporating  $N^G, N^G$ -dimethylarginine (ADMA) as a non-canonical amino acid into the enhanced green fluorescent protein (eGFP). We constructed a novel plasmid library by ligating a backbone vector with synthetic cloning fragments. The fragments harbored highly mutated pyrrolysyl-tRNA synthetases from *Methanosarcina mazei* (*MmPylRSs*). The mutated *MmPylRSs* were designed by Markus Braun<sup>33</sup>. By conducting fluorescence assays we wanted to monitor functionally expressed eGFPs. Since we did not detect eGFP fluorescence values above background, we also analyzed the expression of the mutated *MmPylRSs* with SDS-PAGE.

Furthermore, we repeated a chloramphenicol growth assay with a plasmid including a mutated pyrrolysyl-tRNA synthetase from *Methanomethylophilus alvus* carrying the mutations Y126S M129D V168L Y206S W239L (*MaPylRS* SDL<sup>SL</sup>). The growth assay was performed to analyze whether ADMA could be incorporated at the in-frame amber position of a mutated chloramphenicol acetyltransferase (*CAT*) gene and whether *CAT* could be functionally expressed.

### **3. Materials**

#### **3.1 Chemicals and components**

Chemicals and components used during this master thesis were purchased at Carl Roth, VWR, Sigma Aldrich, Iris Biotech, Thermo Scientific, abcr or fluorochem unless indicated differently. Filter sterilization of liquids was done with a 0.22  $\mu\text{m}$  or 0.45  $\mu\text{m}$  tip filter from Merck Millipore.

#### **3.2 Stocks, solutions and buffers**

##### **T10E1 buffer (pH 8)**

T10E1 buffer was used as a DNA solvent. For the preparation of the T10E1 buffer 60.57 mg Tris-HCl and 18.61 mg EDTA were mixed in 50 mL ddH<sub>2</sub>O and adjusted to a pH of 8.

##### **TfB1 and TfB2 buffer**

TfB1 and TfB2 buffers were used for the preparation of chemocompetent cells. For the preparation of 500 mL TfB1 buffer 1.47225 g KCH<sub>3</sub>COO, 3.7275 g KCl, 0.55493 g CaCl<sub>2</sub> and 65 mL 13% (v/v) glycerol were mixed in dH<sub>2</sub>O. For the preparation of 500 mL TfB2 buffer, 0.37275 g KCl, 0.55493 g CaCl<sub>2</sub>, 1.04635 g MOPS and 65 mL 13% (v/v) glycerol were mixed in dH<sub>2</sub>O.

##### **50 g/L kanamycin stock**

The kanamycin stocks with a final concentration of 50 g/L were prepared by dissolving kanamycin sulfate in dH<sub>2</sub>O. The stock solution was then filter sterilized, dispensed in 1 mL aliquots and stored in 1.5 mL tubes at -20 °C.

##### **100 g/L ampicillin stock**

The ampicillin stocks with a final concentration of 100 g/L were prepared by dissolving ampicillin in dH<sub>2</sub>O. The stock solutions were then filter sterilized, aliquoted in 1.5 mL tubes and stored at -20 °C.

### **60% (v/v) glycerol**

To prepare 100 mL of a 60% (v/v) glycerol stock, 60 mL of 100% glycerol were mixed with 40 mL of dH<sub>2</sub>O. For 10% and 80% (v/v) glycerol stocks it was done accordingly. Then the stock was autoclaved and stored at room temperature.

### **1 M glucose stock**

The 1 M glucose stock was prepared by weighing D-(+)-glucose monohydrate and dissolving it in dH<sub>2</sub>O. Afterwards the stock solution was autoclaved and stored at room temperature.

### **1 M magnesium sulfate stock**

The 1 M MgSO<sub>4</sub> stock was prepared by dissolving MgSO<sub>4</sub>•7H<sub>2</sub>O in dH<sub>2</sub>O. Afterwards the stock was autoclaved and stored at room temperature.

### **1 g/L calcium chloride stock**

The CaCl<sub>2</sub> stock with a final concentration of 1 g/L was prepared by dissolving CaCl<sub>2</sub> in dH<sub>2</sub>O. The solution was filter sterilized and stored at room temperature.

### **0.5 g/L biotin stock**

The biotin stock with a final concentration of 0.5 g/L was prepared by dissolving biotin in dH<sub>2</sub>O. Afterwards the solution was filter sterilized and stored at 6 °C.

### **1 g/L thiamine hydrochloride stock**

The thiamine hydrochloride stock was prepared by dissolving thiamine hydrochloride in dH<sub>2</sub>O. Afterwards the solution was filter sterilized and stored at 6 °C.

### **20% (w/v) arabinose stock**

The 20% (w/v) arabinose stock was prepared by dissolving L(+)-arabinose in dH<sub>2</sub>O. Afterwards the solution was filter sterilized and stored at room temperature.

### **Chloramphenicol stocks**

The chloramphenicol (Cm) stocks with a final concentration of 25 g/L were prepared by dissolving Cm in 100% ethanol. Stocks with a final concentration of 50 g/L were prepared

accordingly. The stock solution was then filter sterilized, dispensed in 1 mL aliquots and stored in 1.5 mL tubes at -20 °C.

### **5x M9-salt stock**

The 5x-M9 salt stock was either prepared by dissolving 56.4 g of M9-salts in 1 L of dH<sub>2</sub>O or by combining the chemicals listed in Table 2 and dissolving them in dH<sub>2</sub>O to a final volume of 1 L. Afterwards the stock solution was autoclaved and stored at room temperature.

**Table 2 Components and final concentrations in the 5x M9-salt stock.**

<b>Component</b>	<b>Final concentration</b>
Na <sub>2</sub> HPO <sub>4</sub>	33.91 g/L
KH <sub>2</sub> PO <sub>4</sub>	15.1 g/L
NaCl	2.56 g/L
NH <sub>4</sub> Cl	5.01 g/L

### **0.1 M NaP<sub>i</sub> buffer (pH 8)**

To prepare 50 mL of a 0.1 M NaP<sub>i</sub> buffer (pH 8), 4.66 mL of 1 M Na<sub>2</sub>HPO<sub>4</sub> were mixed with 0.34 mL of 1 M NaH<sub>2</sub>PO<sub>4</sub>. The 1 M Na<sub>2</sub>HPO<sub>4</sub> stock was prepared by dissolving 2.13 g Na<sub>2</sub>HPO<sub>4</sub>·2·H<sub>2</sub>O in 15 mL of dH<sub>2</sub>O. The 1 M NaH<sub>2</sub>PO<sub>4</sub> was prepared by dissolving 0.24 g NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O in 2 mL of dH<sub>2</sub>O.

### **Glycerol stocks**

Each glycerol stock contained 500 µL of the regarding cell culture and 500 µL of 60% glycerol. They were stored at -80 °C.

### **0.0952 g/mL MgCl<sub>2</sub> stock**

The MgCl<sub>2</sub> stock was prepared by dissolving MgCl<sub>2</sub>·4H<sub>2</sub>O in dH<sub>2</sub>O. The solution was filter sterilized and stored at 4 °C.

### **5x SDS sample buffer stock**

The SDS sample buffer was available as a 5x stock. The final concentrations in the 1x SDS sample buffer were 50 mM Tris-HCl (pH 6.8), 2% SDS, 10% glycerol, 1% beta-mercaptoethanol, 12.5 mM EDTA and 0.02% bromophenol blue.

### **5 mM IPTG stock**

The 1 M IPTG stock was diluted with ddH<sub>2</sub>O to a final concentration of 5 mM.

### **50 mM BocK stock**

The 50 mM BocK stocks were always prepared freshly before application and depending on the volume required for the screening. The BocK stock with a final concentration of 50 mM was prepared by dissolving Boc-Lys-OH (Fluorochem, LOT number: FCB030006) in 0.1 M NaP<sub>i</sub> buffer (pH 8).

### **50 mM ADMA stock**

The 95% *N*<sup>G</sup>,*N*<sup>G</sup>-dimethylarginine stock with a final concentration of 50 mM was prepared by dissolving *N*<sup>G</sup>,*N*<sup>G</sup>-dimethylarginine (abcr, LOT number: 1408632) in ddH<sub>2</sub>O. Afterwards, the stock solution was filter sterilized and stored at 4 °C or at -20 °C.

### **50 mM Arg stock**

The L-arginine stock with a final concentration of 50 mM was prepared by dissolving L-arginine (Carl Roth, LOT number: 301174565) in ddH<sub>2</sub>O. Afterwards the stock solution was filter sterilized and stored at 4 °C.

## **3.3 Liquid media**

### **LB-medium**

LB-medium was prepared by dissolving LB-medium (Lennox) to a final concentration of 20 g/L in dH<sub>2</sub>O. Afterwards the medium was autoclaved and stored at room temperature.

### **LB-kanamycin (LB-Kan) medium**

For the preparation of LB-Kan medium, kanamycin was added to LB-medium (Lennox) to a final concentration of 50 mg/L.

### **SOB- and SOC-media**

SOB medium was prepared by dissolving SOB-medium in dH<sub>2</sub>O to a final concentration of 26.6 g/L and by autoclaving it afterwards. SOC medium was prepared by adding 1 M glucose to SOB medium to a final concentration of 20 mM glucose.

### **M9-kanamycin (M9-Kan) medium and overnight culture (ONC)**

To prepare M9-medium, the components listed in Table 3 were combined and filled up with autoclaved dH<sub>2</sub>O until the calculated final volume was reached.

**Table 3 Components added to prepare M9-medium and M9-plates.** ncAA either indicates the addition of BocK, ADMA or Arg. Components highlighted in grey are added depending on the later use of the medium or plate.

<b>Compound/stock concentration</b>	<b>End concentration</b>
Glucose [1 M]	20 mM
MgSO <sub>4</sub> [1 M]	1 mM
CaCl <sub>2</sub> [1 g/L]	1 mg/L
Thiamine [1 g/L]	1 mg/L
Biotine [0.5 g/L]	1 mg/L
Yeast extract [30 g/L]	0 – 1.5 g/L
Kanamycin (Kan) [50 g/L]	50 mg/mL
Chloramphenicol (Cm) [25 g/L]	25-70 mg/L
Trace elements	-
IPTG [1 M]	0.5 mM
Arabinose [20%]	0.2% (w/v)
ncAA [50 mM]	5 mM

For cultivation in 96-deep well plates we added 30 g/L yeast extract to a final concentration of 0-1.5 g/L. The ONCs were prepared in either LB-Kan or M9-Kan medium and had a total volume of either 5 mL or 10 mL. They were incubated overnight at 130 rpm and 37 °C.

### **3.4 Solid media**

#### **LB-Kan plates**

LB-Kan plates were prepared by dissolving LB-agar (Lennox) in dH<sub>2</sub>O to a final concentration of 30 g/L. Then the medium was autoclaved and cooled to approx. 60 °C. Kanamycin was added to a final concentration of 50 mg/L and the plates were poured.

#### **Growth assay plates: M9-Kan, M9-Kan-ADMA, M9-Kan-BocK, M9-Kan-Arg**

The plates were prepared by dissolving agar-agar in dH<sub>2</sub>O to a final concentration of 15 g/L. Afterwards the dissolved agar-agar as well as 5x M9-salt stock (calculated to a 1x concentration) were autoclaved separately to prevent the salts from precipitating. After cooling to approx. 60 °C, the agar-agar and the M9-salts were combined and the components listed in Table 3 were added. Then the plates were poured. The negative plate either contained dH<sub>2</sub>O or 0.1 M NaP<sub>i</sub> buffer (pH 8) instead of the AA.

### **3.5 Marker**

For sizing and approximate quantification of dsDNA on agarose gels, Thermo Scientific™ GeneRuler 1 kb DNA Ladder was used. For sizing of proteins in protein electrophoresis (SDS-PAGE), Thermo Scientific™ PageRuler™ Prestained Protein Ladder was used.

### **3.6 Kits**

For plasmid DNA isolation, Thermo Scientific™ GeneJET Plasmid Miniprep Kit or Wizard® Plus SV Minipreps DNA Purification System A1460 (Promega) were used. For plasmid DNA purification, Wizard® SV Gel and PCR Clean-Up System (Promega) was used.

### **3.7 Bacterial strains**

For screening, *E. coli* Top10F<sup>'</sup>, *E. coli* BL21(DE3) and *E. coli*® EXPRESS BL21(DE3) (Lucigen) were employed. Table 4 lists the bacterial strains and corresponding genotypes.

**Table 4 Bacterial strains and the corresponding genotypes.** *E. coli* genotypes have been taken from OpenWetWare<sup>27</sup>. *E. cloni* genotype has been taken from Lucigen<sup>28</sup>.

Bacterial strain	Genotype
<i>E. coli</i> Top10 F <sup>'</sup>	F <sup>'</sup> [lacI <sup>q</sup> Tn10(tet <sup>R</sup> )] mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 deoR nupG recA1 araD139 Δ(ara-leu)7697 galU galK rpsL(Str <sup>R</sup> ) endA1 λ <sup>-</sup>
<i>E. coli</i> BL21(DE3)	<i>E. coli</i> str. B F <sup>-</sup> ompT gal dcm lon hsdSB(rB <sup>-</sup> mB <sup>-</sup> ) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB <sup>+</sup> ] <sub>K-12</sub> (λ <sup>S</sup> )
<i>E. cloni</i> <sup>®</sup> EXPRESS BL21(DE3)	F <sup>-</sup> ompT hsdSB (rB- mB-) gal dcm lon λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])

### Synthetic cloning fragments

The synthetic DNA cloning fragments harboring highly mutated pyrrolysyl-tRNA synthetases from *Methanosarcina mazei* were ordered at Twist Bioscience and used as inserts in cloning. Table 5 summarizes the number, abbreviations and properties of the cloning fragments.

**Table 5 ID, abbreviations and properties of synthetic DNA cloning fragments ordered at Twist Bioscience.** dist\_2\_4 and dist\_4\_0, distance between ligand and AMP; hydrophobic only, hydrophobic amino acid residues in the binding pocket were used in the Rosetta script, which designed the proteins.

ID	Abbreviation	Properties
1	ADMARS1cf	dist_2_4 and hydrophobic_only
2	ADMARS2cf	dist_2_4 and hydrophobic_only
3	ADMARS3cf	dist_2_4 and hydrophobic_only
4	ADMARS4cf	dist_2_4 and hydrophobic_only
6	ADMARS6cf	dist_2_4
8	ADMARS8cf	dist_2_4
9	ADMARS9cf	dist_2_4
10	ADMARS10cf	dist_2_4
11	ADMARS11cf	dist_4_0 and hydrophobic_only
12	ADMARS12cf	dist_4_0 and hydrophobic_only
13	ADMARS13cf	dist_4_0



15	ADMARS15cf	dist_4_0
16	ADMARS16cf	dist_4_0
17	ADMARS17cf	AMP
19	ADMARS19cf	AMP
20	ADMARS20cf	AMP
21	ADMARS21cf	AMP
22	ADMARS22cf	dist_4_0
23	ADMARS23cf	dist_4_0 and hydrophobic_only
25	ADMARS25cf	dist_4_0 and hydrophobic_only

### 3.8 Enzymes

The restriction enzymes FastDigest *EcoRI* and FastDigest *BspTI* (= *AflIII*) from Thermo Fisher Scientific were used for DNA digestion. FastDigest *HindIII* from Thermo Fisher Scientific was used for restriction analysis. The 10x FastDigest Green Buffer from Thermo Fisher Scientific was used as a universal buffer for DNA digestion.

T4 DNA ligase (1 U/ $\mu$ L) from Thermo Fisher Scientific was used for ligation. T4 DNA ligase buffer (10x) from Thermo Fisher Scientific was used with T4 DNA ligase.

## 4. Methods

### 4.1 Plasmids

#### Screening plasmids

The first version of the plasmid used for screening was termed pT7x3-MmOP eGFP Y40am and had the abbreviation plasmid GAM67. The second version was termed pT7x3-ADMARS##OP eGFP Y40am. ## indicated the number of the inserted mutated cloning fragment within the *MmPylRS* gene. Plasmid pT7x3-ADMARS##OP eGFP Y40am contained a mutated *MmPylRS* gene instead of the wild type *MmPylRS* but was otherwise like pT7x3-MmOP eGFP Y40am. Both contained an eGFP gene amber mutated at position Y40. Plasmid GAM67 was used as a backbone vector for cloning and as positive and negative control in the fluorescence assay while the ligation product pT7x3-ADMARS##OP eGFP Y40am was used as a sample in the fluorescence assay. pT7x3-MmOP eGFP WT contained the wild type eGFP gene and was used as a wild type control in the fluorescence assay.

#### Growth assay plasmids

pCAT1K4am-MaOP contained the wild type *MaPylRS* and was used as a negative control in the chloramphenicol growth assay. pCAT1WT-MaOP contained the wild type *CAT* gene and was used as a positive control in the chloramphenicol growth assay.

### 4.2 Preparation of competent cells

#### Preparation of chemocompetent cells

100 mL LB-medium were inoculated using an overnight culture of *E. coli* Top10 F' or *E. coli* BL21(DE3). The culture was incubated at 37 °C and 200 rpm to a cell density of 0.4-0.5. The cell suspension was centrifuged at 4 °C and 4000 x g for 15 min. Tfb1 buffer, MgCl<sub>2</sub> solution and Tfb2 buffer were precooled: The pellet was resuspended in 30 mL Tfb1 buffer and 3.2 mL 1 M MgCl<sub>2</sub> and incubated on ice for 15 min. The cells were centrifuged at 4 °C and 4000 x g for 10 min. The supernatant was discarded and the pellet was resuspended in 4 mL Tfb2 buffer. The solution was incubated on ice for 15 min, pipetted in 50 µL aliquots and stored at -80 °C.

## Preparation of electrocompetent cells

500 mL LB-medium were inoculated with the respective overnight culture of *E. coli* Top10 F' or *E. coli* BL21(DE3). The cells were incubated at 37 °C and 180 rpm to a cell density of 0.5-0.7. The cells were cooled on ice for 20 min and centrifuged at 4 °C and 4500 x g for 30 min. The supernatant was discarded, and the pellet was resuspended in 500 mL pre-cooled 10% glycerol solution. The solution was centrifuged at 4 °C and 4500 x g for 30 min. This step was repeated twice with 250 mL and 20 mL pre-cooled 10% glycerol solution, respectively. The supernatant was discarded, and the pellets were resuspended in 1.5 mL pre-cooled 10% glycerol solution. Cells were pipetted in 50 µL aliquots and stored at -80 °C.

## 4.3 Agarose gel electrophoresis

For analysis and purifications as well as for concentration estimations, 0.5-1% (w/v) agarose gels were used. The agarose was dissolved in 1x TAE Tris buffer (1 mM EDTA, 40 mM, 20 mM acetic acid in dH<sub>2</sub>O, pH 8) and heated in the microwave under maximum power until the agarose was completely dissolved. After cooling, ~0.1 mM ethidium bromide was added for band visualization and the gel was poured. Each gel was run at 110-150 V for 30-90 min. As a molecular size marker GeneRuler 1 kb plus DNA ladder was used.

