Lisa Ramona Offner, MSc

Synthetic Biology - Looking at it from a chemist's and an engineer's point of view

MASTER'S THESIS

to achieve the university degree of

Diplom-Ingenieurin

Master's degree programme: Biotechnology

submitted to

Graz University of Technology

Supervisor

Univ.-Prof. Dipl.-Ing. Dr.techn. Helmut Schwab

Institute of Molecular Biotechnology
Graz University of Technology

Dipl.-Ing. Dr.techn. Birgit Wiltschi
Austrian Centre of Industrial Biotechnology

Graz, December 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

ACKNOWLEDGEMENTS

First of all, I thank my family and my friends for encouraging me throughout my entire studies. A special thanks goes to my parents, without whose support it would not have been possible to attend and finally graduate from university.

I also want to thank my colleges of the Junior Group Synthetic Biology for their help in the lab, especially in the beginning of my thesis. I really enjoyed working with you!

In particular, I thank Patrik Fladischer who always had a sympathetic ear and good advice for my problems and Kathrin Heckenbichler for working with me on my project and continuing it until completion.

Special thanks go to my supervisor, Dipl.-Ing. Dr.techn. Birgit Wiltschi, for her guidance throughout my master thesis.

I also would like to express my thanks to Univ.-Prof. Dipl.-Ing. Dr.techn. Helmut Schwab for giving me the opportunity to conduct my thesis at the Institute of Molecular Biotechnology in the Junior Group Synthetic Biology.

Last but not least, I thank the Austrian Centre of Industrial Biotechnology and the FFG for funding my project.

ABSTRACT

Synthetic biology is a rapidly evolving, interdisciplinary field of biology, chemistry and engineering. With this thesis, we intended to take the chemist's and the engineer's points of view and to exploit the tools developed in both disciplines of synthetic biology for an integrated research approach. Applying the chemical synthetic biology approach, we evaluated a single protein production system for the expression of synthetic variants of the *C*-methyltransferase NovO. The enzyme catalyzes a chemo- and regioselective Friedel-Crafts reaction and was, therefore, an attractive candidate for protein engineering. By incorporating non-canonical amino acids, we were able to improve the biophysical properties and the catalytic activity of the enzyme. Like NovO, its homolog CouO is also capable of acting as alkyltransferase, when provided with *S*-adenosyl-L-methionine analogs as non-natural alkyl group donors. To bypass the problems associated with the chemical synthesis of the artificial cofactors, we reached into the bio-engineer's tool box and coupled the biosynthesis of the cofactor analogs to the Friedel-Crafts alkylation reaction catalyzed by the methyltransferase. A combination of chemical synthetic biology and metabolic engineering allowed the biosynthesis and incorporation of alkene-containing amino acids. This functionalization enabled the target protein, flavodoxin to participate in olefin metathesis reactions which can be exploited for the production of biopolymers.

KURZFASSUNG

Die synthetische Biologie ist ein sich schnell weiterentwickelndes, interdisziplinäres Forschungsfeld, mit Einflüssen aus der Biologie, Chemie und Ingenieurwissenschaft. Ziel dieser Arbeit ist es, an das Gebiet mit der Sichtweise eines Chemikers als auch eines Technikers heranzugehen und die neuesten Entwicklungen in beiden Bereichen der synthetischen Biologie für einen integrativen Forschungsansatz zu nutzen. Mit Hilfe des chemische synthetische Biologie-Ansatzes evaluierten wir ein "single protein production system" zur Expression von synthetischen Varianten der *C*-Methyltransferase NovO. Das Enzym katalysiert eine chemo- und regioselektive Friedel-Crafts Reaktion und war daher ein attraktiver Kandidat für Protein Engineering. Durch den Einbau von nicht-kanonischen Aminosäuren konnten wir die biophysikalischen Eigenschaften und die katalytische Aktivität des Enzyms zu verbessern. Wie NovO, ist auch die homologe *C*-Methyltransferase CouO fähig als Alkyltransferase zu agieren, wenn ihr *S*-adenosyl-L-methionin Analoga als nicht-natürliche Alkylgruppendonoren zur Verfügung gestellt werden. Um die Probleme, die mit der chemischen Synthese der künstlichen Cofaktoren verbunden sind, zu umgehen, griffen wir in die Werkzeugkiste des Bio-Ingenieurs und verknüpften die Biosynthese der Cofaktor Analoga mit der von der Methyltransferase katalysierten Friedel-Crafts Alkylierung. Die Kombination der chemischen synthetischen Biologie und des Metabolic Engineering ermöglichte die Biosynthese und den Einbau von Aminosäuren mit Alkengruppen. Diese Funktionalisierung befähigte das Zielprotein, Flavodoxin sich an Olefinmetathese Reaktionen zu beteiligen, die wiederum für die Produktion von Biopolymeren genutzt werden können.

TABLE OF CONTENTS

INTRODUCTION

Basically, synthetic biology may be divided in two major research areas. On the one hand, there is the bio-engineering aspect based on the genetic modification of existing forms of life, mainly microorganisms like bacteria and yeasts. Here, the synthetic biologist uses natural molecules and assembles them to create an unnatural system.[1] This may be genes (or parts) from different organisms combined in a cascade or pathway (or devices) that does not exist in nature. The cell (or system) is then able to produce chemicals of interest and acts as nature-based chemical reaction chassis.

While the goal of bio-engineering is to enrich the functionality of microorganisms (and maybe multi-cellular organisms in future) for production,[2] chemical synthetic biology is devoted to basic research. Here, the synthetic biologist uses unnatural molecules that are chemically similar to the natural ones and redesigns natural systems.[1] This may be the incorporation of non-canonical amino acids (or parts) into proteins (or devices). Chemical synthetic biology is trying to answer the question: “why did nature this and not that?” by doing that and comparing it with this. By analyzing organisms in this way, we may be able to understand and to engineer them more effectively, one day.[3]

Incorporation of non-canonical amino acids into proteins - looking at synthetic biology from a chemist's point of view

Proteins are capable of carrying out a remarkable variety of functions with only 20 natural amino acids. However, the canonical amino acids (cAA) included by the genetic code do not feature attractive chemistries like halogens, keto, aldehyde and azido groups or C-C double and triple bonds. By expanding the amino acid repertoire, as basic building blocks for polypeptide biosynthesis, tailor-made proteins with improved or even new biological or physicochemical properties and catalytic activities can be designed.[4] To change the interpretation of the genetic code and to incorporate non-canonical amino acids (ncAA) in place of their canonical counterparts, the cellular protein translation machinery has to be reprogrammed. This delicate mission can be accomplished by suppressing the termination codons or non-triplet coding units using an orthogonal pair of suppressor tRNA and aminoacyl tRNA synthetase specific for the ncAA (site-specific incorporation) or by reassigning the sense codons.[5] The second approach is referred to as residue-specific, supplementation-based or selective pressure incorporation (SPI) and is dealt with in this thesis. When this pressure in form of starvation for the cAA, which should be replaced by the

supplemented ncAA, is exerted on auxotrophic microorganisms, the amino acid supply controls the selection of the building blocks for protein biosynthesis. Furthermore, the incorporation of an ncAA is facilitated by the inability of the key enzymes in the translation of the genetic code (the aminoacyl tRNA synthetases) to distinguish between the cAA and their chemically similar analogs as substrates of the amino acylation reaction.[6]

Among the ncAA, fluorinated amino acids (FAA) have gained great importance.[7] For a century, chemists have been aware of the unique characteristics of fluorinated molecules. Although their qualities were exploited to develop outstanding poly-fluorinated materials like the polymer Teflon (polytetrafluoroethylene), FAA were added to the amino acid repertoire for protein engineering only a decade ago.[8]

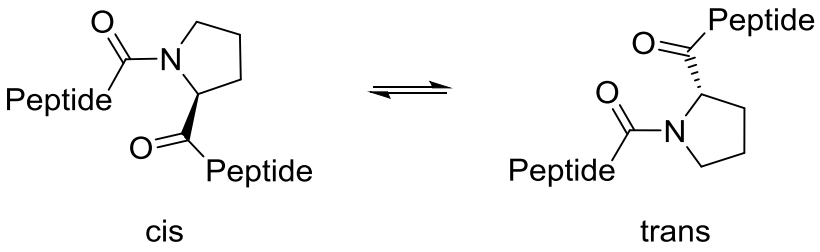


Figure 1: *Cis/trans* equilibrium of the prolyl peptide bond.[9]

Improved solubility or stability have been reported for a variety of proteins when Pro was substituted with F-Pro.[10] The effects of F-Pro on the biophysical properties of proteins are due to the unique structure of the prolyl peptide bond and to the special characteristics of the fluorine substituent. In the Xaa-F-Pro bond, the fluorine atom of the proline ring and the nitrogen atom of the peptide bond favor a *gauche* over an *anti* orientation to allow a good overlap between the best donor and the best acceptor orbital for hyperconjugation (*gauche* effect). Depending on the configuration of the fluorine substituent on the C4 stereo center ((4*R*)-FPro or (4*S*)-FPro), the proline ring has to be in the C^{γ}-*exo* or C^{γ}-*endo* conformation to be able to arrange the electronegative atoms in *gauche* orientation. Due to its cyclic structure, the conformation of the proline ring has an effect on the configuration of the peptide bond, which may be *cis* or *trans* (Figure 1).[9] This means that there is a correlation between the stereochemistry of F-Pro and the configuration of the fluoroprolyl peptide bond which in turn changes the biophysical properties of a protein. Therefore, it is critical to find the ratio of both F-Pro diastereomers that results in the most stable, soluble and active variant of the target protein.

Another highly interesting class of ncAA are amino acids featuring C-C double bonds, as they enable proteins to participate in bio-orthogonal olefin metathesis reactions. The mechanism according to the Chauvin model is shown in Figure 2.[11]

While the reaction has been known in petro- and polymer chemistry for decades and was used in industrial processes early on, its implementation in synthetic chemistry was long in coming.[12] The first generation of strongly Lewis-acidic metathesis catalysts were poorly compatible with polar functional groups which restricted the applicability of the reaction to the production of unfunctionalized polymers.[13] Only the development of new stable, water and oxygen insensitive catalysts with a high functional group tolerance rendered olefin metathesis a useful reaction in organic chemistry.[14] Today, it is even applied in the total synthesis of complex natural products.[15]

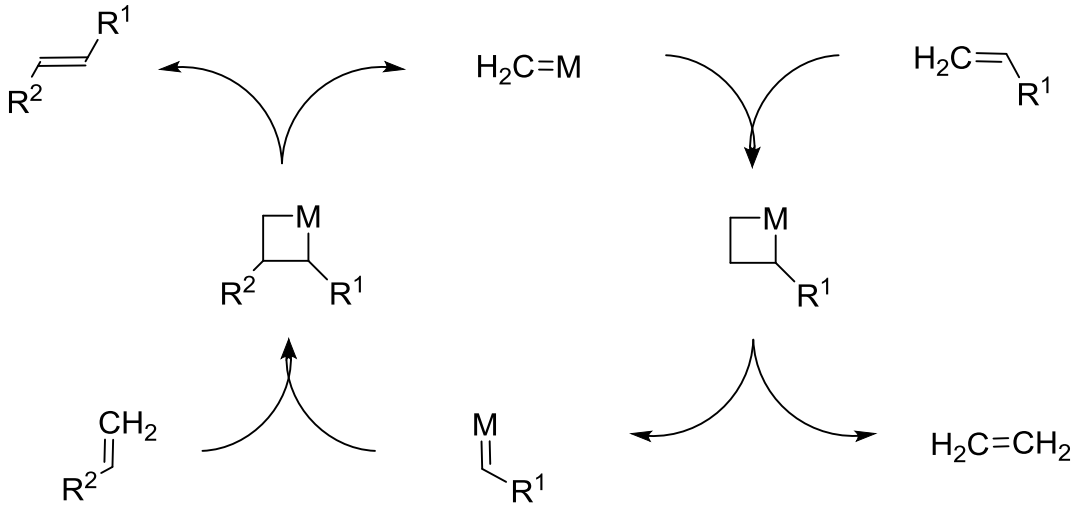


Figure 2: Catalytic cycle of the metathesis reaction.[11] In the case of peptide or protein coupling, R^1 and R^2 refer to the alkene-containing amino acid side chains.

Also, metathesis has long been considered unsuitable for protein engineering, as organic solvents and reaction conditions required for the reaction are not favorable for protein solubility or stability. Owing to substantial research, metathesis reactions in aqueous solution were made possible by water-soluble catalysts modified with polar or charged moieties. Davis *et al.* showed that the reaction is also catalyzed by conventional hydrophobic metathesis catalysts (e.g. Hoveyda-Grubbs second generation catalyst) in aqueous-organic solutions allowing full maintenance of the protein structure.[16] While the side chains of alkene-containing amino acids in peptides and small proteins could already be coupled by intramolecular olefin metathesis,[17] the potential of the intermolecular reaction for the production of biopolymers has not been exploited yet.

Genetic engineering of reconstituted pathways - looking at synthetic biology from an engineer's point of view

Their applicability in multiple-step one-pot processes is one of the greatest advantages of biocatalysts over chemocatalysts, which often cannot be combined due to their incompatibility with each other. In the cell, enzymatic reactions are carried out in a common medium, the cytosol. The cellular pathways or cascades act as a role model for the production of interesting chemicals in organic synthesis.[18] By combining two or more biotransformations in multi-enzyme catalytic systems, clean and sustainable processes that obey to the principles of green chemistry are developed (Figure 3).[19]

Figure 3: The 12 principles of green chemistry.[20]

Multi-enzyme systems are inherently green, as biocatalysts are bio-based and biodegradable, and act in an aqueous reaction medium. They display high chemo-, regio- and stereoselectivity, which minimizes the formation of by-products in side reactions. Further waste production is prevented using multi-step systems, as the isolation of intermediates becomes redundant. By reducing the number of downstream processing steps, also less of the intermediates is lost which leads to higher overall yields of the process. Additionally, higher conversions can be achieved for reversible reactions, as the product is constantly removed from the equilibrium to react as substrate in the following step. Unstable or hazardous intermediates do not accumulate but are converted to the final product, which makes processes safer.[18]

Aims of the Master Thesis

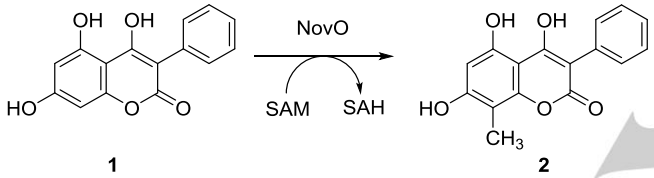
With the new tools developed in chemical synthetic biology and bio-engineering, we are able to master difficult tasks. In this thesis, the SPI method will be exploited to engineer the *C*-methyltransferase NovO with regard to its catalytic activity and biophysical properties. To make this enzyme engineering technique more economical, a single protein production system will be evaluated for the expression of fluorinated variants. NovO and its homolog CouO are also capable of acting as alkyltransferases, when provided with *S*-adenosyl-L-methionine (SAM) analogs as non-natural alkyl group donors. To bypass the problems associated with the chemical synthesis of the artificial cofactors, we will reach into the engineer's tool box and couple the biosynthesis of the cofactor analogs to the Friedel-Crafts alkylation reaction catalyzed by the methyltransferase. A combination of chemical synthetic biology and metabolic engineering will allow the biosynthesis and incorporation of alkene-containing amino acids. This functionalization will render the target protein, flavodoxin a substrate for intermolecular olefin metathesis and a potential candidate for biopolymer production.

EVALUATION OF A SINGLE PROTEIN PRODUCTION SYSTEM FOR THE EXPRESSION OF FLUORINATED BIOCATALYSTS

Evaluation of a Single Protein Production System for the Expression of Fluorinated Biocatalysts

Lisa Offner,[a,b] Benoît David,[b] Mandana Gruber-Khadjawi,[b] Christian Gruber,[c] Karl Gruber,[c] Birgit Wiltschi*[b]

Abstract: We evaluated a single protein production system for the expression of the *C*-methyltransferase NovO from *Streptomyces spheroides* and its fluorinated variants. The enzyme acts as methyl- and alkyltransferase and catalyzes a chemo- and regioselective Friedel-Crafts reaction (Scheme 1).[1] The reduction of the culture volume had no negative effect on the expression level of the parent protein, but led to a decreased expression of the fluorinated protein variants. By incorporating (2*S*,4*R*)-4-fluoroproline, the activity and stability of the enzyme were improved, while incorporation of (2*S*,4*S*)-4-fluoroproline and the diastereomeric mixture caused the methyltransferase to be insoluble.

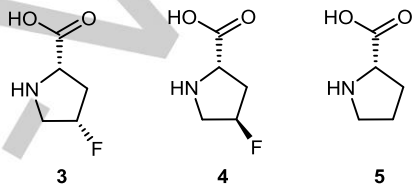


1 2

Scheme 1. Methylation of the substrate, 4,5,7trihydroxy-3-phenylcoumarin (**1**) catalyzed by the (*S*)-adenosyl-L-methionine (SAM)-dependent *C*-methyltransferase NovO. (*S*)-adenosyl-L-homoserine (SAH) is the by-product of the reaction.

The methyl group constitutes one of the most common functional groups in pharmaceutically relevant molecules.[2] Therefore, there is a great demand for chemo- and regioselective reactions by which a methyl or alkyl group can be introduced into a molecule. While chemical alkylation reactions often do not tolerate the presence of other functional groups, methyltransferases are highly selective with regard to the atom they transfer the methyl group to.[3] This renders biocatalytic alkylation reactions valuable and environmentally benign alternatives to their chemical equivalents.

However, the catalytic activities and biophysical properties of wild-type enzymes do not always comply with the requirements of industrial applications. These proteins are often engineered to acquire the desired properties. In the past decade, the incorporation of non-canonical amino acids (ncAA) has evolved into an effective, cutting-edge enzyme engineering technique by which the solubility, catalytic activity and stability of proteins may be improved.[4] Among the ncAAs, fluorinated amino acids (FAA) have gained great importance.[5] To date, the incorporation of FAA has been subject to substantial research and has been reported to have significant effects on the properties of proteins.[6] Therefore, FAA have earned their special status as valuable building blocks in the protein engineer's toolbox.



3 4 5

Scheme 2. Structures of the non-canonical amino acids (2*S*,4*S*)-4-fluoroproline (**3**) and (2*S*,4*R*)-4-fluoroproline (**4**), and the isostructural canonical amino acid (*S*)-proline (**5**).

In the present study, we intended to develop fluorinated variants of the *C*-methyltransferase NovO with improved properties and potential for biotechnological applications. For the expression of these synthetic enzymes or synzymes in *E. coli*, supplementation incorporation (SPI) is the method of choice.[7] The selected FAA (2*S*,4*S*)-4-fluoroproline (S, **3**) and (2*S*,4*R*)-4-fluoroproline (R, **4**) are isostructural to the canonical amino acid (cAA) (*S*)-proline (P, **5**) (Scheme 2). Therefore, they are recognized by the *E. coli* prolyl-tRNA synthetase and suitable for SPI.[8]

To make this protein engineering technique more economical, we tested the so-called single protein production system (SPPS) for the expression of the methyltransferase. SPPS is based on SPI and has been shown to be a suitable method for the incorporation of non-canonical amino acids into proteins.[9] The system was designed to produce only a single protein of interest in living *E. coli* cells. This is accomplished by co-expressing the endoribonuclease MazF which cleaves single-stranded RNAs specifically at ACA sequences. To prevent the degradation of the target mRNA, all ACA sequences must be changed to alternative sequences in the gene of interest yet the corresponding amino acid sequence must not be altered. MazF cleaves the endogenous *E. coli* mRNAs and thus inhibits the biosynthesis of cellular proteins but leaves the transcript of interest intact for translation and expression of the target protein. As a consequence of the action of MazF cell growth is stopped, which allows an up to 40-fold compression of the culture volume and, hence, of the amount of high-priced ncAA. For a more detailed description of the SPPS methodology, the reader is referred to the literature.[9b]

[a] MSc L. Offner
Institute of Molecular Biotechnology, Graz University of Technology
Petersgasse 14, A-8010 Graz, Austria
E-mail: lisa.offner@student.tugraz.at
[b] acib - Austrian Centre of Industrial Biotechnology
Petersgasse 14, A-8010 Graz, Austria
[c] Institute of Molecular Biosciences, Karl Franzens University Graz
Humboldtstrasse 50/3, A-8010 Graz, Austria
* corresponding author

Supporting information for this article is available.

NovO is expressed at high levels in soluble form using the pET-26b(+) expression system.[1] However, this expression system is not suitable for SPPS. NovO is expressed from a T7 promoter, which is specifically transcribed by the T7 RNA polymerase. During SPPS, the T7 RNA polymerase would be co-expressed with NovO and MazF. While an ACA-free coding sequence of NovO can be cloned into the pET vector, the coding sequence of the T7 RNA polymerase on the DE3 lysogen in the *E. coli* genome is not ACA-free. Consequently, the mRNA encoding the T7 RNA polymerase would be cleaved by MazF after induction with IPTG.[10]

Therefore, we screened different systems for a comparable or even better expression and solubility of the methyltransferase. Since Suzuki *et al.* developed the SPPS protocol using pCOLD plasmids, we first tested the system with the described vectors.[9b] Although the plasmids required for SPPS are commercially available, we constructed them from parts using different assembly tools of synthetic biology.[11] The gene encoding the endoribonuclease is controlled by an IPTG inducible *lacUV5* promoter on the pMazF plasmid. The pCOLD NovO Strep plasmid carries the ACA-free *novO* gene under the control of a cold shock inducible *cspA* promoter and a *lac* operator. This makes the expression of NovO inducible by a combination of cold shock and IPTG. The temperature is decreased to 15°C for expression and we expected better folding and an improved solubility of the protein.[12] To test the effect of a translation enhancing element (TEE) on the expression of NovO, an additional plasmid, pCOLD TEE NovO Strep was assembled. To evaluate the new protein production system with the methyltransferase, we decided to express only the parent protein under SPPS conditions before applying the technique for the incorporation of the ncAA.

pCOLD NovO Strep and pCOLD TEE NovO Strep were introduced into a proline-auxotrophic *E. coli* strain and the cells were grown in 500 mL minimal medium with limiting amounts of P until it was depleted. At this point the cultures were split in two. The cells of one half of the cultures were suspended in 6.25 mL medium (40-fold compressed), while the cells of the other half were suspended in 250 mL medium (non-compressed). After supplementation with P and induction of the co-expression of NovO and MazF, growth did not re-start. Since neither compressed nor non-compressed cultures were able to grow, the high cell density and the limited supply with nutrients after the reduction of the culture volume could be excluded as reasons for the growth stop. This indicates that the endoribonuclease MazF was functionally expressed and inhibited the biosynthesis of cellular proteins and hence, cell growth. However, contrary to our expectations from literature reports,[9b] NovO was not the only expressed protein (Figure 1, Figure 2, Figure 3). The 40-fold reduction of the culture volume did not negatively influence the NovO titer, yet the translation enhancing element (TEE) had no beneficial effect on the expression of the enzyme.

To test SPPS with other expression systems, we constructed five more plasmids carrying the ACA-free *novO* gene. The untranslated regions (UTR) upstream and downstream of the *novO* open reading frame are also present on the transcript and hence, could be subject to cleavage by the endoribonuclease. To increase the NovO expression level, we substituted the ACA sequences in the UTR on the pCOLD plasmids. Unexpectedly, the expression of NovO was lower with the pCOLD ACA-free UTR plasmids than with the pCOLD plasmids (Figure 1, lanes 1, 2, 4 and 5). We cloned the ACA-free *novO* gene into the pQE80L plasmid that allows expression at 28 °C. The protein background in these cultures was constant throughout the expression and was comparable to that of the cultures expressing NovO from the pCOLD constructs (Figure 1, lanes 3 and 6). This indicates that the low temperature did not inhibit the degradation of the cellular proteins synthesized before induction of the endoribonuclease. Furthermore, the expression level of the parent protein NovO[P] was weaker with the pCOLD plasmids than with the pQE80L plasmids (Figure 1). Since we expected even lower expression levels for the fluorinated variant proteins, the pCOLD plasmids were not suitable for the incorporation of **3** and **4** into NovO. The C-terminal Strep or Car9 tags had no effect on the expression (Figure 1).

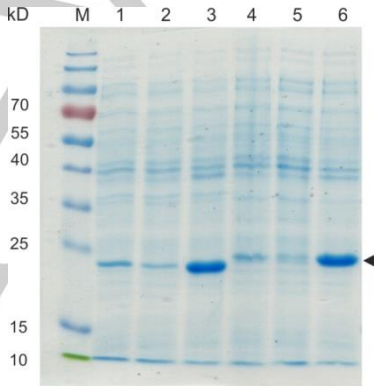


Figure 1. The highest expression level of NovO was achieved with the pQE80L plasmids. ACA sequences in the UTR had a beneficial effect while the tag had no influence on the target protein biosynthesis. 12% SDS-polyacrylamide gel was loaded with the total protein of cells which were co-transformed with pMazF and one of the following plasmids: pCOLD NovO Strep (1), pCOLD ACA free UTR NovO Strep (2), pQE80L NovO Strep (3), pCOLD NovO Car9 (4), pCOLD ACA free UTR NovO Car9 (5) or pQE80L NovO Car9 (6). Culture samples were collected 44 h after induction (10 µL, normalized to a D_{600} of 0.5). M, molecular weight marker (5 µL Page Ruler™ Prestained Protein Ladder). Bands of the target protein NovO are marked with a triangle (MW_{calc} 26.6 kD). Coomassie Brilliant Blue stain.

The constant protein background throughout the experiment (44 h after induction) indicates that cellular proteins were synthesized after induction of the endoribonuclease and that the enzyme was potentially inactive although the cells did not grow. To analyze if the action of MazF is indeed required for expression in a reduced culture volume, we compared the target protein production in concentrated and non-concentrated cultures using the pET-26b(+) NovO (P5A) Strep plasmid, and the pQE80L NovO Strep or pQE80L NovO Car9 plasmids with the pMazF plasmid (Figure 2). While the expression levels of NovO in the concentrated and non-concentrated cultures did not differ when MazF was co-expressed, the concentrated culture with the pET-26b(+) plasmid produced significantly less of the methyltransferase than the corresponding non-concentrated control culture. The Strep and Car9 tags did neither have an effect on the expression level nor on the solubility of NovO.

