Maria Koshanskaya, BSc

Combinatorial Cloning of Multipart Expression Constructs using Modular Cloning

MASTER'S THESIS

to achieve the university degree of

Master of Science (MSc)

Master's degree programme: Biochemistry and Molecular Biomedical Sciences

submitted to

Graz University of Technology

Supervisor

Univ.-Prof. Dipl.-Ing. Dr. techn. Helmut Schwab
Institut of Molecular Biotechnology,
TU Graz, Austria

DIpl.-Ing. Dr. techn. Birgit Wiltschi,
Junior Group Synthetic Biology
Austrian Centre of Industrial Biotechnology

Graz, August 2015

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, 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 TUGRAZonline is identical to the present master‘s thesis dissertation.

Date	Signature

Table of Contents

Acknowledgments

I would like to thank Univ.-Prof. Dipl.-Ing. Dr. techn. Helmut Schwab, the main examiner of this work, for giving me the opportunity to conduct my research within the Junior Group Synthetic Biology.

I would also like to express my gratitude to my supervisor, Dipl.-Ing. Dr. techn. Birgit Wiltschi, for her useful comments, remarks and engagement through the learning process of this master thesis. I also want to thank Dr. Wiltschi for her rewarding advices concerning scientific writing, which helped me to prepare my document in a professional manner.

Special and kind thanks go to Corinna Odar, Dipl.-Ing. for being such a great supervisor and for introducing me to this topic. Your support at every step of my way and the never-ending willingness to help always encouraged and inspired me.

I also want to thank the members of the Junior Group of Synthetic Biology for their friendliness and support in discussing lab issues. It has been a pleasure working with everyone.

Furthermore, I would like to thank the Austrian Centre of Industrial Biotechnology and FFG for funding this project.

Last but not the least, I would like to thank my loved ones, who have supported and encouraged me throughout entire process. I will be grateful forever for your love.

Graz, August 2015

Maria Koshanskaya, BSc

1. Abstract

Fluorinated chemicals are becoming more and more important for the pharmaceutical, agricultural and material industries. Currently, most fluorinations are performed chemically under harsh conditions leading to non-specific byproducts. The enzyme fluorinase is currently the only known enzyme which can introduce inorganic fluoride into organic compounds. Fluorinase was first isolated from *Streptomyces cattleya* and catalyzes the formation of 5-fluoro-5´-deoxyadenosine (FDA). This fluorinated compound represents the first metabolite of a pathway leading to a non-canonical amino acid 4-fluorothreonine (4-FT) and fluoroacetate. In this study, we aimed to reproduce the first step of the 4FT pathway in *E.coli*. Our aim was to improve the FDA production by enhancing the functional expression of the enzyme fluorinase in *E.coli.* On one hand, we codon optimized and harmonized the *flA* gene sequence from *S.cattleya* with the intent to increase its soluble expression in *E. coli*. On the other hand, we co-expressed the enzyme rat liver SAM synthetase (RLSS) in order to retrieve the L-methionine from the reaction equilibrium and hence push the production of FDA. According to our results, the applied codon optimization and harmonization strategy did neither significantly improve the expression of FlA nor its solubility. The co-expression of the enzymes RLSS and fluorinase led to an increased production of the FDA compared to the single expression of the FlA. The cloning of the parts for the reaction cascade was performed by modular cloning (MoClo). MoClo is a highly efficient and directional assembly technique which allows assembly of multiple DNA fragments in a one-pot reaction. The relevant destination vectors and the parts for the co-expression experiments were successfully designed and partially produced. Due to the encountered issues with the MoClo assembly, we produced the cascade expression constructs via Gibson cloning.

2. Kurzfassung

Fluorierte Chemikalien gewinnen heutzutage immer mehr an Bedeutung, vor allem in Bereichen der Arzneimittel-, Agrar- und Materialwissenschaften. Zur Zeit wird der Großteil der Fluorverbindungen mittels chemischer Synthese unter harschen Reaktionsbedingungen hergestellt, wobei viele nicht spezifische Nebenprodukte entstehen. Das Enzym Fluorinase ist momentan das einzige, welches anorganisches Fluor in organische Verbindungen einbauen kann. Die erste identifizierte Fluorinase wurde aus dem Organismus *Streptomyces cattleya* isoliert und katalysiert die Bildung von 5´-Fluoro-5´-deoxyadenosin (FDA). Diese fluorierte Organoverbindung ist der erste Metabolit des Stoffwechselweges, welcher zu 4-Fluorothreonin (4-FT), einer nicht kanonischen Aminosäure, und Fluoroacetat führt. Ziel dieser Studie war es den ersten Schritt des 4-FT Stoffwechselweges in *E.coli* zu reproduzieren. Unsere Absicht war es die FDA Produktion zu steigern, indem wir die funktionelle Expression der Fluorinase in *E. coli* verbesserten. Einerseits haben wir die Sequenz des *flA* Gens aus *S.cattleya* Codon-optimiert und harmonisiert, um die lösliche Expression der Fluorinase zu verbessern. Andererseits haben wir das Enzym rat liver SAM synthetase (RLSS) mit der Fluorinase coexprimiert, um das entstehende L-Methionin aus dem Reaktionsgleichgewicht zu entfernen und somit die Produktion von FDA zu steigern. Laut unseren Ergebnissen führte die angewandte Strategie der Codon-Optimierung und Harmonisierung weder zu Erhöhung der Expression noch der Löslichkeit der Fluorinase. Die Coexpression der Enzyme RLSS und Fluorinase hat zu einer erhöhten Ausbeute an FDA geführt, verglichen zur Expression der Fluorinase allein. Die Klonierungsschritte der einzelnen Bestandteile der Reaktionskaskade wurden mittels modular cloning (MoClo) durchgeführt. MoClo ist eine hoch effiziente und gerichtete Assemblierungsmethode von multiplen DNA-Fragementen, die als eine Eintopfreaktion durchgeführt wird. Die DNA Komponenten, welche für die Coexpressionsexperimente benötigt wurden, und die Zielvektoren wurden erfolgreich konstruiert und teilweise produziert. Aufgrund der Tatsache, dass einige DNA Komponenten nicht in der für die MoClo-Methode erforderlichen Weise kloniert werden konnten, wurde die Assemlierung der Multigen-Expressionskonstrukte mit Hilfe von Gibson-Klonierung durchgeführt.

3. Introduction

Despite the fact that fluorine is one of the most abundant elements in the crust of the Earth fluorinated organic compounds are rare in nature (Dong et al. 2004; Murphy et al. 2003). However, it was shown that fluorine is a remarkable element since its incorporation into biological molecules allows modulating their activity and studying their mechanism of action without gross alterations of the structure (Marsh 2014). Also proteins containing fluorinated amino acids (FAA) are of scientific and biotechnological interest (Biava & Budisa 2014). To date only one naturally occurring FAA, 4-fluorothreonine (4FT), is known. It is produced biosynthetically in *Streptomyces cattleya*. Chemical approaches for the synthesis of this non-canonical amino acid in high amounts and for a reasonable price are still not available (Deng et al. 2014).The first step of the 4FT pathway is performed by the enzyme fluorinase which catalyzes the bond formation between carbon and inorganic fluoride. In *E. coli*, the expression of the fluorinase led to mostly insoluble protein (M. Koshanskaya, C. Odar, unpublished observations). We hypothesized that this might impede the successful reconstitution of the 4FT pathway in this host organism, which represents an easy, cheap and quick expression and production host for recombinant proteins (Rosano & Ceccarelli 2014). Hence, in this study we focused on the first step of the 4FT pathway and our aim was to optimize the production of FDA through heterologous expression of FlA in *E.coli*.

3.1. Biosynthesis of organofluorines compounds in *S. cattleya*

Organic compounds containing halogens comprise a diverse group of chemicals that display a wide range of biological properties, which includes anticancer and antibiotic activity. Over the centuries they played a key role in natural sciences and to date, more than 4000 naturally occurring halogenated compounds have been discovered. Only 30 of these compounds contain fluorine (Gribble 2004) and only five of these compounds were securely identified as secondary metabolites produced by some classes of organisms (Hagan & Deng 2015). The fact that organofluorines are extremely rare in nature is due to their unique physicochemical properties such as the highest electronegativity, size of this element and the strength of the carbon-fluorine bond (C-F). A comprehensive overview of the organofluorine chemistry is given by O'Hagan (O'Hagan 2008).These properties also render fluorinated compounds as well as fluorinated amino acids of increasing interest for the biosciences, especially in the pharmaceutical and medical sector (Murphy et al. 2003).For instance, more than 20% of the currently used drugs contain at least one fluorine atom (Purser et al. 2008). Nonetheless, the approaches for

fluorination are limited. Fluorination is mostly performed chemically under harsh conditions leading to non-specific byproducts (Murphy et al. 2003). As well, naturally occurring fluorination approaches are confined. With the exception of CF_4 and chlorofluorocarbons (CFCs), which are abiotically formed during geothermal processes, there are only few species of tropical and subtropical plants and some microorganisms which perform enzymatic fluorination in nature (Murphy et al. 2003).

The bacterium *Streptomyces cattleya* was first recognized to produce fluorine-containing organo-compounds. When incubated in a medium supplemented with fluoride ions this microorganism secretes in particular fluoroacetate and 4-fluorothreonine (4FT) (Deng et al. 2006). Both these secondary metabolites are produced through a five-step reaction cascade which is represented in Figure 1 and was comprehensively described by O'Hagan and co-workers (Deng et al. 2008). In the first step, the fluorinating enzyme 5'-deoxy-5'-fluoroadenosine synthase, also known as the fluorinase, catalyzes the C-F bond formation between S-adenosyl-L-methionine (SAM) and fluoride to generate 5'-deoxy-5'-fluoroadenosine (FDA). In the next step, the adenine base of FDA is displaced upon phosphorylation by the enzyme FDA phosphorylase. This generates 5-deoxy-5-fluororibose phosphate (FDRP) which then undergoes an isomerization step to 5-deoxy-5-fluororibulosephosphate (FDRibulP). The next enzyme of the cascade, an aldolase, cleaves FDRibulP generating fluoroacetaldehyde (F-acetaldehyde), the last common intermediate of the pathway. F-acetaldehyde is transformed either to fluoroacetone by a dehydrogenase or to the non-canonical amino acid 4FT by a PLP dependent transaldolase. 4FT exhibits antibiotic properties and fluoroacetate is a well-known toxin acting as an inhibitor of aconitase, one of the key enzymes of the citric acid cycle (O'Hagan 2006). Fluoroacetate also represents a branching point for other fluorinated compounds such as fluorinated steroids or eicosanoids (Walker et al. 2013).

The 4FT pathway of *S.cattleya* was successfully reconstituted *in vitro* by Deng et al. by overexpressing the pathway genes either from *S. cattleya* or a surrogate source (Deng et al. 2008).

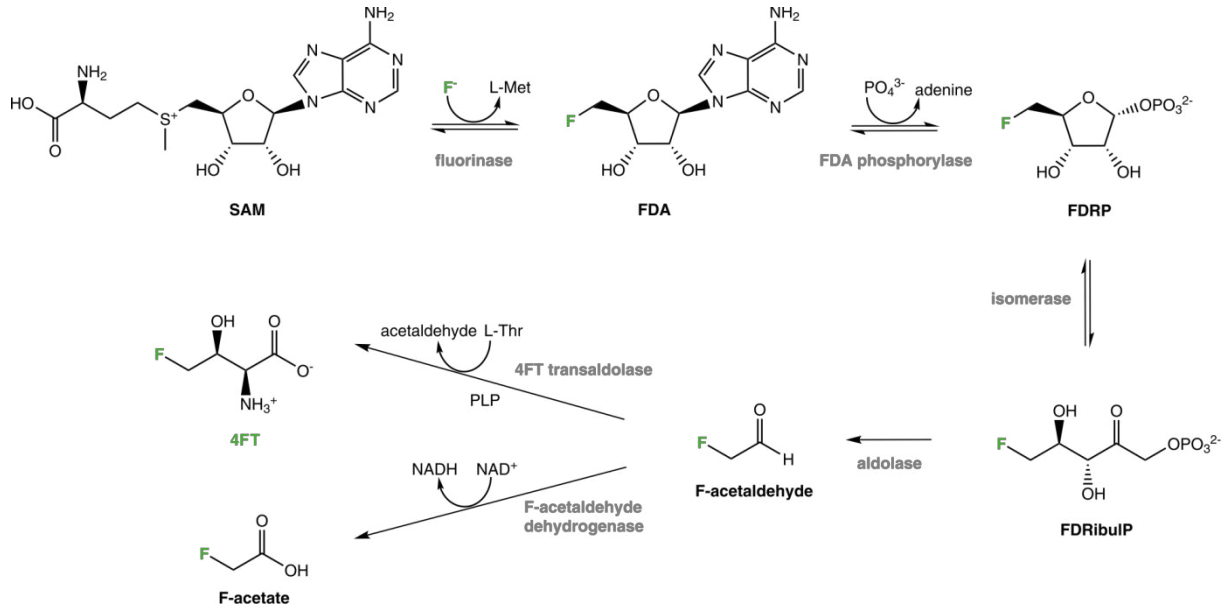

Figure 1: Biosynthetic pathway for 4-fluorothreonine (4FT) and fluoroacetate (F-acetate) in *S.cattleya*.
SAM, S-adenosyl-L-methionine; FDA, 5'-deoxy-5'-fluoroadenosine; FDRP, 5-deoxy-5-fluororibose phosphate; FDRibulP, 5-deoxy-5-fluororibulosephosphate; F-acetaldehyde, fluoroacetaldehyde; F-acetate, fluoroacetate. Reproduced from (Odar et al. 2015).

3.2. FDA production using an enzyme cascade

We intended to reconstruct the 4FT pathway *in vivo* using *E.coli* as the expression host.

In this study we focused on the first step of the pathway which produces FDA by the action of the fluorinase (EC 2.5.1.63; see Figure 2). The expression level of the fluorinase in its natural host *S.cattleya* is high (Schaffrath et al. 2003) but according to Deng et al., the heterologous expression in *E.coli* might be problematic (Deng et al. 2008) e.g. due to a high GC content of *Streptomyces*. The codon usage of the heterologous host *E.coli* might be incompatible with the one of the parent organism (Wright & Bibb 1992). As reported by Dong et al., substantial amounts of FIA are expressed in insoluble form (Dong et al. 2004). In general, to overcome these issues the codon optimization and harmonization are often the methods of choice (Gustafsson et al. 2004; Angov et al. 2008). The fluorinase catalyzes the formation of FDA from the substrates SAM and fluoride. As it was shown previously, the presence of SAM might positively influence the stability and activity of FIA (Dong et al. 2004). We hypothesize that by increasing the intracellular SAM concentrations we might enhance the activity of the fluorinase and hence increase the production of FDA.

SAM is a central metabolite and the main cellular methyl donor which is involved in numerous cellular metabolic reactions (Lieber 2002). *E. coli* is incapable to take up exogenously supplied

SAM (Manavathu & Thomas 1982), which means that in order to ensure sufficient supplementation of SAM for the reaction by the fluorinase (K_M for SAM is 74 µM) SAM needs to be recycled. In *E. coli*, the recycling is performed by the enzyme MetK, an endogenous SAM synthase. MetK uses L-methionine and ATP for the SAM synthesis. However, high expression levels of MetK would deplete the *E.coli* cells of L-methionine and thus disturb the cellular metabolism leading to a growth disadvantage. For this reason MetK expression is strictly controlled in *E. coli* and its activity is rather low. Additionally, MetK is susceptible to feedback inhibition (Wei & Newman 2002).

The above mentioned reasons render the endogenous MetK not efficient as a recycling enzyme for the FlA-catalyzed FDA production. Studies showed that the overexpression of the SAM synthetase from *Rattus Norvegicus* (RLSS) highly increased the intracellular SAM levels (Alvarez et al. 1994).

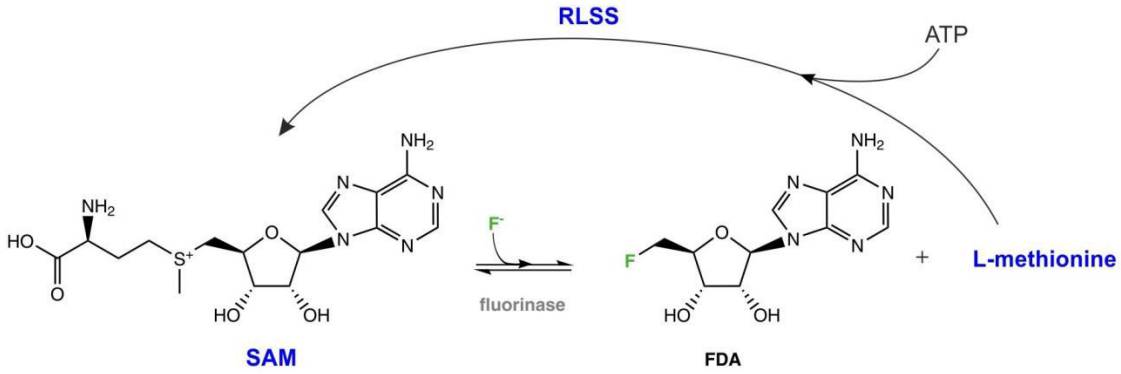

Figure 2: Enzymatic cascade for production of 5'-deoxy-5'-fluoroadenosine (FDA) and recycling of S-adenosyl-L-methionine (SAM).
RLSS, rat liver SAM synthase. Adapted from Corinna Odar (ACIB GmbH, Graz).

RLSS is highly homologous to MetK of *E.coli* but it is not susceptible to feedback inhibition. This means that RLSS is largely unaffected by high SAM concentrations (Posnick & Samson 1999). Similar to MetK, RLSS uses L-methionine together with adenosine triphosphate (ATP) to generate SAM. As represented in Figure 2, the reaction of the fluorinase leads to L-methionine as the second product. We hypothesize that by recycling L-methionine we would not only elevate the SAM levels but also remove one product from the reaction equilibrium which might push the reaction towards the desired product FDA.

The second substrate of the fluorinase is fluoride. The fluorinase exhibits a rather low affinity for it, however, substantially differing K_m values have been published (K_m 2 mM; (Wang et al. 2014), 8.56 mM (Schaffrath et al. 2003)). Thus, a sufficient intracellular concentration of fluoride is needed for an efficient *in vivo* conversion to FDA. Fluoride is cytotoxic to *E. coli* what might be a

reason why initial trials to biosynthesize FDA *in vivo* in *E. coli* were unsuccessful (Corinna Odar, unpublished results). To circumvent this issue we decided to perform the experiments of the FDA production using lyophilized cells. Lyophilized cells are not living so they are not affected by high fluoride concentration and exhibit preserved enzyme activity. Additionally, the cell membrane is made porous during the lyophilization process. Thus, the substrates SAM and L-methionine can pass through the permealized cell membrane (Gurney 1984).

3.3. Modular cloning – a short introduction

Modular cloning (MoClo) is a recently developed cloning approach, which allows the assembly of multiple genetic elements in combinatorial manner. It was first published by Ernst Weber and Carola Engler in 2011 and represents a hierarchical and highly directed DNA assembly method, that can be applied in high throughput. It combines the use of type IIS restriction enzymes with ligation (Weber et al. 2011). The advantage of MoClo is its ability to assemble multiple DNA parts into complex constructs in one reaction. This drastically increases the speed and decreases the number of steps needed for the assembly of the multi-part constructs. Weber and Engler assembled a 33 kb construct encoding 11 transcriptional units made from 44 individual basic modules using only three successive assembly reactions (Weber et al. 2011).

The directed and hierarchical character of MoClo relies on type IIS restriction enzymes. These restriction enzymes comprise two distinct domains, one DNA binding domain, and another domain for DNA cleavage. Consequently, they have separate recognition and cleavage sites (Figure 3A). Their recognition sites are non-palindromic and the enzymes cut the DNA at a precise distance from them (Szybalski et al. 1991). The cut occurs only on one side of the recognition site, depending on its orientation on the DNA, either up or downstream of it (as exemplified for BpiI in Figure 3A). Figure 3B shows the recognition and cleavage sites of two more type IIS restriction enzymes used in this study, BsaI and BsmBI.

To generate fragments suitable for MoClo the recognition sites are placed at the 5`and 3`ends of each DNA fragment in opposite direction (Engler et al. 2008). Since the recognition site is separated from the cleavage site, the recognition sites are removed during the cleavage process. Cleavage with BpiI, BsaI and BsmBI generates cohesive ends. The single stranded overhangs of the cohesive ends are referred to as fusion sites. This is because their sequence can be randomly chosen as it is neither essential for recognition nor for cleavage. The fusion sites are designed such that they are compatible on two adjacent fragments (Figure 3C). This strategy ensures the seamless and highly directed assembly of multiple fragments.

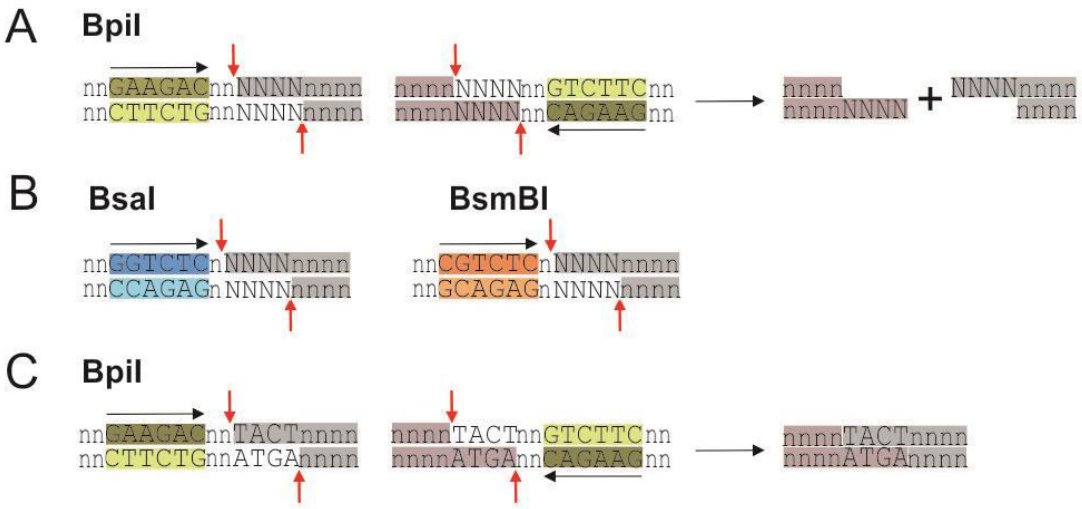

Figure 3: Recognition and cleavage sites of the type IIS restriction enzymes. A: general overview. Restriction sites, indicated by green color, and cleavage sites, indicated by red arrows, represent separated domains. B: Restriction enzymes cut outside of their recognition sites (blue and orange colors for BsaI and BsmBI, respectively; light and dark hues indicate the asymmetry of the recognition site) as indicated by the arrows. This generates custom fusion sites for seamless ligation. C: An example of the one pot restriction/ligation of two fragments featuring the same fusion site TACT using BpiI.