## 4.4 Cloning

*In silico* cloning was done using the computer software SnapGene. For cloning, *E. coli* Top10 F' were transformed with the isolated plasmid GAM67 and the mutated cloning fragments ordered at Twist Bioscience. Overnight cultures were grown, and the plasmids were isolated from the cultures the next day. Backbone and inserts were obtained through an enzymatic digest with FastDigest *EcoRI* and FastDigest *BspTI*. For the backbone, 40 µL of plasmid DNA were mixed with 2.5 µL of *BspTI* and 5 µL 10x FD green buffer. The digest was incubated at 37 °C for 1 h before 2.5 µL of *EcoRI* were added to a total volume of 50 µL. The digest was further incubated at 37 °C for 30 min. For the inserts ID-1, -2, -11, -12, -23 and -25, 40 µL of plasmid DNA were mixed with 2.5 µL of *BspTI*, 2.5 µL of *EcoRI* and 5 µL 10x FD green buffer to a total volume of 50 µL. The digest was incubated at 37 °C for 25 min. For the inserts ID-3 and -4, only 30 µL of plasmid DNA were used in the digest approach due to less DNA available. Restriction enzymes and buffer were added accordingly. The residual 10 µL to the total 50 µL were filled up with ddH<sub>2</sub>O. The digest was incubated at 37 °C for 25 min. The digested backbone was purified twice by agarose gel electrophoresis, the digested inserts once. The DNA

concentration was estimated on the gel and measured by UV/VIS spectrophotometry. Backbone and inserts were stored at -20 °C.

#### **4.5 Transformation of *E. coli* Top10 F' and *E. coli* BL21(DE3) and transformation efficiency**

Electrocompetent cells were transformed as follows: 40 µL of electrocompetent cells were mixed with 1 µL of DNA (10-100 ng) in a pre-cooled cuvette. The electroporator was set to 1.5-2.0 kV, 200 Ω and 25 µF. Immediately after electroporation, 1 mL of SOC medium was added and the cell mixture was transferred into a fresh 1.5 mL tube. The cells were incubated at 37 °C and 350-400 rpm for 1 h for cell regeneration. After cell regeneration, 100-1000 µL of the cell mixture were streaked out on LB-Kan plates and the plates were incubated over night at 37 °C.

Chemocompetent cells were transformed as following: 50 µL of chemocompetent cells were mixed with 5 µL of DNA in a 1.5 mL tube and incubated on ice for 30 min. The temperature was increased to 42 °C for 1 min. Immediately after the heat shock, 1 mL of SOC medium was added and the cell mixture was incubated at 37 °C and 350-400 rpm for 1 h for cell regeneration.

After electrocompetent and chemocompetent cell regeneration, 500 µL of the cell mixture were streaked out on LB-Kan plates and the plates were incubated at 37 °C over night. The rest of the transformation mixes was cultivated in 5-10 mL LB-Kan medium and cultured at 37 °C and 130 rpm over night. The ONC was used for the preparation of glycerol stocks (1:1 60% glycerol and culture). The glycerol stocks were stored at -80 °C and could be used for further applications.

The colony forming units (cfu) on the LB-Kan plates were counted to calculate the transformation efficiency as shown in Equation 1.

$$\text{Transformation efficiency} = \frac{\text{cfu}}{\mu\text{g DNA}} \cdot \frac{V_{\text{tot}}}{V_{\text{plated}}}$$

**Equation 1 Calculation of transformation efficiency.** cfu, colony forming units;  $\mu\text{g DNA}$ ,  $\mu\text{g DNA}$  transformed;  $V_{\text{tot}}$ , total regeneration volume (1 mL);  $V_{\text{plated}}$ , volume plated (dilution factor included).

## 4.6 Ligation

100 ng backbone were mixed with 40-80 ng of insert DNA, 1.5  $\mu\text{L}$  T4 DNA ligase and 1.5  $\mu\text{L}$  T4 DNA ligase buffer to a total volume of 15  $\mu\text{L}$ . If necessary, ddH<sub>2</sub>O was used to fill up to the total volume. The ligation mix was incubated at 4 °C over night. The religation control was performed accordingly omitting the inserts. The next day, the ligation mixes were examined on agarose gels. 2  $\mu\text{L}$  of the ligation mix with 10x loading dye were applied onto an agarose gel. When the successful ligation was confirmed, the ligation mixes were used to transform *E. coli* Top10 F'. The transformed cells were streaked out on LB-Kan plates and cultivated at 37 °C over night. The next day, the plasmids were isolated from the culture. DNA concentrations were estimated by agarose gel and measured by UV/VIS spectrophotometry. The plasmids were stored at -20 °C.

## 4.7 Restriction analysis with *EcoRI* and *HindIII*

Plasmids obtained after ligation were enzymatically digested with FastDigest *EcoRI* and FastDigest *HindIII* to examine if the ligation was successful. 1  $\mu\text{L}$  of plasmid DNA was mixed with 1  $\mu\text{L}$  *EcoRI*, 1  $\mu\text{L}$  *HindIII* and 2  $\mu\text{L}$  FD Green buffer. ddH<sub>2</sub>O was used to fill up to a total volume of 20  $\mu\text{L}$ . The digests were incubated at 37 °C for 30 min and put on ice afterwards. After mixing with 10x loading dye, the digests were applied onto an agarose gel.

## 4.8 Sequencing

DNA was sequenced at Microsynth AG. 3  $\mu\text{L}$  of 20  $\mu\text{M}$  primer and 40-100 ng/ $\mu\text{L}$  DNA (optimal concentration was 80 ng/ $\mu\text{L}$ ) were mixed in a 1.5 mL sterile tube and send to Microsynth AG. A total volume of 15  $\mu\text{L}$  was not exceeded. After ligation, the resulting plasmids were sequence verified using the primers listed in Table 6.

**Table 6 Primers used for plasmid verification after ligation of backbone and inserts.** Genes to be analyzed and primer sequences from 5' to 3' terminus are given.

Primer	Gene to be analyzed	Sequence 5'-3' terminus
pBP562	ADMARS	GGCGCGAGATTTAATCG
pBP2217	eGFP	AAACCACCCTGGCGCCCAATAC
pBP2518	eGFP	CAGCAGATTACGCGCAG
pBP2332	ADMARS	CGAATATCATGGTGGAAAATGGCCGCTTTTC
pBP2357(2)	ADMARS	CAAGCATTCTACAGGAGCAACTGCATC

#### 4.9 Expression of *E. coli* BL21(DE3) and fluorescence assay

*E. coli* BL21(DE3) were transformed with the isolated plasmids obtained after ligation. 500  $\mu$ L of the transformation mix were streaked out on LB-Kan plates and incubated at 37 °C over night. The remaining 500  $\mu$ L were cultivated in 5 mL LB-medium at 37 °C and 350 rpm over night. The single colonies grown on the LB-Kan plates were used to inoculate an ONC of 5 mL LB-medium. The culture was incubated at 37 °C over night. The next day, the ONCs of the transformation mixes as well as of the single colonies were pelleted. In the wells of a 96-well plate, 50  $\mu$ L pelleted cells were mixed with 50  $\mu$ L 60% glycerol. Two 96-well master plates were prepared equally, one working plate (stored at -20 °C) and one backup plate (stored at -80 °C). *E. coli* BL21(DE3) were also transformed with the control plasmids. The transformation mixes were used to inoculate ONCs. The next day, cells were pelleted and stored as glycerol stocks. A 96-well plate with 900  $\mu$ L LB-medium per well was prepared as a pre-culture. A 96-pin stamp was used to inoculate the pre-culture with the working master plate. The pre-culture was then cultivated at 37 °C and 350 rpm over night. The next day, a 96-well plate with ~900  $\mu$ L M9-Kan medium per well was prepared as a main culture. The cell density of the pre-culture was measured with a plate reader. Based on the measured cell density, the inoculation volume for a starting cell density of 0.1 for the main culture was calculated. Immediately after inoculation of the main culture, cell density and eGFP fluorescence were measured. The culture was then incubated at 37 °C for 3-4 h. Depending on the starting cell density of the main culture, the second measurement of cell density and eGFP fluorescence was performed 3-4 h after inoculation. The cells were incubated until they reached a cell density of 0.6-0.8. When the cells reached the intended cell density, they were induced with 5 mM IPTG

to a final concentration of 0.5 mM IPTG per well. Cell density and eGFP fluorescence were measured again 2, 4 and 16 h after induction.

#### **4.10 Protein expression analysis**

To analyze the expression of the ADMARS variants, they were cultivated in 24-well plate cultures. The cell cultivation for the protein expression analysis was done accordingly to the cell cultivation of *E. coli* BL21(DE3) for the fluorescence assay. Supplementation with amino acids was omitted. The cell densities were measured 2, 4 and 6.5 h after inoculation as well as immediately, 2 and 4 h after induction. 0.25 cell density units were collected immediately before and 4 h after induction. The cells were centrifuged at 14000 x g for 10 min and stored at -20 °C.

#### **4.11 SDS-polyacrylamide gel electrophoresis**

SDS-polyacrylamide gel electrophoresis was used for protein analysis after expression of ADMARSs, *MmPylRS* and *MaPylRS* SDLSL in *E. coli* BL21(DE3). The cultures to be analyzed were grown in a 24-well plate according to the cultivation described in section 4.10. 0.25 cell density units of the cell cultures were pelleted and frozen at -20 °C.

For preparation of protein samples, SDS sample buffer was used. The next day, the pellets were thawed, mixed with 75 µL of 1x SDS sample buffer and boiled at 95 °C for 10 min. Afterwards, they were treated with ultrasound for 30 min. The samples were centrifuged at 14000 x g for 1 min to collect the whole sample on the bottom of the tube. A Pre-cast gel NuPAGE™ 4 to 12% Bis-Tris, 1.0 mm, Mini Protein Gel, 2D-well purchased from Thermo Fisher Scientific was used. The gel chamber was assembled as described in the manual. 10 µL of the prepared samples and 3 µL of the protein marker were pipetted into the slots of the gel. The gel was run at 200 V, 120 mA for ~50 min. As a staining solution, Expedeon InstantBlue™ Stain Protein Stain from Thermo Fisher Scientific was used. The gel was covered completely with the staining solution and kept on a seesaw for about 1.5 h. The staining solution could be reused and was filled back into the bottle. The gel was washed with water and photographed afterwards.

## 4.12 Chloramphenicol growth assay

The M9-Kan plates for the chloramphenicol growth assay were prepared according to Table 7. Three plates supplemented with ADMA but varying Cm concentrations, one plate supplemented with Arg and one plate without AA addition were prepared.

**Table 7 Ingredients of M9-Kan plates for chloramphenicol growth assay.** Compounds and corresponding stock concentrations are given. Plates ADMA1, 2 and 3 were supplemented with ADMA. Plate Arg was supplemented with arginine. Plate (-) contained no amino acid. Kan, kanamycin; Cm, chloramphenicol; ADMA, *N*<sup>G</sup>,*N*<sup>G</sup>-dimethylarginine; Arg, arginine.

Compound/stock conc.	Plate ADMA1	Plate ADMA2	Plate ADMA3	Plate Arg	Plate (-)
Agar-Agar	15 g/L	15 g/L	15 g/L	15 g/L	15 g/L
M9-salts 5X	1x	1x	1x	1x	1x
Glucose [1 M]	20 mM	20 mM	20 mM	20 mM	20 mM
MgSO <sub>4</sub> [1 M]	1 mM	1 mM	1 mM	1 mM	1 mM
CaCl [1 mg/mL]	1 µg/mL	1 µg/mL	1 µg/mL	1 µg/mL	1 µg/mL
Thiamine [1 mg/mL]	1 µg/mL	1 µg/mL	1 µg/mL	1 µg/mL	1 µg/mL
Biotin [1 mg/mL]	1 µg/mL	1 µg/mL	1 µg/mL	1 µg/mL	1 µg/mL
Kan [50 mg/mL]	50 µg/mL	50 µg/mL	50 µg/mL	50 µg/mL	50 µg/mL
Cm [25 g/L]	-	25 µg/mL	50 µg/mL	25 µg/mL	25 µg/mL
ADMA [50 mM]	5 mM	5 mM	5 mM	-	-
Arg [50 mM]	-	-	-	5 mM	-
IPTG [1 M]	0.5 mM	0.5 mM	0.5 mM	0.5 mM	0.5 mM
Arabinose [20%]	0.2 %	0.2 %	0.2 %	0.2 %	0.2 %

The *E. coli* as well as *E. cloni* samples were prepared as following: We prepared ONCs from the glycerol stocks *E. cloni* {pCAT1-MaOP SDLSL} clone 2 and *E. cloni* {pCAT1-MaOP SDLSL} clone 3 in LB-Kan medium incubated at 37 °C with shaking. *E. coli* BL21(DE3) and *E. coli* Top10F' were transformed with the isolated plasmids from the ONCs. Transformed *E. coli* Top10F' were not used in the assay but stored as glycerol stocks. The transformed *E. coli* BL21(DE3) cells were used to inoculate ONCs in LB-Kan medium incubated at 37 °C with shaking. The next day, half of each ONC was frozen as a glycerol stock at -80 °C. 100 µL of cell suspension from the ONC were mixed with 900 µL LB medium and the cell densities were



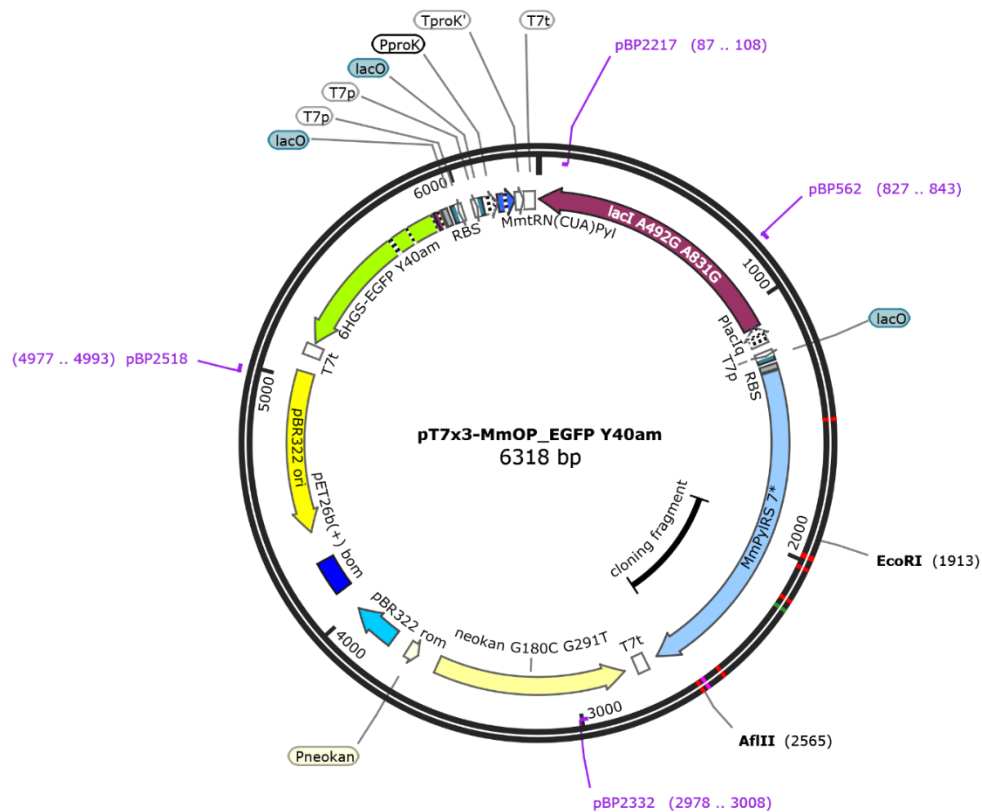
measured. A volume equal to a  $D_{600}$  of 1.0 in 1 mL cell suspension was harvested. The cell densities were measured again to confirm a  $D_{600}$  of 1.0 in 1 mL cell suspension. The cell suspensions were centrifuged at 4000 x g for 10 min and the pellets were collected. The pellets were washed twice with and resuspended in 1 mL of 0.9% saline solution. The resuspended pellets were pipetted into the wells of a DWP and diluted by the factors  $10^2$ ,  $10^4$  and  $10^6$ . 3  $\mu$ L of each dilution were dropped onto the corresponding section on the M9-Kan plates as depicted in section 5.5. The plates were incubated at 37 °C over night and evaluated 24 and 45 h afterwards.

## 5. Results and Discussion

### 5.1 Construction of the plasmid library

#### 5.1.1 Preparation of plasmid backbone pT7x3-MmOP eGFP Y40am

The isolated plasmid pT7x3-MmOP-eGFP Y40am<sup>37</sup> (plasmid GAM67) was used as a backbone vector since it comprised all necessary parts for the construction of the plasmid library. As can be seen in the plasmid map (Figure 2), it carried a high-copy origin of replication, a kanamycin resistance gene and eGFP with an amber mutation Y40am under the control of the T7<sub>lacO</sub> promoter. The expression of *MmPylRS* and tRNA<sub>CUA</sub><sup>Pyl</sup> (*MmOP*) were also under the control of a T7 promoter. In the *MmPylRS* gene, mutations were inserted encompassing nucleotide 1913-2570 (cloning fragment) to greatly expand the substrate binding scope. The region spanned by these nucleotides encoded the amino acid binding pocket in the *MmPylRS* gene. The plasmid backbone as well as the cloning fragments had to be enzymatically digested. *EcoRI* and *BspTI* (*AflIII*) were selected as suitable restriction sites.



**Figure 2 Plasmid map pT7x3-MmOP eGFP Y40am (plasmid GAM67).** The plasmid was used as a backbone vector in the cloning experiment. *MmPylRS* 7\*, pyrrolysyl-tRNA synthetase from *Methanosarcina mazei* Go1 mutated by insertion of synthetic DNA fragments (cloning fragment); *lacI* A492G A831G, lac repressor;

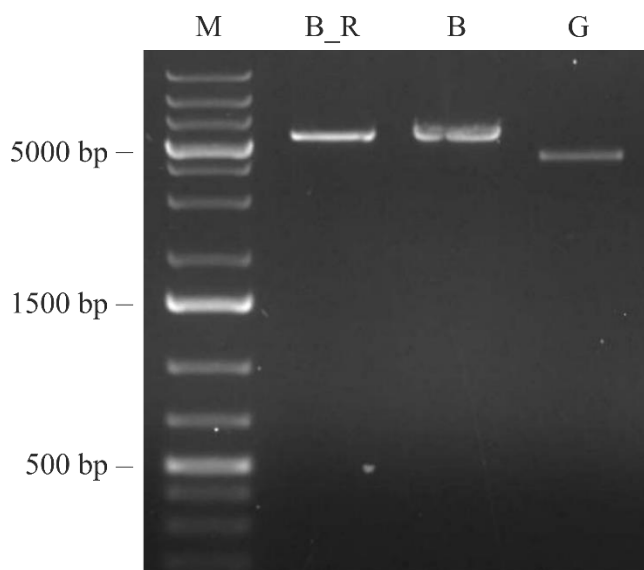
MmtRN<sub>CUA</sub><sup>Py1</sup>, amber suppressor tRNA<sub>CUA</sub><sup>Py1</sup> from *Methanosarcina mazei* Go1; 6HGS-EGFP Y40am, hexahistidine-tagged enhanced green fluorescent protein EGFP F64L S65T with an amber mutation at position Y40; pBR322 ori, origin of replication from *E. coli* plasmid pBR322; *neokan* G180C G291T, aminoglycoside-3'-O-phosphotransferase from transposon Tn5; *EcoRI* and *AflIII* (*BspTI*), restriction sites.

We aimed to excise a 652 bp fragment of the plasmid GAM67 by enzymatic digest with *EcoRI* and *BspTI*. The resulting double-cut backbone of 5670 bp was used for the construction of the plasmid library. According to the manuals of *EcoRI* and *BspTI*, they should have been compatible. Still, the combination of both was a challenge. We tried different protocols for the enzymatic digest. We varied the incubation time and examined *EcoRI*'s star activity. According to the manual, *BspTI* did not show any star activity. Therefore, we started the digest with *BspTI* already before adding *EcoRI*. We used different amounts of plasmid DNA and tried with and without heat inactivation at 80 °C after the digest. Agarose gel electrophoresis was used to monitor the digests as well as to estimate DNA concentrations. Alternatively, UV/VIS spectrophotometry was used to measure DNA concentrations. Concentration measurements by UV/VIS spectrophotometry exceeded the gel estimation by 2- to 3-fold. We observed this difference repeatedly for DNA after plasmid isolation with DNA purification kits.