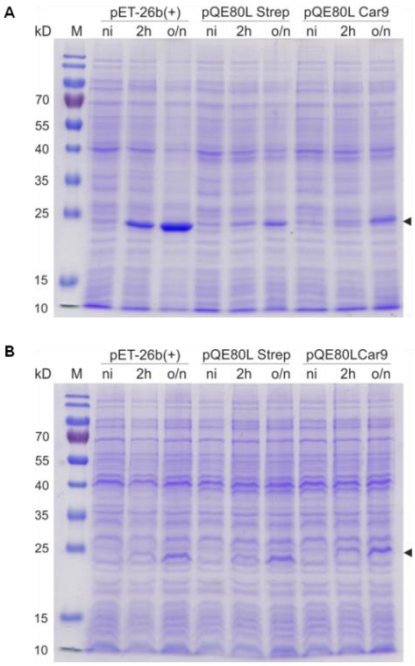


Figure 2: Reduction of the culture volume decreased the level of NovO expressed from the pET-26b(+) NovO (P5A) Strep (pET-26b(+)), while it did not affect the level of the methyltransferase expressed by cells co-transformed with pMazF and pQE80L NovO Strep (pQE80L Strep) or pQE80L NovO Car9 (pQE80L Car9). The tags had no effect on the expression level. A: non-concentrated cultures, **B**: concentrated cultures. 12% SDS-polyacrylamide gels were loaded with the total protein of cells which were transformed with pET-26b(+) NovO (P5A) Strep or with pMazF and pQE80L NovO Strep or pQE80L NovO Car9. Samples were collected before induction (ni), 2 h and over night (o/n) after induction (10 µL, normalized to a $D_{600}$of 0.5). M, molecular weight marker (5 µL Page Ruler^{TM} Prestained Protein Ladder). Bands of the target protein NovO are marked with a triangle (MW_{calc} 26.6 kD). Coomassie Brilliant Blue stain.

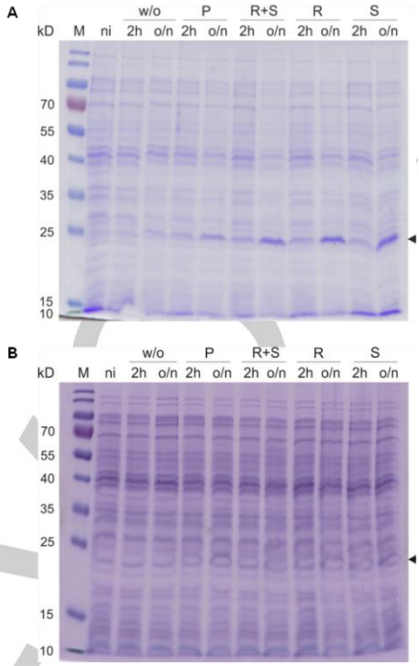


Figure 3. Reduction of the culture volume decreased the expression level of the variants. A: non-concentrated cultures, **B**: concentrated cultures. 12% SDS-polyacrylamide gels were loaded with the total protein of cells which were supplemented with **3**, **4**, **5** and a diastereomeric mixture of **3** and **4**. Samples were collected before induction (ni), 2 h and over night (o/n) after induction (10 µL, normalized to a D_{600} of 0.5). M, molecular weight marker (5 µL Page Ruler^{TM} Prestained Protein Ladder). Bands of the target protein NovO are marked with a triangle (MW_{calc} 26.6 kD). Coomassie Brilliant Blue stain.

Using the pQE80L NovO Strep plasmid, NovO was successfully overexpressed, when the cultures were supplemented with the cAA **5**, the FAA **3** and **4** or a 50:50 mixture of both diastereomers, after depletion of **5** (Figure 3). Unexpectedly, the expression level of the parent protein was lower than that of the variants in the non-concentrated culture. In accordance with the previous experiment, the concentrated and non-concentrated cultures produced the same amount of parent protein. However, the reduction of the culture volume caused a significant decrease of the expression level for the variants.

While the solubility of the parent protein and the NovO[R] variant was comparable, incorporation of **3** and the diastereomeric mixture of **3** and **4** caused the enzyme to be insoluble. Due to insufficient amounts of soluble NovO[S] and NovO[R+S], the variants were not included in the further analysis.

The mass spectrometric analysis (ESI-MS) revealed that the parent protein was subject to methionine oxidation either during expression, lysis, purification or mass analysis. The NovO[R] variant protein preparation was inhomogeneous. Besides the fully substituted variant and the parent protein, species with two, four, six, seven or eight **4** residues were also detected.

Despite the incomplete substitution of **5** with **4**, the catalytic activity (Scheme 1) and stability of the parent and the variant proteins were analyzed and compared.

Table 1. Activity of the parent protein NovO[P] and the variant NovO[R] towards the methylation of substrate 1. The substrate conversion was calculated by comparing the substrate concentrations of the blank and the samples after 2 or 24 h. In the case of a conversion of >99%, no substrate peak was detected. The mean of three technical replicates is given.

	conversion [%] after 2 h	conversion [%] after 24 h
NovO[P]	40	>99
NovO[R]	61	>99

By incorporating **4**, the methyl transfer activity of NovO was enhanced (Table 1). Furthermore, a higher stability of NovO[R] compared to NovO[P] was noticed in 50 mM sodium phosphate buffer pH 6.5. Within three weeks at 4°C, the concentration of NovO[P] decreased due to precipitation of the protein, while the concentration of NovO[R] was almost constant.

Experimental Section

The expressions with SPPS were carried out according to the protocol of Suzuki *et al.*[9b] which is described, with all deviations, in full detail in the supporting information. For the incorporation of **3**, **4** and **5**, the proline-depleted cultures were supplemented with 1 mM of the amino acids or a 50:50 mixture of the diastereomers and expression was induced with 0.5 mM IPTG. The cells were lysed by sonication and the soluble protein fractions were purified by affinity chromatography.

The biotransformation (Scheme 1) was carried out in 100 µL scale in an Eppendorf thermomixer at 35 °C and 1000 rpm for 2 h or 24 h. The reaction mix contained a total amount of 50 µg purified NovO[P] and NovO[R] prepared in 50 mM sodium phosphate buffer pH 6.5 (reaction buffer), 0.5 mM substrate **1** (from a 10 mM solution in dimethyl sulfoxide for full solubility), 2 mM (*S*)-adenosyl-L-methionine (SAM) (from a 20 mM solution in the reaction buffer) and 1 mg/mL bovine serum albumin (from a 20 mg/mL solution in the reaction buffer). The reaction was started by adding the enzyme solution to the reaction mix and stopped by heating it to 85°C for 10 min. The denatured protein was removed prior to HPLC-MS analysis by centrifugation at 13200 rpm for 10 min. A detailed description of the HPLC-MS analysis is given in the supporting information.

Acknowledgements

We gratefully acknowledge Dr. Martin Tengg for providing the pET-26b(+) NovO(P5A) Strep plasmid.

Keywords: synthetic biology • enzyme engineering • non-canonical amino acids • biocatalysis • C-C bond formation

[1] H. Stecher, M. Tengg, B. J. Ueberbacher, P. Remler, H. Schwab, H. Griengl, M. Gruber-Khadjawi, *Angewandte Chemie International Edition* **2009**, *48*, 9546-9548.

[2] H. Schönherr, T. Cernak, *Angewandte Chemie International Edition* **2013**, *52*, 12256-12267.

[3] a) C. Dalhoff, E. Weinhold in *Modified Nucleosides*, Wiley-VCH Verlag GmbH & Co. KGaA, **2008**, pp. 223-247; b) S. Klimasauskas, G. Lukinavicius, T. P. Begley in *Wiley Encyclopedia of Chemical Biology*, John Wiley & Sons, Inc., **2007**, pp. 8-17.

[4] N. Budisa, *Angewandte Chemie International Edition* **2004**, *43*, 6426-6463.

[5] C. Odar, M. Winkler, B. Wiltschi, *Biotechnology Journal* **2015**, *10*, 427-446.

[6] a) M. D. Crespo, M. Rubini, *PLoS ONE* **2011**, *6*, e19425; b) B. Holzberger, A. Marx, *Journal of the American Chemical Society* **2010**, *132*, 15708-15713; c) B. Holzberger, S. Obeid, W. Welte, K. Diederichs, A. Marx, *Chemical Science* **2012**, *3*, 2924-2931; d) N. Budisa, W. Wenger, B. Wiltschi, *Molecular BioSystems* **2010**, *6*, 1630-1639.

[7] L. Merkel, M. Schauer, G. Antranikian, N. Budisa, *ChemBioChem* **2010**, *11*, 1505-1507.

[8] K. Deepankumar, M. Shon, S. P. Nadarajan, G. Shin, S. Mathew, N. Ayyadurai, B.-G. Kim, S.-H. Choi, S.-H. Lee, H. Yun, *Advanced Synthesis & Catalysis* **2014**, *356*, 993-998.

[9] a) L. Falzon, M. Suzuki, M. Inouye, *Current Opinion in Biotechnology* **2006**, *17*, 347-352; b) M. Suzuki, L. Mao, M. Inouye, *Nat. Protocols* **2007**, *2*, 1802-1810.

[10] H. P. Sørensen, K. K. Mortensen, *Journal of Biotechnology* **2005**, *115*, 113-128.

[11] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, *Nature Methods* **2009**, *6*, 343-345.

[12] G. Qing, L.-C. Ma, A. Khorchid, G. V. T. Swapna, T. K. Mal, M. M. Takayama, B. Xia, S. Phadtare, H. Ke, T. Acton, G. T. Montelione, M. Ikura, M. Inouye, *Nat Biotech* **2004**, *22*, 877-882.

Supporting Information

Evaluation of a Single Protein Production System for the Expression of Fluorinated Biocatalysts

Lisa Offner,[a,b] Benoît David,[b] Mandana Gruber-Khadjawi,[b] Christian Gruber,[c] Karl Gruber,[c] Birgit Wiltschi*[b]

E-mail: birgit.wiltschi@acib.at

[a] Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, A-8010 Graz, Austria

[b] acib - Austrian Centre of Industrial Biotechnology, Petersgasse 14, A-8010 Graz, Austria

[c] Institute of Molecular Biosciences, Karl Franzens University Graz, Humboldtstrasse 50/3, A-8010 Graz, Austria

Supporting Information

Table of Content

1) Nucleotide and amino acid sequences

1.1) Nucleotide sequence of *novO* on the pCOLD and pQE80L plasmids

The nucleotide sequence of the *C*-methyltransferase NovO from *Streptomyces spheroides* with a P5A mutation is given. For SPPS, all ACA sequences were substituted with other nucleotides without altering the amino acid sequence of NovO. The C-terminal Strep tag II is underlined and the substituted ACA sequences are highlighted in bold.

atgaagattgaagcgattacgggatccgaggccgaagcatttcaccggatgggctcccaggcgtccc**ata**ggtacgacgaattc
gtggatctgctcgtcggggccggcatcgccgacggccagaccgtggtcgacc**ttt**gc**tgc**ggctcgggtgagctggaagtcatc
ctcagctcccggttcccctcgctgaatctggtcgg**cgt**cgacctctcggagg**ata**tggtccgtatcgcccgggagtacgcggccg
agcagggcaaggcgctcgaattccggcacggtgacgcgcaactcctggccggtatggaggatctggccggaaaagcggatct
ggtggtctcacggaacgcgttccaccggctcacccggctgccggcagcgttcg**ata**ccatgctgcgtctggccaagccgggag
gggccgtcctcaactgctccttcat**tca**cccgagcgacttcgacgagtccgggttccgggcgtgggtaacgttcctcaaccagcg
gccctgggactcggagatgcagatcgtctgggcgctggctcaccattacgcgccccggcttgacgactaccgcgaggccctggc
ccaggcggcccgtgagactccgg**ttt**ccgagcagcgcgtctggatcgacgaccagggatacggcgttcccacggtcaaatgctt
cgcccgtcgggccgctgccagcgctt<u>tggagccacccgcagttcgaaaaat</u>ga

1.2) Amino acid sequence of NovO

The amino acid sequence of NovO is given. The C-terminal Strep tag II is underlined and the nine proline residues are highlighted in bold.

M K I E A I T G S E A E A F H R M G S Q A S H R Y D E F V D L L V G A G I A D G Q T V V
D L C C G S G E L E V I L S S R F **P** S L N L V G V D L S E D M V R I A R E Y A A E Q G K
A L E F R H G D A Q L L A G M E D L A G K A D L V V S R N A F H R L T R L **P** A A F D T
M L R L A K **P** G G A V L N C S F I H **P** S D F D E S G F R A W V T F L N Q R **P** W D S E M
Q I V W A L A H H Y A **P** R L D D Y R E A L A Q A A R E T **P** V S E Q R V W I D D Q G Y G
V **P** T V K C F A R R A AA S A <u>W S H **P** Q F E K</u>*

2) Construction of plasmids

2.1) pMazF

The pMazF plasmid is composed of a p15A origin of replication, a *lacI* gene with a constitutive promoter, a kanamycin resistance gene with a constitute promoter and a T7 terminator, the *mazF* gene under a *lacUV5* promoter and a λ t0 terminator. The sequences of the *lacUV5* promoter and of the gene encoding MazF were ordered on one synthetic gBlock (Integrated DNA Technologies, Coralville, IA). For sequence verification the gBlock was cloned into a pJET1.2 vector (Thermo Fisher Scientific Inc., Waltham, MA) by blunt end ligation following the protocol of the manufacturer and amplified in *E. coli* TOP10F' cells. The vector served as a template for the amplification of the *mazF* gene and the *lacUV5* promoter. The other parts for the pMazF plasmid were amplified from the commercially available pACYC plasmid and from pET26b(+)-novO.[1] The primers (Integrated DNA technologies) for the PCR amplifications were designed to introduce 40 bp homologous overlaps to the adjacent parts for overlap extension PCR and Gibson isothermal assembly. The reactions contained 10 ng of template DNA, 0.5 µM of the forward and reverse primers, 0.2 mM of each deoxyribonucleotide triphosphate, 1X *Phusion* HF Buffer and 1 U *Phusion* high-fidelity DNA Polymerase (Thermo Fisher Scientific Inc.) in a total volume of 50 µL. The PCR program included an initial denaturation (98 °C for 30 sec), followed by 30 cycles of denaturation (98 °C for 10 sec), annealing (56 °C for 20 sec) and elongation (72 °C for 1 min) and the final elongation (72 °C for 7 min).

The p15A origin of replication and the T7 terminator as well as the *lacUV5* promoter with the *mazF* gene and the λ t0 terminator were joined via overlap extension PCR. For the assembly of the pMazF plasmid Gibson isothermal assembly was applied. We followed the method described by Gibson *et al.*.[2] 50 ng of the smallest fragment and all other fragments in equimolar quantities were added to 7.5 µL of the assembly master mix which is composed of 10 U/µL T5 exonuclease (Biozym Scientific GmbH, Hessisch Oldendorf, Germany), 2 U/µL *Phusion* high-fidelity DNA Polymerase (Thermo Fisher Scientific Inc., Waltham, MA), 40 U/µL *Taq* DNA ligase (New England BioLabs Inc., MA) and 1X isothermal (ISO) reaction buffer (5% (w/v) PEG-8000, 100 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 0.2 mM of each deoxyribonucleotide triphosphate, 1 mM NAD in water). The mixture was incubated at 50 °C for 60 min without shaking. The resulting plasmid was amplified in *E. coli* TOP10F' cells on LB agar plates containing 50 µg/mL kanamycin and it was verified by sequencing (Microsynth, Balgach, Switzerland). The plasmid map of pMazF is shown in Figure S1.

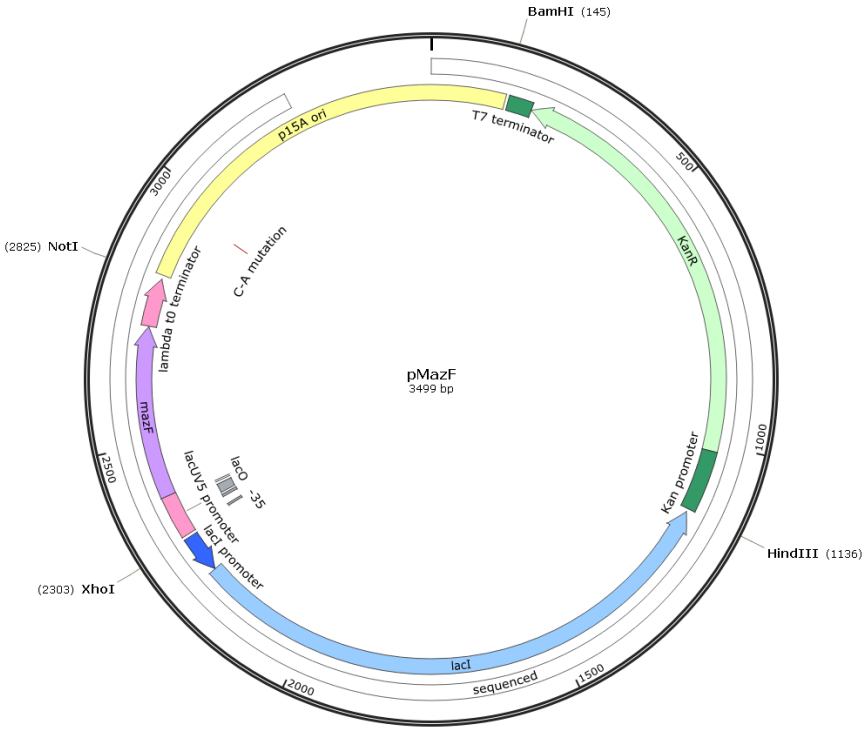


Figure S1: Plasmid map of pMazF harboring the gene encoding the single-stranded RNA- and ACA-sequence-specific endoribonuclease MazF. The plasmid carries a medium copy number p15A origin of replication, a kanamycin resistance cassette with a T7 terminator and a *lacI* expression cassette, both controlled by constitutive promoters. The *mazF* gene is located downstream of an IPTG inducible *lacUV5* promoter and upstream of a λ t0 terminator. The MazF expression cassette is flanked by XhoI and NotI restriction sites, which allow its excision. The sequence marked in white was verified, the p15A ori had a C-A mutation in position 128.

2.2) pCOLD NovO Strep and pCOLD TEE NovO Strep

All ACA sequences of the *novO* gene were substituted with MazF-insensitive sequences without altering the amino acid sequence of the protein. A synthetic gBlock composed of a PacI restriction site, a cold shock inducible *cspA* promoter, a *lac* operator, a *cspA* 5'UTR (untranslated region), an NdeI restriction site, a ribosome binding site (RBS), a translation enhancing element (TEE),[4] the modified *novO* gene, a C-terminal Strep tag II, a multiple cloning site (MCS), a *cspA* 3'UTR and a NcoI restriction site, was ordered (Integrated DNA Technologies). The sequences of the promoter, the operator, the UTRs, the RBS, the TEE and the MCS were adopted from TAKARA's pCOLD plasmids. For sequence verification the gBlock was cloned into a pJET1.2 vector (Thermo Fisher Scientific Inc.) by blunt end ligation following the protocol of the manufacturer and amplified in *E. coli* TOP10F' cells. The vector and the pCAS1 plasmid[3] served as templates for the PCR amplification of the pCOLD plasmid parts. For the assembly of the pCOLD NovO Strep and pCOLD TEE NovO Strep plasmids Gibson isothermal assembly was applied. The resulting plasmids were amplified in

E. coli TOP10F' cells on LB agar plates containing 100 µg/mL ampicillin and they were verified by sequencing (Microsynth). The plasmid maps of pCOLD NovO Strep and pCOLD TEE NovO Strep are shown in Figure S2 and Figure S5, A.

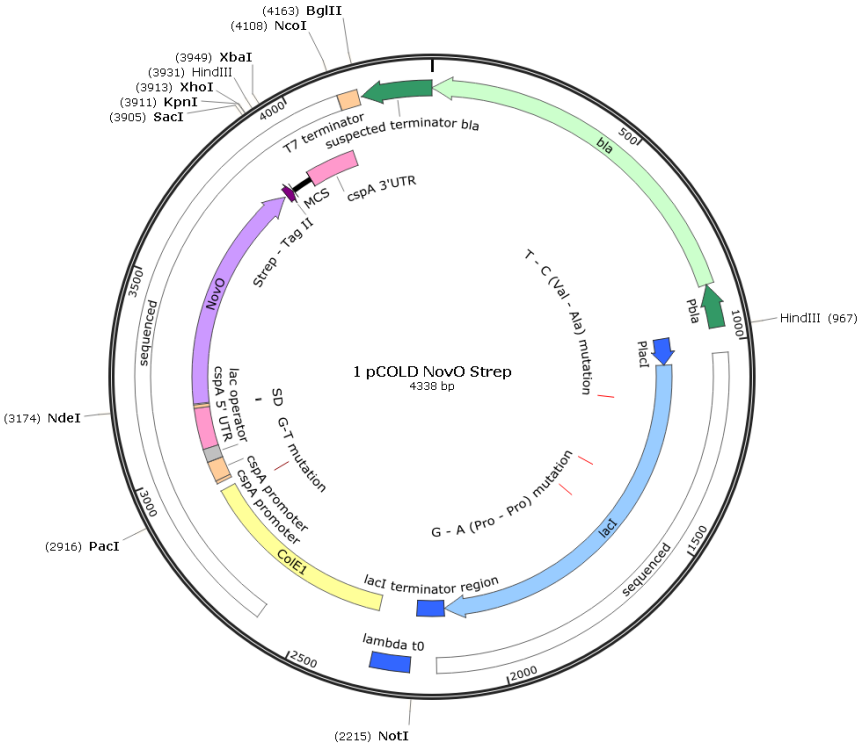


Figure S2: Plasmid map of pCOLD NovO Strep harboring the ACA-free *novO* gene and a Strep tag II. The plasmid carries a medium copy number ColE1 origin of replication, a *lacI* expression cassette with a λt0 terminator and an ampicillin resistance cassette, both are constitutively expressed. The *novO* gene is located downstream of a cold shock inducible *cspA* promoter, a *lac* operator, a *cspA* 5'UTR and an RBS and carries a C-terminal Strep tag II. A MCS, a *cspA* 3'UTR and a T7 terminator follow downstream of *novO*. The expression of NovO is inducible by a combination of cold shock and IPTG. The *novO* gene and the sequence of the C-terminal Strep tag II are flanked by NdeI and SacI restriction sites which allow their excision. The sequence marked in white was verified, the ColE1 ori has a G-T mutation in position 562 and the *lacI* gene has three mutations (V38A, P127P, L174P). As shown in Figure S7 and Figure S9, the two amino acid exchanges did not affect its function as Lac repressor protein.

2.3) pCOLD ACA free UTR NovO Strep

To construct a pCOLD NovO Strep plasmid with ACA free UTRs (Figure S5, B), the corresponding sequences were ordered as a synthetic gBlock (Integrated DNA Technologies). For sequence verification the gBlock was cloned into a pJET1.2 vector (Thermo Fisher Scientific Inc.) by blunt end ligation following the protocol of the manufacturer and amplified in *E. coli* TOP10F' cells. The 5'UTR was excised with BbsI and used as a mega primer for QuikChange PCR with pCOLD NovO Strep as a template. The PCR was performed with 1 U Phusion high fidelity DNA Polymerase (Thermo Fisher Scientific Inc.) and the following program: 95 °C for 2 min (initial denaturation), 95 °C for 1 min (denaturation), 60 °C for 20 sec (annealing), 68 °C for 15 min (elongation), 25 cycles of denaturation, annealing and elongation. After PCR, the template was degraded using DpnI (Thermo Fisher Scientific Inc.) and the newly synthesized plasmid was amplified in *E. coli* TOP10F' cells on LB agar plates containing 100 µg/mL ampicillin. Its sequence was verified by sequencing (Microsynth). The pCOLD NovO Strep plasmid with an ACA free 5'UTR was used as a template for the second round of QuikChange PCR with the 3'UTR as a mega primer which was excised from the pJET1.2 vector with BsaI.

2.4) pQE80L NovO Strep

All parts of pQE80L NovO Strep (Figure S3) were PCR amplified from the commercially available pQE-80L plasmid (Qiagen, Venlo, the Netherlands) and pCOLD NovO Strep. For the assembly of the pQE80L NovO Strep plasmid Gibson isothermal assembly was applied. The resulting plasmid was amplified in *E. coli* TOP10F' cells on LB agar plates containing 100 µg/mL ampicillin and its sequence was verified by sequencing (Microsynth).

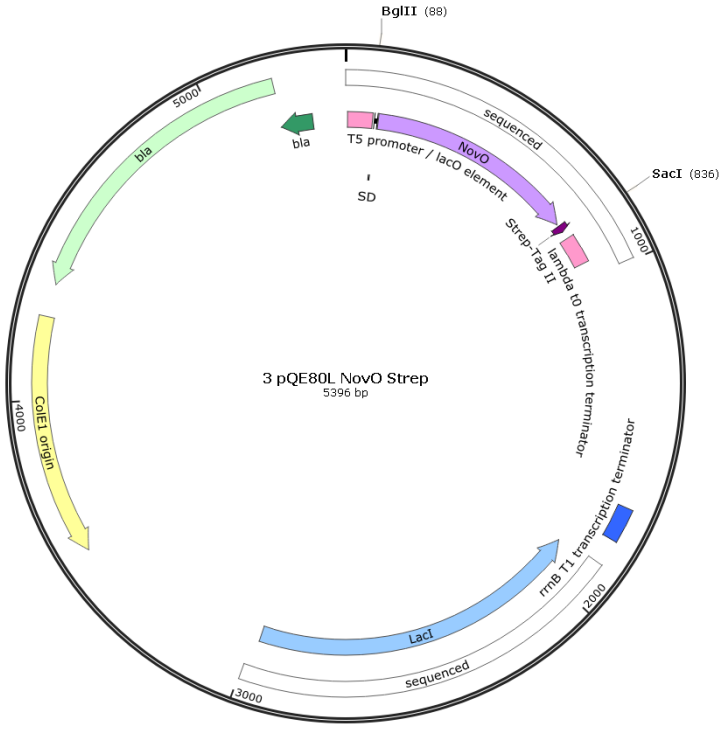


Figure S3: Plasmid map of pQE80L NovO Strep harboring the ACA-free *novO* gene and a Strep tag II. The plasmid carries a medium copy number ColE1 origin of replication, a *lacI* expression cassette and an ampicillin resistance cassette, both are constitutively expressed. The *novO* gene is located downstream of an IPTG inducible T5 promoter, a *lac* operator and an RBS and upstream of a C-terminal Strep tag II, a MCS and a lambda t0 terminator. The expression of NovO is inducible by IPTG. The *novO* gene and the sequence of the C-terminal Strep tag II are flanked by BglII and SacI restriction sites which allow their excision. The sequence marked in white was verified.

2.5) pCOLD NovO Car9, pCOLD ACA free UTR NovO Car9 and pQE80L NovO Car9

For the construction of pCOLD NovO Car9 (Figure S5, C), pCOLD ACA free UTR NovO Car9 (Figure S5, D) and pQE80L NovO Car9 (Figure S5, E), the *novO* gene was amplified from the corresponding plasmids carrying the Strep II tag (Figure S2, Figure S5, B and Figure S3) with primers designed to introduce the Car9 tag and the restriction sites for cloning. pCOLD NovO Strep (Figure S2), pCOLD ACA free UTR NovO Strep (Figure S5, B) and the corresponding PCR product were treated with NdeI (FastDigest) and SacI (FastDigest). pQE80L NovO Strep (Figure S3) and the PCR product were treated with BglII (FastDigest) and SacI. The plasmids were dephosphorylated using alkaline phosphatase (FastAP). The plasmid parts were gel purified and ligated using T4 DNA ligase. The restriction, dephosphorylation and ligation were performed according to the protocol of the manufacturer (Thermo Fisher Scientific Inc.).