Basically, MoClo comprises three hierarchical levels which are displayed in Figure 4. At the lowest level, level°0, basic biological parts such as promoters, coding sequences, RBSs, terminators etc. are cloned individually into the level°0 destination vector to yield part plasmids. These part plasmids constitute the level°0 library of basic DNA modules from which all subsequent levels are assembled. Selected parts encoded on part plasmids are assembled into the level°1 destination vector to give device plasmids, which, for instance, contain transcriptional units. These device plasmids can be directly used for expression of single transcriptional units. Otherwise, device plasmids can be introduced into a further cloning step resulting in a complex multi-part level°2 assembly, which might comprise an entire biosynthesis pathway. In conclusion, the constructs that are generated in one level become the donor vectors for the next level.

While classical restriction-ligation cloning of multiple inserts involves several individual steps of restriction and ligation, MoClo facilitates even complex assemblies by a one-pot reaction. Basically, all selected donor vectors are mixed in a vessel and cleaved simultaneously with the appropriate type IIS restriction enzyme. The mix also contains a ligase to ligate the fragments that anneal via their fusion sites. Since the recognition site is lost during cleavage, the resulting fragments cannot be re-cleaved by the enzyme. Thus, restriction and ligation occur in parallel and tedious intermittent purification steps of the generated fragments are obsolete.

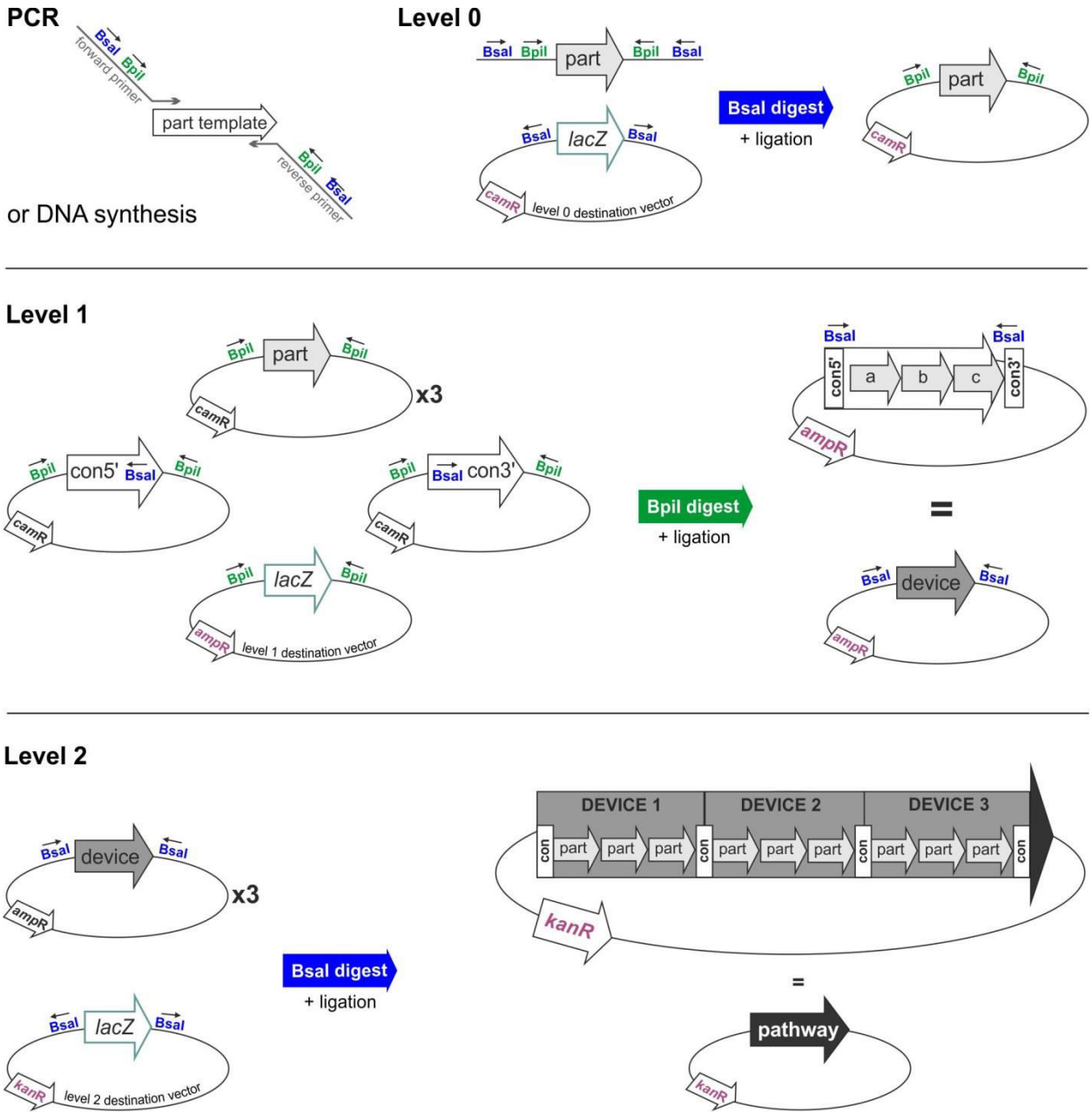

Figure 4: Schematic representation of the three MoClo levels.
Parts, e.g. promoter, coding sequences, RBS, terminator, connectors, etc. are produced by PCR amplification or by DNA synthesis and contain the overhangs with BpiI and BsaI recognition and cleavage sites. Parts are assembled into part plasmids upon BsaI digest and are selected by chloramphenicol resistance (camR). Upon BpiI restriction, part and connector plasmids are assembled to a device plasmid, which carries an ampicillin resistance (ampR). Level 2 is the last assembly reaction with BsaI leading to a multi-gene plasmid which is selected by kanamycin (kanR) resistance. Figure adapted from Corinna Odar (ACIB GmbH, Graz).

It is important that the MoClo destination vectors carry individual antibiotic markers to ensure their selection on the different levels. Each level requires a different type IIS restriction enzyme. To make the selection even more stringent, all destination vectors carry the *lacZ* gene for blue/white selection of the inserts. During the assembly reactions, the *lacZ* gene is replaced by

the basic biological part or one or more devices and the positive clones can be easily selected by their white color, as represented by Figure 5.

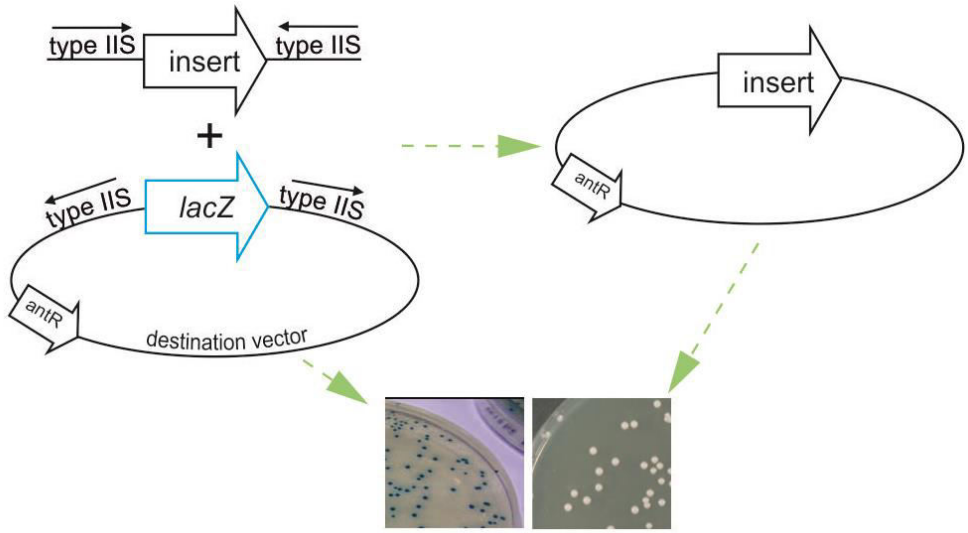

Figure 5: Schematic representation of the blue/white selection of vectors carrying an insert. The insert can be a basic biological part or one or more devices and depends on the level. The insert and the *lacZ* transcriptional unit are both flanked by the same type IIS restriction enzyme recognition sites which depend on the level. During the digest of both insert and destination vector, followed by ligation, the *lacZ* gene is replaced by the insert giving rise to white clones. antR: antibiotic resistance marker.

As shown in Figure 4, each device is flanked by a connector at each terminus. A general structure of a connector is shown in Figure 6. After restriction with BsmBI each connector can ligate via the level°0 fusion sites with the level°0 destination vector resulting in connector plasmids. They are equal to the part plasmids containing basic biological parts. Upon digest with BpiI in level°1 reaction each connector can ligate with both the destination vector and a part and thus, directs the assembly of the parts in a device plasmid. Additionally, connectors carry internal fusion sites for the level°2 destination vector. This facilitates the assembly of devices at level°2 and ensures a directed and modular assembly of the multi-gene constructs.

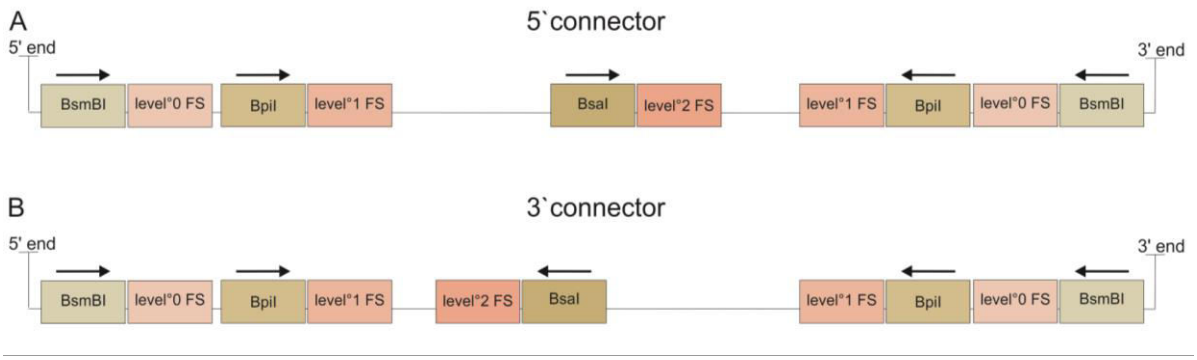

Figure 6: Scheme of the connector's structure. A and B represent the connectors placed at the 5´and 3´end of the corresponding device, respectively. Restriction with the type IIS endonucleases BsmBI, BpiI and BsaI produce fusion sites (FS) for the three levels of assembly, 0, °1 and °2.

3.4. Thesis objectives

The aim of this study was the heterologous production in *E. coli* of FDA, the first metabolite of the 4FT pathway. To achieve this, we intended to improve the functional expression of the fluorinase from *Streptomyces cattleya* in *E.coli* by codon optimization and harmonization of the *flA* gene sequence. Additionally, by overexpressing the RLSS from *Rattus Norvegicus* we aimed to retrieve the L-methionine from the reaction equilibrium. On the one hand, the idea was to recycle the L-methionine to SAM and hence, increase the intracellular SAM concentration. This, in turn, should stabilize the fluorinase and we expected a higher activity of the enzyme. On the other hand, by retrieving the byproduct L-methionine from the reaction equilibrium we wanted to push the reaction to the product side and expected an increased yield of FDA. To overcome the issue of the cytotoxicity of fluoride, one of the substrates of the fluorinase, we performed the assays using lyophilized *E. coli* cells and analyzed the product formation by HPLC-UV. For the assembly of the desired expression constructs we planned to implement MoClo, which we adopted to our own needs. As this assembly strategy was not successful for the assembly of the enzyme cascade we switched to Gibson cloning, another assembly strategy established in the group.

4. Materials

4.1.1. Bacterial strains and plasmids

Expression was performed in the *E. coli* strain BL21(DE3)Gold (Invitrogen, Carlsbad, CA) with the genotype B F- *dcm*+ Hte *ompT hsdS(rB- mB-) gal λ* (DE3) *end*A Tet^r or in the *E. coli* strain BWEC46, which is a threonine auxotrophic derivative of BL21(DE3)Gold carrying the genetic deletion Δ*thrC*::0.

For the plasmid amplification the *E. coli* strain TOP10F' was used. The genotype was F´*{lacI*^{q}, Tn*10(*Tet^r*)} mcr*A *Δ(mrr-hsd*RMS-*mcr*BC*)* Φ80*lac*ZΔM15 *Δlac*X74 *rec*A1 *ara*D139 *Δ(ara leu)7697 gal*U *gal*K *rps*L *(*Str^R*) end*A1 *nup*G and it was purchased from Invitrogen (Carlsbad, CA).

The genes and primers used in this study are listed in the supporting material (9.1 and 9.5). Genes were ordered as synthetic gBlocks® from Integrated DNA Technologies (Coralville, IA) and subcloned for further work into the pJET1.2/blunt cloning vector from Thermo Scientific (Waltham, MA).

Primers (Table S1) were designed with SnapGene (http://www.snapgene.com) and ordered from IDT (Integrated DNA Technologies, Coraville, IA).

The expression vector pQE-80L was purchased from Qiagen (Venlo, Netherlands). The plasmids pQE80L-FIA and p15aARA for the co-expression experiments were assembled and kindly provided by Corinna Odar (ACIB GmbH, Graz). The first destination vectors pMC0.1 and pMC0.3 were assembled during an earlier laboratory project (Koshanskaya 2014). The vector maps of all plasmids used for this study are summarized in section 9.2.

Plasmid isolation was performed using the GeneJET Plasmid Miniprep Kit from Thermo Scientific (Waltham, MA). For the extraction and purification of DNA and PCR fragments, respectively, the Wizard® SV Gel and PCR Clean-Up System from Promega (Madison, WI) were used. The manufacture's protocols were followed without any deviations, except where indicated.

Sequencing was performed at Microsynth GmbH (Balgach, Switzerland).

4.1.2. Instruments, chemicals and enzymes

All instruments, devices, chemicals and enzymes used for this study are listed in the supporting materials (Table S5, Table S6 and Table S7).

4.2. Methods

4.2.1. Codon optimization

The genes of FlA from *S. cattleya* and RLSS from *R. norvegicus* were codon optimized for the expression in *E. coli*. The rare codons used in these host organisms were detected using the program Graphical Codon Usage Analyser (http://gcua.schoedl.de/) and the codon usage table of *S. cattleya* was taken from the Codon Usage Database Kazusa[1]. The optimization was performed using the program Gene Designer from DNA2.0 Inc according to the codon usage of *E. coli* class I, which is based on a collection of genes involved in most metabolic processes in *E.coli* (Villalobos et al. 2006). The following restriction sites were avoided: BsaI, BpiI, PacI, SbfI, NheI, NotI, SpeI, EcoRI, BglII, HindIII and SbfI. The methylation of these restriction sites was avoided: BsaI, ClaI, XbaI. Additionally, rare codons used by the host organisms were manually replaced by rare codons used by *E.coli*. The codons which were optimized manually are summarized in Table S10 and the optimized sequences are listed in section 9.5.

4.2.2. QuikChange PCR

In order to exchange the incorrectly optimized codons of the enzyme FlA one forward and one reverse primer were designed such that the primer sequences were not completely overlapping. Only the sequence containing the desired mutations was overlapping, lowering the possibility of the primer pair to form self-dimers (Zheng et al. 2004).

QuikChange PCR was carried out with Phusion® High-Fidelity DNA polymerase (Thermo Scientific; Waltham, MA) and addition of 3% DMSO. The vector carrying the not codon optimized FlA was used as the template. The PCR amplification reaction was initiated by pre-heating the reaction mixture to 95°C for 2 min; 25 cycles of 95°C for 1 min, 64°C for 20 sec and 68°C for 20 min; followed by incubation at 72°C for 5 min.

The PCR amplification product was treated with restriction enzyme DpnI for 7 h. An aliquot of 10 µL of PCR product was transformed into *E. coli* Top10F` chemocompetent cells and spread on LB plates containing 100 mg/ml ampicillin. A total of 4 colonies were selected and their plasmids were isolated using the GeneJET Plasmid Miniprep Kit (Thermo Scientific; Waltham,

[1] http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=29303&aa=1&style=N
Accessed 29 September 2014.

MA). The resulting plasmids were analyzed by sequencing using the primer QE_for (for sequence see Table S1).

4.2.3. Preparation of chemically competent *E. coli* cells and transformation

A glycerol stock of the desired *E, coli* strain was used to inoculate 10 mL LB medium. Cells were grown over night at 37°C and this culture was used to inoculate the main culture to a D_{600} of 0.05. Cells were incubated at 37°C and 120 rpm until the D_{600} of 0.7 was reached. The growth was stopped by putting the cells on ice and shaking for 10 min. Afterwards, cells were harvested by centrifugation in pre-chilled falcon tubes at 4000 rpm and 4°C for 10 min. The resulting pellet was resuspended in 900 µL ice-cold TSS (for composition see Table S7) and 90 µL aliquots of the resuspended cells were flash-frozen in liquid nitrogen and stored at -80°C.

The chemical transformation was performed according to the protocol "Transformation of chemically competent cells[2]" (Corinna Odar, ACIB). 160 µL of the chemically competent *E. coli* TOP10F`cells were thawed on ice and mixed with KCM (for composition see Table S7) to a final concentration of 1X KCM. The pre-chilled DNA was mixed with the cells and put on ice for 2-10 min. After a heat shock for 90 sec at 42° C the mixture was put back on ice for 1 min.

The electroporation was performed according to the protocol BIW#35[2]. Prior the electroporation the reaction mixture was desalted by dialysis (protocol BIW#23[2]) against water. 160 µL electrocompetent *E.coli* TOP10F`cells were thawed on ice and mixed with1-3 µL of the desalted reaction mixture. The transformation happened by an electric pulse (1.5 kV, 100 Ω, 25 µF).

Transformed cells were regenerated in SOC medium (for composition see Table S7) at 37°C and 650 rpm for 1 h. Transformed cells containing the destination vectors were regenerated in SOC medium with addition of 20 mM glucose. The regenerated cells were used to inoculate LB medium containing the appropriate antibiotics and X-Gal. Glucose (20 mM) was added in case of cells containing destination vectors.

The resulting clones were analyzed by restriction analysis or by sequencing.

4.2.4. Expression in *E. coli*

The *E. coli* strain BWEC46 was used for the expression of the codon optimized FIA variants as well as not codon optimized FIA. The overnight cultures in LB medium contained the appropriate

[2] The protocol is available from the author upon request.

vector expressing one of the FlA variants and were used to inoculate the main culture of LB medium containing ampicillin (100 mg/mL) to a starting D_{600} of 0.1. Cells were incubated at 37°C and 120 rpm until the D_{600} of 0.5 was reached. The expression of FlA was induced by addition of 0.1 mM IPTG and performed at 28°C and 120 rpm for 3.5 h or 5 h. Samples for the analysis by SDS-PAGE were taken 3.5 h or 5 h after induction of the FlA and contained the amount of cells equivalent to a cell density of 1 D_{600}.

4.2.5. Cell disruption with BugBuster™

Prior the cell disruption, the BugBuster 10X Protein Extraction Reagent (Merck Millipore, MA) and benzonase were each diluted 1:10 using HEPES buffer (10 mM HEPES, pH 7). The working solution was prepared by mixing the diluted benzonase and the extraction reagent 1:100.

A cell amount equaling 1 D_{600} unit was harvested by centrifugation at 4°C and 13,000 rpm for 5 min. The supernatant was discarded and the pellet resuspended in 35 µL BugBuster working solution. Samples were vortexed properly and incubated at 24°C and 900 rpm for 20 min. After centrifugation for 10 min at 4°C and 13,000 rpm the supernatant containing the soluble protein fraction was collected. The pellet was resuspended in 35 µL 6 M urea and contained the insoluble protein fraction. All samples were stored at -20°C.

4.2.6. SDS PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to the protocol described by Laemmli (Laemmli 1970). All SDS gels with the exception of NuPAGE® 4-12% Bis-Tris precast gels (NuPAGE®, Life Technologies, Carlsbad, CA) were self-made. The stacking gel and the separating gel contained 4% and 12% polyacrylamide, respectively. The components of the gels are summarized in Table S8.

The samples containing the lysate and pellet produced by cell disruption with BugBuster™ were treated prior to the SDS-PAGE. To each sample 40 µL H_2O and 25 µL loading dye (self-made, for composition see Table S7) were added. Samples were heated at 95°C for 5 min and centrifuged at 13,000 rpm for 2 min before loading. The amount of total protein loaded per lane equaled a cell density of 0.05 or 0.15 D_{600}. In order to estimate the size of the protein bands, a protein standard was loaded on each gel (PageRuler pre-stained protein ladder, Thermo Scientific, Waltham, MA). The gels were run at 100 V for 15 min, followed by 180 V for 40 min. Gels were stained with Coomassie Brilliant Blue G250 (Roth, Kalsruhe, Germany). Picture

editing such as adding text and numbering was performed with CorelDRAW Graphics Suite from Corel Corporation.

4.2.7. Assembly reactions

The assembly reactions of destination vectors were carried out in one pot with approximately 100 ng of each PCR amplified subpart (as described in Table S2 and Table S3) in a total volume of 20 µl, 10 U of restriction enzyme (FastDigest, Thermo Scientific or High Fidelity, New England Biolabs), 25 mM ATP, 10 U of T4 or T7 DNA ligase (New England Biolabs) and 2 µL appropriate buffer. 10 mM DTT was added to assembly reactions carried out with BpiI and BsmBI.

The vector pMC0.3 was used for the assembly of plasmids containing basic biological parts. Roughly 100 ng vector, 500 ng of PCR amplified biological part, 10 U of BsaI (FastDigest, Thermo Scientific or New England Biolabs), 25 mM ATP, 10 U of T7 DNA ligase (New England Biolabs) and 2 µL appropriate buffer were combined per assembly reaction. The basic biological parts FlA and *lacI* comprising three and four subparts, respectively, were pre-assembled using a total amount of 1.5 µg DNA. No destination vector was added to these reaction mixes.

For the assembly of connector plasmids, the vector backbone of pMC0.3 without the *lacZ* gene was used. The assembly reactions were carried out using the connectors as gBlocks® with a vector to insert ratio of 2:1; or subcloned into pJET1.2 with a vector to insert ratio of 1:5; or after restriction from pJET1.2 and gel-purification with a vector to insert ratio of 1:1; or after PCR amplification using the gBlocks as the templates with a vector to insert ratio of 1:2. Additional components of the reaction mix were 10 U of BsmBI (FastDigest, Thermo Scientific), 25 mM ATP, 10 U of T4 or T7 DNA ligase (New England Biolabs) and 2 µL FastDigest buffer. The assembly reaction was performed with all components combined in one Eppendorf tube.