Regardless of the applied digestion protocol, a portion of the plasmid always remained uncut or single-cut. Since the double-cut and single-cut backbones did not differ by many base pairs, they migrated very closely to each other on the agarose gel. Therefore, the separation of incompletely from completely digested DNA fragments and their excision from the gel was very challenging. We suggested that the restriction digest with *EcoRI* and *BspTI* was incomplete due to decreased restriction enzyme activity. The decreased activity could have been an effect of the chaotropic and strongly denaturing properties of guanidine hydrochloride<sup>29</sup> which is a component of the neutralization solution of the Wizard® Plus SV Miniprep DNA Purification System<sup>30</sup>. It exhibits an absorbance at 230 nm<sup>31</sup>. Since we detected a contamination at 230 nm by UV/VIS spectrophotometry measurements, this could have corresponded to an increased guanidine hydrochloride concentration in the DNA samples due to insufficient washing during plasmid isolation.

Since plasmid GAM67 was amplified in *E. coli* Top10F' cells and then isolated, the plasmid DNA could have been present in circular and supercoiled form. On an agarose gel, supercoiled DNA migrates fastest followed by linearized DNA (double- or single-cut) and circular unwound DNA<sup>32</sup>. Since we had troubles obtaining completely digested backbone, we purified the

backbone after enzymatic digest from an agarose gel, applied it onto another gel and extracted it again. Figure 3 shows two backbone samples as well as the uncut plasmid GAM67.



**Figure 3 Enzymatic digest of backbone.** GAM67 cut with restriction enzymes *EcoRI* and *BspTI* migrated at the expected band size of 5660 bp. B\_R, backbone pT7x3-MmOP eGFP Y40am was repurified after previous gel purification; B, backbone pT7x3-MmOP eGFP Y40am was applied after digest; G, uncut plasmid GAM67 pT7x3-MmOP eGFP Y40am was applied for comparison (6318 bp). 1% agarose gel. M, GeneRuler 1 kb Plus DNA Ladder.

Backbone sample B (Figure 3, lane 3) contained double-cut plasmid GAM67 and was applied onto the gel after the enzymatic digest. Backbone sample B\_R (lane 2) contained double-cut plasmid GAM67 that had been purified from the gel before and was repurified to improve its purity for downstream ligation. We did not use digested backbone that was gel-purified only once to avoid the risk of religation of incompletely digested backbone.

### 5.1.2 Preparation of inserts

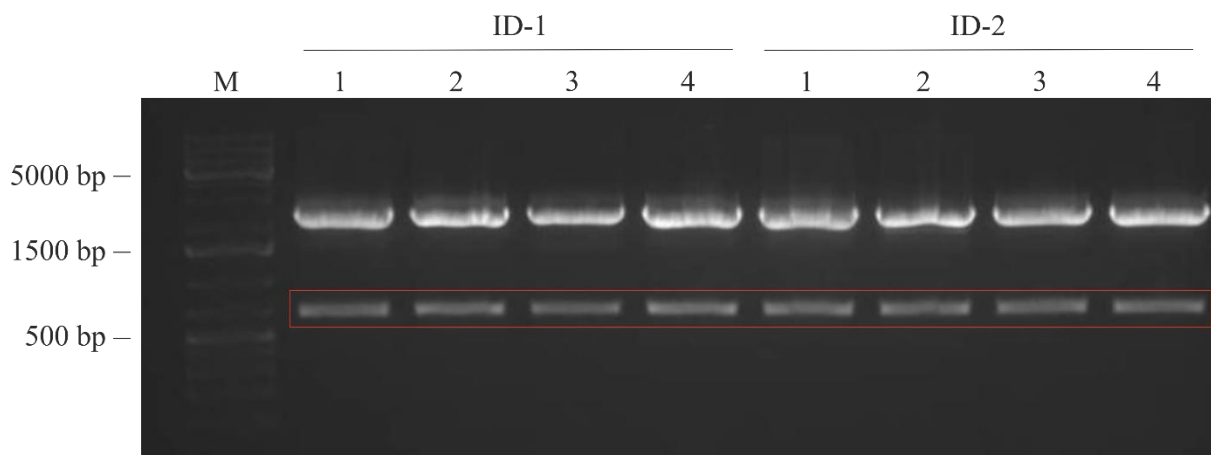
In the protein design of the master thesis of Markus Braun<sup>33</sup>, the pyrrolysyl-tRNA synthetase from *Methanosarcina mazei* was hypermutated to greatly expand its substrate binding scope. 20 synthetic cloning fragments harboring different mutations were inserted in the plasmid backbone. Important parameters for the designs were the respective ncAA (ligand) to be recognized by the synthetase, the design of the binding pocket and the distance between the amino acid and the adenosine monophosphate (AMP) after AA activation by the synthetase. According to the designs, I focused on eight DNA sequences labeled with the numbers ID-1, -2, -3, -4, -11, -12, -23 and -25. We categorized those sequences into the two design groups A and B (Table 8).

**Table 8 Properties of the design groups A and B.** DNA sequences used as inserts covered in this thesis are listed according to the design groups A and B. Ligand, respective amino acid; dist\_2\_4 and dist\_4\_0, distance between ligand and AMP; ADMA, *N<sup>G</sup>,N<sup>G</sup>*-dimethylarginine.

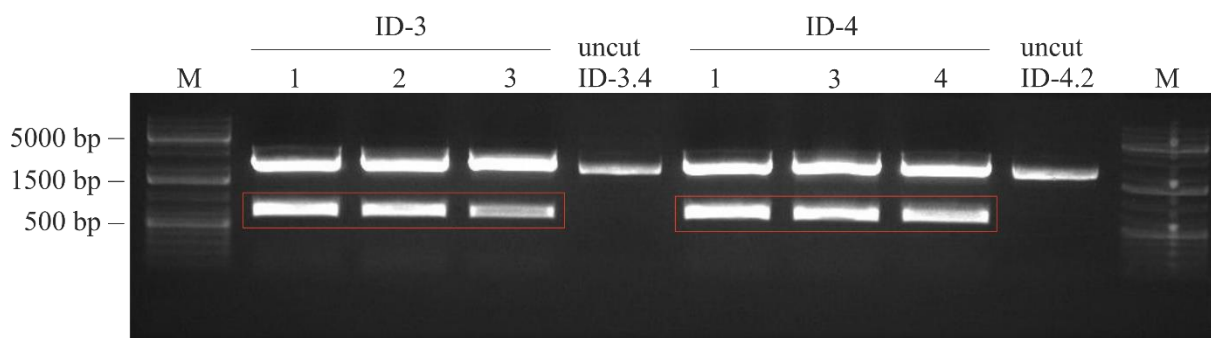
<b>Design group</b>	<b>A</b>	<b>B</b>
Sequences assigned to design groups	ID-1, -2, -3, -4	ID-11, -12, -23, -25
Abbreviation	dist_2_4 and hydrophobic only	dist_4_0 and hydrophobic only
Ligand	ADMA	ADMA
Distance of ligand to AMP	2.4 Å	4.0 Å
Amino acids in the binding pocket	only hydrophobic amino acid residues, amino acids in the beta hairpin around the residue 384 were excluded	only hydrophobic amino acid residues, amino acids in the beta hairpin around the residue 384 were excluded

In both design groups A and B, ADMA was used as a ligand. In group A, the distance between the ligand and AMP was 2.4 Å (Table 8, “dist\_2\_4”), in group B, the distance was 4.0 Å (“dist\_4\_0”). Only hydrophobic AA residues in the binding pocket were used in the Rosetta script, with which the proteins were designed (“hydrophobic only”). AAs in the beta hairpin around the residue 384 were excluded.

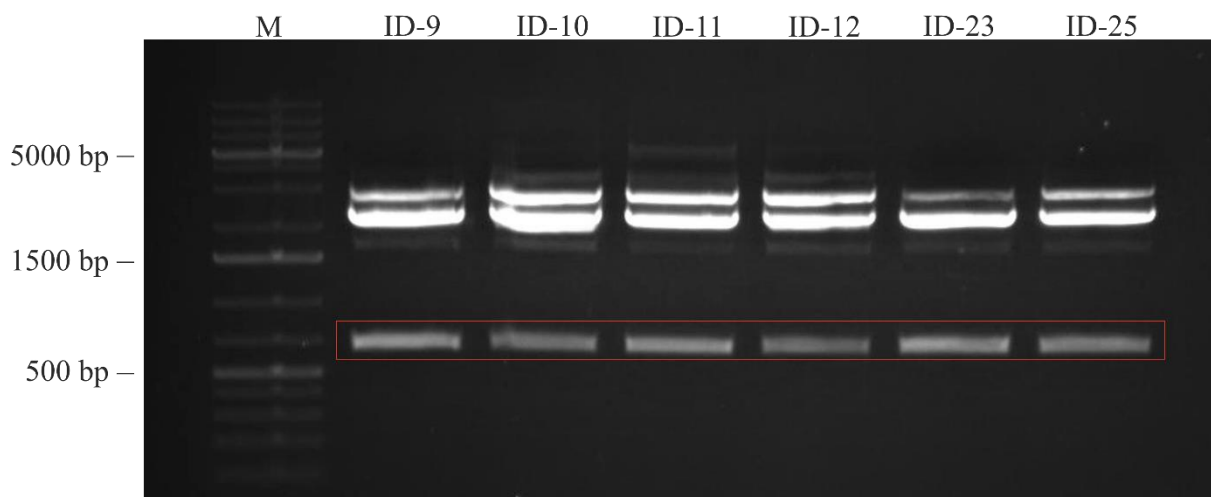
The aim was the isolation of the cloning fragments from the synthetic DNA fragments for further ligation. Through the enzymatic digest of the inserts, we aimed for a 654 bp insert which could be distinguished well from the rest of the DNA fragment after digest. As mentioned above, *EcoRI* and *BspTI* were used as restriction enzymes. Agarose gel electrophoresis was used to monitor the digests as can be seen in Figure 4, Figure 5 and Figure 6 for all inserts. In lane 5 and lane 9 of Figure 5, uncut DNA sequences were applied for comparison. The expected band sizes were 654 bp for the excised insert and 2221 bp for the remaining DNA fragment.



**Figure 4 Enzymatic digest of inserts ID-1 and ID-2.** Inserts ID-1 and ID-2 were cut with restriction enzymes *EcoRI* and *BspTI* and migrated at the expected band size of 654 bp (red box). 1% agarose gel. M, GeneRuler 1 kb Plus DNA Ladder. Numbers (1-4) indicate biological replicates.



**Figure 5 Enzymatic digest of inserts ID-3 and ID-4.** Inserts ID-3 and ID-4 were cut with restriction enzymes *EcoRI* and *BspTI* and migrated at the expected band size of 654 bp (red box). Uncut DNA sequences ID-3.4 and ID-4.2 were applied for comparison (2879 bp). 1% agarose gel. M, GeneRuler 1 kb Plus DNA Ladder. Numbers (1-4) indicate biological replicates.



**Figure 6 Enzymatic digest of inserts ID-11, ID-12, ID-23 and ID-25.** Inserts were cut with restriction enzymes *EcoRI* and *BspTI* and migrated at the expected band size of 654 bp (red box). Inserts ID-9 and ID-10 were applied for comparison. 1% agarose gel. M, GeneRuler 1 kb Plus DNA Ladder.

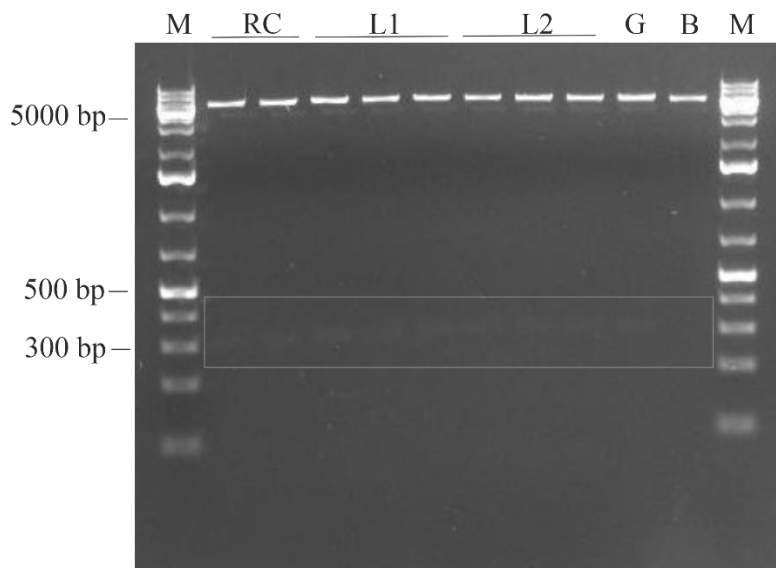
DNA concentrations of all eight inserts were estimated on the gels and the DNA was purified from the gels. DNA concentrations were also measured by UV/VIS spectrophotometry.

### 5.1.3 Ligation of backbone vectors and inserts

We ligated the completely digested plasmid backbone and the inserts to obtain closed and fully functional plasmids. For each ligation, positive, negative and religation controls were included. Uncut plasmid GAM67 was the positive control and *E. coli* Top10 F' cells transformed with ddH<sub>2</sub>O instead of any DNA was the negative control. The religation control mix contained all components of the ligation samples, except that the corresponding volume of water instead of insert DNA was added. The expectations for the plated transformants were as following: The positive control should yield a large quantity of colonies. The negative control should not yield any colonies. The religation control should yield no or very few colonies. Only if the ligation constructs yielded substantially more colonies than the religation control, it could be concluded that the ligation was successful.

We started with the ligation of the backbone and the insert ID-4. In the first round of ligation, the negative control showed no colonies but the religation control more than any of the ligation plates. We concluded that presumably the backbone was not completely digested. Single-cut backbone which still carried the resistance gene could close by forming a nicked circular plasmid in the cells after transformation. If the religation control yielded no or very few colonies, it presumably included unligated double-cut backbone. With no insert DNA present, the double-cut backbone could not form a closed plasmid since we used two different restriction

enzymes generating incompatible sticky ends. In this case, the resistance gene could not be expressed and no colonies could appear. We wanted to examine the colonies on both religation and ligation plates. To distinguish between single-cut religated backbone and ligated backbone with insert we performed a restriction analysis with the restriction enzymes *EcoRI* and *HindIII* (Figure 7). *EcoRI* was used for linearization whereas *HindIII* was used for discrimination. The *HindIII* restriction site was only present on the uncut plasmid GAM67, namely in the fragment that was excised by digestion with *EcoRI* and *BspTI*. Therefore, the site was still present on uncut or single-cut backbone but not on double-cut backbone or rather ligated backbone and insert. Uncut or single-cut backbone digested with *HindIII* appeared on the agarose gel as two bands (Figure 7, 301 bp and 6017 bp) whereas ligated backbone and insert appeared as one band (6318 bp).

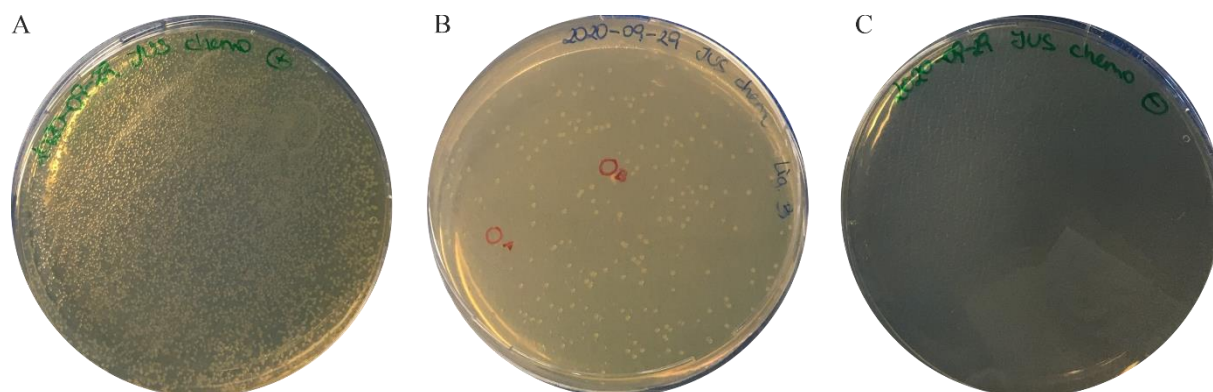


**Figure 7 Restriction analysis of ligation of backbone and insert ID-4 with *EcoRI* and *HindIII*.** Plasmids obtained from religation and ligation plates as well as controls were cut with the restriction enzymes *EcoRI* and *HindIII*. Bands appeared at 301 bp (grey box), 6017 bp and 6318 bp. RC, restriction control; L1 and L2, ligation of backbone and ID-4 from two plates; G, plasmid GAM67; B, backbone previously double-cut with *EcoRI* and *BspTI*. 1% agarose gel. M, GeneRuler 1 kb Plus DNA Ladder.

As can be seen in Figure 7, one band appeared for the backbone (B) (previously double-cut with *EcoRI* and *BspTI*) and two for the plasmid GAM67 (G). Backbone and plasmid GAM67 were applied as controls. As expected, two bands appeared for the religation samples (RC). Two bands also appeared for the ligation samples (L1 and L2). By this, we verified that the plasmids obtained from the ligation plates were religated constructs and not ligated backbone and insert. To avoid religation constructs in the next round of ligation, we repeated the preparation of the backbone to achieve a higher purity of completely digested backbone.



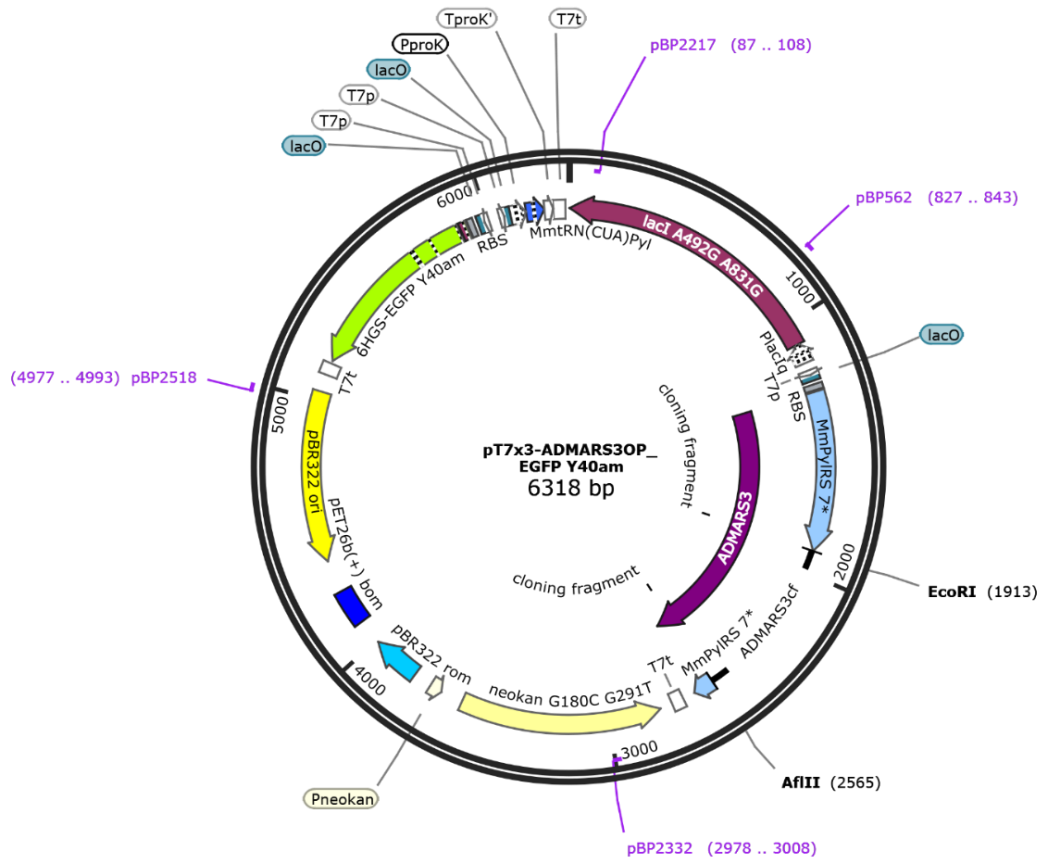
In the second round of ligation, a freshly prepared plasmid backbone was used and the ligation protocol was adapted. The amount of backbone remained the same (100 ng). We increased the amount of inserts from 30 ng to 40-80 ng and the total volume from 10  $\mu$ L to 15  $\mu$ L. Incubation was done at 4  $^{\circ}$ C over night instead of 22  $^{\circ}$ C for 3.5 h. Positive, negative and religation controls were included as before. Representatively for all eight ligation products, transformation plates of positive and negative control as well as the ligation plate are shown for the ligation of backbone and insert ID-3 (Figure 8).