2.6) pET-26b(+) NovO (P5A) Strep

The *novO* gene with a P5A mutation and a Strep tag II was cloned into a commercially available pET-26b(+) plasmid.[1]

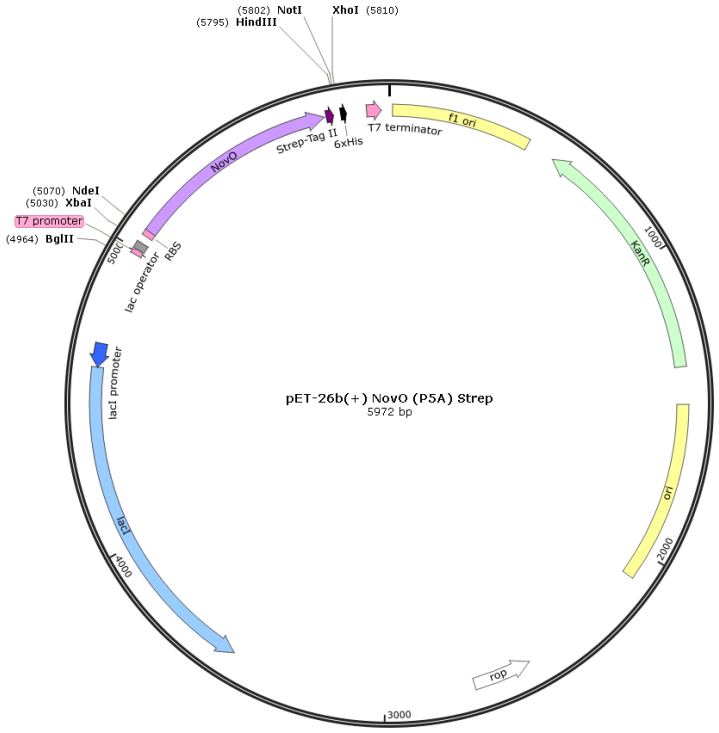


Figure S4: Plasmid map of pET-26b(+) NovO(P5A) Strep harboring the *novO* gene and a Strep tag II. The plasmid carries two origins of replication, a *lacI* expression cassette and an kanamycin resistance cassette, both are constitutively expressed. The *novO* gene is located downstream of an IPTG inducible T7 promoter, a *lac* operator and an RBS and upstream of a C-terminal Strep tag II and a T7 terminator. The expression of NovO is inducible by IPTG. The *novO* gene and the sequence of the C-terminal Strep tag II are flanked by restriction sites which allow their excision.

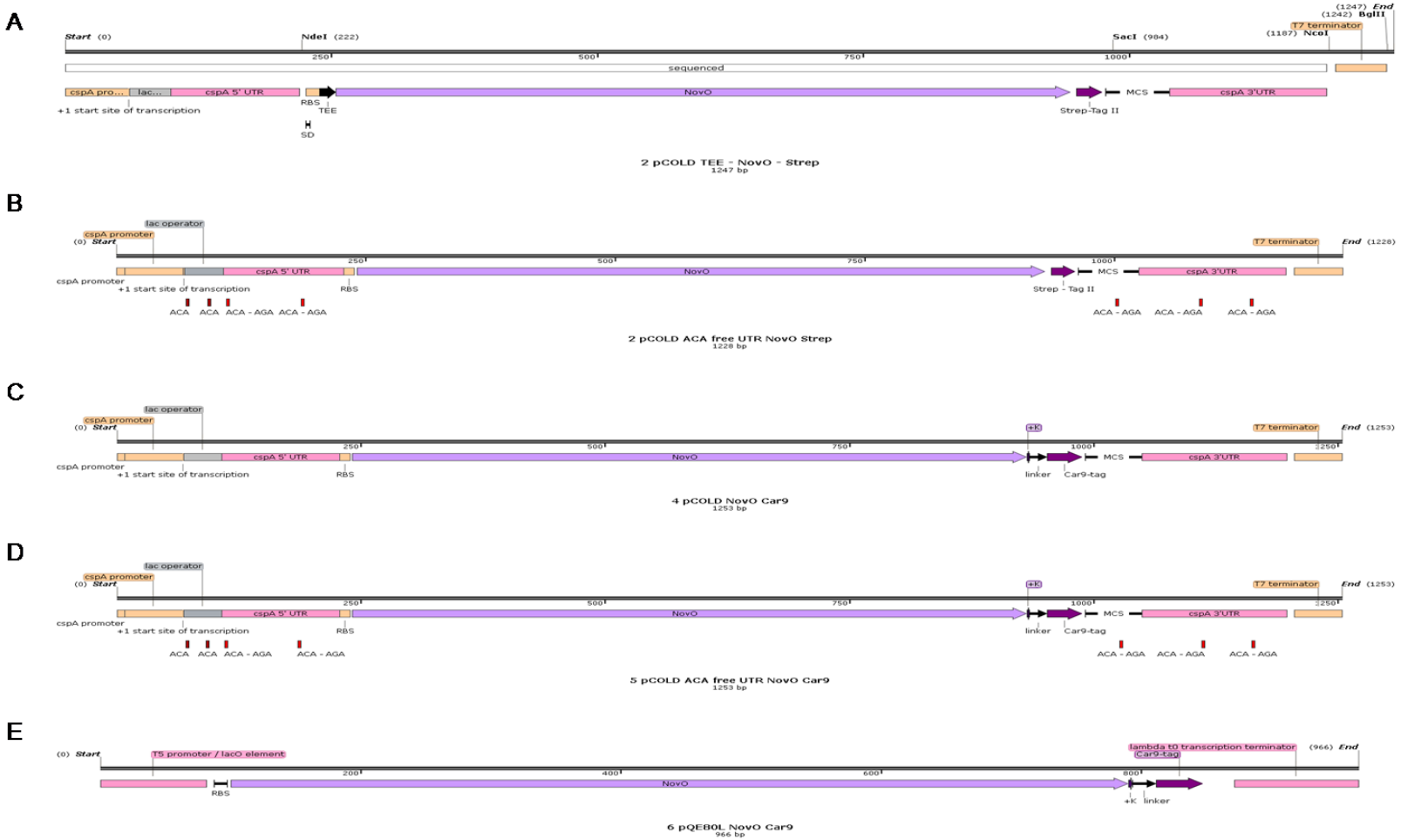


Figure S5: Sections of the plasmid maps of pCOLD TEE NovO Strep (A), pCOLD ACA free UTR NovO Strep (B), pCOLD NovO Car9 (C), pCOLD ACA free UTR NovO Car9 (D) and pQE80L NovO Car9 (E).

3) Supplementary Methods

3.1) Expression with the SPPS Protocol and the pCOLD Plasmids

Proline auxotrophic *E. coli* BL21 (DE3) Gold (BWEC44) cells were co-transformed with pMazF and one of the pCOLD plasmids and incubated over night in LB medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) supplemented with 0.4% (w/v) glucose. All media contained 50 µg/mL kanamycin and 100 µg/mL ampicillin for maintenance of the plasmids. M9 medium (composed according to Suzuki *et al.*[4]) supplemented with 0.1% (w/v) casamino acids was inoculated to a D_{600} of 0.1 and the cultures were incubated at 37 °C. To record a growth curve (Figure S6, Figure S8), D_{600} samples were collected in regular time intervals throughout the expression. When the cultures reached a D_{600} of 0.5, the medium was changed from M9 medium supplemented with casamino acids to M9 medium with lysine, phenylalanine, threonine, isoleucine, leucine and valine to achieve depletion for proline and to avoid activation of the endogenous *E. coli* MazF induced by starvation for other amino acids.[5] Therefore, the cultures were centrifuged at 4000 rpm and 15 °C for 10 min and the cells were resuspended in the same volume (non-concentrated cultures) or in a 40-fold reduced volume (concentrated cultures) of M9 medium with amino acid supplementation (KFTILV: Lys (100 µg/mL), Phe (100 µg/mL), Thr (100 µg/mL), Ile (50 µg/mL), Leu (50 µg/mL), Val (50 µg/mL)). After 1 h of depletion at 15 °C, the cultures were supplemented with 1 mM proline and incubated at 15 °C for 20 min for amino acid uptake. The co-expression of NovO and MazF was induced by adding 0.5 mM IPTG and the cultures were incubated at 15 °C. To analyze the expression by SDS-PAGE, whole-cell samples were collected before induction by cold shock (n1), before induction by IPTG (n2), and throughout the expression after induction (Figure S7, Figure S9). Deviations from this protocol are described in the figure legends.

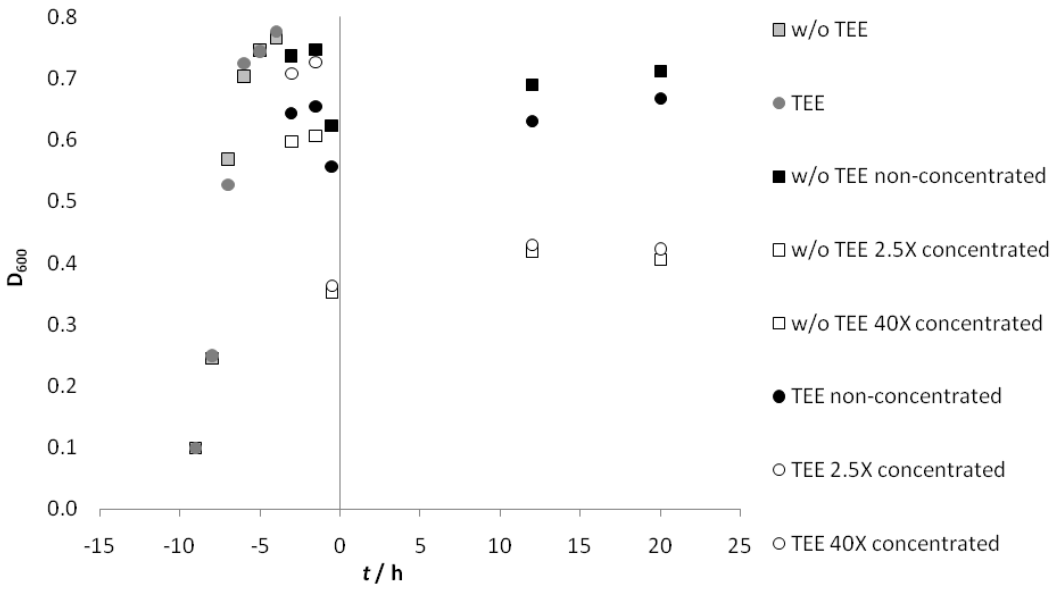


Figure S6: After supplementing the proline-depleted cells with proline and inducing the co-expression of NovO and MazF, growth did not re-start. Proline auxotrophic *E. coli* BWEC44 cells were co-transformed with pMazF and either pCOLD NovO Strep (w/o TEE, without translation enhancing element) or pCOLD TEE NovO Strep (TEE, translation enhancing element) and incubated in 500 mL M9 medium supplemented with casamino acids. For depletion of proline, the two cultures were divided in two and the cells were resuspended in either 250 mL (non-concentrated) or 100 mL (2.5X concentrated) M9 medium with amino acid supplementation (KFTILV). The cultures were cooled to 15 °C (cold shock). For the expression, the cells were resuspended in either 250 mL or 6.25 mL (40X concentrated) M9 medium. The cultures were supplemented with 1 mM proline and co-expression of NovO and MazF was induced with 1 mM IPTG (indicated by the vertical line). The D_{600} of the 100 mL and the 6.25 mL cultures were actually 2.5-fold and 40-fold higher. The samples collected from the concentrated cultures were diluted. Data points represent an average of three technical replicates.

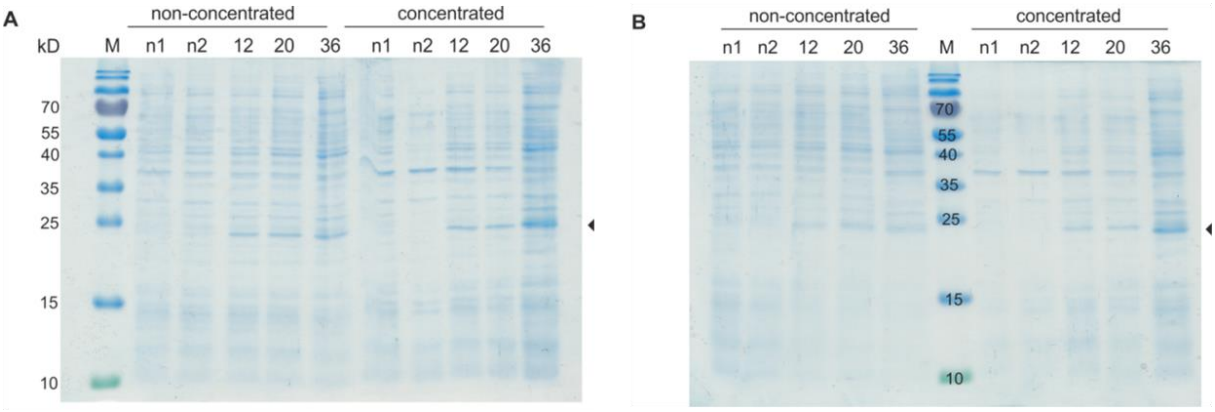


Figure S7. The expression level of NovO was neither increased by the translation enhancing element (TEE), nor decreased by the 40-fold reduction of the culture volume. 12 % SDS-polyacrylamide gels were loaded with the total protein of cells which were transformed with pMazF and pCOLD NovO Strep (A) or pCOLD TEE NovO Strep (B). Samples were collected before induction by cold shock (n1), before induction by IPTG (n2), and 12 h, 20 h and 36 h after induction (10 µL, normalized to a D_{600} of 0.25). M, molecular weight marker (5 µl Page Ruler^{TM} Prestained Protein Ladder, Thermo Fisher Scientific Inc., Waltham, MA). Bands of the target protein NovO are marked with a triangle (MW_{calc} 26.6 kD). Coomassie Brilliant Blue stain.

3.2) Expression with the SPPS Protocol and the pQE80L and pET Plasmids

Proline auxotrophic *E. coli* BWEC44 cells were co-transformed with pMazF and one of the pQE80L plasmids or with the pET-26b(+) plasmid and incubated over night in LB medium supplemented with 0.4% (w/v) glucose. All media contained 50 µg/mL kanamycin and 100 µg/mL ampicillin (pMazF and pQE80L plasmids) or 50 µg/mL kanamycin (pET26b(+)) for maintenance of the plasmids. M9 medium (composed according to Suzuki *et al.*[4]) supplemented with 0.1% (w/v) casamino acids was inoculated to a D_{600} of 0.1 and the cultures were incubated at 37 °C. To record a growth curve (Figure S8, Figure S10), D_{600} samples were collected in regular time intervals throughout the expression. After reaching a D_{600} of 0.5, the cultures were centrifuged at 4000 rpm and 20 °C for 10 min. The cells were resuspended in the same volume (non-concentrated cultures) or in a 40-fold reduced volume (concentrated cultures) of M9 medium with amino acid supplementation (KFTILV: Lys (100 µg/mL), Phe (100 µg/mL), Thr (100 µg/mL), Ile (50 µg/mL), Leu (50 µg/mL), Val (50 µg/mL)). After 1 h of depletion at 37 °C, the cultures were supplemented with 1 mM proline, (2*S*,4*S*)-4-fluoroproline, (2*S*,4*R*)-4-fluoroproline or a diastereomeric mixture and incubated at 28 °C for 20 min for amino acid uptake. The co-expression of NovO and MazF was induced by adding 0.5 mM IPTG and the cultures were incubated at 28 °C. To analyze the expression by SDS-PAGE, whole-cell samples were collected before induction (ni) and throughout the expression after induction (Figure S9, Figure 1, Figure 2, Figure 3).

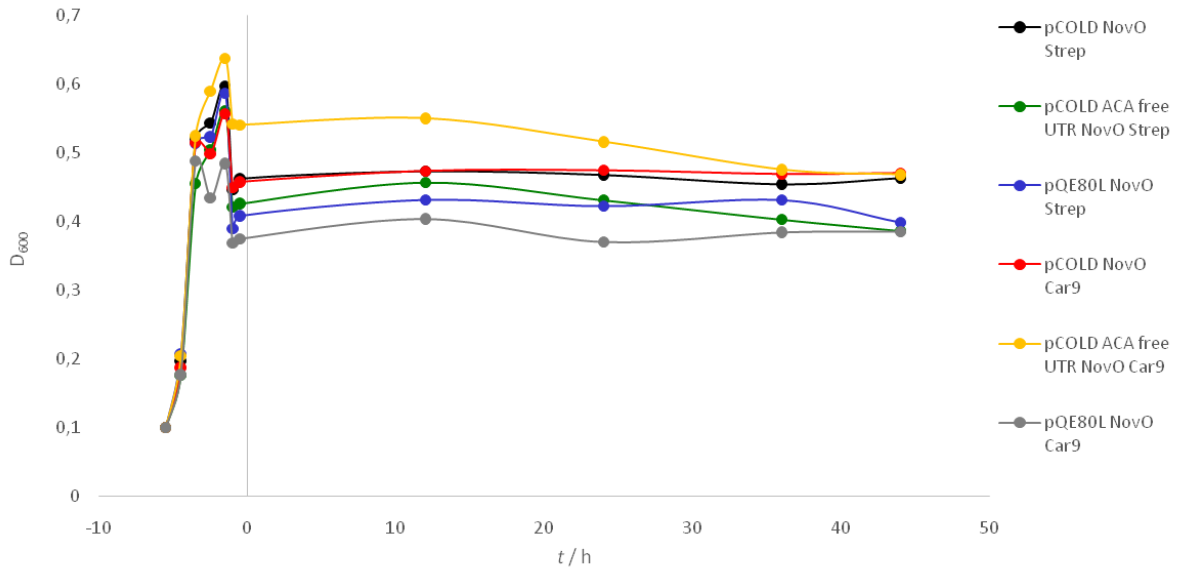


Figure S8: Also at 28 °C, growth did not re-start after supplementation. Proline auxotrophic *E. coli* BWEC44 cells were co-transformed with pMazF and one of the following plasmids: pCOLD NovO Strep, pCOLD ACA free UTR NovO Strep, pQE80L NovO Strep, pCOLD NovO Car9, pCOLD ACA free UTR NovO Car9 or pQE80L NovO Car9. After depletion, the cultures were supplemented with 1 mM proline and co-expression of NovO and MazF was induced by adding 0.5 mM IPTG (indicated by the vertical line). The cultures with the pCOLD plasmids were incubated at 15 °C and the cultures with the pQE80L plasmids at 28 °C for 44 h. The D_{600} of the concentrated cultures were actually 40-fold higher. The samples collected from these cultures were diluted. Data points represent an average of three technical replicates.

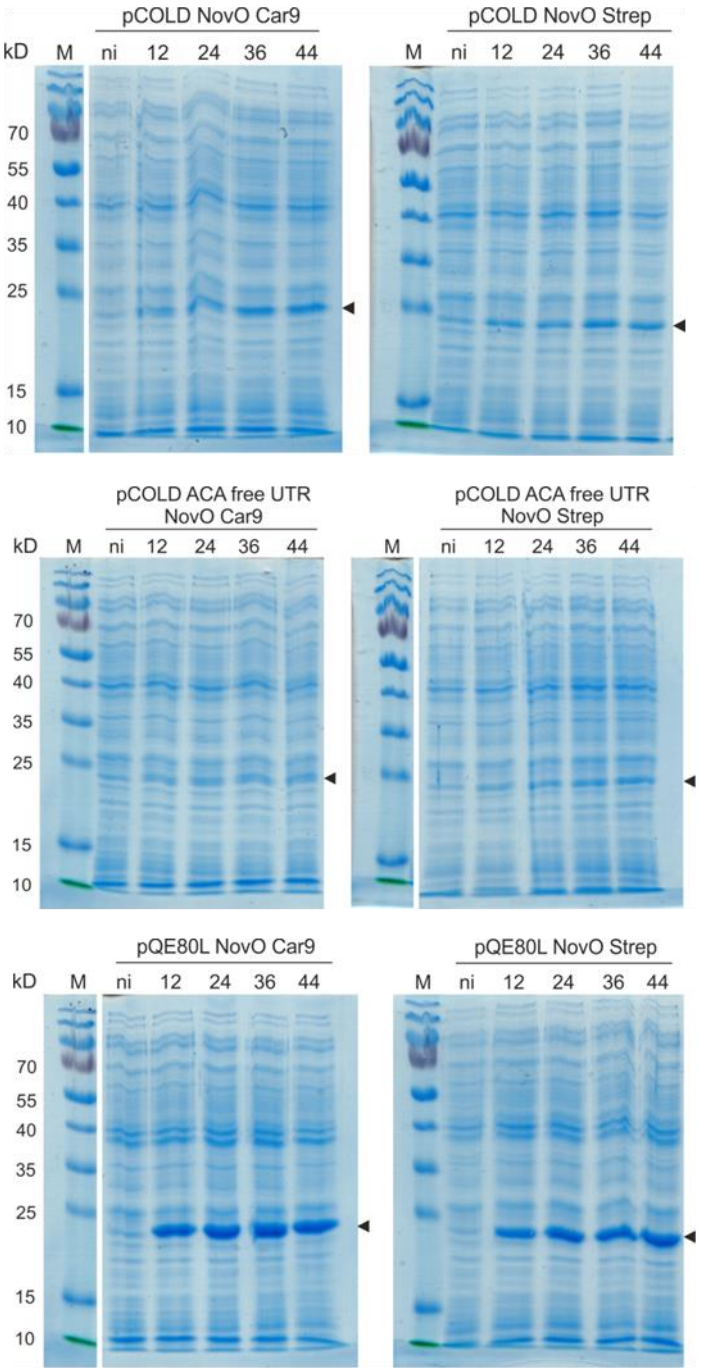


Figure S9: The highest expression level of NovO was achieved with the pQE80L plasmids. ACA nucleotide sequences in the untranslated regions (UTR) had a beneficial effect while the tag had no influence on the target protein biosynthesis. 12% SDS-polyacrylamide gels were loaded with the total protein of cells which were transformed with pMazF and one of the following plasmids: pCOLD NovO Car9 or pCOLD NovO Strep (top), pCOLD ACA free UTR NovO Car9 or pCOLD ACA free UTR NovO Strep (middle), pQE80L NovO Car9 and pQE80L NovO Strep (bottom). Samples were collected before induction (ni), 12 h, 24 h, 36 h and 44 h after induction (10 µL, normalized to a D_{600} of 0.5). M, molecular weight marker (5 µL Page Ruler^{TM} Prestained Protein Ladder, Thermo Fisher Scientific Inc., Waltham, MA). Irrelevant lanes were cut. Bands of the target protein NovO are marked with a triangle (MW_{calc} 26.6 kD). Coomassie Brilliant Blue stain.

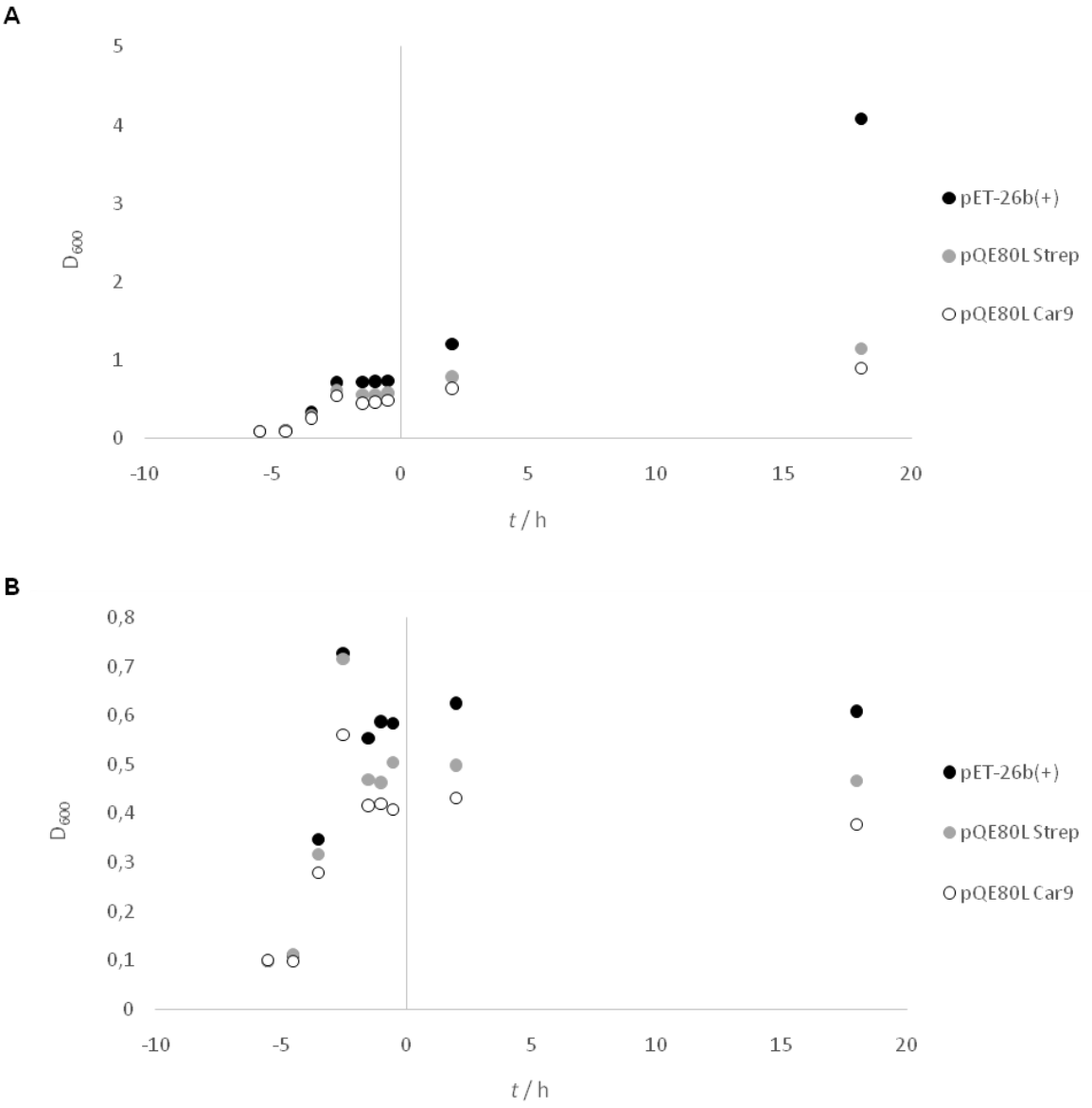


Figure S10: Since the concentrated pET-26b(+) culture did not re-start growing after supplementation, the growth stop was partly due to the 40-fold reduction of the culture volume. Proline auxotrophic *E. coli* BWEC44 cells were co-transformed with pMazF and pQE80L NovO Strep or pQE80L NovO Car9 or transformed with pET-26b(+) NovO (P5A) Strep, without MazF. After depletion, the cultures were supplemented with 1 mM proline and co-expression of NovO and MazF was induced by adding 0.5 mM IPTG (indicated by the vertical line). The D_{600} of the concentrated cultures were actually 40-fold higher. The samples collected from these cultures were diluted. **A**: non-concentrated cultures, **B**: concentrated cultures. Data points represent an average of three technical replicates.

