Claudia Gafko, BSc

Genome editing of *Escherichia coli* improves the incorporation efficiency of non-canonical amino acids

Master's Thesis

to achieve the university degree of

Diplomingenieurin

Master's degree programme: Biotechnology

submitted to

Graz University of Technology

Supervisor

Assoc.-Prof. Dipl-Ing. Dr.techn. Harald Pichler

Institute of Molecular Biotechnology

Dipl.-Ing. Dr.techn. Birgit Wiltschi

Austrian Centre of Industrial Biotechnology

Graz, March 2018

1 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 UniGRAZonline and TUGRAZonline is identical to the present master‘s thesis.

Date, Signature

2 Acknowledgements

First, I would like to thank Assoc.-Prof. Dipl.-Ing. Dr.techn. Harald Pichler, the main examiner of this thesis, for giving me the opportunity to perform my master's thesis at the Austrian Centre of Industrial Biotechnology.

My special thanks go to Dipl.-Ing. Dr.techn. Birgit Wiltschi, who was my supervisor. I particularly want to thank Dr. Wiltschi for her continuous support, patience and advice on how to conduct scientific research and critically question its results. Her expertise and passion for natural sciences inspired me to pursue a career in scientific research, which I am grateful for.

I also want to thank all my colleagues and friends from university, with whom even the most demanding study periods were manageable and entertaining. I know it is rare to spend your student life with so many awesome and inspiring people and I am thankful for every late-night reunion and every lecture we still managed to attend. A special thanks goes to all my colleagues from the Synthetic Biology Group and the 4^{th} floor of acib. Their constant support and and valuable discussions about my results definitely helped me to progress in my work. Particularly I want to thank Felix Tobola, MSc., who introduced me to the lab and topic and always had a sympathetic ear for all of my questions.

Furthermore I would like to thank Patrik Fladischer, MSc., for kindly providing the strains, plasmids and his expertise on the topic. His groundwork definitely helped me to progress faster in my thesis. I also want extend my gratitude to acib, for funding this project.

I sincerely want to thank Jakob Lidl for his relentless support and love. I am utterly grateful that he is always by my side and helps me to find solutions when I cannot find them myself and that we complement each other in the most positive ways.

Am allermeisten möchte ich jedoch meinen Eltern, Edith und Walter Gafko, danken. Es ist und war immer ihr größtes Anliegen mir eine unbeschwerte Kindheit, Jugend und beste Ausbildung zu ermöglichen. Ich bin unendlich dankbar dafür, dass sie mir in jeder Lebenslage zur Seite stehen und für alles, was sie auf sich genommen haben um mir ein sorgenfreies Leben zu ermöglichen.

3 Abstract

The incorporation of non-canonical amino acids (ncAAs) is an excellent strategy for the introduction of novel chemical and physical properties into target proteins. For site-specific ncAA incorporation by amber stop codon suppression, an archaeal aminoacyl-tRNA synthetase/ suppressor tRNA pair is utilised. Despite its great potential, incorporation efficiencies are rather poor, because the suppressor tRNA competes with endogenous release factors. To make this strategy desirable for industry, it is essential to improve the incorporation efficiency of ncAAs. It is reported that the competition of the suppressor tRNA and release factors can be eliminated by deleting release factor 1 (RF-1). RF-1 is not essential in *E. coli*, however its knock-out causes growth deficiencies, due to improper translation termination.

We generated the *E. coli* BL21 strain BWEC72, which contained a tunable RF-1. In the context of this study, "tunable RF-1" meant that RF-1 was expressed during the growth phase and degraded upon induction of target protein expression. To controllably degrade RF-1 we applied the *ssrA*/SspB-ClpXP degradation system, by fusing a *ssrA*-tag to the C-terminus of the RF-1 encoding gene *prfA* and integrating an inducible *sspB* expression construct at the *melAB* locus of *E. coli*. We confirmed that RF-1 degradation was controllable by the expression of SspB.

The degradation of RF-1 leads to an efficient read through of in-frame amber stop codons, yet it consequently results in aberrant cellular proteins. Approximately 300 open-reading frames in *E. coli* end at amber stop codons, of which seven are essential. Read through of those seven essential genes might cause growth deficiencies. To avoid negative effects on the growth behaviour we performed stop codon exchange in the seven essential genes from amber to ochre or opal codons by CRISPR/Cas9.

We tested whether the incorporation efficiency of the ncAA BocK into the amber mutant reporter protein eGFPx was improved by the tunable degradation of RF-1. The generated strain BWEC72 exhibited an approximately 1.5 fold higher fluorescence signal than the wild-type strain BL21. Subsequently, we used the strain BWEC72 to incorporate the reactive ncAA AzK into the subunit B shiga toxin (Stx1B) from *Shigella dysenteriae*. Stx1B specifically binds the glycosphingolipid Gb3, which is overexpressed in certain tumor cells. This property makes Stx1B particularly attractive for the application in targeted cancer therapy. We obtained the variant protein Stx1B-K8[AzK] in high purity and confirmed the accessibility of the chemical modification by Copper(I)-catalysed azide-alkyne cycloaddition.

4 Kurzfassung

Der Einbau von nicht-kanonischen Aminosäuren (nkAS) stellt eine vielversprechende Methode dar, um die chemischen und physikalischen Eigenschaften von industriell relevanten Proteinen zu verbessern. nkAS sind nicht im genetischen Code codiert und können daher von *Escherichia coli* nicht natürlich in Proteine eingebaut werden. Um nkAS trotzdem ortsspezifisch einbauen zu können, wird ein orthogonales Aminoacyl-tRNA-Synthetase/ tRNA Paar verwendet, welches natürlich in Archäen vorkommt. Mithilfe dieses orthogonalen Paares wird die gewünschte nkAS an der Position eines amber Stopcodons in das Zielprotein eingebaut. Dabei konkurriert die orthogonale tRNA, welche mit der nkAS beladen ist, mit endogenen Freisetzungsfaktoren (*release factors*, RF) um die Bindungsstelle am amber Stopcodon. Es wurde bereits nachgewiesen, dass eine Deletion des Freisetzungsfaktors 1 (RF-1) die Einbaueffizienz der nkAS erhöht, da keine Konkurrenz zwischen beiden Komponenten mehr besteht. RF-1 ist nicht essenziell für *E. coli*, jedoch wurden starke Wachstumsdefizite beobachtet, da die Translation nicht mehr exakt terminiert wird. In dieser Arbeit generierten wir den *E.coli* Stamm BWEC72, welcher einen regulierbaren RF-1 enthielt. RF-1 wurde während der exponentiellen Wachstumsphase exprimiert und erst abgebaut, sobald die Expression des Zielproteins induziert wurde. Um RF-1 kontrollierbar abbauen zu können, verwendeten wir das zelluläre *ssrA*/SspB-ClpXP Degradationssystem. Dazu wurde ein *ssrA*-Tag am C-terminus des RF-1 codierenden Gens *prfA* angefügt und eine induzierbare *sspB* Expressionskassette in den *melAB* locus von *E. coli* integriert. Somit konnte sichergestellt werden, dass der Abbau von RF-1, durch die Expression von SspB gesteuert wurde.

Der Abbau von RF-1 erhöhte die Effizienz an überlesenen amber Stopcodons, was unausweichlich zur Expression von aberranten Proteinen führte. Ungefähr 300 Gene von *E. coli* enden mit einem amber Stopcodon. Davon sind sieben Gene essentiell. Eine inkorrekte Expression dieser sieben Gene würde eine große Bürde für den Organismus darstellen und zu Wachstumsnachteilen führen. Um dies zu verhindern, wurden die sieben amber Stopcodons mithilfe des CRISPR/Cas9 Systems zu ochre oder opal Stopcodons mutiert.

Wir untersuchten ob die Einbaueffizienz, in der Abwesenheit von RF-1, tatsächlich erhöht werden kann, indem wir die nkAS BocK ortsspezifisch in das Reportergen eGFPx einbauten. Der generierte Stamm BWEC72 wies ein ungefähr 1.5-fach höhreres Fluoreszenzsignal auf, als der Wildtyp Stamm BL21. Aufgrund der vielversprechenden Resultate des Stammes, verwendeten wir BWEC72 für den Einbau der reaktiven nkAS AzK in die Untereinheit B des Shigatoxins (Stx1B) von *Shigella dysenteriae*. Stx1B besitzt eine spezifische Affinität für Globotriasylceramid (Gb3), welches in einigen Tumorzellen überexprimiert wird. Dadurch bietet sich Stx1B als vielversprechende Möglichkeit zur gerichteten Krebstherapie an. Es gelang uns, Stx1B-K8[AzK] in sehr hoher Reinheit und guter Ausbeute zu exprimieren. Des Weiteren konnten wir durch Kupfer(I)-katalysierte Azid-Alkin Cycloaddition, die sterische Zugänglichkeit der chemischen Modifizierung von Stx1B-K8[AzK] nachweisen.

Contents

1 Introduction

1.1 Relevance of non-canonical amino acids

Natural proteins and enzymes are generated by a strikingly small set of 20 canonical amino acids, which are encoded by the genetic code. These amino acids display highly diverse functionalities. For instance, their side chains must be stable under physiological conditions, yet reactive when used for selective catalysis. They also need to contribute to molecular interactions of sufficient diversity to constitute well-defined, convoluted three-dimensional structures. However, wild-type proteins are limited to their rather moderate side-chain chemistry. Hence, many proteins do not display properties, which are required for certain industrial applications [1], [2].

Non-canonical amino acids (ncAAs) exhibit a more diverse side chain chemistry, which substantially expands the spectrum of chemical protein modifications. Thus, addition of ncAAs to the genetic code might result in proteins with novel properties and an expanded range of functions [3]. In this study we focused on the lysine derivatives, which are depicted in Figure 1. Panel A-C show the canonical amino acid L-lysine in comparison to its derivatives Boc-lysine (BocK) and azido-lysine (AzK). The incorporation of AzK into proteins is particularly attractive, since its azide group is reactive and can be used for bioorthogonal conjugations with other biomolecules [4].

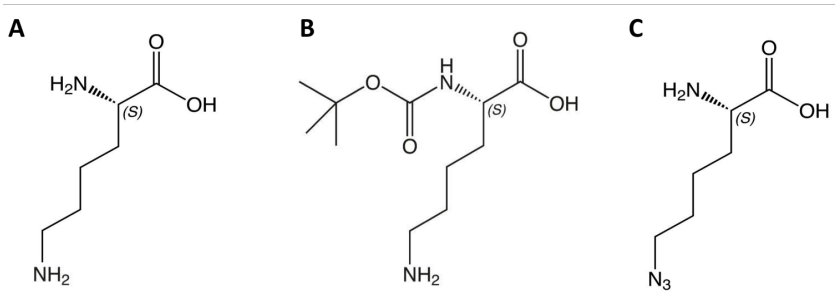


Figure 1: Structures of the non-canonical lysine derivatives used in this study. Structures of L-lysine (**A**), non-reactive BocK (Boc-Lys-OH), where Boc functions as a protective group (**B**) and reactive AzK (H-L-Lys(EO-N3)-OH, **C**).

Non-canonical amino acids are not encoded by the genetic code, hence they are not naturally incorporated into proteins. In 1959 Munier and Cohen first showed the quantitative incorporation of the ncAA seleno-methionine instead of methionine into a bacterial protein [5]. To date, there are two complementary techniques for the incorporation of ncAAs into proteins *in vivo*. Residue-specific incorporation facilitates global substitution of a canonical amino acid with its non-canonical analog. This method allows incorporation of ncAAs at multiple sites, which can substantially alter the chemical and physical behavior of the engineered protein. Alternatively, site-specific incorporation of ncAAs enables introduction of point mutations, with minimal disruption of the target proteins's structure [6]. In this study we exclusively performed site-specific incorporation.

1.2 Site-specific incorporation of non-canonical amino acids

All proteins are synthesised by the cellular ribosomal translation machinery. The ribosome carries out the translation process. Prokaryotic ribosomes are built of a 50S and 30S subunit, which when assembled into a complex, translate the mRNA into a full length protein. The mRNA encodes the genetic information and determines the order of the amino acids, which are ultimately assembled to a protein. Transfer RNAs (tRNAs) play a crucial role in protein synthesis, by delivering the amino acids to the ribosome. They contain a set of three nucleotides, the anticodon, by which they bind to the complementary codon on the mRNA. The tRNA also carries an amino acid that specifically is encoded by the codon to which the tRNA binds. The charging of the amino acid onto its corresponding tRNA is executed by an enzyme called aminoacyl-tRNA synthetase. A pair of aminoacyl-tRNA synthetase and tRNA is specific for each canoncical amino acid. Protein expression terminates at either an amber (TAG), ochre (TAA) or opal (TGA) stop codon [7].

Non-canonical amino acids are not encoded by the genetic code, which is a major challenge for ncAA incorporation by ribosomal translation. They do not possess an aminoacyl-tRNA synthetase/ tRNA pair, hence they cannot naturally be transferred and incorporated into proteins by the ribosome. Additionally, the translation machinery contains several quality control mechanisms, which prevent incorporation of the "wrong" amino acid at a specific codon. For instance, one important checkpoint is the binding of the amino acid to the active-site pocket of the aminoacyl-tRNA synthetase. It recognises the wrong shape and chemical properties of the amino acid and binding to the aminoacyl-tRNA synthetase is prevented [8].

It is known that archaeal *Methanosarcina* species naturally incorporate the ncAA pyrrolysine into methyltransferases. The methyltransferase encoding genes contain in-frame amber stop codons, which are read through during synthesis of the full length methlytransferase. Interestingly, pyrrolysine is found to be incorporated at these in-frame amber stop codons [9]. Incorporation of the ncAA at an amber stop codon is achieved by a pyrrolysyl-tRNA synthetase (PylRS), which aminoacylates a suppressor Pyl-tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$ with pyrrolysine. Pyrrolysyl-tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$, carries the anticodon CUA and incorporates the ncAA in response to an in-frame amber stop codon (UAG). In addition to its natural substrate pyrrolysine, the PylRS/ suppressor Pyl-tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$ pair is able to charge and incorporate a diverse set of pyrrolysine- and lysine derivatives into target proteins [10]. This makes this mechansism highly suitable to perform site-specific incorporation of ncAAs into recombinant proteins [11].

To ensure incorporation of ncAA into proteins expressed in *E. coli*, it is crucial that the PylRS/ Pyl-tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$ pair is orthogonal, because it must not interact with other cellular aminoacyl-tRNA-synthetase/ tRNA pairs in *E. coli*. In this study, we used the pyrrosyl-tRNA synthetase from *Methanosarzina mazei* (*Mm*PylRS) and the suppressor Pyl-tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$ from *Methanomethylophilus alvus* (*Mma*Pyl-tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$). Despite the great potential of protein engineering with ncAAs, the incorporation efficiencies of ncAAs are rather low, because the suppressor Pyl-tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$ competes with the cellular release fac-

tors for amber codons. The native function of release factors is to recognise stop codons and terminate translation [12]. *Escherichia coli* possesses two release factors, RF-1 and RF-2. RF-1 recognises the the amber and ochre stop codons TAG and TAA and RF-2 recognises the ochre and opal stop codons TAA and TGA [13]. Figure 2 illustrates the incorporation of an ncAA using the orthogonal *Mm*PylRS/ *Mma*Pyl-tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$ pair and the competition of the charged Pyl-tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$ with RF-1.

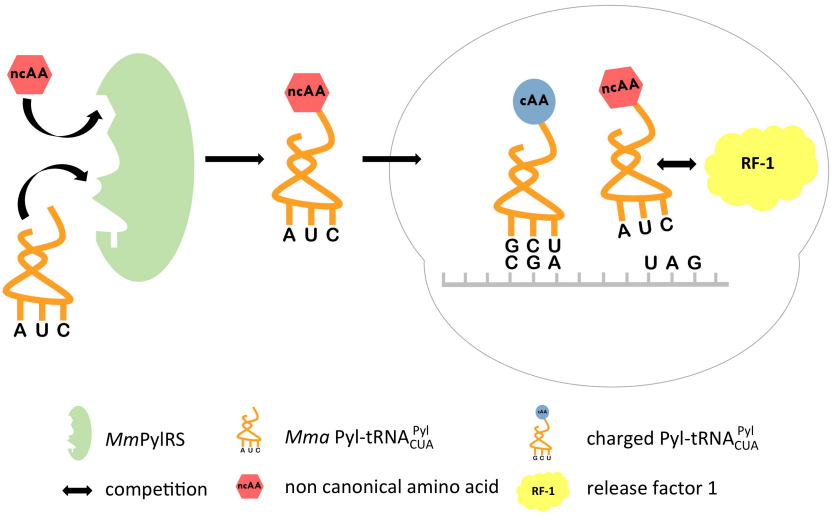


Figure 2: Principle of ncAA incorporation by an orthogonal *Mm*PylRS/ *Mma*Pyl-tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$ pair. The pyrrolysyl-tRNA synthetase from *Methanosarcina mazei* *Mm*PylRS loads the ncAA onto a suppressor Pyl-tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$ from *Methanomethylophilus alvus* (*Mma*). The charged suppressor Pyl-tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$ is then able to incorporate the ncAA in response to an in-frame amber stop codon. However, incorporation efficiencies are rather low, because the Pyl-tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$ competes with RF-1 for the recognition of the amber codon.

Several studies have demonstrated that RF-1 is not essential for *E. coli*. More importantly, a knock out of RF-1 enhances the incorporation efficiency of non-canonical amino acids at amber stop codons [14]. Knock-out of RF-1 leads to an efficient read through at multiple in-frame amber stop codons. The effect can be augmented by switching selected [16] or all [15] amber stop codons of *E. coli* to ochre or opal codons. The reason why stop codons are preferred for codon reassignment is because they occur much more rarely than sense codons. Compared to ochre (UAA) and opal (UGA) stop codons, amber stop codons terminate the smallest number of genes. Therefore they are preferred for stop codon reassignment [12]. Even though only a small number of genes are terminated by amber stop codons, read-through of those codons consequently results in aberrant cellular proteins. Aberrant proteins of essential genes might have a negative effect on the growth behaviour of the cells. To avoid this limitation, amber stop codons can be exchanged for ochre or opal stop codons. *E. coli* BL21 carries seven essential genes that terminate at an amber codon, therefore these seven amber codons are primary targets for a switch to opal or ochre [17].

1.3 Protein degradation by the SspB-dependent ClpXP degradation tag *ssrA*

Protein degradation is an essential regulatory mechanism, which is for instance responsible, for instance, for cell-cylce progression [18] or apoptosis [19]. It is also crucial for quality control of intracellular proteins, by selective breakdown of misfolded or damaged polypeptides [20]. The *ssrA* tag from *Escherichia coli* is a well characterized degradation signal, which is recognised by the SspB adapter protein and the ClpXP protease-complex [21]. Figure 3 illustrates the underlying principle of the SspB mediated ClpXP protein degradation. To degrade a certain protein of interest in a controlled manner, the protein must be equipped with an N- or C-terminal *ssrA*-tag. This tag contains a recognition site for the adapter protein SspB as well as the ATPase regulatory subunit ClpX [22]. When SspB is expressed, it recognises the *ssrA*-degradation signal and guides the tagged protein to the ClpXP protease-complex. After binding to ClpX, the protein is degraded by ClpP (Figure 3A) [23].

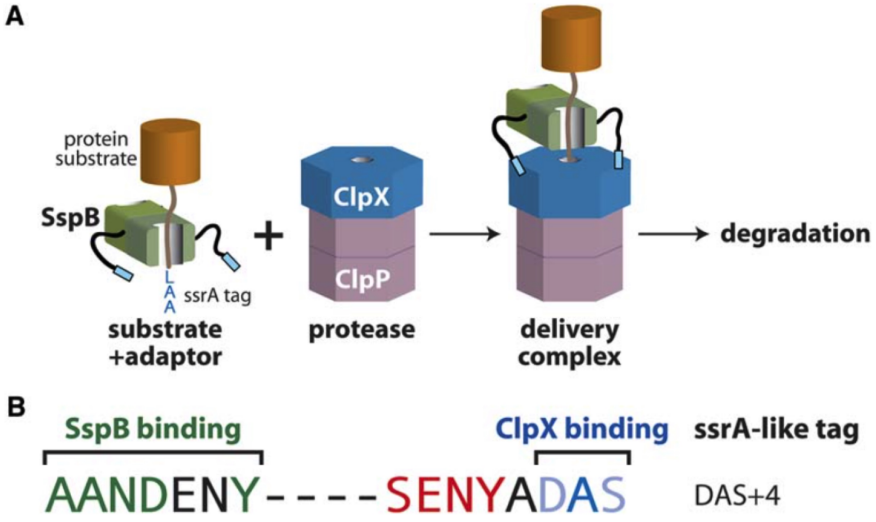


Figure 3: Principle of the SspB-dependent ClpXP protein degradation system. Proteins that carry a *ssrA*-degradation tag are recognized and bound by the adapter protein SspB. The SspB-protein complex then binds the ClpXP protease, which degrades the protein. Image source [23].

In this study, we utilized th *ssrA*/SspB-ClpX degradation system to regulate the intracellular levels of RF-1. We fused the *ssrA*-tag to the C-terminus of the RF-1 encoding gene *prfA*. As illustrated in Figure 3B, we chose the *ssrA*-like tag DAS+4, which contained four amino acid residues between the SspB and ClpX binding sites. Degradation rates of *ssrA*-tagged proteins are dependent on the absence or presence of SspB. Hence, we aimed to regulate RF-1 levels, by controlling the expression of SspB. Since the SspB-mediated ClpXP degradation system is natural in *E. coli*, we deleted the endogenous *sspB* allele and replaced it with an IPTG inducible copy [24].

1.4 CRISPR/Cas9 mediated genome editing

As described above, we deleted the genomic copy of *sspB* copy and integrated a controllable, genomic *sspB* expression construct in the *E. coli* BL21 strain. For stable genomic integration, we used the CRISPR/Cas9 system in combination with λ-Red recombineering. In 2012, Jinek et al. published how the clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas system can be transformed into a programmable tool for in vivo site-directed mutagenesis [25]. This discovery fundamentally increased the ease of genome editing in eukaryotic and prokaryotic cells. Basically, the CRISPR/Cas9 system introduces precise double-strand (ds) breaks in DNA. The system relies on the complex formation of a synthetic single guide RNA (sg-RNA) and an RNA-guided DNA endonuclease Cas9. The sg-RNA consists of a 20 bp long, target-specific CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA). The particular structure of the tracrRNA enables Cas9-binding and guides the complex to the targeted DNA sequence. This target sequence is also referred to as protospacer, with a protospacer adjacent motive (PAM) at the 3' end (in case of *Streptococcus pyogenes* = NGG). The sg-RNA/Cas9-complex introduces the ds-break three base pairs upstream of the PAM sequence [25],[26]. Double-strand breaks have fatal effects on *E. coli*, since non-homologous end joining is a very rare event [27]. In this study, we applied the λ-Red recombineering system to repair the introduced gap. λ-Red recombination is a phage protein-mediated homologous recombination. The phage proteins Exo, Beta and Gamma are crucial for λ-Red recombineering [28]. Exo is a 5'→3' dsDNA dependent exonuclease, which degrades linear dsDNA from the 5'-end. Beta binds the ssDNA created by Exo and promotes the annealing to a complementary ssDNA present in the cell. Gamma prevents the RecBCD system from digesting linear DNA [29]. For gap repair, a DNA fragment, which carries homologies to the adjacent downstream and upstream sequences is introduced to the cells. Figure 4 provides an overview of genome editing using CRISPR/Cas9 in combination with λ-Red recombination.

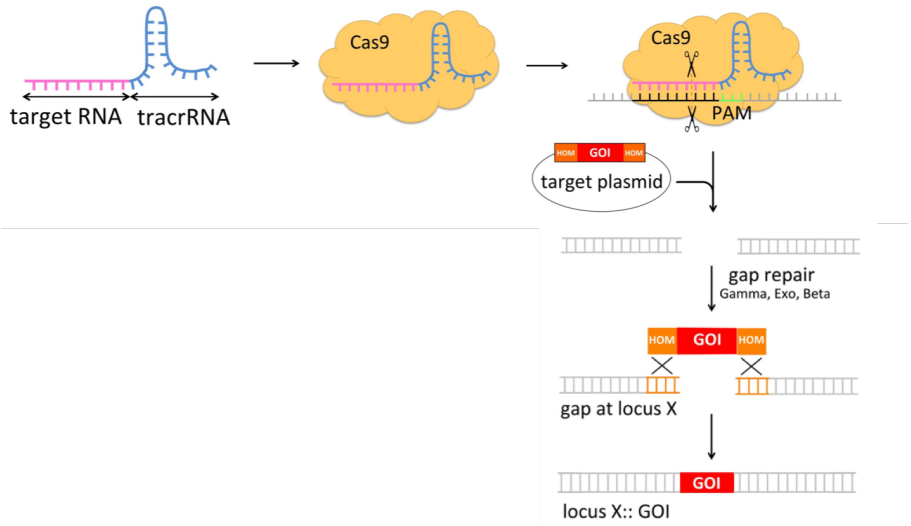


Figure 4: Principle of the directed ds-break repair using CRISPR/Cas9 in combination with λ-Red recombineering. The CRISPR/Cas9 system relies on the formation of the endonuclease Cas9 and a sg-RNA. The sg-RNA consists of a 20 bp target sequence and a tracrRNA sequence. The complex recognises the PAM sequence, adjacent to the target sequence on the genome and introduces a ds-break 3 bp's upstream of the PAM region. To repair this double-strand break a so called target plasmid is introduced into the cells. The target plasmid carries an integration cassette with flanking sequences that are homologous to the cut locus. The gap was repaired by homologous recombination of the Λ-Red system. After genome editing the locus carried the desired integration fragment.

In this project, we integrated a controllable *sspB* construct at the melibiose operon *melAB* in the genome of *E. coli*. The *melAB* locus consists of two structural genes, which code for α-galactosidase (*melA*) and the melibiose transport carrier (*melB*). These proteins are necessary for the melibiose metabolism, which is however, not an essential pathway in *E. coli* [30]. Hence, we chose this locus for genomic integration of the *sspB* construct.
For genomic integration we applied a CRISPR/Cas9 two-plasmid system described by Jiang et al [26]. It consists of the plasmid pCas and a target plasmid. pCas carries the endonuclease Cas9 and the λ-Red recombineering genes *exo, bet* and *gam*. The target plasmid carries the target-specific sg-RNA and the DNA fragment, with the respective homologies, for gap repair. In a recent study we devised a workflow for the two-vector CRISPR/Cas9 system in combination with λ-Red recombineering [31].