**Figure 8 Colonies of *E. coli* Top10 F' on LB-Kan plates after transformation with ligation product of backbone and insert ID-3.** *E. coli* Top10 F' were transformed with ligation mixes and streaked out on LB-Kan plates. (A), positive control (uncut plasmid GAM67); (B), ligation products of backbone and insert ID-3; (C), negative control (ddH<sub>2</sub>O). All three panels were extracted from the same picture.

As expected, a high number of colonies appeared for the positive control (Figure 8, panel A) and no colonies for the negative control (panel C). The ligation plate showed prominent colonies (panel B) and we concluded that the ligation was successful.

With the revised ligation protocol all eight ligation products could be obtained as intended. Hereinafter, the successfully ligated backbones with corresponding inserts are called ADMA-tRNA synthetase (ADMARS) variants. The plasmid map for the ADMARS 03 variant (pT7x3-ADMARS03OP eGFP Y40am) is shown in Figure 9 as a representation for all ADMARS variants.



**Figure 9 Plasmid map of pT7x3-ADMARS03OP eGFP Y40am (ADMARS 03).** Plasmid map of the final ligation product of plasmid backbone and insert ID-3. ADMARS3, mutated pyrrolysyl-tRNA synthetase from *Methanosarcina mazei* Go1; ADMARS3cf, cloning fragment ID-3; *lacI* A492G A831G, lac repressor; *MmtRN*<sub>CUA</sub><sup>Pyl</sup>, amber suppressor tRNA<sub>CUA</sub><sup>Pyl</sup> from *Methanosarcina mazei* Go1; 6HGS-EGFP Y40am, hexahistidine-tagged enhanced green fluorescent protein EGFP F64L S65T with an amber mutation at position Y40; pBR322 ori, origin of replication from *E. coli* plasmid pBR322; *neokan* G180C G291T, aminoglycoside-3'-O-phosphotransferase from transposon Tn5; *EcoRI* and *AflIII* (*BspTI*), restriction sites.

After ligation, all ADMARS variants were verified by sequencing. The sequencing electropherograms were compared to the annotated plasmid maps and mismatches or gaps could be excluded.

## 5.2 Fluorescence assay with ADMARS variants

Together with other ADMARS variants comprising different mutations, the previously mentioned eight ADMARS variants as well as controls were cultivated as quadruplicates in a 96-deep-well plate. The layout of the samples on the 96-deep-well plate can be seen in Table 9. Replicates 1-3 were individual clones, replicate 4 was the liquid culture of the regenerated transformation mix. Cell density as well as eGFP fluorescence were measured in constant intervals of 3-4 h with a plate reader.

**Table 9 Sample layout for screening of ADMARS variants in 96-well plate cultures.** ADMARS variants of design groups A and B are indicated in blue and bold. WT-1 to WT-4, *E. coli* BL21(DE3) {pT7x3-MmOP eGFP WT} without addition of any amino acid (4 replicates); PC-1 to PC-4, *E. coli* BL21(DE3) {pT7x3-MmOP eGFP Y40am} with addition of Bock (4 replicates); NC-1 to NC-4, *E. coli* BL21(DE3) {pT7x3-MmOP eGFP Y40am} without addition of any amino acid (4 replicates); sterile control, M9-Kan medium not inoculated; ##-1 to ##-4, *E. coli* BL21(DE3) {pT7x3-ADMARS##OP eGFP Y40am}, ADMARS mutants 1-25; replicates 1-3, individual clones; replicate 4, liquid culture of regenerated transformation mix.

	1	2	3	4	5	6	7	8	9	10	11	12
<b>A</b>	WT-1	01-1	02-1	03-1	04-1	06-1	08-1	09-1	10-1	11-1	12-1	PC-1
<b>B</b>	WT-2	01-2	02-2	03-2	04-2	06-2	08-2	09-2	10-2	11-2	12-2	PC-2
<b>C</b>	WT-3	01-3	02-3	03-3	04-3	06-3	08-3	09-3	10-3	11-3	12-3	PC-3
<b>D</b>	WT-4	01-4	02-4	03-4	04-4	06-4	08-4	09-4	10-4	11-4	12-4	PC-4
<b>E</b>	Sterile control	13-4	15-4	16-4	17-4	19-4	20-4	21-4	22-4	23-4	25-4	NC-1
<b>F</b>	Sterile control	13-3	15-3	16-3	17-3	19-3	20-3	21-3	22-3	23-3	25-3	NC-2
<b>G</b>	Sterile control	13-2	15-2	16-2	17-2	19-2	20-2	21-2	22-2	23-2	25-2	NC-3
<b>H</b>	Sterile control	13-1	15-1	16-1	17-1	19-1	20-1	21-1	22-1	23-1	25-1	NC-4

### 5.2.1 Cultivation in 96-deep-well plates

We aimed to incorporate ADMA site-specifically at the amber position in the eGFP gene. For the cultivation, minimal or synthetic media are preferred rather than complex media since oligopeptides and amino acids are overrepresented in these and can compete with ADMA for incorporation<sup>34</sup>. We used minimal M9 medium supplemented with kanamycin (M9-Kan) for cultivation. It only contained salt, glucose and vitamins to allow the microorganism to produce AAs and proteins but was free of any animal-derived components. In contrast, yeast extract (YE) contains AAs, peptides, vitamins and carbohydrates and is often used as a media additive<sup>35</sup>. We tried different YE concentrations in the cultivation medium: 0, 0.5, 1.0, 1.5 and 3.0 g/L YE. To improve cell growth, we decided to add 0.5 g/L YE to the M9-Kan medium. In addition to varying YE concentrations, we tried different cultivation conditions. We added plastic beads with a diameter of approx. 2 mm to each well to improve mixing. We cultivated with different stirring speed as well as at 28 °C and 37 °C. The best cultivation condition was

M9-Kan medium with 0.5 g/L YE, plastic beads in each well, at 37 °C and 320 rpm. After running several rounds of screening, we concluded that 3 h after inoculation was a good point in time for the first measurement of cell density and eGFP fluorescence. We found out that the approximate time until IPTG induction was 3-4 h after inoculation.

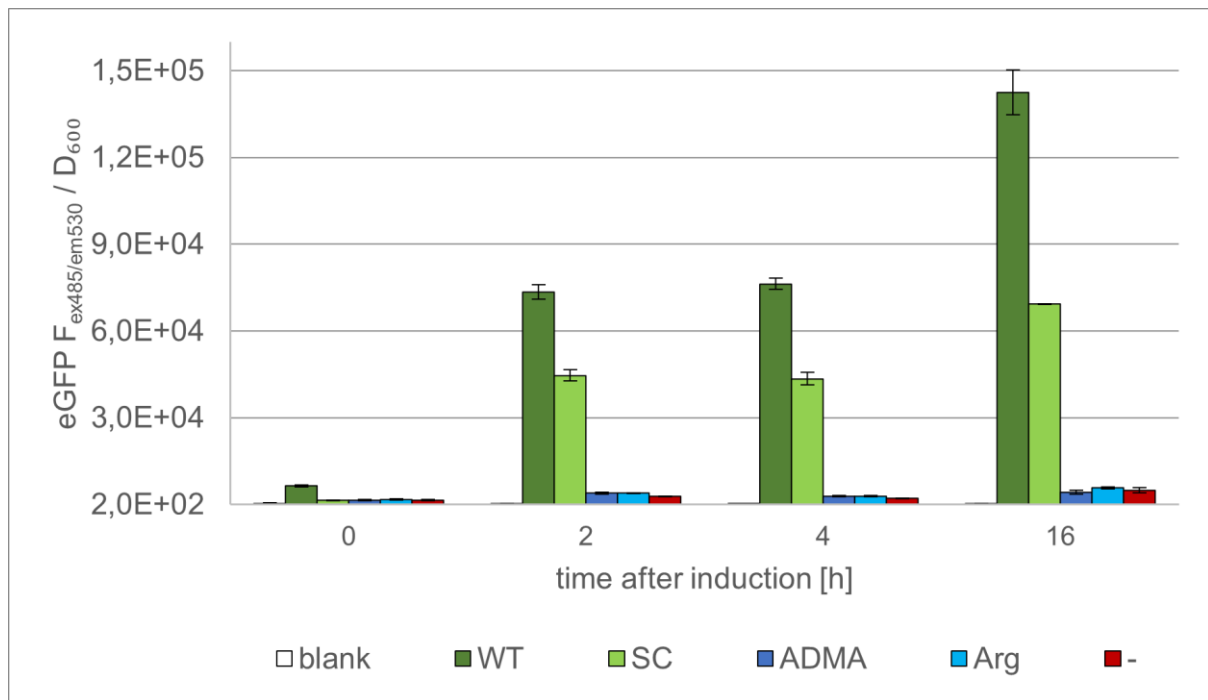
### 5.2.2 Cell density and fluorescence assay

The data of cell density and eGFP fluorescence measurements was collected in constant intervals. With the layout of the samples on the 96-well plate being defined, eGFP fluorescence per cell density values were calculated by the plate reader software. In the first screening experiment induction failed and we had to start a second round of screening.

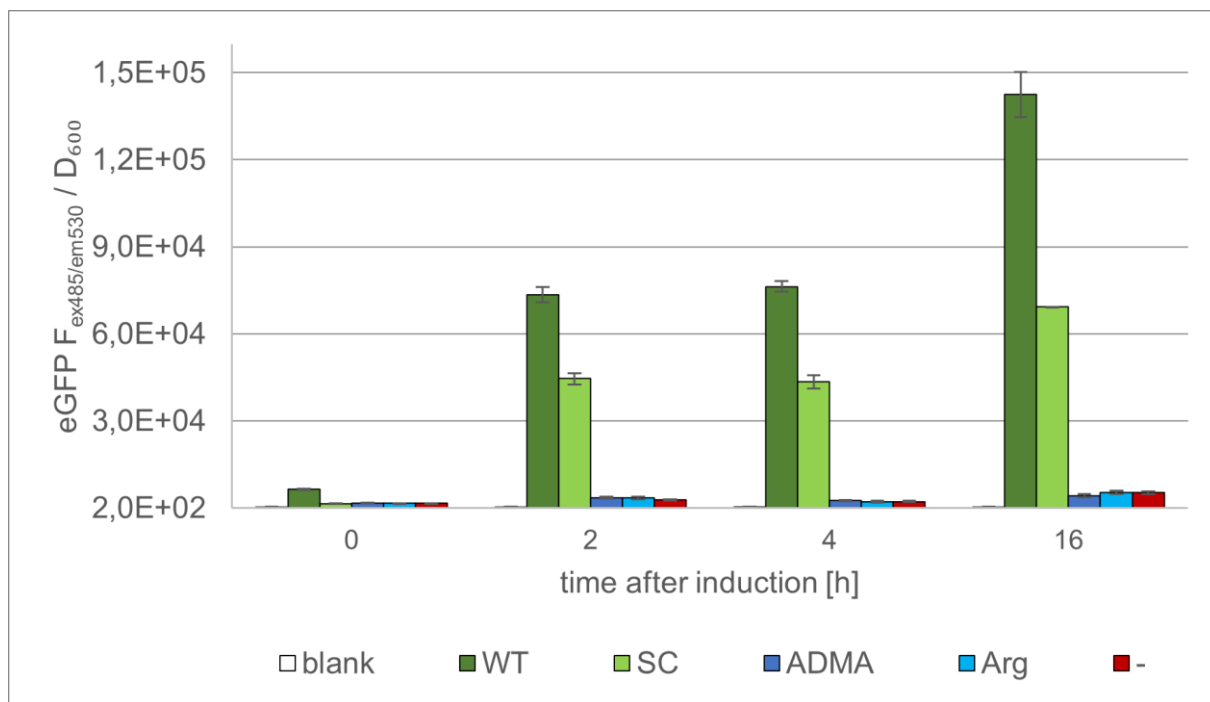
The data of the second screening experiment was validated. Representative for the ADMARS variants of design groups A and B, the diagrams show eGFP fluorescence per cell density of ADMARS 01 (design group A, Figure 10) and ADMARS 11 (design group B, Figure 11) immediately and 2, 4 and 16 h after induction. The average of four replicates is shown each. One outlier was excluded in the calculation of the average of the suppression control (SC). The diagrams of the other ADMARS variants are shown in the appendix (Supplementary figures 17, 18, 19, 20, 21 and 22).

Three screening plates were processed in one experiment, one supplemented with ADMA (positive plate, ADMA), one supplemented with arginine (negative plate, Arg) and one without addition of any AA (background plate, -). 95%  $N^G, N^G$ -dimethylarginine (ADMA) has not been tested before as a substrate for *MmPylRS*.  $N_\alpha$ -(*tert.*-butoxycarbonyl)-L-lysine (BocK) is a commercially available surrogate for pyrrolysine<sup>36</sup> and was used as a model substrate for *MmPylRS*. ADMA, Arg and BocK were provided at the same concentrations.

The diagrams include the results of the ADMARS variants as well as of the controls. The controls were present on all three plates and included sterile control (M9-Kan medium, not inoculated, blank), suppression control (eGFP Y40am mutant with *MmOP*) and wild type control (wild type eGFP with *MmOP*). The sterile controls were not supplemented with any AA. The suppression control (Figure 10 and Figure 11, SC) was supplemented with BocK. The wild type control (Figure 10 and Figure 11, WT) was supplemented with arginine. All samples except the sterile controls were induced with IPTG.



**Figure 10 eGFP fluorescence normalized to cell density for ADMARS 01 at various points in time after induction.** eGFP fluorescence was excited at 485 nm and detected at 530 nm. Blank, sterile control M9-Kan medium; WT, wild type control *E. coli* BL21(DE3) {pT7x3-MmOP eGFP WT}; SC, suppression control *E. coli* BL21(DE3) {pT7x3-MmOP eGFP Y40am} + 5 mM BocK at induction; ADMA, *E. coli* BL21(DE3) {pT7x3-ADMARS01OP eGFP Y40am} + 5 mM ADMA at induction; Arg, *E. coli* BL21(DE3) { pT7x3-ADMARS01OP eGFP Y40am } + 5 mM Arg at induction; -, *E. coli* BL21(DE3) { pT7x3-ADMARS01OP eGFP Y40am } without addition of any amino acid. The average of four replicates is shown each; error bars indicate the standard deviation.



**Figure 11 eGFP fluorescence normalized to cell density of ADMARS 11 at various points in time after induction.** ADMA, *E. coli* BL21(DE3) {pT7x3-ADMARS11OP eGFP Y40am} + 5 mM ADMA at induction; Arg, *E. coli* BL21(DE3) {pT7x3-ADMARS11OP eGFP Y40am} + 5 mM Arg at induction; -, *E. coli* BL21(DE3) {pT7x3-ADMARS11OP eGFP Y40am} without addition of any amino acid. All other abbreviations are as defined in Figure 10.

Before induction we made an interesting observation. We could already measure eGFP fluorescence at a noticeable level. For the wild type control, we could even detect eGFP fluorescence visually. We did not expect this but when scrutinizing the plasmid map of the ADMARS variants (Figure 9), this phenomenon could be explained. The plasmid map shows that the eGFP gene was located near the *lacI* gene which had its own promoter yet no terminator. The expression of *lacI* was driven by a constitutive promoter and due to the lack of a terminator the transcription could have read through into the downstream eGFP gene. Most probably, an mRNA was transcribed which included the *lacI* and also the eGFP gene. This could be a reason why we detected a certain background signal of eGFP fluorescence already before IPTG induction.

In Figure 10 and Figure 11, the eGFP fluorescence per cell density average values of the quadruplicates of each sample are depicted. In all diagrams, a comparable trend for the controls could be observed. The sterile controls did not show any fluorescence. The fluorescence of the wild type control (WT, dark green) increased with the induction time. Fluorescence values of the wild type eGFP expressed under T7 promoter should be around  $10^6$ <sup>37</sup>. According to the

data, values around  $10^5$  were the highest. The fluorescence of the suppression control (SC, light green) also clearly increased within the first 2 h after induction. In comparison to the wild type control, the fluorescence of the suppression control stagnated between 2 and 4 h after induction. The fluorescence increase between 4 and 16 h after induction was lower than of the wild type control. Induction was clearly successful. When looking at the ADMARS variants supplemented with ADMA (ADMA, dark blue), a different trend was observed. The fluorescence increase observed within the first 2 h after induction was considerably lower than that of the controls. No significant increase in fluorescence could be measured henceforward. This was also true for the ADMARS variants supplemented with arginine (Arg, light blue). No difference to the variants supplemented with ADMA was observed. This indicates that the ADMA-OP neither incorporated ADMA nor arginine into eGFP. When compared to the background results (-, red), both ADMARS variants supplemented with ADMA and arginine did not yield higher fluorescence values. The eGFP fluorescence for all ADMARS variants of both design groups was not above the background (Figure 10 and Figure 11, ADMA vs. Arg vs. -).

When normalized to the wild type eGFP, fluorescence of the suppression control supplemented with BocK reached on average 72.2% of the fluorescence of the wild type control. In contrast to that, the ADMARS variants supplemented with ADMA reached on average only 4.9% of the fluorescence of the wild type. This again did not exceed the background values. In summary, all eight ADMARS variants did not show eGFP fluorescence above the background, so presumably eGFP was not expressed. This could have several reasons. The ADMA concentration could have been too low or ADMA could not have been taken up by the cells from the media. Even if ADMA was taken up by the cells it could have been immediately degraded by peptidases. Assuming that ADMA was not degraded after the cellular uptake, the expression of ADMARS might not have been successful. A cause for low expression levels of ADMARS could have been the hypermutation of the designs. At least 20 positions within the enzyme have been modified and by this, the enzyme could have gotten instable and might not have been expressed successfully. If ADMARS was expressed successfully, it could not have recognized ADMA and therefore could not or not successfully have loaded the  $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$  with ADMA. To make the incorporation possible at all, either more or a more active  $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$  over the host's tRNAs was necessary. Since the charging of tRNAs with their cognate amino acids is among the most inaccurate steps of translation<sup>38</sup>, a more strongly represented  $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$  could increase the chances of correct charging with ADMA. If  $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$  was

successfully loaded with ADMA, the elongation factor Tu (EF-Tu) might not have recognized and brought the charged tRNA<sub>CUA</sub><sup>Pyl</sup> to the ribosome. The latter is highly unlikely since EF-Tu has a broad substrate tolerance<sup>39</sup>. The incorporation of ADMA at the amber position in the eGFP gene could not have occurred because the amber stop codon could have been recognized as such by the release factor. Transcription would have been terminated and the eGFP would have been expressed as a truncated nonfluorescent protein. If ADMA was incorporated, eGFP might have been expressed in its full length but the protein folding could have been impaired by the incorporation of ADMA or arginine at the amber position. By this, the eGFP fluorescence could have been inhibited. To examine this, the amber codon in the eGFP gene would have to be changed to an arginine codon. A subsequent protein expression analysis would reveal if the incorporation of arginine influences the eGFP fluorescence. A supporting experiment to examine whether ADMARS still recognizes BocK and BocK could be incorporated into eGFP could be done.