The assembly reactions were carried out using one of two different approaches. For the cyclic approach (cyc), restriction (37°C, 2 min) and ligation (22°C for T4 and 25°C for T7 DNA ligase, 5 min) were run in cycles. After 24 cycles the samples were heat inactivated (65°C, 10 min).

For the second, the non-cyclic approach, restriction and ligation were performed successively for either 45 min or 2 h each (nc45 and nc2h, respectively) followed by heating up to 65°C for 10 min. It was confirmed in an earlier laboratory project (Koshanskaya 2014) that the ligase maintains its activity after heating to 37°C for 2 h.

4.2.8. Agarose gel electrophoresis

DNA fragments and plasmids were analyzed and purified by agarose gel electrophoresis. It was performed according to a standard protocol (Armstrong & Schulz 2008). The agarose gels were run in 1 X TAE buffer (for composition see Table S7) and the size of the DNA fragments was estimated by comparison with the DNA ladder (GeneRuler™ 1 kb Plus DNA Ladder, Thermos Scientific), which was loaded on each gel.

4.2.9. Gibson cloning of RLSS into p15aARA

The Gibson cloning was performed according to the standard protocol as described by Gibson et al. (Gibson 2011). The PCR amplified RLSS and the linearized vector p15aARA were added in equimolar amounts to 15 µL of Gibson assembly master mix (for composition see Table S7). The assembly reaction was performed at 50° C for 1 h and followed by chemical transformation into *E.coli* strain Top10F` as described above. Resulting clones were analyzed by sequencing.

4.2.10. Co-expression experiment of FlA and RLSS in *E. coli*

For the co-expression experiment the FlA, which was not codon optimized for *E. coli* and the RLSS, which was codon optimized, were used. Four cell cultures in LB medium were produced by co-transforming different vector combinations (see section 5.3.1) into the *E coli* strain BWEC46. Cell cultures were incubated at 37°C and 120 rpm over night and used to inoculate the main culture of LB medium containing the appropriate antibiotics to a starting D_{600} of 0.1. Cells were incubated at 37°C and 120 rpm until the D_{600} of 0.5 was reached. The expression of RLSS was induced by addition of 0.2*% v/v* arabinose and performed at 28°C and 120 rpm. When D_{600} of 0.9 was reached, the expression of FlA was induced by addition of 0.1 mM IPTG. Samples for the analysis by SDS-PAGE were taken 5 h after induction of the RLSS and contained the amount of cells equivalent to a cell density of 1 D_{600}. Cells were disrupted by BugBuster and treated prior to SDS-PAGE as described above in sections 4.2.5 and 4.2.6.

4.2.11. Preparation of lyophilized cells

E.coli BL21(DE3)Gold cells carrying different vector constructs were used to inoculate four overnight cultures (ONC; C1, C2, C3 and C4 as indicated in Table 1). These ONCs were used on the next day to inoculate 500 mL LB medium in baffled shake flasks containing the appropriate antibiotics (50 mg/mL kanamycin and/or 100 mg/mL ampicillin) to a starting D_{600} of 0.05. ONC of C2 was also used to inoculate C5 (Table 1). Cells were cultivated at 37°C and 120 rpm until they reached a D_{600} of 0.4. For the subsequent protein expression the temperature

was lowered to 28°C and the expression of RLSS was induced by addition of 0.4% arabinose to all five cultures. Cells were cultivated at 28°C and 120 rpm until the D_{600} reached 0.9. Then the expression of FlA was induced by addition of 0.1 mM IPTG to cultures 1 to 4. The expression of FlA in culture 5, carrying pQE80L-FlA and p15aARA-RLSS, was induced at D_{600} of 4.5. Cells were cultivated overnight under the same conditions and were then harvested by centrifugation at 4°C and 4000 rpm for 15 min. Each cell culture was washed with 250 mL ice-cold Tris/HCl buffer (pH 7.8), followed by a centrifugation step at 4°C and 4000 rpm for 15 min. The resulting pellets were resuspended in 50 mL ice-cold ddH_2O, transferred into pre-chilled round bottom flasks and chilled in liquid nitrogen by gentle shaking. Cells were freeze dried over night. The lyophilized cells were transferred into Greiner tubes and stored at -20° C.

Samples for the analysis by SDS-PAGE were collected before the induction of RLSS, before the induction of FlA at D_{600} of 0.9 and 4.5 and after the overnight expression and contained the amount of cells equivalent to a cell density of 1 D_{600}. Cells were disrupted by BugBuster and treated prior to SDS-PAGE like described above in sections 4.2.5 and 4.2.6.

4.2.12. FDA assay using lyophilized cells

To study the influence of the co-expression of the RLSS on the FDA production by the five different cell cultures (as described in 4.2.11.) the assay was performed using 50 mg lyophilized cells resuspended in 350 µL of 50 mM Tris/HCl buffer (pH 7.8) for each reaction. In order to lower the experimental error, the amount of cells of each culture, which was needed for all reactions, was diluted with the corresponding volume of 50 mM Tris/HCl buffer (pH 7.8) and then 350 µL of cell suspension were distributed to the single reaction tubes. The reactions were supplemented with 200 mM KF (in ddH_2O) and 2 mM SAM (in 1 mM HCl) and incubated at 37°C and 400 rpm for 1 h. Reactions were stopped by protein precipitation at 95°C and 900 rpm for 5 min, followed by a centrifugation step at 13,000 rpm for 5 min. The supernatant was collected and the pellet discarded. 150 µL of the supernatant were filtered using the MultiScreen® Filter Plate (Merck Millipore, MA). 100 µL of the filtered supernatant were transferred into HPLC vials and stored at -20°C until further analysis.

In order to compare the FDA production by the five different cell cultures (as described in 4.2.11.) depending on the amount of the enzymes present in the cells, the assay was performed with a varying amount of each lyophilized cell culture. For this a stock suspension with 150 mg/mL of each cell culture was prepared in 50 mM Tris/HCl buffer. Each stock suspension was used to prepare a dilution series with the following concentrations (mg/mL): 100; 75; 50; 25; 10

and 5. Each sample was supplemented with 200 mM KF (in ddH_2O), 2 mM SAM (in 1 mM HCl) and processed as described above.

In order to compare the intracellular SAM levels of the five different cell cultures (as described in 4.2.11.) the assay was performed using 47 mg of each lyophilized cell culture resuspended in 350 µL of 50 mM Tris/HCl buffer (pH 7.8) for each reaction. Each reaction was supplemented with 200 mM KF (in ddH_2O) and 20 mM L-methionine (in 1 mM HCl). 20 mM ATP (in ddH_2O) was added or not. Samples were processed like described above.

4.2.13. HPLC-UV measurements and data evaluation of the analysis of FDA production

The samples of the FDA assay were measured using a HPLC-UV equipment from Agilent Technologies and a ZORBAX Eclipse XDB-C18, 5 µm, 4.6 x 150 mm cart column, both from Agilent Technologies (Santa Clara, CA). The settings for the HPLC method are summarized in Table S9.

Data evaluation and visualization was done by Microsoft Excel (Microsoft, Redmond, WA). The retention times for FDA and SAM were evaluated by analyzing pure compounds. The measured area of the relevant peaks was automatically integrated by the HPLC software. The relative amounts of FDA and SAM produced by lyophilized cells were correlated with the integrated area of the peaks appearing at the expected retention time. The bigger the peak area was the more of the corresponding compound was present.

For the data evaluation the blank values for SAM and FDA, meaning the SAM and FDA values of cell cultures 1-5 without addition of SAM and KF, were subtracted from the results of the reactions.

5. Results

5.1. Comparison of the codon optimized variants of FlA concerning the solubility and the expression levels

The enzyme fluorinase from *S. cattleya* was codon optimized and harmonized for the expression in *E.coli* but the applied approach did not significantly improve the solubility and expression levels of the FlA.

The optimization was performed using the program Gene Designer from DNA2.0 Inc. It is a stand-alone software for design of synthetic DNA segments and is described in detail by Villalobos et al. (Villalobos et al. 2006). The optimization was performed according to the codon usage of *E. coli* class I, which is based on a collection of genes involved in most metabolic processes in *E.coli* (Villalobos et al. 2006). Certain restriction and methylation sites were avoided as described in section 4.2.1. Additionally, codons with a probability score lower than 10% were omitted by the algorithm. We avoided mRNA secondary structures and DNA repeats by setting the minimum repeat size to 12. This ensured that the resulting mRNA is not forming double-stranded RNA stems and the DNA is not predominantly self-annealing during the assembly. The algorithm also optimized the codon usage of the first 15 codons which is important for the efficient expression of a gene (Stenström et al. 2001).

Additionally, nine codons were determined to be rare in the host organism *S. cattleya* by the program Graphical Codon Usage Analyser and were manually exchanged against correspondingly rare codons used by *E.coli* (see Table S10) to harmonize the sequence. The codon optimized version of the FlA (FlA c.o1) is represented in section 9.5. It was ordered as gBlock® and cloned into the vector pQE80L by Gibson cloning. The resulting plasmids were sequence-verified and the expression was performed like described in section 4.2.4 and followed by SDS PAGE analysis.

According to the results of the SDS-PAGE, the codon optimized version of FlA (FlA c.o1) was not expressed (Figure 7). Neither the pellet nor the lysate contained a visible band of the expected size (34 kDa). In order to find the reason, the strategy of the codon harmonization was reviewed and it was found that the manual exchange of the rare codons had been performed incorrectly. Erroneously, some rare codons used by *S. cattleya* were exchanged against abundantly used *E. coli* codons. This, of course, did not comply with codon harmonization. Especially, the incorrectly optimized codons at the 5'-terminus encoding for amino acids alanine, arginine and proline at positions 2, 7 and 9, respectively, were considered critical since they

might negatively influence the secondary structure of the FlA at the ribosome binding site or start codon and lead to the inhibition of the translation. Hence, these three codons were exchanged against less abundant codons of *E. coli* by QuikChange PCR (see 4.2.2) giving rise to the second variant of the codon optimized FlA (FlA c.o2). The expression of FlA c.o2 was performed like described in section 4.2.4 and followed by SDS PAGE.

The expression levels of the three different FlA variants are shown in Figure 7. It was clearly visible that the FlA, which was not codon optimized (FlA n.o, lanes 3 and 4), showed the highest expression level as compared to the optimized variants. According to this gel, there was no expression of FlA after the first optimization (FlA c.o1, lanes 5 and 6). The harmonization of the three codons encoding Ala, Arg and Pro at positions 2, 7 and 9 restored the expression of FlA c.o2. However, the expression level was lower than that of the non-optimized FlA (FlA c.o2, lanes 7 and 8). Most FlA n.o and FlA c.o2 ended up in the insoluble fraction. Hence, the codon optimization did not significantly influence the solubility of the FlA.

The described codon optimization approach was not suitable for the improvement of the FlA expression. Based on these results, the FlA, which was not optimized for the expression in *E. coli*, was used for all further work. Additionally, we found out that the usage of rare codons, especially in the region of the translation start, turned out to crucially influence the expression of the FlA.

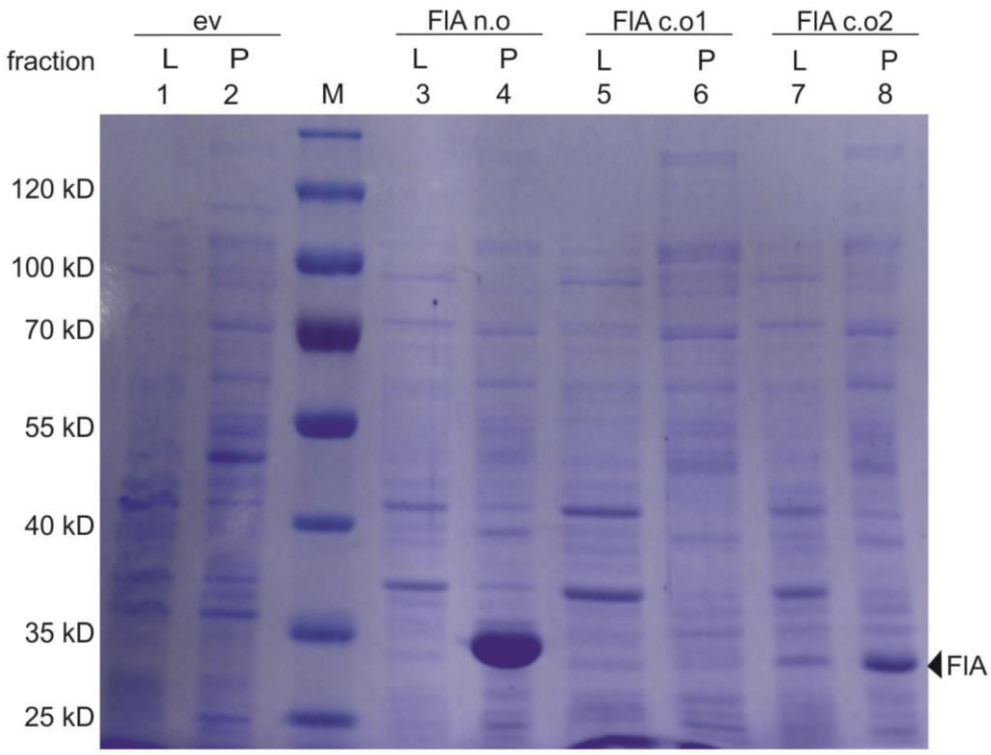

Figure 7: Comparison of the expression levels of fluorinase which was not codon optimized for E. coli (FIA n.o) and two codon optimized variants (FIA c.o1 and FIA c.o2).
The amount of total protein loaded per lane equaled a cell density of 0.05 D_{600}. Ev, empty vector pQE80L; L, lysate (soluble protein fraction); P, pellet (insoluble protein fraction); M, molecular weight marker. The arrow indicates the calculated molecular weight of FIA (34 kDa). 12% Coomassie-stained SDS gel.

5.2. Modular cloning approach

We used the modular cloning approach for the combinatorial and directed assembly of an enzyme cascade producing the fluormetabolite FDA. The implementation of this task comprised of several steps. First of all, we had to design and assemble the necessary destination vectors for levels °1 and °2. The level °0 destination vectors were assembled during an earlier laboratory project (Koshanskaya 2014). As well, the parts comprising the transcriptional units of FIA and RLSS, i.e., the coding sequences of the genes, promoters and terminators etc., needed to be chosen and produced. The next step involved the assembly of part and connector plasmids followed by the second cloning step of those to give device plasmids of FIA and RLSS. In the last assembly reaction the two devices had to be assembled in the level°2 destination vector giving an enzyme cascade for the production for FDA.

5.2.1. Design and construction of the relevant destination vectors

The destination vectors for level 1, pMC1.1 and pMC1.2, and for level 2, pMC2.1 and pMC2.2, were designed as described in the Introduction. The corresponding vector maps are shown in section 9.2. The destination vector pMC1.1 was designed for the assembly of transcriptional units directly used for protein expression. It carried a ColE1 origin of replication and an ampicillin resistance marker. The destination vector pMC1.2 was designed as the level 1 shuttle vector and should be used for further assembly of devices into the level 2 destination vector to give multi-gene constructs. Compared to pMC1.1, this vector was carrying a high copy number pMB1 origin of replication and a kanamycin resistance marker. The level 1 destination vector pMC1.2 should be used with the level 2 destination vector pMC2.2, which comprised a ColE1 origin of replication and an ampicillin resistance marker.

All destination vectors carried the *lacZ* gene for the blue/white selection of inserts. The used strain, *E. coli* TOP10F`, carried a mutant β-galactosidase lacking the N-terminal amino acids 11-41 which disabled oligomerization for a functional enzyme (Huo et al. 2009). Destination vectors carried the *lacZ* gene encoding the missing part of β-galactosidase. When LacZ was expressed, the intact enzyme could be produced and X-gal in the agar plates was degraded giving rise to blue colonies. This means that cells carrying destination vectors without an insert exhibit blue color.

The subparts comprising the level 1 destination vectors pMC1.1 and pMC1.2 are summarized in Table S2. Both vectors were assembled using BsmBI. Two different DNA ligases were used for the assembly reactions. On the one hand, unexplainable assembly structures occurred in former experiments with the T4 DNA ligase. Therefore, we also used the T7 DNA ligase, which is not an efficient catalyst of blunt end ligations, in order to test if this reduces random sequence assemblies. The transformation gave 60 transformants when using the T4 DNA ligase and 45 transformants when using the T7 DNA ligase. In total, four clones (two of each) were analyzed by restriction and sequencing. All clones showed a similar restriction and sequencing pattern and led to correct vector constructs. We concluded that there is no difference between these two DNA ligases for this application.

The first round of assembly reactions of pMC1.1 and pMC1.2 were performed using the cyclic approach. The transformation was performed chemically and gave around 50 clones for pMC1.1 and no clones for pMC1.2. In total, four clones were analyzed by streaking out on agar plates containing X-Gal. Differently than expected, all clones retained the white color. They were analyzed by restriction and sequencing. The agarose gel of the digested samples showed

bands of bigger size than expected (data not shown). According to the sequencing results, the analyzed sequence of all two clones corresponded to the sequence of pSYN53. This vector had been used as the template to PCR amplify the λt_0 terminator, one of the subparts of pMC1.1. It seemed that this vector was still present in the reaction mix and needed to be digested and the λt_0 terminator part gel-purified.

In principle, gel purification of DNA fragments is not necessary for MoClo. On the contrary, it is omitted because the parts are introduced into the assembly reaction in the form of circular plasmids. However, since the plasmid that carried the λt_0 terminator part would not get lost, we excised the λt_0 terminator subpart and gel-purified it. Afterwards, the assemblies of pMC1.1 and pMC1.2 were performed using the non-cyclic approach with 2 h for restriction and ligation each. 10 mM DTT was added to the reaction mix and the assemblies were electroporated into the *E. coli* cells. Around 100 clones of pMC1.1 and 10 clones of pMC1.2 were observed. Four clones of each vector were streaked out on agar plates containing X-Gal. Clones of pMC1.1 turned blue, as expected but the clones of pMC1.2 remained white. The analysis of all restreaked clones by sequencing showed that none of the pMC1.2 clones contained the *lacZ* transcriptional unit, as suggested by the white colony color. Two clones of pMC1.1 contained the correct sequence but also an insert within the promoter region of the *lacZ* gene (see Figure S10). Both clones were considered positive since the restriction and fusion sites, which are necessary for the construction of the next level constructs, were not affected. Moreover, the promoter was functional as indicated by the blue color of the clones on X-Gal plates.

The subparts comprising the level 2 destination vectors pMC2.1 and pMC2.2 are summarized in Table S3. Both vectors were assembled using BpiI. The vector pMC2.1 was assembled using the T4 DNA ligase and three different approaches: cyc, nc45min and nc2h (for details see section 4.2.7). The transformation was performed chemically and no glucose was added to the regeneration medium. Growth plates contained X-Gal. In case of the cyclic approach, 16 transformants were observed; only one in case of nc45 min and 12 in case of nc2h. All transformants were white. In total, seven clones were analyzed by restriction with BsaI, HindIII and NdeI and undigested samples were also loaded onto the agarose gel. The results are shown in Figure S9. The restriction with HindIII and NdeI showed a very similar pattern and indicated a vector size of around 1000 bp. The calculated size of pMC2.1 is around 5000 bp. In contrast to the HindIII/NdeI digest, the restriction pattern with BsaI seemed to be correct for the three cyc clones and two nc2h (clones 1 and 2). These five clones were analyzed by

sequencing but they did not contain the *lacZ* transcriptional unit. Since this vector was not relevant for further work, the cloning was put on hold.

To summarize the efficiency of the different assembly approaches, the results of the cyc and nc2h approaches yielded comparable numbers of transformants. As well, according to the restriction and sequencing analysis, these both approaches led to similar results. The nc45 min approach showed a low number of transformants and, according to BsaI restriction analysis, incorrect restriction pattern. This approach was therefore no longer used.

The vector pMC2.2 was assembled using the T7 DNA ligase and the nc2h approach. The assembly reaction was performed with and without the addition of 10 mM DTT, in order to investigate the influence of DTT on the activity of BpiI, the restriction enzyme used for the assembly of this vector. With addition of DTT, 100 clones of pMC2.2 were observed and no growth was observed without addition of DTT. In total, 20 clones were analyzed by streaking out on growth plates containing X-Gal. Two clones turned blue. They were analyzed by sequencing and represented the expected sequence.

In conclusion, the destination vectors pMC1.1 and pMC2.2 were successfully designed and assembled. It seemed that both, T4 and T7 DNA ligases led to similar results. The nc2h approach seemed to give better results concerning the assembly efficiency. The addition of 10 mM DTT clearly enhanced the assembly efficiency.

5.2.2. Modular assembly of an enzyme cascade for the production of FDA and SAM recycling in *E. coli*

The above described modular cloning strategy was used to set up a concept for the assembly of an enzyme cascade for the production of FDA and SAM recycling in *E. coli*. The modularity of this system would be applied for the assembly of three different versions of device I for the single expression of the FlA, as shown in Figure 8A. We decided to express the gene of interest, the fluorinase, under the control of the arabinose-inducible P_{BAD} promoter. It would be untagged or tagged with the glutathione-S-transferase tag (GST) or maltose binding protein tag (MBP) at the N-terminus. These tags were chosen because it was shown previously that they, specifically MBP, enhance the folding and soluble expression of proteins (Deng et al. 2008; Esposito & Chatterjee 2006). Device I is flanked by two different connectors which allow its seamless ligation into the level°1 destination vectors. The version of device I leading to the best expression levels and solubility of the FlA would be used for the co-expression experiment with the RLSS. Here, a two-gene construct containing devices I and II should be assembled into a

level°2 destination vector (Figure 8B). The connectors between these devices had compatible fusion sites and ensured a directional assembly of the transcriptional units.