1.5 Shiga toxin

Shiga toxin from *Shigella dysenteriae* is a multimeric, cell-associated protein toxin, which is composed of one subunit A and five subunits B [32]. Figure 5 illustrates the structure of the pentameric subunit B of shiga toxin. Shiga toxin subunit A is cytotoxic and lethal to eukaryotic cells by inhibiting protein biosynthesis. The mechanism relies on the cleavage of the N-glycosidic bond at adenine 4324 in the 28S rRNA [33]. The non-toxic subunit B is responsible for the attachment of the toxin to target cell surfaces, by binding onto N-linked glycoproteins of the target cell surface [34]. Shiga toxin subunit B specifically binds to the glycosphingolipid globotriasylceramide (Gb3). Gb3 is overexpressed in membranes of certain tumor cells, such as metastatic colon or gastric cancer [35]. The affinity of Stx1B

to Gb3 is particularly promising, because of its applicability targeted drug ferry in cancer therapy.

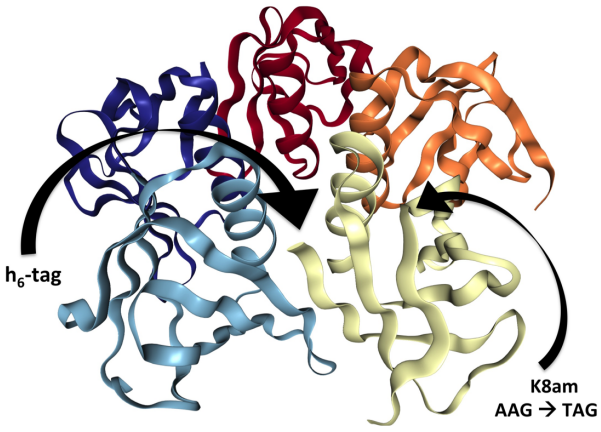


Figure 5: Structure and modifications of shiga toxin 1 subunit B (Stx1B). Stx1B has a calculated molecular weight of 8.7 kDa and forms a pentamer, which possesses a specific affinity for Gb3. We added a C-terminal hexahistidine tag, to purify Stx1B by Ni-chelate affinity chromatography. For incorporation of AzK we exchanged the codon for lysine (AAG) at postion at for an amber stop codon (TAG).

In this study, we aimed to incorporate a reactive ncAA into Stx1B. For this, we exchanged the sequence coding for lysine (CAA) at position eight, with an in-frame amber stop codon (Figure 5). We intended to use a reactive ncAA for bioconjugation of Stx1B with other biomolecules by "click chemistry".
We also added a C-terminal hexahistidine tag, to purifiy Stx1B by Ni-chelate affinity chromatography. It was crucial to add the hexahistidine tag on the C-terminal end, to ensure that solely full-length Stx1B, which arose from the read through at the in-frame amber codon, was purified.

1.6 Main thesis objectives and tasks

Non-canonical amino acids display a more diverse side chain chemistry than canonical amino acids. Incorporation of ncAAs into proteins substantially expands the spectrum of chemical protein modifications. These modifications can result in proteins with new properties and functions [36]. However, incorporation efficiencies of ncAAs are rather poor, because suppressor tRNAs are competing with endogenous release factors. To eliminate this competition, protein expression can be performed in RF-1 knockout strains [12]. RF-1 is not essential in *Escherichia coli*, yet its deletion creates a burden for the cells due to improper translation termination, which causes growth deficiencies [14].
The ultimate goal of this study was to generate an *E. coli* BL21 strain, which carried a tunable release factor 1. Tunable in this regard means that RF-1 was expressed during cell growth and degraded upon induction of target protein expression. To achieve this goal, we fused a C-terminal SspB dependent ClpXP degradation tag *ssrA* (Section 1.3) to the RF-1 encoding gene *prfA*. Our hypothesis was that the ssrA-tagged RF-1 would be

recognised by the adapter protein SspB and would be delivered to the ClpXP protease, which would degrade it. We anticipated that controllable expression of SspB would lead to degradation of RF-1 and consequently to a higher ncAA incorporation efficiency. In this study we tested five different genomic *sspB* constructs, which we stably integrated into the *melAB* locus. Genomic integration was performed by the CRISPR/Cas9 system in combination with λ-Red recombineering. A major challenge was to ensure that *sspB* was tightly regulated, because basal levels of SspB led to a growth deficiency of the strain. After we had generated the strain we tested the incorporation efficiency of BocK and AzK into eGFPx and shiga toxin 1 subunit B, respectively.

Even though a knock-out of RF-1 leads to an efficient read through of the in-frame amber stop codons, this read through consequently results in aberrant cellular proteins. Aberrant proteins of essential genes might have a negative effect on growth behaviour of the cells. To overcome this limitation, we exchanged the stop codons of seven essential genes from amber to ochre/opal stop codons.

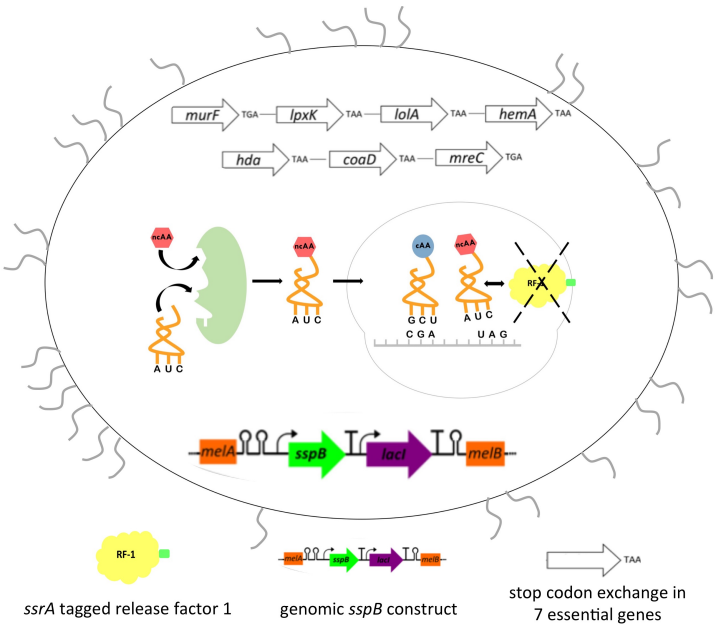


Figure 6: Schematic figure of the generated *E. coli* strain with a tunable release factor 1 for ncAA incorporation by amber stop codon suppression. We integrated lacI and an IPTG inducible *sspB* into the *melAB* locus of *E. coli*. SspB expression triggers the degradation of ssrA-tagged RF-1. When RF-1 is degraded, the competition with the suppressor disappears and the ncAA can be incorporated. To prevent accidental read through at amber termination signals, which results in aberrant cellular proteins, we switched the amber codons of the seven essential genes *murF, lolA, lpxK, hemA, hda, mreC* and *coaD* to either ochre or opal.

The sketch in Figure 6 depicts the schematic features of the modified *E. coli* strain, carrying a tunable RF-1 and the stop codon switch in seven essential genes. The use of a strain with a tunable RF-1 for recombinant protein expression is a promising strategy to incorporate ncAAs more efficiently into industrially relevant proteins.

2 Methods and Material

2.1 Strains

In this project we used the strain *E. coli* BL21 (*E. coli* B F^- *ompT gal dcm lon hsd*S_B ($r_B^- m_B^-$)[$malB^+$]$_{K-12}$(λ^S)) as background for genome editing and as expression host for the incorporation of ncAAs into target proteins. The strain BWEC60 is a descendant of BL21, where the *sspB* gene was replaced with the chloramphenicol resistance gene (Δ*sspB*::Cm^R). The strain BWEC70 carried the gene knockout Δ*sspB*::Cm^R and a C-terminal *ssrA*-degradation tag on the *prfA* gene. Both strains were used with the kind permission of Patrik Fladischer. All subsequent genomic modifications were performed in these two strains. Descriptions of all generated strains are provided in Table 12.
The strain *E. coli* Top10F' (F′ [$lacI^q$ Tn10(tet^R)] mcrA Δ(mrr- hsdRMS- mcrBC) φ80lacZ ΔM15 ΔlacX74 deoR nupG recA1 araD139 Δ(ara-leu)7697 galU galK rpsL(Str^R) endA1 λ^-) was used for cloning.

2.2 Plasmids

For CRISPR/Cas9 mediated genome editing we used a two-vector system that consisted of pTargetF (Addgene #62226) and pCas (Addgene #62225) [26]. For integration of the genomic *sspB* expression construct we modified pTargetF to obtain the desired target plasmids pMEL(500)RFP, pMEL(500)lacI-WsspB, pMEL(500) lacI-MsspB, pMEL(500)lacI-sspB, pMEL(500) $P_{LlacO-1}$-*sspB-lacI*, pMEL(500) P_{lacUV}5-*sspB-lacI* and pMEL(500)lacI. We used the plasmid pUC19 [37] as backbone for the plasmid pUC19-GFPssrA-RFP as well as control for the protein degradation experiment. The plasmids pSCSara-empty, pSCSara-1am, pSCScum-1am, pSCScum-2am and pSCScum-3am were used for expression of recombinant eGFPx with one to three amber stop mutations. pSCSara-stx1B and pSCSara-stx1B-K8am were used for expression and ncAA incorporation into Stx1B. For stop codon exchange we generated the tandem target plasmids pT-murF-lolA-lpxK, pT-hemA-hda and pT-mreC-coaD. All used and generated plasmids are listed with their respective plasmid map in Supplementary Section A.1.

2.3 Primers, enzymes, chemicals and software

All gene constructs and primers were designed with Snapgene (GSL Biotech LLC, Chicago, IL) and ordered from IDT (Integrated DNA Technologies, Coraville, IA). Full lists of primers, gBlocks and gene sequences are provided in Supplementary Table 2, 3 and 4, respectively. All enzymes and chemicals were purchased from Thermo Fisher Scientific (Waltham, MA), Sigma Aldrich (St. Louis, MO), Promega (Fitchburg, WI) or Carl Roth GmbH (Karlsruhe, Germany) unless indicated otherwise. A full list of enzymes, chemicals, buffer solutions, growth media, reagents and instruments is provided in Supplementary Section A.3. The sequence specific sg-RNAs were generated with the webpage ATUM [38]

by setting the preferences for *E. coli* K12 MG1655, wild-type Cas9 with the PAM sequence NGG.

2.4 Molecular biology methods

2.4.1 Preparation of electro-competent *E. coli* cells

The preparation of electro-competent *E. coli* cells was performed according to protocol #35 [39]. For the preparation of electro-competent *E. coli* cells, 10 mL LB-medium were inoculated with the desired *E. coli* strain and incubated over night at 37 °C. If the strain contained a plasmid, the appropriate antibiotic was added to the overnight culture (ONC). The next day a main culture of 500 mL LB-medium was inoculated with the ONC to reach a start D_{600} of 0.1 and incubated at 37 °C under vigorous shaking. After the main culture reached a D_{600} of 0.8, the cells were transferred to 2x 450 mL sterile, pre-cooled centrifuge bottles. The cells were harvested by centrifugation at 6000 rpm for 15 minutes at 4 °C (Avanti J-20XP centrifuge, rotor: JA10). We discarded the supernatant and gently resuspended the cells in 2x 250 mL sterile, ice-cold 10 % (v/v) glycerol. The resuspended cells were again centrifuged at 6000 rpm for 15 minutes at 4 °C and the cell pellet was resuspended in 2x250 mL ice-cold 10 % (v/v) glycerol. The centrifugation step was repeated. The remaining pellet was resuspended in $\sim$ 3 mL 10 % (v/v) glycerol and 50 µL aliquots were stored at -80 °C.

In order to prepare Cas9 containing *E. coli* recombineering cells, 50 µL of the desired electro-competent *E. coli* cells were transformed with 50 ng pCas. Transformation was performed as described in Section 2.4.2 and protocol #35. An ONC with 10 mL LB_{kan50}-medium, 50 µg/mL and the pCas containing cells was incubated at 28 °C. The next day a main culture of 500 mL LB_{kan50}-medium was inoculated with the ONC to reach a start D_{600} of 0.1 and incubated at 28 °C under vigorous shaking. After the main culture reached a D_{600} of 0.4 the expression of the λ-Red recombineering proteins Gamma, Exo and Beta was induced with 25 mL 1 M L-arabinose (50 mM end concentration, Carl Roth) and incubated for another 20 minutes. Subsequently, cell harvest, wash and centrifugation steps were performed as described above.

2.4.2 Electroporation of *E. coli* cells

All plasmids used in this project were brought into electro-competent *E. coli* cells by electroporation. We used $\sim$ 100 ng plasmid for standard electroporations. For genome editing $\sim$ 400 ng of the target plasmid were introduced into *E. coli* recombineering cells harbouring pCas. The competent *E. coli* cells were thawed on ice, mixed with the desired amount of plasmid and transferred into a pre-cooled Gene Pulser®/Micropulser™ electroporation cuvette (Bio-Rad, Hercules, CA). The cuvettes were placed into the Bio-Rad micropulser with the settings EC2, 2.50 kV, 1 pulse. Immediately after the pulse the cells were transferred to pre-warmed SOC medium and regenerated for 45 minutes at 37 °C. The transformed

cells carrying pCas were incubated for 1 hour at 28 °C. After regeneration the cells were plated onto selective plates and incubated over night at 37 °C or 28 °C.

2.4.3 Gel electrophoresis purification of DNA

The generated PCR fragments were separated and visualised via agarose gel electrophoresis. The samples were loaded onto 1 % (w/v) agarose gels (2.5 g agarose, 250 mL 1x TAE-buffer) and separated for 50 minutes at 140 V and 400 mA. GeneRuler 1 kb Plus DNA Ladder (Thermo Fisher Scientific) served as a molecular weight marker. The DNA bands were visualised by UV-exposure of the gels. The desired DNA bands were excised from the gel and purified with the Promega Wizard SV gel and PCR clean-up system according to manufacturer's protocol [40]. Deviations to the protocol were the elution of DNA with 2x 20 µL ddH_2O, instead of elution buffer. The concentration of the samples was estimated via absorbance measurements at 260 nm in an UV/VIS spectrophotometer (Nanodrop, Thermo Fisher Scientific).

2.4.4 Restriction digestion and ligation of DNA

Restriction enzymes (RE) were ordered as FastDigest enzymes from Thermo Fisher Scientific. Each restriction digestion reaction contained 1-2 µg DNA, 2 µL of each FastDigest RE, 2 µL FastDigest Green Buffer and ddH_2O filled up to a total volume of 40 µL. Restriction digestion was performed for 1 h at 37 °C according to manufacturer's protocol. All samples were analysed by gel electrophoresis, purified and quantified as mentioned in Section 2.4.3.

The inserts and plasmid backbones were joined together by T4-DNA-ligation. 200 ng plasmid backbone and a five times molar excess of the insert were used for ligation. The ligation mix contained 1 µL T4 DNA ligase (1 U/µL), 2 µL 10x T4 DNA ligase buffer, the DNA fragments and was filled up with ddH_2O to a total volume of 20 µL. In case the volumes of insert and backbone exceeded the total volume, a volume reduction to 10 µL was performed with a vacuum-centrifuge (Christ, RVC 2-18; 42 °C in five minute steps). Ligation was performed for 20 minutes at 22 °C, followed by heat inactivation of the ligase for 10 minutes at 65 °C. 50 µL electro-competent *E. coli* Top10F' cells were transformed with 4 µL of the ligation mix, regenerated in SOC medium for 45 minutes at 37 °C, plated onto selective plates and incubated over night at 37°C as described above.

2.4.5 Gibson assembly

We performed *in vitro* DNA assembly as described by Gibson et al [41] and according to protocol #122 [42] to generate the three target plasmids for amber stop codon exchange (pT-murF-lolA-lpxK, pT-hemA-hda and pT-mreC-coaD), pUC19-GFP*ssrA*-RFP, pMEL(500)lacI and the expression plasmids pSCSara-stx1B and pSCSara-stx1B-K8am.

The DNA fragments had to carry overlapping homologies with a T_m of 60 °C to ensure efficient assembly. Homologous sequences were added as 5'-extensions onto primers and the fragments were generated by PCR. The vector backbones were either linearized by PCR

or restriction digestion (Section 2.4.4). We mixed the matching DNA fragments at an equimolar concentration, with the smallest fragment having a concentration of 50 ng/µL. The total volume of the DNA fragments was 2.5 µL. If this volume was exceeded we reduced the volume to 2.5 µL by vacuum-centrifugation. We added 7.5 µL "Gibson assembly master mix", which contained 1x ISO buffer, 10 U/µL T5 exonuclease, 2 U/µL Phusion® High-Fidelity DNA Polymerase and 40 U/µL Taq DNA ligase. The assembly mix was incubated for 1 hour at 50 °C. 4 µL of the "Gibson assembly mix" were introduced into electro-competent *E. coli* Top10F' cells, as described previously.

2.4.6 Plasmid isolation, quantification and sequencing

Plasmid isolation was performed with Promega PureYield™ Plasmid Miniprep System according to manufacturer's protocol [43]. Deviations to the protocol were that we used 6 mL ONC and eluted with 2x 20 µL ddH_2O instead of elution buffer. We estimated the concentration of the samples via absorbance measurements at 260 nm in an UV/VIS spectrophotometer (Nanodrop).
All plasmids and DNA constructs were verified by sequencing (Microsynth AG, Balgach, Switzerland). 12 µL DNA with a total concentration of $\sim$ 1 µg and 3 µL primer (10 ng/µL) were mixed together and sent for sequencing. We either used standard primers from Microsynth or designed sequencing primers (Supplementary Table 2, 3). It was essential to use primers that bound approximately 30 bp upstream of the desired sequence, because the sequencing reaction starts after the first $\sim$ 30 bp.

2.5 PCR methods

2.5.1 Standard PCR

All PCR reactions were prepared on ice with the components stated in Table 1. PCR amplification was performed in the GeneAmp® PCR System 2720 thermal cycler (Applied Biosystems, Waltham, CA).

Table 1: Composition of a single reaction for standard PCR.

compound	end concentration	volume [µL]
5x Phusion HF buffer	1x	10
2.5 mM dNTP's	0.2 mM	4
10 ng DNA template	0.2 ng	1
10 µM forward primer	0.5 µM	2.5
10 µM reverse primer	0.5 µM	2.5
2 U/µL Phusion® High-Fidelity DNA polymerase	0.02 U/µL	0.5
ddH_2O	-	29.5
total		50

The temperature profile is stated below in Table 2. According to the manufacturer 2 U/µL

Phusion® High-Fidelity DNA polymerase has an extension time of 15-30 seconds/kb, thus the elongation time was set accordingly to the length of the DNA fragment between ten seconds and three minutes. 10 µL 6x DNA Gel loading dye (Thermo Fisher Scientific) were added to the 50 µL samples before agarose gel-electrophoresis (Section 2.4.3).

Table 2: PCR temperature profile

stage 1	stage 2 (25 cycles)			stage 3
initial denaturation	denaturation	annealing	elongation	final extention
98 °C	98 °C	55 °C- 65 °C	72 °C	72 °C
30 s	10 s	30 s	10 s - 3 min	5 min

2.5.2 Colony PCR

Colony PCR was performed by using the OneTaq Quick-Load 2x Master Mix with Standard buffer from NEB (New England Biolabs, Ipswich, MA) as stated in Table 3. For the screening of positive clones, a total volume of 10 µL was used. To excise and sequence the generated PCR fragments, the reaction was performed as described above, but in a total volume of 50 µL.

Table 3: Composition of a single colony PCR reaction. A single colony was picked from the plate with a sterile pipet tip and dispensed in the colony PCR mix.

compound	end concentration	volume [µL]
OneTaq Quick-Load 2x master mix	-	5
10 µM forward primer	0.2 µM	0.2
10 µM reverse primer	0.2 µM	0.2
ddH_2O	-	4.5
total		10

Colony PCR was primarily performed to analyse genomic integrations at the *melAB* locus or to analyse the correct assembly of the generated plasmids. To check whether integration at the *melAB* locus was successful the primers pBP2082 and pBP2023 were used. Both primers bound outside of the 500 bp homology hooks. To analyse the correct assembly of the plasmids, we designed primers that bound the plasmid backbone as well as the insert. The temperature profile was set according to the protocol, which is shown in Table 4.

Table 4: colony PCR temperature profile

stage 1	stage 2 (30 cycles)			stage 3
initial denaturation	denaturation	annealing	elongation	final extention
94 °C	94 °C	56 °C - 57 °C	68 °C	68 °C
30 s	20 s	30 s	1 min - 4 min	5 min

2.5.3 Overlap extension PCR (OexPCR)

Overlap extension PCR (OexPCR) was performed in two separated PCR reactions, which are listed in Table 5. In the first PCR reaction the overlapping DNA fragments were joined together. These fragments served as template for the second PCR reaction, in which the newly created DNA fragment was amplified.

Table 5: colony PCR temperature profile

	1. reaction		2. reaction	
compound	end conc.	volume [µL]	end conc.	volume [µL]
5x Phusion HF buffer	1x	10	1x	10
2.5 mM dNTP's	0.2 mM	4	0.2 mM	4
10 ng of each DNA template	0.2 ng	1	1. reaction	50
10 µM forward primer	-	-	0.5 µM	2.5
10 µM reverse primer	-	-	0.5 µM	2.5
2 U/µL Phusion® High- Fidelity DNA polymerase	0.02 U/µL	0.5	0.02 U/µL	0.5
ddH_2O		x	-	30.5
Total		50		100

The two respective temperature profiles are stated below in Table 6.

Table 6: Overlap extension PCR temperature profile

stage 1	stage 2 (10 cycles)			stage 3
initial denaturation	denaturation	annealing	elongation	final extention
98 °C	98 °C	55 °C- 65 °C	72 °C	72 °C
30 s	10 s	30 s	30 s - 90 s	5 min
stage 1	stage 2 (25 cycles)			stage 3
initial denaturation	denaturation	annealing	elongation	final extention
98 °C	98 °C	55 °C- 65 °C	72 °C	72 °C
30 s	10 s	30 s	30 s - 90 s	10 min

2.5.4 MASC PCR

To screen clones by MASC PCR two PCR reactions were performed. A single PCR reaction mix was prepared as stated in Table 3. One PCR reaction contained forward primers with TAG at the 3'-end, the other PCR reaction contained the identical forward primers but with TAA/TGA at the 3'-end. We performed genome editing with tandem target plasmids, therefore theoretically up to three genes carried point mutations. To screen single clones for all three point mutations multiplex MASC PCR was performed. Table 7 gives an example of a multiplex MASC PCR reaction for screening of a colony from BWEC72{pCas, pT-murF-lolA-lpxK}.

Table 7: Composition of a MASC PCR reaction to screen single clones.

Reaction 1	Reaction 2	end conc. [µM]	expected PCR product size
pBP2062 *murF* fwd TAG	pBP2063 *murF* fwd TGA	0.2	300 bp
pBP2064 *murF* rev	pBP2064 *murF* rev	0.2	
pBP2065 *lolA* fwd TAG	pBP2066 *lolA* fwd TAA	0.2	600 bp
pBP2066 *lolA* rev	pBP2066 *lolA* rev	0.2	
pBP2068 *lpxK* fwd TAG	pBP2069 *lpxK* fwd TAA	0.2	400 bp
pBP2070 *lpxK* rev	pBP2070 *lpxK* rev	0.2	
5 µL OneTaq Quick-Load 2x Master Mix	5 µL OneTaq Quick-Load 2x Master Mix		
4.6 µL ddH_2O	4.6 µL ddH_2O		
total		10	

2.6 CRISPR/Cas9 mediated genome editing and plasmid curing

2.6.1 Genome editing by CRISPR/Cas9

pCas containing electro-competent recombineering cells were used for genome editing. 50 µL electro-competent recombineering cells were transformed with ~ 400 ng of the desired target plasmid. The transformation mixture was plated onto selective plates and incubated at 28 °C for two days. Clones carrying the correct genomic integration were identified by colony PCR (Section 2.5.2). Bands that migrated at the expected size were excised, purified and sent for sequencing.

2.6.2 Target plasmid and pCas curing

We performed plasmid curing in one sequence verified clone of each strain in order to prepare it for further experiments. To cure the cells from the target plasmid we inoculated 10 mL LB_{kan50}-medium containing 0.5 mM IPTG with a clone and incubated over night at 28 °C. To ensure the growth of single cells on the LB-agar plates, appropriate dilutions of the ONC were made, plated onto LB_{kan} agar plates and incubated for two days at

28 °C. To test whether the growing cells solely contain pCas, single colonies were streaked onto $LB_{kan,spec}$ agar plates and incubated over night at 28 °C. No growth on these plates indicated that the cells were cured from the target plasmid. Additional to streaking the clones on selective agar plates, we also performed colony PCR, for the verification of the absence of the target plasmid. The forward primer pBP2026, which bound the plasmid backbone and a reverse primer that bound inside the editing fragment were used. Clones that neither grew on $LB_{kan,spec}$ selective agar plates nor showed a visible band on the agarose gel after colony PCR were picked and an ONC was prepared. In the following step, the cells were cured from the heat-sensitive plasmid pCas. Therefore, 10 mL LB-medium were inoculated with the target plasmid free clone and incubated over night at 37 °C. The next day the ONC was diluted appropriately, plated onto LB agar plates and incubated over night at 37 °C. To check if the clones were actually cured from pCas, we picked 10 clones, streaked them onto LB_{kan} agar plates and incubated them for two days at 28 °C. Additionally, colony PCR of those 10 clones was performed with primers that bound specifically pCas. We used the forward primer pBP2110 that bound within Kan^R and the reverse primer pBP2111 that bound within the Cas9 gene. A clone that neither grew on LB_{kan} selective agar plates nor showed a visible band on the agarose gel after colony PCR was chosen. We prepared electro-competent cells from this clone to use it for subsequent experiments.