3.3) Downstream Processing

After the expression, the medium was removed by centrifugation at 4000 rpm and 4°C for 20 min. The harvested cells were resuspended in phosphate buffer (50 mM, pH 6.5) and lysed by sonication. The solubility of NovO was analyzed by SDS-PAGE[6] (Figure S11, Figure S12). The soluble protein fractions were purified by Strep-Tactin affinity chromatography following the protocol of the manufacturer (IBA, Göttingen, Germany). The fractions, enriched in NovO[P] and NovO[R], were pooled and re-buffered from 2.5 mM desthiobiotin (100 mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA) elution buffer to phosphate buffer (50 mM, pH 6.5) using gravity flow PD-10 desalting columns (GE Healthcare Life Sciences, Little Chalfont, Great Britain). The protein concentrations were determined by Bradford assay (Bio-Rad Laboratories Inc., Hercules, CA). The solubility of the proteins and the quality of the purification were analyzed by SDS-PAGE (Figure S12).

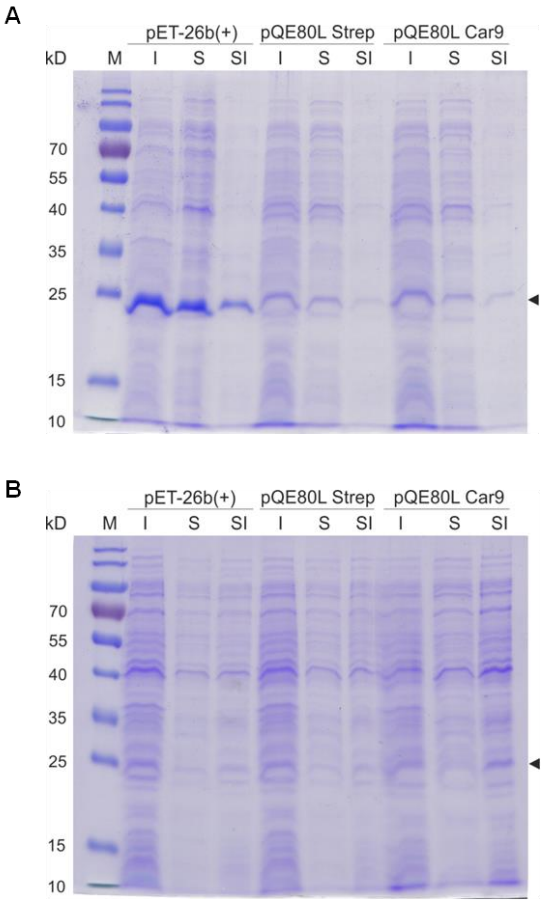


Figure S11: Reduction of the culture volume decreased the solubility of NovO[P]. The tags neither had an effect on the expression level (I) nor on the solubility (S, SI) of NovO. A: non-concentrated cultures, **B**: concentrated cultures. Whole cell samples taken after induction (I, 10 µL, normalized to a D_{600} of 0.5), and soluble (S) and insoluble (SI) protein fractions were loaded (4 µg lysate). M, molecular weight marker (5 µL Page Ruler^{TM} Prestained Protein Ladder, Thermo Fisher Scientific Inc., Waltham, MA). Bands of the target protein NovO are marked with a triangle (MW_{calc} 26.6 kD). 12% SDS-polyacrylamide gels. Coomassie Brilliant Blue stain.

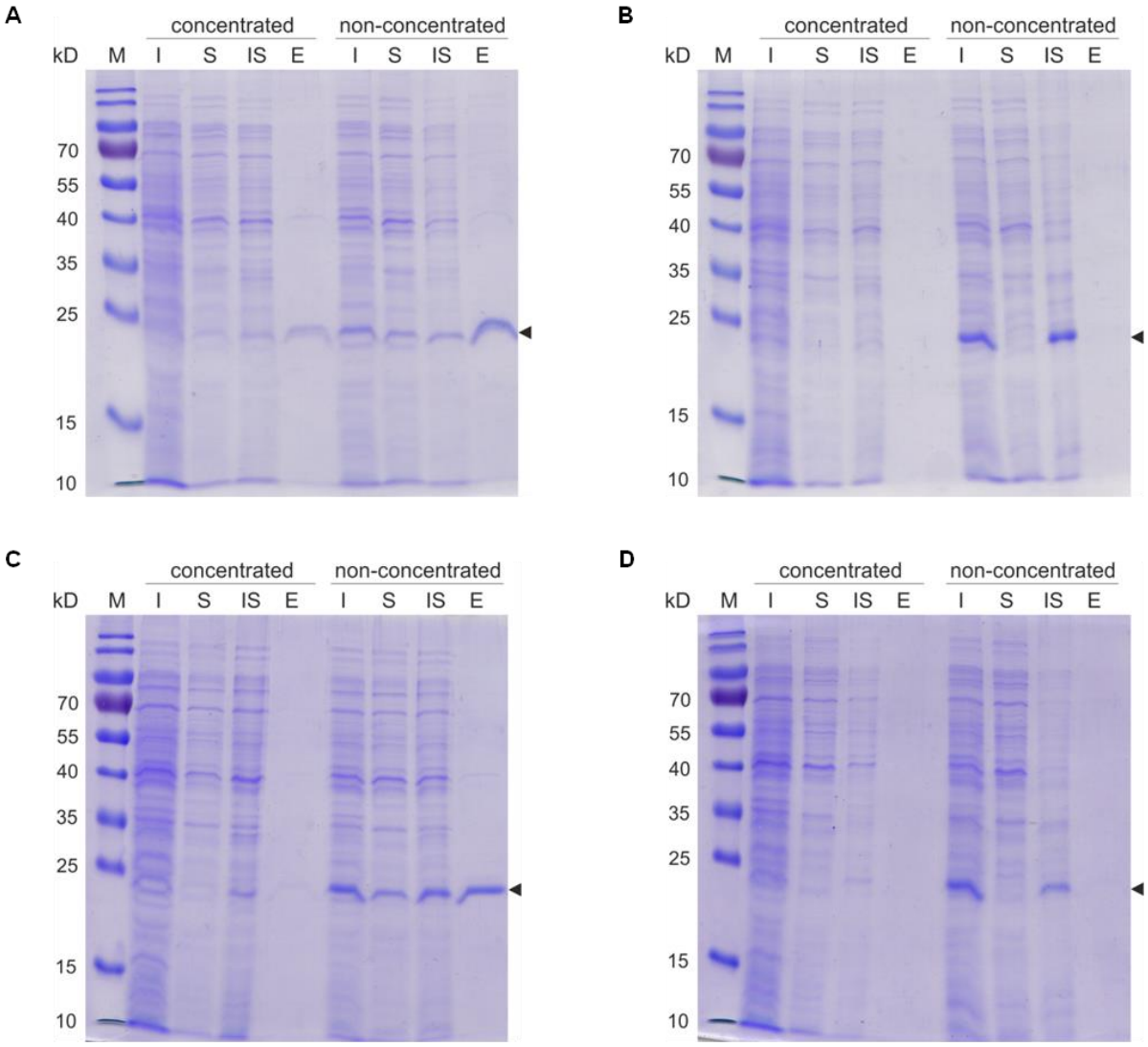


Figure S12: While the solubility of the parent protein and the (4*R*)-FPro variant was comparable, incorporation of (4*S*)-FPro and the diastereomeric mixture caused the enzyme to be insoluble. Whole cell samples taken after induction (I, 10 µL, normalized to a D_{600} of 0.5), and soluble (S), insoluble (IS) and purified (E) fractions of NovO[P] (**A**), NovO[R+S] (**B**), Novo[R] (**C**) and NovO[S] (**D**) were loaded (4 µg lysate, 2 µg purified protein). M, molecular weight marker (5 µL Page Ruler^{TM} Prestained Protein Ladder, Thermo Fisher Scientific Inc., Waltham, MA). Bands of the target protein NovO are marked with a triangle (MW_{calc} 26.6 kD). 12% SDS-polyacrylamide gels. Coomassie Brilliant Blue stain.

3.4) Activity Assay

The biotransformations were performed as described in the main text. The components of the reaction mix were analyzed on an Agilent Technologies HPLC-System 1200 equipped with a Merck Millipore Chromolith® RP-18e 100-4.6 column. The components were separated isocratically applying a 1.5 mL/min flow rate of 90% ammonium acetate (10 mM, pH 5.5) and 10% acetonitrile (HPLC-MS grade) for 15 min. The conversion of the substrate was measured by UV detection at 254, 260 and 330 nm. The mass spectrometer (ESI-MS positive mode) was set to scan mode (100-600 m/z) with single ion monitoring (SIM).

3.5) Mass Analysis

Analysis of the incorporation efficiency was performed by Electrospray Ionisation Mass Spectrometry (ESI-MS) at the ACIB Core Facility Proteomics, Medical University, Graz. Intact protein samples of NovO[P] and NovO[R] were analyzed on a Thermo Fischer Xcalibur mass spectrometer (Figure S13, Table 1, Table 2).

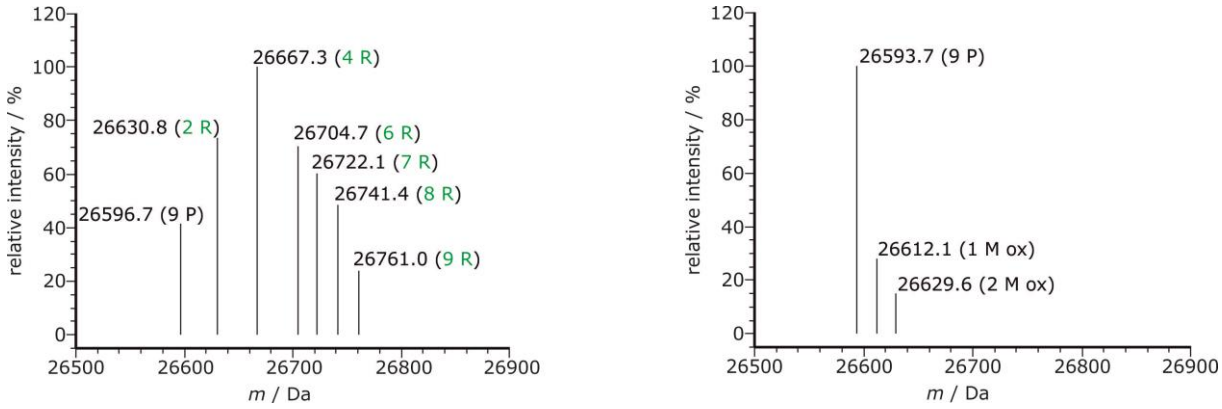


Figure S13: ESI-MS profile of the variant NovO[R] and the parent protein NovO[P]. Mass analysis revealed that differently substituted species of the variant protein were expressed. The parent protein was subject to methionine oxidation (M ox). The mass (m) is given in Dalton (Da).

Table 1: Mass analysis results for NovO[P].

protein species	average mass [Da]	$mass_{found}$ [Da]	Δmass [Da]
NovO	26596.7754	26593.7	3.0
NovO ($1M_{ox}$)	26612.7748	26612.1	0.7
NovO ($2M_{ox}$)	26628.7742	26629.6	0.8
NovO (1R)	26614.7659	26613.1	1.7
NovO (2R)	26632.7564	26630.4	1.7

Table 2: Mass analysis results for NovO[R].

protein species	average mass [Da]	$mass_{found}$ [Da]	Δmass[Da]
NovO	26596.7754	26596.7	0.1
NovO (1R)	26614.7659		
NovO (2R)	26632.7564	26630.8	2.0
NovO (3R)	26650.7468		
NovO (4R)	26668.7373	26667.3	1.4
NovO (5R)	26686.7277		
NovO (6R)	26704.7182	26704.7	0.0
NovO (7R)	26722.7087	26722.1	0.6
NovO (8R)	26740.6991	26741.4	0.7
NovO (9R)	26758.6896	26761.0	2.3
NovO (9R, $1M_{ox}$)	26774.6890	26778.2	3.5

4) Supplementary References

[1] H. Stecher, M. Tengg, B. J. Ueberbacher, P. Remler, H. Schwab, H. Griengl, M. Gruber-Khadjawi, *Angewandte Chemie International Edition* **2009**, *48*, 9546-9548.

[2] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, *Nature Methods* **2009**, *6*, 343-345.

[3] G. Gourinchas, E. Busto, M. Killinger, N. Richter, B. Wiltschi, W. Kroutil, *Chemical Communications* **2015**, *51*, 2828-2831.

[4] M. Suzuki, L. Mao, M. Inouye, *Nat. Protocols* **2007**, *2*, 1802-1810.

[5] X. Li, M. Yagi, T. Morita, H. Aiba, *Molecular Microbiology* **2008**, *68*, 462-473.

[6] U. K. Laemmli, *Nature* **1970**, *227,* 680-685.

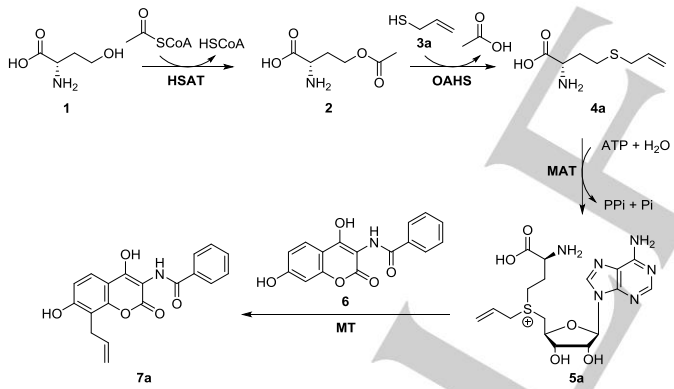
MULTI-ENZYME CASCADE COUPLING A MODULAR COFACTOR BIOSYNTHESIS TO A BIOCATALYTIC FRIEDEL-CRAFTS ALKYLATION

Multi-Enzyme Cascade Coupling a Modular Cofactor Biosynthesis to a Biocatalytic Friedel-Crafts Alkylation

Lisa Offner,[a, c] Kathrin Heckenbichler,[b] Patrik Fladischer,[a, c] Mandana Gruber-Khadjawi,[c] Rolf Breinbauer,[b] Birgit Wiltschi*[c]

Abstract: We developed a multi-enzyme cascade for the modular biosynthesis of *S*-adenosyl-L-methionine analogs that were used as artificial cofactors in a biocatalytic Friedel–Crafts alkylation. The enzymes involved in the multi-step reaction were expressed from combinatorially assembled plasmids. Different bottle necks of the pathway could be identified for the *in vitro* and *in vivo* biotransformations. Neither the cofactor analog, nor the alkylation product could be detected or identified by HPLC analysis.

Methylation is one of the most common chemical modifications in the cell. Hence, a substantial set of enzymes catalyzing the transfer of methyl groups has evolved in nature. These methyltransferases are cofactor-dependent enzymes and *S*-adenosyl-L-methionine (SAM) is by far their main natural methyl group donor.[1] Owing to their participation in many biological processes, their acceptor substrates are manifold, covering a scope from small metabolites to large biopolymers.[2]



Scheme 1: The multi-enzyme cascade couples the biosynthesis of the cofactor to the alkylation of the substrate. It is composed of the L-homoserine *O*-acetyltransferase (HSAT, **A**) from *Deinococcus radiodurans*, the *O*-acetyl-L-homoserine sulfhydrylase (OAHS, **B**) from *Bacillus stearothermophilus* CN3, the methionine adenosyltransferase (MAT, **C**) from *Sulfolobus solfataricus* and the *C*-methyltransferase (MT, **D**) from *Streptomyces rishiriensis*.

[a] MSc L. Offner
Institute of Molecular Biotechnology, Graz University of Technology
Petersgasse 14, A-8010 Graz, Austria
E-mail: lisa.offner@student.tugraz.at
[b] Institute of Organic Chemistry, Graz University of Technology
Stremayrgasse 9, A-8010 Graz, Austria
[c] acib - Austrian Centre of Industrial Biotechnology
Petersgasse 14, A-8010 Graz, Austria
* corresponding author

Supporting information for this article is available.

The small molecule *C*-methyltransferase CouO from *Streptomyces rishiriensis*[3] catalyzes a C-C bond formation by transferring a methyl group to its aromatic substrate in a biocatalytic Friedel-Crafts alkylation reaction. In previous studies, we showed that the methyltransferase is able to act as an alkyltransferase when provided with cofactor analogs as non-natural alkyl group donors.[3] The current state-of-the-art is to synthesize the SAM analogs chemically from *S*-adenosyl-L-homocysteine (SAH) in moderate yields and in 1:1 mixtures of the biologically active (*S*,*S*) diastereomer and the potentially inhibiting (*R*,*S*) diastereomer.[4] SAM and its analogs are inherently instable, impeding the isolation of the active diastereomer and their use as synthetic reagents.[5] By coupling the enzymatic cofactor synthesis to the alkylation of a substrate, the stability, selectivity and activity problems were addressed.[6] However, these approaches require the synthesis of different L-methionine analogs as precursors of the SAM analogs. Using an artificial multi-enzyme cascade, we developed a method that allows a modular cofactor biosynthesis (Scheme 1). Furthermore, avoiding the chemical synthesis of the cofactors and their amino acid precursors renders our biocatalytic Friedel-Crafts alkylation reaction even greener.

The cascade plasmids are derived from the customized pCAS vector backbone.[7] They carry two inducible promoters permitting the induction of the cofactor synthesis and the alkylation enzyme expression at different times. This allows the incorporation of non-canonical amino acids into the *C*-methyltransferase CouO (MT, **D**) which is expressed from a T7 promoter. Engineering this enzyme bears potential to increase its catalytic activity which we were already able to achieve for its homolog NovO.[8] The genes for the enzymatic cofactor synthesis, L-homoserine *O*-acetyltransferase (HSAT, **A**), *O*-acetyl-L-homoserine sulfhydrylase (OAHS, **B**) and methionine adenosyltransferase (MAT, **C**), are organized in a polycistronic operon whose expression is controlled by the p*araBAD* promoter. As we have shown for the transformation of L-α-amino acids to the corresponding enantiopure (*R*)- or (*S*)-α-hydroxy acids,[7] finding the right balance of the individual steps is critical for an ideal flux through the pathway and a high overall yield. One way to harmonize the rate of the steps catalyzed by the co-expressed enzymes (via their concentrations) is to vary the order of the corresponding genes, as their position in the operon influences their expression levels.[9] Therefore, we intended to construct six cascade plasmids using a combinatorial assembly approach but those with the orders BAC and BCA could not be obtained.

The other four plasmids were used for the expression of the multi-enzyme cascade. The first step of the alkylation cascade is the acetylation of L-homoserine **1** catalyzed by the acetyl-CoA dependent enzyme HSAT. Since the cellular concentration of coenzyme A (CoA) is directly associated with the concentration

of its precursor, D-pantothenate in the medium,[10] we supplemented the medium with D-pantothenate and acetate to enhance the acetyl-CoA production.[11] The second step of the pathway is the conversion of *O*-acetyl-L-homoserine **2** to the methionine analogs **4** catalyzed by OAHS which makes the cofactor analog biosynthesis modular, depending on which mercaptan **3** is supplemented. Since the methyltransferases CouO and NovO are insensitive to inhibition by the alkylation reaction's by-product *S*-adenosyl-L-homocysteine (SAH),[3, 12] we did not include an SAH hydrolase in the multi-enzyme cascade.

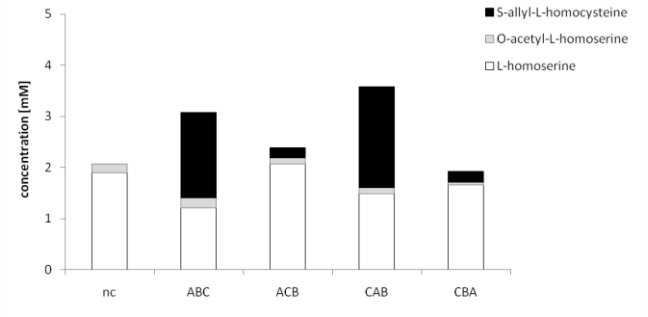


Figure 1: *In vitro* biotransformation with the ABC and CAB alkylation cascade plasmids was more productive with regard to the biosynthesis of S-allyl-L-homocysteine than with the ACB and CBA plasmids. O-acetyl-L-homoserine was also detected in samples carrying the empty backbone vector (nc, negative control).

The biocatalysts for the *in vitro* biotransformation were prepared by expression of the cascade enzymes from the four combinatorial constructs and homogenization of the cells, to avoid mass transport problems. The biosynthesis of the methionine analog **4a** was successful with all cascade plasmids. However, the concentrations reached with the constructs that carry the genes in the orders ACB and CBA were significantly lower than with the ABC and CAB plasmids (Figure 1). Interestingly, cells carrying the empty backbone vector (nc, negative control) were able to produce intermediate **2**, although HSAT was not expressed by these cells. The cofactor analog **5a** could not be detected by HPLC analysis with any of the plasmids. This indicates that the third step, catalyzed by the adenosine triphosphate (ATP) dependent enzyme MAT, may be the bottle neck of the pathway. This reaction creates a cationic sulfonium center at the cofactor that facilitates the Friedel-Crafts alkylation catalyzed by the MT. Phillips *et al.* reported a turnover of 27% **4a** to the corresponding cofactor[13] which is probably even lower when the enzyme is part of a cascade. Furthermore, the MAT originates from the thermophilic organism *Sulfolobus solfataricus* and has a temperature optimum of 65 °C. Therefore, we tested the cascade, at 35 °C, 45 °C, 55 °C, and 65 °C, and exposed samples to gradually increasing temperature (35°C - 65°C for a period of 8 hours), cooled them to 35 °C and added fresh homogenate for 2 more hours of reaction time. Since this temperature screening did not lead to a detectable biosynthesis of the cofactor analog, we tested one of the two cascade plasmids, that carry the MAT gene (**C**) at the first position in the operon, in an *in vivo* biotransformation. We used the CBA alkylation cascade plasmid because we hoped, that the biosynthesis of the methionine analog **4a** could be enhanced in the *in vivo* biotransformation. Due to a lack of the compound as a reference for HPLC analysis, we have not been able to identify the alkylation product **7a** yet.

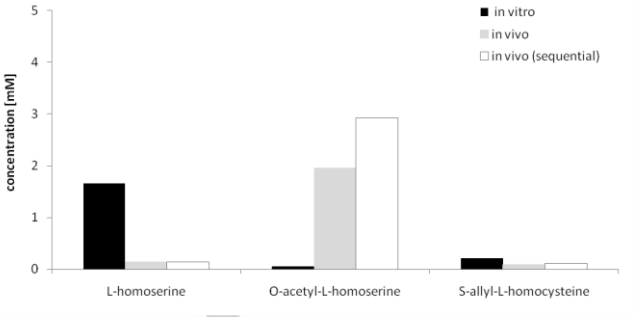


Figure 2: Different bottle necks of the multi-enzyme cascade could be identified for the *in vitro* and *in vivo* biotransformations with the CBA alkylation cascade plasmid.

As illustrated in Figure 2, the bottle neck of the *in vitro* biotransformation with the CBA alkylation cascade plasmid is the acetylation of L-homoserine **1**. This may be due to a lack of the co-substrate acetyl-CoA. Adding D-pantothenate to the medium did not lead to a sufficient overexpression of the enzymes involved in acetyl-CoA biosynthesis for its regeneration from coenzyme A and acetate. However, the addition of D-pantothenate elicited the desired effect in the *in vivo* biotransformation, as most of the starting material **1** was consumed. Here, the biosynthesis of the methionine analog **4a** was clearly the bottle neck of the pathway. Since the mercaptan is volatile, it may have evaporated from the shake flask cultures before it could enter the cells for conversion. This problem may be solved by process engineering, e.g. by applying an oxygen overpressure to increase the solubility of the mercaptan in the medium. Otherwise, the corresponding disulfides could be used to increase the availability of the mercaptan.

By supplementing L-homoserine **1** and allyl mercaptan **3a** before substrate **6** and inducing the expression of the enzymes for cofactor biosynthesis before that of the MT ("sequential"), the production of **2** was increased while the concentration of **4a** in the "sequential" cultures was comparable to that achieved by the "all in" *in vivo* biotransformation (Figure 2). For the "all in" biotransformation **1**, **3a** and **6** were added at the same time and the expression of all four enzymes was induced simultaneously. The expression protocols are described in detail in the supporting information.

Acknowledgements

We gratefully acknowledge the Austrian Research Promotion Agency (FFG) for funding the project.

Keywords: synthetic biology • genetic engineering • cofactor biosynthesis • biocatalysis • C-C bond formation

[1] H. Schönherr, T. Cernak, *Angewandte Chemie International Edition* **2013**, *52*, 12256-12267.

[2] L. Wessjohann, M. Dippe, M. Tengg, M. Gruber-Khadjawi in *Cascade Biocatalysis*, Wiley-VCH Verlag GmbH & Co. KGaA, **2014**, pp. 393-426.

[3] H. Stecher, M. Tengg, B. J. Ueberbacher, P. Remler, H. Schwab, H. Griengl, M. Gruber-Khadjawi, *Angewandte Chemie International Edition* **2009**, *48*, 9546-9548.

[4] C. Dalhoff, G. Lukinavicius, S. Klimasauskas, E. Weinhold, *Nature Chemical Biology* **2006**, *2*, 31-32.

[5] J. L. Hoffman, *Biochemistry* **1986**, *25*, 4444-4449.

[6] a) S. Singh, J. Zhang, T. D. Huber, M. Sunkara, K. Hurley, R. D. Goff, G. Wang, W. Zhang, C. Liu, J. Rohr, S. G. Van Lanen, A. J. Morris, J. S. Thorson, *Angewandte Chemie International Edition* **2014**, *53*, 3965-3969; b) M. Thomsen, S. B. Vogensen, J. Buchardt, M. D. Burkart, R. P. Clausen, *Organic & Biomolecular Chemistry* **2013**, *11*, 7606-7610.

[7] G. Gourinchas, E. Busto, M. Killinger, N. Richter, B. Wiltschi, W. Kroutil, *Chemical Communications* **2015**, *51*, 2828-2831.