A

con	DEVICE I					con
con 5`	**promoter**	**RBS**	**tag**	**GoI**	**terminator**	con 3`
conSa	P_{BAD}	RBS-IV	-	FIA	rrnB T1	conEc
conSa	P_{BAD}	RBS-IV	GST	FIA	rrnB T1	conEc
conSa	P_{BAD}	RBS-IV	MBP	FIA	rrnB T1	conEc

B

con	DEVICE I					con	con	DEVICE II					con
con 5`	**promoter**	**RBS**	**tag**	**GoI**	**terminator**	con 3`	con 5`	**promoter**	**RBS**	**tag**	**GoI**	**terminator**	con 3`
conSa	P_{BAD}	RBS-IV	-	FIA	rrnB T1	con1c	con2a	P_{T5}	RBS-III	-	RLSS	T7	conEc
conSa	P_{BAD}	RBS-IV	?	FIA	rrnB T1	con1c	con2a	P_{T5}	RBS-III	-	RLSS	T7	conEc

Figure 8: Expression constructs for the production of FDA.
A: Device I for the single expression of the fluorinase (FIA). It is flanked by connectors containing fusion sites compatible to the level°1 destination vector. B: Multi-gene constructs for the co-expression of FIA and rat liver SAM synthetase (RLSS) comprising devices I and II. Each device is flanked by the appropriate connectors in order to insure a directed cloning of the devices. The tag for the FIA was to be chosen depending on the results of the single expression experiment of the FIA which is indicated by the question mark. Con, connector; conSa and con2a connectors at the 5`terminus; conEc and con1c, connectors at the 3´terminus; RBS, ribosome binding site; GoI, gene of interest; P_{BAD}, arabinose inducible BAD promoter; GST, glutathione-S-transferase tag; MBP, maltose binding protein; P_{T5}, T5 promoter.

The parts for the modular assembly are summarized in Table S4. The parts FIA and *lacI* were divided in three and four subparts, respectively, in order to remove internal BpiI and BsaI restriction sites. For this, one base pair of the corresponding restriction site was exchanged such that the same amino acid was produced but the restriction site was destroyed. The mutation was introduced by PCR reaction with the original FIA sequence as template using primers carrying the mutations. After the PCR amplification the single subparts containing the mutations were produced. The correct ligation of the subparts to give the whole fragment was ensured by additional internal fusion sites. For better understanding the strategy is shown in Figure 9.

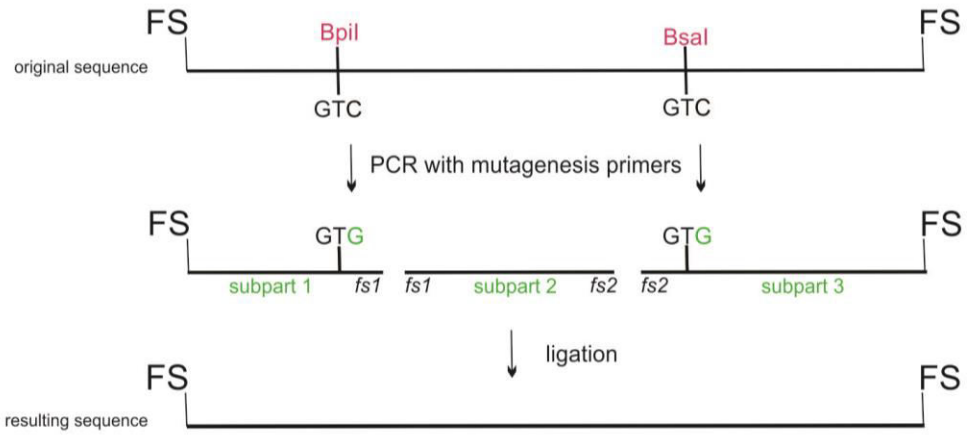

Figure 9: Schematic representation of the separation principle of the part FIA into three subparts in order to remove internal BpiI and BsaI restriction sites.
FS, outer fusion sites; fs1, fs2, additionally created internal fusion sites for directed ligation of the subparts. The number indicates the compatibility of the fusion sites.

The RLSS from *R. norvegicus* was codon optimized for the expression in *E.coli*. The optimization was performed using the program Gene Designer from DNA2.0 Inc. Additionally, three rarely used codons were harmonized manually (data not shown). The successful expression of the codon optimized RLSS in *E.coli* was confirmed by SDS PAGE (see Figure 10, lanes 8 and 9).

The connectors consisted of defined sequences including the fusion and restriction sites for all three cloning levels. Additionally, they contained 160 bp random sequences generated by the r2oDNA online tool (http://www.r2odna.com/). These random sequences represented biologically neutral orthogonal DNA spacer sequences rendering all connectors of the same size. This ensured that the cloning efficiency of the connectors was not affected by their size.

All parts needed for the modular assembly were successfully PCR amplified with the exception of the RLSS and the connectors. These parts were ordered as gBlocks® and subcloned into the pJET1.2 vector.

5.2.3. Assembly of part plasmids

Part plasmids are level°0 constructs comprise a level°0 destination vector and an integrated basic biological part or connector. Nine out of 13 basic biological parts, which are shown in Table S4, were successfully assembled into pMC0.3 (vector map represented in Figure S3. However, the assembly of the basic biological parts FIA, *lacI*, RLSS and bS3_RBSIV did not work.

The part plasmids were assembled as described in section 4.2.7. As indicated in Table S4, the assembly reactions of the part plasmids led to a variable number of transformants. The assembly reaction of RLSS into pMC0.3 gave no clones. The integration of the parts *lacI* and FIA into the destination vector yielded four transformants each, while the integration of the part GST led to 90 transformants. The ratio of blue to white clones was also variable. Some assembly reactions led to predominantly white transformants such as those of the part GST where the ratio of white to blue clones was 90:3. The assembly of the part bS1_RBSIV into pMC0.3, on the contrary, led to seven blue clones out of eleven clones in total.

Roughly two white clones of each part plasmid were analyzed by sequencing. Nine out of 13 part plasmids carried the correct inserts and were sequence verified. In case of the parts *lacI* and bS3_RBSIV the sequencing revealed that the destination vector pMC0.3 was not digested by BsaI and was still carrying the *lacZ* fragment. In case of FIA, only one of the three subparts was inserted into the destination vector (data not shown). The assemblies of these part plasmids were repeated again using the nc2h approach in order to increase the restriction time but the sequencing pattern remained the same. This hints at systematic difficulties with the sequences in digestion and/or ligation.

Since the assembly of part plasmids containing FIA and *lacI* did not work, it was tried to pre-assemble these parts using only the subparts but no destination vector. The remaining conditions were the same like for the assembly of the part plasmids. The outcome of the assembly reactions was controlled by agarose gel electrophoresis. The resulting gel showed the single bands of the subparts but no entire parts FIA and *lacI*. Thus, the pre-assembly reactions did not work either.

For the directed assembly of the devices four different connectors were used in this work: conSa, con2a, con1c and conEc. ConSa and con2a represented the connectors at the 5´terminus; con1c and conEc the connectors at the 3´terminus of a device. They carried the appropriate fusion and restriction sites, as represented in Figure 6. Connector plasmids were assembled using the pMC0.1 destination vector (Figure S2) and BsmBI. Since connectors contained internal BsaI restriction sites, which were essential for the level°0 and level°2 assemblies as represented in Figure 6, it was important to pre-cut the destination vector with BsaI and gel-purify the backbone prior to the assembly reaction. The restriction digest of pMC0.3 led to three fragments because of an additional BsaI restriction site within the *lacZ* gene. Since these fragments were hard to separate on the gel due to a similar size, we decided on using the pMC0.1 vector for the assembly of connector plasmids. This vector featured the

same backbone but was carrying the *lacZα* fragment instead of the whole *lacZ* gene, leading to two distinct fragments after digestion with BsaI.

The assembly of the part plasmids containing connectors was performed as described in section 4.2.7. The assembly reaction performed with connectors as gBlocks led to 50-90 clones; the use of subcloned connectors lead to around 10 clones. All clones were, as expected, white and in total, 20 of them were analyzed by sequencing. The sequencing results showed that connectors were not assembled into the destination vector. The resulting sequences represented illegitimately religated backbone of the destination vector pMC0.1. The attempts to assemble pMC0.1 and the connectors after restriction from pJET1.2 and gel-purification or after PCR amplification gave no transformants.

In conclusion, nine out of 13 basic biological parts for the modular assembly of a biosynthetic pathway for the production of FDA and substrate recycling in *E. coli* were successfully assembled into pMC0.3. The assembly of parts FlA, *lacI*, RLSS and bS3_RBSIV into the first destination vector pMC0.3 did not work out as well as the pre-assembly of the parts FlA and *lacI*. Also the assembly of connector plasmids failed.

5.3. Production of FDA with lyophilized cells and co-expression of RLSS

Since the work progress on the modular assembly of a biosynthetic pathway for the production of FDA and substrate recycling in *E. coli* was behind schedule and the most important parts such as the *flA* and *RLSS* genes were not available as part plasmids, we decided on using a different strategy in order to accomplish the co-expression experiments.

We decided on using a two plasmid borne expression system with compatible vector constructs. Each of the plasmids was carrying one of the genes for the expression of the enzymes FlA or RLSS. The vector pQE80L encoded the enzyme FlA which was not codon optimized for the expression in *E. coli* and the expression of FlA was induced by addition of IPTG. The second vector p15Aara encoded the enzyme RLSS which was codon optimized and the expression of RLSS was induced by addition of arabinose. For more details of the expression conditions and process see sections 4.2.10 and 4.2.11.

Before the experiments with the lyophilized cells could be conducted, it was important to perform a test expression in order to make sure that the RLSS from *R. norvegicus*, which was codon optimized for the expression in *E.coli,* was expressed.

In order to perform the assays, the lyophilized cells were prepared as described in 4.2.11. The dried weight of the cells varied from 0.6 to 1.0 g per 500 mL culture. As we showed previously the preparation of the reactions for the HPLC measurements (see section 4.2.12) was not harmful to FDA (Corinna Odar, ACIB GmbH, Graz, unpublished results). Additionally, we measured following standards and blanks: 50 mM Tris/HCl buffer (pH 7.8); FDA (50 µg/mL), SAM (10 mM), KF (2 M), cell cultures 1 to 5 (no SAM or KF added), cell culture 1 supplied with KF (no SAM added) and cell culture 1 supplied with SAM (no KF added); ATP (100 mM, in ddH_2O), L-methionine (335 mM in 1 M HCl), 1 mM HCl, all reactions without supplementation with SAM. They showed that none of the reaction components or the lyophilized cells themselves gave a signal at the same retention time as FDA.

5.3.1. Comparison of the expression levels and solubility of FlA with and without co-expression of RLSS

To demonstrate that the RLSS, which was codon optimized for the expression in *E. coli*, was expressed at a detectable level and to get the first insights about the expression behavior of the FlA with the co-expression of RLSS, a test expression in *E. coli* strain BWEC46 was performed. As represented in Table 1, the main cultures C1-C4 were used for this experiment. Culture C5 was used for a later experiment.

Table 1: Overview of the cultures and their components used in the following experiments.

	Culture C1	Culture C2	Culture C3	Culture C4	Culture C5
FlA induced at D_{600}	-	+ 0.9	+ 0.9	-	+ 4.5
RLSS	-	+	-	+	+

The samples for the analysis by SDS PAGE were taken 5 h after the induction of the RLSS. The results are shown in Figure 10, which represents an SDS gel of the soluble and insoluble protein fractions of each of the four cultures. In the first two lanes, the soluble protein fraction (Figure 10, lane 1) and the insoluble fraction (Figure 10, lane 2) of the cells carrying both empty vectors are shown. As expected, neither a FlA nor a RLSS band are visible. Culture C3, where the cells carried the FlA overexpression construct but did not express RLSS, showed an overexpression band of the FlA at around 34 kDa. It was clearly visible that the greater portion of the expressed protein was insoluble (Figure 10, lane 4 *vs* lane 3, soluble protein fraction). C4,

that is cells carrying the RLSS overexpression construct, showed an expression band of RLSS at around 46 kDa (Figure 10, lanes 5 and 6). According to this gel, the RLSS was present in both, the soluble (Figure 10, lane 5) and the insoluble protein fractions (Figure 10, lane 6), yet the protein was mainly soluble. In C 2 where the cells co-expressed FlA and RLSS, the FlA band was quite pronounced indicating a relatively higher expression level of this enzyme (Figure 10, lanes 7 and 8). However, even with the co-expression of the RLSS, FlA was mainly found in the insoluble protein fraction (Figure 10, lane 8). Conversely, the co-expression of FlA did not change the expression behavior of the RLSS.

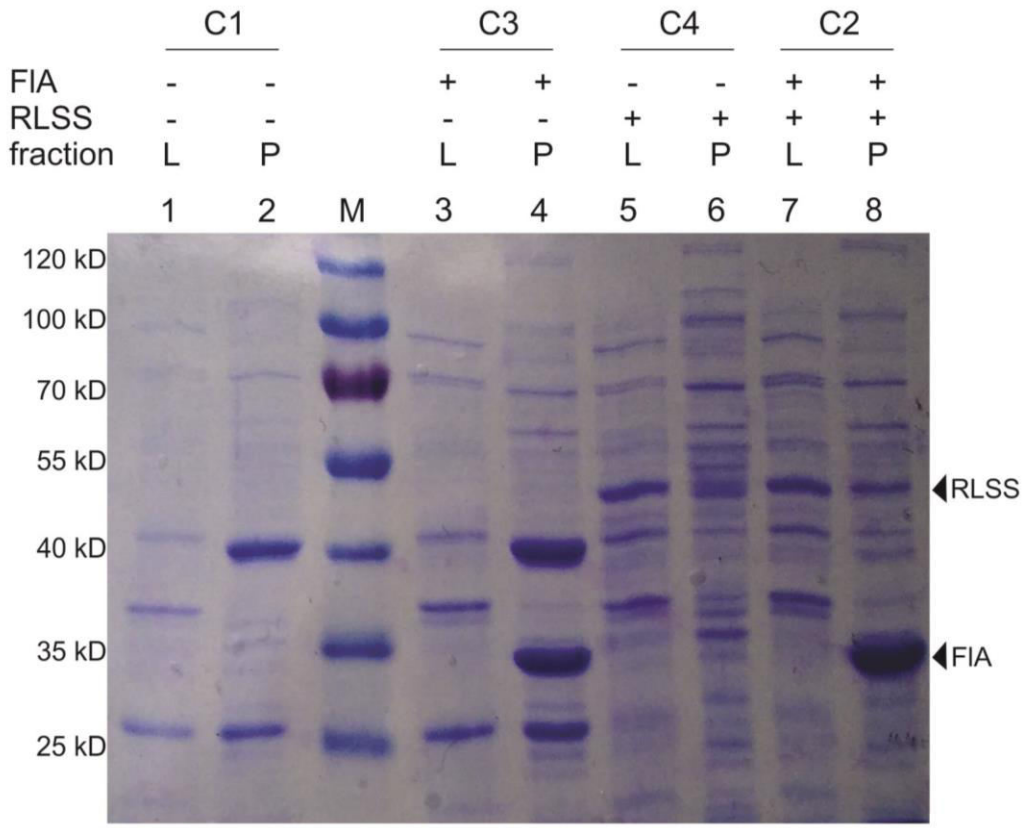

Figure 10: Comparison of the FIA expression levels with and without co-expression of RLSS.
Amount of total protein loaded per lane equaled a cell density of 0.05 D_{600}. C1-C4 are detailed in Table 1. L: lysate (soluble protein fraction); P: pellet (insoluble protein fraction); M: molecular weight marker. The arrows indicate the calculated molecular weight of FIA (34 kDa) and RLSS (46 kDa). 12% Coomassie-stained SDS gel.

We wanted to analyze whether the co-expression of RLSS affected the solubility of FIA depending on the time point of the expression initiation of the FIA. For this, four cultures C1-C4 carrying the above mentioned constructs were used, with addition of a fifth culture, where the expression of the FIA was induced at a D_{600} 4.5 (Table 1, C5). Samples for the analysis by SDS PAGE were taken 4.5 and 18 h after the induction of the FIA. The results are summarized in

Figure 11. Since in this experiment we were only interested whether the solubility of the FIA was enhanced we loaded only the soluble protein fractions on the gel. As expected, no FIA band was detected at any time in cells that carried the empty pQE80L and p15aARA vectors (Figure 11, lanes 2 and 7) or that overexpressed only RLSS (Figure 11, lanes 5 and 10). The highest amount of the soluble FIA protein was detected 18 h after the FIA induction in cells carrying the constructs for the co-expression of FIA and RLSS (Figure 11, lanes 8 and 11). However, the amounts of the FIA in these fractions were much lower than those in the insoluble protein fraction, as shown in Figure 11, lane 1.

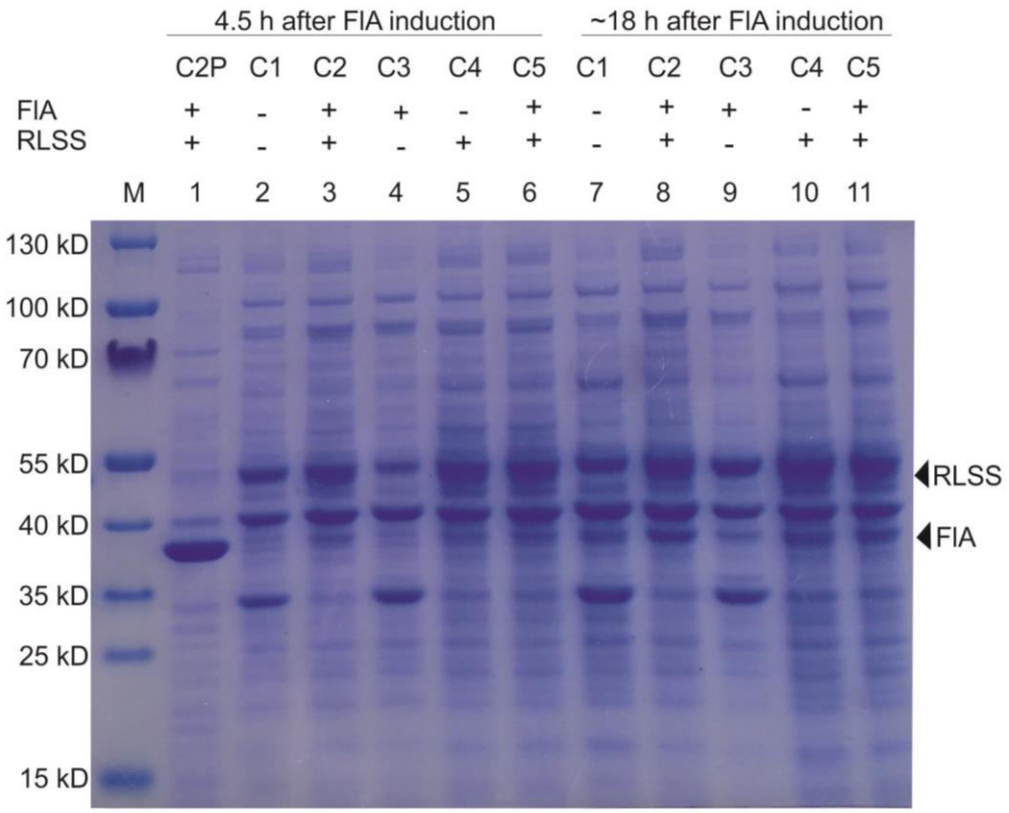

Figure 11: Co-expression of FIA and RLSS provokes no significant improvement in solubility of the FIA.
12% Coomassie-stained SDS gel picturing the lysate (soluble protein fraction). Samples were taken 4.5 h and 18 h after induction of FIA. An amount of total protein equaling a cell density of 0.15 D_{600} was loaded per each lane. Cultures C1-C5 were as described in **Table 1**; RLSS was induced at D_{600} 0.5 in all relevant cultures. For comparison, lane 1 indicates the overexpression of FIA in the insoluble protein fraction of cultureC2. . M: molecular weight marker. The arrows indicate the calculated molecular weight of FIA (34 kDa) and RLSS (46 kDa).

In conclusion, there was no significant influence of the solubility of the FIA with the co-expression of RLSS. Compared to other unspecific proteins, FIA was the most abundant protein in the insoluble protein fraction and less abundant in the soluble fraction. Concerning the

expression levels of FlA, we observed slightly increased amounts of the protein with the co-expression of RLSS, as shown in Figure 10, lane 8.

5.3.2. Comparison of the FDA levels produced by lyophilized cells with and without co-expression of RLSS

The lyophilized cells converted fluoride and SAM to FDA under the reaction conditions described in section 4.2.12. The results are summarized in Figure 12. For data evaluation, the measured amounts of SAM and FDA present in cultures C1-C5 (Table 1) without any additives were subtracted from the results of the reactions. This ensured that the represented data showed only the amounts produced during the reactions.

Panel A of Figure 12 shows the amounts of FDA produced by the lyophilized cells. As expected, no FDA was produced by cultures where the cells carried the empty vectors control and the RLSS expression construct (Figure 12A, C1 and C4). Comparing the cultures where the FlA was induced at the same time, one can see that the amount of FDA produced by the cells co-expressing the RLSS (Figure 12A, C2) is only negligibly higher compared to those without the co-expression (Figure 12A, C3). Comparing the amount of FDA produced by the cells carrying the co-expression constructs, it was visible that the cells, where the FlA was induced at a D_{600} of 0.9 (Figure 12A, C2), produced slightly more FDA compared to those, where the FlA was induced at D_{600} of 4.5 (Figure 12A, C5).