2.7 Plasmid construction

2.7.1 Construction of the target plasmids for integration of the genomic *sspB* construct at the *melAB* locus

The integration fragments were assembled from multiple parts, which were either generated by PCR (Table 10) using the primers shown in Supplementary Table 2 and 3 or ordered as gBlocks from IDT (Integrated DNA technologies, Coralville, IA). The editing fragments were generated by OexPCR of the PCR fragments and are shown in Table 10. The *sspB* adapter protein sequence was PCR amplified from p15a_pT5cym_SspB and the *lacI* repressor sequence was PCR amplified from pQE80L. Each time 10 ng of the plasmid were used as template. 2 µg of pMEL(500)RFP were digested with PaeI (GCATG'C) and KpnI (GGTAC'C), to release the RFP insert, purified by gel purification and subsequently ligated with the integration fragment digested with the same restriction enzymes. 5 µL of the ligation mix were used for transformation of electro-competent *E. coli* Top10F' cells, as described previously.

2.7.2 Construction of the tandem target plasmids for stop codon exchange

We performed amber stop codon exchange to ochre or opal stop codons in seven essential genes by CRISPR/Cas9. For this we ordered the seven editing cassettes, which consisted of locus specific sg-RNAs and mutation fragments as gBlocks from IDT. Each editing cassette

carried overhangs to either the plasmid backbone or another editing cassette. We used the seven editing cassettes and pTargetF to assemble the three tandem target plasmids pT-hemA-hda, pT-mreC-coaD and pT-murF-lolA-lpxK. Plasmid maps and of all three target plasmids and the sequences of the editing cassettes are depicted in Supplementary Figure 16, 17, 15 and Supplementary Table 4, respectively.
Gibson assembly was performed to generate the three target plasmids. pTargetF was digested with BamHI and HindIII (A'AGCTT) to remove the sg-RNA and linearise the plasmid. Linear pTargetF and the respective editing fragments were joined together by Gibson assembly (see Section 2.4.5). Correctly assembled plasmids were analysed by colony PCR and sequencing.

2.7.3 Construction of pUC19-GFP*ssrA*-RFP

To analyse SspB-mediated protein degradation, we designed and constructed the plasmid pUC19-GFP*ssrA*-RFP, which carried a *ssrA*-tagged sfGFP and TagRFP. The plasmid was generated from multiple parts, by Gibson assembly. The sf*gfp-ssrA* gene sequence was ordered as gBlock from IDT and carried overhangs for the Tag*rfp* gene and the pUC19 plasmid backbone. The Tag*rfp* sequence was PCR amplified from pMEL(500)RFP by using the primer pBP2096 and PBP2097. pUC19 was linearised by PCR, by using the primers pBP2100 and pBP2101. The resulting DNA fragments were joined together by Gibson assembly as stated in Section 2.4.5. Correctly assembled plasmids were analysed by colony PCR and sequencing.

2.7.4 Construction of the Stx1B expression plasmids

For wild-type Stx1B and Stx1B-K8am expression we generated the plasmid pSCSara-stx1B and pSCSara-stx1B-K8am by Gibson assembly. pSCSara-1am was digested with BglII (A'GATCT) to release GFPx40. The genes *stx1B* and *stx1B*-K8am contained a C-terminal hexahistidine tag and were ordered as gBlocks from IDT. The linearised plasmid backbone and PCR-amplified *stx1B* fragment were joined together by Gibson assembly as stated in Section 2.4.5. Correctly assembled plasmids were analysed by colony PCR and sequencing.

2.8 Fluorescence spectrophotometry

To test protein degradation efficiency 50 µL electro-competent BWEC60, BWEC61, BWEC 62, BWEC63, BWEC64 and BWEC65 cells were transformed with pUC19 and pUC19-GFP*ssrA*-RFP, respectively. For BocK incorporation into eGFPx 50 µL electro-competent BWEC71, BWEC72, BWEC65, BWEC73 and wild-type BL21 were transformed with pSCScum-empty and pSCScum-1am, respectively. ONCs were prepared by inoculating M9 medium supplemented with 0.05 g/L amino acids and 100 µg/mL ampicillin or 50 µg/mL kanamycin with each of the desired strains and incubation over night at 37 °C. The next day main cultures were prepared with a start D_{600} of 0.1 and incubated at 37 °C under vigorous shaking.
The plate reader Synergy Mx Microplate reader (BioTek Instruments Inc, Winooski, VT)

with the software program Gen5 2.09 was used to measure the change of optical density and fluorescence of the samples over time. We measured D_{600}, GFP fluorescence emission at 540 nm (excitation at 485 nm) and RFP fluorescence emission at 590 nm (excitation at 510 nm). The gain was set at 87. Six replicates of each sample were measured in 96-well microplates for fluorescence-based assays (Thermo Fisher Scientific).
For the protein degradation experiment the samples were measured over a time course of seven hours. D_{600}, GFP and RFP fluorescence were measured at the start of incubation, before induction, two hours after induction and at the end of incubation. To investigate the incorporation efficiency of BocK into eGFPx, D_{600} and GFP fluorescence were measured over the time course of 24 hours, before induction, 4 hours after induction and at the end of the incubation.

2.9 Protein methods

2.9.1 Protein expression and cell harvest

All protein expressions were performed in a total volume of 500 mL. 25 mL LB_{kan50}-medium were inoculated with a single colony of an *E. coli* strain containing the expression plasmid and incubated at 37 °C over night. The next day the main culture of 500 mL LB_{kan50}-medium was inoculated with the ONC to reach a start D_{600} of 0.1 and incubated at 37 °C and 125 rounds per minute (rpm). When the main culture reached a D_{600} of 0.8 target protein and SspB expression were induced by addition of 0.5 mM IPTG. For incorporation of a non-canonical amino acid we also induced the aminoacyl-tRNA synthetase expression with 0.2% (v/v) L-arabinose or 0.01 mM cumate and addition of the desired ncAA (5 mM). The considerably small volume of cumate was dispensed, by pipetting a small drop onto the surface of the cultures. 50 mM BocK (Bachem Holding AG, Bubendorf, Switzerland) or AzK (Iris Biotech GmbH, Marktredwitz, Germany) solutions were prepared by dissolving the respective amino acid in 10 mM sodium-phosphate buffer. After induction the cells were incubated at 28 °C over night. For SDS-PAGE analysis samples were taken before induction that contained the number of cells equivalent to a cell density D_{600} of 1.
The next day the cells were centrifuged at 4 °C and 3500 rpm for 20 minutes (rotor JA-10), the supernatant discarded and the pellet resuspended in 10 mL 0.9 % NaCl. The resuspended cell solution was transferred to 50 mL reaction tubes and centrifuged again at 4 °C and 3500 rpm for 20 minutes. The supernatant was discarded and the cell pellet was stored at -21 °C until protein purification.

2.9.2 Cell disruption by sonication

The cells were disrupted by sonication. Therefore, the cell pellet was resuspended in 25 mL lysis buffer (50 mM NaPi, 150 mM NaCl, 10 mM imidazole, pH 7.2) and subsequently sonicated for 10 minutes with an output power of 7 and 75 % duty cycle for up to 10 minutes with the Branson Sonifier 250 (Emerson Electric, St. Louis, MO). After sonication

the cell debris was removed by centrifugation at 4 °C and 20000 rpm for 30 minutes (rotor JA-25.50). Both, soluble and insoluble protein fractions were analysed by SDS-PAGE and the soluble fraction was used for immobilized metal chelate affinity chromatography (IMAC).

2.9.3 Protein purification by Ni-chelate affinity chromatography

All proteins in this project carried a hexahistidine tag and were purified by Ni-chelate affinity chromatography. For gravity flow purification disposable 10 mL polypropylene columns (Thermo Fisher Scientific) were packed with 2 mL Ni-NTA agarose beads (Qiagen). Separate columns were used for each variant protein.
The column was washed with 3x 35 mL ddH_2O and equilibrated with 3x 35 mL lysis buffer. The column was loaded with 25 mL of the soluble fraction and the flow through was collected in a 50 mL reaction tube. The column was washed with 20 mL wash buffer (50 mM NaPi, 150 mM NaCl, 30 mM imidazole, pH 7.2) and the flow through was collected. The target protein was eluted with 10x 1 mL elution buffer (50 mM NaPi, 150 mM NaCl, 300 mM imidazole, pH 7.2) and 4x 1 mL cleaning buffer (50 mM NaPi, 150 mM NaCl, 500 mM imidazole, pH 7.2). The resin was washed and regenerated with 20 mL cleaning buffer, 20 mL 0.5 mM NaOH and 20 mL 20 % (v/v) ethanol and stored at 4 °C. Buffer exchange was performed with all elution samples. The elution fractions were pooled and concentrated with VivaSpin 3000 MWCO columns (Sartorius AG, Göttingen, Germany). Concentration of the proteins was performed according to manufacturer's protocol [44]. 1x PBS buffer (137 mM NaCl, 12 mM phosphate, 2.7 mM KCl, pH 7.4) was used for buffer exchange and desalting. The concentrated protein solution was transferred to Zeba^{TM} Spin desalting columns 7K MWCO (Thermo Fisher Scientific) and buffer exchange was performed according to manufacturer's protocol [45]. The purified protein samples were stored in aliquots of 50 µL at 4 °C as well as -21 °C.

2.9.4 Protein concentration determination

Protein concentration was measured either with Bradford assay or UV/Vis spectrometry by NanoDrop. For concentration measurement by NanoDrop, the setting, in which an absorption of 1 corresponded to 1 mg/mL was used. Protein concentration of the lysate, soluble fraction, insoluble fraction, flow through after load, wash fraction, elution fraction and cleaning fraction was determined. Before each measurement we blanked with the appropriate buffers.
For protein concentration measurements by the Bradford assay, the protein assay dye reagent concentrate (Biorad) was used. The reagent was diluted 1:5 and filtered with a 0.2 µm non-sterile PES membrane. Bovine serum albumin (BSA) served as protein standard. BSA was either dissolved in lysis buffer or in 6 M urea. Protein concentration of the protein standards was measured by UV/Vis spectrometry (1 Abs = 1 mg/mL). Three technical replicates were measured each and an average concentration was calculated. Dilutions of both protein standards ranging from 1 mg/mL to 0.5, 0.25, 0.125 and 0.0625 mg/mL

were prepared. For the Bradford assay, all samples were measured in triplicates in transparent 96-well plates. 10 µL of appropriately diluted sample were mixed with 200 µL of the Bradford reagent and measured at 595 nm in a EON plate reader (BioTek). The protein concentration of the samples was calculated from the standard curve.

2.9.5 SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed as described by Laemmli [46]. The protein samples were diluted to an end concentration of 2 µg/µL and a volume of 18.75 µL and mixed with 6.25 µL 4x SDS loading dye to a total sample volume of 25 µL. The previously taken cell sample (before induction with a D_{600} of 1), was mixed with 50 µL 1x SDS loading dye. All samples were heated at 95 °C for 7 minutes and 12 µL were loaded on a NuPAGE 4-12 % Bis-Tris gel (Invitrogen AG, Carlsbad, CA). 5 µL PageRuler Prestained Protein ladder (Thermo Fisher Scientific) was used as a molecular weight marker. The samples were separated for 35 minutes at 200 V by a vertical electrophoresis unit SE250 Hoefer (Hoefer Inc., Holliston, MA), which was filled with 1x MES buffer (Thermo Fisher Scientific).
After SDS-PAGE, the gel was stained with Comassie Staining solution (2.5 g/L Brilliant-G250, 7.5 % (v/v) acetic acid, 50 % (v/v) EtOH) for 20 minutes and destained two times destaining solution (7.5 % (v/v) acetic acid, 20 % (v/v) EtOH).

2.9.6 Western blot and immunodetection

For western blot analysis we performed SDS-PAGE analysis of the samples (Section 2.9.5) and blotted the gels onto nitrocellulose membranes with a Hoefer^{TM} TE22 Mini Tank Blotting Unit. The inner chamber of the blotting unit was filled with transfer buffer (25 mM Tris, 192 mM glycine, 10 % (v/v) methanol). The blotting parameters were set to 500 V, 500 mA, 50 W and 90 minutes. After blotting the nitrocellulose membrane was stained with PonceauS to evaluate transfer efficiency and subsequently destained with 1x tris-buffered saline (TBST, 25 mM Tris, 150 mM NaCl, 0.03 % (v/v) Tween 20, pH 7.5).
For immunodetection the nitrocellulose membrane was blocked for 45 minutes and vigorous shaking, with 5 % (w/v) milk powder (Roth) dissolved in TBST. Subsequently, the membrane was rinsed with TBST. The membrane was incubated with a primary anti-His monoclonal antibody solution (1:3000 dilution in 3 % (w/v) BSA+TBS, 20 mL) at 4 °C over night. The next day, the membrane was washed three times with 1x TBST buffer for 10 minutes and the secondary antibody (20 mL, anti-mouse HRP, 1:10000 in TBS-BSA 3 % (w/v)) was added. The membrane was incubated for 1 hour at room temperature and washed twice for 10 minutes with 1x TBST buffer. For immunodetection the SuperSignal West Dura Extended Duration Substrate (Thermo Fisher Scientific) was used. The membrane was incubated with 1.5 ml of the solution and the chemiluminescent signal was then captured with the G:BOX Bioimaging System (Syngene, Cambridge, UK).

2.9.7 Protein activity assay

Activity of wild-type Stx1B and Stx1B-K8[AzK] were analysed by differential scanning fluorimetry (DSF). Stx1B binds globotriasylceramide (Gb3), thus it was used as ligand for this assay. Gb3 (Elicityl Oligotech, Crolles, France) was diluted with sterile ddH_2O to a concentration of 50 mM. DSF reactions were prepared according to Table 8.

Table 8: Composition of a single reaction for the thermal shift assay. We also prepared samples that did not contain Gb3. Instead of Gb3 we added 1x PBS to the reaction.

compound	end concentration	volume [µL]
50 mM Gb3	1 mM	0.5
Stx1B, Stx1B-K8[AzK]	0.3 mg/mL	2.8 WT, 3.04 K8[Azk]
1x PBS	-	20.4 to WT, 20.16 to K8[AzK]
100x SYPROorange protein gel stain	5x	1.32
total		25

DSF was performed with a 7500 Real Time PCR System (Applied Biosystems Perkin-Elmer Corp., Foster City, CA). Protein samples of 0.3 mM were measured in 1xPBS with 8x Sypro Orange (Sigma-Aldrich) in the presence (1 mM) or absence of Gb3. Melting temperatures of triplicate measurements were analysed with the Protein Thermal Shift software v1.3 (Applied Biosystems).

2.9.8 Copper(I)-catalysed azide-alkyne cycloaddition (CuAAC)

We performed copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) to confirm AzK incorporation into Stx1B. Therefore, Stx1B-K8[AzK] was conjugated with the fluorescence dye, sulfo-cyanine-3-alkyne. The wild-type protein served as a negative control. Table 9 shows the used compounds for a standard click reaction. Sodium ascorbate should be freshly prepared prior to running the reaction. Since $CuSO_4$ is harmful for the proteins, it was mixed with THPTA before adding it to the reaction. The fluorescence dye was light sensitive, hence the click reaction was performed under light protection. The samples were incubated at 22 °C for 1 hour and constant shaking (550 rpm). At the end of the reaction EDTA was added to a final concentration of 5 mM and the samples were prepared for SDS-PAGE analysis as already described in Section 2.9.5. Since we used a fluorescence dye, we were able to confirm successful bioconjugation by exposing the gel to UV-light.

Table 9: Composition of a standard reaction for copper(1)catalyzed azide-alkyne cycloaddition. Reaction mixes were prepared with either wild-type Stx1B or Stx1B-K8[AzK].

compound	end concentration	volume [µL]
10 mM $CuSO_4$	0.5 mM	1
100 mM sodium ascorbate	5 mM	1
1 mM sulfo-cyanine-3-alkyne	0.07 mM	1.5
Stx1B, Stx1B-K8[AzK]	0.19 mg/mL	15.5
50 mM THPTA	2.5 mM	1
total		20

3 Results

3.1 The design of the genomic *sspB* construct allowed controllable RF-1 degradation

We chose the SspB-dependent ClpXP protein degradation system to perform controlled RF-1 degradation. As described in Section 1.3 the adapter protein SspB recognizes proteins carrying an *ssrA*-tag and facilitates degradation of the tagged protein by the ClpXP protease. To regulate protein degradation, we knocked out the endogenous *sspB* and integrated a controllable sspB allele. We chose the melibiose operon (*melAB*) as locus for genomic integration. In a preliminary study we confirmed that *melAB* was a suitable locus for stable integration of the genomic sspB construct, because it is not essential for the metabolism of *E. coli*. We used the site-specific sg-RNA, which we designed and constructed in the previous study [31]. The sg-RNA was provided on the plasmid pMEL(500)RFP. As depicted in Figure 7a, the sg-RNA consisted of the locus-specific 20 bp target sequence and a tracr-RNA and was under the control of the promoter $P_{J23119(SpeI)}$ and terminator $T_{spy-tem}$. Genomic integration at the *melAB* locus was designed such that 2539 bp spanning bases 140-1355 of *melA* and bases 1-1210 of *melB* were deleted.

For genomic integration we designed editing fragments, which consisted of the two genes *sspB* and *lacI*. *sspB* was placed under the control of an IPTG inducible *tac*-promoter and the *rrnB* T1 terminator [47], [48]. To ensure tight regulation of the *tac*-promoter we added the *lacI* suppressor gene upstream of *sspB*. *lacI* was regulated by the constitutive $lacI^q$ promoter and *rrnB* T1 terminator. We designed three different editing fragments, with ribosome binding sites (RBSs) of different strengths for the *sspB* gene. Figure 7d-f illustrates the design of the editing fragments with three different RBSs.

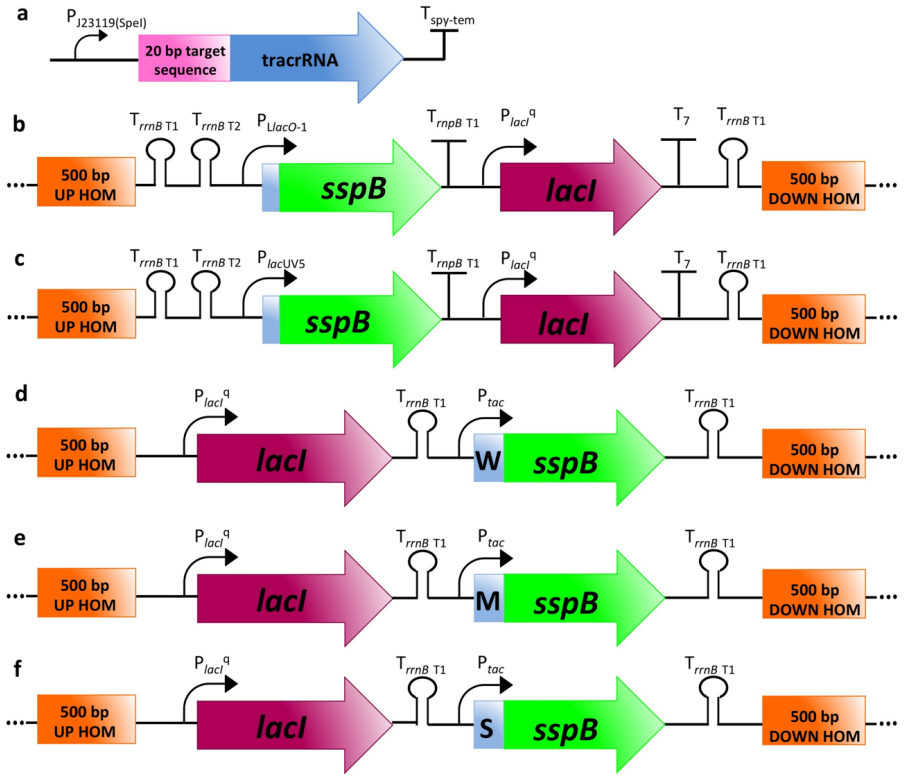


Figure 7: Editing fragments for the construction of the different target plasmids (b-f). Each editing fragment was combined with a *melAB* specific sgRNA cassette, which is depicted in panel a (see text for details). The resulting editing cassettes were inserted into pTargetF as outlined in the text.

Tight regulation of *sspB* was crucial to avoid premature degradation of RF-1. The initial constructs (Figure 7 d-f) showed leaky expression, which will be explained in more detail in Section 3.3.1. To overcome the leaky expression of SspB, we designed two alternative editing fragments. Figure 7 b-c illustrates the design of the two alternative editing fragments. The main difference in design was the changed order of *sspB* and *lacI*. *lacI* was initially located upstream of *sspB* and since it was constitutively transcribed, read-through at the terminator led to accidental transcription of *sspB*. To assure tight regulation of SspB expression, accidental *sspB* transcription had to be prevented. Positioning *sspB* upstream of *lacI*, we eliminated this flaw. Furthermore we flanked the editing fragments with double terminators, to avoid polar effects at the *melAB* locus. The *rrnB* T1T2 terminator and T7 and *rrnB* T1 terminator were positioned upstream and downstream of the *sspB-lacI* cassette, respectively.

One editing fragment carried the *sspB* gene under the control of the $PL_{lacO} - 1$ promoter, which contained two *lacO* binding sites and the endogenous *sspB* ribosome binding site. The *sspB* gene of the second newly designed editing fragment was controlled by the *lacUV5* promoter. We used the endogenous *rnpB* T1 terminator for *sspB*, in both editing fragments. *lacI* was regulated by the constitutive promoter $lacI^q$ and the terminators T7 and *rrnB* T1. We added an ochre stop codon after *lacI* and an opal stop codon after *sspB*, to ensure proper translation termination.

As control we used a strain with a tagged RF-1 and the integrated editing fragment, but did not contain the *sspB* copy. For this we designed an editing fragment that solely con-

tained the *lacI* gene.

To construct the target plasmids, all plasmid parts were generated by PCR and spliced together by overlap extension PCR. This resulted in six editing fragments. Table 10 gives an overview of the used primers and generated plasmid parts. The editing fragments were cloned into pMEL(500)RFP to obtain the six target plasmids pMEL(500)lacI-WsspB, pMEL(500) lacI-MsspB, pMEL (500)lacI-SsspB, pMEL(500)PLlacO-1-sspB-lacI, pMEL(500)PlacUV5-sspB-lacI and pMEL(500)lacI. Maps of all seven generated target plasmids are provided in Supplementary Section A.1.1.

Table 10: PCR fragments. Each editing fragment was assembled from multiple parts, which were generated by PCR.

Name	Forward primer	Reverse primer
P_{tac}-W*sspB*	pBP2075	pBP2078
P_{tac}-M*sspB*	pBP2076	pBP2078
P_{tac}-S*sspB*	pBP2077	pBP2078
lacI-rrnB T1	pBP2079	pBP2080
UP HOM-*lacI*	pBP2132	pBP2133
lacI-rrnB T1	pBP2134	pBP2135
gBlock $P_{L lacO-1}$-*sspB-lacI*	pBP2104	pBP2105
UP HOM-P_{lacUV5} – *sspB*	pBP2104	pBP2106
P_{lacUV5}-*sspB-rrnB* T1	pBP2107	pBP2105

Table 11 shows which components and primers were used for each editing fragment. For assembly of the target plasmids we cut pMEL(500)RFP and the six editing fragments with the restriction enzymes KpnI and BcuI, ligated with a T4-DNA ligase and electroporated the resulting constructs into electro-competent *E. coli* Top10F' cells. Clones carrying the correctly assembled target plasmids were analysed by colony PCR and sequence verified.

Table 11: Overlap extension PCR fragments for restriction cloning.The generated single parts were joined together by overlap extension PCR and resulted in the six editing fragments *lacI-WsspB*, *lacI-MsspB*, *lacI-SsspB*, *lacI*, $P_{L lacO-1}$-*sspB-lacI* and P_{lacUV5}-*sspB-lacI* .

Name	components	Forward primer	Reverse primer
lacI-W*sspB*	P*tac*-W*sspB*, *lacI-rrnB* T1	pBP2079	pBP2078
lacI-M*sspB*	P*tac*-M*sspB*, *lacI-rrnB* T1	pBP2079	pBP2078
lacI-S*sspB*	P*tac*-M*sspB*, *lacI-rrnB* T1	pBP2079	pBP2078
P_{lacUV5}-*sspB-lacI*	UP HOM-P_{lacUV5}-*sspB*, P_{lacUV5}-*sspB-rrnB* T1	pBP2104	pBP2105
lacI	UP HOM-*lacI*, *lacI-rrnB* T1	pBP2132	pBP2135

3.2 The genomic *sspB* construct stably integrated at the *melAB* locus by CRISPR/ Cas9

We integrated the genomic *sspB* construct at the *melAB* locus by CRISPR/Cas9 combined with λ-Red recombineering. For genome editing we used the two strains BWEC70 and BWEC60, which were descendants of *E. coli* BL21. BWEC70 carried the gene knock-out Δ*sspB*::Cm^R and a C-terminal *ssrA*-degradation tag on the *prfA* gene encoding RF-1. By integration of an IPTG inducible *sspB* copy, RF-1 became degradable. BWEC60 also carried the gene knock-out Δ*sspB*::Cm^R but RF-1 did not contain an *ssrA*-tag and could thus not be degraded by the *ssrA*-tag dependent ClpXP protease. We transformed the strains BWEC70 and BWEC60 with pCas, which was necessary for CRISPR/Cas9 mediated genome editing, and generated electro-competent recombineering cells as described in Section 2.4.1. pCas encoded the endonuclease Cas9, which introduced the double strand break. As well, the genes exo, bet and gam were encoded on pCas, which were necessary for gap repair by λ-Red recombineering.