[8] L. Offner, B. David, C. Gruber, K. Gruber, B. Wiltschi, *ChemCatChem* **2015**.

[9] H. N. Lim, Y. Lee, R. Hussein, *Proceedings of the National Academy of Sciences* **2011**, *108*, 10626-10631.

[10] A. W. Alberts, R. M. Bell, R. P. Vagelos, *Journal of Biological Chemistry* **1972**, *247*, 3190-3198.

[11] J. E. Cronan, *Journal of Bacteriology* **1980**, *141*, 1291-1297.

[12] P. M. Ueland, *Pharmacological Reviews* **1982**, *34*, 223-253.

[13] F. Wang, S. Singh, J. Zhang, T. D. Huber, K. E. Helmich, M. Sunkara, K. A. Hurley, R. D. Goff, C. A. Bingman, A. J. Morris, J. S. Thorson, G. N. Phillips, *FEBS Journal* **2014**, *281*, 4224-4239.

Supporting Information

Multi-Enzyme Cascade Coupling a Modular Cofactor Biosynthesis to a Biocatalytic Friedel-Crafts Alkylation

Lisa Offner,[a, c] Kathrin Heckenbichler,[b] Patrik Fladischer,[a, c] Mandana Gruber-Khadjawi,[c] Rolf Breinbauer,[b] Birgit Wiltschi*[c]

E-mail: birgit.wiltschi@acib.at

[a] Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, A-8010 Graz, Austria

[b] Institute of Organic Chemistry, Graz University of Technology Stremayrgasse 9, A-8010 Graz, Austria

[c] acib - Austrian Centre of Industrial Biotechnology Petersgasse 14, A-8010 Graz, Austria

Supporting Information

Table of Content

1) DNA and protein sequences of the cascade enzymes

1.1) Homoserine O-acetyltransferase from *Deinococcus radiodurans* (A)

DNA sequence (codon harmonized for *E. coli*):

ATGACCGCCGTGCTCGCGGGCCAtGCCtcagcaCTGCTGCTGACCGAAGAAccggatTGTtca
GGaCCaCAGactGTagtattaTTtcgtcgaGAGCCGCTGCTGttaGAtTGCGGACGaGCGCTGAGC
gatGTGCGaGTtgcaTTTCAtACCtatGGtactCCGCGCgcagatgcaactCTGGTGCTGCAtgcaCTG
ACCGGtgatAGCGCGGTGCAtGAGTGGTGGccggatTTTCTGggtGCGGGtCGGcccCTGGAC
CCGgctgatgattatGTGGTGTGCgcaAACGTCttaggtGGaTGCgcaGGtactactAGCgcaGCTGAAc
ttgcagcaACCTGTTCCGGACCGGTGCCGttaAGCCTGCGCgatATGgcaCGGGTGGGaCGCg
caCTGCTGgactcgttgGGtGTGCGACGGGTGCGGGTCattGGtGCGAGCATGGGtGGaATGtta
gcgtatgcgTGGCTGCTGgaaTGCccggatCTGGTGGAAaaagcgGTGATTATAGGAGCCCCGG
CGCGGCAttcaccgTGGGCTATTGGACTGAACactGCGGCCCGCAGCGCCataGCCttaGCTcc
gGGtGGtgaaGGaCTGaaaGTtGCGCGtCAGataGCaATGCTttcatatCGtAGCccgGAAAGCCTA
AGCCGtactCAGGCGGGaCAGCGtGTtCCGGGaGTtccgGCCgtaacttcatatCTGCAtTACCAAG
GtGAAaagttagcggcgCGCtttgatgaaCAGACCtatTGCGCCttaACCTGGGCGATGgatGCCttcCA
GCCGAGCAGCGCCgatttaaagGCGGTGCGCGCGCCGGTAttaGTCGTCGGtattTCCAGCgac
CTGttatatccgGCCGCCgaaGTCCGCGCCTGCGCCGCCgaaCTTccgCAtGCCgattatTGGGAA
CTGGGtAGCataCAtGGtCAtgatGCCttcctcATGGACcccCAGgatttaCCGgaaCGGGTGGGGgcc
ttcCTCAGGAGTtaa

Protein sequence:

MTAVLAGHASALLLTEEPDCSGPQTVVLFRREPLLLDCGRALSDVRVAFHTYGTPRADATLV
LHALTGDSAVHEWWPDFLGAGRPLDPADDYVVCANVLGGCAGTTSAAELAATCSGPVPLS
LRDMARVGRALLDSLGVRRVRVIGASMGGMLAYAWLLECPDLVEKAVIIGAPARHSPWAIG
LNTAARSAIALAPGGEGLKVARQIAMLSYRSPESLSRTQAGQRVPGVPAVTSYLHYQGEKL
AARFDEQTYCALTWAMDAFQPSSADLKAVRAPVLVVGISSDLLYPAAEVRACAAELPHADY
WELGSIHGHDAFLMDPQDLPERVGAFLRS*

1.1) O-acetylhomoserine sulfhydrylase from *Bacillus stearothermophilus CN3* (B)

DNA sequence (codon harmonized for *E. coli*):

atgagcaatgaacaaactttccgcccggagacgctcgccatccacgccgggcaaaaaccggatgcggaaacgggtgcgcgg
gcggtgccgatttaccaaacaagctcgtatgtattccgcgacagcgagcatgcggccaatttgtttggcttgaaagaggaagggtt
tatttacacgcgcattatgaacccgacaaacgatgtcttggaaaaacggatcgcggcgcttgaaggcggcattggagcgcttgcg
ctgtcatcggggcaggcggcggtgttttattcgatcatcaacatcgcctcggcgggcgatgaaatcgtcagttcgtcatccatttacg
gcggaacgtacaacttgttcgcccatacgctgcgcaagttcggcattacggtgaagtttgtcgatccgtccgatccggaaaactttg
agcgggcgatcaccgacaaaacgaaagcgttgtttgcggaaacgatcggcaacccgaaaaacgatgtgctcgacatcgaag
cggtggccgacatcgcccatcgccatgccattccgctcatcgtagacaacacggtggccagtccatacttattgcggccgattgaa
ttcggtgccgatatcgtcgtccactcagcgacgaagttcatcggcgggcacggcaattcgatcggcggtgtgattgtggacagcg
gcaagttcgactggaaagggagcggcaagtttccggagttcaccgagccagacccaagctaccacgggttggtgtatgtggac
gccgtcggcgaagcggcgtacatcacgaaagcacgcattcagctcttgcgcgacttaggagcggcactgtcgccgtttaatgcgt
ttttgcttttgcaaggattagagacgcttcacttgcggatgcagcgtcatagcgaaaatgcgctcgctgttgctaagtttttagaagag
gaagaagccgtcgaatcggtcaactacccagggcttccgagccatccgtcgcatgaactggcgaaaaagtatttgccgaatgg
gcagggcgcgatcgtcacgtttgaaatcaaaggcggcgtcgaagccgggaaaaaattgatcgactcggtgaagctgttctcgca
tttggccaacatcggcgattcgaaatcgctcatcatccacccagccagcacgacgcatgagcagctgaccccggaagaacagc
tgtccgccggcgtcaccccaggccttgtgcgtctgtctgtcggcaccgaagcgatcgacgatattttggacgacttgcgccaagcc
attcgccaaagccagacggtgggggtgaagtaa

Protein sequence:

MSNEQTFRPETLAIHAGQKPDAETGARAVPIYQTSSYVFRDSEHAANLFGLKEEGFIYTRIM
NPTNDVLEKRIAALEGGIGALALSSGQAAVFYSIINIASAGDEIVSSSSIYGGTYNLFAHTLRKF
GITVKFVDPSDPENFERAITDKTKALFAETIGNPKNDVLDIEAVADIAHRHAIPLIVDNTVASPY
LLRPIEFGADIVVHSATKFIGGHGNSIGGVIVDSGKFDWKGSGKFPEFTEPDPSYHGLVYVD
AVGEAAYITKARIQLLRDLGAALSPFNAFLLLQGLETLHLRMQRHSENALAVAKFLEEEEAVE
SVNYPGLPSHPSHELAKKYLPNGQGAIVTFEIKGGVEAGKKLIDSVKLFSHLANIGDSKSLIIH
PASTTHEQLTPEEQLSAGVTPGLVRLSVGTEAIDDILDDLRQAIRQSQTVGVK*

1.1) Methionine adenosyltransferase from *Sulfolobus solfataricus* (C)

DNA sequence (codon harmonized for *E. coli*):

ATGcgcAATatcAATGTGCAATTAAATCCCCTTTCAGATATAGAGAAACTTCAAGTAGAACT
AGTAGAGAGAAAAGGATTAGGTCATCCAGATTATATTGCGGATGCAGTTGCTGAAGAAG
CTAGCAGAAAGttaTCCTTGTATTATCTTAAAAAATATGGTGTGATTTTACATCATAATTTAG
ATAAGACATTGGTGGTTGGAGGACAAGCTACACCTCGTTTTAAAGGTGGAGATATAATA
CAACCAATATATATCATAGTTGCGGGCAGAGCTACGACTGAGGTAAAAACAGAAAGCGG
GATAGACCAAATCCCCGTAGGTACTATCATTATAGAGAGCGTAAAAGAGTGGATCAGAA
ATAACTTTAGATATCTCGATGCGGAAAGGCATGTTATAGTTGACTATAAGATAGGCAAAG
GTTCAAGTGATTTAGTGGGTATATTTGAGGCAAGCAAGAGAGTTCCACTATCTAATGATA
CTAGTTTTGGAGTAGGTTTTGCTCCATTAACTAAATTAGAAAAACTGGTTTATGAAACGG
AGAGGCATTTGAACTCAAAGCAATTCAAAGCCAAACTACCCGAAGTAGGAGAAGATATC
AAAGTAATGGGGTTAAGAAGAGGGAATGAGGTAGACCTTACGATAGCGATGGCAACAAT
TAGCGAACTAATAGAAGACGTTAACCACTATATTAACGTAAAAGAACAAGTAAGGAATCA
AatcctcGATCTCGCATCAAAGATAGCTCCAGGCTACAATGTAAGAGTTTATGTTAATACTG
GAGATAAAATAGATAAGAATATACTTTATTTAACCGTAACTGGTACTTCTGCTGAACATG
GTGACGATGGAATGACAGGGAGAGGAAATAGAGGTGTTGGGCTAATAACACCAATGAG
GCCTATGTCATTAGAAGCTACTGCTGGAAAGAATCCCGTTAATCATGTTGGTAAACTATA
TAATGTCTTAGCTAATCTTATAGCTAATAAAATAGCTCAAGAAGTAAAGGATGTGAAATTC
TCGCAAGTTCAAGTTCTAGGGCAAATAGGGAGACCAATAGACGATCCTTTAATAGCTAAT
GTTGATGTAATTACTTATGATGGCAAACTTACTGATGAGACTAAAAATGAAATAAGCGGG
ATTGTAGATGAAATGTTAAGTTCCTTCAATAAATTAACTGAACTAATATTAGAAGGGAAAG
CTACTCTCTTTTAA

Protein sequence:

MRNINVQLNPLSDIEKLQVELVERKGLGHPDYIADAVAEEASRKLSLYYLKKYGVILHHNLDK
TLVVGGQATPRFKGGDIIQPIYIIVAGRATTEVKTESGIDQIPVGTIIIESVKEWIRNNFRYLDAE
RHVIVDYKIGKGSSDLVGIFEASKRVPLSNDTSFGVGFAPLTKLEKLVYETERHLNSKQFKAK
LPEVGEDIKVMGLRRGNEVDLTIAMATISELIEDVNHYINVKEQVRNQILDLASKIAPGYNVRV
YVNTGDKIDKNILYLTVTGTSAEHGDDGMTGRGNRGVGLITPMRPMSLEATAGKNPVNHVG
KLYNVLANLIANKIAQEVKDVKFSQVQVLGQIGRPIDDPLIANVDVITYDGKLTDETKNEISGIV
DEMLSSFNKLTELILEGKATLF*

1.2) *C*-methyltransferase CouO from *Streptomyces rishiriensis*

DNA sequence:

atgaagattgaaccgattacgggatcagaggccgaagcctttcaccggatgggctcccgggcgttcgaaaggtacaacgagttc
gtggacctgctggtcggcgccggcatcgccgatggccagaccgtggtcgacctgtgttgtggttcgggggagctggagatcatcc
tcacctcccggttcccatcgctgaacctggtcggcgtcgacctctcggaggacatggtccgaattgcccgggactatgccgccga
gcagggcaaggagctggagttccgtcacggggacgcgcagtccccggccggcatggaggatctgctcggaaaagcggatct
ggtggtctcgcggcacgcgttccaccggctcacccggctcccggcggggttcgacaccatgttgcgcctggtcaagccaggggg
agccatcctcaacgtctccttcctgcatctcagcgacttcgacgagcccgggttccggacgtgggttcggtttctgaaggagcggcc
ttgggacgcggagatgcaggtcgcctgggcgctggctcactactacgcgccccggctccaggactaccgcgatgccctggccc
aggcggccgatgagacgccggtgtccgagcagagaatctgggttgatgaccagggctacggcgttgccacggtcaagtgttttg
cccgtcgggccgctgcctga

Protein sequence:

MKIEPITGSEAEAFHRMGSRAFERYNEFVDLLVGAGIADGQTVVDLCCGSGELEIILTSRFPS
LNLVGVDLSEDMVRIARDYAAEQGKELEFRHGDAQSPAGMEDLLGKADLVVSRHAFHRLTR
LPAGFDTMLRLVKPGGAILNVSFLHLSDFDEPGFRTWVRFLKERPWDAEMQVAWALAHYY
APRLQDYRDALAQAADETPVSEQRIWVDDQGYGVATVKCFARRAAA*

2) Construction of the cascade plasmids

The cascade genes L-homoserine *O*-acetyltransferase (A), *O*-acetyl-L-homoserine sulfhydrylase (B) and methionine adenosyltransferase (C) were ordered as synthetic gBlocks (Integrated DNA Technologies, Coralville, IA). For sequence verification the gBlocks were cloned into pJET1.2 vectors (Thermo Fisher Scientific Inc., Waltham, MA) by blunt end ligation following the protocol of the manufacturer and amplified in *E. coli* TOP10F' cells. Vectors with the correct gene sequences served as templates for the PCR amplification of the genes. The other parts for the pCAS2 alkylation cascade plasmids were amplified from pCAS2[1] and from pET26b(+)-couO[2]. The primers (Integrated DNA technologies (IDT)) for the PCR reactions were designed to introduce terminal 40 bp homologous overlaps to the adjacent parts which is required for Gibson isothermal assembly. The PCR reactions contained 10 ng of plasmid template DNA, 0.5 µM of the forward and reverse primers, 0.2 mM of each deoxyribonucleotide triphosphate, 1X *Phusion* HF Buffer and 1 U *Phusion* high-fidelity DNA Polymerase (Thermo Fisher Scientific Inc., Waltham, MA) in a total volume of 50 µL. The PCR program included an initial denaturation (98 °C for 30 sec), followed by 30 cycles of denaturation (98 °C for 10 sec), annealing (56-60 °C, depending on the melting temperature, for 20 sec) and elongation (72 °C for 1-3 min, depending on the length) and the final elongation (72 °C for 7 min).

For the assembly of the alkylation cascade plasmids, Gibson isothermal assembly was applied. We followed the method described by Gibson *et al.*.[3] 50 ng of the smallest fragment and all other fragments in equimolar quantities were added to 7.5 µL of the assembly master mix which is composed of 10 U/µL T5 exonuclease (Biozym Scientific GmbH, Hessisch Oldendorf, Germany), 2 U/µL *Phusion* high-fidelity DNA Polymerase (Thermo Fisher Scientific Inc., Waltham, MA), 40 U/µL *Taq* DNA ligase (New England BioLabs Inc., MA) and 1X isothermal (ISO) reaction buffer (5% (w/v) PEG-8000, Tris-HCl (100 mM, pH 7.5), 10 mM $MgCl_2$, 10 mM DTT, 0.2 mM of each deoxyribonucleotide triphosphate, 1 mM NAD in water). The mixture was incubated at 50 °C for 60 min without shaking. The resulting plasmids were amplified in *E. coli* TOP10F' cells on LB agar plates containing 100 µg/mL ampicillin and they were sequence verified (Microsynth, Balgach, Switzerland).

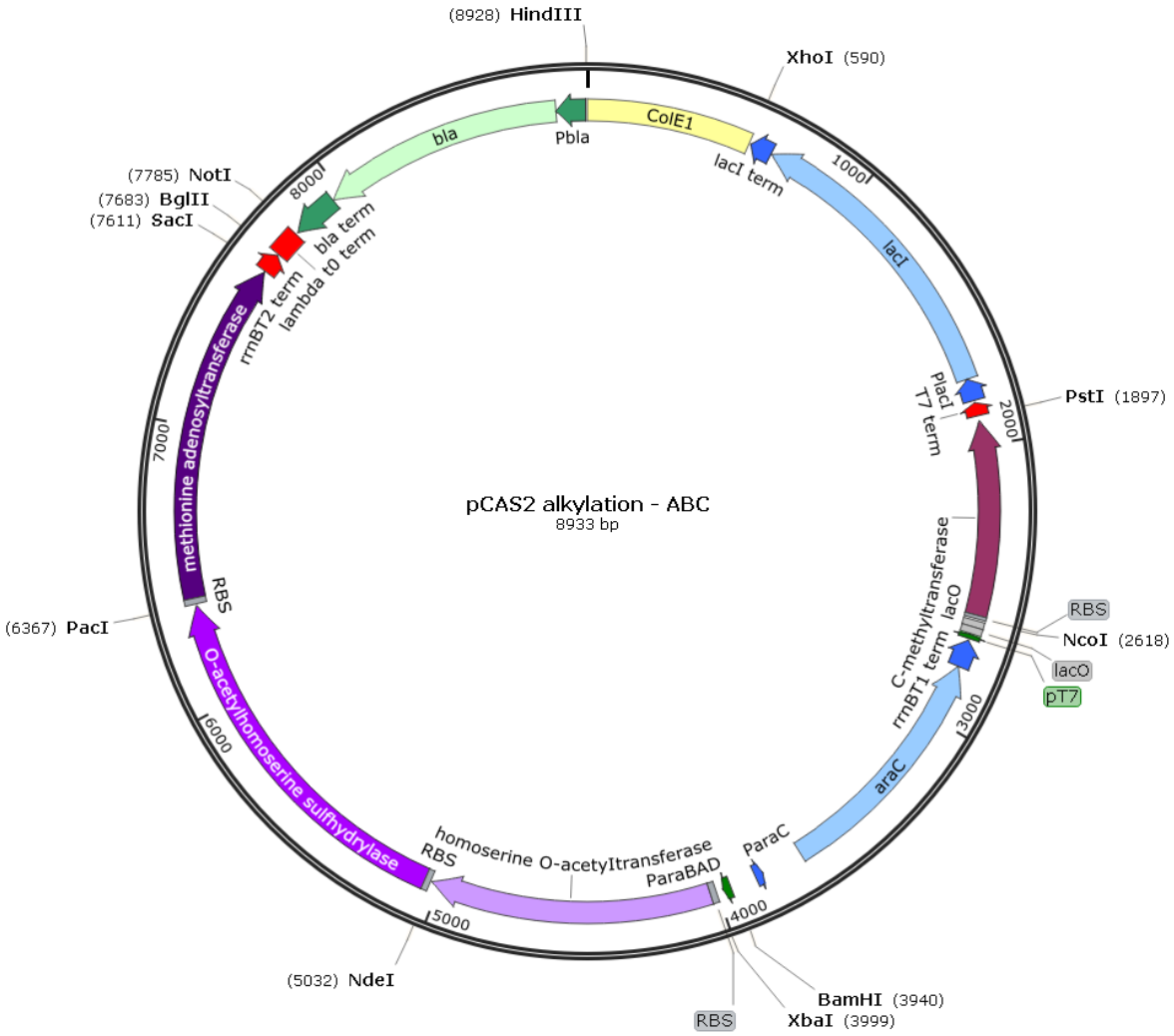


Figure S1: Map of the alkylation cascade plasmid pCAS2 alkylation ABC. The L-homoserine *O*-acetyltransferase (**A**), *O*-acetyl-L-homoserine sulfhydrylase (**B**) and methionine adenosyltransferase (**C**) genes are under the control of an arabinose-inducible p*araBAD* promoter while the *C*-methyltransferase (CouO) gene is controlled by an IPTG-inducible T7/*lacO* promoter (pT7 and *lacO*). All cascade genes carry the same RBS sequence from the pET21a(+) plasmid. The plasmid carries the genes in the polycistronic operon in the order ABC. All constructs feature a T7 terminator (T7 term) and an *rrnB*T2 terminator (*rrnB*T2 term), the lac repressor (*lacI*) with a lac repressor promoter (p*lacI*) and a lac repressor terminator (*lacI* term), the ara repressor (*araC*) with an ara repressor promoter (p*araC*) and an *rrnB*T1 terminator (*rrnB*T1 term) and the β-lactamase gene (*bla*) with a β-lactamase promoter (p*bla*), a β-lactamase terminator (*bla* term) and a λt0 terminator (lambda t0 term).

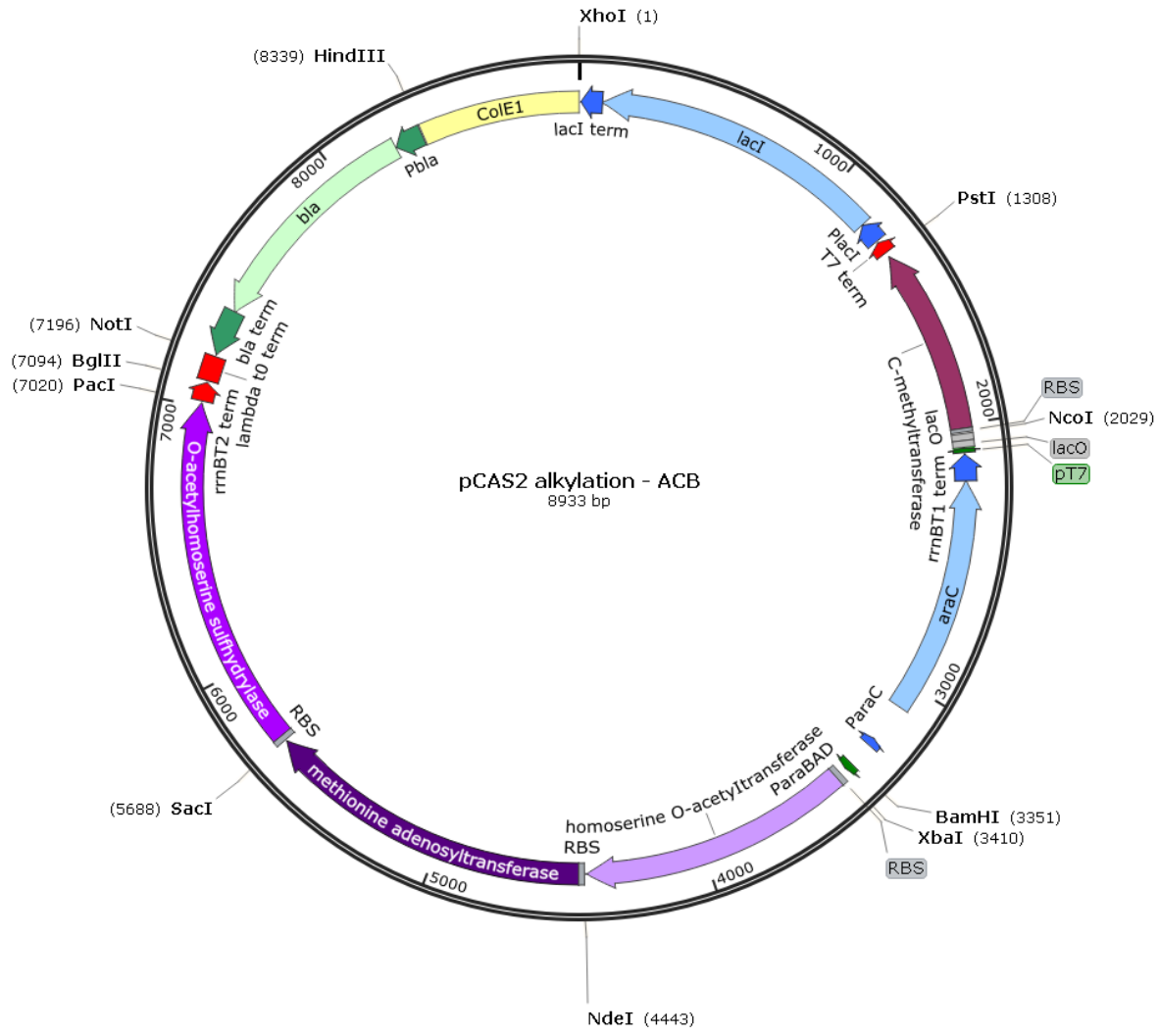


Figure S2: Map of the alkylation cascade plasmid pCAS2 alkylation ACB. The plasmid carries the genes in the polycistronic operon in the order ACB.