The Panel B of Figure 12 shows the amounts of SAM present in the reactions with the different strains. These SAM amounts were composed of the SAM, which was added to the reactions and the SAM, which was regenerated. The regeneration was performed, on the one hand, by the RLSS and, on the other hand, by the endogenous MetK, the SAM synthase of *E. coli*. In case of C1, the cells carried the empty vectors. Consequently, the comparably high level of SAM was due to the added SAM and the SAM, which was regenerated by the endogenous MetK using residual ATP and L-methionine present in the cells (Figure 12B, C1). Here, SAM was not used up by the FlA for the FDA synthesis and could accumulate. In case of C3, where only the FlA was expressed, the added SAM and SAM, which was regenerated by the endogenous MetK, was most probably used for the FDA production resulting in a comparably low intracellular SAM level (Figure 12B, C3). C2 and C5, where the cells carried the co-expression constructs, showed relatively low SAM levels (Figure 12B, C2 and C5). The SAM level of C2, where the FlA was induced at a lower D_{600} of 0.9, was lower compared to the SAM level of C5, where the FlA was induced at D_{600} of 4.5 (Figure 12B, C2 *vs* C5). The highest SAM

level was observed when only the RLSS was overexpressed. Presumably, this amount was composed of SAM, which was added to the reaction augmented by the SAM, which was

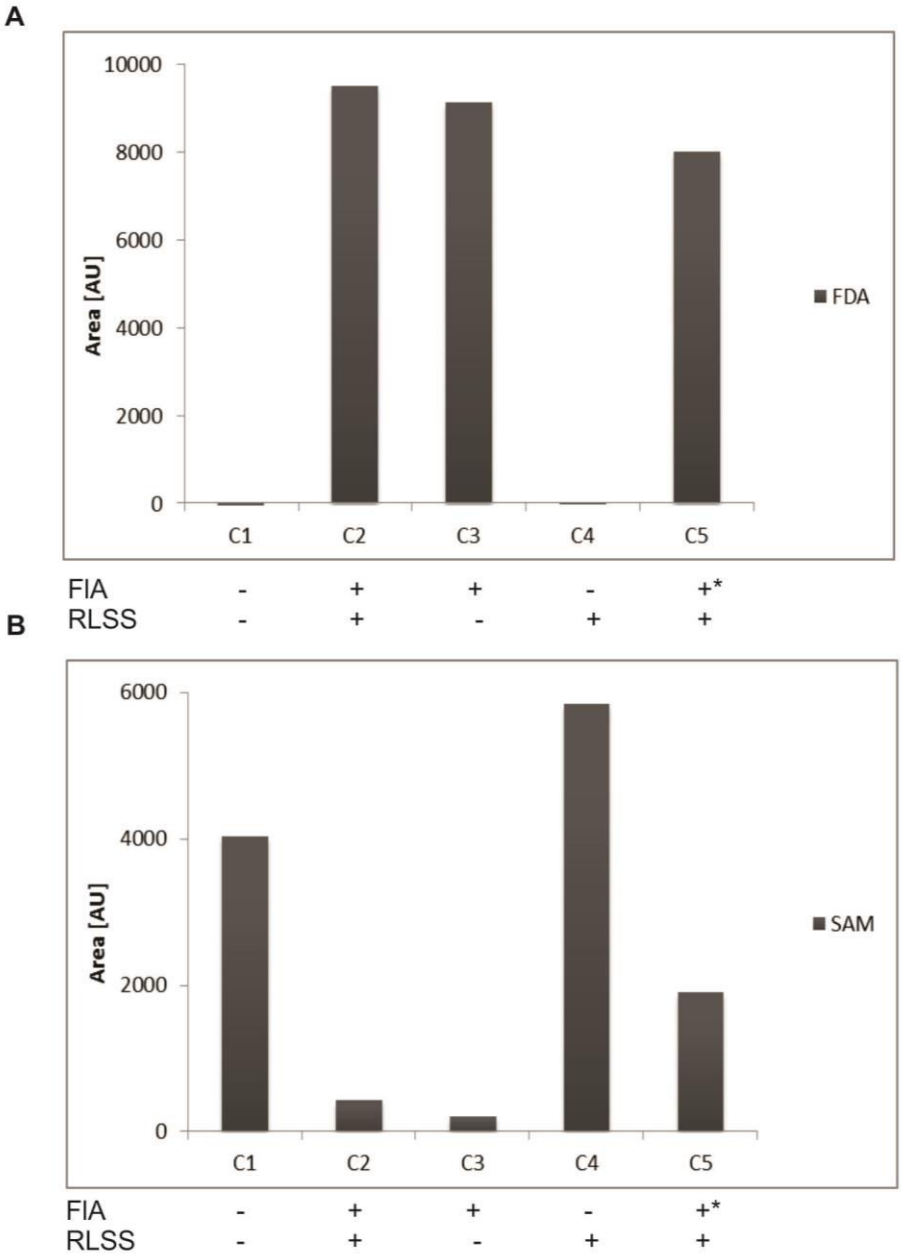

Figure 12: Comparison of the FDA (A) and SAM (B) levels produced by lyophilized cells after the supplementation with 200 mM KF and 2 mM SAM.
C1-C5 as described in Table 1. RLSS was induced at D_{600} 0.5 in all relevant cultures. Star (*) indicated a different time point of FIA induction (D_{600} 4.5 *vs.* D_{600} 0.9 in C2). Area indicates the integrated area of the relevant HPLC peak (see Materials and Methods for details of data evaluation). Data represent a single experiment.

regenerated by the overexpressed RLSS and the endogenous MetK. Since neither ATP nor L-methionine was added to the reaction, the enzymes used the chemicals residually present in the

cells. In case of the culture carrying the empty vectors (C1) and the one only expressing the RLSS (C4), the impact of the additional SAM and SAM regenerated by the endogenous MetK should be the same for both cultures, since no FDA could be produced. Hence, the reason for the elevated SAM level of the C4 expressing the RLSS was due to the regeneration of SAM by this enzyme. It was clearly visible that the regeneration efficiency of the overexpressed RLSS was much higher than the regeneration efficiency of the endogenous MetK.

Since the impact of the SAM regeneration by the enzyme RLSS did not lead to a pronounced elevation of the FDA levels, we decided to perform an experiment where we varied the amount of the lyophilized cells per reaction and, hence, the amount of FlA. We assayed the FDA production by FlA with (C2) and without (C3) the co-expression of the RLSS and the assay conditions were as described above with the exception that the amounts of lyophilized cells varied between 100 and 5 mg per milliliter. Figure 13 shows the results of this experiment.

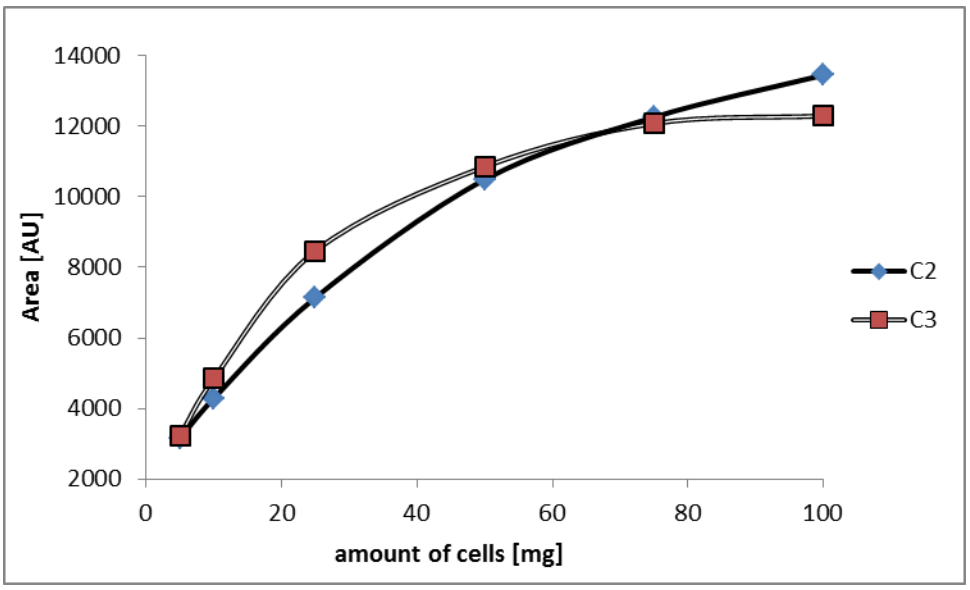

Figure 13: FDA production was directly related to the amount of lyophilized cells and therefore FlA and the co-expression of RLSS had a beneficial effect only at high amounts of cells.
The analyzed cells carried double overexpression constructs pQE80L-FlA and p15aAra-RLSS, FlA was induced at D_{600} 0.9 (C2); and the FlA overexpression construct pQE80L-FlA (C3). The integrated areas of the relevant HPLC peaks are shown as a function of the amount of lyophilized cells. Data represent a single experiment.

As expected, the level of the produced FDA increased with the amount of cells per reaction. Up to 75 mg/mL lyophilized cells, the enforced regeneration of SAM by RLSS had the opposite effect as intended: It led to slightly lower FDA levels (Figure 13, C2) than with FlA alone (Figure 13, C3). However, when upper amounts of cells expressing only FlA and no RLSS were used the FDA production stagnated: 100 mg C3 cells produced approximately the same level of

FDA as 75 mg (Figure 13, C3, 100 mg *vs* 75 mg). In these cases, SAM regeneration became limiting. Obviously, the co-expression of RLSS alleviated this effect as it was less pronounced in C2. The overall amount of the synthesized FDA was only marginally higher with C2 than with C3. According to these results, the amount of produced FDA directly correlated with the amount of FlA and RLSS was beneficial for SAM recycling only at upper FlA levels.

5.3.3. *In situ* synthesis of SAM from L-methionine and ATP and its use for FDA production

In order to compare the SAM levels of cultures with and without co-expression of the RLSS, the assay was performed by the supplementation of the lyophilized cells with 200 mM KF and 20 mM L-methionine which is required for the SAM synthesis. Upon supplementation with ATP, L-methionine is converted to SAM by the enzyme RLSS. The reactions were performed with and without addition of 20 mM ATP and the samples were processed under the same conditions as described above. The results of the reactions which were supplemented with ATP are represented in Figure 14A. The reactions without ATP exhibited the same pattern but the amount of produced SAM was 10-fold lower.

The panel A of Figure 14 depicts the SAM levels of the reactions. Since no SAM was added to the reactions in this experiment, the detected amounts were composed of SAM produced by the RLSS and the endogenous MetK from the added L-methionine and ATP. C1 and C3 produced comparable amounts of SAM. Since these cultures did not express RLSS, the SAM levels were generated by the cellular MetK and were comparably low (Figure 14A, C1 and C3). C4 expressed RLSS and displayed an approximately 20-fold higher SAM level (Figure 14A, C4). Even if the SAM level of C1 (SAM production only by MetK) was subtracted from that of C4 (SAM production by RLSS plus MetK), it was clearly visible that the RLSS was much more active and efficient than the endogenous MetK. C2 expressed both the RLSS and FlA and it showed a lower amount of SAM compared to C4 expressing only the RLSS (Figure 14A, C2 *vs.* C4) This can be explained by the activity of the FlA, which was using up SAM for the synthesis of FDA (see below). C5, which carried the expression constructs for RLSS and FlA and where FlA was induced later (D_{600} 4.5), showed a higher SAM level compared to the analogous C2 where FlA had been induced earlier (D_{600} 0.9). The explanation is that due to the induction of FlA at a higher cell density, the cells had more time to accumulate SAM until its consumption for FDA synthesis by FlA, which started later.

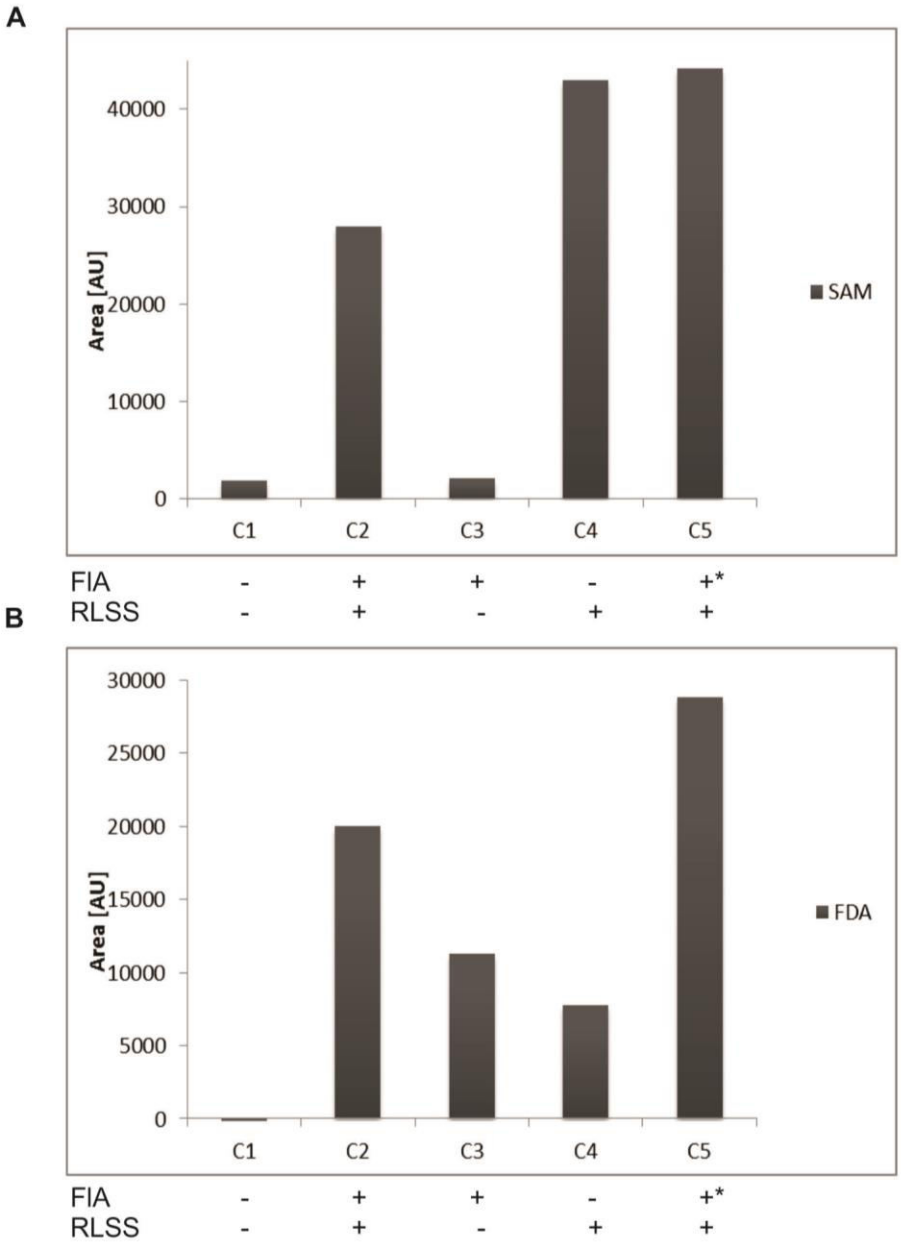

Figure 14: SAM (A) and FDA (B) levels when SAM was synthesized *in situ* from ATP and L-methionine. Cells were supplemented with 20 mM L-methionine and 20 mM ATP. C1-C5 as indicated in **Table 1**. RLSS was induced at D_{600} 0.5 in all relevant cultures. Star (*) indicated a different time point of FIA induction (D_{600} 4.5 *vs.* D_{600} 0.9 in C2). Area indicates the integrated area of the relevant HPLC peak (see Materials and Methods for details of data evaluation.) Data represent a single experiment.

To trace the *in situ* synthesized SAM we assessed the FDA levels of the samples shown in Figure 14B. The highest level of FDA was observed with C5 co-expressing RLSS and late-induced FIA (Figure 14B, C5). This might be due to the high amounts of SAM in C5 (Figure 14A, C5) before the induction of the FIA. Elevated SAM levels might positively influence the activity of the FIA as SAM (Figure 14A, C2 *vs* C5) as well as FDA (Figure 14B, C2 *vs* C5)

levels were lower in C2 (RLSS plus FlA) than in C5 (RLSS plus late-induced FlA). C3, which expressed only FlA had low SAM levels (Figure 14A, C3) and also comparably low FDA levels (Figure 14B, C3). When SAM was added to the reaction (see section 5.3.2 above), C4 expressing only RLSS showed high SAM levels (Figure 12A, C4) but no FDA production (Figure 12B, C4) as expected. As well, C4 accumulated SAM when it was synthesized *in situ* from ATP and L-methionine (Figure 14A, C4). However, C4 yielded also a quite substantial FDA signal (Figure 14B, C4) although this strain did not express FlA and should, therefore, not be able to produce FDA. Most likely, this observation was an outlier of the measurement. To validate this hypothesis the assay must be repeated.

In conclusion, we demonstrated that heterologously expressed RLSS was more effective for the *in situ* synthesis of SAM from supplied ATP and L-methionine than the endogenous MetK, an *E. coli* SAM synthase. We also showed that the co-expression of RLSS with FlA increased the production of FDA from exogenously supplied KF, ATP and L-methionine. All presented results were from single experiments and the analysis should be repeated in order to validate the results.

6. Discussion

6.1. Codon optimization of *flA* gene for expression in *E. coli*

The enzyme fluorinase from *S. cattleya* catalyzes the carbon-fluoride bond formation between SAM and KF to produce FDA, the first intermediate of the 4-fluorothreonine pathway. Like all enzymes of this pathway, FlA showed very low soluble expression levels in *E.coli* (Corinna Odar, unpublished results).

As described in the literature (Deng et al. 2014), overexpression in *E.coli* of a codon optimized FlA from *Streptomyces* sp.MA37 with 87% homology to the *flA* gene from *S. cattleya* led to a soluble protein. Our idea was to perform a similar codon optimization of the original FlA from *S. cattleya* in order to improve the expression of the enzyme in soluble form in *E.coli*.

The results of the codon optimization experiment are represented in Figure 7 showing an SDS gel of the expression of the FlA variants in the soluble and insoluble protein fractions. It was clearly visible that the first codon optimized FlA variant (FlA c.o1) was not expressed under the given conditions. Most probably, the reason for the failure to express were the incorrectly optimized codons at the 5´-terminus of the FlA. The rare codons used by *S. cattleya* were exchanged against abundant codons of *E.coli* unsettling the theory of codon harmonization. This theory is based on the concept that protein synthesis and folding in *E. coli* happens co-translationally and the sequence-dependent change in translation kinetics might influence the nascent protein folding (Angov et al. 2008). This means that in order to ensure efficient expression of a gene from *S. cattleya* in *E. coli*, the codon usage frequencies must be adapted to closely match those of the native host. Briefly, codons that are rare in *S. cattleya* should be substituted by synonymous rare codons in *E. coli*. A change in codon usage frequencies can affect protein folding and function, and disparities in codon usage frequencies may lead to misfolded proteins (Angov et al. 2008). A high level of inconformity in codon usage between the species of the target gene and the expression host can lead to poor heterologous expression levels and solubility (Angov et al. 2011).

Besides that it was previously shown that the use of rare codons at the N-terminus instead of more common ones leads to an increase in expression because theses rare codons slow down the ribosome in the initial phase of translation (Goodman et al. 2013).

By exchanging the incorrectly optimized codons at the 5´-terminus by QuikChange PCR we produced a second variant of the codon optimized enzyme (FlA c.o2). As represented in

Figure 7, this variant showed lower expression as compared to the non-optimized variant (FlA n.o) which displayed the highest expression level among these three variants. Comparing the soluble protein fractions of FlA n.o and FlA c.o2, it seemed that FlA c.o2 contained more soluble protein, relative to the total protein amount. Since not all incorrectly optimized codons were exchanged by QuikChange PCR, the relative high amount of soluble protein might be due to the inconformity in the codon harmonization. On the other hand, the protein amount, which was loaded on each lane, was normalized to the cell density D_{600}. As relative D_{600} is less exact than relative protein concentrations the protein load was probably not exactly the same for all samples. Consequently, the FlA band intensities shown in Figure 7 might only roughly reflect the distribution of the enzyme in the soluble and insoluble protein fractions. Nevertheless, the codon optimization of the FlA for the expression in *E. coli* did not significantly improve the solubility of this enzyme. Maybe the application of a different codon optimization program or algorithm would enhance the FlA expression.

6.2. Modular cloning assembly of an enzyme cascade for the production of FDA in *E. coli*

During this study, we set up a modular cloning strategy for the assembly of a biosynthetic pathway for the production of the FDA in *E. coli*. As described in Figure 8A., three different constructs were designed for the single expression of the FlA. The enzyme was untagged or tagged at the N-terminus with a GST or a MBP tag in order to evaluate the potential influence of a tag on the expression efficiency and solubility of the FlA.

The enzyme FlA converts KF and SAM to the desired product FDA and as the byproduct L-methionine is formed. The enzyme RLSS can remove the L-methionine from the reaction equilibrium as it converts it back to SAM in the presence of ATP (Figure 2). The idea behind this study was to co-express these two enzymes in order to elevate the intracellular SAM levels and hence increase the FDA production. In Figure 8B we represented a two-device construct for the co-expression of FlA and RLSS under individual promoters which are differently induced. Each device encoded one of the enzymes, FlA or RLSS, together with its individually inducible promoter, if applicable a tag, and a terminator. The tagged FlA construct yielding the best (soluble) expression efficiency of the single enzyme was intended to be used for the co-expression construct.

In order to fulfill the task of the modular assembly, it was important to produce the destination vectors, clone the parts for the pathway and assemble them into the desired expression construct.

For the assemblies of the relevant destination vectors we tried different assembly strategies with the objective to come up with the optimal ones.

We performed the assemblies of the destination vectors using DNA ligases from two different bacteriophages. On the one hand, we used the T4 DNA ligase which is capable of joining blunt ends as well as sticky ends. Since the designed fusion sites represented sticky ends, blunt end ligations were undesired and would lead to incorrect constructs. Blunt ends might be generated by incorrect or incomplete restriction during MoClo, though no literature was found to corroborate this notion. As well, PCR fragments of destination vector subparts contained blunt ends. Hence we performed the assembly reactions using the T7 DNA ligase, which does not efficiently catalyze blunt-end ligations. The results were compared with respect to the transformation and assembly efficiency and showed that both DNA ligases led to correct constructs and were both suitable for this application.

Undesired blunt-end ligation was observed only in one case where we used subparts of the destination vectors, which were not gel-purified after PCR amplification. The PCR template of one of the subparts represented a plasmid and this subpart was incorrectly blunt-end cloned into the destination vector due to the residual template vector within the reaction mixture. However, after digestion by DpnI and gel-purification of this subpart, we never encountered the same problem again. Since the blunt-end ligation was observed only once, it seemed that the restriction with the typIIS enzymes during the assembly reactions produced the correct sticky ends in most cases, lowering the probability of blunt-end cloning. However, in case this problem recurs in future we would recommend to gel-purify all subparts of the destination vectors prior to the assembly reactions. Albeit, this considerably increases the workload for modular cloning.

As described in section 4.2.7, we performed the assemblies of the destination vectors using two different approaches. The cyclic approach comprised 24 cycles of restriction for 2 min, followed by ligation for 5 min. It was possible to run this reaction in cycles because the recognition sites of the typeIIS enzymes are removed after restriction and the ligated fragments are insensitive to the subsequent restriction step. The overall time of restriction and ligation of the cyclic approach was around 50 min and 2 h, respectively. The second, non-cyclic, approach included either 45 min or 2 h *en bloc* for restriction and ligation each. The results showed that the non-cyclic approach with 45 min reaction and ligation (nc45min) led to the lowest number of transformants

and incorrect constructs. The reason might be that 45 min for the restriction were not sufficient to produce the correct fusion sites hence leading to incorrect assemblies. The cyclic approach (cyc) and the non-cyclic approach with 2 h for restriction and ligation each (nc2h), showed similar results. These conditions yielded better results than nc45min. The number of transformants was comparable and both approaches led to correctly assembled destination vectors. It seemed that longer reaction times and alternating restriction and ligation had positive influence on the assembly efficiency.

Inspired by an article in a scientific internet blog[3], which postulated that the enzyme BpiI was exhibiting a higher activity when supplied with DTT, we conducted the assembly reactions of pMC2.2 with the addition of 10 mM DTT. We obtained roughly 100 transformants with addition of DTT and no transformants without DTT. It seemed that the addition of DTT crucially enhanced the assembly efficiency. For this reason, all assembly reactions performed with BpiI were supplemented with 10 mM DTT in order to increase the activity of this enzyme. The reason might be that DTT was stabilizing the enzyme but we could not find any literature clarifying this issue.