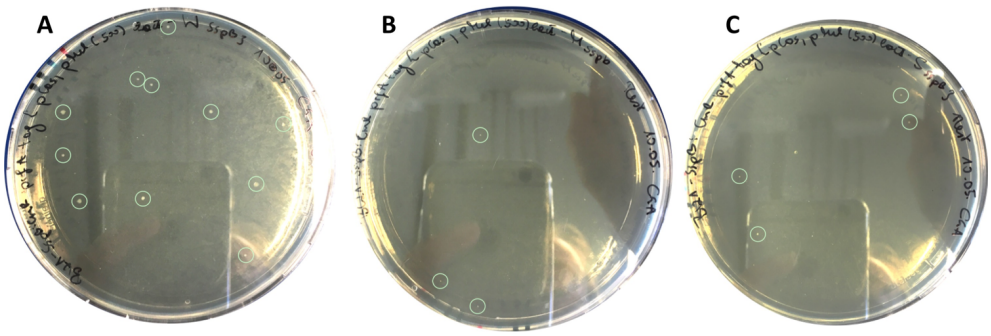


Figure 8: Increasing strengths of the RBSs cause growth deficiencies and problems for genome editing of BWEC70. A: Colonies of BWEC74 showed normal growth and contained the correctly integrated genomic *sspB* construct. **B:** Colonies of BWEC75 were smaller than **A** and not every clone contained the genomic *sspB* construct. **C:** Colonies of BWEC76 were smaller than **A** and did not contain the genomic *sspB* construct. However, those clones carry the *sspB* gene on the target plasmid and leaky expression of this copy leads to the growth phenotype.

We performed genome editing by transforming the recombineering cells with each of the six previously assembled target plasmids. We confirmed the correct integration of the genomic *sspB* construct and *lacI* by colony PCR and sequencing. By using CRISPR/Cas9 together with λ-Red recombineering we obtained the desired strains with high integration efficiencies. However, it is noteworthy that genome editing of BWEC70 showed considerable differences when using the target plasmids pMEL(500)lacI-WsspB, pMEL(500)lacI-MsspB and pMEL(500)lacI-SsspB. All three target plasmids were identical, except that the translation of *sspB* was controlled by RBSs of three different strengths. Cells containing pMEL(500)lacI-SsspB showed a considerable growth deficiency compared to cells carrying pMEL(500)lacI-WsspB (Figure 8). Sequencing also confirmed that the clones did not carry the desired integration of *lacI-SsspB*. Since *sspB* was provided on the target plasmid, it was leaky expressed and caused the growth deficiency. Obviously, high levels of SspB were

fatal for *E. coli* cell growth. The growth deficiency of BWEC75 and BWEC76 strongly indicated that SspB was basally expressed and thus harmed *E. coli* cell growth.

Eventually, we were able to generate BWEC61, BWEC62, BWEC63, BWEC64, BWEC65, which carried the genomic *sspB* expression construct, but no *ssrA*-tagged RF-1, hence RF-1 could not be degraded by the *ssrA*/SspB-ClpXP degradation system. We also generated BWEC71, BWEC72, which carried the *ssrA*-tagged RF-1 and the inducible, genomic *sspB* construct. These strains, were theoretically able to degrade RF-1 by the *ssrA*/SspB-ClpXP degradation system. BWEC73 carried a genomic copy of the *lacI* repressor gene at the *melAB* locus and a *ssrA*-tagged RF-1, but since it lacked an *sspB* copy, RF-1 could not be degraded. Table 12 summarises all generated and used strains in this study.

Table 12: Description of all strains generated and used over the course of this project. All strains had a BL21 background and carried the gene knockout Δ*sspB*::Cm^R .

Name	genotype		note
BWEC60[1)]	*prfA*		wild-type *prfA*
BWEC70[1)]	*prfA-ssrA*		*ssrA*-tagged RF-1
BWEC61	*prfA*	Δ*melAB*::*lacI*-W*sspB*	inducible *sspB* construct[2)]
BWEC62	*prfA*	Δ*melAB*::*lacI*-M*sspB*	"
BWEC63	*prfA*	Δ*melAB*::*lacI*-S*sspB*	"
BWEC64	*prfA*	Δ*melAB*::$P_{L lacO-1}$*sspB-lacI*	"
BWEC65	*prfA*	Δ*melAB*::P_{lacUV5}*sspB-lacI*	"
BWEC71	*prfA-ssrA*	Δ*melAB*::$P_{L lacO-1}$*sspB-lacI*	tunable RF-1[3)]
BWEC72	*prfA-ssrA*	Δ*melAB*::P_{lacUV5}*sspB-lacI*	tunable RF-1
BWEC73[4)]	*prfA-ssrA*	Δ*melAB*::*lacI*	*ssrA*-tagged RF-1
BWEC74	*prfA-ssrA*	Δ*melAB*::*lacI*-W*sspB*	
BWEC75	*prfA-ssrA*	Δ*melAB*::*lacI*-M*sspB*	
BWEC76	*prfA-ssrA*	Δ*melAB*::*lacI*-S*sspB*	
BL21			wild-type

[1)] all genomic modifications were performed in this strain

[2)] this strain carried an inducible *sspB* construct, but RF-1 could not be degraded, due to the missing *ssrA*-tag; this strain was used for protein degradation

[3)] this strain carried an inducible *sspB* construct and an *ssrA*-tagged RF-1, thus RF-1 was degradable; this strain was used for protein expression and incorporation of ncAAs

[4)] this strain carried a *ssrA*-tag, but no *sspB* construct; this strain was used for protein expression and ncAA incorporation

To perform further experiments, it was necessary to cure the generated strains from the target plasmids and pCas. Curing in this context meant removing the plasmids from the cells after genome editing. pCas encoded a sg-RNA targeting the origin of replication of the target plasmid. Transcription of this sg-RNA was regulated by an IPTG inducible promoter. At first transcription of the sg-RNA was induced by IPTG to cure the cells from the target plasmid. pCas also contained a heat-sensitive origin of replication. The

shift of incubation temperature from 30 °C to 37 °C led to the loss of pCas. After we confirmed the loss of the target plasmid, we incubated the cells at 37 °C to cure pCas. All strains were cured from pCas and the target plasmids. Only BWEC71 could not be cured, because the strain lost pCas before the target plasmid could be cured. As described above, pCas encodes a sg-RNA necessary for target plasmid curing.

3.3 Protein degradation was tunable by SspB expression

We could not directly analyse if RF-1 was degraded and whether the degradation was caused by the induced expression of SspB. Therefore we designed a proof of concept experiment that confirmed controllable *sspB* mediated protein degradation in the generated strains. We also tested the degradation efficiency of the system and if SspB was basally expressed. To observe protein degradation we used superfolder green fluorescence protein sf*gfp* carrying a C-terminal *ssrA*-degradation tag. Degradation of GFP-*ssrA* would thus result in a decreased fluorescence signal. It was crucial to assure that the decrease of the signal was solely caused by protein degradation and not by premature mRNA degradation. For this we co-expressed the red fluorescence protein TagRFP. TagRFP was not *ssrA*-tagged and therefore steadily expressed throughout the whole experiment. Only GFP-*ssrA* was degradable by the SspB mediated ClpXP degradation system. In this way, the bicistronic mRNA assured that a decrease in fluorescence signal was solely caused by SspB mediated protein degradation. The sfGFP-*ssrA*-RFP construct was controlled by the constitutive *lac* promoter and the *rrnB* T1 terminator. We provided the construct on plasmid pUC19-GFP*ssrA*-RFP (Section 2.7.3). Figure 9 illustrates the principle of the sfGFP-*ssrA* degradation experiment. Before induction of SspB both sfGFP-*ssrA* and TagRFP would be constitutively expressed. After induction, the *ssrA*-tagged sfGFP would be degraded, which would result in a decrease of GFP fluorescence signal.

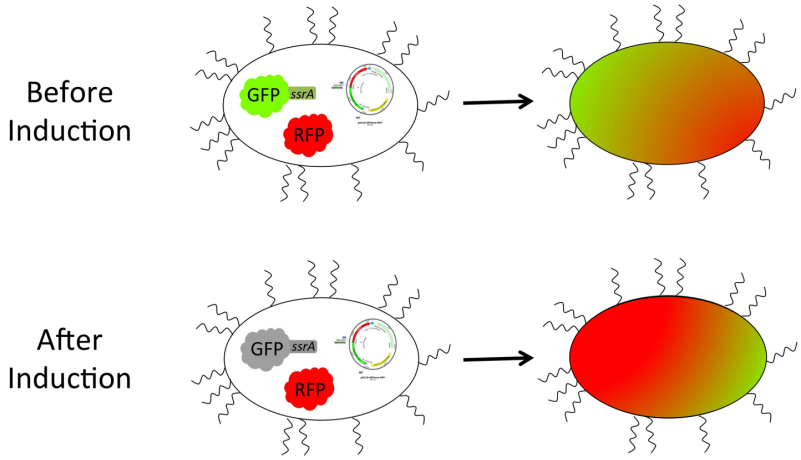


Figure 9: SspB mediated sfGFP-*ssrA* degradation was determined by fluorescence measurements. We transformed the generated strains with pUC19-GFP*ssrA*-RFP and analysed the inducible degradation of sfGFP-*ssrA*. Before induction of SspB, sfGFP-*ssrA* and TagRFP were constitutively expressed, thus the cells exhibited a mix of red and green fluorescence. After induction of SspB, sfGFP-ssrA was degraded and decrease in green fluorescence was visible.

The respective strains were transformed with either pUC19 or pUC19-GFP*ssrA*-RFP. Strains containing pUC19 were used as control to measure the autofluorescence signal of the cells. The measured signal of these strains was later subtracted from the fluorescence signal of the strains expressing sfGFP-*ssrA* and TagRFP. We used the strain BWEC60 as additional control, since it completely lacked the *sspB* gene. We expected this control strain to be unable to degrade sfGFP-*ssrA* and hence to show the same fluorescence signal throughout the experiment, regardless of induction. We also used the fluorescence signal of this control strain to check if the tested strains showed leaky expression of SspB. Basal expression of SspB would have led to premature degradation of sfGFP-*ssrA* and an overall lower fluorescence signal compared to the control strain.

3.3.1 The inducible SspB regulated GFP degradation, yet its expression was leaky

In this experiment, we analysed whether the degradation of sfGFP-*ssrA* could be controlled by the induction of SspB expression and whether the RBSs of different strength had an impact. We used the strains BWEC61, BWEC62, BWEC63 and BWEC60 and transformed them with the plasmid pUC19-GFP*ssrA*-RFP. As described in Section 2.8 we inoculated main cultures of all four strains to reach a start D_{600} of 0.1 and incubated them at 37 °C for 7 hours. Figure 10 illustrates the GFP fluorescence profiles of the tested strains over the time course of the experiment. Panel A shows the time course of the samples with induction of SspB expression, while in panel B SspB was not induced. All strains contained the same pUC19-GFP*ssrA*-RFP plasmid, hence we expected all strains to have the same

fluorescence levels at the beginning of the incubation. As depicted in Figure 10, only the control strain BWEC60, which lacked the *sspB* gene, showed a fluorescence signal. This strongly indicates that sfGFP-*ssrA* was already degraded over night, due to leaky SspB expression. After three hours of incubation we measured an increase of fluorescence in BWEC61, but no fluorescence signal in BWEC62 and BWEC63. It appears that the basal expression of SspB driven by an RBS of medium strength was sufficient to completely degrade sfGFP-*ssrA*, which in turn indicates that the SspB-mediated ClpXP degradation system worked highly efficiently.

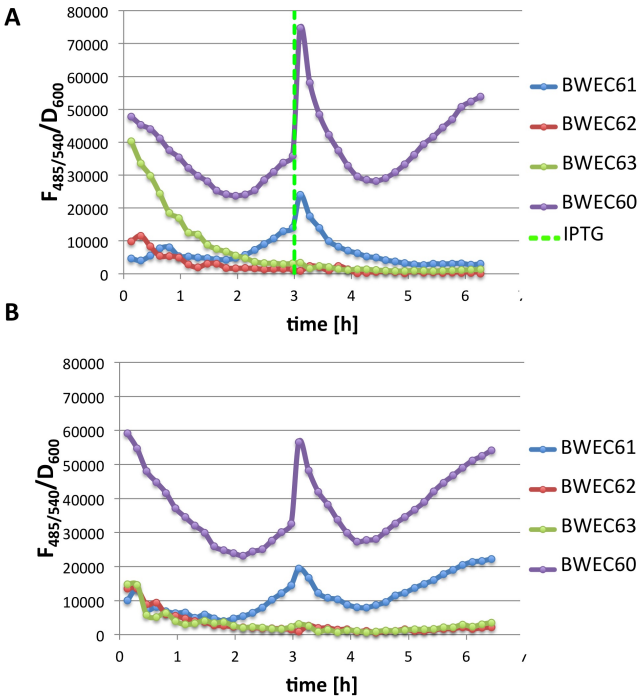


Figure 10: GFP degradation was regulated by the IPTG inducible adapter protein SspB. Three *E. coli* strains containing an IPTG-inducible SspB under a weak (WsspB), medium (MsspB) or strong (SsspB) ribosome binding site and a *sspB* knock-out strain (Δ*sspB*) carried the plasmid pUC19-GFP*ssrA*-RFP. This plasmid constitutively expressed *ssrA*-tagged GFP. Fluorescence was excited at 485 nm and recorded at 540 nm (F485/540). We measured the fluorescence profile over 6.5 hours. IPTG induction is indicated by the dashed, green line. **A:** Time course of GFP fluorescence with induction of SspB expression. **B:** Time course of GFP fluorescence without induction of SspB expression. There was no fluorescence detectable in the samples with a medium and strong RBS. Solely ΔsspB and and W*sspB* showed fluorescence development over time. Subsequent to SspB induction the GFP levels decreased in W*sspB*, while fluorescence remained in Δ*sspB* and not-induced W*sspB*. The cultures were diluted after $\sim$ 3 h of growth to keep the cell density in the linear range of the plate reader, which most probably caused the decline in fluorescence.

Consequently, we evaluated the degradation of sfGFP-*ssrA* in the strains BWEC61 (Figure 10 blue line) and BWEC60 (Figure 10 purple line). Before induction, the GFP fluorescence levels of both strains rose (Figure 10A). After the induction of SspB expression, the fluorescence signal of BWEC60 increased, while the fluorescence signal of BWEC61 rapidly declined and was undetectable two hours after the induction. The drastic decline in fluorescence at the point of induction was caused by a dilution of the cultures. The dilution was necessary to keep the cell density in the linear range of the plate reader.

As a negative control, we did not induce SspB in both strains (Figure 10B). As expected, the fluorescence signal in the *sspB* deficient strain BWEC60 increased. The fluorescence signal of BWEC61 increased as well without induction. This result strongly suggests that sfGFP-*ssrA* degradation was regulated by SspB expression.
We detected RFP fluorescence throughout the entire experiment, which confirmed that sfGFP-*ssrA* degradation was caused by the action of the ClpXP protease rather than by premature degradation of the bicistronic mRNA.

3.3.2 The alternative genomic *sspB* constructs tightly regulated SspB expression and GFP degradation

As described above, sfGFP-*ssrA* degradation was controllable by SspB expression. However, the initial constructs showed leaky expression of SspB, which resulted in premature degradation of sfGFP-*ssrA*. To overcome this limitation we designed two alternative genomic *sspB* constructs $P_{L lacO-1}$*sspB-lacI* and P_{lacUV5}*sspB-lacI* (Figure 7). In these constructs the expression of *sspB* was controlled by the $P_{L lacO-1}$ or P_{lacUV5} promoters and the expression cassette was positioned upstream of *lacI*. We integrated the constructs in BWEC70 and BWEC60 and generated the strains BWEC71, BWEC72 and BWEC65. As described in Section 3.2, we were not able to cure BWEC64 from pCas and the target plasmid, therefore we did not analyse this strain. We used the strain BWEC60 as a negative control, since it completely lacked *sspB*.
Figure 11 shows the GFP fluorescence profiles of the strains over the time course of eight hours. We used the same experimental conditions as described in Section 3.3.1. In contrast to the experiment shown in Figure 10, all strains showed the same fluoresence level at the beginning of the incubation, which confirmed that these constructs did not basally express SspB and thus did not degrade sfGFP-*ssrA* before induction.
The strains containing an IPTG inducible *sspB* were expected to show a decline in fluorescence signals after induction, whereas the *sspB* deficiant strain BWEC60 was expected to steadily fluoresce throughout the experiment. As visible in Figure 11A, the fluorescence signal of the strains BWEC71 and BWEC65 rapidly decreased after induction. By comparison of the induced (panel A) and non-induced samples (panel B) it became clear that sfGFP-ssrA degradation was caused by SspB expression.
Surprisingly, the fluorescence signal of the strain BWEC72 did not decline after IPTG induction. It is notable that BWEC65 and BWEC71 carried the same genomic *sspB* construct, the only difference was that the latter strain carried an *ssrA*-tagged RF-1. Thus we had expected the same fluorescence behavior of both strains. Because of this unexpected behavior of BWEC72 we analysed the strain in more detail, more precisely its *melAB* locus (Section 3.5).

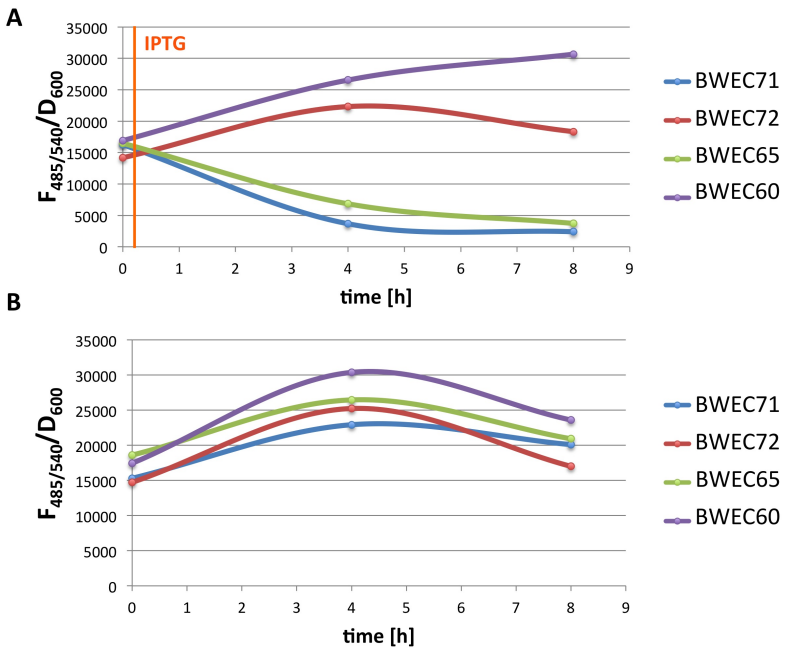


Figure 11: SspB expression tightly regulated GFP degradation, after the redesign of the expression cassette (see Figure 7). Fluorescence of the three *E. coli* strains BWEC71, BWEC72, BWEC65 and the *sspB* knock-out strain (BWEC60) was excited at 485 nm and recorded at 540 nm (F485/540). IPTG induction is indicated by the orange line. **A:** Fluorescence profile of GFP with induction of SspB expression. **B:** Fluorescence profile of GFP without induction of SspB expression. Subsequent to SspB induction GFP fluorescence levels rapidly decreased in BWEC71 and BWEC65, while the fluorescence remained unaltered in BWEC60 and BWEC72.

3.4 The site-specific incorporation of BocK into the reporter protein eGFPx was enhanced by the controlled degradation of RF-1

The ultimate goal of this project was the enhanced site-specific incorporation of an ncAA into target proteins via an orthogonal aminoacyl-tRNA synthetase/ suppressor tRNA pair during the controlled degradation of RF-1. As described in Section 1.2, the suppressor tRNA incorporates its loaded ncAA in response to an amber stop codon. However, incorporation efficiencies are fairly low, because the suppressor tRNA competes with RF-1. By degradation of RF-1, incorporation efficiency of the ncAA should be enhanced. To test this hypothesis, we performed site-specific incorporation of the ncAA BocK-lysine (BocK) in the amber mutant eGFPx. We tested the incorporation efficiency in eGFP variants containing up to three in-frame amber stop mutations. We chose the positions 40, 134 and 213 for introduction of amber stop mutations. Particularly, we put the in-frame amber stop codon at position 40, because it was upstream of the fluorophore. Thus, if translation stopped at position 40, a truncated protein would be formed that would not fluoresce. The incorporation of BocK at the in-frame amber stop codons would result in the expression of full length, fluorescent protein.

In this experiment eGFPx was provided on the expression plasmid pSCScum-1am (Supplementary Figure 12). eGFPx was controlled by an IPTG inducible P_{T5} promoter and

the λt_0 terminator. The expression plasmid also carried the orthogonal aminoacyl-tRNA-synthetase and suppressor tRNA pair. The aminoacyl-tRNA-synthetase, more precisely, pyrrolysyl-tRNA synthetase, derived from *Methanosarcina mazei* (*Mm*PylRS) was controlled by the P_{T5} – cym promoter. Regulation of this promoter was mediated by the binding of the cumate repressor protein to an operator site. The operator was placed downstream of the P_{T5} promoter [49]. Addition of cumate caused the cym-repressor to dissociate from the operator, which enabled transcription of *Mm*PylRS. The suppressor tRNA derived from the archeon *Methanomethylophilus alvus* (*Mma*tRNA) and was regulated by the constitutive P_{ProK} promoter.
As described in Section 2.9.1, we introduced the plasmids pSCScum-1am and pSCScum-empty to the respective strains. pSCScum-empty did not carry the amber mutant eGFPx. Hence strains, which carried pSCScum-empty were used as a negative control to determine the autofluorescence of the cells. The measured fluorescence signal was later subtracted from the fluorescence signal of the cells expressing eGFPx. We inoculated main cultures of the strains to reach a start D_{600} of 0.1 and incubated the cells for 24 hours. After the cultures reached a D_{600} of ~ 0.8 we induced expression of SspB, eGFPx and *Mm*PylRS. Addition of IPTG induced SspB and eGFPx expression. *Mm*PylRS was induced by addition of cumate. For incorporation of the ncAA, we added BocK to the cultures at the point of induction. After 24 hours of incubation and eGFPx expression we monitored the GFP fluorescence readout of each strain.
It is reported that the absence of RF-1 improves incorporation efficiency of ncAAs into target proteins [12]. Hence, we expected higher fluorescence signals in the strains that were able to degrade RF-1 than in the control strains, which lacked a degradable RF-1.

3.4.1 RF-1 degradation enhanced incorporation of BocK into eGFPx with one amber stop codon

In this approach, we performed site-specific incorporation of BocK into eGFPx at one amber stop mutation at position 40. We tested the strains BWEC71, BWEC72, BWEC65, BWEC73 and BL21, which all carried the expression plasmid pSCScum-1am. Figure 12 illustrates the GFP fluorescence signals after 24 hours incubation and eGFPx[BocK] expression. The samples, which contained all compounds necessary for protein expression and BocK incorporation are depicted in blue bars. Figure 12 reveals that BWEC72 showed the highest GFP fluorescence signal. Therefore we used it for all subsequent protein expression experiments. The results also revealed that incorporation of BocK was more efficient in the strain, where *sspB* was controlled by P_{lacUV5} than the strain with $P_{LlacO-1}$ controlled *sspB*. It is also notable that BWEC72, which carried a *ssrA*-tagged RF-1, showed a fluorescence signal that was almost 3-fold higher than BWEC65, which carried the same genomic *sspB* construct, but lacked a *ssrA*-tagged RF-1.
Additionally to the fluorescence measurements, we performed SDS-PAGE and a western blot analysis, to confirm that indeed eGFPx was expressed and detected. eGFPx carried an N-terminal hexahistidine tag and could thus be immunodected by an anti-hexahistidine

antibody. Both the SDS-gel and western blot revealed very intense bands at the expected size, which strongly indicated that eGFPx was expressed in high yields. These results were consistent with the detected strong fluorescence signals of the analysed strains.
As a control, we used the strain BWEC73, which contained a *ssrA*-tag on RF-1, but could not be degraded by ClpXP due to the missing *sspB*. As a second control, we used the wild-type BL21, which carried the endogenous sspB allele, but lacked a *ssrA*-tagged RF-1. Both controls showed a lower fluorescence signal than BWEC72, indicating that BocK incorporation into eGFPx was more efficient in the strain with a degradable RF-1 than in the strains with a non-degradable RF-1. The inducible degradation of RF-1 appeared to enhance site-directed incorporation of BocK, which was most probably because of to the reduced competition of RF-1 with the suppressor tRNA.
We included three controls, where each strain lacked one of the three compounds necessary for protein expression. In Figure 12, the samples lacking BocK are indicated by red bars. The fluorescence output in the absence of BocK was almost not detectable, which indicated that the orthogonal pair specifically incorporated the non-canonical amino acid rather than a canonical amino acid. All samples that lacked cumate (purple bar), showed considerable levels of GFP fluoresence. This indicates that the cumate repressor did not tightly repress the transcription of *Mm*PylRS, hence basal levels of *Mm*PylRS were present even without cumate induction. Apart from BWEC72, all strains carrying an additional *lacI* copy, showed no fluorescence signal in the absence of IPTG (green bars). BL21, which lacked an additional *lacI*, showed a fluorescence signal without IPTG induction. This suggests that the IPTG inducible promoters were tightly regulated by LacI. It was peculiar that BWEC72 showed GFP fluorescence without IPTG induction, even though it contained an additional *lacI* copy. As described in Section 3.3.2, the strain also showed unexpected results in the sfGFP-*ssrA* degradation experiment. These two independent unexpected results suggested that the genomic *sspB* construct was most probably altered or mutated. Hence, we analysed the *melAB* locus of this strain in more detail, by sequencing (Section 3.5).