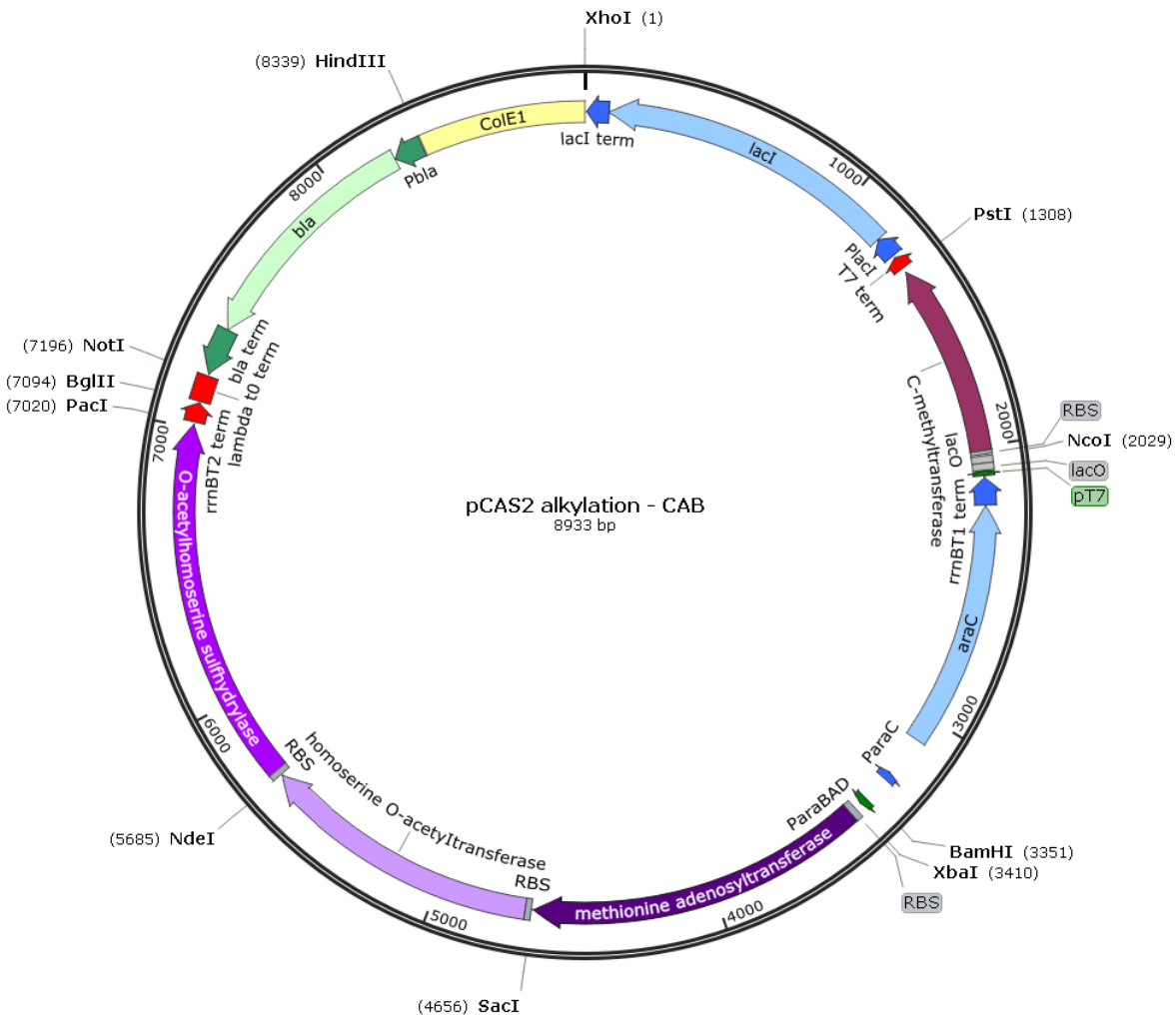


Figure S3: Map of the alkylation cascade plasmid pCAS2 alkylation CAB. The plasmid carries the genes in the polycistronic operon in the order CAB.

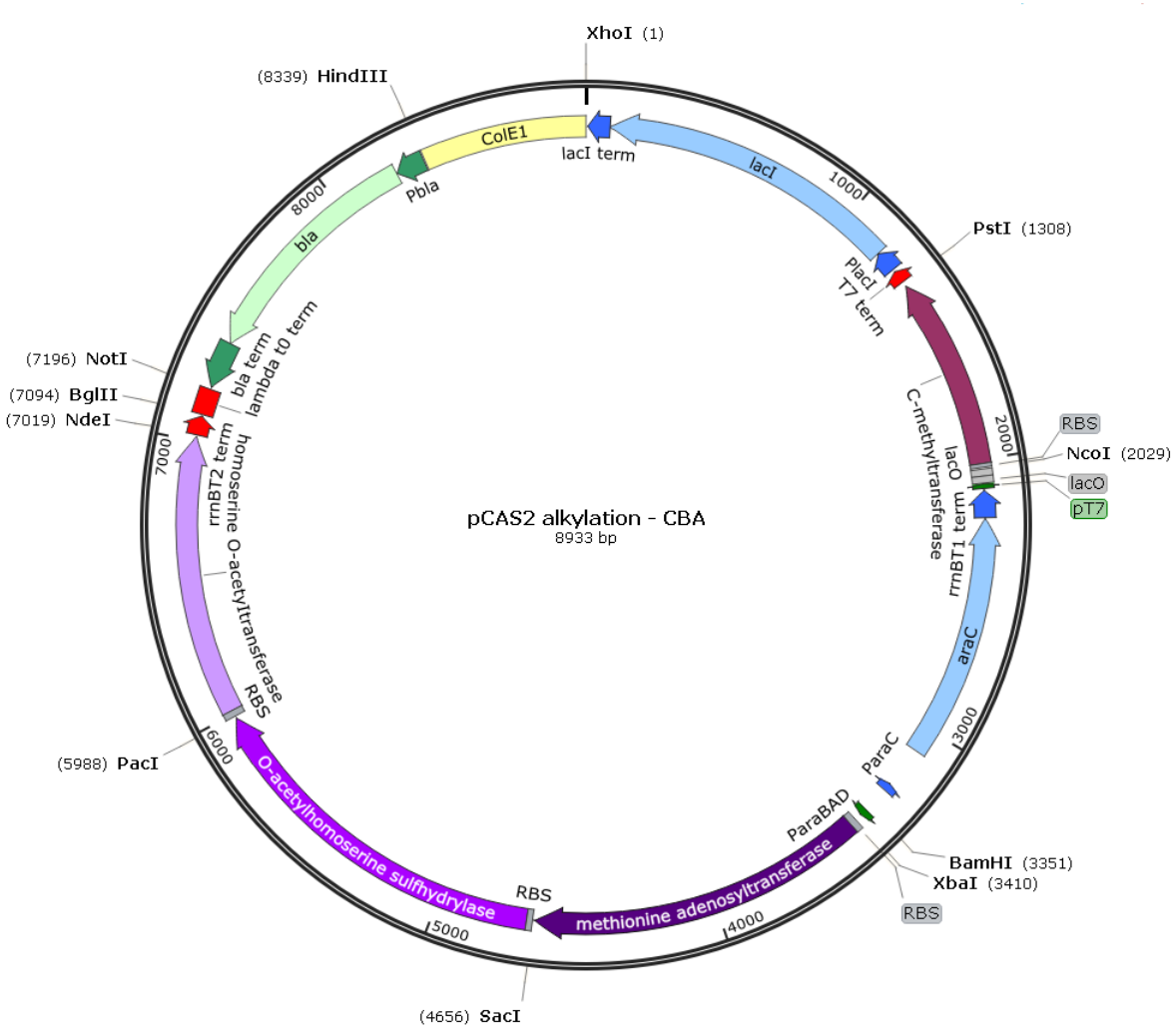


Figure S4: Map of the alkylation cascade plasmid pCAS2 alkylation CBA. The plasmid carries the genes in the polycistronic operon in the order CBA.

3) Supplementary Methods

3.1) Expression of the multi-enzyme cascade for the *in vitro* biotransformation

Chemocompetent *E. coli* BL21 (DE3) Gold cells were transformed with one of the alkylation cascade plasmids or pCAS2 (empty vector backbone) and incubated over night in LB medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) containing 100 µg/mL ampicillin. 250 mL pre-warmed LB-amp medium supplemented with 0.01% (w/v) D-pantothenate and 0.01% (w/v) acetate were inoculated to an initial D_{600} of 0.1. The cultures were shaken at 37 °C until they reached a D_{600} of 0.6-0.8. Expression of all four enzymes was induced with 0.2% (w/v) L-arabinose (A, B and C) and 1 mM IPTG (CouO) and performed for 16 h at 28 °C shaking. To analyze the expression by SDS-PAGE, culture samples were collected before induction (ni) and 4 h and 16 h after induction (Figure S5). After the expression, the cells were harvested by centrifugation at 4000 rpm and 4°C for 20 min, suspended in phosphate buffer (50 mM, pH 6.5) and lysed by sonication (6 min, output 80%). The homogenates were used for the activity assay.

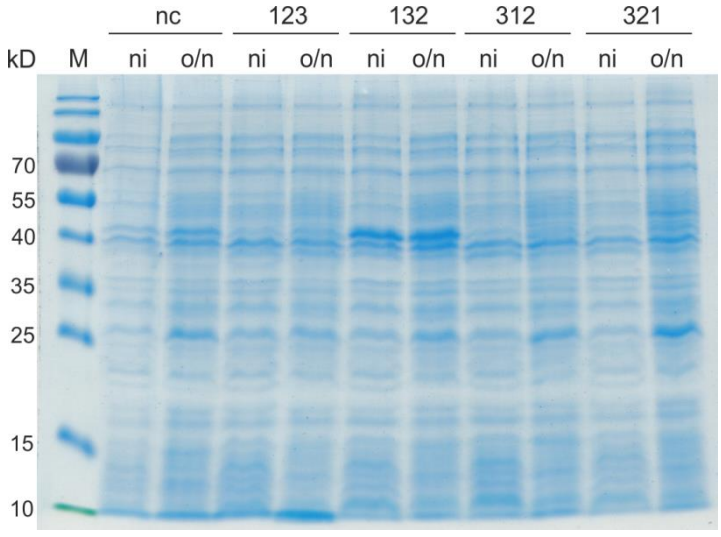


Figure S5: Expression of the cascade enzymes, homoserine *O*-acetyltransferase (1, 35 kD), *O*-acetylhomoserine sulfhydrylase (2, 47 kD), methionine adenosyltransferase (3, 45 kD) and *C*-methyltransferase CouO (26 kD). 12 % SDS-polyacrylamide gel was loaded with the total protein of cells which were transformed with the different alkylation cascade plasmids. Samples were collected before induction (ni) and 16 h after induction (o/n) (10 µL, normalized to a D_{600} of 0.5). M, molecular weight marker (5 µl Page Ruler^{TM} Prestained Protein Ladder, Thermo Fisher Scientific Inc., Waltham, MA). Coomassie Brilliant Blue stain.

3.2) *In vitro* biotransformation

The reaction was carried out in 500 µL scale in an Eppendorf thermomixer at 35 °C (and 45 °C, 55 °C and 65 °C for a temperature screening) and 1000 rpm for 3 h. The reaction mix contained 450 µL homogenate prepared in 50 mM phosphate buffer pH 6.5 (reaction buffer), 1 mM substrate **6** (from a 100 mM solution in dimethyl sulfoxide for full solubility), 5 mM or 0.06% (w/v) L-homoserine (from a 500 mM aqueous solution), 0.05% (w/v) D-pantothenate (from a 1% (w/v) aqueous solution), 0.05% (w/v) acetate (from a 10% (w/v) aqueous solution), 5 mM allyl mercaptan and 5 mM ATP (from a 500 mM aqueous solution). The reaction was started by adding the homogenate solution to the reaction mix and stopped by heating it to 90°C for 10 min. The denatured protein was removed by centrifugation at 13200 rpm for 10 min.

Analysis of the *in vitro* biotransformation samples was performed on an Agilent Technologies HPLC-System 1200 equipped with a Merck Millipore Purospher® STAR RP-18e (5 µm) 250-4.6 column. The components were separated applying a 0.7 mL/min flow rate of 0.1% formic acid (A) and acetonitrile (B). The following gradient was applied: 0-0.5 min 2% B, 0.5-9 min to 98% B, 9-10 min to 2% B. The mass spectrometer (MS-ESI positive mode) was set to scan mode (100-600 m/z) with single ion monitoring (SIM). Chemically synthesized L-homoserine, *O*-acetyl-L-homoserine and *S*-allyl-L-homocysteine served as reference substances (Figure 1 and Figure 2).

3.3) Expression of the multi-enzyme cascade for the *in vivo* biotransformation

Chemocompetent *E. coli* BL21 (DE3) Gold cells were transformed with pCAS2 alkylation CBA or pCAS2 (empty vector backbone) and incubated over night in LB medium containing 100 µg/mL ampicillin. 50 mL pre-warmed LB-amp medium supplemented with 0.05% (w/v) D-pantothenate were inoculated to an initial D_{600} of 0.1. The cultures were shaken at 37 °C until they reached D_{600} ~ 1. All cultures were supplemented with 5 mM L-homoserine and 5 mM allyl mercaptan and expression of the cofactor biosynthesis enzymes, homoserine *O*-acetyltransferase, *O*-acetylhomoserine sulfhydrylase and methionine adenosyltransferase, was induced with 0.2% (w/v) L-arabinose. Additionally, the "all in" cultures were supplemented with 2 mM substrate **6** and expression of the methyltransferase was induced with 1 mM IPTG. After 2 h shaking at 28 °C, the "sequential" cultures were also supplemented with 2 mM substrate **6** and expression of the methyltransferase was induced with 1 mM IPTG. 6 h after the induction of the "sequential" cultures, 5 mM L-homoserine and 5 mM allyl mercaptan were added. Expression was performed for 20 h at 28 °C shaking.

To analyze the *in vivo* biotransformation by HPLC (Figure 1 and Figure 2), whole-cell samples were collected at the end of the expression and frozen at -20 °C. For analysis, they were thawed and incubated in an equal volume of 75% ethanol at 80 °C and 1000 rpm for 10 min. The samples were cooled on ice for 10 min and centrifuged at 13000 rpm and 4 °C for 15 min. The cell extracts were directly used for HPLC analysis.

HPLC analysis of the *in vivo* biotransformation samples was performed as described for the *in vitro* biotransformation samples (3.2)).

4) Supplementary References

[1] G. Gourinchas, E. Busto, M. Killinger, N. Richter, B. Wiltschi, W. Kroutil, *Chemical Communications* **2015**, *51*, 2828-2831.

[2] H. Stecher, M. Tengg, B. J. Ueberbacher, P. Remler, H. Schwab, H. Griengl, M. Gruber-Khadjawi, *Angewandte Chemie International Edition* **2009**, *48*, 9546-9548.

[3] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, *Nature Methods* **2009**, *6*, 343-345.

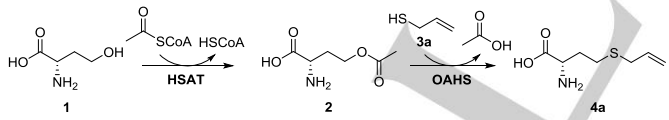
BIOSYNTHESIS AND INCORPORATION OF ALKENE-CONTAINING AMINO ACIDS

Biosynthesis and Incorporation of Alkene-Containing Amino Acids**

Lisa Offner,[a,b] Patrik Fladischer,[a,b] Birgit Wiltschi*[b]

Abstract: We developed a two-enzyme cascade for the modular biosynthesis of non-canonical amino acids which enable proteins to take part in bio-orthogonal reactions, e.g. olefin metathesis. By supplementing the cultures with allyl mercaptan, the alkene-containing amino acid *S*-allyl-L-homocysteine (AHcy) could be synthesized and incorporated into a target protein. Furthermore, we were able to functionalize the protein with 2-amino-5-hexenoic acid (AHA). However, we were not able to detect multimers of the alkene-functionalized proteins, Fdx[AHA] and Fdx[AHcy] after the olefin metathesis reaction by SDS-PAGE.

Recently, non-canonical amino acids (ncAA) featuring C-C double bonds have gotten into the focus of the synthetic biology community, as they enable proteins to participate in bio-orthogonal reactions. The reactive ncAA were synthesized by chemically modification of canonical amino acids (cAA) on the protein surface[1] or incorporated by supplementation of the ncAA in the course of protein biosynthesis.[2] Subsequently, the alkene moieties were coupled with ligands or other amino acid side chains containing reactive chemistries by different reactions, e.g. thiol-ene click chemistry[3] or Diels-Alder reactions[4]. Special attention was devoted to olefin metathesis which was applied for the post-translational modification of alkene-functionalized proteins with other olefins,[1, 5] their copolymerization with acrylamide[6] and the intramolecular coupling of two alkene-containing amino acids in peptides and small proteins, resulting in improved stability.[2c, 7]



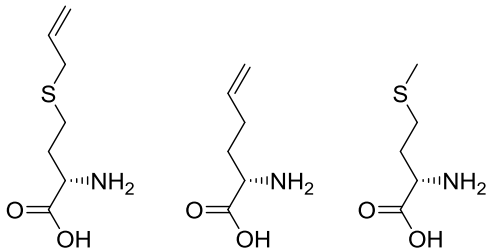
Scheme 1. The two-enzyme cascade couples the biosynthesis of the methionine analogs to their incorporation into proteins. It is composed of the L-homoserine *O*-acetyltransferase (HSAT) from *Deinococcus radiodurans* and the *O*-acetyl-L-homoserine sulfhydrylase (OAHS) from *Bacillus stearothermophilus* CN3.

[a] MSc L. Offner
Institute of Molecular Biotechnology, Graz University of Technology
Petersgasse 14, A-8010 Graz, Austria
E-mail: lisa.offner@student.tugraz.at
[b] acib - Austrian Centre of Industrial Biotechnology
Petersgasse 14, A-8010 Graz, Austria
* corresponding author
** L. Offner and P. Fladischer contributed equally to this work.

Supporting information for this article is available.

We based the present study on the pioneering work of Davis *et al*.[1] who showed that allyl thioether-functionalized proteins are reactive substrates for cross metathesis using the ruthenium-based Hoveyda-Grubbs second generation catalyst. They stated that the unique reactivity of allyl sulfides in the metathesis reaction is due to sulfur coordination to the ruthenium center. Furthermore, they expanded the substrate scope of olefin metathesis to alkene-containing proteins by developing a reaction protocol in aqueous-organic solution. While they chemically converted a cysteine residue on the protein surface into *S*-allyl-L-cysteine, its higher homolog *S*-allyl-L-homocysteine was supplemented and incorporated residue-specifically into proteins in *E. coli* (Scheme 2).[8] However, the latter method requires the chemical synthesis of the supplemented amino acids which inevitably involves protection and deprotection steps.

By genetic engineering, we developed a greener protocol for the *in vivo* biosynthesis of *S*-allyl-L-homocysteine **4a** starting from L-homoserine **1** in a modular two-enzyme cascade which allows us to synthesize also other methionine analogs (Scheme 1). This intracellular production of the non-canonical amino acids was coupled to their residue-specific incorporation into proteins in a methionine-auxotrophic *E. coli* strain. Furthermore, we applied the residue-specific incorporation method to incorporate the methionine analog, 2-amino-5-hexenoic acid **5**, which was synthesized in a multi-step *in vitro* biotransformation from 5-hexenoic acid by our co-workers (Scheme 2). With these alkene-functionalized proteins, we intended to exploit the potential of olefin metathesis for the intermolecular coupling and hence, the polymerization of proteins and the production of biopolymers.



Scheme 2: Structures of the non-canonical amino acids *S*-allyl-L-homocysteine (AHcy) **4a** and 2-amino-5-hexenoic acid (AHA) **5**, and the canonical amino acid L-methionine **6**.

The cascade plasmid, pCAS5 Met-analog_Fdx is derived from the customized pCAS2 vector backbone.[9] To allow sequential induction of the amino acid biosynthesis and the target gene expression, we put the genes under the control of an arabinose-

inducible p*araBAD* promoter and IPTG-inducible T5 promoter, respectively. This strategy gives the cells time for the *in vivo* production of the ncAA that is then incorporated into the target protein.

The first step of the two-enzyme cascade is the acetylation of L-homoserine **1** catalyzed by the acetyl-CoA dependent enzyme HSAT. The cellular concentration of coenzyme A (CoA) is directly associated with the concentration of its precursor, D-pantothenate in the medium.[10] Therefore, we supplemented the medium of cultures B, C and D with D-pantothenate and acetate to enhance the acetyl-CoA production.[11] Furthermore, we constructed a plasmid (pAcetylCoA) which carries the acetyl-CoA synthetase gene and allows its co-expression in cultures C and D. Cells in culture A were neither transformed with pAcetylCoA nor supplemented with D-pantothenate and acetate, and served as a control. The conditions for the four cultures A-D are summarized in table 1.

The second step of the cascade is the conversion of *O*-acetyl-L-homoserine **2** to the methionine analogs **4** catalyzed by OAHS which makes their biosynthesis modular, depending on which nucleophile **3** is supplemented. This enzyme has already been successfully used for the *in vivo* synthesis of L-azidohomoalanine, starting from chemically synthesized *O*-acetyl-L-homoserine **2,** and its incorporation replacing L-methionine **6**.[12] For the biosynthesis of *S*-allyl-L-homocysteine **4a**, the cultures were supplemented with 5 mM L-homoserine **1** and 5 mM allyl mercaptan **3a** (A, B and D) or 1 mM **1** and 1 mM **3a** (C).

Table 1. All four cultures A-D were transformed with pCAS5 Met-analog_Fdx. Additional supplements are given in this table.

	pAcetylCoA	pantothenate	acetate	L-homoserine	allyl mercaptan
A	-	-	-	5 mM	5 mM
B	-	+	+	5 mM	5 mM
C	+	+	+	1 mM	1 mM
D	+	+	+	5 mM	5 mM

The amino acid biosynthesis was analyzed by TLC and HPLC-MS. According to TLC, the concentration of L-homoserine **1** decreased and the *S*-allyl-L-homocysteine **4a** concentration increased in all cultures over the expression time. While the cultures A and C produced mainly *O*-acetyl-L-homoserine **2**, *S*-allyl-L-homocysteine **4a** was the main product in culture B at the end of the expression. The concentrations of **2** and **4a** in culture D were comparable. The start concentration of **1** in culture C was too low, as all **1** was already completely converted to **2** in the beginning of the expression.

The results of TLC analysis were partly confirmed by HPLC-MS analysis, but Figure 1 illustrates that *O*-acetyl-L-homoserine **2** was the main product in all cultures and that the concentration of *S*-allyl-L-homocysteine **4a** did not increase significantly throughout the expression, while most of the L-homoserine **1** was consumed at the end. This indicates that regeneration of acetyl-CoA from D-pantothenate and acetate work well and that the biosynthesis of the methionine analog **4a** was clearly the bottle neck of the two-enzyme cascade. The allyl mercaptan **3a** is highly volatile, it may have evaporated from the shake flask cultures before it could enter the cells for conversion. This problem may be solved by process engineering, e.g. by applying an oxygen overpressure to increase the solubility of the mercaptan in the medium.

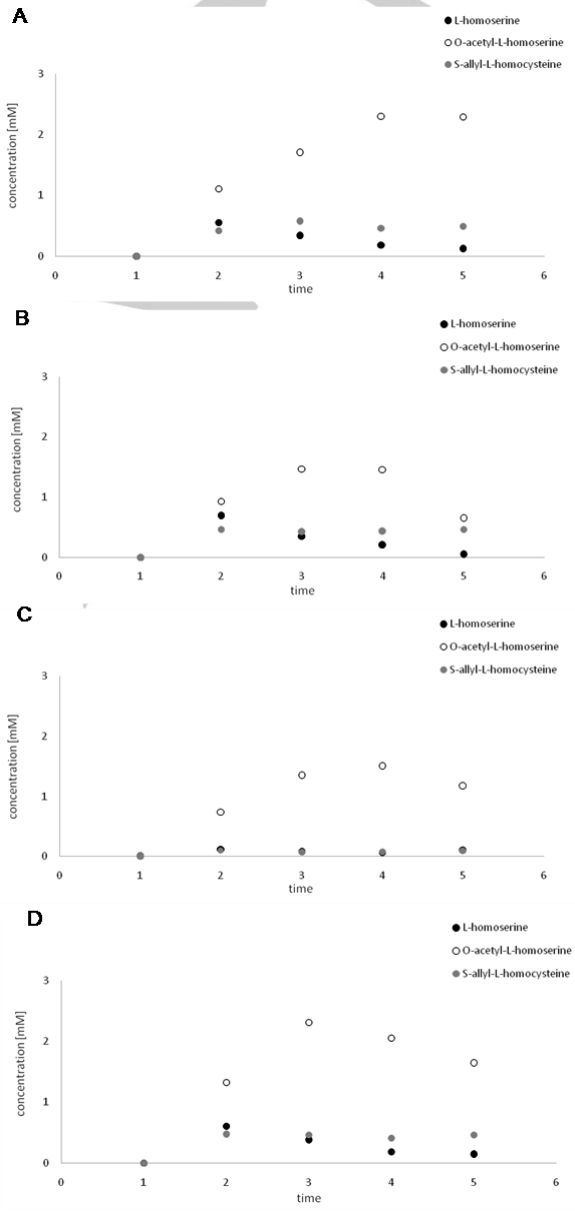


Figure 1. HPLC-MS analysis of the amino acid biosynthesis indicates that the second step of the two-enzyme cascade catalyzed by OAHS is its bottle neck.

According to mass analysis *S*-allyl-L-homocysteine (AHcy) **4a** and 2-amino-5-hexenoic acid (AHA) **5** were incorporated into Fdx. However, we were not able to detect multimers of the alkene-functionalized proteins, Fdx[AHA] and Fdx[AHcy] after

the olefin metathesis reaction by SDS-PAGE. Although allyl sulfide-functionalized proteins were identified to be reactive substrates for olefin metathesis by Davis *et al.*,[1] attempts to perform the reaction with the incorporated alkene-containing amino acids have not only been unsuccessful in our hands.[8] From our experiments, we could not conclude if the reaction did not work due to the double bonds not being accessible for the catalyst or if the reaction worked but SDS-PAGE was not the appropriate detection method. Therefore, we intend to analyze the samples by mass spectrometry.

Acknowledgements

We gratefully acknowledge Kathrin Heckenbichler and Prof. Rolf Breinbauer for providing the starting material, L-homoserine and the reference substances for HPLC analysis.

Keywords: synthetic biology • protein engineering• amino acid biosynthesis • olefin metathesis • biopolymer production

[1] Y. A. Lin, J. M. Chalker, N. Floyd, G. J. L. Bernardes, B. G. Davis, *Journal of the American Chemical Society* **2008**, *130*, 9642-9643.

[2] a) J. C. M. van Hest, K. L. Kiick, D. A. Tirrell, *Journal of the American Chemical Society* **2000**, *122*, 1282-1288; b) Z. Zhang, L. Wang, A. Brock, P. G. Schultz, *Angewandte Chemie International Edition* **2002**, *41*, 2840-2842; c) H.-w. Ai, W. Shen, E. Brustad, P. G. Schultz, *Angewandte Chemie International Edition* **2010**, *49*, 935-937.

[3] N. Floyd, B. Vijayakrishnan, J. R. Koeppe, B. G. Davis, *Angewandte Chemie International Edition* **2009**, *48*, 7798-7802.

[4] M. L. Blackman, M. Royzen, J. M. Fox, *Journal of the American Chemical Society* **2008**, *130*, 13518-13519.

[5] R. S. Al Toma, A. Kuthning, M. P. Exner, A. Denisiuk, J. Ziegler, N. Budisa, R. D. Süssmuth, *ChemBioChem* **2015**, *16*, 503-509.

[6] J. Xiao, T. J. Tolbert, *Biomacromolecules* **2009**, *10*, 1939-1946.

[7] a) L. K. Henchey, A. L. Jochim, P. S. Arora, *Current Opinion in Chemical Biology* **2008**, *12*, 692-697; b) Y.-W. Kim, T. N. Grossmann, G. L. Verdine, *Nat. Protocols* **2011**, *6*, 761-771.

[8] F. Truong, PhD Thesis thesis, California Institute of Technology **2013**.

[9] G. Gourinchas, E. Busto, M. Killinger, N. Richter, B. Wiltschi, W. Kroutil, *Chemical Communications* **2015**, *51*, 2828-2831.

[10] A. W. Alberts, R. M. Bell, R. P. Vagelos, *Journal of Biological Chemistry* **1972**, *247*, 3190-3198.

[11] J. E. Cronan, *Journal of Bacteriology* **1980**, *141*, 1291-1297.