To summarize the optimal conditions for the assembly of destination vectors, we would recommend to gel-purify the subparts of the destination vectors prior to the assembly reactions. Both T4 and T7 DNA ligases proved to be suitable for this application, as well as the cyc and nc2h assembly approaches. When using the enzyme BpiI we highly recommend supplying the reaction mix with DTT.

The devices for the modular cloning of a biosynthetic pathway for the production of FDA comprised 13 different parts which are shown in Table S4. Additionally, four connectors were designed for this study. All parts, except FIA and *lacI* could be successfully produced by PCR, overlap extension PCR or ordered as gBlocks.

The parts FIA and *lacI* contained internal BsaI and BpiI restriction sites which had to be removed. We decided to split FIA into three and *lacI* into four subparts, which lacked the undesired restriction sites but could be linked together by type IIS cloning as outlined in Figure 9. We amplified the subparts by PCR with primers containing the corresponding mutations. The idea was to use all subparts in equal amounts for the assembly reactions of the part plasmids.

[3] http://2012.igem.org/Team:Freiburg/Project/Golden; accessed in January 2015.

As indicated in Table S4, nine out of 13 basic biological parts could be successfully cloned into the level 0 vector pMC0.3. The parts which failed to insert were FlA, *lacI*, RLSS and bS3_RBSIV.

Surprisingly, it was not possible to assemble the part bS3_RBSIV, although the part bS1_RBSIV could be successfully introduced into pMC0.3. These parts contained the same RBS sequence but a different fusion site at the 3`end, meaning that both parts shared a very high sequence similarity of 99%. Apparently, the efficiency of the assembly reactions was dependent on the sequence of the insert. Especially the sequence at the 5'- and 3'-termini of the part seemed to be important, since the difference in the above parts occurred there. This hypothesis was also confirmed by the fact that assembly reactions of certain inserts led to a high number of transformants carrying the correct constructs. Due to time constraints, we could not further investigate the correlation between the sequence and the efficiency of the assembly reactions.

The assembly reactions failed when we used the subparts of the FlA and *lacI* for the assembly of the corresponding part plasmids. The reason might be that the mutagenesis PCR failed and the newly created fusion sites contained mutations making the flanking fusion sites not compatible anymore. Hence, correct ligation between the subparts and of the assembled subparts into the destination vector could have been prevented (see Figure 9). This might be the reason why the pre-assembly reactions of these parts using the subparts and no destination vector did not work either. To solve this problem, it would be possible to order the FlA and *lacI* parts as gBlocks without the undesired restriction sites and use them for the assembly reactions of the corresponding part plasmids.

The assembly of the part plasmid containing the RLSS failed as well. The RLSS was used as a gBlock for this application. As observed previously, some parts, when used as gBlocks, could be assembled into the destination vector without any problems but some assembly reactions using gBlocks did not work at all. We presume, that the efficiency of the assembly reactions using gBlocks was depended, on the one hand, on the quality of the gBlock and, on the other hand, on the sequence of the gBlock. However, we did not further investigate the correlation.

None of the assembly reactions including the connectors was successful. The results of all four connectors were similar concerning the number of transformants and sequencing pattern. According to the sequencing results, no connectors were inserted or the destination vector was not digested. The reason might be a mistake in the connector design or the sequence dependence of the assembly reactions.

The assembly reactions of some part plasmids were performed using both assembly approaches (cyc and nc2h). For some inserts we observed that a particular approach was leading to a higher number of transformants and better assembly results. Since the ligation time for both approaches was 2 h, it seemed that the restriction time made the difference. Maybe the longer restriction time was better for some inserts because of their sequence complexity and quality. As well, the purity of the insert might play a role. Some undesired contaminations might lower or inhibit the activity of the restriction enzyme. In these cases, we would recommend using the non-cyclic approach with 2 h for restriction.

6.3. An approach for the production of FDA with lyophilized cells and the co-expression of the RLSS

Since we could not produce the constructs for the co-expression of the FlA and RLSS by the modular cloning technique in time, we decided on using a different approach. A two vector system was implemented. Each vector carried one of the enzymes. The FlA expression was controlled by the IPTG-inducible T5/*lacO* promoter and the expression of the RLSS by the arabinose-inducible pBAD promoter. The implementation of this system was possible because both vectors featured compatible replications of origin and different antibiotic resistance markers.

The dry weight of lyophilized cells varied from 0.6 to 1.0 g per 500 mL culture. The highest amount of cells was produced by the culture carrying the empty vectors thus not expressing any proteins. The lowest amount of cells was achieved by the culture carrying the co-expression constructs. Obviously, the co-expression of the recombinant proteins represented a growth disadvantage for the cells since they had to allocate some metabolic energy to the expression of the proteins and could not use all for growth.

Before the lyophilized cells were produced and the assays conducted, we performed a test expression as described in section 4.2.4. The aim was to make sure that the codon optimized RLSS was expressed in *E.coli* and to get the first insights of how far the co-expression of the RLSS was influencing the expression levels and solubility of the FlA. The results of the test expression were evaluated by SDS-PAGE analysis and the resulting gels are shown in Figure 10 and Figure 11. It was clearly visible that the codon optimized RLSS was expressed at a detectable level. Thus, the codon optimization was successful for this enzyme. The FlA showed a slightly more intense over-expression band with the co-expression of the RLSS. As described in the literature (Dong et al. 2004), SAM is not only a substrate for the FlA but is

tightly bound to the enzyme and cannot be removed during purification. According to this paper, SAM is an important structural component of the FlA and it seemed reasonable that RLSS co-expression for the increase of intracellular SAM levels might enhance the stability of the FlA and hence lead to a higher amount of FlA visible on the gel. However, according to Figure 11 showing the soluble protein fractions of the FlA, the co-expression of the RLSS did not lead to an increased soluble expression of the FlA. In order to obtain more reliable results it would be necessary to load equal protein amounts onto the gel and perform a Western Blot of the soluble protein fractions.

Summarizing the results of this experiment one can confirm that FlA was the most abundant protein of the insoluble protein fraction. Only a low amount of the FlA was expressed in soluble form. Additionally, the co-expression with the SAM-recycling enzyme RLSS led to higher expression levels of the FlA. The FDA levels produced by the lyophilized cells with and without the co-expression of RLSS are represented in Figure 12. It was visible that the culture expressing both the FlA and the RLSS showed slightly higher amounts of FDA than the culture expressing only the FlA. As described by Dong et al. (Dong et al. 2004), the extensive interactions between SAM and the FlA are leading to the formation of the active site of the enzyme. Thus, SAM is important for the activity of the FlA and hence, higher intracellular SAM concentrations might lead to an increased production of FDA. It was also clearly visible that more FDA was produced when the FlA was induced at an earlier D_{600}. This finding seems reasonable because there was more time for the FlA to convert the substrates KF and SAM to the desired product FDA.

As well, the SAM levels of the reactions seemed reasonable (Figure 12B). The highest amount of SAM was observed by C4 expressing only the RLSS and most probably represented the SAM, which was added to the reaction and the SAM, which was regenerated by the RLSS and the endogenous MetK.Since no FlA was expressed by this culture no L-methionine could be produced for the regeneration of SAM. It might be possible that SAM was hydrolyzed or degraded to L-methionine by the action of other enzymes, such as e.g. biotin synthase (Choudens et al. 2002). Hence, RLSS and MetK could use this L-methionine for the SAM regeneration. Expectedly, the activity of the overexpressed RLSS was higher (Figure 12B, C4) than the activity of the endogenous MetK (Figure 12B, C1), since the SAM amount of C1 was most likely composed of the SAM added to the reaction and the SAM regenerated only by the MetK. As described in the literature (Wei & Newman 2002), an increased expression of MetK by *E. coli* depletes the cells of methionine and thus interferes with the cell metabolism. For this

reason, the excessive expression of MetK represents a growth disadvantage for *E. coli* and needs to be strictly regulated. Most probably, the comparably lower expression level leads to a lower activity of the MetK compared to the over-expressed RLSS. The SAM level of the C3, which expressed only the FlA, was the lowest: It consisted of added SAM and SAM, which was regenerated by the endogenous MetK. Since the activity of the MetK was lower compared to the RLSS only little SAM could be regenerated by the MetK but the reaction was pushed by the excess of L-methionine from the FlA reaction and possible side reactions cleaving SAM to L-methionine. The SAM concentration of the cultures expressing both enzymes seemed to be lower than expected. In C5 the FlA was induced at a D_{600} 4.5 meaning that the RLSS was the expressed for a longer time than the FLA. We expected higher SAM amount than we observed due to a better regeneration of SAM by the high levels of RLSS. We thought that the RLSS was probably inhibited by high intracellular SAM levels but this hypothesis was contradictory to the results of the culture expressing only the RLSS, which showed the highest SAM level and hence no inhibition of the RLSS. Additional experiments are necessary to clarify this issue, which cannot be satisfactorily explained by the results of this study. The SAM amount of C2 expressing both enzymes but where the FlA was induced earlier (D_{600} 0.9) seemed reasonable. It was low because the SAM was used up for the FDA production. It seemed that the FDA production happened faster than the SAM regeneration.

Summarizing the results of this experiment, it was clearly visible that the amount of FDA produced with and without the co-expression of the RLSS did not significantly differ and we concluded that we probably used an inappropriate amount of FlA in these reactions. For this reason it might not have been possible to see a pronounced difference in the FDA levels depending on the RLSS.

Therefore, we performed the assay using varying amounts of lyophilized cells for each reaction. The results are represented in Figure 13 and it was clearly visible that the amounts of FDA produced by lyophilized cells with and without the co-expression of the RLSS showed no significant difference up to 75 mg cells. We concluded that the amount of synthesized FDA directly correlated to the amount of used cells. Coexpression of the RLSS was beneficial for SAM recycling only at upper FlA levels. Further work might address the correlation between the SAM amounts and the produced FDA.

SAM production from L-methionine and ATP with and without the co-expression of RLSS is represented in Figure 14A. As described in the results section, the reactions which were not supplemented with ATP showed the same pattern as reactions with ATP supplementation but

the amounts of synthesized SAM were 10-fold lower. The production of SAM in this case was possible due to the residual intracellular ATP but its concentration was low and limiting.

Regarding the SAM amounts produced by the reactions, which were supplemented with ATP and L-methionine, the highest amounts of SAM were produced by cultures C4 and C5. This seemed reasonable since culture C4 expressed only the RLSS so the synthesized SAM was not used up for the production of the FDA and culture C5 expressed both enzymes but the FlA was induced at a late stage. The amounts of FDA produced during these reactions are shown in Figure 14B. It was clearly visible that the cultures C2 and C5, both expressing the FlA and RLSS, showed significantly higher FDA amounts compared to culture C3, expressing only the FlA. This result confirmed the hypothesis that an elevated SAM level enhanced the FDA production. We speculate that this effect, at least in part, could be provoked by the interactions between SAM and the FlA (Dong et al. 2004), thus increasing the activity of the FlA. This was clearly illustrated by the fact that C5, where the FlA was induced at higher D_{600}, yielded much more FDA than C2, where the FlA was induced at an earlier D_{600}.The concentration of RLSS in C5 was higher at the time point of the FlA induction than in C2, thus enhancing the SAM regeneration.

Summarizing the results of the assays, we demonstrated that the FlA was active and synthesized the desired product FDA. Since the majority of the FlA was insoluble, we hypothesize that the FlA in the soluble protein fraction might be hyperactive or the FlA in the insoluble protein fraction showed residual activity. Additionally, we showed that the co-expression of the RLSS slightly increased the amounts of produced FDA. However, the experiments of the FDA production using lyophilized cells represented initial trials and should be repeated in order to gain more reliable results. The assay reactions were measured only once and occasionally two different HPLC machines but the same column had to be used which increases the error. The integrated area of the desired peaks was correlated with the amount of the corresponding substances. It would be necessary to measure a calibration curve of each substance e.g. SAM and FDA in order to make sure that there is a linear correlation between the amount of the substance and the peak area. For this reasons the described results should be considered with care.

7. Conclusion

In this study, we showed that the applied approach of the codon optimization of the FlA for the expression in *E. coli* did not improve the expression levels and the solubility of this enzyme. Besides, we set up a modular cloning strategy for the assembly of a biosynthetic pathway for the production of FDA and SAM substrate recycling in *E. coli*. We produced all relevant destination vectors and almost all parts of the pathway and showed that the assembly reactions worked. However, the assembly efficiency of the part plasmids was dependent on the sequence of the insert. Future work in this project might address this correlation in more detail. We performed the first experiments with lyophilized cells and obtained first insights how the co-expression of the RLSS affected the expression levels and the solubility of the FlA as well as the production of FDA. We found that there was no significant increase in the expression of soluble FlA. The co-expression of the RLSS led to enhanced FDA production. Nevertheless, the conducted experiments represented initial trials and more experimental data is needed to draw more reliable conclusions.

8. References

Alvarez, L. et al., 1994. Expression of rat liver S-adenosylmethionine synthetase in Escherichia coli results in two active oligomeric forms. *The Biochemical journal*, 301 (Pt 2, pp.557–561.

Angov, E. et al., 2008. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS ONE*, 3(5), pp.1–10.

Angov, E., Legler, P.M. & Mease, R.M., 2011. Adjustment of Codon Usage Frequencies by Codon Harmonization Improves Protein Expression and Folding. In *Heterologous Gene Expression in E.coli; Methods and Protocols*. pp. 1–14.

Armstrong, J.A. & Schulz, J.R., 2008. Agarose gel electrophoresis. In *Current Protocols Essential Laboratory Techniques*. pp. 7.2.1–7.2.22.

Biava, H. & Budisa, N., 2014. Evolution of fluorinated enzymes: An emerging trend for biocatalyst stabilization. *Engineering in Life Sciences*, 14(4), pp.340–351.

Choudens, S.O. De et al., 2002. Reductive cleavage of S-adenosylmethionine by biotin synthase from Escherichia coli. *Journal of Biological Chemistry*, 277(16), pp.13449–13454.

Deng, H. et al., 2014. Identification of fluorinases from streptomyces sp MA37, norcardia brasiliensis, and actinoplanes sp N902-109 by genome mining. *ChemBioChem*, 15(3), pp.364–368.

Deng, H. et al., 2008. In Vitro Reconstituted Biotransformation of 4-Fluorothreonine from Fluoride Ion: Application of the Fluorinase. *Chemistry and Biology*, 15(12), pp.1268–1276. Available at: http://dx.doi.org/10.1016/j.chembiol.2008.10.012.

Deng, H. et al., 2006. The fluorinase from Streptomyces cattleya is also a chlorinase. *Angewandte Chemie - International Edition*, 45(5), pp.759–762.

Dong, C. et al., 2004. Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature*, 427(February), pp.561–565.

Esposito, D. & Chatterjee, D.K., 2006. Enhancement of soluble protein expression through the use of fusion tags. *Current Opinion in Biotechnology*, 17(4), pp.353–358.

Gibson, D.G., 2011. Enzymatic assembly of overlapping DNA fragments. *Methods in Enzymology*, 498, pp.349–361.

Goodman, D.B., Church, G.M. & Kosuri, S., 2013. Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science*, 342, pp.475–479. Available at: http://www.sciencemag.org/content/early/2013/09/25/science.1241934.abstract.

Gribble, G.W., 2004. Natural Organohalogens: A New Frontier for Medicinal Agents? *Journal of Chemical Education*, 81(10), pp.1441–1449.

Gurney, T., 1984. Preparation of Lyophilized Cells to Preserve Enzyme Activities and High Molecular Weight Nucleic Acids. In *Nucleic Acids. Methods in Molecular Biology*. pp. 35–42.

Gustafsson, C., Govindarajan, S. & Minshull, J., 2004. Codon bias and heterologous protein expression. *Trends in Biotechnology*, 22(7), pp.346–353.

Hagan, D.O. & Deng, H., 2015. Enzymatic Fluorination and Biotechnological Developments of the Fluorinase.

Huo, B.R. et al., 2009. Analysis and Comparison of β -galactosidase Production in Escherichia coli BW25993 / Δ lacI and MG1655 / Δ lacI Mutants During Conditions of Growth and Catabolite Repression. , 13(April), pp.79–82.

Koshanskaya, M., 2014. *Establish MoClo In JG SynBio – Level 0*,

Laemmli, 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 227, pp.680–685.

Lieber, C.S., 2002. S-Adenosyl- L -methionine : its role in the treatment of liver disorders. *American Journal of Clinical Nutrition*, 76(suppl), pp.1183–1187S.

Manavathu, E.K. & Thomas, D.D.S., 1982. The uptake of S-adenosyl-L-methionine in the aquatic fungus Achlya ambisexualis. *FEBS Letters*, 137(1), pp.14–18.

Marsh, E.N.G., 2014. Fluorinated Proteins: From Design and Synthesis to Structure and Stability. *Accounts of chemical research*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24883933.

Murphy, C.D., Schaffrath, C. & O'Hagan, D., 2003. Fluorinated natural products: The biosynthesis of fluoroacetate and 4-fluorothreonine in Streptomyces cattleya. *Chemosphere*, 52(2), pp.455–461.

O'Hagan, D., 2006. Recent developments on the fluorinase from Streptomyces cattleya. *Journal of Fluorine Chemistry*, 127(11), pp.1479–1483.

O'Hagan, D., 2008. Understanding organofluorine chemistry. An introduction to the C-F bond. *Chemical Society reviews*, 37(2), pp.308–319.

Odar, C., Winkler, M. & Wiltschi, B., 2015. Fluoro amino acids: A rarity in nature, yet a prospect for protein engineering. *Biotechnology Journal*, 201400587, p.n/a–n/a. Available at: http://doi.wiley.com/10.1002/biot.201400587.

Posnick, L.M. & Samson, L.D., 1999. Influence of S -Adenosylmethionine Pool Size on Spontaneous Mutation , Dam Methylation , and Cell Growth of Escherichia coli Influence of S -Adenosylmethionine Pool Size on Spontaneous Mutation , Dam Methylation , and Cell Growth of Escherichia coli Downl. , 181(21), pp.6756–6762.

Purser, S. et al., 2008. Fluorine in medicinal chemistry. *Chemical Society Reviews*, 37(2), pp.320–330. Available at: <Go to ISI>://WOS:000252411800007\nhttp://www.ncbi.nlm.nih.gov/pubmed/18197348.

Rosano, G.L. & Ceccarelli, E. a., 2014. Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*, 5(APR), pp.1–17.

Schaffrath, C., Deng, H. & O'Hagan, D., 2003. Isolation and characterisation of 5´-fluorodeoxyadenosine synthase, a fluorination enzyme from Streptomyces cattleya. *FEBS Letters*, 547(1-3), pp.111–114.

Stenström, C.M., Holmgren, E. & Isaksson, L. a, 2001. Cooperative effects by the initiation codon and its flanking regions on translation initiation. *Gene*, 273(2), pp.259–265.

Szybalski, W. et al., 1991. Class-IIS restriction enzymes - a review probes ; recognition and cleavage domains ; trimming vectors) Class-IRS restriction enzymes (ENases-IIS) interact with two discrete sites on double-stranded DNA : the recognition site , which is 4-7 bp long , an. *Gene*, 109(1), pp.13–26.

Villalobos, A. et al., 2006. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC bioinformatics*, 7, p.285.

Walker, M.C. et al., 2013. Expanding the fluorine chemistry of living systems using engineered pelyketide synthase pathways. *Science*, 341(6150), pp.1089–1094.

Wang, Y., Deng, Z. & Qu, X., 2014. Characterization of a SAM-dependent fluorinase from a latent biosynthetic pathway for fluoroacetate and 4-fluorothreonine formation in Nocardia brasiliensis. *F1000Research*, 3, p.61. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3999930&tool=pmcentrez&rendertype=abstract.

Weber, E. et al., 2011. A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE*, 6(2).

Wei, Y. & Newman, E.B., 2002. Studies on the role of the metK gene product of Escherichia coli K-12. *Molecular Microbiology*, 43(6), pp.1651–1656. Available at: http://doi.wiley.com/10.1046/j.1365-2958.2002.02856.x [Accessed June 17, 2015].

Wright, F. & Bibb, M.J., 1992. Codon usage in the G+C-rich Streptomyces genome. , 113, pp.55–65.

Zheng, L., Baumann, U. & Reymond, J.-L., 2004. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic acids research*, 32(14), p.e115.

9. Supporting material

9.1. Primer table

Table S1: List of primers used in this study. Homology regions are underlined, restriction sites are uppercase and italicized, fusion sites are bold and mutagenesis sites are uppercase and colored in red.