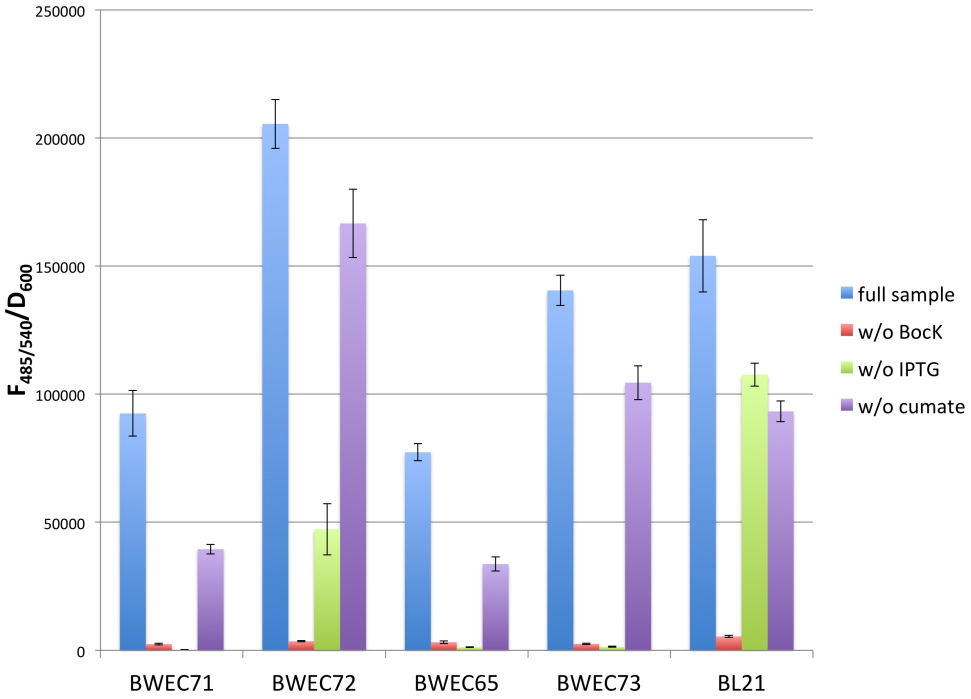


Figure 12: Incorporation of BocK at amino acid position 40 into eGFPx was enhanced in BWEC72. The blue bars indicate the "full samples", which contained BocK, IPTG and cumate and were able to express eGFPx[BocK]. The red, green and purple bars show controls in which the strains lacked one of the compounds necessary for eGFPx expression. IPTG was added to induce the expression of SspB and eGFPx, cumate induced the expression of *Mm*PylRS. We excited GFP fluorescence at 485 nm and recorded it at 540 nm. The strain BWEC72, which contained a degradable RF-1 showed the highest fluorescene signal. The fluorescence signal was almost 3-fold higher than in the same strain containing a non-degradable RF-1 (BWEC65).

3.4.2 BocK could not be incorporated in eGFPx with two or three amber stop codons

As stated previously, the generated strain BWEC72 showed enhanced incorporation efficiency of BocK into the reporter protein eGFPx with one in-frame amber stop codon mutation. It is reported that incorporation efficiency of multiple ncAAs into target proteins is enhanced when RF-1 is absent [12]. Thus, we tested incorporation efficiencies of BocK into eGFPx with two or three amber stop codon mutations. We placed the amber stop codons at the positions 40, 134 and 213 of eGFPx. The genes for eGFPx40, eGFPx40,134 and eGFPx40,134,213 were provided on the expression plasmids pSCScum-1am, pSCScum-2am and pSCScum-3am. Once again, we used the strains BWEC71, BWEC72, BWEC65, BWEC73 and BL21 which carried the respective expression plasmids, to test BocK incorporation efficiency.

Figure 13 shows the GFP fluorescence signal of each strain, upon integration of BocK at either one, two or three amber stop codons. The results revealed that incorporation of BocK into eGFPx at multiple amber stop codons was not successful. In all tested strains, the fluorescence signal of eGFPx40,134 and eGFPx40,134,213 was considerably lower compared to the fluorescence signal of eGFPx40.

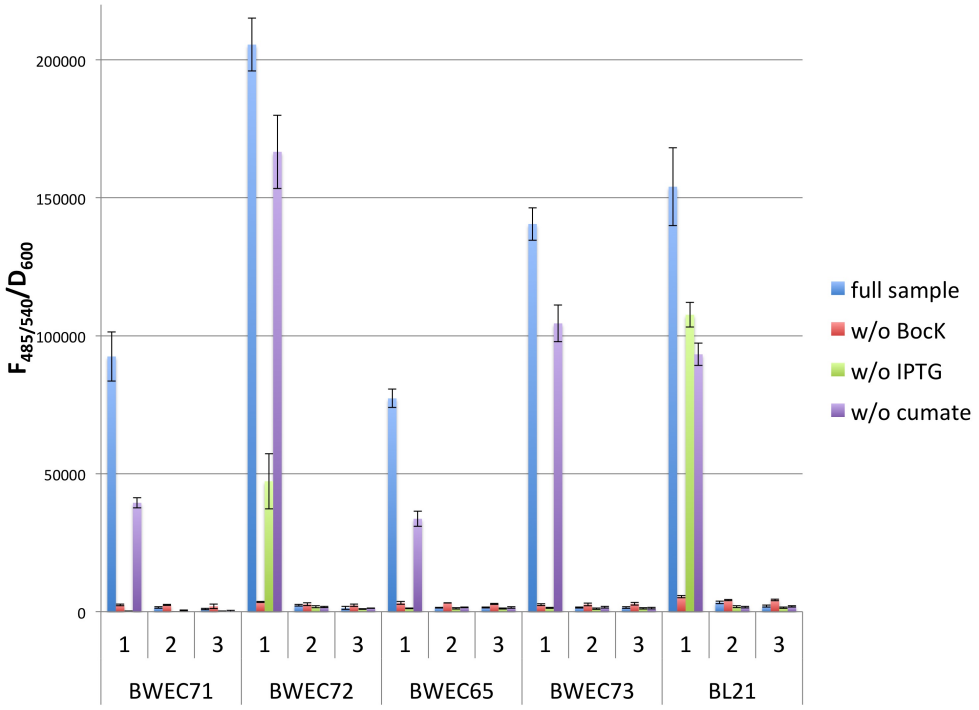


Figure 13: BocK incorporation into eGFPx at two or three amber stop codons was not successful. The blue bars show the samples that contained all compounds necessary for expression of eGFPx40,134 and eGFPx40,134,213. We also made three controls (red, green and purple bar), where the strains each lacked one of the compounds necessary for eGFPx expression. IPTG was added to induce the expression of SspB and eGFPx. Cumate induced the expression of the orthogonal *Mm*PylRS. Fluorescence was excited at 485 nm and recorded at 540n nm. As previously shown, BWEC72 showed the highest fluorescence level, upon incorporation of BocK at one amber stop codon. Incorporation of BocK at two or three stop codons was not successful in any of the five strains.

3.5 Sequencing of BWEC72 revealed a mixed genotype

We integrated the genomic *sspB* construct into the *melAB* locus of BWEC70 by the CRISPR/Cas9 system. For this we used the two-vector system, consisting of pCas and the target plasmid (Section 1.4). To use the generated strain for further experiments, we cured it from both plasmids (Section 3.2). Curing of the target plasmid involved IPTG addition, which also induced SspB expression. As described in Figure 8, over expression of SspB had detrimental effects on the cell growth. We suspected that IPTG addition for target plasmid curing caused stress and hence changes in BWEC72. The unexpected results of the strain from the sfGFP-*ssrA* degradation (Section 11) and BocK incorporation experiment (Section 12) supported this hypothesis. Consequently, we sequenced the *melAB* locus of BWEC72 at three different stages of plasmid curing. We sequenced the strain before induction, when it carried both plasmids, after target plasmid curing, when it only carried pCas and after pCas curing, when the strain was plasmid-free.

We chose sequencing primers that bound outside of the homology hooks, to sequence the entire *melAB* locus and primers adjacent to the *sspB* and *lacI* gene. Sequencing revealed that the strain, which carried both plasmids contained the correctly integrated genomic *sspB* construct. However, we also detected the contaminating sequence of the non-modified locus. This suggested that BWEC72 was present as mixed strain, exhibiting both the wild-

type and the genomically edited genotype. We did not obtain any sequencing results of the strain carrying pCas or the plasmid free strain. Since the primers bound outside the homologies, it was rather surprising that no sequencing result could be obtained. This strain definitely needed further analysis to identify its actual genotype. However, since this strain exhibited the best ncAA incorporation efficiency of all tested strains, we continued to use it for subsequent experiments.

3.6 Wild-type Stx1B and Stx1B-K8[AzK] were obtained in high purity

We demonstrated enhanced incorporation efficiencies of one ncAA into the reporter protein eGFPx. In the next step we tested incorporation of the reactive ncAA into an industrially relevant protein. We selected the *Shigella dysenteriae* descendent lectin shiga toxin. The toxin consists of two subunits A and B. Subunit A is cytotoxic and disrupts protein biosynthesis in target cells, while subunit B (Stx1B) is responsible for the binding onto the target cell surface. Stx1B is a particularly appealing target protein, because it has specific affinity for the glycosphingolipid globotriasylceramide (Gb3). Gb3 is overexpressed in metastatic cancer cells [50], which makes it suitable for targeted cancer therapy.

We added a C-terminal hexahistidine tag to purify the protein by Ni-chelate affinity chromatography. It was crucial to add the purification tag on the C-terminus, to assure that only full length protein was purified, after successful readthrough at the in-frame amber codon. For incorporation of the ncAA we exchanged the sequence coding for lysine at position eight, with an in-frame amber stop codon. We chose the reactive ncAA azido-lysine (AzK), because it is known to be accepted by the *Mm*PylRS/ *Mma*tRNA-orthogonal pair [51]. Its azide group can be used for bioconjugation with other molecules by copper(I)-catalysed azide-alkyne cycloaddition (CuAAC) [52] or strain-promoted azide-alkyne conjugation (SpAAC). We used CuAAC to confirm the incorporation of AzK into Stx1B_K8am. We expressed wild-type protein (Stx1B) as well as the variant protein (Stx1B-K8am). Therefore, we constructed the two expression plasmids pSCSara-stx1B and pSCSara-stx1B-K8am, which were introduced into the strain BWEC72.

Stx1B expression was carried out in 500 mL cultures. The strains containing either of the expression plasmids were incubated in the presence of AzK for 24 hours. As the cultures reached a D_{600} of 0.8 we induced SspB, Stx1B and *Mm*PylRS expression by IPTG and arabinose, respectively. For the site-specific incorporation of the ncAA, we added AzK to the cultures and incubated the cells over night. The cells were harvested, disintegrated and wild-type Stx1B and Stx1B-K8[AzK] were purified by Ni-chelate affinity chromatograph as described in the methods section.

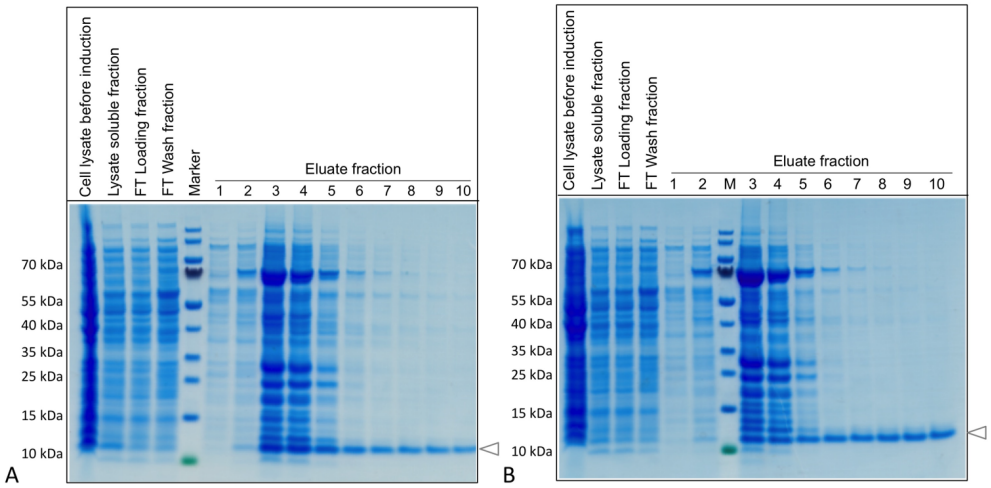


Figure 14: Wild-type Stx1B and Stx1B-K8[AzK] were obtained in high purity. Coomassie-brilliant-blue stained 4-12 % SDS-gel of Stx1B expression and purification by Ni-chelate affinity chromatography (calculated molecular weight of 8.7 kDa). **A:** SDS-gel after purification of wild-type Stx1B. **B:** SDS-gel after purification of Stx1B-K8[AzK]. Migration bands of Stx1B are indicated with white triangles. PageRuler Prestained Protein ladder was used as a molecular weight marker.

Figure 14 shows the SDS-gels after Stx1B purification. Panel A and B show the SDS gels of wild-type Stx1B and Stx1B-K8[AzK], respectively. By comparison of both gels, an apparent observation was that expression levels of the wild-type and variant protein were similar. This was remarkable, because expression levels of variant proteins are usually considerably lower than expression levels of the wild-type proteins [53]. Figure 14 also showed that both proteins could be obtained in high purity, which is promising for the upscaling of the expression.

3.7 Copper(I)-catalysed azide-alkyne cycloaddition (CuAAC) in stx1B-K8[AzK]

To ascertain AzK incorporation into Stx1B_K8am, we performed a bio-orthogonal conjugation, which is collectively termed "click chemistry". Click chemistry intends to development powerful, selective and modular "blocks" by joining small units together with heteroatom links. Common characteristics of a click reaction is that it is modular, wide in scope, requires simple reaction conditions, simple product isolation and inoffensive by-products. In this way it provides a foundation for the fast assembly of new molecular entities [54]. More specifically, we used copper(I)-catalysed-azide-alkyne cycloaddition (CuAAC) to conjugate Stx1B-K8[AzK] with the fluorescent dye sulfo-cyanine-3-alkyne (AlkD) [52],[55]. Figure 15 illustrates the principle of CuAAC. Basically, the azide-functionalised Stx1B and the alkyne-functionalised fluorophore were crosslinked via a triazole moiety in the presence of copper(I). The reaction is very specific, because azide and alkyne groups are rather unreactive under physiological conditions [56]. The reaction can only occur in the presence of copper(I). We provided copper(I) by addition of $CuSO_4$ and added ascorbate as reducing agent. Copper can be harmful for proteins, therefore we also added Tris((1-hydroxy-propyl- 1H-1,2,3-triazol-4-yl)methyl)amine (THPTA). THPTA on

the one hand chelates the copper ion and thus protects the protein from oxidative damage. On the other hand it accelerates the "click reaction" by maintaining the copper(I) oxidative state of $CuSO_4$ [57].

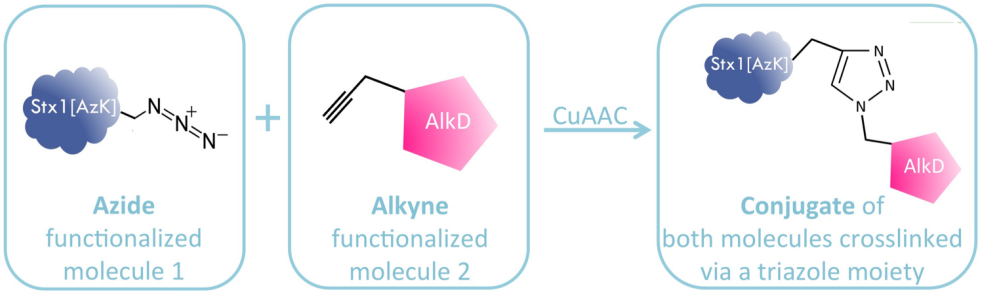


Figure 15: Principle of the Copper(I)-catalysed azide-alkyne cycloaddition (CuAAC). The azide-functionalised Stx1B-K8[AzK] was conjugated with an alkyne-functionalised fluorophore AlkD via a triazole moeity in the presence of copper(I).

As visible in Figure 14, the elution fractions were rather impure at the beginning of elution and became highly pure towards the end of elution. For click chemistry we pooled the elution fractions 3-5 and 6-10 and labelled them impure and pure, respectively. As negative control, we used wild-type Stx1B, since it would not be able to form a conjugate with the fluorescent dye, due to the missing azide group.

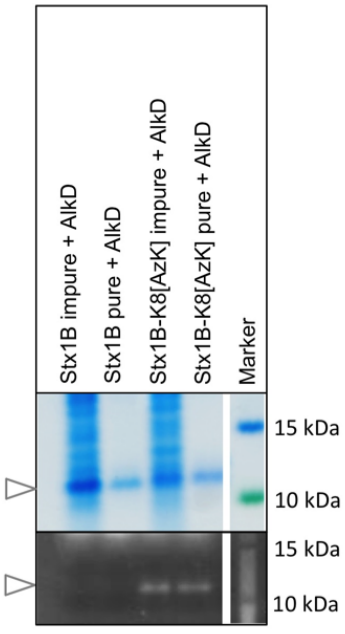


Figure 16: Stx1B-K8[AzK] was conjugated with the fluorescent dye AlkD. SDS-PAGE analysis of the click reaction. All Stx1B and Stx1B-K8[AzK] samples indicated bands at the expected size of $\tilde{8}.7$ kDa. The terms "impure" and "pure" refer to the pooled elution fractions 3-5 and 6-10, respectively. Under UV exposure, the Stx1B-K8[AzK]-fluorophore conjugate showed a fluorescent band at the expected size. Expectedly, there was no fluorescence signal of wild-type Stx1B.

Successful conjugation resulted in a fluorescence signal and a visible Stx1B-K8[AzK] band under UV exposure. Figure 16 depicts the SDS-PAGE analysis of the CuAAC and the fluorescence signal of the Stx1B-K8[AzK]-fluorophore conjugate under UV-exposure. The SDS gel revealed that all four samples indicated bands at the expected size of $\sim$ 8.7 kDa. Only the Stx1B-K8[AzK]-fluorophore conjugate showed a visible fluorescent band under UV-exposure. As expected, the wild-type Stx1B did not give a fluorescence signal. This result confirmed the successful bioconjugation of Stx1B-K8[AzK] and AlkD and also that the azide group was sterically accessible. This result is very promising for potential bioconjugations with other biomolecules.

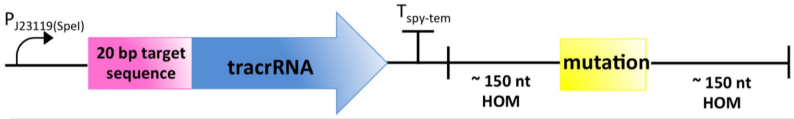
Subsequently, we aimed to analyse the activity of wild-type Stx1B and Stx1B-K8[AzK], by differential scanning fluorimetry (DSF) [58]. This analysis method relies on the fact that proteins denature at a certain temperature. When they are bound to a ligand, they denature at a higher temperature. We hypothesised that Stx1B would show an increased denaturation temperature, when it actively bound its ligand Gb3. We repeatedly performed DSF with wild-type Stx1B and Stx1B-K8[AzK] in the presence as well as the absence of Gb3. However, we did not obtain valid results and it remains obscure whether we had produced active StxB1. As well, our results did not reveal whether the incorporation of AzK into StxB at position K8 had an impact on ligand binding.

3.8 Amber stop codon exchange in seven essential genes should prevent expression of aberrant proteins

As previously mentioned, RF-1 degradation improved incorporation efficiency of ncAAs into proteins, but it consequently suppressed every amber stop codon. In the *E. coli* genome, 300 open-reading frames end with an amber stop codon [12]. Efficient suppression of these amber stop codons might prevent the accurate termination of the translation of the corresponding proteins, ultimately leading to aberrant proteins. Out of those 300 genes, seven are essential, *murF, lolA, lpxK, hemA, hda, mreC* and *coaD*. Improper translation of those proteins could cause fatal events in *E. coli*. To evade this apparent disadvantage, we performed stop codon switches from amber to ochre or opal stop codons in these seven essential genes. The stop codons were exchanged by introducing point mutations by the CRISPR/Cas9 system in combination with λ-Red recombineering.
Editing of every single stop codon would have been rather labour intensive and time consuming, thus we aimed to perform simultaneous genome editing of two to three stop codons. For this we designed and constructed three tandem target plasmids. Each target plasmid carried the pTargetF vector backbone (Supplementary Figure 3), with the pMB1 origin of replication and a spectinomycin resistance marker and two or three editing cassettes. Each editing cassette consisted of a target specific sg-RNA and an $\sim$ 300 bp mutation fragment. Figure 17 shows the schematic set-up of the editing cassettes and the introduced point mutations. Since we used CRISPR/Cas9 for genome editing, it was essential to mutate the

PAM region from NGG to NAG or mutate the 20 bp target specific sequence on the editing cassette. This prevented the endonuclease Cas9 from continuously introducing ds-breaks at the mutated locus. We particularly took care to introduce silent mutations when altering the PAM region or target sequence, to keep the locus as unaffected as possible. Table 13 lists the introduced point mutations that were introduced at the stop codon of each essential gene and a potential impact on the amino acid sequence of the encoded protein. The amber stop codons of *lolA, lpxK, hemA, hda* and *coaD* were switched to an opal stop codon. The stop codons of *murF* and *mreC* overlapped with the open-reading frames of *marY* and *mreC*, respectively. As depicted in Figure 17, we switched both amber codons to ochre codons and introduced additional mutations, as described by Isaacs et. al. [59].

For the construction of the three tandem target plasmids, we had the seven editing cassettes synthesised. pTargetF was linearised using the restriction enzymes BamHI and HindIII and the editing cassettes were introduced by Gibson assembly. We generated the three tandem target plasmids pT-murF-lolA-lpxK, pT-hemA-hda and pT-mreC-coaD and performed genome editing by transforming pCas-containing BWEC72 recombineering cells with each of the target plasmids. In our first approach we tested the simultaneous stop codon exchange of the stop codon at three loci, by introducing pT-murF-lolA-lpxK into the electro-competent cells. We analysed wheter the stop codon exchange at the respective loci had been successful by multi allel-specific colony PCR (MASC PCR), as first described by Isaacs et. al. [59].



gene	wild-type	after mutation
mreC	***Gln **GATAAC** **CGAGCGGTG** SerAlaMet	***Gln **AGTGAC** **TCTTCGGTA** SerAlaMet
murF	Cys*** **TGTTAG** **ATGTTAGTT** MetLeuVal	Cys*** **TGCTGA** **ATGTTGATC** MetLeuIle
lolA	Lys*** **AAGTAG**	Lys*** **AAGTAA**
lpxK	Asn*** **AACTAG**	Asn*** **AACTAA**
hda	Leu*** **TTGTAG**	Leu*** **TTGTAA**
coaD	Ala*** **GCGTAG**	Ala*** **GCGTAA**
hemA	Glu*** **GAGTAG**	Glu*** **GAGTAA**

Figure 17: Mutation fragments for the construction of tandem target plasmids. The illustration at the top depicts the general design of the editing cassette, which consisted of the target specific sg-RNA and a $\sim$ 300 bp DNA fragment carrying the mutation flanked by 150 bp overlapping homologies. The table illustrates the introduced mutations.

Table 13: List of the mutations introduced at the stop codons of the seven essential genes. The seven essential genes are highlighted in bold letters. The amber stop codons of *murF* and *mreC* were changed to opal, the amber stop codons of *lolA, lpxK, hemA, hda* and *coaD* were changed to ochre.

gene	DNA[a)]	amino acid[b)]	note
murF	T1356C, A1358G, G1359A		
marY	T4C, A6G, G7A, T9C	V3I	overlaps with *murF*
mreC	T1356C, A1358G, G1359A		
mreD	G1A, G6T, A7T, G8C, C9T		overlaps with *mreC*
lpxK	G987A		
lolA	G615A		
hemA	G1257A		
hda	G747A		
coaD	G480A		

a) DNA sequence is numbered from start (A of ATG = +1) to last base of stop codon
b) Numbering starts with first translated amino acid

3.8.1 MASC PCR enabled analysis of successful stop codon exchange

When we performed stop codon exchange from TAG (amber) to either TAA (ochre) or TGA (opal) in seven essential genes of *E. coli*, we used tandem target plasmids to introduce the point mutations. This meant that theoretically in one genome editing step, up to three stop codon switches could occur. To avoid excessive sequencing of a variety of clones, we performed (MASC PCR). This method allowed the screening of positive clones carrying the desired point mutations without sequencing.

For MASC PCR we designed two forward primers for each locus that contained either TAG or TAA/TGA at the 3'-end. We designed one reverse primer for each locus. It was essential that the reverse primers bound outside of the editing cassettes to avoid binding of the target plasmids. We designed the primer pairs in a way that each PCR product had a distinct fragment length. PCR products with discriminable lengths enabled us to perform multiplex MASC PCR, which meant that up to three PCR fragments were amplified, using up to three primer pairs in a single PCR reaction. By using the tandem target plasmid pT-murF-lolA-lpxK, we could analyse three loci simultaneously by MASC PCR. Hence, we designed primers that resulted in three PCR products of distinct lengths.

Theoretically, amplification of the DNA fragment was only possible, when the 3'-end of the forward primer perfectly annealed to the template. Depending on the stop codon of the amplified gene, only one of the two primers could perfectly anneal to the template under stringent PCR conditions, thus MASC PCR should result in only one PCR product. In reality, however, when we applied a standard PCR temperature profile, the DNA fragment could also be amplified when the 3'-end of the primer did not perfectly anneal to the

template. This meant that the reaction conditions needed to be aggravated to an extend that only perfectly annealing primers could elongate the PCR fragment. Figure 18 shows an agarose gel after MASC PCR of four clones BWEC72{pCas, pT-murF-lolA-lpxK}. As mentioned previously, we performed two PCR reactions per clone, by using the forward primer that either bound the wild-type amber stop codon or the mutated ochre or opal stop codon. The gel image revealed that multiplex MASC PCR was successful, since all three bands were visible and migrated at the expected sizes. Figure 18 shows that the samples in lane "a" of each clone corresponded to the wild-type locus, while the samples in lane "b" corresponded to the mutated locus It was visible that only MASC PCR of the wild-type loci was successful, since only the left lanes of each clone showed visible bands. In this first attempt we mainly focussed on finding the most suitable temperature profile for multiplex MASC PCR, rather than analysing successful genome editing. After an appropriate protocol for MASC PCR has now been established, we will screen a higher number of clones to analyse the stop codon exchange efficiency of the system.

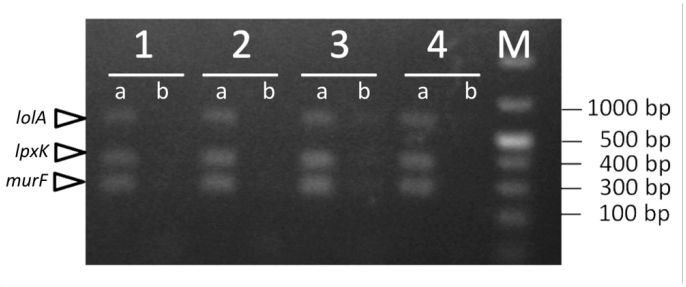


Figure 18: Agarose gel after multiplex MASC PCR of BWEC72{pCas, pT-murF-lolA-lpxK}. We analysed whether the exchange from amber to ochre or opal stop codons had been successful in four clones. We performed two PCR reactions per clone, using either a forward primer containing an amber stop codon at the 3'-end (lanes "a") or an opal/ochre stop codon at the 3'-end (lanes "b"). GeneRuler 1 kb plus DNA ladder was used as a molecular weight marker.