[12] Y. Ma, H. Biava, R. Contestabile, N. Budisa, M. di Salvo, *Molecules* **2014**, *19*, 1004-1022.

Supporting Information

Biosynthesis and Incorporation of Alkene-Containing Amino Acids

Lisa Offner,[a, b] Patrik Fladischer,[a, b] Birgit Wiltschi*[b]

E-mail: birgit.wiltschi@acib.at

[a] Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, A-8010 Graz, Austria

[b] acib - Austrian Centre of Industrial Biotechnology Petersgasse 14, A-8010 Graz, Austria

Supporting Information

Table of Content

1) DNA and protein sequences

1.1) Flavodoxin (Fdx) from *Clostridium beijerinckii*

DNA sequence (codon optimized for *E. coli*[1]):

ATGAAAATCGTATATTGGTCTGGTACCGGCAACACTGAGAAAATGGCAGAGCTCATCGC
TAAAGGTATCATCGAATCTGGTAAAGACGTCAACACCATCAACGTGTCTGACGTTAACAT
CGATGAACTGCTGAACGAAGATATCCTGATCCTGGGTTGCTCTGCCATGGGCGATGAA
GTTCTCGAGGAAAGCGAATTTGAACCGTTCATCGAAGAGATCTCTACCAAAATCTCTGG
TAAGAAGGTTGCGCTGTTCGGTTCTTATGGTTGGGGCGACGGTAAGTGGATGCGTGAC
TTCGAAGAACGTATGAACGGCTATGGTTGCGTTGTTGTTGAGACCCCGCTGATCGTTCA
GAACGAGCCGGACGAAGCTGAGCAGGACTGCATCGAATTTGGTAAGAAGATCGCGAAC
ATC

Protein sequence (The methionine residues are highlighted in bold.)

MKIVYWSGTGNTEK**M**AELIAKGIIESGKDVNTINVSDVNIDELLNEDILILGCSA**M**GDEVLEES
EFEPFIEEISTKISGKKVALFGSYGWGDGKW**M**RDFEER**M**NGYGCVVVETPLIVQNEPDEAE
QDCIEFGKKIANI

1.2) Homoserine O-acetyltransferase from *Deinococcus radiodurans*

DNA sequence (codon harmonized for *E. coli*):

ATGACCGCCGTGCTCGCGGGCCAtGCCtcagcaCTGCTGCTGACCGAAGAAccggatTGTtca
GGaCCaCAGactGTagtattaTTtcgtcgaGAGCCGCTGCTGttaGAtTGCGGACGaGCGCTGAGC
gatGTGCGaGTtgcaTTTCAtACCtatGGtactCCGCGCgcagatgcaactCTGGTGCTGCAtgcaCTG
ACCGGtgatAGCGCGGTGCAtGAGTGGTGGccggatTTTCTGggtGCGGGtCGGcccCTGGAC
CCGgctgatgattatGTGGTGTGCgcaAACGTCttaggtGGaTGCgcaGGtactactAGCgcaGCTGAAc
ttgcagcaACCTGTTCCGGACCGGTGCCGttaAGCCTGCGCgatATGgcaCGGGTGGGaCGCg
caCTGCTGgactcgttgGGtGTGCGACGGGTGCGGGTCattGGtGCGAGCATGGGtGGaATGtta
gcgtatgcgTGGCTGCTGgaaTGCccggatCTGGTGGAAaaagcgGTGATTATAGGAGCCCCGG
CGCGGCAttcaccgTGGGCTATTGGACTGAACactGCGGCCCGCAGCGCCataGCCttaGCTcc
gGGtGGtgaaGGaCTGaaaGTtGCGCGtCAGataGCaATGCTttcatatCGtAGCccgGAAAGCCTA
AGCCGtactCAGGCGGGaCAGCGtGTtCCGGGaGTtccgGCCgtaacttcatatCTGCAtTACCAAG
GtGAAaagttagcggcgCGCtttgatgaaCAGACCtatTGCGCCttaACCTGGGCGATGgatGCCttcCA
GCCGAGCAGCGCCgatttaaagGCGGTGCGCGCGCCGGTAttaGTCGTCGGtattTCCAGCgac
CTGttatatccgGCCGCCgaaGTCCGCGCCTGCGCCGCCgaaCTTccgCAtGCCgattatTGGGAA
CTGGGtAGCataCAtGGtCAtgatGCCttcctcATGGACcccCAGgatttaCCGgaaCGGGTGGGGgcc
ttcCTCAGGAGTtaa

Protein sequence:

MTAVLAGHASALLLTEEPDCSGPQTVVLFRREPLLLDCGRALSDVRVAFHTYGTPRADATLV
LHALTGDSAVHEWWPDFLGAGRPLDPADDYVVCANVLGGCAGTTSAAELAATCSGPVPLS
LRDMARVGRALLDSLGVRRVRVIGASMGGMLAYAWLLECPDLVEKAVIIGAPARHSPWAIG
LNTAARSAIALAPGGEGLKVARQIAMLSYRSPESLSRTQAGQRVPGVPAVTSYLHYQGEKL
AARFDEQTYCALTWAMDAFQPSSADLKAVRAPVLVVGISSDLLYPAAEVRACAAELPHADY
WELGSIHGHDAFLMDPQDLPERVGAFLRS*

1.3) O-acetylhomoserine sulfhydrylase from *Bacillus stearothermophilus CN3*

DNA sequence (codon harmonized for *E. coli*):

atgagcaatgaacaaactttccgcccggagacgctcgccatccacgccgggcaaaaaccggatgcggaaacgggtgcgcgg
gcggtgccgatttaccaaacaagctcgtatgtattccgcgacagcgagcatgcggccaatttgtttggcttgaaagaggaagggtt
tatttacacgcgcattatgaacccgacaaacgatgtcttggaaaaacggatcgcggcgcttgaaggcggcattggagcgcttgcg
ctgtcatcggggcaggcggcggtgttttattcgatcatcaacatcgcctcggcgggcgatgaaatcgtcagttcgtcatccatttacg
gcggaacgtacaacttgttcgcccatacgctgcgcaagttcggcattacggtgaagtttgtcgatccgtccgatccggaaaactttg
agcgggcgatcaccgacaaaacgaaagcgttgtttgcggaaacgatcggcaacccgaaaaacgatgtgctcgacatcgaag
cggtggccgacatcgcccatcgccatgccattccgctcatcgtagacaacacggtggccagtccatacttattgcggccgattgaa
ttcggtgccgatatcgtcgtccactcagcgacgaagttcatcggcgggcacggcaattcgatcggcggtgtgattgtggacagcg
gcaagttcgactggaaagggagcggcaagtttccggagttcaccgagccagacccaagctaccacgggttggtgtatgtggac
gccgtcggcgaagcggcgtacatcacgaaagcacgcattcagctcttgcgcgacttaggagcggcactgtcgccgtttaatgcgt
ttttgcttttgcaaggattagagacgcttcacttgcggatgcagcgtcatagcgaaaatgcgctcgctgttgctaagtttttagaagag
gaagaagccgtcgaatcggtcaactacccagggcttccgagccatccgtcgcatgaactggcgaaaaagtatttgccgaatgg
gcagggcgcgatcgtcacgtttgaaatcaaaggcggcgtcgaagccgggaaaaaattgatcgactcggtgaagctgttctcgca
tttggccaacatcggcgattcgaaatcgctcatcatccacccagccagcacgacgcatgagcagctgaccccggaagaacagc
tgtccgccggcgtcaccccaggccttgtgcgtctgtctgtcggcaccgaagcgatcgacgatattttggacgacttgcgccaagcc
attcgccaaagccagacggtgggggtgaagtaa

Protein sequence:

MSNEQTFRPETLAIHAGQKPDAETGARAVPIYQTSSYVFRDSEHAANLFGLKEEGFIYTRIM
NPTNDVLEKRIAALEGGIGALALSSGQAAVFYSIINIASAGDEIVSSSSIYGGTYNLFAHTLRKF
GITVKFVDPSDPENFERAITDKTKALFAETIGNPKNDVLDIEAVADIAHRHAIPLIVDNTVASPY
LLRPIEFGADIVVHSATKFIGGHGNSIGGVIVDSGKFDWKGSGKFPEFTEPDPSYHGLVYVD
AVGEAAYITKARIQLLRDLGAALSPFNAFLLLQGLETLHLRMQRHSENALAVAKFLEEEEAVE
SVNYPGLPSHPSHELAKKYLPNGQGAIVTFEIKGGVEAGKKLIDSVKLFSHLANIGDSKSLIIH
PASTTHEQLTPEEQLSAGVTPGLVRLSVGTEAIDDILDDLRQAIRQSQTVGVK*

1.4) Acetyl-CoA synthetase from *Escherichia coli* (K12)

DNA sequence:

ATGAGCCAAATTCACAAACACACCATTCCTGCCAACATCGCAGACCGTTGCCTGATAAA
CCCTCAGCAGTACGAGGCGATGTATCAACAATCTATTAACGTACCTGATACCTTCTGGG
GCGAACAGGGAAAAATTCTTGACTGGATCAAACCTTACCAGAAGGTGAAAAACACCTCC
TTTGCCCCCGGTAATGTGTCCATTAAATGGTACGAGGACGGCACGCTGAATCTGGCGG
CAAACTGCCTTGACCGCCATCTGCAAGAAAACGGCGATCGTACCGCCATCATCTGGGA
AGGCGACGACGCCAGCCAGAGCAAACATATCAGCTATAAAGAGCTGCACCGCGACGTC
TGCCGCTTCGCCAATACCCTGCTCGAGCTGGGCATTAAAAAAGGTGATGTGGTGGCGA
TTTATATGCCGATGGTGCCGGAAGCCGCGGTTGCGATGCTGGCCTGCGCCCGCATTGG
CGCGGTGCATTCGGTGATTTTCGGCGGCTTCTCGCCGGAAGCCGTTGCCGGGCGCATT
ATTGATTCCAACTCACGACTGGTGATCACTTCCGACGAAGGTGTGCGTGCCGGGCGCA
GTATTCCGCTGAAGAAAAACGTTGATGACGCGCTGAAAAACCCGAACGTCACCAGCGTA
GAGCATGTGGTGGTACTGAAGCGTACTGGCGGGAAAATTGACTGGCAGGAAGGGCGC
GACCTGTGGTGGCACGACCTGGTTGAGCAAGCGAGCGATCAGCACCAGGCGGAAGAG
ATGAACGCCGAAGATCCGCTGTTTATTCTCTACACCTCCGGTTCTACCGGTAAGCCAAA
AGGTGTGCTGCATACTACCGGCGGTTATCTGGTGTACGCGGCGCTGACCTTTAAATATG
TCTTTGATTATCATCCGGGTGATATCTACTGGTGCACCGCCGATGTGGGCTGGGTGACC
GGACACAGTTACTTGCTGTACGGCCCGCTGGCCTGCGGTGCGACCACGCTGATGTTTG
AAGGCGTACCCAACTGGCCGACGCCTGCCCGTATGGCGCAGGTGGTGGACAAGCATC
AGGTCAATATTCTCTATACCGCACCCACGGCGATCCGCGCGCTGATGGCGGAAGGCGA
TAAAGCGATCGAAGGCACCGACCGTTCGTCGCTGCGCATTCTCGGTTCCGTGGGCGAG
CCAATTAACCCGGAAGCGTGGGAGTGGTACTGGAAAAAAATCGGCAACGAGAAATGTC
CGGTGGTCGATACCTGGTGGCAGACCGAAACCGGCGGTTTCATGATCACCCCGCTGCC
TGGCGCTACCGAGCTGAAAGCCGGTTCGGCAACACGTCCGTTCTTCGGCGTGCAACCG
GCGCTGGTCGATAACGAAGGTAACCCGCTGGAGGGGGCCACCGAAGGTAGCCTGGTA
ATCACCGACTCCTGGCCGGGTCAGGCGCGTACGCTGTTTGGCGATCACGAACGTTTTG
AACAGACCTACTTCTCCACCTTCAAAAATATGTATTTCAGCGGCGACGGCGCGCGTCGC
GATGAAGATGGCTATTACTGGATAACCGGGCGTGTGGACGACGTGCTGAACGTCTCCG
GTCACCGTCTGGGGACGGCAGAGATTGAGTCGGCGCTGGTGGCGCATCCGAAGATTG
CCGAAGCCGCCGTAGTAGGTATTCCGCACAATATTAAAGGTCAGGCGATCTACGCCTAC
GTCACGCTTAATCACGGGGAGGAACCGTCACCAGAACTGTACGCAGAAGTCCGCAACT
GGGTGCGTAAAGAGATTGGCCCGCTGGCGACGCCAGACGTGCTGCACTGGACCGACT
CCCTGCCTAAAACCCGCTCCGGCAAAATTATGCGCCGTATTCTGCGCAAAATTGCGGCG

GGCGATACCAGCAACCTGGGCGATACCTCGACGCTTGCCGATCCTGGCGTAGTCGAGA
AGCTGCTTGAAGAGAAGCAGGCTATCGCGATGCCATCGTAA

Protein sequence:

MSQIHKHTIPANIADRCLINPQQYEAMYQQSINVPDTFWGEQGKILDWIKPYQKVKNTSFAP
GNVSIKWYEDGTLNLAANCLDRHLQENGDRTAIIWEGDDASQSKHISYKELHRDVCRFANTL
LELGIKKGDVVAIYMPMVPEAAVAMLACARIGAVHSVIFGGFSPEAVAGRIIDSNSRLVITSDE
GVRAGRSIPLKKNVDDALKNPNVTSVEHVVVLKRTGGKIDWQEGRDLWWHDLVEQASDQH
QAEEMNAEDPLFILYTSGSTGKPKGVLHTTGGYLVYAALTFKYVFDYHPGDIYWCTADVGW
VTGHSYLLYGPLACGATTLMFEGVPNWPTPARMAQVVDKHQVNILYTAPTAIRALMAEGDK
AIEGTDRSSLRILGSVGEPINPEAWEWYWKKIGNEKCPVVDTWWQTETGGFMITPLPGATE
LKAGSATRPFFGVQPALVDNEGNPLEGATEGSLVITDSWPGQARTLFGDHERFEQTYFSTF
KNMYFSGDGARRDEDGYYWITGRVDDVLNVSGHRLGTAEIESALVAHPKIAEAAVVGIPHNI
KGQAIYAYVTLNHGEEPSPELYAEVRNWVRKEIGPLATPDVLHWTDSLPKTRSGKIMRRILR
KIAAGDTSNLGDTSTLADPGVVEKLLEEKQAIAMPS*

2) Construction of plasmids

2.1) pQ Strep_Fdx_H_6

The *fdx* gene was ordered as a synthetic gBlock (Integrated DNA Technologies, Coralville, IA). For sequence verification the gBlock was cloned into a pJET1.2 vector (Thermo Fisher Scientific Inc., Waltham, MA) by blunt end ligation following the protocol of the manufacturer and amplified in *E. coli* TOP10F' cells. A vector with the correct gene sequence served as template for the PCR amplification of the gene with primers (Integrated DNA Technologies) designed to introduce restriction sites. The gene was cloned into a commercially available pQE-80L plasmid. The restriction and ligation were performed according to the protocol of the manufacturer (Thermo Fisher Scientific Inc.).

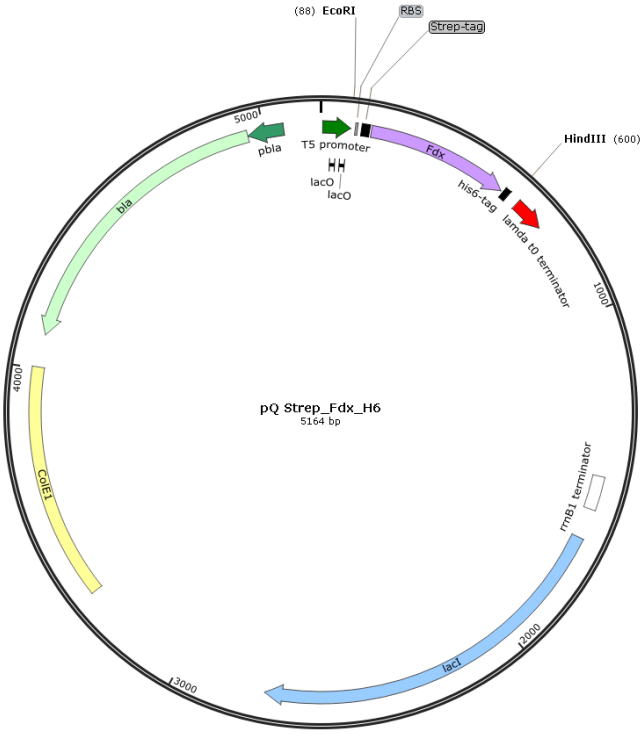


Figure S1: Plasmid map of pQ Strep_Fdx_H_6 harboring the gene encoding Fdx. The plasmid carries a medium copy number ColE1 origin of replication, the β-lactamase gene (*bla*) with a β-lactamase promoter (p*bla*) and the constitutively expressed lac repressor gene (*lacI*). The *fdx* gene is under control of an IPTG-inducible T5/*lacO* promoter and a λt0 terminator (lambda t0 term). It carries an N-terminal Strep tag II and a C-terminal polyhistidine tag (his6-tag). The *fdx* gene including the tags is flanked by EcoRI and HindIII restriction sites, which allow its excision.

.

2.2) pCAS5 Met-analog_Fdx

For the construction of pCAS5 Met-analog_Fdx (Figure S2), the pCAS2 alkylation cascade plasmid CBA[2] was treated with XbaI (FastDigest) and SacI (FastDigest), to excise the methionine adenosyltransferase gene. The parts were gel purified, treated with a DNA blunting enzyme and ligated with a T4 DNA ligase. The restriction, blunting and ligation were performed according to the protocol of the manufacturer (Thermo Fisher Scientific Inc.). The resulting plasmid was amplified in *E. coli* TOP10F' cells on LB agar plates containing 100 µg/mL ampicillin and treated with SalI (FastDigest) and PstI (FastDigest), to excise the methyltransferase gene.

The other parts for the pCAS5 Met-analog_Fdx plasmid were PCR amplified from pQ Strep_Fdx_H_6. The primers (Integrated DNA technologies) for the PCR amplifications were designed to introduce 40 bp homologous overlaps to the adjacent parts which is required for Gibson isothermal assembly. The reactions contained 10 ng of template DNA, 0.5 µM of the forward and reverse primers, 0.2 mM of each deoxyribonucleotide triphosphate, 1X *Phusion* HF Buffer and 1 U *Phusion* high-fidelity DNA Polymerase (Thermo Fisher Scientific Inc., Waltham, MA) in a total volume of 50 µL. The PCR program included an initial denaturation (98 °C for 30 sec), followed by 30 cycles of denaturation (98 °C for 10 sec), annealing (56-58 °C, depending on the melting temperature for 20 sec) and elongation (72 °C for 0.5-2 min, depending on the length) and the final elongation (72 °C for 7 min).

For the assembly of the pCAS5 Met-analog_Fdx plasmid, Gibson isothermal assembly was applied. We followed the method described by Gibson *et al.*.[3] 50 ng of the smallest fragment and all other fragments in equimolar quantities were added to 7.5 µL of the assembly master mix which is composed of 10 U/µL T5 exonuclease (Biozym Scientific GmbH, Hessisch Oldendorf, Germany), 2 U/µL *Phusion* high-fidelity DNA Polymerase (Thermo Fisher Scientific Inc., Waltham, MA), 40 U/µL *Taq* DNA ligase (New England BioLabs Inc., MA) and 1X isothermal (ISO) reaction buffer (5% (w/v) PEG-8000, Tris-HCl (100 mM, pH 7.5), 10 mM $MgCl_2$, 10 mM DTT, 0.2 mM of each deoxyribonucleotide triphosphate, 1 mM NAD in water). The mixture was incubated at 50 °C for 60 min without shaking. The resulting plasmid was amplified in *E. coli* TOP10F' cells on LB agar plates containing 100 µg/mL ampicillin and it was sequence verified (Microsynth, Balgach, Switzerland).

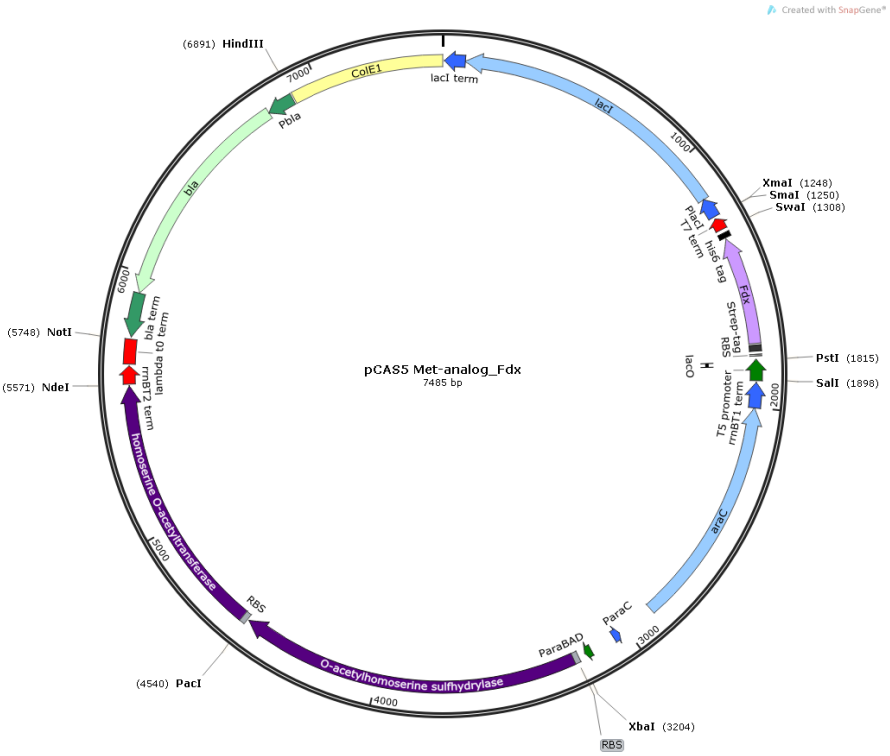


Figure S2: Plasmid map of pCAS5 Met-analog_Fdx. The L-homoserine *O*-acetyltransferase and *O*-acetyl-L-homoserine sulfhydrylase genes are under the control of an arabinose-inducible p*araBAD* promoter and the *rrnB*T2 terminator (*rrnB*T2 term) while the *fdx* gene is controlled by an IPTG-inducible T5/*lacO* promoter (T5 promoter and *lacO*) and a T7 terminator (T7 term). All genes carry the same RBS sequence from the pET21a(+) plasmid. Besides the parts already mentioned, the plasmid features the lac repressor gene (*lacI*) with a lac repressor promoter (p*lacI*) and a lac repressor terminator (*lacI* term), the ara repressor gene (*araC*) with an ara repressor promoter (p*araC*) and an *rrnB*T1 terminator (*rrnB*T1 term) and the β-lactamase gene (*bla*) with a β-lactamase promoter (p*bla*), a β-lactamase terminator (*bla* term) and a λt0 terminator (lambda t0 term).

2.3) pAcetylCoA

The parts for the pAcetylCoA plasmid were PCR-amplified from the pMazF plasmid[4] and from *E. coli* K12 genomic DNA. The PCR amplifications and the Gibson isothermal assembly were carried out as described for the construction of pCAS5 Met-analog_Fdx (2.2)).

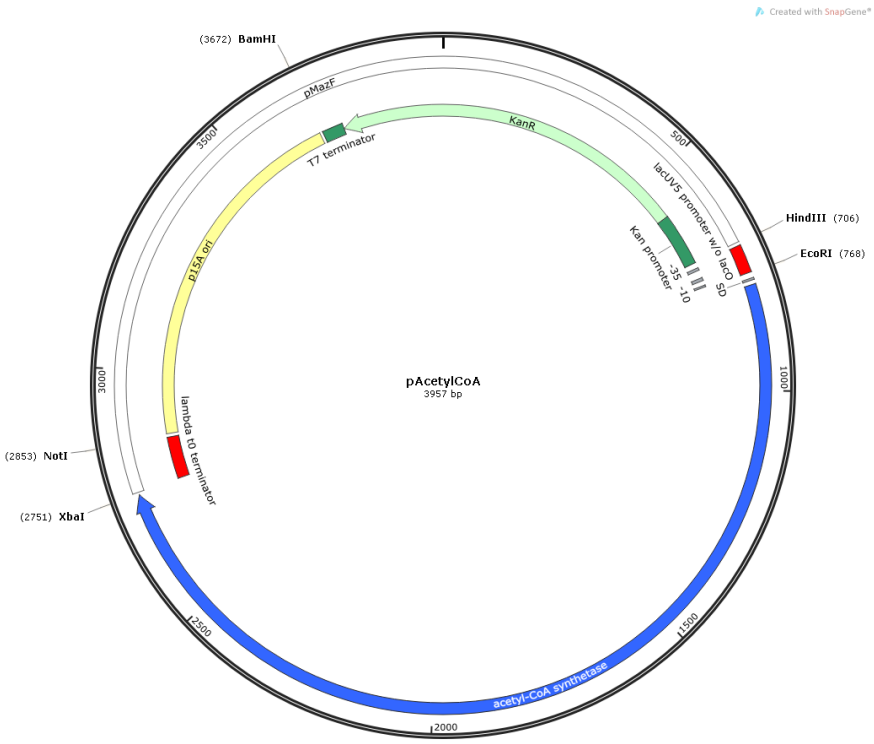


Figure S3: Plasmid map of pAcetylCoA harboring the gene encoding the acetyl-CoA synthetase. The plasmid carries a low copy number p15A origin of replication and a kanamycin resistance gene (*KanR*) with a constitutive kanamycin promoter and a T7 terminator. The acetyl-CoA synthetase gene is controlled by a constitutive lacUV5 promoter (without *lacO*) and a λ t0 terminator. The acetyl-CoA synthetase gene is flanked by EcoRI and XbaI restriction sites which allow its excision. The sequence marked in white was amplified from the pMazF plasmid.

3) Supplementary Methods

3.1) Expression of Fdx[AHA]

Methionine auxotrophic *E. coli* BL21 (DE3) Gold cells were transformed with pQ Strep_Fdx_H_6 and incubated in LB medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl), over night. All media contained 100 µg/mL ampicillin for maintenance of the plasmids. 50 mL M9 medium (6.78 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl, 20 mM glucose, 1 mM $MgSO_4$, 1 mg/mL $CaCl_2$, trace elements) supplemented with 3 g/L yeast extract (YE) were inoculated to an initial D_{600} of 0.2 and the cultures were shaken at 37 °C until methionine depletion at a D_{600} ~ 3 (Figure S4). The cultures were supplemented with 1 mM methionine, or 1 mM or 3 mM 2-amino-5-hexenoic acid and shaken at 28 °C for 20 min for amino acid uptake. The expression of Fdx was induced with 0.5 mM IPTG and the cultures were shaken at 28 °C for 24 °h. To analyze the expression by SDS-PAGE, culture samples were collected before induction (ni), and 4 h and 24 h after induction (Figure S5).