Primer	F	R	Description	Used for	Sequence
pBP1051	+		lacZ_gBlock	PCR	ct*GAAGAC*taaggcggcggagccgac
pBP1096			pMC0.1_ori	Sequencing	gggaaacgcctggtatctttatagtc
pBP1097			pMC0.1_λt0	Sequencing	gcaaccgagcgttctgaac
pBP1098			pMC0.1_CmRPr	Sequencing	actcaaaaaatacgcccggtagtg
pBP1110			pMC0.1_ori	Sequencing	gtagctcttgatccggcaaaca
pBP1130			lacZ	Sequencing	caatgcgggtcgcttcacttac
pBP1131			lacZ	Sequencing	gatcgacagatttgatccagcgatac
pBP1132			lacZ	Sequencing	catcgataatttcaccgccgaaag
pBP1133	+		AmpR+pr level 1	PCR	catgca*CGTCTCa***gctt**ttaccaatgcttaatcagtgaggcacc
pBP1134		+	AmpR+pr level 1	PCR	tgaaca*CGTCTCt***gttc**cgcggaacccctatttgtttatttttctaaatac
pBP1135	+		lacZ TU level 1	PCR	catgca*CGTCTCa***gaac**ct*GTCTTC*gagctgttgacaattaatc
pBP1136		+	lacZ TU level 1	PCR	tgaaca*CGTCTCt***gcga**ta*GTCTTC*atttgtcctactcaggagagcg
pBP1137	+		lacZ1	oePCR	gcggcattttccgtgacgt**G**tc**C**ttgctgcataaaccgactacacaaatc

pBP1138	+	lacZ1	oePCR	gatttgtgtagtcggtttatgcagcaa**G**ga**C**acgtcacggaaaatgccgc
pBP1139	+	λt0 ter level 1	PCR	catgca*CGTCTC*a**tact**accaataaaaaacgcccg
pBP1140	+	λt0 ter level 1	PCR	tgaaca*CGTCTC*t**aagc**gactcctgttgatagatcc
pBP1141	+	pMBori level 1	PCR	catgca*CGTCTC*a**tcgc**tttccataggctccgc
pBP1142	+	pMBori level 1	PCR	tgaaca*CGTCTC*t**agta**ttgagatccttttttctgcgcg
pBP1143	+	flA_codopt	Gibson	caattataatagattcaattgtgagcggataacaatttcacacagaattcattaaagaggagaaattaactatggcggctaactccacac
pBP1144	+	flA_codopt	Gibson	ctggatctatcaacaggagtccaagctcagctaattaagctttcaacgagcttcaacacg
pBP1146	+	KanR+pr level 1	PCR	catgca*CGTCTC*a**gctt**ttagaaaaactcatcgagcatcaaatgaaactg
pBP1147	+	KanR+pr level 1	PCR	tgaaca*CGTCTC*t**gttc**gaagatcctttgatcttttctacgggtc
pBP1148	+	KanR+pr level 1	oePCR	gcctgagcga**AG**cgaaatacgcgatcgctgttaaaaggacaattacaaac
pBP1149	+	KanR+pr level 1	oePCR	gcgtatttcg**CT**tcgctcaggcgcaatcacgaatgaataacggtttg
pBP1150	+	Col1E ori level 1	PCR	catgca*CGTCTC*a**tcgc**ggccgcgttgctggcgtttttc
pBP1151	+	Col1E ori level 1	PCR	tgaaca*CGTCTC*t**agta**tcatgaccaaaatcccttaacgtgagttttcgttccactgag
pBP1158	+	Col1E ori level 2	PCR	ct*GAAGAC*ta**tcgc**ggccgcgttgctggcgtttttc

pBP1159	+	Col1E ori level 2	PCR	tgaaca*GAAGAC*ta**agta**tcatgaccaaaatcccttaac gtgagttttcgttccactgagcg
pBP1160	+	KanR+pr level 2	PCR	catgca*GAAGAC*ta**gctt**ttagaaaaactcatcgagc
pBP1161	+	KanR+pr level 2	PCR	tgaaca*GAAGAC*ta**gttc**gaagatcctttgatcttttctac
pBP1152	+	aS1_lacI	PCR	catgca*GGTCTC*t**gaac**ta*GAAGAC*at**tact**attctc accaataaaaaacgcc
pBP1153	+	aS1_lacI	PCR	tgaaca*GGTCTC*g**atac**cgcagatagctcatg
pBP1154	+	a23_lacI	PCR	catgca*GGTCTC*c**gtat**cgtcgtatcccac
pBP1155	+	a23_lacI	PCR	tgaaca*GGTCTC*c**ggac**ggtacgcgactg
pBP1156	+	a4E_lacI	PCR	catgca*GGTCTC*t**gtcc**tcatgggagaaaataatac
pBP1157	+	a4E_lacI	PCR	tgaaca*GGTCTC*c**gcga**ct*GAAGAC*at**aagc**tgtgt gaaattgttatccg
pBP1183	+	aSE_araC_pBAD	PCR	catgca*GGTCTC*t**gaac**ta*GAAGAC*at**tact**ttatga caacttgacggctac
pBP1184	+	aSE_araC_pBAD	PCR	tgaaca*GGTCTC*c**gcga**ct*GAAGAC*at**aagc**aaa aaaacgggtatggagaaac
pBP1185	+	b23_FLAG	oePCR	catgca*GGTCTC*a**gaac**at*GAAGAC*tc**caat**gggt gactacaaggacgacgatgac
pBP1186	+	b23_FLAG	oePCR	tgaaca*GGTCTC*a**gcga**ct*GAAGAC*at**catt**gaac ctttgtcatcgtcgtccttgtagtcac
pBP1187	+	b23_GST	PCR	catgca*GGTCTC*a**gaac**at*GAAGAC*tc**caat**gtccc ctatactaggttattg
pBP1188	+	b23_GST	PCR	tgaaca*GGTCTC*a**gcga**ct*GAAGAC*at**catt**gaac

				cttttggaggatggtcgcc
pBP1189	+	b23_MBP	PCR	catgca*GGTCTC*a**gaac**at*GAAGAC*tc**caat**gaaa atcgaagaaggtaaactggtaatctggattaacggcgataa aggctataacggac
pBP1190	+	b23_MBP	PCR	tgaaca*GGTCTC*a**gcga**ct*GAAGAC*at**catt**gagc cacttcctgaaccattagtctgcgcgtctttc
pBP1191	+	bS1_RBSIII	oePCR	catgca*GGTCTC*a**gaac**at*GAAGAC*tc**gctt**aatac tgttcggacactcagaaggctttatattaaagaggag
pBP1192	+	bS1_RBSIII	oePCR	tgaaca*GGTCTC*a**gcga**ct*GAAGAC*at**attg**ttaatt tctcctctttaatataaagccttctgag
pBP1193	+	bS1_RBSIV	oePCR	catgca*GGTCTC*a**gaac**at*GAAGAC*tc**gctt**ttctta cgaatcaacttattggagcggaacaagataag
pBP1194	+	bS1_RBSIV	oePCR	tgaaca*GGTCTC*a**gcga**ct*GAAGAC*at**attg**tatat ctccttcttatcttgttccgctccaataag
pBP1195	+	bS3_RBSIII	oePCR	catgca*GGTCTC*a**gaac**at*GAAGAC*tc**gctt**aatac tgttcggacactcagaaggctttatattaaagagg
pBP1196	+	bS3_RBSIII	oePCR	tgaaca*GGTCTC*a**gcga**ct*GAAGAC*at**catt**gttaa tttctcctctttaatataaagccttctgagtg
pBP1197	+	bS3_RBSIV	oePCR	catgca*GGTCTC*a**gaac**at*GAAGAC*tc**gctt**ttctta cgaatcaacttattggagcggaacaagataag
pBP1198	+	bS3_RBSIV	oePCR	tgaaca*GGTCTC*a**gcga**ct*GAAGAC*atc**att**gtatat ctccttcttatcttgttccgctccaataag
pBP1199	+	cSE_rrnBT1	PCR	catgca*GGTCTC*t**gaac**ta*GAAGAC*at**ggag**aggc atcaaataaaacgaaagg
pBP1200	+	cSE_rrnBT1	PCR	tgaaca*GGTCTC*c**gcga**ct*GAAGAC*at**gcct**ggcg

				gatttgtcctactc
pBP1201	+	cSE_T7	oePCR	catgca*GGTCTC*t**gaac**ta*GAAGAC*at**ggag**agca taacccttggggcctctaaacgggtc
pBP1202	+	cSE_T7	oePCR	tgaaca*GGTCTC*c**gcga**ct*GAAGAC*at**gcct**caaa aaacccctcaagacccgtttagaggccc
pBP1203	+	b4E_flAnotOpt1	PCR	ct*GGTCTC*t**gaac**ta*GAAGAC*at**aatg**gctgcgaac agc
pBP1204	+	b4E_flAnotOpt1	PCR	ca*GGTCTC*c**gaac**accgttccctcgg
pBP1205	+	b4E_flAnotOpt2	PCR	ct*GGTCTC*t**gttc**gccaccaccac
pBP1206	+	b4E_flAnotOpt2	PCR	ca*GGTCTC*t**acac**cacgcccacc
pBP1207	+	b4E_flAnotOpt3	PCR	ct*GGTCTC*g**gtgt**ccgccatcgaccacccgttcg
pBP1208	+	b4E_flAnotOpt3	PCR	tc*GGTCTC*c**gcga**ct*GAAGAC*at**ctcc**tcagcggg cctcg
pBP1209	+	lacZ TU level 2	PCR	ct*GAAGAC*ta**gaac**t*GAGACC*gagctgttgacaatta atcatcg
pBP1210	+	lacZ TU level 2	PCR	tc*GAAGAC*at**gcga**c*GAGACC*<u>atttgtcctactcagg agag</u>
pBP1246	+	AMpR+pr level 2	oePCR	ccgcg**T**gacccacgctcaccg
pBP1247	+	AMpR+pr level 2	oePCR	gtgggtc**A**cgcggtatcattgcagcactg
pBP1248	+	AMpR+pr level 2	PCR	catgca*GAAGAC*ta**gctt**ttaccaatgcttaatcagtgag g
pBP1249	+	AMpR+pr level 2	PCR	tgaaca*GAAGAC*ta**gttc**cgcggaacccctatttg
pBP1250	+	aS1_lacI	PCR	tgaaca*GGTCTC*g**atac**cgcagatagctcatgttatatc

				cc
pBP1251	+	a4E_lacI	PCR	catgca*GGTCTC*t**gtcc**tcatgggagaaaataatactgttgatgggtg
pBP1252	+	a4E_lacI	PCR	tgaaca*GGTCTC*c**gcga**ct*GAAGAC*ataagctgtgtgaaattgttatccgctcacaattgaatc
pBP1288	+	lacI 3	PCR	tgaaca*GGTCTC*c**gtaa**gttggacaccatcgaatggtgcaaaacctttc
pBP1289	+	T5+2xlacO	oePCR	catgca*GGTCTC*a**ttac**gagtgataaatcataaaaaattatttgctttgtgagcggataacaattataatagattcaattg
pBP1290	+	T5+2xlacO	oePCR	tgaaca*GGTCTC*c**gcga**ct*GAAGAC*ataagctgtgtgaaattgttatccgctcacaattgaatctattataattgttatccg
pBP1327	+	flA_codopt	QuikChange	ctatggc**C**gctaactccacacg**G**cgtcc**C**attattgcatttatgtctgacctgggtacc
pBP1328	+	flA_codopt	QuikChange	caataat**G**ggacg**C**cgtgtggagttagc**G**gccatagttaatttctcctctttaatgaattctgtgtgaaattgttatccgctcac
pBP1340	+	RLSS	Gibson	gcaactctctactgtttctccatacccgttttttt*GGTACC*ggaaaaaggagatctgcatatgaatggccctgttgacgg
pBP1341	+	RLSS	Gibson	ggcaaattctgttttatcagaccgcttctgcgttctgatttaatc*TTAATTAA*ctaaaaaaccagtttcttcggcacttc
pBP1363	+	connectors	PCR	catgca*CGTCTC*t**gaact**a*GAAGAC*
pBP1364	+	connectors	PCR	tgaaca*CGTCTC*a**gcga**ctgaag

9.2. Vector maps

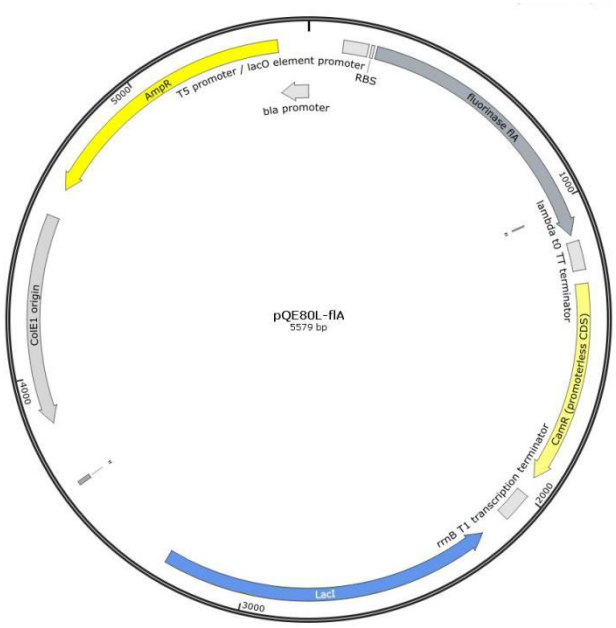

Figure S1: Overview of the vector pQE80L-FIA used for the overexpression of the FIA variants (not optimized and codon optimized).
Fluorinase was expressed under the control of IPTG-inducible T5/*lacO* promoter. Other important features are the high copy number origin of replication ColE1 and ampicillin resistance marker.

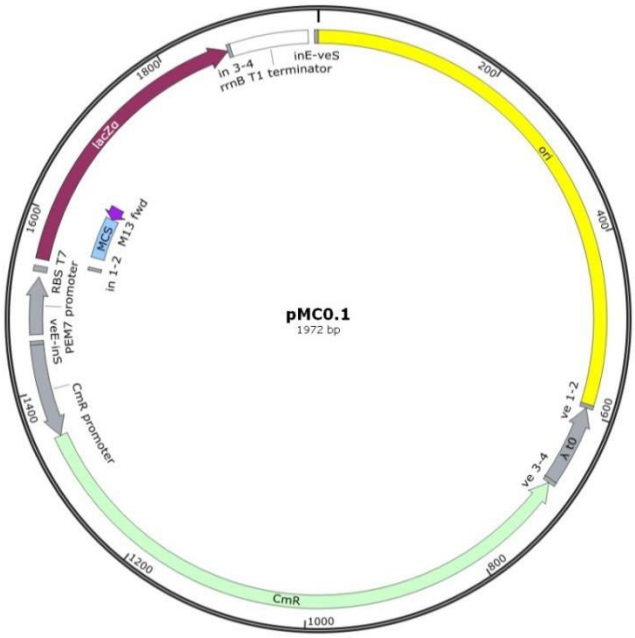

Figure S2: Overview of the vector pMC0.1 used for the compilation of the part plasmids containing connectors.
LacZα was expressed under the control of constitutive EM7 promoter. Other important features are the high copy number origin of replication pMB1 and chloramphenicol resistance marker.

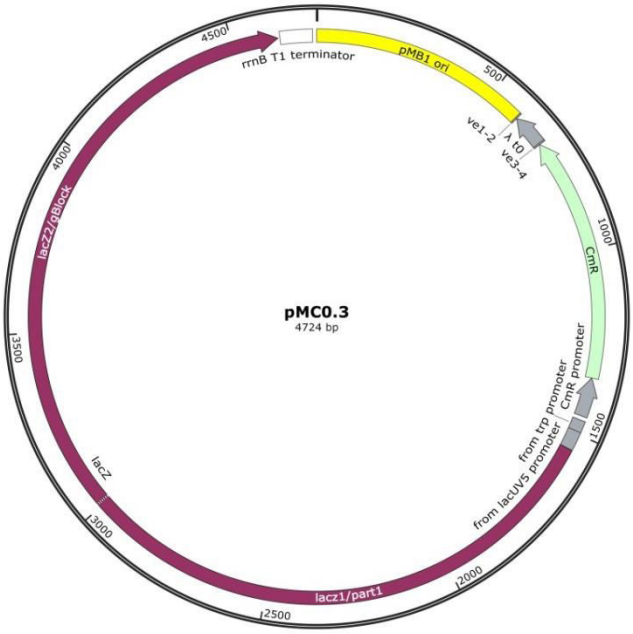

Figure S3: Overview of the vector pMC0.3 used for the compilation of the part plasmids library.
LacZ was expressed under the control of inducible TacII promoter. Remaining features are the same as of pMC0.1.

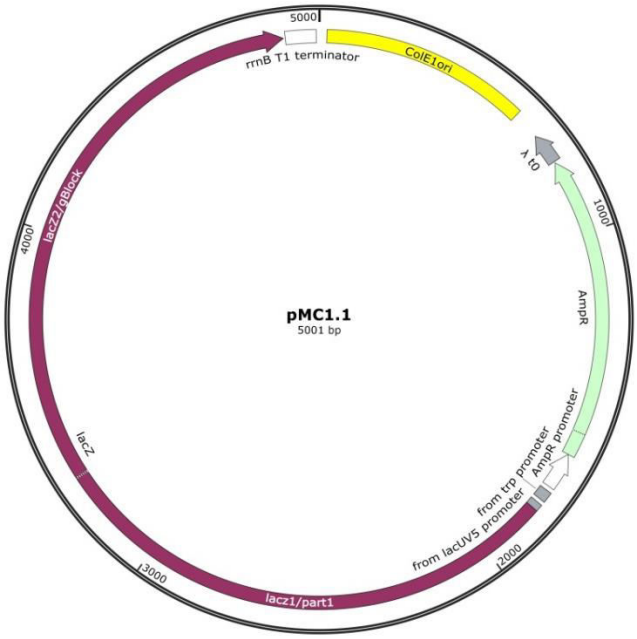

Figure S4: Overview of the vector pMC1.1 which was designed and successfully assembled by modular cloning.
LacZ was expressed under the control of inducible TacII promoter. Other important features are the high copy number origin of replication ColE1 and ampicillin resistance marker.

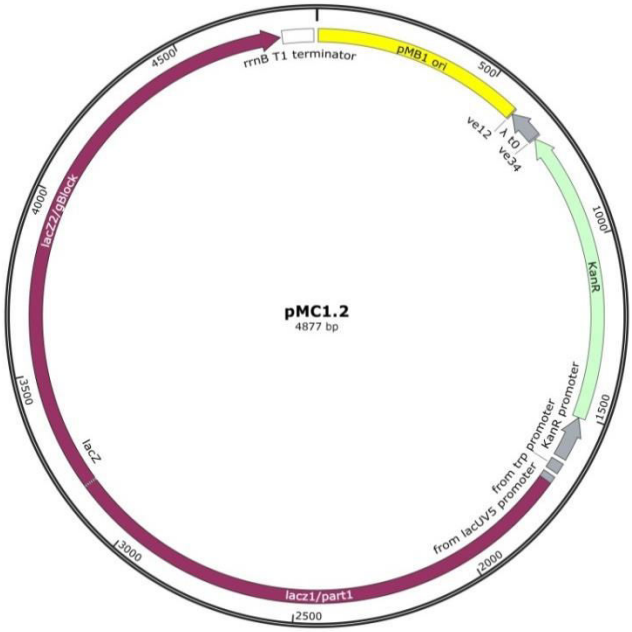

Figure S5: Overview of the vector pMC1.2 which was designed for MoClo.
The assembly did not succeed. Different to pMC1.1 this vector featured a high copy number origin of replication pMB1 and a kanamycin resistance marker.

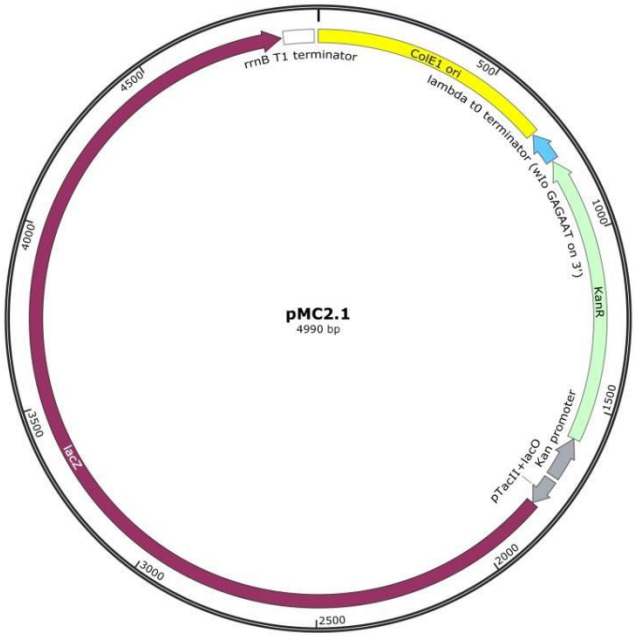

Figure S6: Overview of the vector pMC2.1 which was designed for MoClo.
The assembly did not succeed. Important features are the high copy number origin of replication ColE1 and kanamycin resistance marker.

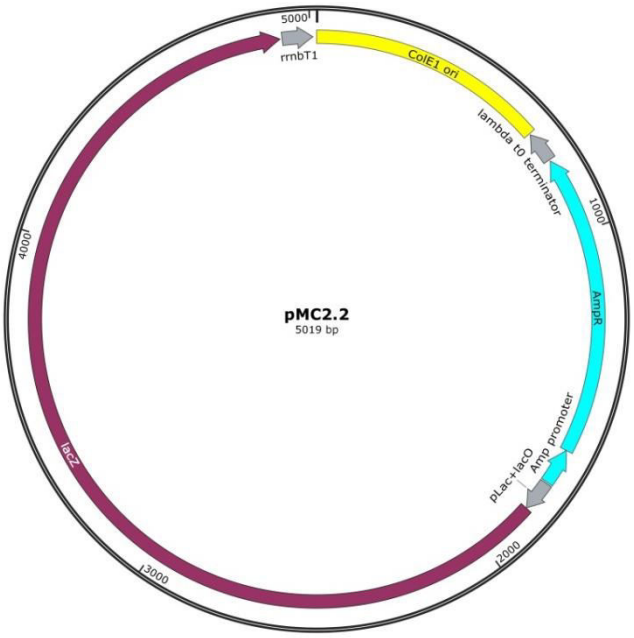

Figure S7: Overview of the vector pMC2.2 which was designed and successfully assembled for MoClo.
Different to pMC2.1 this vector featured an ampicillin resistance marker.

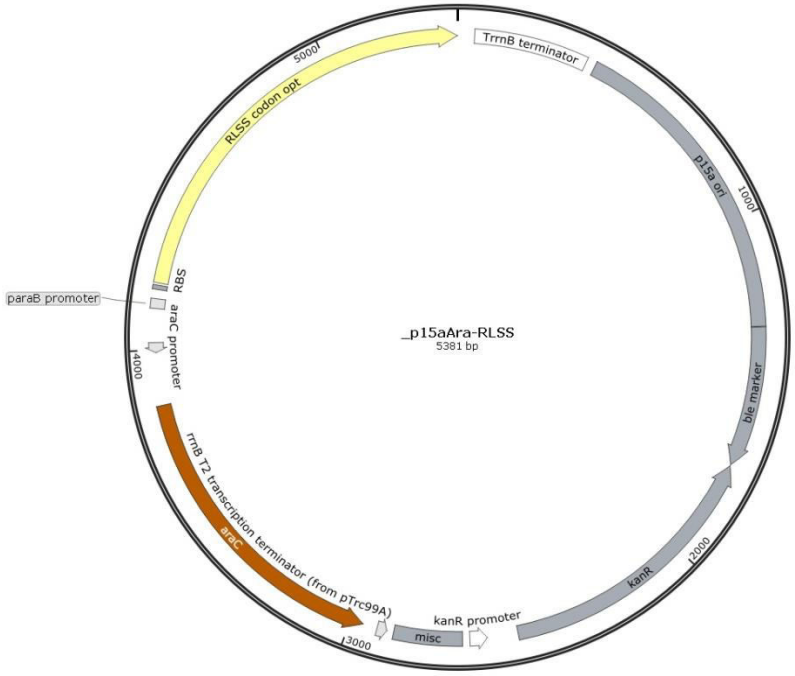

Figure S 8: Overview of vector p15aAra used for the overexpression of RLSS.
RLSS was expressed under the control of the P_{BAD} promoter. Other important features are the medium copy number origin of replication p15a and kanamycin resistance marker

9.3. Subparts of the destination vectors and parts for the assembly of a biosynthetic pathway

Table S2: Overview of the subparts of the level 1 destination vectors.
TU: transcriptional unit.

pMC1.1			pMC1.2		
Subpart	Template	Location [bp]	**Subpart**	Template	Location[4] [bp]
ColE1 ori	pQE80L	5..687	pMB1 ori	pUC18	5..593
λt_0 terminator	pSYN53	692..780	λt_0 terminator	pSYN53	598..686
AmpR+promoter	pUC18	785..1750	KanR+promoter	pET28_flA_his	691..1627
lacZ TU	pMC0.3	1757..4999	lacZ TU	pMC0.3	1634..4875

Table S3: Overview of the subparts of the level 2 destination vectors.
TU: transcriptional unit.

pMC2.1			pMC2.2		
Subpart	Template	Location [bp]	**Subpart**	Template	Location [bp]
ColE1 ori	pQE80L	1..683	ColE1 ori	pQE80L	1..683
λt_0 terminator	pSYN53	688..776	λt_0 terminator	pSYN53	688..776
KanR+promoter	pET28_flA_his	781..1717	AmpR+promoter	pUC18	781..1746
lacZ TU	pMC0.3	1723..4986	lacZ TU	pMC0.3	1752..5015

[4] Referred to the location on the resulting plasmid

Table S4: Overview of the parts for the modular assembly.
The parts FIA and lacI were divided in three and four subparts (not indicated here). oePCR, overlap extension PCR using two overlapping primers. Parts indicated with a star * could not be assembled.