4 Discussion

4.1 Regulation of the degradation of GFP-*ssrA* by the controlled expression of SspB

We integrated three different genomic *sspB* constructs at the *melAB* locus of *E. coli* BL21. Each construct contained an *sspB* gene, which was controlled by an RBS of different strength. We aimed to regulate the degradation of *ssrA*-tagged RF-1 by controlling SspB expression. Since we could not directly detect RF-1 degradation, we designed a proof-of-concept experiment. The reporter gene sfGFP was *ssrA*-tagged and could hence be degraded by the SspB-dependent ClpXP machinery. In this way, we were able to observe the degradation of sfGFP over time by monitoring the decay of its fluorescence. As shown in Figure 10, it was fairly obvious that SspB was basally expressed in all tested strains, which led to premature degradation of sfGFP-*ssrA*. The basal levels of SspB were sufficient to prevent the cells from exhibiting any fluorescence signal. Solely the strain in which *sspB* was controlled by a weak RBS, was able to recover a fluorescence signal. This result strongly indicated that the SspB-dependent protein degradation machinery worked highly efficiently and that the use of a weak RBS was sufficient for protein degradation. In fact, the expression of SspB under control of the weak RBS, led to a rapid decline of the fluorescence signal and no signal remained after one hour of induction of SspB expression. This result suggested that an even weaker RBS could potentially be used to efficiently degrade *ssrA*-tagged proteins.

After we had confirmed that inducible SspB expression regulated the degradation of sfGFP-*ssrA*, however we still had to overcome the leaky expression of SspB. Tight regulation of *sspB* was crucial to assure that RF-1 was only degraded upon induction of SspB. We ascribed the leaky expression to several factors. First of all the *tac*-promoter, which regulated *sspB* transcription, is a strong promoter, yet it is considerably leaky [60]. We evaded this limitation by switching to the $P_{L}lacO\text{–}1$ and the P_{lacUV5} promoters. $P_{L}lacO\text{–}1$ contains an additional *lacO* operator, which enabled binding of two LacI repressor molecules, which we expected to lower leaky expression. P_{lacUV5} is reportedly a weaker promoter than P_{tac}. By lowering the overall mRNA level of *sspB*, we reasoned to reduce the level of basal SspB expressed. Secondly, in this study the *lacI* gene was regulated by the P_{lacI^q} promoter, which efficiently expresses LacI. The *lacI* gene was positioned upstream of *sspB*, which potentially led to transcriptional readthrough into the downstream *sspB* gene. Even though the *lacI* expression construct contained the *rrnB* T1 terminator, transcription termination is not always tight. Positioning the constitutively expressed *lacI* gene downstream of *sspB* we prevented the accidental read through into the *sspB* gene. Furthermore we added an additional terminator T_{T7} after the *rrnB* T1 terminator of *lacI*. As described by Mairhofer et. al., the coupling of the two terminators increases the termination efficiency to 99 % [61]. We integrated the genomic *sspB* construct at the *melAB* locus, which genes are also transcriptionally active. We deleted the *melA* gene by integration of the *sspB* construct, yet the *melA* promoter was still intact. To avoid polar effects at the locus, we flanked

the genomic *sspB* construct with the already mentioned *rrnB* T1 and T7 terminators and the *rrnB* T1 and T2 double terminator. As shown in Figure 11 the strains exhibited no leaky expression, which strongly indicated that the implemented measures enabled tight regulation of the genomic *sspB* construct.

We were able to tightly regulate SspB expression, however, the generated strains showed unexpected behaviour. We expected that upon IPTG induction, SspB would be expressed in the three strains BWEC71, BWEC72 and BWEC65 and degrade sfGFP-*ssrA*. As visible in Figure 11. The strains BWEC71 and BWEC65 showed the expected result, but BWEC72 was not able to degrade sfGFP-*ssrA*. We could eliminate the possibility that the genomic *sspB* construct was not designed properly, because both BWEC72 and BWEC65 carried the same genomic *sspB* construct. As described in Section 3.5, sequencing results of the *melAB* locus of BWEC72 showed that it carried a contaminating sequence of the non-edited *melAB* locus. These results strongly indicated that BWEC72 and BWEC65 did not carry the same genomic *sspB* constructs. The altered situation at the *melAB* locus of BWEC72 could for instance have led to a diminished degradation of sfGFP-*ssrA*, which was not detected over the time course of eight hours.

Another non-negligible aspect is that SspB-mediated degradation of sfGFP-*ssrA* solely confirmed that the degradation machinery was working controllably and efficiently. The regulatable degradation of sfGFP-*ssrA* was not necessarily coherent with SspB-dependent degradation of *ssrA*-tagged RF-1. For instance, the C-terminal *ssrA-tag* could not be accessible for SspB or ClpX. We did not analyse whether SspB or ClpX were sterically hindered from binding the tag. Subsequent experiments should scrutinise the cellular levels of RF-1 before and after induction of SspB expression. A rather straight forward experiment would be to perform a western blot of the samples during several instances with an antibody specific for RF-1. Alternatively, an N-terminal hexahistidine-tag could be added to the *ssrA*-tagged *prfA* gene. In this way it could be analysed, whether RF-1 is actually degraded upon SspB expression.

Even though BWEC72 showed unexpected results, we confirmed that *ssrA*-tag mediated protein degradation was controllable by the regulated expression of SspB.

4.2 Incorporation of BocK into eGFPx was enhanced in BWEC72

After we confirmed that degradation of *ssrA*-tagged proteins was controllable by SspB expression, we analysed whether non-canonical amino acids can be incorporated more efficiently in the strains carrying a controllable genomic *sspB* construct. As described previously, incorporation efficiency of ncAAs is improved in the absence of RF-1 [12]. Hence, we expected that the generated strains, which controllably degrade RF-1, exhibit an enhanced incorporation efficiency. We used the amber mutant reporter gene eGFPx, hence we expected a higher fluorescence signal in the strains that contained a degradable RF-1 than in the control strains that were not able to degrade RF-1. This means that BWEC71 and BWEC72 should have displayed the highest fluorescence signal, compared to all other

strains. Indeed, strain BWEC72 displayed the highest fluorescence signal (see Figure 12). It was roughly 3-fold higher than the fluorescence signal of its immediate control BWEC65, which lacked a degradable RF-1. Apparently, this result was very promising, because it strongly indicated that ncAA incorporation was improved approximately 1.5 fold compared to wild-type BL21. However, sequencing of this strain revealed revealed an ambiguous sequence at the melAB locus which hinted at mixed strains. Some cells carried the inducible, genomic *sspB* construct and thus a degradable RF-1, while other cells did not. Expectedly, cells lacking the genomic *sspB* construct exhibit a lower fluorescence signal than cells lacking RF-1-*ssrA*. We normalised the fluorescence signal to cell density (D_{600}), thus it is possible that an unknown proportion of the cells showed a lower fluorescence level than expected and thus reduced the overall $F_{485/540}/D_{600}$ signal. If the strain had been pure, the fluorescence signal potentially could have been higher. It is thus mandatory to obtain a pure BWEC72 strain and unequivocally confirm the sequence at the *prfA* and *melAB* loci before the incorporation experiment can be repeated.

As mentioned above, we expected BWEC71 to exhibit an increased fluorescence signal in the ncAA incorporation experiment similar to BWEC72. Surprisingly, it showed almost the lowest $F_{485/540}/D_{600}$ signal. In the sfGFP-*ssrA* degradation experiment, this strain showed a rapid degradation of the protein, which suggested that RF-1 could have been rapidly degraded as well. However, the low fluorescence signal after incorporation of BocK into eGFPx, endorses the theory that the *ssrA*-tag of RF-1 of BWEC71 was not degradable.

The amber stop mutation was put upstream of the fluorophore of eGFPx, thus only read-through of the stop codon would lead to full length, fluorescing protein. We had to make sure that the fluorescence signal only derived from eGFPx, with incorporated BocK and not a canonical amino acid. Thus, we prepared a control, in which BocK was missing. The samples lacking BocK in Figure 12 (red bars) gave virtually no fluoresence signal, which strongly indicated that the amber stop mutation could only be read over by incorporation of the ncAA BocK.

We confirmed that incorporation efficiency at one amber stop mutation was improved in a strain, which carried a degradable RF-1. Hence, we tested if incorporation efficiency would be improved at to three amber stop mutations. As shown in Figure 13, incorporation of a ncAA at up to three amber stop mutations was not successful. This was rather unexpected, since in the absence of RF-1 there should be no competition with the suppressor tRNA. To confirm that cells with no RF-1 exhibit improved incorporation efficiencies, an additional control strain, with a deleted *prfA* gene should be tested. In this way, it could be assessed whether the presence of RF-1 was indeed responsible for the low incorporation efficiency of BocK.

Even though the strain BWEC72 revealed an ambiguous sequence at the *melAB* locus, it incorporated BocK most efficiently into eGFPx. Hence, we used it for incorporation of azido-lysine into the target protein Stx1B.

4.3 Wild-type Stx1B and Stx1B-K8[AzK] were expressed in similar yields and obtained in high purity

Ultimately, we aimed to improve the incorporation efficiency of a ncAA into an industrially relevant protein. We chose to incorporate the reactive ncAA AzK into the subunit B of the *Shigella dysenteriae* descendent shiga-toxin (Stx1B). Figure 14 shows that both the wild-type and variant protein of Stx1B were obtained in high purity. Yet, it was even more remarkable that both proteins were expressed in similar yields. Due to low incorporation efficiencies of ncAAs at amber stop mutations, the expression levels of the variant protein are usually considerably lower than the expression levels of the wild-type protein [62]. We only performed protein expression once, hence the expression of both proteins definitely needs to be repeated, to confirm that they are expressed at similar levels. Additionally, it would be worthwhile to perform the expression of wild-type Stx1B and Stx1B-K8[AzK] in the wild-type strain BL21. If Stx1B-K8[AzK] expression in the BL21 strain results in lower yields than in BWEC72, this would strongly indicate that the degradable RF-1 was responsible for improved incorporation of AzK.

Stx1B reportedly exhibits a high denaturing temperature above 65 °C [64]. This makes it suitable for purification by heat precipitation. As reported by Oh et. al, after heat treatment of the cells at 65 °C for 30 minutes, almost all *E. coli* proteins were removed. Solely Stx1B was still present in the cell free extract and obtained with a purity of ~ 90 %. In subsequent Stx1B expressions, protein purity could be improved by applying heat precipitation of contaminating heat-labile host proteins.

Even though both, the wild-type and the variant Stx1B were obtained in decent yields and purity, we had to confirm that AzK had been incorporated into Stx1B. AzK is a reactive ncAA and can be used for bio-orthogonal conjugations. We decided to crosslink Stx1B-K8[AzK] with an alkyne-functionalised fluorescent dye AlkD. After the click reaction we performed SDS-PAGE analysis and exposed the gel to UV. As depicted in Figure 16, only the variant Stx1B-K8[AzK] formed a conjugate with the fluorophore and thus showed a visible band at the expected size under UV-exposure. This result strongly indicated that AzK was incorporated into Stx1B. Nonetheless, additional mass analysis will be crucial to definitely confirm that the desired protein Stx1B was expressed and AzK incorporated at position 8.

After protein purification and confirmation of AzK incorporation, we aimed to analyse whether the two Stx1B proteins were active. Stx1B has a strong affinity to bind the glycosphingolipid Gb3, hence our first approach was to compare the binding of Gb3 by the wildtype and variant proteins using differential scan fluorimetry (DSF). Basically, the active Stx1B binds Gb3, which shifts its denaturing temperature towards a higher temperature. We repeated DSF with the wild-type and variant protein and Gb3 several times, but were not able to obtain a convincing melting curve. Stx1B reportedly exhibits a high melting temperature above 65 °C [53]. Since the maximum temperature of DSF is 100 °C, it was possible that Stx1B was still not fully denatured at 100 °C and DSF was an inappropriate method to analyse the activity of Stx1B. An alternative approach could be

an enzyme-linked immunosorbent assay (ELISA) as proposed by Gallegos et. al. In this method, the globotriasylceramide Gb3 is coated onto microtiter plates and the binding of Stx1B is detected by an anti-hexahistidine antibody [63].

4.4 Amber stop codon exchange in seven essential genes by CRISPR/Cas9

So far, we have designed and constructed the three tandem target plasmids to exchange the amber stop codons at the seven essential *E. coli* genes, *murF, lolA, lpxK, hemA, hda, mreC* and *coaD* for ochre or opal stop codons. We have also performed genome editing, by transforming BWEC72 with pT-murF-lolA-lpxK and have analysed the stop codon exchange at those three loci by MASC PCR. However, there still remain several open tasks. First ol all the PCR temperature profile for MASC PCR needs to be optimized. It is essential to find a balance between aggravating the reaction conditions to a point, where only the perfectly annealing primer can elongate the PCR fragment and preserving the efficiency of the PCR reaction. Since we perform a multiplex PCR, this might be challenging, because the reaction conditions need to be optimized for up to six primers simultaneously. As soon as the MASC PCR protocol is established, we will focus on the stop codon exchange efficiency at multiple loci. We will address the following important issues: (I) Is it possible to introduce point mutations in *E. coli* by using CRISPR/Cas9 in combination with λ-Red recombineering, or does *E. coli* possess mechanisms to repair those point mutations? (II) If it is possible to exchange stop codons by CRISPR/Cas9, can two or three loci be edited simultaneously?

We aim to adress these questions by performing genome editing at first with target plasmids that only carry one editing cassette, to confirm that the system works. Subsequently, we will increase the number of editing cassettes to three and analyse, whether stop codon exchange is still feasible. We will screen multiple clones by MASC PCR and assess the point mutations by sequencing.

Mutating seven essential genes one after another by CRISPR/Cas9 mediated genome editing is fairly labour intensive, because after each locus is edited, the target plasmid needs to be cured and the edited strain needs to be made electro-competent for the next stop codon exchange. By minimising these steps to three "genome editing rounds", plenty of time could be saved and the risk of mutations could be reduced. Hence, we aim to optimise the CRISPR/Cas9 genome editing protocol, to enable simultaneous genome editing at multiple loci in *E. coli*.

5 Conclusion

This study aimed to generate an *E. coli* strain that exhibits improved incorporation efficiency of ncAAs into target proteins. This was achieved by regulating the cellular RF-1 levels by controlled expression of the adapter protein SspB. We generated an *E. coli* BL21 strain, which carried a genomic, controllable *sspB* construct and an *ssrA*-tagged RF-1. We confirmed that the SspB-dependent ClpXP protein degradation machinery was working highly efficient and that *ssrA*-tagged protein degradation was tunable by SspB expression. Eventually, we generated the strain BWEC72, which exhibited an improved incorporation efficiency of the ncAA BocK into the amber mutant reporter protein eGFPx. So far, incorporation efficiency could be improved at one amber stop mutation. Read through of several amber stop codons could not be improved and requires further efforts. Particularly, the confirmation that SspB-controlled degradation of *ssrA*-tagged RF-1 occurs represents a key task.

We used the strain BWEC72 for incorporation of the reactive ncAA AzK into the subunit of shiga toxin (Stx1B) and obtained the protein in high purity. It was also remarkable that both wild-type and variant Stx1B were expressed in similar yields, which is promising for the up-scale of Stx1B-K8[AzK] expression in the future. Further on, we successfully conjugated Stx1B-K8[AzK] with a fluorescent dye by CuAAC and confirmed that the chemical modification is accessible for bio-conjugations with other molecules. The results of Stx1B-K8[AzK] expression and purification are very promising for targeted cancer therapy.

References

[1] Johnson JA, Lu YY, Van Deventer JA & Tirrell DA (2010) Residue-specific incorporation of non canonical amino acids into proteins: recent developments and applications *Curr Opin Chem Biol.* **14**(6), 774–780.

[2] Kirk O, Borchert TV & Fuglsang CC (2002) Industrial enzyme applications. *Curr. Opin. Biotechnol.* **13**, 345-351.

[3] Young TS & Schultz PG (2010) Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *J Biol Chem* **285**(15), 11039-11044.

[4] Nguyen DP, Lusic H, Neumann H, Kapadnis PB, Deiters A & Chin JW (2009) Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA(CUA) pair and click chemistry *J Am Chem Soc* **131**(25), 8720.

[5] Munier R & Cohen GN (1959) Incorporation of structural analogues of amino acid into the bacterial proteins during their synthesis in vivo *Biochim. Biophys. Acta* **31**(2), 378-391.

[6] Link AJ, Mock ML & Tirrell DA (2003) Non-canonical amino acids in protein engineering *Curr Opin Biotechnol.* **14**(6), 603-609.

[7] Berg, JM, Stryer L, Tymoczko JL & Held A (2012) Stryer Biochemie[7] *Springer Verlag.*

[8] Budisa N (2006) Engineering the Genetic Code: Expanding the Amino Acid Repertoire for the Design of Novel Proteins. *Wiley-VCH Verlag* Weinheim.

[9] Srinivasan G, James CM & Krzycki JA (2002) Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA *Science* **296**(5572), 1459-1462.

[10] Liu CC & Schultz PG (2010) Adding New Chemistries to the Genetic Code *Annu. Rev. Biochem.* **79**, 413-444.

[11] Nore CJ, Anthony-Cahill SJ, Griffith MC & Schultz PG (1989) A general method for site-specific incorporation of unnatural amino acids into proteins *Science* **244**(4901), 182-188.

[12] Johnson DBF, Xu J, Shen Z, Takimoto JK, Schultz MD, Schmitz R J, Xiang Z, Ecker JR, Briggs SP & Wang L (2011) RF-1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* **7**, 779-786.

[13] Heurgué-Hamard V, Champ S, Engström Å, Ehrenberg M & Buckingham RH (2002) The hemK gene in Escherichia coli encodes the N^5-glutamine methyltransferase that modifies peptide release factors *EMBO J.* **21**(4), 769-778.

[14] Johnson DBF, Wang C, Xu J, Schultz MD, Schmitz RJ, Ecker JR & Wang L (2012) Release Factor One Is Nonessential in Escherichia coli. *ACS Chem. Biol.* **7**, 1337-1344.

[15] Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, Rohland N, Schultz PG, Jacobson JM, Rinehart J, Church GM & Isaacs FJ (2013) Genomically recoded organisms expand biological functions. *Science* **342**, 357-360.

[16] Mukai T, Hoshi H, Ohtake K, Takahashi M, Yamaguchi A, Hayashi A, Yokoyama S & Sakamoto K (2015) Highly reproductive Escherichia coli cells with no specific assignment to the UAG codon. *Scientific reports* **5**.

[17] Mukai T, Hayashi A, Iraha F, Sato A, Ohtake K, Yokoyama S & Sakamoto K (2010) Codon reassignment in the Escherichia coli genetic code, *Nucleic Acids Res.* **38**, 8188-8195.

[18] King RW, Deshaies RJ, Peters JM & Kirschner MW (1996) How Proteolysis Drives the Cell Cycle. *Science* **274**(5293), 1652-1659.

[19] Grandgirad D, Studer E, Monney L, Belser T, Fellay I, Borner C & Michel MR (1998) Alphaviruses induce apoptosis in Bcl-2-overexpressing cells: evidence for a caspase-mediated, proteolytic inactivation of Bcl-2. *EMBO J* **17**, 1268-1278.

[20] Goldberg AL (2003) Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**, 895-899.

[21] Gonzalez M, Rasulova F, Maurizi MR & Woodgate R (2000) Subunit-specific degradation of the UmuD/D' heterodimer by the ClpXP protease: the role of trans recognition in UmuD' stability. *EMBO J*, **19**, 5251-5258.

[22] Gottesman S, Roche E, Zhou Y & Sauer RT (1998) The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **12**, 1338-1347.

[23] McGuinness KE, Baker TA & Sauer RT (2006) Engineering Controllable Protein Degradation. *Mol. Cell* **22**, 701-707.

[24] Brockman IM & Prather KLJ (2015) Dynamic knockdown of E. coli central metabolism for redirecting fluxes of primary metabolites. *Metab. Eng* **28**, 104-113.

[25] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna J & Charpentier E (2012) A Programmable Dual-RNA–Guided Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816-821.

[26] Jiang Y, Chen B, Duan C, Sun B, Yang J & Yanga S (2015) Multigene Editing in the Escherichia coli Genome via the CRISPR-Cas9 System. *Appl Environ Microbiol* **81**, 2506-2514.

[27] Chayot R, Montagne B, Mazel D & Ricchetti M (2010) An end-joining repair mechanism in Escherichia coli. *Proc Natl Acad Sci* **107**(5), 2141-2146.

[28] Yu D, Ellis HM, Lee EC, Jenkins Na, Copeland NG & Court DL (2000) An efficient recombination system for chromosome engineering in Escherichia coli *PNAS* **97**(11), 5978-5983.

[29] Cassuto E, Lash T, Sriprakash KS & Radding CM (1971) Role of Exonuclease and βProtein of Phage λ in Genetic Recombination, V. Recombination of λ DNA in Vitro. *Proc Natl Acad Sci* **68**(7), 1639-1643.

[30] Hanatani M, Yazyu H, Shiota-Niiya S, Moriyama Y, Kanazawa H, Futai M & Tshuchiya T (1984) Physical and Genetic Characterization of the Melibiose Operon and Identification of the Gene Products in Escherichia coli. *J Biol Chem* **259**, 1807-1812.

[31] Gafko C, Tobola F & Wiltschi B (2016) Project Lab Report. CRISPR-Cas9 mediated genome editing.

[32] Strockbine N, Jackson MP, Sung LM, Holmes RK & O'Brien AD (1988) Cloning and Sequencing of the Genes for Shiga Toxin from Shigella dysenteriae Type 1. *J. Bacteriol* **179**(3), 1116-1122.

[33] Endo Y, Tsurugt K, Yutsudo T, Takeda Y, Ogasawara T & Igarashi K (1988) Site of action of a Vero toxin (VT2) from Escherichia coli O157:H7 and of Shiga toxin on eukaryotic ribosomes RNA N-glycosidase activity of the toxins. *Eur. J. Biochem.* **171**, 45-50.

[34] Jacewicz M, Clausen H, Nudelman E, Donohue-Rolfe A & Keusch GT (1986) Pathogenesis of shigella diarrhea. XI. Isolation of a shigella toxin-binding glycolipid from rabbit jejunum and HeLa cells and its identification as globotriaosylceramide. *J Exp Med.* **163**(6), 1391-1404.

[35] Viel T, Dransart E, Nemati F, Henry E, Thézé B, Decaudin D, Lewandowski D, Boisgard R, Johannes L & Tavitian B (2008) In vivo tmor targeting by the B-subunit of shiga toxin. *Mol Imaging.* **7**(6), 239-247.

[36] Voloshchuk N & Montclare JK (2010) Incorporation of unnatural amino acids for synthetic biology. *Mol. Biosyst.* **6**(1), 65-80.

[37] Yanish-Perron C, Vieira J & Messing Joachim (1985) Improved M13 Phage Cloning Vectors and Host Strains: Nucleotide Sequences of the M13mp18 and pUC19 Vectors. *Gene 33*(1), 103-119.

[38] ATUM. CRISPR gRNA design tool. `https://www.atum.bio/eCommerce/cas9/%-input`. collected 2017-01-18.

[39] Electroporation of *E. coli*: BIW#35

[40] Wizard® SV Gel and PCR Clean-Up System from Promega: `https://www.promega.com/-/media/files/resources/protcards/`

wizard-sv-gel-and-pcr-clean-up-system-quick-protocol.pdf collected 2017-01-18.

[41] Gibson DG, Young L, Chuan RY, Venter JC, Hutchison CA & Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343-345.

[42] Gibson assembly protocol: BIW#122

[43] PureYield^{TM} Plasmid Miniprep System from Promega: https://at.promega.com/-/media/files/resources/protcards/pureyield-plasmid-miniprep-system-quick-protocol.pdf collected 2017-01-18.

[44] VivaSpin 3000 MWCO from Sartorius: https://www.sartorius.com/_ui/images/haa/h90/8871294009374.pdf collected 2018-02-22.

[45] Zeba^{TM} Spin for desalting from Thermo Fisher Scientific: https://www.thermofisher.com/order/catalog/product/89935?SID=srch-srp-89935 collected 2018-02-22.

[46] Laemmli UK (1970) Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**, 680-685.

[47] De Boer HA, Comstock LJ & Vasser M (1983) The tac promoter: A functional hybrid derived from the trp and lac promoters. *Proc. Natl. Acad. Sci.* **80**, 21-25.

[48] Orosz A, Boros I & Venetianer P (1991) Analysis of the complex transcription termination region of the *Escherichia coli* rrnB gene. *Eur. J. Biochem.* **201**, 653-659.

[49] Mullick A, Xu Y, Warren R, Koutroumanis M, Guilbault C, Broussau S, Malenfant F, Bourget L, Lamoureux L, Lo R, Caron AW, Pilotte A & Massie B (2006) The cumate gene-switch: a system for regulated expression in mammalian cells. *BMC Biotechnol* **6**(43).

[50] Kovbasnjuk O, Mourtazina R, Baibakov B, Wang T, Elowsky C, Choti MA, Kane A & Donowitz M (2005) The glycosphingolipid globotriaosylceramide in the metastatic transformation of colon cancer. *PNAS* **102**(52), 19087-19092.

[51] Fladischer P, Blamauer J, Pavkov-Keller T, Schweiger K, Darnhofer B, Birner-Gruenberger R, Gruber K & Wiltschi B A new archaeal pyrrolysyl-tRNA synthetase/amber suppressor tRNA pair for orthogonal protein translation. In preparation.

[52] Rostovtsev VV, Green LG, Fokin VV & Sharpless BK (2002) Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal alkynes. *Angew. Chem. Int. Ed.* **114**(14), 2708-1711.