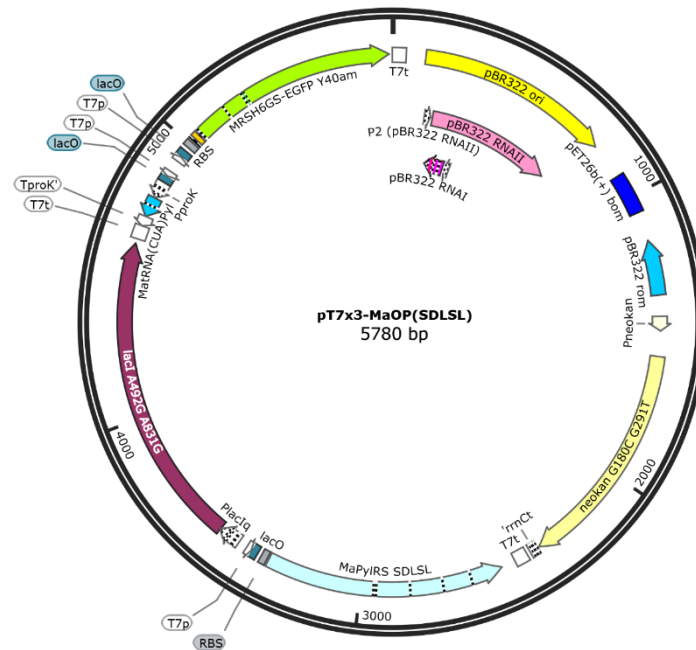
Since all these hypotheses had to be considered, we examined the possible reasons for our observation systematically. We continued with a different orthogonal pair to examine whether eGFP would be expressed and whether we would measure higher fluorescence values. We repeated the fluorescence assay under the same conditions with a plasmid carrying the SDLSL mutant of the *MaPylRS*.

### **5.3 Fluorescence assay with SDLSL mutant**

The fluorescence assay as conducted with the ADMARS variants was repeated with the plasmid pT7x3-MaOP(SDLSL) clone 2 that had emerged in a previous ADMA screen.

The plasmid pT7x3-MaOP(SDLSL) (SDLSL mutant) contained a pyrrolysyl-tRNA synthetase from *Methanomethylophilus alvus* harboring five mutations: Y126S M129D V168L Y206S W239L (Figure 12). The *MaPylRS* SDLSL gene was regulated by the *lac* operator.





**Figure 12 Plasmid map of pT7x3-MaOP(SDLSL).** The plasmid was used in the fluorescence assay of the SDLSL mutant. *MaPyIRS* SDLSL, pyrrolysyl-tRNA synthetase obtained from *Methanomethylophilus alvus* with five mutations Y126S M129D V168L Y206S W239L; *lacI* A492G A831G, lac repressor; *MatRNA*<sub>CUA</sub><sup>Pyl</sup>, amber suppressor tRNA<sub>CUA</sub><sup>Pyl</sup> from *Methanomethylophilus alvus*; MRSH6GS-EGFP Y40am, enhanced green fluorescent protein EGFP F64L S65T Y40X; pBR322 ori, origin of replication from *E. coli* plasmid pBR322; *neokan* G180C, aminoglycoside-3'-O-phosphotransferase from Transposon Tn5.

To conduct the fluorescence assay, the SDLSL mutant as well as the controls were cultivated in a 96-deep-well plate. The controls were the same as mentioned in the assay with ADMARS variants. The SDLSL mutant was supplemented with different concentrations of different amino acids. The layout of the samples on the 96-deep-well plate is shown in Table 10.

**Table 10 Sample layout for screening of SDLSL mutant in 96-well plate cultures.** Blue wells indicate the used SDLSL mutant, *E. coli* BL21(DE3) {pT7x3-MaOP(SDLSL)} clone 2. Supplementation with amino acids is given for each blue row; -, no addition of any amino acid; WT-1 to WT-4, *E. coli* BL21(DE3) {pT7x3-MmOP eGFP WT} without addition of any amino acid (4 replicates); PC-1 to PC-4, *E. coli* BL21(DE3) {pT7x3-MmOP eGFP Y40am} with addition of BocK (4 replicates); NC-1 to NC-4, *E. coli* BL21(DE3) {pT7x3-MmOP eGFP Y40am} without addition of any amino acid (4 replicates); sterile control, M9-Kan medium not inoculated.

	1	2	3	4	5	6	7	8	9	10	11	12
<b>A</b>	WT-1	5 mM ADMA										PC-1
<b>B</b>	WT-2	10 mM ADMA										PC-2
<b>C</b>	WT-3	5 mM Arg										PC-3
<b>D</b>	WT-4	10 mM Arg										PC-4
<b>E</b>	Sterile control	5 mM BocK										NC-1
<b>F</b>	Sterile control	10 mM BocK										NC-2
<b>G</b>	Sterile control	-										NC-3
<b>H</b>	Sterile control	-										NC-4

The fluorescence assay was conducted twice as previously described for the ADMARS variants. The obtained data was validated accordingly. In both assays no significant difference in eGFP fluorescence from 4 to 20.5 h after induction could be observed. The measured eGFP fluorescence values were not higher than in the previous screening with the ADMARS variants. We inoculated with a high volume per well (~1.5 mL each) which might have influenced the oxygen supply. The growth of *E. coli* cells was slower than in the previous cultivations and we did not reach the desired cell density.

Even though we conducted the fluorescence assay with a different orthogonal pair, we did not measure higher eGFP fluorescence values. We concluded that ADMA most probably was neither incorporated in the ADMARS variants nor in the SDLSL mutant. We wanted to examine whether ADMARS of the ADMA-OP was expressed. We decided to perform a protein expression analysis of ADMARS. In contrast to the previous cultivations, we decided to cultivate in a 24-well plate to exclude possible oxygen limitations.

## 5.4 SDS-PAGE

Since we did not see eGFP fluorescence above background in the previous screening, we hypothesized that the ADMARSs were not expressed to the desired extent. We analyzed the expression of the ADMARS variants in 24-well plate cultures. The layout of the samples on the 24-well plate can be seen in Table 11.

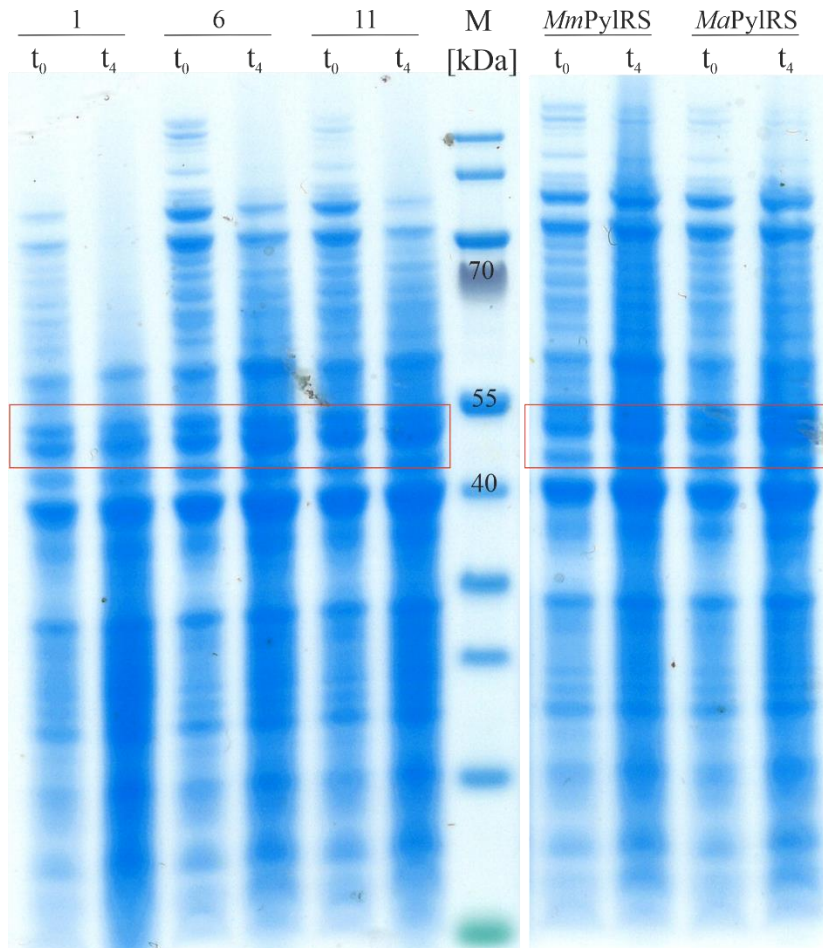
**Table 11 Sample layout for expression analysis of ADMARS variants in 24-well plate cultures.** ADMARS variants as well as controls are shown. ADMARS variants of design groups A and B are indicated in blue and bold. Sterile control, M9Kan medium not inoculated; ADMARS ##, *E. coli* BL21(DE3) {pT7x3-ADMARS##OP eGFP Y40am}; *MmPylRS* wt, *E. coli* BL21(DE3) {pT7x3-MmOP eGFP Y40am}; *MaPylRS* SDLSSL, *E. coli* BL21(DE3) {pCAT1-MaOP SDLSSL} clone 2.

	1	2	3	4	5	6
A	<b>ADMARS 01</b>	ADMARS 06	<b>ADMARS 11</b>	<b>ADMARS 12</b>	ADMARS 17	<i>MmPylRS</i> wt
B	<b>ADMARS 02</b>	ADMARS 08	ADMARS 13	ADMARS 15	ADMARS 19	Sterile control
C	<b>ADMARS 03</b>	ADMARS 09	<b>ADMARS 23</b>	ADMARS 16	ADMARS 20	<i>MaPylRS</i> SDLSSL
D	<b>ADMARS 04</b>	ADMARS 10	<b>ADMARS 25</b>	ADMARS 22	ADMARS 21	Sterile control

ADMARS variants 01, 02, 03, 04, 11, 12, 23 and 25 (Table 11) were analyzed. In addition, we analyzed *MaPylRS* SDLSSL. As a positive control we expressed the *MmPylRS* wild type (*MmPylRS* wt). Supplementation with AAs at induction was omitted.

SDS-PAGE was used to analyze whether ADMARS, *MmPylRS* and *MaPylRS* SDLSSL were expressed. Cell pellets from the cultures were collected immediately before and 4 h after induction. This included both soluble and insoluble proteins from the pellets. Analysis of expressed ADMARS on the gel is possible because of the enzymatic overexpression. This is due to the multicopy plasmid with a high copy origin of replication and the T7 promoter which controls the expression of the ADMARS. The calculated molecular weights for ADMARS, *MmPylRS* and *MaPylRS* SDLSSL on the SDS gel are 50.7 kDa, 50.9 kDa and 30.6 kDa, respectively. We hypothesized that if a remarkable difference between the protein bands before and after induction could be observed, ADMARS, *MmPylRS* and *MaPylRS* SDLSSL were expressed. This would indicate that ADMA had not been incorporated at the amber position in the eGFP gene and eGFP was therefore not expressed. If no difference between the protein bands could be observed, ADMARS, *MmPylRS* and *MaPylRS* SDLSSL were expressed only to

a very low extend or not at all. In this case, ADMA could not be incorporated into the eGFP gene and eGFP would also not be expressed. The gel picture (Figure 13) includes the ADMARS variants 01 and 11 as well as the controls ADMARS variant 06, *MmPylRS* and *MaPylRS* SDLSL immediately ( $t_0$ ) and 4 h after induction ( $t_4$ ).



**Figure 13 Expression of *MmPylRS* of ADMARS variants 01, 06 and 11.** Calculated molecular weights of ADMARS, *MmPylRS* and *MaPylRS* SDLSL (red box) are 50.7 kDa, 50.9 kDa and 30.6 kDa, respectively. Variant 01, {pT7x3-ADMARS01OP eGFP Y40am}; variant 11, {pT7x3-ADMARS11OP eGFP Y40am}. Variant 06, {pT7x3-ADMARS06OP eGFP Y40am} and *MaPylRS*, ({pCAT1-MaOP SDLSL} clone 2), were analyzed additionally. *MmPylRS* ({pT7x3-MmOP eGFP Y40am}) was applied as a positive control.  $t_0$ , right before induction;  $t_4$ , 4 h after induction. NuPAGE™ 4-12% BT 1.0 SDS gel, InstantBlue™ stained. M, PageRuler™ Prestained Protein Ladder. Sizes of relevant marker bands are indicated; irrelevant bands were cut.

The red box indicates the molecular size range where we expected the ADMARS overexpression bands. ADMARS is difficult to detect in the gel picture because the band would appear with other *E. coli* proteins at around 50 kDa. According to the gel picture, we did not observe a difference in protein expression between the begin of induction ( $t_0$ ) and 4 h later ( $t_4$ ) for all variants and controls. As can be seen in the plasmid map (Figure 9), the ADMARS gene

is located near the *lacI* gene which is constitutively expressed. An accidental expression of ADMARS before induction can be excluded since the *lacI* gene is oriented inversely than the ADMARS gene. ADMARS is under control of a tight T7 promoter. It can be concluded that ADMARS, *MmPylRS* and *MaPylRS* SDLSL were not expressed to the desired extent.

Concerning the ADMARS variants, it could have several reasons that ADMARS is not expressed. The design of the synthetic cloning fragments was based on the structure of *MmPylRS* wild type and was intended to greatly expand the substrate binding capacities of the enzyme. The mutated *MmPylRS* of the ADMARS variants should accept ncAAs like ADMA as substrates. Arginine should be excluded as far as possible. Several screening experiments and a protein expression analysis revealed no such enzyme. In fact, the expression of a mutated *MmPylRS* could not be verified. This raises the question if the design inhibits the expression of ADMARS. In a typical mutagenesis experiment, a few positions within a gene are simultaneously and randomly mutated. In comparison to that, our designs were systematically hypermutated. At least 20 positions within the enzyme have been modified with a clearly defined amino acid. By mutating the gene in such a distinctive way, the enzyme can get instable and might not be expressed successfully or folded correctly. The Rosetta energy function used in the design pictures the energy state of the molecule as accurate as possible. The energy score was optimized to design the molecule's structure with the most stable protein folding. However, a stable expression and correct folding of the protein cannot be guaranteed based on the design.

In the protein expression analysis, we also processed *MmPylRS* wild type as a positive control and *MaPylRS* SDLSL clone 2 as a comparison to the ADMARS variants. Both were processed under the same conditions as the ADMARS variants. The bands did also not indicate a difference in protein expression between the begin of induction and 4 h later. This indicates that the *MmPylRS* and *MaPylRS* SDLSL genes of the controls were also not expressed to the desired extent. For this reason, we must take account of a systematic error.

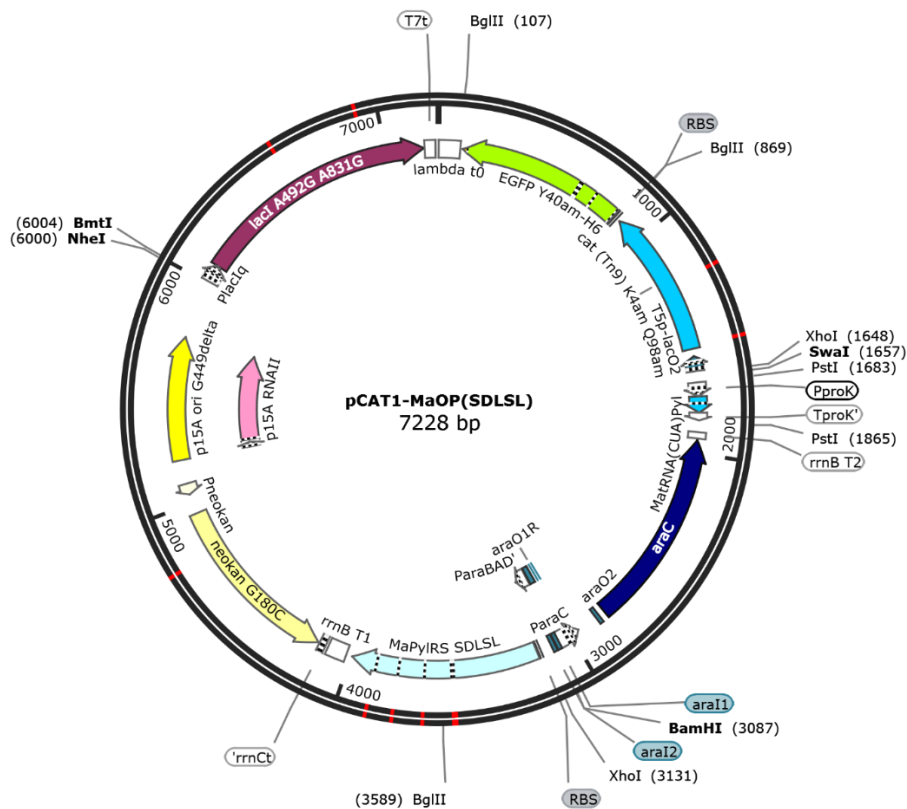
Another reason could be the ADMA-OP plasmid (pT7x3-ADMARS##OP eGFP Y40am). The ADMARS variants were sequenced and any further mutations through the processing of the genes could be excluded. Since the design cannot pose a reasonable explanation, the used expression strain should be examined. We intended to use *E. coli* BL21(DE3) as an expression strain since several genes on our plasmid including ADMARS were under control of a T7 promoter. *E. coli* BL21 can also be used for recombinant protein expression. Both *E. coli* BL21(DE3) and *E. coli* BL21 are B-strains and deficient in Lon protease and OmpT protease<sup>40</sup>.

DE3 indicates the lambdaDE3 lysogen that includes the gene for T7 RNA polymerase to express recombinant genes under control of a T7 promoter. BL21 does not carry the gene for T7 RNA polymerase<sup>41</sup>. We could have accidentally interchanged the *E. coli* expression strains. The strain we used might not have been BL21(DE3) but only BL21. If the strain used by us had not been BL21(DE3) but BL21, we would have not achieved the desired expression of ADMARS which is downstream of a T7 promoter and would have not been able to detect ADMARS on the gel.

As a conclusion it can be said that freshly prepared competent *E. coli* BL21(DE3) must be transformed again with the ADMA-OP plasmids to ensure that the correct expression strain is used. This would exceed the scope of my thesis and therefore I repeated the chloramphenicol growth assay with the SDLSL mutants from previous experiments.

## 5.5 Chloramphenicol growth assay

With the chloramphenicol growth assay, we wanted to analyze *E. coli* BL21(DE3) {pCAT1-MaOP(SDLSL)} and *E. coli* EXPRESS BL21(DE3) {pCAT1-MaOP(SDLSL)} for their ability to incorporate ADMA into the chloramphenicol (Cm) resistance marker. The plasmid (Figure 14) included the pyrrolysyl-tRNA synthetase from *Methanomethylophilus alvus* carrying the mutations Y126S M129D V168L Y206S W239L (*MaPylRS* SDLSL). *MaPylRS* SDLSL was selected from a 5NNK library for the incorporation of ADMA. The growth assay had been performed previously and had revealed functional expression of the chloramphenicol acetyltransferase (*CAT*) gene. To confirm these results, we repeated the growth assay under the exact same conditions. Two clones of the plasmid pCAT1-MaOP SDLSL (clone 2 and 3) were examined. Both carried two amber mutations K4am and Q98am in the *CAT* gene, which confers Cm resistance.