	Fusion sites	Part	Template	Assembly of part plasmids led to:	
				Total number	White clones
1	aSE	araC_pBAD	pQEara_gbd(woscori)	10	10
2	aSE	lacI_PT5_lacO*	pQE80L	4	4
3	b4E	FIA*	pQE80L-his-FIA	4	4
4	b4E	RLSS*	gBlock	-	-
5	b23	FLAG	produced by oePCR	6	2
6	b23	GST	pYEX_EV	90	87
7	b23	MBP	pMAL_flB	6	6
8	bS1	RBSIII	produced by oePCR	6	6
9	bS1	RBSIV	produced by oePCR	11	4
10	bS3	RBSIII	produced by oePCR	2	1
11	bS3	RBSIV*	produced by oePCR	33	6
12	cSE	rrnB T1	pSYN53	17	16
13	cSE	T7	produced by oePCR	11	11

9.4. Instruments and chemicals

Table S5: List of instruments and devices used in this study.

Instrument	Supplier
Analytical scale	Sartorius; Göttingen, Germany
Avanti J-20 XP centrifuge	Beckmann Coulter Inc.; Brea, CA
Centrifuge tubes, 500 mL	Thermo Scientific Inc.; Waltham, MA
Centrifuges	Centrifuge 5415R: Eppendorf; Hamburg, Germany Centrifuge 5424: Eppendorf; Centrifuge 5810 R; Eppendorf
Sub-cell® GT Cell and PowerPac Basic Power Supply	BioRad, Hercules, USA
BioRad Micropulser	BioRad
Electroporation cuvettes	Peqlab Biotechnology GmbH, Erlangen, Germany
Kits	CloneJET PCR cloning Kit from Thermo Scientific GeneJET Plasmid Miniprep Kit form Thermo Scientific Wizard® SV Gel and PCR Clean-Up System from Promega
Eppendorf tubes	Sarstedt, Nümbrecht, Germany
Flasks	2000 mL (Schott Duran); Bartelt; Graz, Austria 1000 mL (SIMAX), Bartelt 250 mL (Schott Duran); Bartelt
ddH_20 device (arium®basic)	Sartorius

Incubator HT Multitron II	InforsAG, Bottmingen, Switzerland
Lab scale	Binder GmbH, Tuttlingen, Germany
Laminar air flow hood	AirClean, Woerden, The Netherlands
NanoDrop	Thermo Scientific
PCR machines (GeneAmp®PCR System 2700)	Applied Biosystems, California, USA
PCR tubes	Greiner bio-one International AG
Petri dishes	Greiner Bio-one International AG; Kremsmünster, Austria
Photometer	Beckman Clouter Inc. BioPhotometer; Eppendorf
Pipette tips	Greiner Bio-one International AG
Pipettes	1000 µL, 200 µL, 20 µL; Denville; South Plainfield, NJ 10 µL (Biohit); Sartorius
Scanner	Tevion USB Scanner; Mühlheim, Germany
Sterile syringe filters	Scientific Strategies; Yukon, OK
Sterile filters	Sartorius
Vortex	IKA®-Werke GembH & Co. K; Staufen, Germany

Table S6: List of chemicals and enzymes used for this study.

Reagent	Cat. Nr.	Supplier
Acetonitrile	A998-212	Fisher Chemical
Agarose LE	840004	Biozyme,Hessisch-Oldendorf, Germany
Albumin fraction V (BSA)	T844.2	Roth
Ampicillin	A0166	Sigma-Aldrich, St.Louis, USA
Ammonium persulfate (APS)	13375.01	Serva, Heidelberg, Germany
Acrylamide/Bis solution, 37.5:1 (30 % w/v)	10688.01	Serva
Arabinose	5118.1	Roth
Comassie Blue-250R	3862.2	Roth
Chloramphenicol	J01BA01	Roth
dNTPs	R0181	Thermo Scientific
1,4-dithiothreit (DTT)	6908.1	Roth
Ethanol	20821.330	VWR International, Pennsylvania, USA
ethidium bromide	46066	Fluka, St. Louis, USA
***Ex Taq*® DNA Polymerase**	RR001A	TaKaRa
α-D-glucose monohydrate	6780.2	Roth
Glycerol	3908.3	Roth
Hydrochloric acid 32%	4625.2	Roth
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	CN03.3	Roth

Kanamycin sulfate	T832.2	Roth
LB-Agar Lennox	X65.3	Roth
LB-medium Lennox	X964.2	Roth
Magnesium chloride	8.14733.0500	Merck
Magnesium sulfate heptahydrate	A537.4	Roth
β-mercaptoethanol	4227.1	Roth
Methionine	1702.1	Roth
PageRuler prestained protein ladder	SM0671	Thermo Scientific
PEG-8000	P5413-1KG	Sigma-Aldrich
Phusion® High-Fidelity DNA polymerase	M0530S	NEB, Frankfurt am Main, Germany
Potassium chloride (KCl)	60128	Fluka
Di-potassium hydrogen phosphate	T875.2	Roth
Potassium dihydrogen phosphate	P018.2	Carl Roth, Germany
Restriction enzymes (KpnI, NdeI, HindIII, PacI, DpnI)	-	all Thermo Scientific, Fast Digest
SOB medium	AE27.1	Roth
Sodium chloride (NaCl)	9265.1	Roth
Di-sodium hydrogen phosphate	T876.2	Roth
Sodium dihydrogen phosphate monohydrate	T879.2	Roth
Sodium dodecyl sulfate (SDS)	2326.1	Roth
Sodium hydroxid	P031.2	Roth

Sterile water	-	Fresenius Kabi, Graz, Austria
Tetramethylethylendiamin (TEMED)	35930.01	Serva
Triton-X100	3051.3	Roth
Tris	4855.3	Roth
Type IIS restriction enzymes (Eco31I/BsaI; BpiI; Esp3I/BsmBI)	-	Thermo Scientific/NEB

Table S7: Composition of media and buffers used in this study.

Name	Composition
LB (Lennox)	10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl
LB agar	10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 20 g/L agar
SOC medium	5 g/L yeast extract, 20 g/L tryptone, 0.6 g/L NaCl, 0.2 g/L KCL, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose
TSS buffer	20 g/L LB medium, 100 mL/L PEG-3350, 50 mL/L DMSO, 2 mM $MgCl_2$*$6H_2O$
10 X SDS running buffer	1% (w/v) SDS, 1.92 M glycine, 0.25 M Tris
TAE buffer	40mM Tris, 20mM acetic acid, and 1mM EDTA
5x KCM	500 mM KCl, 150 mM $CaCl_2$, 250 mM $MgCl_2$
SDS PAGE loading dye	200 mM Tris/HCl pH 6.8; 40% glycerol; 8% SDS; 400mM DTT; 0,4% bromphenolblue
Gibson assembly master mix	For 1.2 mL: 1 X ISO reaction buffer, 6.4 U T4 exonuclease, 20 U Phusion® High-Fidelity DNA Polymerase, 6.4 kU Taq DNA ligase. :

Table S8: The components of the self-made SDS PAGE stacking (4%) and separating (12%) gels.

Separating gel		Stacking gel	
30% Acrylamid/Bis	2.01 mL	**30% Acrylamid/Bis**	325 µL
1.5 M Tris/HCl pH 8.8 +0.4 SDS	1.25 mL	**0.5 M Tris/HCl pH 8.8 +0.4 SDS**	500 µL
ddH_2O	1.72 mL	**ddH_2O**	1.185 mL
10 % APS	25 µL	**10 % APS**	10 µL
TEMED	2.5 µL	**TEMED**	2 µL

9.5. Gene sequences

LacZ (ordered as gBlock®):

ctGAAGACtaAGGCGGCGGAGCCGACACCACGGCCACCGATATTATTTGCCCGATGTACGCGCGCGTGGATGAAGATCAGCCCTTC
CCGGCTGTGCCGAAATGGTCCATCAAAAAATGGCTTTCGCTACCTGGAGAGACTCGCCCGCTGATCCTTTGCGAATACGCCCACGC
GATGGGTAACAGTCTTGGCGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCCCGTTTACAGGGCGGCTTCGTCTGGGACT
GGGTGGATCAGTCGCTGATTAAATATGATGAAAACGGCAACCCGTGGTCGGCTTACGGCGGTGATTTTGGCGATACGCCGAACGAT
CGCCAGTTCTGTATGAACGGTCTCGTCTTTGCCGACCGCACGCCGCATCCAGCGCTGACGGAAGCAAAACACCAGCAGCAGTTTTT
CCAGTTCCGTTTATCCGGGCAAACCATCGAAGTGACCAGCGAATACCTGTTCCGTCATAGCGATAACGAGCTCCTGCACTGGATGG
TGGCGCTGGATGGTAAGCCGCTGGCAAGCGGTGAAGTGCCTCTGGATGTCGCTCCACAAGGTAAACAGTTGATTGAACTGCCTGAA
CTACCGCAGCCGGAGAGCGCCGGGCAACTCTGGCTCACAGTACGCGTAGTGCAACCGAACGCGACCGCATGGTCAGAAGCCGGGCA
CATCAGCGCCTGGCAGCAGTGGCGTCTGGCGGAAAACCTCAGTGTGACGCTCCCCGCCGCGTCCCACGCCATCCCGCATCTGACCA
CCAGCGAAATGGATTTTTGCATCGAGCTGGGTAATAAGCGTTGGCAATTTAACCGCCAGTCAGGCTTTCTTTCACAGATGTGGATT
GGCGATAAAAAACAACTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGCACCGCTGGATAACGACATTGGCGTAAGTGAAGCGAC
CCGCATTGACCCTAACGCCTGGGTCGAACGCTGGAAAGCTGCAGGCCATTACCAGGCCGAAGCAGCGTTGTTGCAGTGCACGGCAG
ATACACTTGCTGATGCGGTGCTGATTACGACCGCTCACGCGTGGCAGCATCAGGGGAAAACCTTATTTATCAGCCGGAAAACCTAC
CGGATTGATGGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCGAGCGATACACCGCATCCGGCGCGGATTGGCCTGAA
CTGCCAGCTGGCGCAGGTAGCAGAGCGGGTAAACTGGCTCGGATTAGGGCCGCAAGAAAACTATCCCGACCGCCTTACTGCCGCCT
GTTTTGACCGCTGGGATCTGCCATTGTCAGACATGTATACCCCGTACGTCTTCCCGAGCGAAAACGGTCTGCGCTGCGGGACGCGC
GAATTGAATTATGGCCCACACCAGTGGCGCGGCGACTTCCAGTTCAACATCAGCCGCTACAGTCAACAGCAACTGATGGAAACCAG
CCATCGCCATCTGCTGCACGCGGAAGAAGGCACATGGCTGAATATCGACGGTTTCCATATGGGGATTGGTGGCGACGACTCCTGGA
GCCCGTCAGTATCGGCGGAATTCCAGCTGAGCGCCGGTCGCTACCATTACCAGTTGGTCTGGTGTCAAAAATAAAGGCATCAAATA
AAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATGGTCT
CGTCGCATGTCTTCGA

FIA c.o1 (ordered as gBlock®):

ctGGTCTCtGAACtaGAAGACatAATGGCGGCTAACTCCACACGCCGTCCGATTATTGCATTTATGTCTGACCTGGGTACCACTGA
TGACTCTGTAGCTCAATGCAAAGGTCTGATGTACAGCATCTGTCCGGACGTAACCGTCGTTGATGTTTGCCACTCTATGACCCCGT
GGGACGTGGAAGAAGGTGCCCGCTACATCGTTGATTTGCCGCGTTTCTTTCCGGAAGGTACCGTTTTTGCGACCACCACTTATCCA
GCAACTGGTACTACCACTCGTTCTGTGGCGGTTCGTATCAAACAGGCCGCGAAGGGTGGAGCACGCGGCCAGTGGGCTGGCTCCGG
TGCTGGTTTCGAACGCGCTGAAGGTAGCTACATTTACATCGCACCAAACAACGGCCTGCTGACCACCGTTCTGGAGGAACATGGTT
ATCTGGAAGCATATGAAGTAACTTCCCCGAAAGTGATCCCGGAACAGCCGGAGCCGACCTTCTACTCCCGTGAAATGGTAGCAATT
CCGAGCGCGCACCTGGCGGCGGGTTTCCCGCTGTCTGAGGTTGGTCGTCCGCTGGAAGATCACGAGATTGTACGTTTTAACCGTCC
GGCGGTCGAGCAGGATGGTGAAGCTCTGGTAGGTGTTGTATCTGCAATCGATCACCCGTTCGGTAACGTTTGGACCAACATCCATC
GTACCGACCTGGAGAAAGCGGGTATCGGTTACGGTGCGCGCCTGCGTCTGACCCTGGATGGTGTGCTGCCATTCGAGGCTCCGCTG
ACTCCGACCTTTGCAGATGCTGGCGAGATCGGTAACATCGCTATCTATCTGAACTCTCGCGGTTACCTGTCTATTGCTCGTAACGC
CGCAAGCCTGGCTTACCCGTACCATCTGAAGGAAGGTATGTCTGCACGTGTTGAAGCTCGTTGAGGAGatGTCTTCagTCGCgGAG
ACCga

FIA c.o2 (produced by QuikChange PCR):

ATGGCcGCTAACTCCACACGgCGTCCcATTATTGCATTTATGTCTGACCTGGGTACCACTGATGACTCTGTAGCTCAATGCAAAGG
TCTGATGTACAGCATCTGTCCGGACGTAACCGTCGTTGATGTTTGCCACTCTATGACCCCGTGGGACGTGGAAGAAGGTGCCCGCT
ACATCGTTGATTTGCCGCGTTTCTTTCCGGAAGGTACCGTTTTTGCGACCACCACTTATCCAGCAACTGGTACTACCACTCGTTCT
GTGGCGGTTCGTATCAAACAGGCCGCGAAGGGTGGAGCACGCGGCCAGTGGGCTGGCTCCGGTGCTGGTTTCGAACGCGCTGAAGG
TAGCTACATTTACATCGCACCAAACAACGGCCTGCTGACCACCGTTCTGGAGGAACATGGTTATCTGGAAGCATATGAAGTAACTT
CCCCGAAAGTGATCCCGGAACAGCCGGAGCCGACCTTCTACTCCCGTGAAATGGTAGCAATTCCGAGCGCGCACCTGGCGGCGGGT
TTCCCGCTGTCTGAGGTTGGTCGTCCGCTGGAAGATCACGAGATTGTACGTTTTAACCGTCCGGCGGTCGAGCAGGATGGTGAAGC
TCTGGTAGGTGTTGTATCTGCAATCGATCACCCGTTCGGTAACGTTTGGACCAACATCCATCGTACCGACCTGGAGAAAGCGGGTA
TCGGTTACGGTGCGCGCCTGCGTCTGACCCTGGATGGTGTGCTGCCATTCGAGGCTCCGCTGACTCCGACCTTTGCAGATGCTGGC
GAGATCGGTAACATCGCTATCTATCTGAACTCTCGCGGTTACCTGTCTATTGCTCGTAACGCCGCAAGCCTGGCTTACCCGTACCA
TCTGAAGGAAGGTATGTCTGCACGTGTTGAAGCTCGTTGA

RLSS (ordered as gBlock®):

ATGAATGGCCCTGTTGACGGTCTGTGTGACCACTCTCTGTCTGAAGAGGGTGCGTTCATGTTCACCTCTGAGTCCGTTGGCGAAGG
TCATCCTGATAAGATTTGTGACCAAATCAGCGATGCTGTCCTGGACGCCCACCTGAAGCAGGATCCGAACGCAAAAGTGGCATGCG
AAACTGTATGCAAAACCGGCATGGTTCTGCTGTGTGGCGAAATCACCTCTATGGCTATGATCGATTACCAGCGCGTGGTTCGCGAC
ACTATTAAACACATTGGCTACGATGACAGCGCAAAAGGCTTTGATTTCAAAACCTGTAACGTTCTGGTCGCCCTGGAACAGCAATC
CCCGGACATTGCGCAGTGCGTGCACCTGGATCGTAACGAAGAGGACGTAGGTGCGGGCGATCAAGGCCTGATGTTCGGTTATGCTA
CTGACGAAACCGAGGAATGCATGCCGCTGACCATCGTGCTGGCACACAAACTGAACACTCGTATGGCGGACCTGCGCCGTTCCGGT
GTTCTGCCATGGCTGCGTCCGGATAGCAAGACCCAGGTCACTGTTCAGTACGTTCAGGACAACGGTGCTGTTATCCCGGTACGTGT
ACACACCATTGTCATCTCCGTACAGCACAACGAAGATATCACTCTGGAAGCTATGCGCGAGGCTCTGAAAGAACAGGTTATCAAAG
CTGTTGTGCCAGCTAAATACCTGGATGAAGATACCATTTATCACCTGCAGCCTAGCGGTCGCTTCGTTATTGGTGGTCCGCAGGGT
GACGCGGGTGTAACCGGTCGCAAGATTATCGTGGATACTTACGGCGGCTGGGGTGCACATGGCGGTGGTGCTTTCTCTGGTAAAGA
CTACACTAAGGTTGATCGTAGCGCCGCGTATGCCGCGCGTTGGGTTGCTAAGTCTCTGGTCAAAGCAGGTCTGTGTCGTCGTGTTC
TGGTGCAGGTTTCCTACGCTATTGGCGTTGCAGAACCTCTGTCCATCTCTATCTTCACGTACGGTACTAGCAAAAAGACCGAGCGC
GAGCTGCTGGAGGTGGTGAATAAGAACTTCGACCTGCGTCCGGGCGTCATTGTGCGTGATCTGGATCTGAAGAAACCTATTTATCA
GAAAACTGCGTGCTATGGTCATTTCGGTCGTAGCGAATTTCCTTGGGAAGTGCCGAAGAAACTGGTTTTTTAG

9.6. Restriction analysis of pMC2.1.

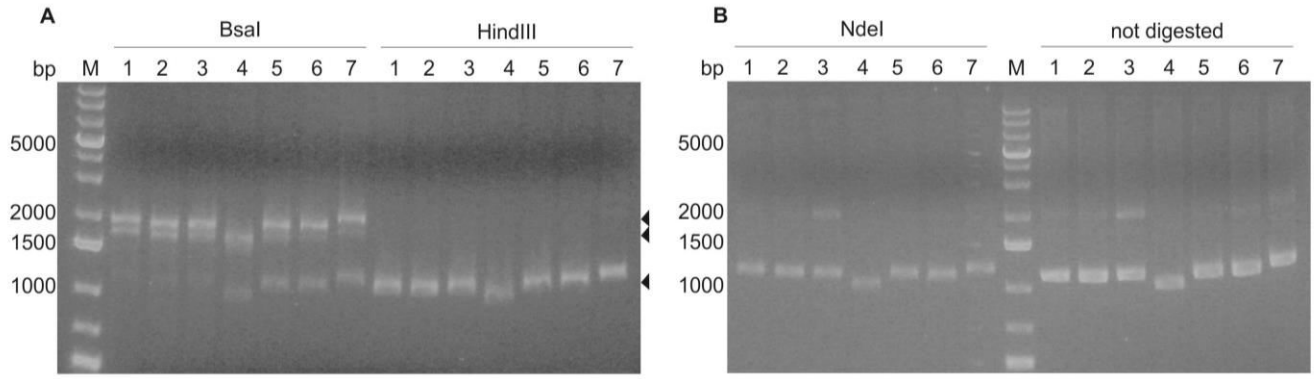

Figure S9: Restriction analysis of the destination vector pMC2.1.
A: Restriction digest with BsaI is leading to three fragments, whose calculated size is indicated by the arrows (1900, 1730 and 1360 bp). HindIII is a single cutter. B: Restriction digest with NdeI (single cutter). The analyzed clones were assembled using the following approaches: 1-3: cyc; 4: nc45min; 5-7: nc2h. Calculated size of the vector is 4990 bp. M, GeneRuler™ 1 kb Plus DNA Ladder.

9.7. Sequencing results of pMC1.1

Figure S10: Sequencing results of two clones of the destination vector pMC1.1 are showing the same insert within the inducible PTacII promoter.
The inserted sequence is TTGTGAGCGGATAACAATTA. A BLAST search gave no meaningful hits.

9.8. HPLC-UV method

Table S9: HPLC method

Solvent	90% 20 mM KH_2PO_4 (pH 8); 10% ACN
Column flow	0.7 mL/min
Stop time	15 min
Maximum pressure	400 bar
Temperature setting	20°C
Injection volume	10 µL
Detection	260 nm

9.9. Codon optimization

Table S10: Rare codons of the FIA which were manually exchanged.
The underlined codons were exchanged incorrectly during the first codon optimization. Codons, which are in bold, indicate the codons exchanged during the second optimization.

Position	Amino acid	Relative adaptiveness[5] (%)	Codon used in optimized sequence	Manually changed to: FIA c.o1	Manually changed to: FIA c.o2
2	A	2	GCA	**GCG**	GCC
6	T	3	ACT	ACA	
7	R	10	CGT	**CGC**	CGG
9	P	9	CCC	**CCG**	CCC
62	L	2	CTG	TTG	
69	G	8	GGC	GGT	
77	Y	4	TAC	TAT	
98	G	9	GGC	GGA	
180	R	10	CGC	CGT	
258	G	9	GGT	GGC	

5 Values are calculated by the program Graphical Codon Usage Analyser. Relative adaptiveness means that the best codon (e.g. GAA for Glu) is set to 100. All others values (e.g. GAG for Glu) are calculated using the rule of proportion (Villalobos et al. 2006).