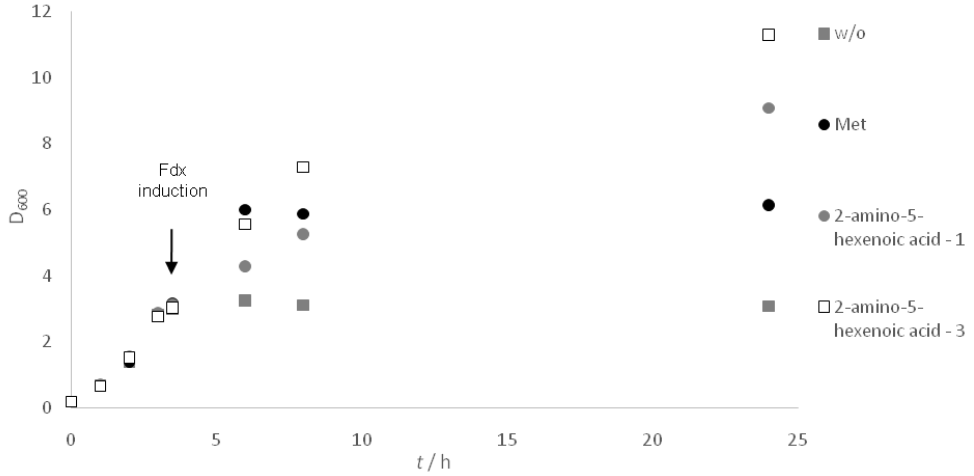


Figure S4: Growth curve of cultures supplemented with no amino acid (w/o), methionine (Met), or 1 mM or 3 mM 2-amino-5-hexonic acid after methionine depletion at a D_{600} ~ 3. Fdx expression was induced 20 min after the supplementation (indicated by the arrow). Data points represent an average of three technical replicates.

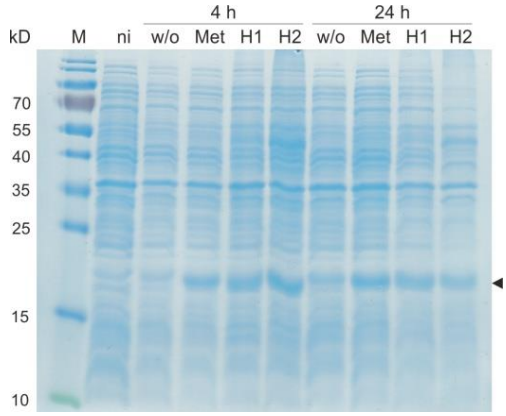


Figure S5: Incorporation of methionine and 2-amino-5-hexenoic acid into Fdx. 14% SDS-polyacrylamide gels were loaded with the total protein of cells which were supplemented with no amino acid (w/o), methionine (Met), or 1 mM (H1) or 3 mM (H2) 2-amino-5-hexonic acid after methionine depletion. Samples were collected before induction (ni), 4 h and 24 h after induction (10 µL, normalized to a D_{600} of 0.5). M, molecular weight marker (5 µL Page Ruler^{TM} Prestained Protein Ladder, Thermo Fisher Scientific Inc., Waltham, MA). Bands of the target protein Fdx are marked with a triangle (MW_{calc} 17.6 kD). Coomassie Brilliant Blue stain.

3.2) Expression of Fdx[AHcy]

A: Methionine auxotrophic *E. coli* BL21 (DE3) Gold cells were transformed with pCAS5 Met-analog_Fdx and incubated in LB medium over night. All media contained 100 µg/mL ampicillin for maintenance of the plasmids. 500 mL M9 medium supplemented with 3 g/L yeast extract (YE) were inoculated to an initial D_{600} of 0.2 and the cultures were shaken at 37 °C until they reached D_{600} ~ 1 (Figure S6).

B-D: Methionine auxotrophic *E. coli* BL21 (DE3) Gold cells were transformed with pCAS5 Met-analog_Fdx (**B**) or co-transformed with pCAS5 Met-analog_Fdx and pAcetylCoA (**C** and **D**) and incubated in LB medium over night. All media contained 100 µg/mL ampicillin (**B**) or 100 µg/mL ampicillin and 50 µg/mL kanamycin (**C** and **D**) for maintenance of the plasmids. 500 mL M9 medium supplemented with 3 g/L yeast extract (YE) and 0.01% (w/v) D-pantothenate and 0.01% (w/v) acetate were inoculated to an initial D_{600} of 0.2 and the cultures were shaken at 37 °C until they reached D_{600} ~ 1 (Figure S6).

A-D: The cultures were supplemented with 5 mM L-homoserine and 5 mM allyl mercaptan (**A**, **B** and **D**) or 1 mM L-homoserine and 1 mM allyl mercaptan (**C**) and expression of the S-allyl-L-homocysteine (AHcy or allyl-Hcy) biosynthesis enzymes, homoserine *O*-acetyltransferase and *O*-acetylhomoserine sulfhydrylase, was induced with 0.1% (w/v) L-arabinose. The cultures were shaken at 37 °C until methionine depletion at a D_{600} ~ 3 (Figure S6). The expression of Fdx was induced with 0.5 mM IPTG and the cultures were shaken at 28 °C for 24 °h. To analyze the amino acid biosynthesis by TLC and HPLC

(Figure S9, Figure 1) and the expression by SDS-PAGE (Figure S7), culture samples were collected before induction of the amino acid biosynthesis enzymes (n1), before induction of Fdx (n2), and throughout the expression after induction of Fdx.

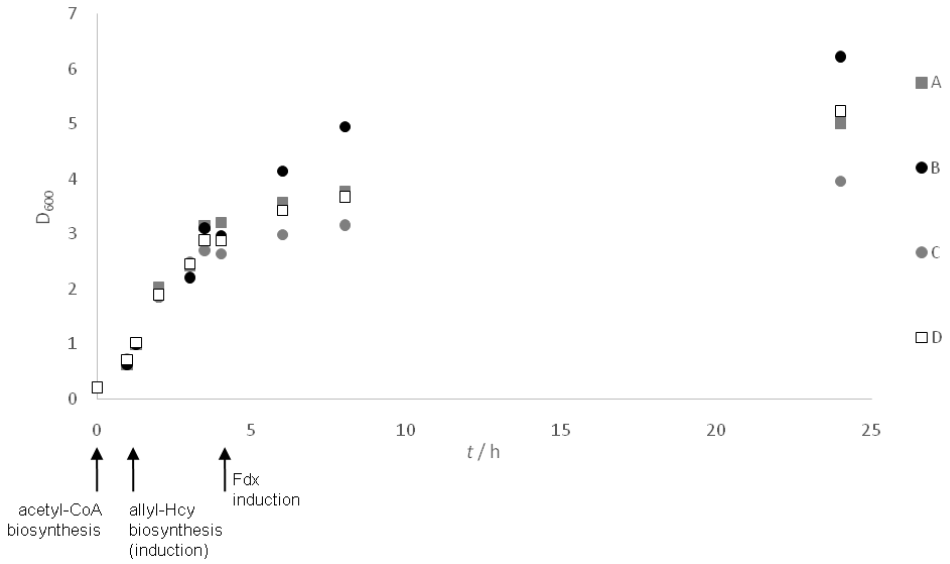


Figure S6: Cells in cultures C and D which were transformed with pAcetylCoA (additionally to pCAS5 Met-analog_Fdx), constitutively expressed acetyl-CoA synthetase (left arrow). At a D_{600} ~ 1, the cultures were supplemented with L-homoserine and allyl mercaptan and expression of the *S*-allyl-L-homocysteine (allyl-Hcy) biosynthesis enzymes was induced (middle arrow). After methionine depletion at a D_{600} ~ 3, expression of Fdx was induced (right arrow). Data points represent an average of three technical replicates.

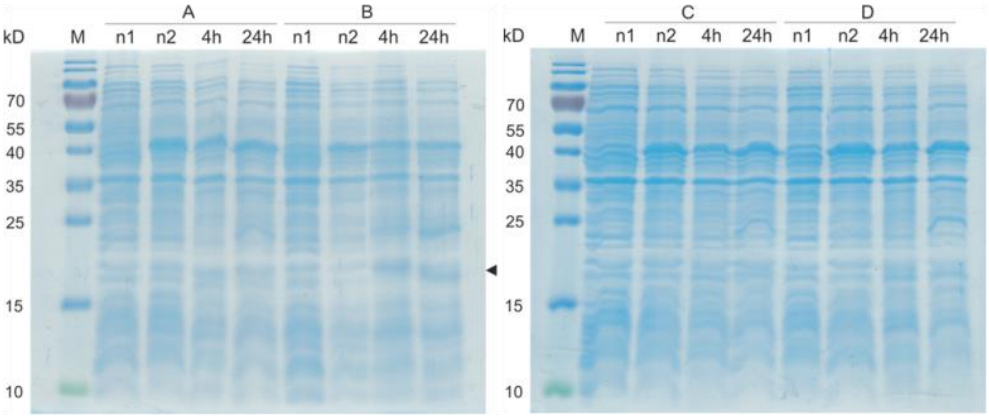


Figure S7: Expression of Fdx (17.6 kD), homoserine *O*-acetyltransferase (35 kD) and *O*-acetylhomoserine sulfhydrylase (47 kD). 14 % SDS-polyacrylamide gels were loaded with the total protein of cells from cultures A-D. Samples were collected before induction of the amino acid biosynthesis enzymes (n1), before induction of Fdx (n2), and 4 h and 24 h after induction of Fdx (10 µL, normalized to a D_{600} of 0.5). M, molecular weight marker (5 µL Page Ruler™ Prestained Protein Ladder, Thermo Fisher Scientific Inc., Waltham, MA). Coomassie Brilliant Blue stain.

3.3) Downstream processing

After the expression, the cells were harvested by centrifugation at 4000 rpm and 4 °C for 20 min, suspended in phosphate buffer (50 mM, pH 8.0) and lysed by sonication (6 min, output 80%). The soluble protein fractions were purified by Strep-Tactin affinity chromatography following the protocol of the manufacturer (IBA, Göttingen, Germany). The fractions, enriched in Fdx[Met], Fdx[AHA] and Fdx[AHcy], were pooled and re-buffered from 2.5 mM desthiobiotin (100 mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA) elution buffer to phosphate buffer (50 mM, pH 8.0) using gravity flow PD-10 desalting columns (GE Healthcare Life Sciences, Little Chalfont, Great Britain). The protein concentrations were determined by Bradford assay (Bio-Rad Laboratories Inc., Hercules, CA). The quality of the purification were analyzed by SDS-PAGE (Figure S8).

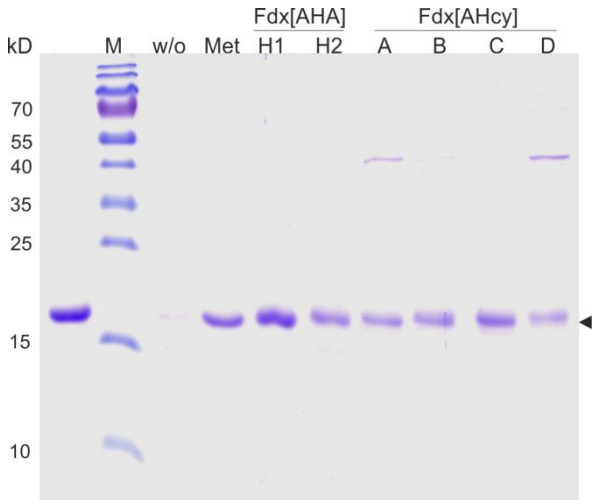


Figure S8: Purification of Fdx[Met], Fdx[AHA] and Fdx[AHcy] 14% SDS-polyacrylamide gel was loaded with 2 µg of purified Fdx (reference), Fdx[Met], Fdx[AHA] from cultures supplemented with 1 mM (H1) or 3 mM (H2) 2-amino-5-hexonic acid and Fdx[AHcy] from cultures A-D. Supplementation with no amino acid (w/o) did not lead to Fdx expression. M, molecular weight marker (5 µL Page Ruler^{TM} Prestained Protein Ladder, Thermo Fisher Scientific Inc., Waltham, MA). Coomassie Brilliant Blue stain.

3.4) TLC analysis (Thin Layer Chromatography)

To analyze the amino acid biosynthesis by TLC and HPLC, whole-cell samples from cultures **A**-**D** were collected throughout the expression and frozen at -20 °C. For analysis, they were thawed and incubated in an equal volume of 75% ethanol at 80 °C and 1000 rpm for 10 min. The samples were cooled on ice for 10 min and centrifuged at 13000 rpm and 4 °C for 15 min. The cell extracts were directly used for TLC and HPLC analysis.

5-10 µL of each sample and of the reference substances were spotted on a TLC plate (silica gel, 200 µm, 60 Å, Merck Millipore, Darmstadt, Germany). TLC was carried out with a n-butanol/acetic acid/water 3:1:1 mixture. The spots were detected by ninhydrin staining.

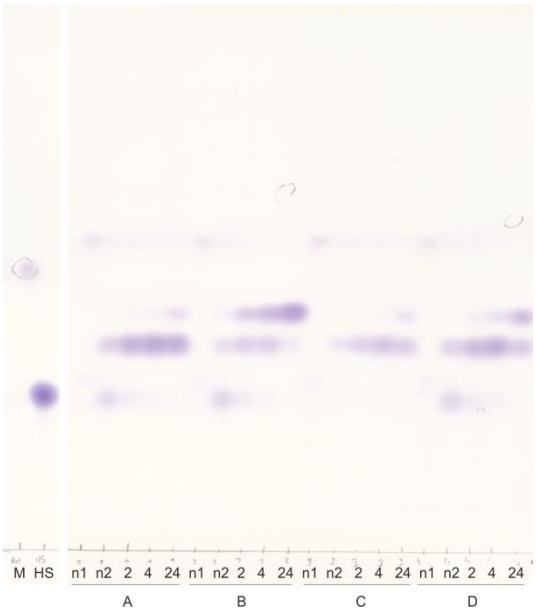


Figure S9: TLC analysis of cells extracts from cultures A-D collected before induction of the amino acid biosynthesis enzyme expression (n1), before induction of Fdx expression (n2), and 2 h, 4 h and 24 h after induction of Fdx expression. Methionine (M) and L-homoserine (HS) served as reference substances. Irrelevant lanes were cut. Ninhydrin stain.

3.5) HPLC analysis (High Performance Liquid Chromatography)

Analysis of the cell extracts was performed on an Agilent Technologies HPLC-System 1200 equipped with a Merck Millipore Purospher® STAR RP-18e (5 µm) 250-4.6 column. The components were separated applying a 0.7 mL/min flow rate of 0.1% formic acid (A) and acetonitrile (B). The following gradient was applied: 0-0.5 min 2% B, 0.5-9 min to 98% B, 9-10 min to 2% B. The mass spectrometer (ESI-MS positive mode) was set to scan mode (100-600 m/z) with single ion monitoring (SIM). Chemically synthesized L-homoserine, *O*-acetyl L-homoserine and *S*-allyl-L-homocysteine served as reference substances (Figure 1).

3.6) Metathesis

Saturated solutions of Hoveyda-Grubbs Catalyst 2nd Generation were prepared by incubating 4 mg of the catalyst in 1 mL water, *t*-butanol or 1,2-dimethoxyethane at 40 °C and 1000 rpm. The metathesis reaction was carried out in 100 µL scale in an Eppendorf thermomixer at 30 °C and 500 rpm for 4 h in open reaction vessels to let the generated ethene escape. The reaction mix contained 60 µL (15 µg) Fdx[Met], Fdx[AHA] or Fdx[AHcy] prepared in phosphate buffer (50 mM, pH 8.0), 10 µL 1M $MgCl_2$, and either 30 µL of the catalyst solution or 10 µL of the catalyst solution and 20 µL phosphate buffer (50 mM, pH 8.0). After 4 h the reaction mix was centrifuged at 13000 rpm and 4 °C for 10 min to separate the soluble (Figure S10) and insoluble fractions containing the undissolved catalyst and the denatured protein.

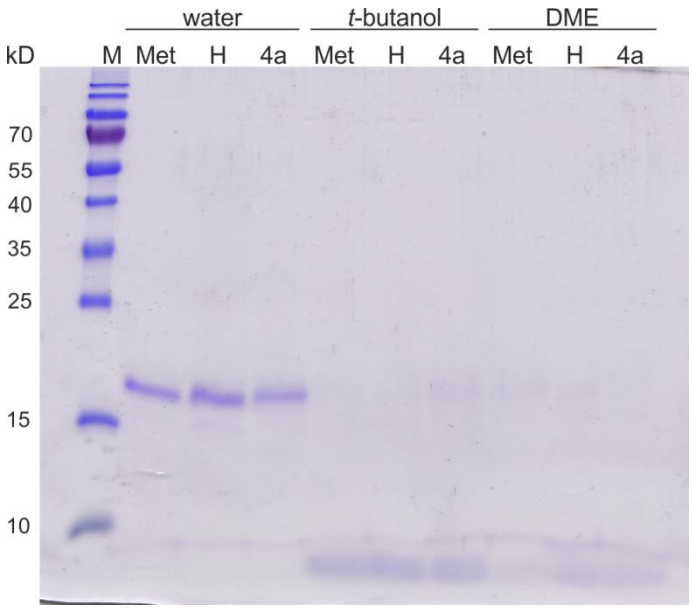


Figure S10: Olefin metathesis of the alkene-functionalized Fdx[AHA] and Fdx[AHcy], Fdx[Met] acted as a negative control. 14% SDS-polyacrylamide gel was loaded with the soluble fractions of the reaction mixes containing water, *t*-butanol or 1,2-dimethoxyethane (DME). M, molecular weight marker (5 µL Page Ruler™ Prestained Protein Ladder, Thermo Fisher Scientific Inc., Waltham, MA). Coomassie Brilliant Blue stain.

4) Supplementary References

[1] M. Eren, R. P. Swenson, *Journal of Biological Chemistry* **1989**, *264*, 14874-14879.

[2] L. Offner, K. Heckenbichler, P. Fladischer, M. Gruber-Khadjawi, R. Breinbauer, B. Wiltschi, *Angewandte Chemie International Edition* **2015**.

[3] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, *Nature Methods* **2009**, *6*, 343-345.

[4] L. Offner, B. David, C. Gruber, K. Gruber, B. Wiltschi, *ChemCatChem* **2015**.

CONCLUDING DISCUSSION & OUTLOOK

Evaluation of a Single Protein Production System for the Expression of Fluorinated Biocatalysts

We evaluated a single protein production system (SPPS) for the expression of the *C*-methyltransferase NovO from *Streptomyces spheroides* and its fluorinated variants. SPPS is based on SPI and bears the potential of making incorporation experiments more economical due to the lower amount of high-priced non-canonical amino acid required. The system was shown to work for a variety of proteins[21] but it was not suitable for the incorporation of (2*S*,4*S*)-4-fluoroproline and (2*S*,4*R*)-4-fluoroproline into the target protein. Indeed, the reduction of the culture volume to a 40-fold concentrated culture had no negative effect on the expression level of the parent protein, but led to a decreased expression of the fluorinated variants. Furthermore, the background concentration of cellular proteins could not be reduced by co-expressing the endoribonuclease MazF. By incorporating (2*S*,4*R*)-4-fluoroproline, the activity and stability of the enzyme were improved, while incorporation of (2*S*,4*S*)-4-fluoroproline and the diastereomeric mixture caused the methyltransferase to be insoluble. However, mass spectrometric analysis revealed that incomplete incorporation of (2*S*,4*R*)-4-fluoroproline into NovO was achieved. To improve the incorporation efficiency, the enzyme charging the amino acids onto their cognate tRNAs, prolyl-tRNA synthetase could be co-expressed.[22]

Multi-Enzyme Cascade Coupling a Modular Cofactor Biosynthesis to a Biocatalytic Friedel-Crafts Alkylation

The methyltransferase CouO from *Streptomyces rishiriensis* is able to act as alkyltransferase in a biocatalytic Friedel–Crafts reaction, when provided with *S*-adenosyl-L-methionine (SAM) analogs as non-natural alkyl group donors. We developed a multi-enzyme cascade that couples the modular biosynthesis of the cofactor analogs to the alkylation of a substrate. The enzymes involved in this multi-step reaction were expressed from different combinatorial plasmids. In an *in vitro* biotransformation, we were able to show that the position of the corresponding genes in the polycistronic operon had an influence on the reaction rates of the individual steps. Different bottle necks of the pathway could be identified for the *in vitro* and *in vivo* biotransformations with the alkylation cascade plasmid carrying the genes in the order CBA (A: L-homoserine *O*-acetyltransferase, B: *O*-acetyl-L-homoserine sulfhydrylase, C: methionine adenosyltransferase). However, neither the cofactor analog, nor the alkylation

product could be detected or identified by HPLC analysis. The preparative scale biosynthesis of the product as a reference for the analysis was still ongoing during the preparation of this thesis. Furthermore, we plan to test other substrates for the alkylation reaction and other mercaptans for the cofactor biosynthesis.

Biosynthesis and Incorporation of Alkene-Containing Amino Acids

With the alkylation cascade plasmids in hand, we developed a two-enzyme cascade for the modular biosynthesis of non-canonical amino acids which enable proteins to take part in bio-orthogonal reactions, e.g. olefin metathesis. By supplementing the cultures with allyl mercaptan, the alkene-containing amino acid *S*-allyl-L-homocysteine (AHcy) could be synthesized and incorporated into the target protein. Furthermore, we were able to functionalize with 2-amino-5-hexenoic acid (AHA). However, we could not detect multimers of the alkene-functionalized proteins, Fdx[AHA] and Fdx[AHcy] after the olefin metathesis reaction by SDS-PAGE. From our experiments, we could not conclude if the reaction did not work due to the double bonds not being accessible for the catalyst or if the reaction worked but SDS-PAGE was not the appropriate detection method. We observed that the proteins showed an altered migration behavior due to components of the metathesis reaction mix. Since their behavior was not affected when the saturated catalyst solution was prepared in water instead of an organic solvent, an effect of the catalyst could be excluded. To clarify if polymerization of the alkene-functionalized proteins occurred, we intend to analyze the samples by mass spectrometry.

Our ultimate goal for the biosynthesis of the artificial cofactor and of the non-canonical amino acids with the modular multi-enzyme cascades is to create completely self-sufficient systems. To make them suitable for application in large scale experiments, we intend to engineer an L-homoserine overproducing strain.

REFERENCES

[1] S. A. Benner, F. Chen, Z. Yang in *Chemical Synthetic Biology* (Eds.: P. L. Luisi, C. Chiarabelli), John Wiley & Sons, Ltd, Chichester, UK, **2011**, pp. 69-106.

[2] J. Becker, C. Wittmann, *Angewandte Chemie* **2015**, *127*, 3383-3407.

[3] C. Chiarabelli, P. Stano, F. Anella, P. Carrara, P. L. Luisi, *FEBS Letters* **2012**, *586*, 2138-2145.

[4] a) K. Deepankumar, S. P. Nadarajan, S. Mathew, S.-G. Lee, T. H. Yoo, E. Y. Hong, B.-G. Kim, H. Yun, *ChemCatChem* **2015**, *7*, 417-421; b) M. G. Hoesl, C. G. Acevedo-Rocha, S. Nehring, M. Royter, C. Wolschner, B. Wiltschi, N. Budisa, G. Antranikian, *ChemCatChem* **2011**, *3*, 213-221.

[5] N. Budisa, *Angewandte Chemie International Edition* **2004**, *43*, 6426-6463.

[6] N. Budisa, C. Minks, S. Alefelder, W. Wenger, F. Dong, L. Moroder, R. Huber, *The FASEB Journal* **1999**, *13*, 41-51.

[7] C. Odar, M. Winkler, B. Wiltschi, *Biotechnology Journal* **2015**, *10*, 427-446.

[8] N. C. Yoder, K. Kumar, *Chemical Society Reviews* **2002**, *31*, 335-341.

[9] M. Salwiczek, E. K. Nyakatura, U. I. M. Gerling, S. Ye, B. Koksch, *Chemical Society Reviews* **2012**, *41*, 2135-2171.

[10] a) B. Holzberger, S. Obeid, W. Welte, K. Diederichs, A. Marx, *Chemical Science* **2012**, *3*, 2924-2931; b) C. Renner, S. Alefelder, J. H. Bae, N. Budisa, R. Huber, L. Moroder, *Angewandte Chemie International Edition* **2001**, *40*, 923-925.

[11] J.-L. P. Hérisson, Y. Chauvin, *Die Makromolekulare Chemie* **1971**, *141*, 161-176.

[12] M. Schuster, S. Blechert, *Angewandte Chemie International Edition in English* **1997**, *36*, 2036-2056.

[13] A. Fürstner, *Angewandte Chemie International Edition* **2000**, *39*, 3012-3043.

[14] a) Y. Chauvin, *Angewandte Chemie International Edition* **2006**, *45*, 3740-3747; b) R. R. Schrock, *Angewandte Chemie International Edition* **2006**, *45*, 3748-3759; c) R. H. Grubbs, *Angewandte Chemie International Edition* **2006**, *45*, 3760-3765; d) R. H. Grubbs, *Tetrahedron* **2004**, *60*, 7117-7140.

[15] K. C. Nicolaou, P. G. Bulger, D. Sarlah, *Angewandte Chemie International Edition* **2005**, *44*, 4490-4527.

[16] a) Y. A. Lin, J. M. Chalker, N. Floyd, G. J. L. Bernardes, B. G. Davis, *Journal of the American Chemical Society* **2008**, *130*, 9642-9643; b) J. B. Binder, R. T. Raines, *Current Opinion in Chemical Biology* **2008**, *12*, 767-773.

[17] a) H.-w. Ai, W. Shen, E. Brustad, P. G. Schultz, *Angewandte Chemie International Edition* **2010**, *49*, 935-937; b) L. K. Henchey, A. L. Jochim, P. S. Arora, *Current Opinion in Chemical Biology* **2008**, *12*, 692-697; c) Y.-W. Kim, T. N. Grossmann, G. L. Verdine, *Nat. Protocols* **2011**, *6*, 761-771.

[18] E. Ricca, B. Brucher, J. H. Schrittwieser, *Advanced Synthesis & Catalysis* **2011**, *353*, 2239-2262.

[19] P. T. Anastas, J. C. Warner in *Green Chemistry: Theory and Practice*, Oxford University Press, **1998**.

[20] R. C. Cioc, E. Ruijter, R. V. A. Orru, *Green Chemistry* **2014**, *16*, 2958-2975.

[21] M. Suzuki, L. Mao, M. Inouye, *Nat. Protocols* **2007**, *2*, 1802-1810.

[22] W. Kim, A. George, M. Evans, V. P. Conticello, *ChemBioChem* **2004**, *5*, 928-936.