**Figure 14 Plasmid map of pCAT1-MaOP(SDLSL).** The plasmid was used to transform *E. coli* BL21(DE3) and *E. coli* EXPRESS BL21(DE3) cells for the chloramphenicol growth assay. *MaPyIRS* SDLSL, pyrrolysyl-tRNA synthetase obtained from *Methanomethylophilus alvus* with five mutations Y126S M129D V168L Y206S W239L. *lacI* A492G A831G, lac repressor. EGFP Y40am-H6, enhanced green fluorescent protein EGFP F64L S65T Y40X. *araC*, DNA-binding transcriptional dual regulator. *cat* (Tn9) K4am Q98am, chloramphenicol acetyltransferase containing two amber mutations K4am and Q98am. *neoKan* G180C, aminoglycoside-3'-O-phosphotransferase from transposon Tn5.

We aimed to select cells by the ability to resist Cm. Cm resistance was possible when the resistance gene *CAT* was expressed functionally. The *MaPyIRS* SDLSL supplied with ADMA should have enabled the production of the *CAT* protein and therefore Cm resistance. Supplementation of cAAs should have not enabled the production of the *CAT* protein. We hypothesized that the higher the efficiency of the *MaOP*, the better the amber mutations were suppressed. The better the amber mutations in the *CAT* gene were suppressed, the higher the Cm resistance and thus the growth rate in presence of Cm. *E. coli* BL21(DE3) cells transformed with the plasmid pCAT1WT-MaOP carrying the wild type *CAT* gene were used as a positive control. *E. coli* BL21(DE3) cells transformed with the plasmid pCAT1K4am-MaOP carrying the wild type *MaPyIRS* and the amber mutated *CAT* gene were used as a negative control. The wild type *MaPyIRS* could only incorporate BocK, Pyl or other ncAAs usually not present in *E. coli*. Both *E. coli* BL21(DE3) and *E. coli* EXPRESS BL21(DE3) cells were transformed

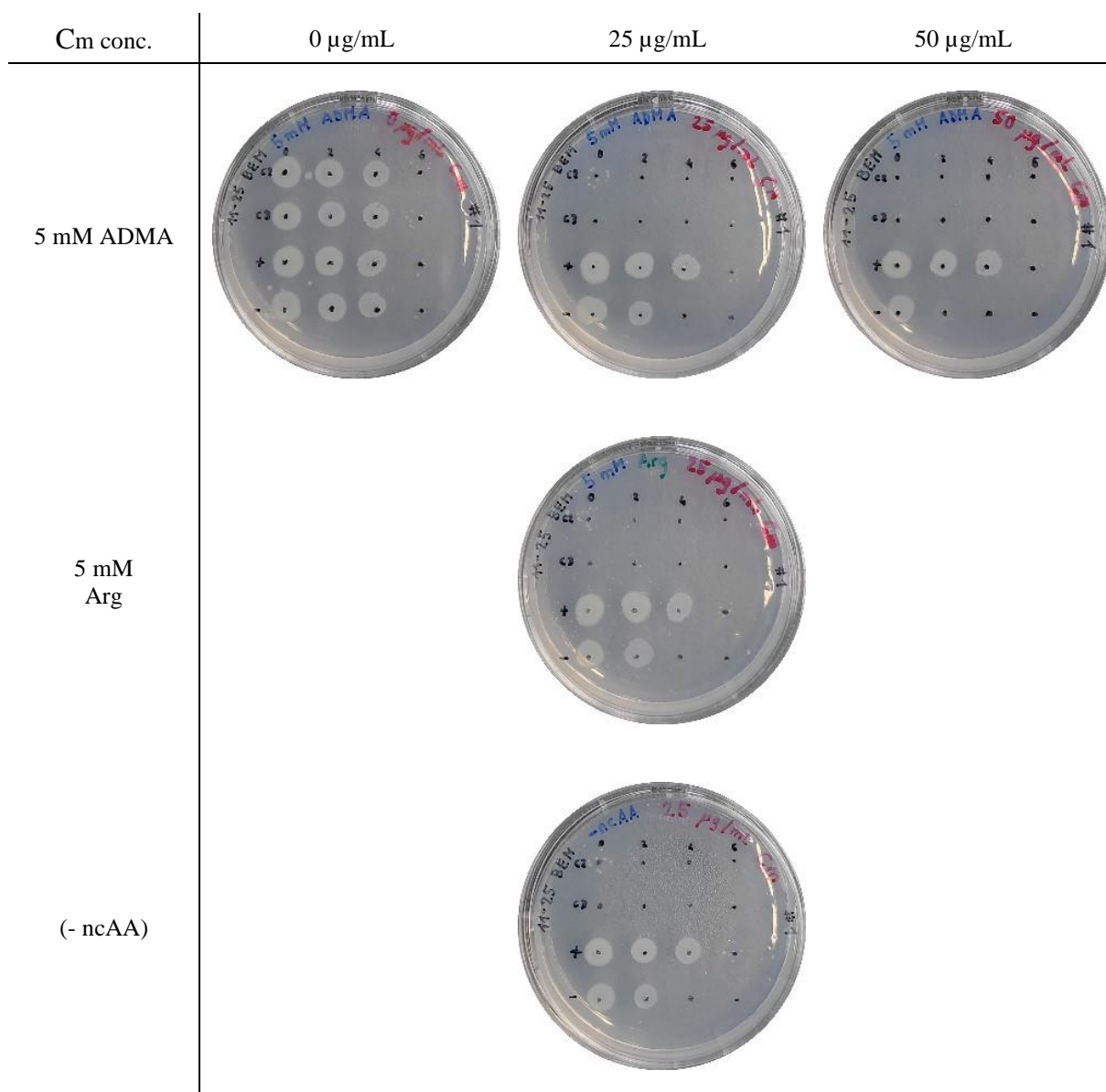
with *MaPylRS* SDLSL clone 2 and 3. The sample layout of the preparation for the growth assay is shown in Table 12.

**Table 12 Sample layout of the preparation for the chloramphenicol growth assay.** Ecl., *E. cloni* EXPRESS BL21(DE3); Eco., *E. coli* BL21(DE3); C2, pCAT1-MaOP SDLSL clone 2; C3, pCAT1-MaOP SDLSL clone 3; PC, positive control pCAT1WT-MaOP; NC, negative control pCAT1K4am-MaOP; /n, dilution factor 10<sup>n</sup>.

	1	2	3	4
A	Ecl. C2/0	Ecl. C2/2	Ecl. C2/4	Ecl. C2/6
B	Ecl. C3/0	Ecl. C3/2	Ecl. C3/4	Ecl. C3/6
C	Eco. C2/0	Eco. C2/2	Eco. C2/4	Eco. C2/6
D	Eco. C3/0	Eco. C3/2	Eco. C3/4	Eco. C3/6
E	Eco. PC/0	Eco. PC/2	Eco. PC/4	Eco. PC/6
F	Eco. NC/0	Eco. NC/2	Eco. NC/4	Eco. NC/6

Cells were diluted according to the dilution factors given in Table 12 and spotted onto the M9-Kan plates in 4x4 and 4x2 schemes. The M9-Kan plates with *E. coli* BL21(DE3) and *E. cloni* EXPRESS BL21(DE3) are shown in Figure 15 and Figure 16.



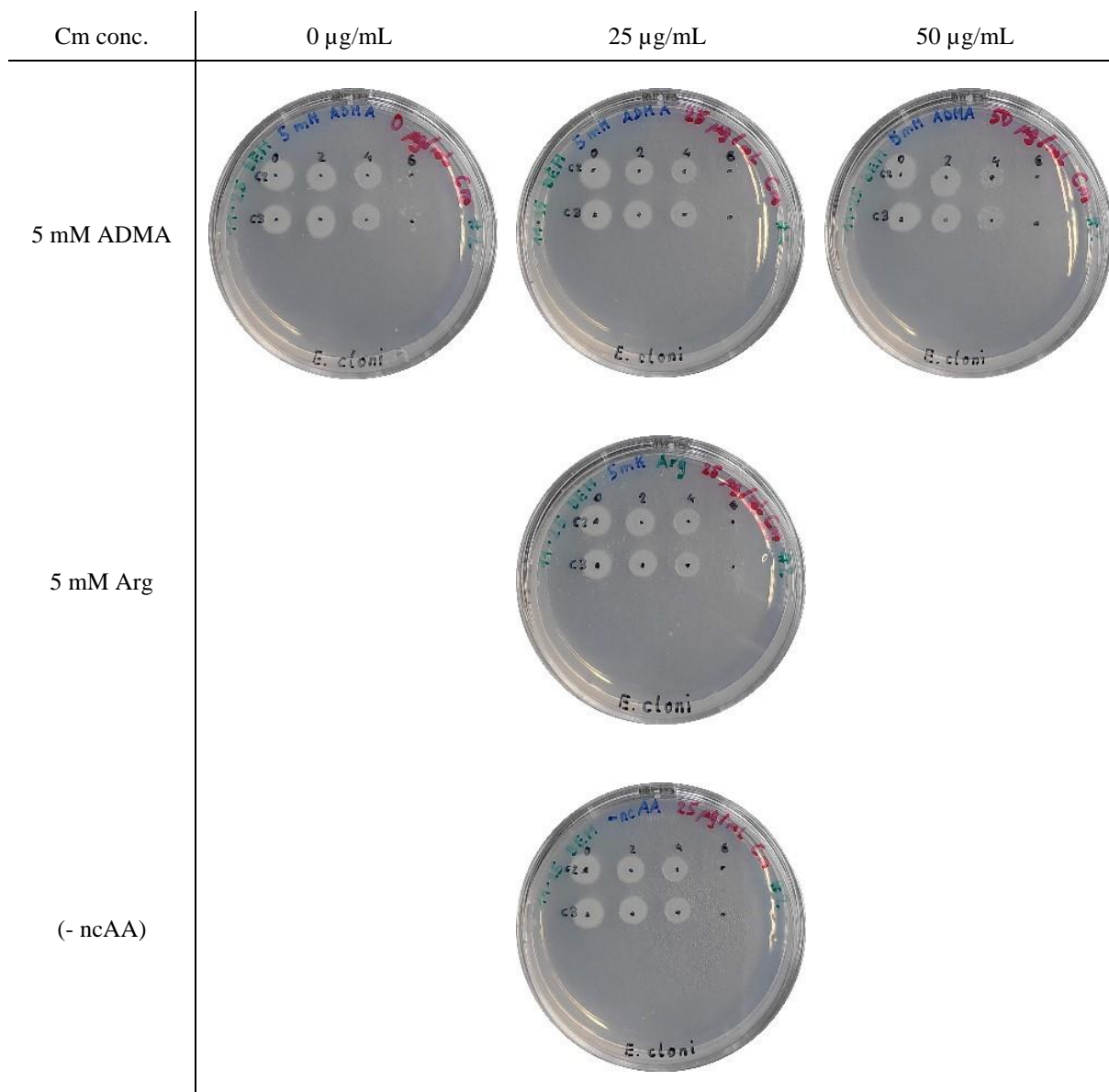


**Figure 15 Chloramphenicol growth assay M9-Kan plates with *E. coli* BL21(DE3) under various conditions after 45 h.** The assay was performed with 5 mM ADMA as an ncAA and different chloramphenicol (Cm) concentrations. Plates with addition of arginine (5 mM Arg) as well as no amino acid supply (- ncAA) were used as controls. C2, *E. coli* BL21(DE3) {pCAT1-MaOP SDLSL} clone 2; C3, *E. coli* BL21(DE3) {pCAT1-MaOP SDLSL} clone 3; +, *E. coli* BL21(DE3) {pCAT1WT-MaOP}; -, *E. coli* BL21(DE3) {pCAT1K4am-MaOP}; 0 to 6 indicate the dilution factors.

As can be seen in Figure 15, the positive control (+) grew equally well under all conditions. Its growth seemed to be unaffected by the Cm concentration. Against our expectation, the negative control (-) also grew under all conditions but varied with the Cm concentration. Growth was reduced at 50  $\mu\text{g/mL}$  Cm and increased at 0  $\mu\text{g/mL}$  Cm showing a clear dependence on Cm. The cells of the negative control grew equally well independent of the amino acid supplementation. The growth of the negative control is, however, highly unlikely when considering the components of the plasmid. pCAT1K4am-MaOP contained the wild type

*MaOP* which could not incorporate ADMA as an ncAA. The plasmid also carried the *CAT* gene with a K4am mutation requiring an incorporation of ADMA at the amber position to confer Cm resistance. Since the wild type *MaOP* could not incorporate ADMA and the K4am mutation lied far upstream in the gene, the *CAT* protein could not have been functionally expressed and the cells could not have grown. Yet colonies appeared indicating that the *CAT* protein somehow was functionally expressed. The amber mutation at position K4 should have been read as a stop codon so that the *CAT* gene was not translated. When scrutinizing the DNA sequence of the *CAT* gene, we found the nucleotide sequence 5'-GGAG-3' at amino acid positions 71 and 72 six bases upstream of an ATG codon. This sequence partially correlates with the Shine-Dalgarno (SD) sequence (5'-AGGAGG-3')<sup>42</sup> revealing a hidden ribosomal binding site (RBS) six bases upstream of the next ATG start codon. Consequently, this RBS could have promoted translational initiation yet less efficient than a complete SD sequence. By this, an N-terminally truncated *CAT* protein could have been expressed which would have been less active than a fully expressed protein but still able to confer Cm resistance. With this insight, the *CAT* resistance gene was not suited for the growth assay. A better choice for the negative control would have been a pCAT vector including two amber mutations within the *CAT* gene.

*E. coli* BL21(DE3) SDLSL clones 2 and 3 did not yet display growth in the presence of chloramphenicol after 24 h. After 45 h, the clones showed equal growth as the controls but only in the absence of Cm (Figure 15, 5 mM ADMA, 0 µg/mL Cm). Cells did neither grow at higher Cm concentrations nor with Arg or no AA supplementation, respectively. This indicated that ADMA was not incorporated at the two amber positions and the *CAT* protein which conferred Cm resistance was not expressed. Still, colonies appeared on the plate supplemented with 5 mM ADMA and without addition of Cm since a functional expression of the *CAT* protein is not mandatory for cell growth.



**Figure 16 Chloramphenicol growth assay M9-Kan plates with *E. coli* EXPRESS BL21(DE3) under various conditions after 45 h.** C2, *E. coli* BL21(DE3) {pCAT1-MaOP SDLSL} clone 2; C3, *E. coli* BL21(DE3) {pCAT1-MaOP SDLSL} clone 3; 0 to 6 indicate dilution factors. All other abbreviations are as defined in Figure 15.

*E. coli* EXPRESS BL21(DE3) SDLSL clones 2 and 3 grew under all conditions (Figure 16). Their growth rate appeared to be negatively correlated to the Cm concentration. We saw moderate growth with 25  $\mu\text{g/mL}$  Cm, more growth in the absence of Cm and less growth with 50  $\mu\text{g/mL}$  Cm. Cells grew on plates with 5 mM ADMA, 5 mM Arg and without AA supplementation. It can be suggested that not only ADMA but also Arg as well as other cAAs were incorporated at the two amber positions in the *CAT* gene. We could not confirm the promising results of the previously performed chloramphenicol growth assay.

When validating the results of our growth assay, we identified a calculation error in the preparation of both *E. coli* BL21(DE3) and *E. coli* EXPRESS BL21(DE3) M9-Kan plates. The volume of the 50 mM ADMA stock necessary to obtain a final concentration of 5 mM in the medium was miscalculated. The ultimate concentration of ADMA in the medium was too low to even expect an incorporation.

The assay did not confirm the ability of the *MaPylRS* SDLSLs to incorporate ADMA at an amber position. The negative control grew under all conditions most likely due to a hidden RBS. A repetition of the experiment would be necessary to examine whether ADMA will be incorporated when ADMA is supplemented at the correct concentration and an appropriate negative control is used.

## 6. Conclusion

Eight synthetic cloning fragments were inserted into the backbone plasmid pT7x3-MmOP-eGFP Y40am to obtain a plasmid library. The synthetic cloning fragments were designed constructs harboring highly mutated pyrrolysyl-tRNA synthetase genes from *Methanosarcina mazei* (*MmPylRS* variants). In the enzymatic digest, compatibility issues and limited restriction activity of the enzymes *EcoRI* and *BspTI* challenged the preparation of completely digested backbone vector. By purifying the backbone twice from an agarose gel after enzymatic digest we could obtain it in an adequate purity to proceed in ligation with the similarly digested inserts. Performing a restriction analysis with the enzymes *EcoRI* and *HindIII* helped us to distinguish religated constructs from ligated backbone and insert. Revision of the ligation protocol also resulted in higher numbers of successfully ligated backbones with corresponding inserts, which were named ADMARS variants. With those ADMARS variants we aimed to incorporate ADMA site-specifically at the amber position Y40 in the eGFP gene. We conducted a fluorescence assay with *E. coli* BL21(DE3) cells transformed with the previously obtained ADMARS variants. We examined different cultivation conditions and concluded that M9-Kan medium with 0.5 g/L yeast extract and incubation at 37 °C and 320 rpm was best suited. Additionally, we added plastic beads in each well of the 96-deep well cultivation plate to improve mixing.

The fluorescence assay was performed twice by measuring cell densities and eGFP fluorescence values in constant intervals of 3-4 h. Both wild type and suppression control showed an increase in fluorescence with the induction time. The sterile controls did not show any fluorescence which confirmed a correct experimental procedure. Still, showing the highest values around  $10^5$ , wild type eGFP fluorescence was 10 times lower than what can be expected<sup>37</sup>. Contrarily, the ADMARS variants did not show any significant increase in eGFP fluorescence after induction. The data obtained from two assays clearly indicated that ADMA was not incorporated into eGFP by the ADMARSs. In fact, ADMARS variants supplemented with ADMA reached on average only 4.9% of the wild type eGFP fluorescence which was not above the background. Since we observed this for all eight ADMARS variants, we presumed that eGFP was not expressed. To scrutinize this, we repeated the fluorescence assay under the same conditions with a different o-pair. PylRS from *Methanomethylophilus alvus* harboring five mutations Y126S M129D V168L Y206S W239L (*MaPylRS* SDL<sup>SL</sup>) was used. Wild type, suppression and sterile control remained the same and showed fluorescence values similar to the results obtained with the ADMARS variants. As before, no significant increase in eGFP

fluorescence after induction could be observed for *MaPylRS* SDLSL. We concluded that ADMA was most probably neither incorporated by the ADMARS variants nor by *MaPylRS* SDLSL. A number of reasons could be given as explanations, but one seemed to be particularly likely: The expression of ADMARS might not have been successful. Based on this hypothesis, we decided to perform a protein expression analysis to examine whether ADMARS was expressed.