[53] Cho H, Daniel T, Buechler YJ, Litzinger DC, Maio Z, Putnam A-MH, Kraynov VS, Sim BC, Bussell S, Javahishvili T, Kaphle S, Viramontes G, Ong M, Chu S, GC B, Lieu R, Knudsen N, Castiglioni P, Norman TC, Axelrod DW, Hoffman AR, Schultz PG, DiMarchi RD & Kimmel BE (2011) Optimized clinical performance of growth hormone with an expanded genetic code. *Proc Natl Acad Sci USA* **108**, 9060-9065.

[54] Kolb HC, Finn MG & Sharpless KB (2001) Click chemistry: Diverse chemical function from a few good reactions *Angew Chem Int Ed* **40**, 2004-2021.

[55] Wang Q, Chan TR, Hilgraf R, Fokin VV, Sharpless BK & Finn MG (2003) Bioconjugation by Copper(I)-Catalyzed Azide-Alkyne [3+2] Cycloaddition. *JACS* **125**(11), 3192-3193.

[56] Spicer CD & Davis BG (2014) Selective chemical protein modification. *Nat. Commun.* **5**.

[57] Presolski SI, Hong VP & Finn MG (2011) Copper-Catalyzed Azide–Alkyne Click Chemistry for Bioconjugation. *Curr Protoc Chem Biol.* **3**(4), 153-162.

[58] Niesen FH, Berglund H & Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc. 2*, 2212-2221.

[59] Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppa NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM & Church GM (2011) Precise Manipulation of Chromosomes in Vivo enables Genome-Wide Codon Replacement. *Science* **333**, 348-353.

[60] Terpe K (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems *Appl Microbiol Biotechnol* **72**, 211-222.

[61] Mairhofer J, Wittwer A, Cserjan-Puschmann M & Striedner G (2015) Preventing T7 RNA Polymerase Read-through Transcrition - A Synthetic Termination Signal Capable of Improving Bioprocess Stability. *ACS Synth. Biol.* **4**, 265-273.

[62] Wang K, Neumann H, Peak-Chew SY & Chin JW (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion *Nat. Biotechnol.* **25**, 770-777.

[63] Gallegos KM, Conrady DG, Karve SS, Gunasekera TS, Herr AB & Weiss AA (2012) Shiga Toxin Binding to Glycolipids and Glycans *PLoS One* **7**(2).

[64] Oh PY, Seong JT, Kim D-W, Kim E-C & Yoon K-H (2002) Simple Purification of Shiga Toxin B Chain from Recombinant Escherichia coli. *J. Microbiol. Biotechnol.* **12**(6), 986-988.

[65] Farabaugh PJ (1978) Sequence of the lacI gene. *Nature* **274**, 765-769.

[66] Lutz R & Bujard H (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **26**(6), 1203-1210.

[67] Calos MP (1978) DNA sequence for a low-level promoter of the lac repressor gene and an "up" promoter mutation. *Nature* **274**, 762-765.

Claudia Gafko, BSc

Genome editing of *Escherichia coli* improves the incorporation efficiency of non-canonical amino acids

Master's Thesis

to achieve the university degree of

Diplomingenieurin

Master's degree programme: Biotechnology

submitted to

Graz University of Technology

Supervisor

Assoc.-Prof. Dipl-Ing. Dr.techn. Harald Pichler

Institute of Molecular Biotechnology

Dipl.-Ing. Dr.techn. Birgit Wiltschi

Austrian Centre of Industrial Biotechnology

Graz, March 2018

A Supplementary material

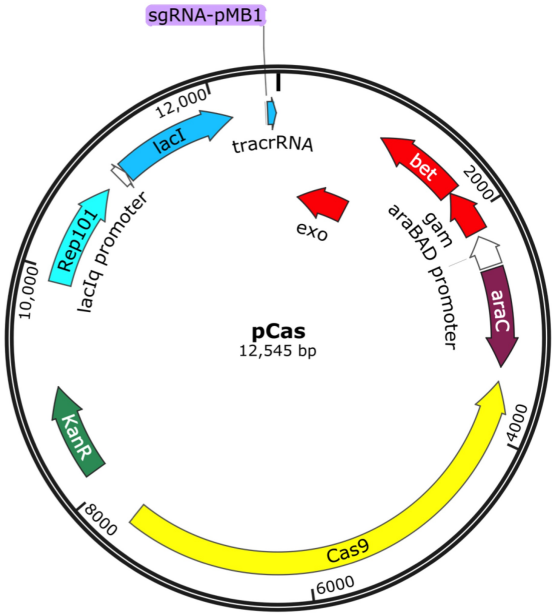
A.1 Plasmids

Supplementary Table 1: Description of all plasmids used and generated over the course of this project.

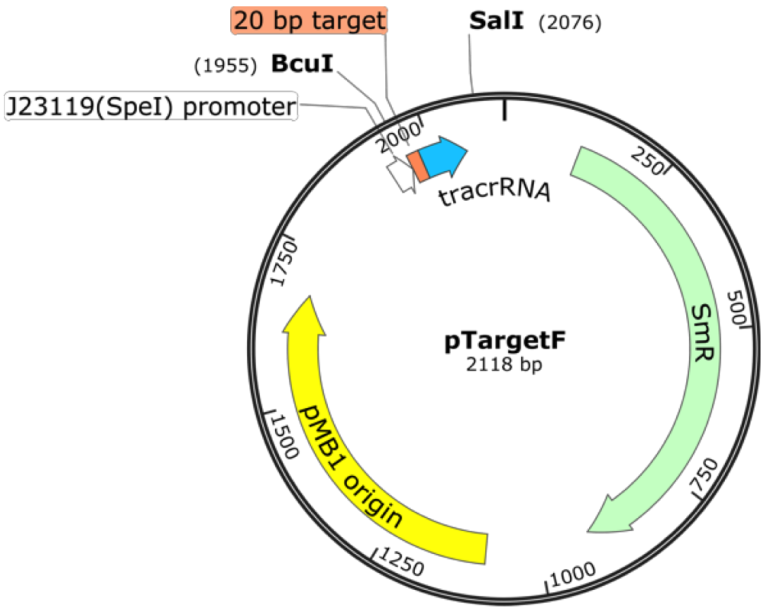
name	properties	note
pCas	*cas9*, *exo, bet* and *gam*	CRISPR/Cas9 genome editing
pMEL(500)RFP	editing fragment targeting the *melAB* locus	CRISPR/Cas9 genome editing
pMEL(500)lacI-WsspB	editing fragment targeting the *melAB* locus, genomic *sspB* construct with a weak RBS	CRISPR/Cas9 genome editing
pMEL(500)lacI-MsspB	editing fragment targeting the *melAB* locus, genomic *sspB* construct with a medium RBS	CRISPR/Cas9 genome editing
pMEL(500)lacI-SsspB	editing fragment targeting the *melAB* locus, genomic *sspB* construct with a strong RBS	CRISPR/Cas9 genome editing
pMEL(500)P–L*lacO* – 1-*sspB-lacI*	editing fragment targeting the *melAB* locus, genomic *sspB* construct with the P–L*lacO* – 1 promoter	CRISPR/Cas9 genome editing
pMEL(500)P–*lacUV*5-*sspB-lacI*	editing fragment targeting the *melAB* locus, genomic *sspB* construct with the P–*lacUV*5	CRISPR/Cas9 genome editing
pMEL(500)lacI	editing fragment targeting the *melAB* locus, *lacI* gene	CRISPR/Cas9 genome editing
pUC19		[37], backbone for pUC19-GFPssrA-RFP
pUC19-GFPssrA-RFP	bicistronic mRNA of GFP-*ssrA* and RFP	protein degradation experiment
pSCSara-empty	*MmMma* o-pair	control
pSCScum-empty	*MmMma* o-pair	control
pSCSara-1am	*MmMma* o-pair, eGFPx40, arabinose inducible *Mm*PylRS	ncAA incorporation
pSCScum-1am	*MmMma* o-pair, eGFPx40, cumate inducible *Mm*PylRS	ncAA incorporation
pSCScum-2am	*MmMma* o-pair, eGFPx40,134, cumate inducible *Mm*PylRS	ncAA incorporation
pSCScum-3am	*MmMma* o-pair, eGFPx40,134,213, cumate inducible *Mm*PylRS	ncAA incorporation

name	properties	note
pSCSara-stx1B	*MmMma* o-pair, wild-type *stx1B*, arabinose inducible *Mm*PylRS	AzK incorporation
pSCSara-stx1B-K8am	*MmMma* o-pair, amber mutant *stx1B*, arabinose inducible *Mm*PylRS	AzK incorporation
pT-murF-lolA-lpxK	mutation fragments for stop codon exchange in *murF, lolA* and *lpxK*	stop codon exchange
pT-hemA-hda	mutation fragments for stop codon exchange in *hemA* and *hda*	stop codon exchange
pT-mreC-coaD	mutation fragments for stop codon exchange in *mreC* and *coaD*	stop codon exchange

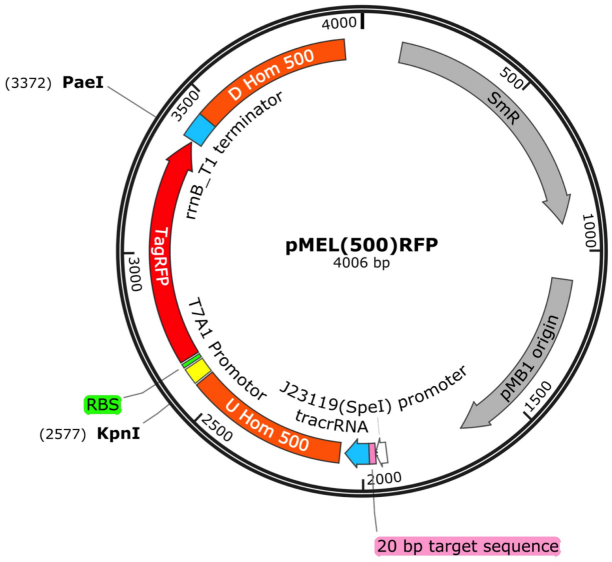
A.1.1 Plasmid maps



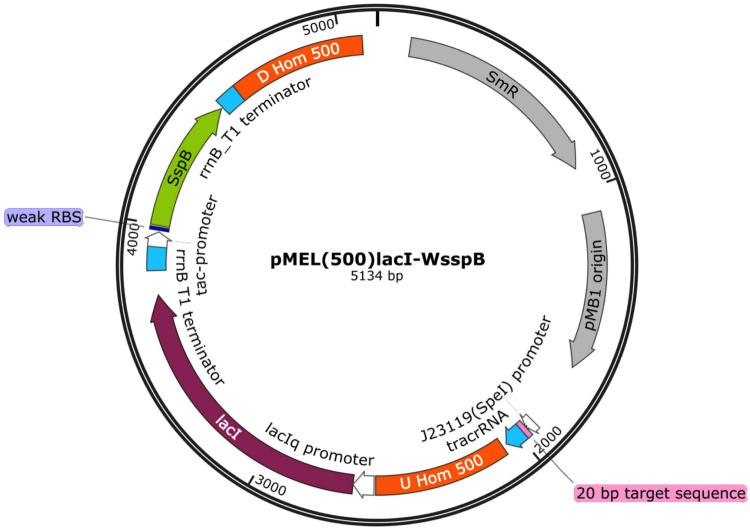
Supplementary Figure 1: Plasmid map of pCas used for CRISPR/Cas9 mediated genome editing. pCas carried the endonuclease Cas9 (yellow), a kanamycin resistance marker (green), the λ-Red genes *exo, bet* and *gam* (red), the heat sensitive origin of replication Rep101 (cyan) and a sg-RNA targeting the pMB1 origin of replication of pTargetF (purple).



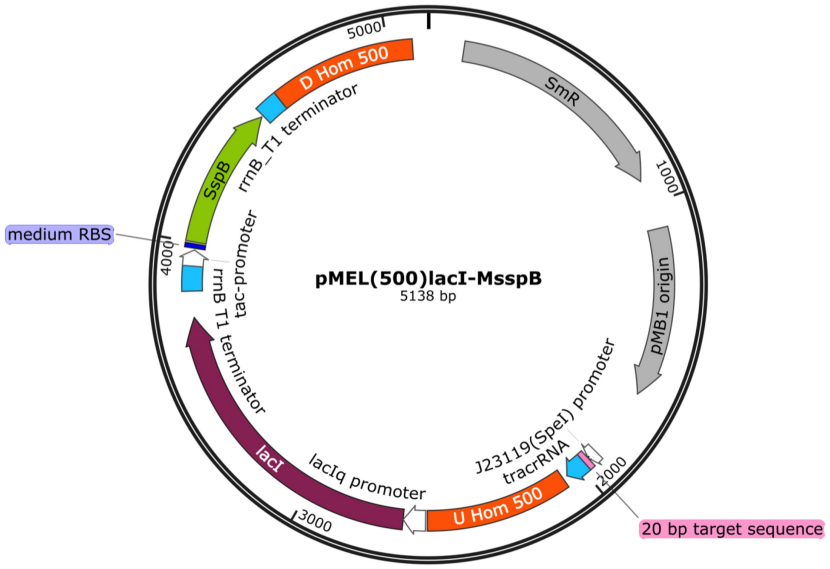
Supplementary Figure 2: Plasmid map of pCas used for CRISPR/Cas9 mediated genome editing. pCas carried the endonuclease Cas9 (yellow), a kanamycin resistance marker (green), the λ-Red genes *exo, bet* and *gam* (red), the heat sensitive origin of replication Rep101 (cyan) and a sg-RNA targeting the pMB1 origin of replication of pTargetF (purple).



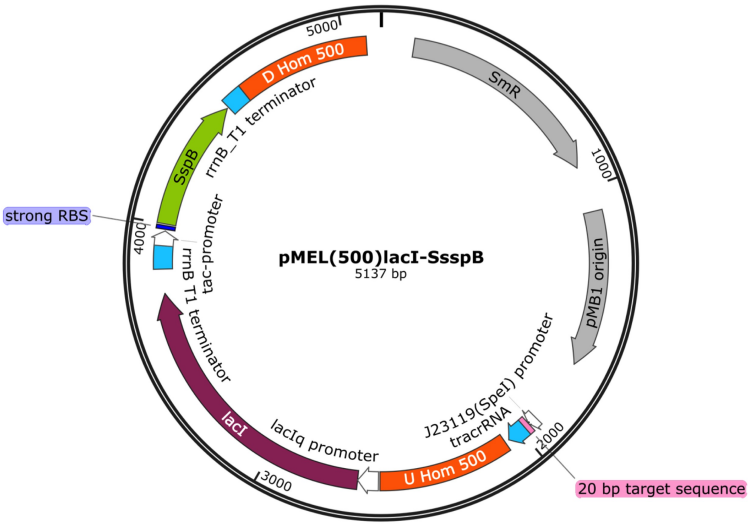
Supplementary Figure 3: Plasmid map of pMEL(500)RFP used as backbone for all target plasmids. pMEL(500)RFP carries a spectinomycin resistance marker and the pMB1 origin of replication (grey). The integration fragment contains the reporter gene TagRFP (red), which is regulated by an constitutive promoter T7A1 and the terminator *rrnB* T1. It is flanked by the 500 bp upstream and downstream homologies, specific for the *melAB* locus (orange). Upstream of the integration fragment, the synthetic sg-RNA for the *melAB* locus is positioned. The sg-RNA consisted of the 20 bp target sequence (pink) and tracrRNA (blue).



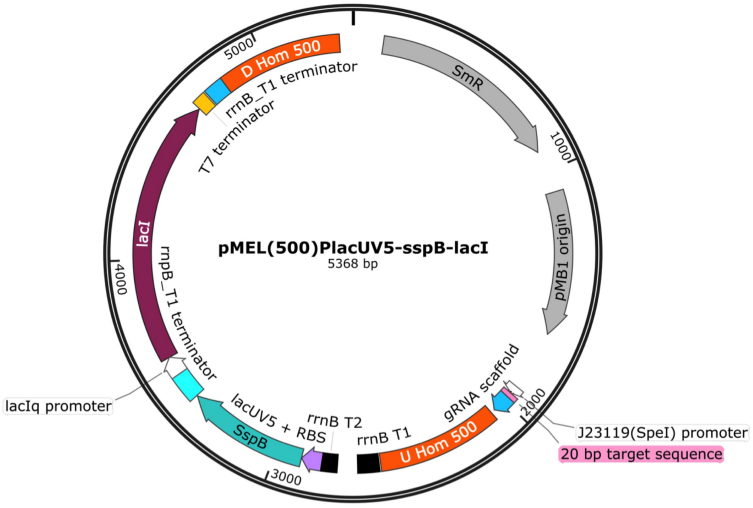
Supplementary Figure 4: Plasmid map of pMEL(500)lacI-WsspB used for stable integration of the genomic *sspB* construct at the *melAB* locus by CRISPR/Cas9 mediated genome editing. This plasmid carried the same origin of replication, resistance marker, upstream and downstream homology hooks and sgRNA as pMEL(500)RFP. The genomic *sspB* construct with a weak RBS is indicated in green. It was regulated by the IPTG-inducible *tac*-promoter and the rrnB T1 terminator (blue). The *lacI* gene is indicated in purple and was regulated by the constitutive P^{q}_{lacI} promoter and *rrnB* T1 terminator.



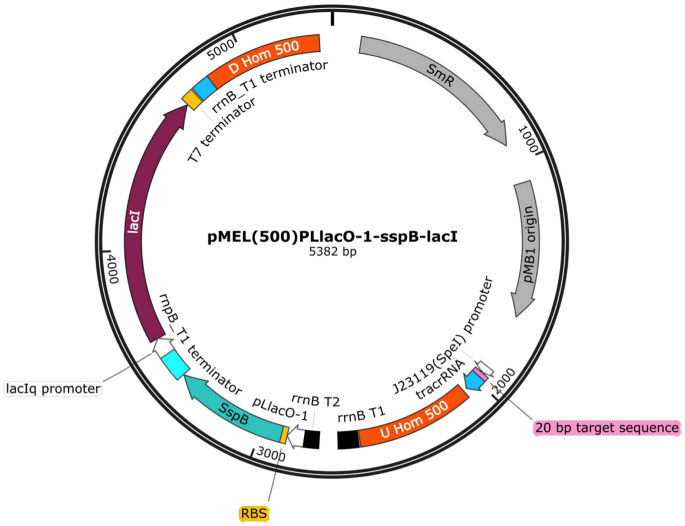
Supplementary Figure 5: Plasmid map of pMEL(500)lacI-MsspB used for stable integration of the genomic *sspB* construct at the *melAB* locus by CRISPR/Cas9 mediated genome editing. This plasmid was identical to pMEL(500)lacI-WsspB, only that the genomic *sspB* construct contained an RBS of medium strength.



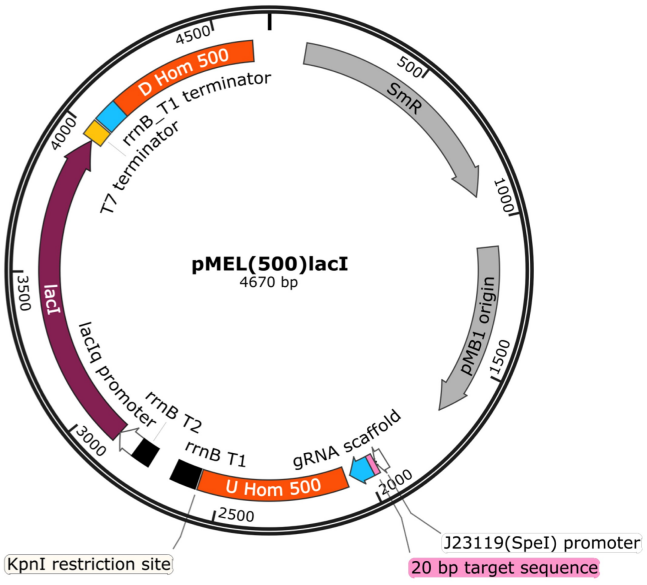
Supplementary Figure 6: Plasmid map of pMEL(500)lacI-SsspB used for stable integration of the genomic *sspB* construct at the *melAB* locus by CRISPR/Cas9 mediated genome editing. This plasmid was identical to pMEL(500)lacI-WsspB, only that the *sspB* gene contained a strong RBS.



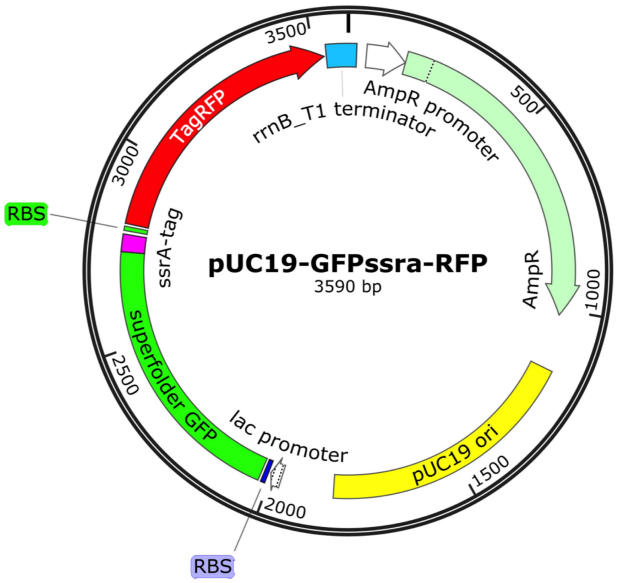
Supplementary Figure 7: Plasmid map of pMEL(500)P$_{lacUV5}$-*sspB-lacI*, which was used for stable integration of the genomic *sspB* construct at the *melAB* locus by CRISPR/Cas9 mediated genome editing. This plasmid carried the same origin of replication, resistance marker, upstream and downstream homology hooks and sg-RNA as pMEL(500)RFP, but the *rfp* sequence was exchanged with the genomic *sspB* construct, consisting of a *lacI* and a *sspB*, which was controlled by P$_{lacUV5}$ promoter.



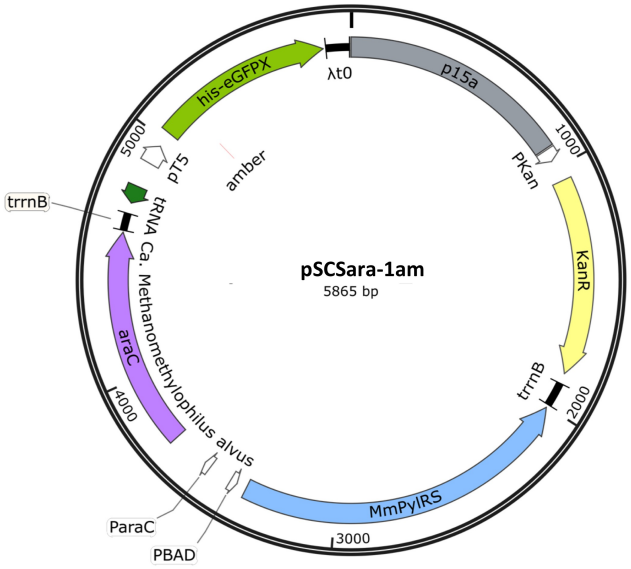
Supplementary Figure 8: Plasmid map of pMEL(500)$\mathrm{P_L}$*lacO*-1-*sspB-lacI* used for stable integration of the genomic *sspB* construct at the *melAB* locus by CRISPR/Cas9 mediated genome editing. This plasmid was identical to pMEL(500)P_{lacUV5}-*sspB-lacI*, but its *sspB* was controlled by the $\mathrm{P_{L}}_{lacO-1}$ promoter.



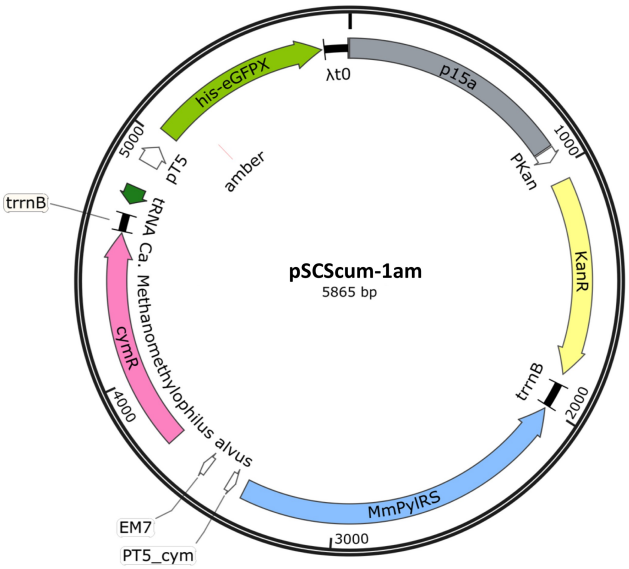
Supplementary Figure 9: Plasmid map of pMEL(500)lacI used for stable integration of *lacI* at the *melAB* locus by CRISPR/Cas9 mediated genome editing. This plasmid carried an editing fragment that consisted of the sequence for *lacI*, which was flanked by double-terminators *rrnB* T1 and T2 and T7 and *rrnB* T1.



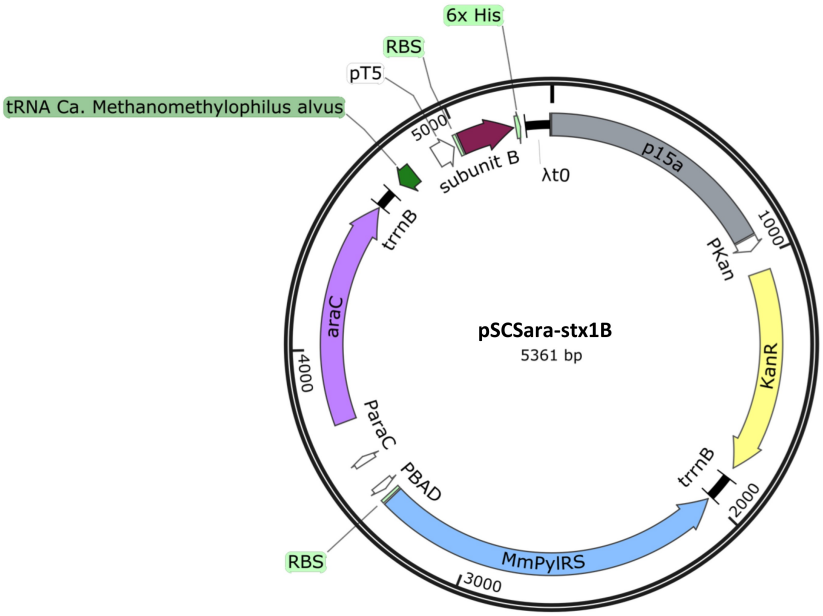
Supplementary Figure 10: Plasmid map of pUC19-GFP*ssrA*-RFP used for the protein degradation proof-of-concept experiment. The plasmid contained the pUC19 backbone, with an ampicillin resistance marker (light green), a pUC19 origin of replication (yellow) and a bi-cistronic sfGFP-*ssrA*-TagRFP construct, which is controlled by a consitutive *lac* promoter and an *rrnB* T1 terminator.