ADMARS variants were cultivated in 24-well plate cultures omitting AA supplementation. Additionally, *MaPylRS* SDLSL as well as *MmPylRS* wild type as a positive control were cultivated. Cell pellets from the cultures were collected immediately before and 4 h after induction. If the protein bands before and after induction clearly differed, we hypothesized an expression of ADMARS, *MaPylRS* SDLSL and *MmPylRS*. The SDS-gel did not verify the enzyme expression since we did not observe a difference between the protein bands before and after induction for all mutants and controls. We concluded that ADMARS, *MaPylRS* SDLSL and *MmPylRS* were not expressed to the desired extent. An explanation could have been the hypermutation of the ADMARSs in which at least 20 positions within the enzyme have been modified. By this, the enzyme could have become instable and might not have been expressed successfully. But we also processed *MmPylRS* wild type as a positive control and *MaPylRS* SDLSL as a comparison to the ADMARS variants and neither saw enzymatic expression. The underlying reason must have been a similarity among the mutants and the controls. Crucial for enzymatic expression is the expression strain. We intended to use *E. coli* BL21(DE3) as an expression strain but *E. coli* BL21 can also be used for recombinant protein expression. However, *E. coli* BL21 does not carry the gene for T7 RNA polymerase. Since several genes on our plasmid including ADMARS were under control of a T7 promoter, *E. coli* BL21 would not have expressed the ADMARSs. An accidental confusion of the *E. coli* expression strains could be an explanation for the unexpressed enzymes. A repetition of the fluorescence assay and the protein expression analysis with newly transformed *E. coli* BL21(DE3) would be necessary to ensure that the correct expression strain is used.

We repeated a chloramphenicol growth assay to analyze *E. coli* BL21(DE3) {pCAT1-MaOP(SDLSL)} and *E. coli* EXPRESS BL21(DE3) {pCAT1-MaOP(SDLSL)} for the ability of *MaPylRS* SDLSL to incorporate ADMA into a chloramphenicol acetyltransferase (*CAT*) gene harboring two amber mutations K4am and Q98am. A plasmid carrying the *MaPylRS* and the wild type *CAT* gene was used as a positive control, whereas a plasmid carrying the wild type *MaPylRS* and the amber mutated *CAT* gene was used as a negative control. Cells were

grown in different dilutions on plates supplemented with ADMA, Arg and without any AA addition as well as different chloramphenicol (Cm) concentrations. *E. coli* BL21(DE3) cells with the positive control grew equally well independent of the AA supplementation and the Cm concentrations. Against our expectation, the negative control also grew independent of the AA supplementation, but growth was reduced at higher Cm concentrations. Four bases usually present in a ribosomal binding site (RBS) were found in the DNA sequence of the *CAT* gene six bases upstream of an ATG start codon. It is likely that this was a hidden RBS causing the expression of the *CAT* protein and enabling cell growth. *E. coli* BL21(DE3) {pCAT1-MaOP(SDLSL)} cells only grew in the absence of Cm indicating no incorporation of ADMA at the amber positions and therefore no expression of the *CAT* protein. *E. coli* EXPRESS BL21(DE3) {pCAT1-MaOP(SDLSL)} cells grew independent of the AA supplementation, but growth was reduced at higher Cm concentrations. We assumed that not only ADMA but also Arg as well as other cAAs were incorporated at the amber positions in the *CAT* gene. Validating the results of the growth assay revealed an erroneous preparation of the plates with an ultimate ADMA concentration too low to even expect an incorporation. We could not confirm the ability of *MaPylRS* SDLSL to incorporate ADMA at amber positions but revealed the plasmid pCAT1-K4am MaOP inapplicable as a negative control due to the expression of a truncated *CAT* protein driven by the hidden RBS.

## 7. Outlook

Within our series of experiments, we did not find an ADMARS variant able to incorporate ADMA as an ncAA into the eGFP gene at an in-frame amber position. Nevertheless, there are next steps that should be considered. The repetition of a protein expression analysis with newly transformed *E. coli* BL21(DE3) would be of great importance and could deliver three results.

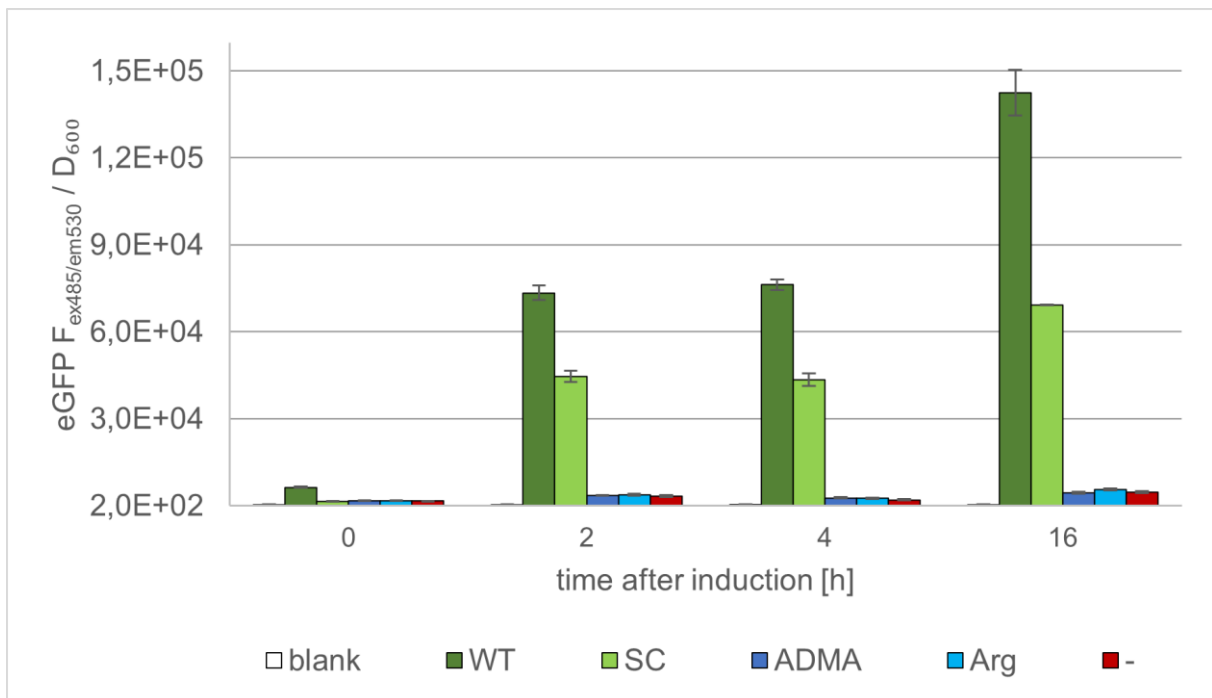
Firstly, ADMARSs could again not be expressed to the desired extent but *MaPylRS* SDLSL and *MmPylRS* as controls would be. This could be due to insolubility or multimerization of ADMARSs but also coagulation with other proteins. ADMARS was under control of a T7<sub>lacO</sub> promoter. The inducer IPTG could have had a decreased binding strength to the repressor, or the inhibitor could have had an increased binding strength to the *lac* operator. Both cases would have negatively affected ADMARS expression. The ADMARS gene could be transferred into another expression vector to see whether the construction of the plasmid pT7x3-ADMARS##OP eGFP Y40am negatively influenced ADMARS expression. A cause for low expression levels of ADMARS could have been the hypermutation of the designs. Even though the energy score of the Rosetta script which was used for the protein design was optimized to obtain a stable protein, a successful expression and correct folding of the enzyme cannot be guaranteed. A possibility would be a redesign of the o-pair. Secondly, it could be expected that ADMARS variants, *MaPylRS* SDLSL and *MmPylRS* would not be expressed to the desired extent. Any other similarity among the mutants than the expression strain would be the reason. The cultivation conditions could have been to the disadvantage of protein production. We could for instance use another cultivation medium, adapt salt concentrations or the pH. Furthermore, we could incubate the cultures longer than 4 h after induction. Thirdly, ADMARS variants, *MaPylRS* SDLSL and *MmPylRS* could be expressed to the desired extent. The accidental use of *E. coli* BL21 instead of *E. coli* BL21(DE3) in the previous experiment could be confirmed and we could repeat the fluorescence assay.

As soon as the protein expression analysis confirms the successful expression of ADMARS variants, *MaPylRS* SDLSL and *MmPylRS*, we could perform another fluorescence assay to examine the incorporation of ADMA at amber position. We could expect two results. Firstly, eGFP fluorescence of the ADMARSs could again not be above background. This would mean that ADMA was not incorporated into the eGFP gene. The cells could have struggled with the uptake of ADMA from the medium or, if ADMA was taken up by the cells, it could have been degraded immediately. HPLC-MS measurements could be done to analyze the intracellular

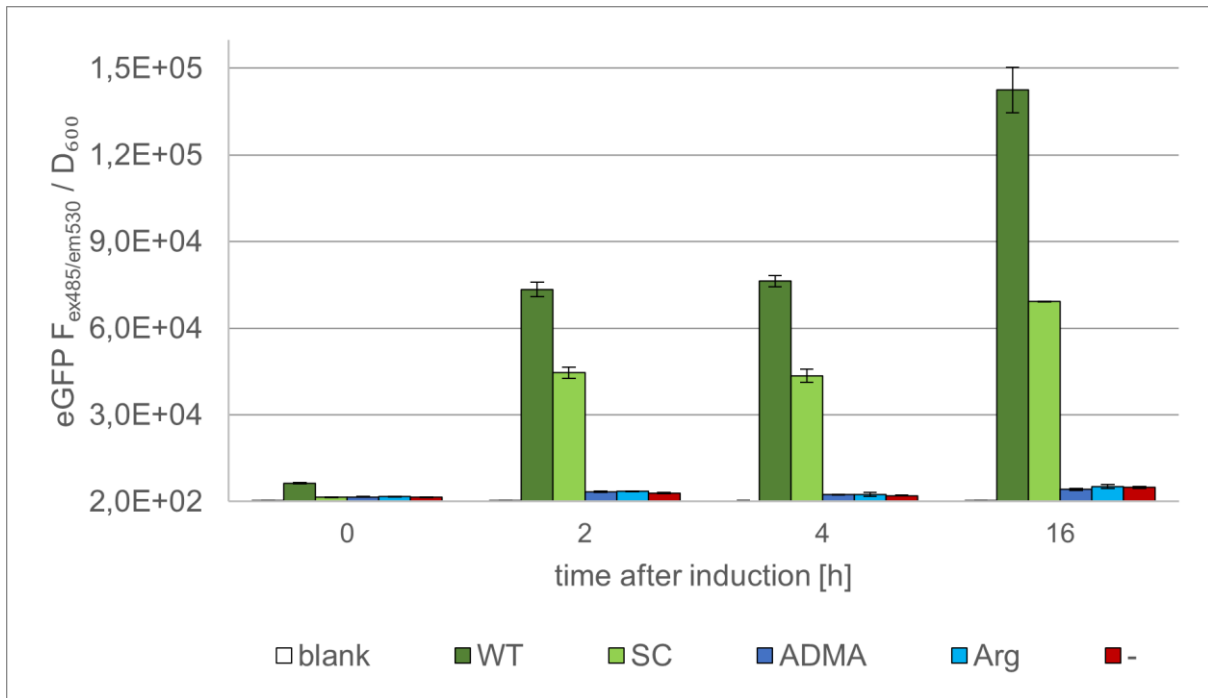


ADMA concentration. Furthermore, we could examine the expression of tRNA<sub>CUA</sub><sup>Py1</sup> and of eGFP. tRNA expression could be analyzed with a Northern blot. A successful expression of tRNA<sub>CUA</sub><sup>Py1</sup> would either indicate that the tRNA<sub>CUA</sub><sup>Py1</sup> is not charged efficiently by the ADMARS or the tRNA<sub>CUA</sub><sup>Py1</sup> is not represented more strongly than the host's tRNAs. No expression of tRNA<sub>CUA</sub><sup>Py1</sup> would open another chapter of troubleshooting. eGFP expression could be analyzed with SDS-PAGE. If eGFP would be expressed as a truncated protein, the amber stop codon would have not been suppressed with ADMA. If eGFP would be expressed in its full length, the incorporation of ADMA might have impaired the protein folding. By this, the eGFP fluorescence could have been inhibited. To examine this, the amber codon in the eGFP gene would have to be changed to an arginine codon. A subsequent protein expression analysis would reveal if the incorporation of arginine influences the eGFP fluorescence. Secondly, it could be expected that the eGFP fluorescence of the ADMARS variants would be above background. ADMA would have been incorporated into eGFP and we would be successful in obtaining an ADMARS/tRNA<sub>CUA</sub><sup>Py1</sup> o-pair able to incorporate ADMA at an amber position.

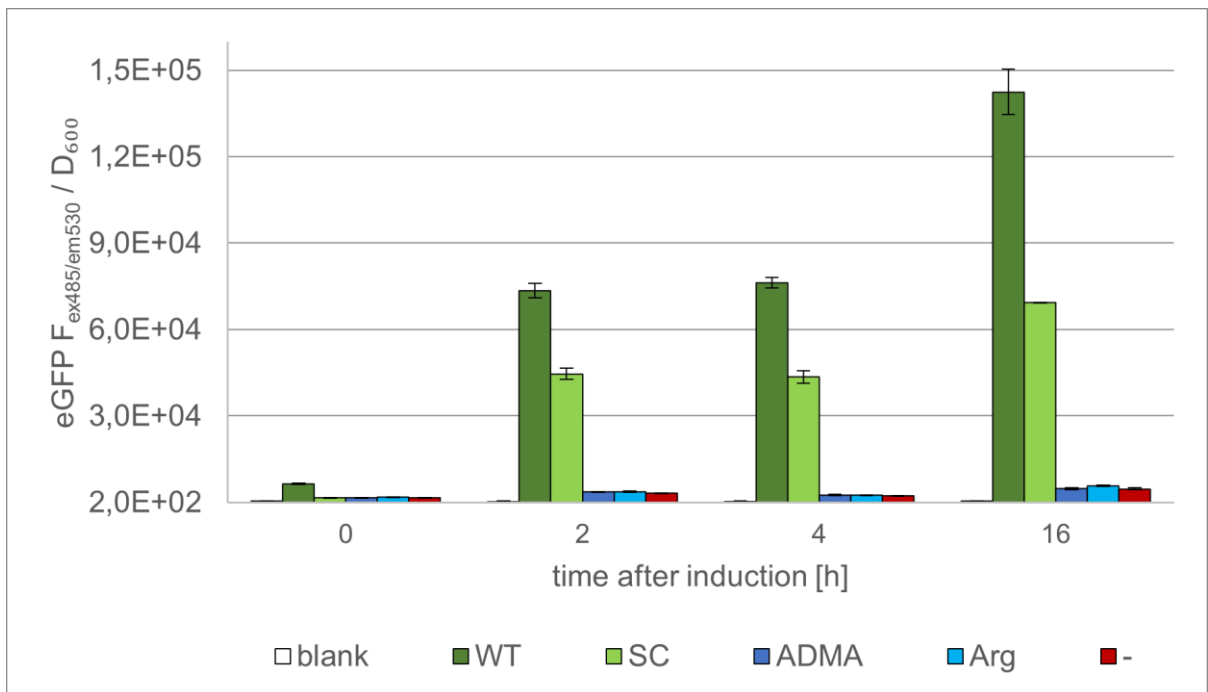
## Supplementary figures



**Figure 17 eGFP fluorescence normalized to cell density of ADMARS 02 at various points in time after induction.** eGFP fluorescence was excited at 485 nm and detected at 530 nm. Blank, sterile control M9Kan medium; WT, wild type control *E. coli* BL21(DE3) {pT7x3-MmOP eGFP WT}; SC, suppression control *E. coli* BL21(DE3) {pT7x3-MmOP eGFP Y40am} + 5 mM BocK at induction; ADMA, *E. coli* BL21(DE3) {pT7x3-ADMARS02OP eGFP Y40am} + 5 mM ADMA at induction; Arg, *E. coli* BL21(DE3) { pT7x3-ADMARS02OP eGFP Y40am } + 5 mM Arg at induction; -, *E. coli* BL21(DE3) { pT7x3-ADMARS02OP eGFP Y40am } without addition of any amino acid. The average of four replicates is shown each; error bars indicate the standard deviation.

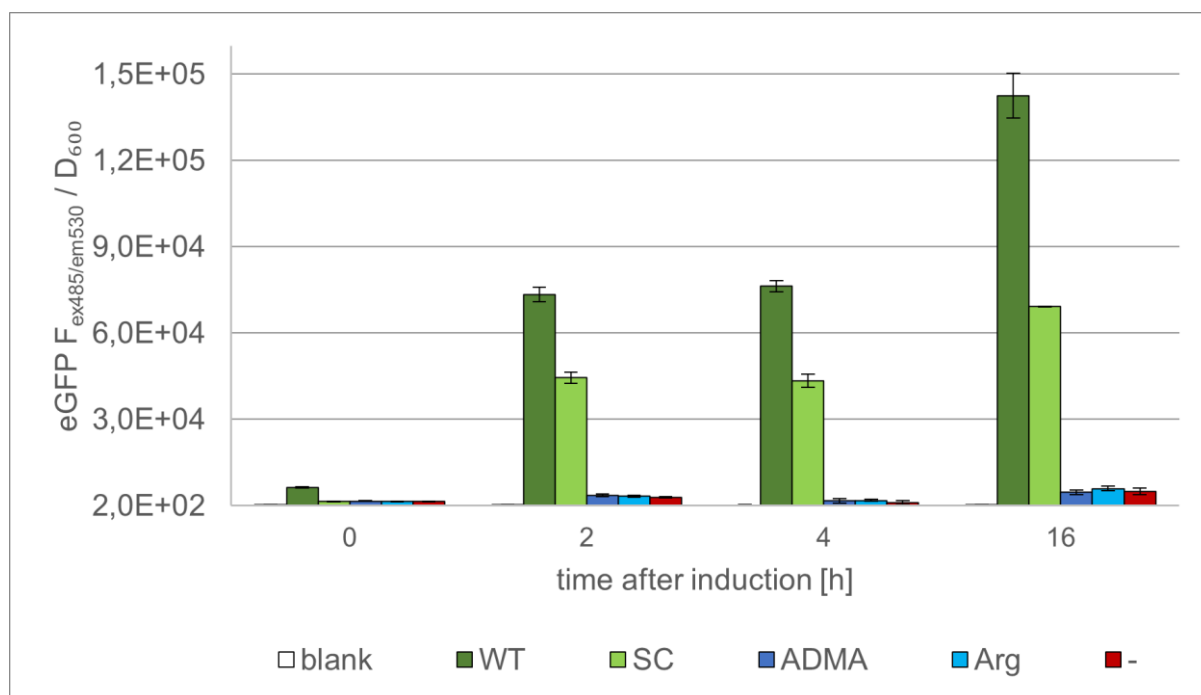


**Figure 18 eGFP fluorescence normalized to cell density of ADMARS 03 at various points in time after induction.** ADMA, *E. coli* BL21(DE3) {pT7x3-ADMARS03OP eGFP Y40am} + 5 mM ADMA at induction; Arg, *E. coli* BL21(DE3) {pT7x3-ADMARS03OP eGFP Y40am} + 5 mM Arg at induction; -, *E. coli* BL21(DE3) {pT7x3-ADMARS03OP eGFP Y40am} without addition of any amino acid. All other abbreviations are as defined in Figure 17.

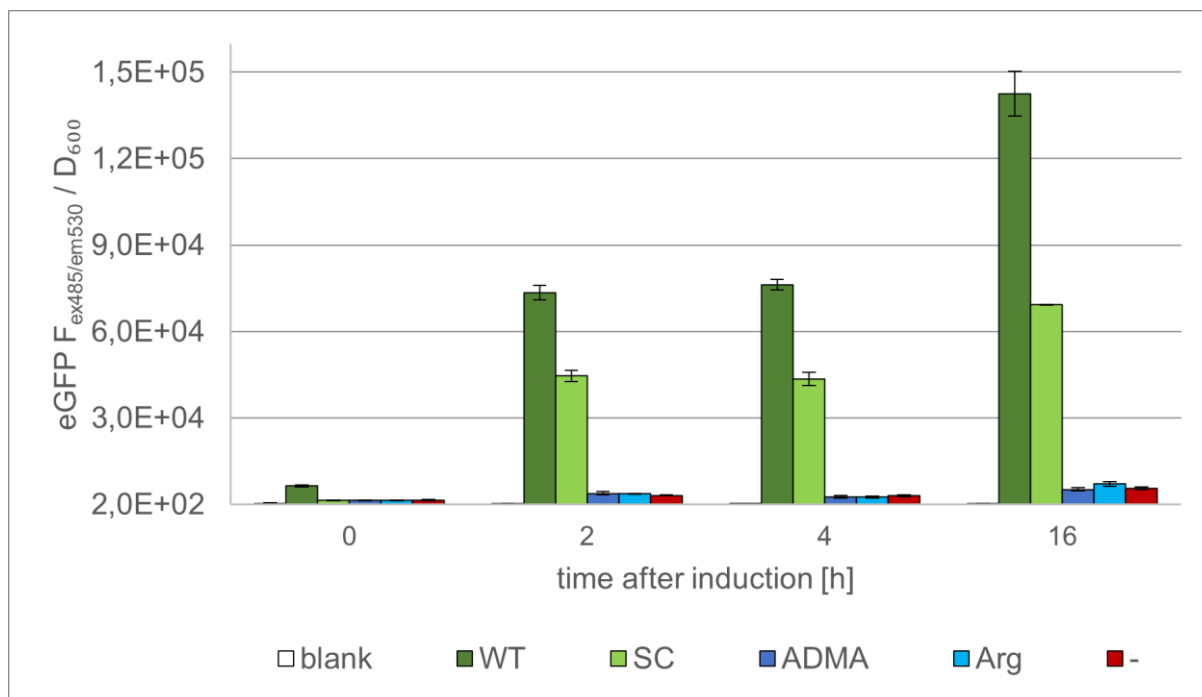


**Figure 19 eGFP fluorescence normalized to cell density of ADMARS 04 at various points in time after induction.** ADMA, *E. coli* BL21(DE3) {pT7x3-ADMARS04OP eGFP Y40am} + 5 mM ADMA at induction; Arg, *E. coli* BL21(DE3) {pT7x3-ADMARS04OP eGFP Y40am} + 5 mM Arg at induction; -, *E. coli* BL21(DE3) {pT7x3-ADMARS04OP eGFP Y40am} without addition of any amino acid. All other abbreviations are as defined in Figure 17.

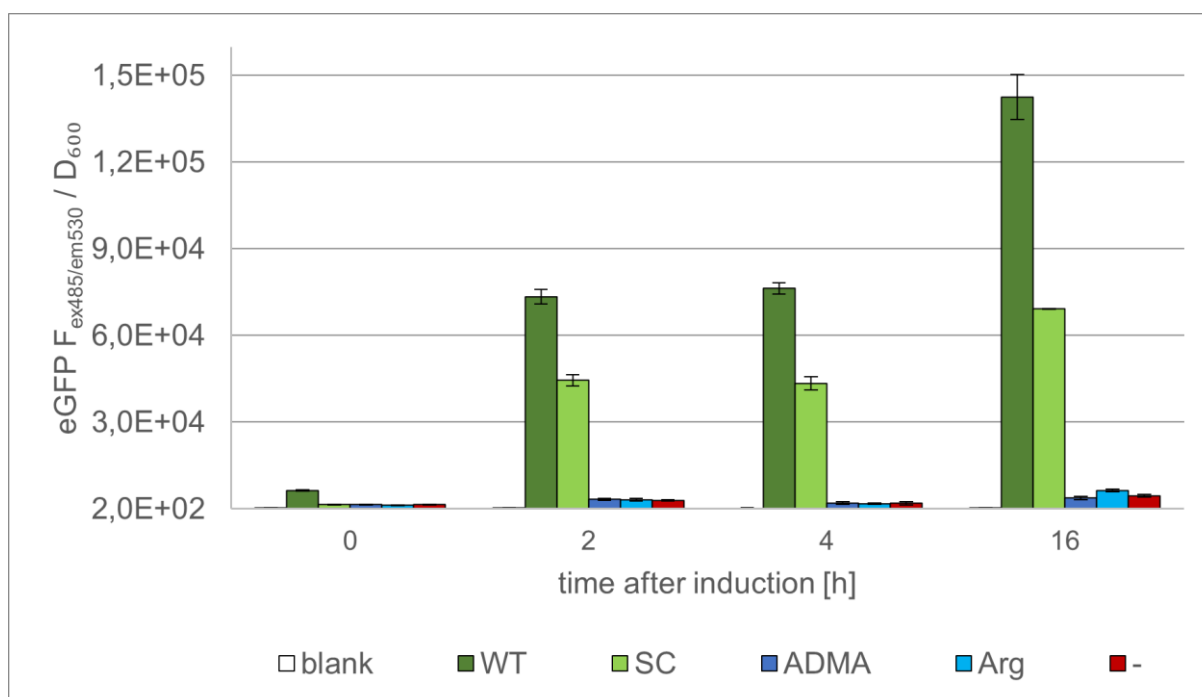
{pT7x3-ADMARS04OP eGFP Y40am} without addition of any amino acid. All other abbreviations are as defined in Figure 17.



**Figure 20 eGFP fluorescence normalized to cell density of ADMARS 12 at various points in time after induction.** ADMA, *E. coli* BL21(DE3) {pT7x3-ADMARS12OP eGFP Y40am} + 5 mM ADMA at induction; Arg, *E. coli* BL21(DE3) {pT7x3-ADMARS12OP eGFP Y40am} + 5 mM Arg at induction; -, *E. coli* BL21(DE3) {pT7x3-ADMARS12OP eGFP Y40am} without addition of any amino acid. All other abbreviations are as defined in Figure 17.



**Figure 21 eGFP fluorescence normalized to cell density of ADMARS 23 at various points in time after induction.** ADMA, *E. coli* BL21(DE3) {pT7x3-ADMARS23OP eGFP Y40am} + 5 mM ADMA at induction; Arg, *E. coli* BL21(DE3) {pT7x3-ADMARS23OP eGFP Y40am} + 5 mM Arg at induction; -, *E. coli* BL21(DE3) {pT7x3-ADMARS23OP eGFP Y40am} without addition of any amino acid. All other abbreviations are as defined in Figure 17.



**Figure 22 eGFP fluorescence normalized to cell density of ADMARS 25 at various points in time after induction.** ADMA, *E. coli* BL21(DE3) {pT7x3-ADMARS25OP eGFP Y40am} + 5 mM ADMA at induction; Arg, *E. coli* BL21(DE3) {pT7x3-ADMARS25OP eGFP Y40am} + 5 mM Arg at induction; -, *E. coli* BL21(DE3) {pT7x3-ADMARS25OP eGFP Y40am} without addition of any amino acid. All other abbreviations are as defined in Figure 17.

{pT7x3-ADMARS25OP eGFP Y40am} without addition of any amino acid. All other abbreviations are as defined in Figure 17.

## List of figures, tables and equations

Figure 1 Aminoacylation of tRNAs. ....	11
Figure 2 Plasmid map pT7x3-MmOP eGFP Y40am (plasmid GAM67).....	34
Figure 3 Enzymatic digest of backbone. ....	36
Figure 4 Enzymatic digest of inserts ID-1 and ID-2. ....	38
Figure 5 Enzymatic digest of inserts ID-3 and ID-4. ....	38
Figure 6 Enzymatic digest of inserts ID-11, ID-12, ID-23 and ID-25. ....	39
Figure 7 Restriction analysis of ligation of backbone and insert ID-4 with <i>EcoRI</i> and <i>HindIII</i> . ....	40
Figure 8 Colonies of <i>E. coli</i> Top10 F' on LB-Kan plates after transformation with ligation product of backbone and insert ID-3. ....	41
Figure 9 Plasmid map of pT7x3-ADMARS03OP eGFP Y40am (ADMARS 03).....	42
Figure 10 eGFP fluorescence normalized to cell density for ADMARS 01 at various points in time after induction. ....	45
Figure 11 eGFP fluorescence normalized to cell density of ADMARS 11 at various points in time after induction. ....	46
Figure 12 Plasmid map of pT7x3-MaOP(SDLSSL). ....	49
Figure 13 Expression of <i>MmPylRS</i> of ADMARS variants 01, 06 and 11. ....	52
Figure 14 Plasmid map of pCAT1-MaOP(SDLSSL).....	55
Figure 15 Chloramphenicol growth assay M9-Kan plates with <i>E. coli</i> BL21(DE3) under various conditions after 45 h. ....	57
Figure 16 Chloramphenicol growth assay M9-Kan plates with <i>E. coli</i> EXPRESS BL21(DE3) under various conditions after 45 h. ....	59
Figure 17 eGFP fluorescence normalized to cell density of ADMARS 02 at various points in time after induction. ....	66
Figure 18 eGFP fluorescence normalized to cell density of ADMARS 03 at various points in time after induction. ....	67
Figure 19 eGFP fluorescence normalized to cell density of ADMARS 04 at various points in time after induction. ....	67
Figure 20 eGFP fluorescence normalized to cell density of ADMARS 12 at various points in time after induction. ....	68
Figure 21 eGFP fluorescence normalized to cell density of ADMARS 23 at various points in time after induction. ....	69

Figure 22 eGFP fluorescence normalized to cell density of ADMARS 25 at various points in time after induction. ....	69
Table 1 Amino acids employed in this thesis.....	15
Table 2 Components and final concentrations in the 5x M9-salt stock. ....	20
Table 3 Components added to prepare M9-medium and M9-plates.....	22
Table 4 Bacterial strains and the corresponding genotypes. ....	24
Table 5 ID, abbreviations and properties of synthetic DNA cloning fragments ordered at Twist Bioscience. ....	24
Table 6 Primers used for plasmid verification after ligation of backbone and inserts.....	30
Table 7 Ingredients of M9-Kan plates for chloramphenicol growth assay. ....	32
Table 8 Properties of the design groups A and B.....	37
Table 9 Sample layout for screening of ADMARS variants in 96-well plate cultures.....	43
Table 10 Sample layout for screening of SDLSL mutant in 96-well plate cultures. ....	50
Table 11 Sample layout for expression analysis of ADMARS variants in 24-well plate cultures. ....	51
Table 12 Sample layout of the preparation for the chloramphenicol growth assay.....	56
Equation 1 Calculation of transformation efficiency. ....	29



## References

- 
- <sup>1</sup> Wiltschi, B. (2016), Protein Building Blocks and the Expansion of the Genetic Code, *Synthetic Biology*, DOI 10.1007/978-3-319-22708-5\_4
- <sup>2</sup> Burbaum, J. J., Schimmel, P. (1991), Structural Relationships and the Classification of Aminoacyl-tRNA Synthetases, *J Biol Chem.* 266(16), 16965-16968
- <sup>3</sup> Ibba, M., Söll, D. (2000), Aminoacyl-tRNA Synthesis, *Annu Rev Biochem.* 69, 617-650
- <sup>4</sup> Moore, P. B., Steitz, T. A. (2011), The roles of RNA in the synthesis of protein, *Cold Spring Harb Perspect Biol.* 3(11)
- <sup>5</sup> Ambrogelly, A., Palioura, S., Söll, D. (2007), Natural expansion of the genetic code, *Nat Chem Biol.* 3(1), 29-35
- <sup>6</sup> Krzycki, J. A. (2005), The direct genetic encoding of pyrrolysine, *Curr Opin Microbiol.* 8, 706-712
- <sup>7</sup> Wang, A., Nairn, N. W., Marelli, M. *et al.* (2012). Protein Engineering with Non-Natural Amino Acids, Protein Engineering, Prof. Pravin Kaumaya (Ed.), ISBN: 978-953-51-0037-9, InTech, Available from: <http://www.intechopen.com/books/protein-engineering/protein-engineering-with-nonnatural-amino-acids>
- <sup>8</sup> Smolskaya, S., Andreev, Y. A. (2019), Site-Specific Incorporation of Unnatural Amino Acids into *Escherichia coli* Recombinant Protein: Methodology Development and Recent Achievement, *Biomolecules* 9, 255
- <sup>9</sup> Wan, W., Huang, Y., Wang, Z., *et al.* (2010), A facile system for genetic incorporation of two different noncanonical amino acids into one protein in *Escherichia coli*, *Angew Chem Int Ed Engl.* 49(18), 3211-3214
- <sup>10</sup> O'Donoghue, P., Prat, L., Heinemann, I. U., *et al.* (2012), Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNA<sup>Pyl</sup> for genetic code expansion, *FEBS Lett.* 586, 3931-3937
- <sup>11</sup> Koutmou, K. S., McDonald, M. E., Brunelle, J. L., *et al.* (2014), RF3: GTP promotes rapid dissociation of the class 1 termination factor, *RNA* 20(5), 609-620
- <sup>12</sup> O'Donoghue, P., Ling, J., Wang, Y., *et al.* (2013), Upgrading protein synthesis for synthetic biology, *Nat Chem Biol.* 9(10), 594-598
- <sup>13</sup> Normanly, J., Masson, J. M. (1986), Construction of two *Escherichia coli* amber suppressor genes: tRNA<sub>CUA</sub><sup>Phe</sup> and tRNA<sub>CUA</sub><sup>Cys</sup>, *Proc Natl Acad Sci U S A.* 83(17), 6548-6552
- <sup>14</sup> Longstaff, D. G., Blight, S. K., Zhang, L., *et al.* (2007), In vivo contextual requirements for UAG translation as pyrrolysine, *Mol Microbiol.* 63(1), 229-241
- <sup>15</sup> Forster, A. C. (2009), Low modularity of aminoacyl-tRNA substrates in polymerization by the ribosome, *Nucleic Acids Res.* 37(11), 3747-3755

- 
- <sup>16</sup> Giege, R., Sissler, M., Florentz, C. (1998), Universal rules and idiosyncratic features in tRNA identity, *Nucleic Acids Res.* 26(22), 5017-5035
- <sup>17</sup> Wan, W., Tharp, J. M., Liu, W. R. (2014), Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool, *Biochim Biophys Acta.* 1844(6), 1059-1070
- <sup>18</sup> Fladischer, P., Weingartner, A., Blamauer, J. *et al.* (2019), A Semi-Rationally Engineered Bacterial Pyrrolysyl-tRNA Synthetase Genetically Encodes Phenyl Azide Chemistry, *Biotechnol J.* 14(3)
- <sup>19</sup> Melançon, C. E., Schultz, P. G. (2009), One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases, *Bioorg Med Chem Lett.* 19(14), 3845-3847
- <sup>20</sup> Liu, C. C., Schultz, P. G. (2010), Adding New Chemistries to the Genetic Code, *Annu Rev Biochem.* 79, 413-444
- <sup>21</sup> Santoro, S. W., Wang, L., Herberich, B. *et al.* (2002), An efficient system for the evolution of aminoacyl-tRNA synthetase specificity, *Nat Biotechnol.* 20(10), 1044-1048
- <sup>22</sup> Wang, L., Xie, J., Schultz, P. G. (2006), Expanding the Genetic Code, *Annu Rev Biophys Biomol Struct.* 35, 225-249
- <sup>23</sup> Kobayashi, T., Yanagisawa, T., Sakamoto, K., *et al.* (2009), Recognition of Non- $\alpha$ -amino Substrates by Pyrrolysyl-tRNA Synthetase, *J Mol Biol.* 385(5), 1352-1360
- <sup>24</sup> Bao, J., Lorenzo, A. D., Lin, K. *et al.* (2019), Mouse models of overexpression reveal distinct oncogenic roles for different type I protein arginine methyltransferases, *Cancer Res.* 79(1), 21-32
- <sup>25</sup> Lunde, B. M., Moore, C., Varani, G. (2007), RNA-binding proteins: modular design for efficient function, *Nat Rev Mol Cell Biol.* 8(6), 479-490
- <sup>26</sup> Pertschy, B., BioTechMed-Graz Flagship proposal: Dynamics of subcellular partitioning through protein modification
- <sup>27</sup> OpenWetWare (2020), *E. coli* genotypes. Retrieved March 29, 2021, from [www.openwetware.org](http://www.openwetware.org)
- <sup>28</sup> Lucigen Corporation, *E. coli*<sup>®</sup> EXPRESS BL21(DE3) Competent Cells. Retrieved March 29, 2021, from [www.lucigen.com](http://www.lucigen.com)
- <sup>29</sup> Ness, F., Ferreira P., Cox, B. S. *et al.* (2002), Guanidine hydrochloride inhibits the generation of prion “seeds” but not prion protein aggregation in yeast, *Mol. Cell. Biol.* 22(15), 5593-5605
- <sup>30</sup> Technical Bulletin Wizard<sup>®</sup> Plus SV Minipreps DNA Purification System, Instructions for Use of Products A1270, A1330, A1340, A1460 and A1470, Promega Corporation, Revised 12/10

- 
- <sup>31</sup> Fuertes, M. A., Pérez, J. M., Alonso, C. (2004), Small amounts of urea and guanidine hydrochloride can be detected by a far-UV spectrophotometric method in dialysed protein solutions, *J Biochem Biophys Methods* 59, 209-216
- <sup>32</sup> Ishido, T., Ishikawa M., Hirano K. (2010), Analysis of supercoiled DNA by agarose gel electrophoresis using low-conducting sodium threonine medium, *Anal Biochem.* 400(1), 148-150
- <sup>33</sup> Braun, M. (2021), Design of a novel PylRS mutant to site-specifically incorporate the ncAA ADMA into proteins (unpublished)
- <sup>34</sup> Neidhardt, F. C., Bloch, P. L., Smith, D. (1974), Culture Medium for Enterobacteria, *J Bacteriol.* 119(3), 736-747
- <sup>35</sup> Grant, C. L., Pramer, D. (1962), Minor Element Composition Of Yeast Extract, *J Bacteriol.* 84(4), 869-870
- <sup>36</sup> Kobayashi, T., Yanagisawa, T., Sakamoto, K., *et al.* (2009), Recognition of Non- $\alpha$ -amino Substrates by Pyrrolysyl-tRNA Synthetase, *J Mol Biol.* 385(5), 1352-1360
- <sup>37</sup> Galindo Casas, M., Stargardt, P., Mairhofer, J., *et al.* (2020), Decoupling Protein Production from Cell Growth Enhances the Site-Specific Incorporation of Noncanonical Amino Acids in *E. coli*, *ACS Synth Biol.* 9(11), 3052-3066
- <sup>38</sup> Freist, W., Sternbach H., Pardowitz I., *et al.* (1998), Accuracy of Protein Biosynthesis: Quasi-species Nature of Proteins and Possibility of Error Catastrophes, *J Theor Biol.* 193(1), 19-38
- <sup>39</sup> Sprinzl, M., Kucharzewski, M., Hobbs, J. B., *et al.* (1977), Specificity of Elongation Factor Tu from *Escherichia coli* with Respect to Attachment to the Amino Acid to the 2' or 3'-Hydroxyl Group of the Terminal Adenosine of tRNA, *Eur J Biochem.* 78(1), 55-61
- <sup>40</sup> Paliy, O., Gunasekera, G. S. (2007), Growth of *E. coli* BL21 in minimal media with different gluconeogenic carbon sources and salt contents, *Appl Microbiol Biotechnol.* 73(5), 1169-1172
- <sup>41</sup> Studier, W. F., Moffatt, B. A. (1986), Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes, *J Mol Biol.* 189, 113-130
- <sup>42</sup> Shine, J., Dalgarno, L. (1974), The 3'-Terminal Sequence of *Escherichia coli* 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites, *Proc Natl Acad Sci USA.* 71(4), 1342-1346