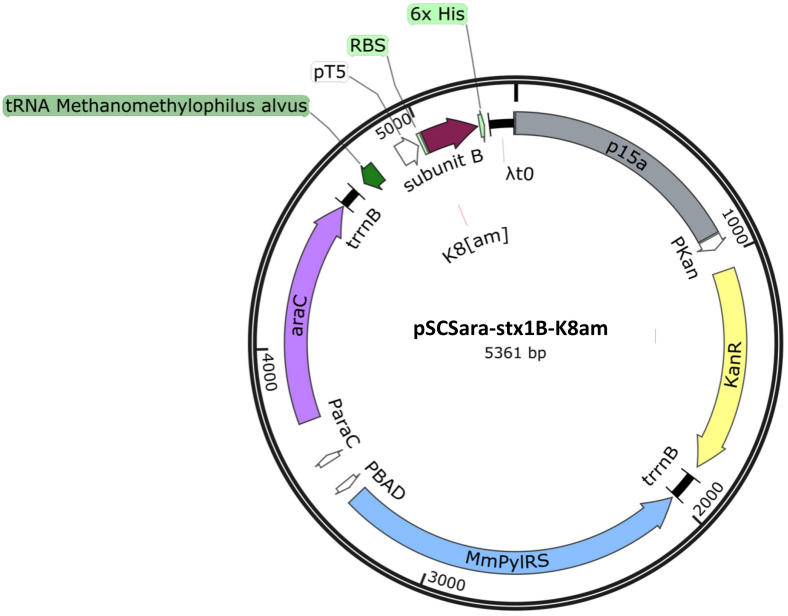
Supplementary Figure 11: Plasmid map of pSCSara-1am used for recombinant GFPx expression and incorporation of BocK. The plasmid carried a p15a origin of replication (grey), a kanamycin resistance marker (yellow), the *Methanosarcina mazei* descendent *Mm*PylRS under the control of an arabinose inducible promoter (blue), the constitutively expressed *Methanomethylophilus alvus* derived *Mma* Pyl-tRNA (dark green) and the amber mutant reporter gene eGFPx40.



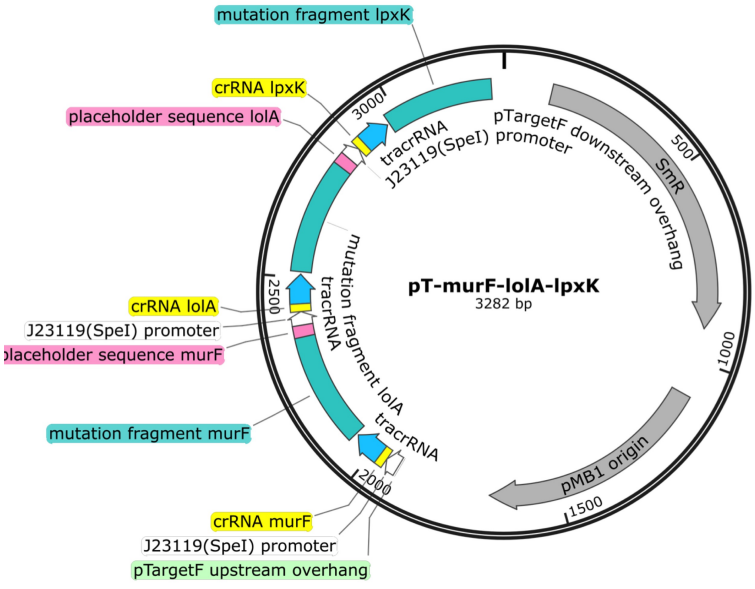
Supplementary Figure 12: Plasmid map of pSCScum-1am used for recombinant GFPx expression and incorporation of BocK. This plasmid was identical to to pSCSara-1am, only that the *Mm*PylRS was controlled by a cumate inducible promoter P_{T5_cym}. The plasmids pSCScum-2am and pSCScum-3am were identical to this plasmid, only the reporter gene eGFPx carried an amber stop mutation either at position 40 and 134 or 40, 134 and 213.



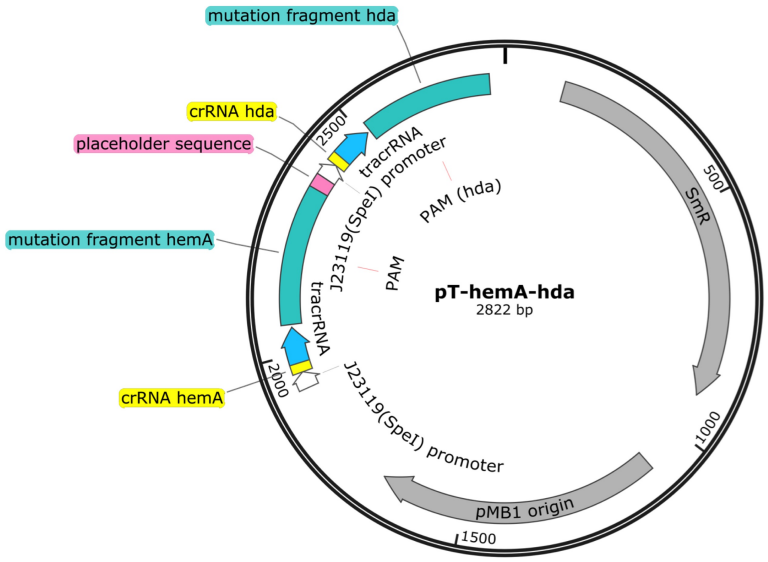
Supplementary Figure 13: Plasmid map of pSCSara-stx1B used for recombinant Stx1B wild-type expression and. This plasmid carried the p15a origin of replication (grey), a kanamycin resistance marker (yellow), an arabinose inducible *Mm*PylRS (blue), a constitutively expressed *Mma*tRNA (dark green) and the wild-type *stx1B* gene (purple). The *stx1B* gene carried a C-terminal hexahistidine purification tag.



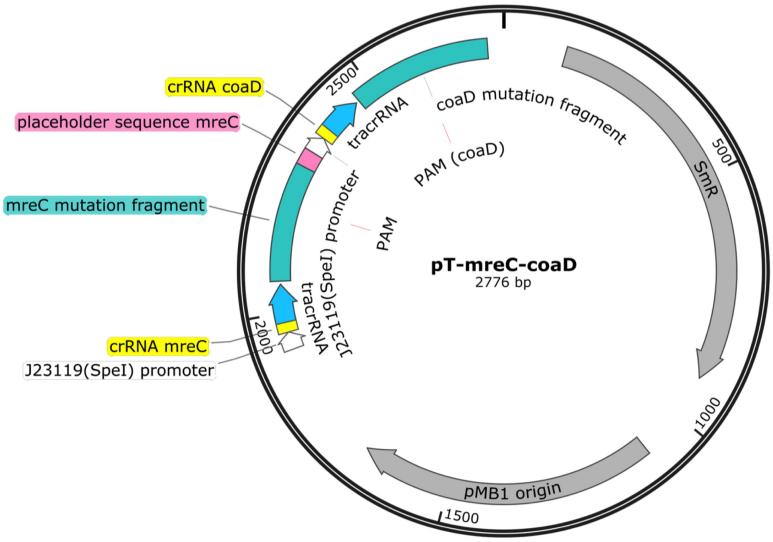
Supplementary Figure 14: Plasmid map of pSCSara-stx1B-K8am used for recombinant stx1B-K8[AzK] expression. This plasmid was identical to pSCSara-stx1B, only that it contained the *stx1B* gene, with an amber stop mutation at position 8 (purple).



Supplementary Figure 15: Tandem target plasmid pT-murF-lolA-lpxK carrying the editing cassettes for stop codon exchange of the essential genes *murF, lolA* and *lpxK*. The plasmid contained the pTargetF backbone, with a spectinomycin resistance marker and the pMB1 origin or replication (grey). The editing cassettes for *murF, lolA* and *lpxK* consisted of the specific sg-RNA (blue) and the respective mutation fragment for each locus (cyan).



Supplementary Figure 16: Tandem target plasmid pT-hemA-hda carrying the editing cassettes for stop codon exchange of the essential genes *hemA* and *hda*. The plasmid was identical to pT-murF-lolA-lpxK, only that it carried the editing cassettes for *hemA* and *hda*.



Supplementary Figure 17: Tandem target plasmid pT-mreC-coaD carrying the editing cassettes for stop codon exchange of the essential genes *mreC* and *coaD*. The plasmid contained the pTargetF backbone and the editing cassettes for *mreC* and *coaD*.

A.2 Sequences and primers

Supplementary Table 2: Primers, which were used throughout this project. Overlapping sequences are indicated in bold letters.

Primer	Target sequence & features (5' → 3')	⇄	Sequence (5' → 3')
number	name	fwd-rev	sequence
pBP2023	melAB locus colony PCR	rev	acgtaaaatggtggcgttcc
pBP2026	pTargetF sequencing	fwd	aagcggaagagcgcc
pBP2062	murF TAG MASC PCR	fwd	ctttacaggagaatgggacatgttag
pBP2063	murF TGA MASC PCR	fwd	ctttacaggagaatgggacatgctga
pBP2064	murF MASC PCR	rev	gacgggtaagcccacagc
pBP2065	lolA TAG MASC PCR	fwd	tcacggtagatgatcaacgtaagtag
pBP2066	lolA TAA MASC PCR	fwd	tcacggtagatgatcaacgtaagtaa
pBP2067	lolA MASC PCR	rev	ctgaccaccatagccacgg
pBP2068	lpxK TAG MASC PCR	fwd	gctggcttctggcaactag
pBP2069	lpxK TAA MASC PCR	fwd	gctggcttctggcaactaa
pBP2070	lpxK MASC PCR	rev	tttgtatttccagcccatttttcag
pBP2071	melAB locus sequencing	fwd	atgactgcacccaaaattacat
pBP2072	melAB locus sequencing	rev	tgcgatatttatccagcagatg
pBP2075	WsspB + Ptac overhang	fwd	**tgtggaattgtgagcgctcacaatt**gacgtctcacacaggactctagaatggatttgtcacagctaacacc
pBP2076	MsspB + Ptac overhang	fwd	**tgtggaattgtgagcgctcacaatt**gacgtcattaaagaggagaaatctagaatggatttgtcacagctaacacc
pBP2077	SsspB + Ptac overhang	fwd	**tgtggaattgtgagcgctcacaatt**gacgtcattaaagaggagaatctagaatggatttgtcacagctaacacc
pBP2078	sspB + rrnB terminator	rev	cgttttatttgatgcctgggcatgcttacttcacaacgcgtaatgc
pBP2079	lacI	fwd	gcctggaaggtaccgacaccatcgaatggtgca
pBP2080	lacI	rev	acattatacgagccgatgattaattgtcaacacaaataaaacgaaaggctcagtcg
. pBP2081	melAB locus inside HOM	fwd	tgatgtgttccatcgcgag
pBP2082	melAB locus outside HOM	fwd	caggtattccacatcgccatc
pBP2083	lacI sequencing	rev	aacagctgattgcccttcac
pBP2084	lacI sequencing	fwd	gtccggttttcaacaaaccatg
pBP2088	target plasmid colony PCR	fwd	cggtatttcacaccgcatatg
pBP2089	target plasmid colony PCR	rev	tcgagtagggataacagggtaata
pBP2096	TagRFP + GFP overhang	fwd	**aagcttaaagaggagaaattaact**atggtgtctaagggc
pBP2097	TagRFP + rrnBT1 terminator	rev	gagagcgttcaccgacaa
pBP2098	sfGFP + Plac	fwd	ttagcgacgtctttacactttatgc
pBP2099	sfGFP + TagRFP overhang	rev	**gacaccat**agttaatttctcctctttaag
pBP2100	pUC19 linear + TagRFP overhang	fwd	**gttgtttgtcggtgaacgctctc**caggtggcacttttcgg
pBP2101	pUC19 linear + sfGFP overhang	rev	**ggaagcataaagtgtaaagacgtcgctaa**ggcgctcttccgct

Supplementary Table 3: Primers, which were used throughout this project. sg-RNA sequences are written in capital, overhangs in bold letters.

Primer	Target sequence & features (5' → 3')	⇄	Sequence (5' → 3')
number	name	fwd-rev	sequence
pBP2104	rrnB T1T2 + overhang UP HOM	fwd	**gcctggaa**ggtacccaaataaaac
pBP2105	T7 and rrnB T1 terminator	rev	atgcctgggcatgcct
pBP2106	PlacUV5-sspB + rrnB T2 overhang	rev	gttatccgctcacaattccacacattatacgagccggaagcataaagtgtaaagcctgg**cgacaggaagagtttgtagaaacgca**
pBP2107	PlacUV5-sspB	fwd	cgtataatgtgtggaattgtgagcggataacaatttcacacaggaaacagaattctatggatttgtcacagctaacaccac
pBP2110	pCas curing	fwd	cattatcgcgagcccatttatacc
pBP2111	pCas curing	rev	tctaagcctattgagtatttcttatccatttttg
pBP2130	stx1B gibson assembly	fwd	gcggataacaatttcacacagaattc
pBP2131	stx1B gibson assembly	rev	ctggatctatcaacaggagtcga
pBP2132	UP HOM + rrnB T1 OexPCR	fwd	ccacccgcctggaagg
pBP2133	rrnB T2 + PlacI OexPCR	rev	**ggttttgcaccattcgatggtgtc**cgacaggaagagtttgtagaaacg
pBP2134	PlacIq OexPCR	fwd	gacaccatcgaatggtgcaa
pBP2135	rrnB T1 OexPCR	rev	tttcgttttatttgatgcctgggc
pBP2136	sspB sequencing	rev	gcgacatcgtataacgttactgg
pBP2137	prfA sequencing	fwd	gggctggagtagcagtacat
pBP2138	prfA sequcning	rev	ccagcaggatttcagcatcac

Supplementary Table 4: Sequences of all genes, gBlocks, DNA parts used in this project. Start codons are indicated in orange, stop codons in red, ribosome binding sites are in bold letters, coding regions in cyan, promoter regions in brown, terminator regions in gray, hexa-histidine tag in green and the ssrA-tag in purple.

name	sequence (5' → 3')	note
ssrA-tag	GCAGCGAACGATGAAAACTATTCTGAAAACTATGCGGATGCCTCT	
gfp Y40am	ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAGGGCAAG CTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACC ACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTC TTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGC AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG AAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAAC AGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATC CGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATC GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAA GACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACT CTCGGCATGGACGAGCTGTACAAGTAA	
gfp Y40am, D134am	ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAGGGCAAG CTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACC ACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTC TTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGC AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG AAGGGCATCGACTTCAAGGAGTAGGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAAC AGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATC CGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATC GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAA GACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACT CTCGGCATGGACGAGCTGTACAAGTAA	
gfp Y40am, D134am, N213am	ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAGGGCAAG CTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACC ACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTC TTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGC AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG AAGGGCATCGACTTCAAGGAGTAGGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAAC AGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATC CGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATC GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAA GACCCCTAGGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACT CTCGGCATGGACGAGCTGTACAAGTAA	
super folder *gfp*	ATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGAT GTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGAAAACTC ACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACT CTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTC AAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGACC TACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAG GGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAACTCGAGTACAACTTTAACTCA CACAATGTATACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGC CACAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGC GATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGAT CCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACAT GGCATGGATGAGCTCTACAAATAA	
lacI	GTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCC CGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATG GCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTG ATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAA TCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAA GCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTAT CCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTT CTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGA CTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTA AGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATT CAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAA ATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGC GCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATAC GACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGC CTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGC AATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACC GCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAA AGCGGGCAGTGA	2), [65]
sspB	ATGGATTTGTCACAGCTAACACCACGTCGTCCCTATCTGCTGCGTGCATTCTATGAGTGGTTG CTGGATAACCAGCTCACGCCGCACCTGGTGGTGGATGTGACGCTCCCTGGCGTGCAGGTTCCT ATGGAATATGCGCGTGACGGGCAAATCGTACTCAACATTGCGCCGCGTGCTGTCGGCAATCTG GAACTGGCGAATGATGAGGTGCGCTTTAACGCGCGCTTTGGTGGCATTCCGCGTCAGGTTTCT GTGCCGCTGGCTGCCGTGCTGGCTATCTACGCCCGTGAAAATGGCGCAGGCACGATGTTTGAG CCTGAAGCTGCCTACGATGAAGATACCAGCATCATGAATGATGAAGAGGCATCGGCAGACAAC GAAACCGTTATGTCGGTTATTGATGGCGACAAGCCAGATCACGATGATGACACTCATCCTGAC GATGAACCTCCGCAGCCACCACGCGGTGGTCGACCGGCATTACGCGTTGTGAAGTAA	2)

TAG*rfp*	ATGGTGTCTAAGGGCGAAGAGCTGATTAAGGAGAACATGCACATGAAGCTGTACATGGAGGGC ACCGTGAACAACCACCACTTCAAGTGCACATCCGAGGGCGAAGGCAAGCCCTACGAGGGCACC CAGACCATGAGAATCAAGGTGGTCGAGGGCGGCCCTCTCCCCTTCGCCTTCGACATCCTGGCT ACCAGCTTCATGTACGGCAGCAGAACCTTCATCAACCACACCCAGGGCATCCCCGACTTCTTT AAGCAGTCCTTCCCTGAGGGCTTCACATGGGAGAGgGTCACCACATACGAAGACGGGGGCGTG CTGACCGCTACCCAGGACACCAGCCTCCAGGACGGCTGCCTCATCTACAACGTCAAGATCAGA GGGGTGAACTTCCCATCCAACGGCCCTGTGATGCAGAAGAAAACACTCGGCTGGGAGGCCAAC ACCGAGATGCTGTACCCCGCTGACGGCGGCCTGGAAGGCAGAAGCGACATGGCCCTGAAGCTC GTGGGCGGGGGCCACCTGATCTGCAACTTCAAGACCACATACAGATCCAAGAAACCCGCTAAG AACCTCAAGATGCCCGGCGTCTACTATGTGGACCACAGACTGGAAAGAATCAAGGAGGCCGAC AAAGAGACCTACGTCGAGCAGCACGAGGTGGCTGTGGCCAGATACTGCGACCTCCCTAGCAAA CTGGGGCACAAACTTAATTGA	
sg-RNA *melAB* locus	CTCACGTTTATCGAGCGTTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGT TATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGAGTCAATTG	
Weak RBS	TCACACAGGAC	
Medium RBS	ATTAAAGAGGAGAAA	
Strong RBS	ATTAAAGAGGAGAA	
tac promoter	TGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGCTCACAATT	[47]
*lac*UV5 promoter	CCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATT TCACACAGGAAACAGAATTCT	[47]
L*lac*O-1 promoter	ATAAATGTGAGCGGATAACATTGACATTGTGAGCGGATAACAAGATACTGAGCACATCAGCAG GACGCACTGACC	[66]
lacI^{q} promoter	GACACCATCGAATGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAA TTCAGGGTGGTGAAT	[2], [67]
*rrnB*T1 T2 terminator	GGTACCCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTG TCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAA CGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAG GCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTCCTGTCG	[2], [48]
*rrnB*T1 terminator	ATTTGTCCTACTCAGGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTT CGACTGAGCCTTTCGTTTTATTTG	[2], [48]
T7 terminator	TCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGG	[61]
rnpB terminator	CCGGCTTATCGGTCAGTTTCACCTGATTTACGTAAAAACCCGCTTCGGCGGGTTTTTGCTTTT GGAGGGGCAGAAAGATGAATGACTGTCCACGACGCTATACCCAAAAGAAA	[2]
murF editing cassette (gBlock)	CCCGGGCGGTATTTCACACCGCATATGCTGGATCCTTGACAGCTAGCTCAGTCCTAGGTATAA TACTAGTGCGTACTACCTCTTCCATGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTACTGCGTTAATTACGCG TCTTAAATCACTGATTGCTGAGCAACAGGTAATTACGATTTTAGTTAAGGGTTCACGTAGTGC AGCCATGGAAGAGGTAGTACGCGCTTTACAGGAGAATGGGACATGCTGATCTGGCTGGCCGAA CATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCCTATCTGACGTTTCGCGCCATCGTC AGCCTGCTGACCGCGCTGTTCATCTCATTGTGGATGGGCCCGCGTATGATTGCTCATTTGCAA AAACTTTCCTTTGGTCGGGTCTCTCAATGTTAATCAATATGACCCGGG	[1]
lolA editing cassette (gBlock)	CCCGGGTCTCTCAATCGTTAATCAATATGACCCTTGACAGCTAGCTCAGTCCTAGGTATAATA CTAGTTAGATGATCAACGTAAGTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTAAATTTACCTTCACCCCGC CGCAAGGCGTCACGGTAGATGATCAACGTAAGTAAAGCCACCTGAGTGAGCAATCTGTCGCTC GATTTTTCGGATAATACTTTTCAACCTCTGGCCGCGCGTATGCGGCCAGAAAATTTAGCACAG TATATCGGCCAGCAACATTTGCTGGCTGCGGGGAAGCCGTTGCCGCGCGCTATCGAAGCCGGG CATTTACATTCTATGATCCTCTGGGGGCCGCCGGGTACCGGCAAAACAACTCTCGCTGAAGTG ATTGCCCGCTATGCGAACGGGGTCAGTAGATTAAGAAATGATATCGCCCGGG	[1]
lpxK editing cassette (gBlock)	CCCGGGTCAGTAGATTAAGAAATGATATCGCCCTTGACAGCTAGCTCAGTCCTAGGTATAATA CTAGTCCGCGGCAGCGTTCGATTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGGTGGTATTTGCCTGTAGA CGCACAGCTTTCAGGTGATGAACCAGCGAAACTGCTTACGCAACTAACCTTGCTGGCTTCTGG CAACTAATTACGCCGCGGCAGCGTTCGATTGATGCAGTCATGAATGTCGCTGCCGCACCTCTC CCTTGCTGATGCGCGTAATCTTCACCTTGCCGCACAAGGCCTGTTAAACAAACCCCGCCGTCG AGCGTCGTTGGAGGATATTCCGGCAACGATCTCCCGCATGTCCTTGCTGCAAATCGATACCAT CAATATTGTAAGCTTAGATCTATTACCCTGTTATCCCTACTCGACCCGGG	[1]
hemA editing cassette (gBlock)	CCCGGGCGGTATTTCACACCGCATATGCTGGATCCTTGACAGCTAGCTCAGTCCTAGGTATAA TACTAGTACATCATTTTCTTTTTTTACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTAGGATCTGGCATGGAAA CTGACTAACCGCTTGATCCATGCGCCAACGAAATCACTTCAACAGGCCGCCCGTGACGGGGAT AACGAACGCCTGAATATTCTGCGCGACAGCCTCGGGCTGGAGTAACAGTACATCATTTTCTTT TTTTACAGCGTGCATTTACGCCTATGAAGCCTTCTATCGTTGCCAAACTGGAAGCCCTGCATG AACGCCATGAAGAAGTTCAGGCGTTGCTGGGTGACGCGCAAACTATCGCCGACCAGGAACGTT TTCGCGCATTATCACGCGAATGGGGACAAGTTTATGACACAATCTAACCCGGG	[1]

hda editing cassette (gBlock)	CCCGGGGACAAGTTTATGACACAATCTAACCCTTGACAGCTAGCTCAGTCCTAGGTATAATAC TAGTTGGATTAACTTCGCGTAGTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGT CCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCTGGAAATCGTGGGTTAAGA TGTGCTCCATCGCTAGTTGAAGCACATTGCCGGATGCGACGCTTGCCACGTCTTATCCGGCCT ACGAATACTATGGATTAACTTCGCGTAGTTCGCATAAGGCGTTCGCGCCGCATCCGACAATAA ACACCTTATTTACAACTTCAGAATTTCTTTCACAAACGGAATGGTCAGCTTACGTTGCGCGGT AATCGACGCACGATCCAACTGATCCAACGTCATAAATAGCGTGCGCATTTCTCTGTCGAGCCG CAAGCTTAGATCTATTACCCTGTTATCCCTACTCGACCCGGG	1)
mreC editing cassette (gBlock)	CCCGGGCGGTATTTCACACCGCATATGCTGGATCCTTGACAGCTAGCTCAGTCCTAGGTATAA TACTAGTGCTCGCCACTATTGCCCTCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTATCCAATACAACAAGAT GAGTAACACCCAGTTTGGCCGGAAAACAATCAGGTTATCCGGCCAGGGCATGATTTGCAGCAA CAGCGCAATGAGGAAAGAGAGCCAGATTACCCAGCGTCCCTGGCTACGATAAGAAGCCATCAC TGCCCTCCCGGCGCACGCGCAGGCGGTTGAGCACCACTTTGCGGCGGCGTAGCCCTTTGTGGA GAGCGATTAGCAGCAGGCTGTGTCGGCGGGTTCATATCTTGCATAGACTGTATTAGTCCCGGG	1)
coaD editing cassette (gBlock)	CCCGGGTTCATATCTTGCATAGACTGTATTAGTCCCTTGACAGCTAGCTCAGTCCTAGGTATA ATACTAGTATGCCGGATGGTATGCCATCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC TAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGAGTGGTCGTTTATCT CTTCATCGTTGGTGAAAGAGGTGGCGCGCCATCAGGGCGATGTCACCCATTTCCTGCCGGAGA ATGTCCATCAGGCGCTGATGGCGAAGTTAGCGTAACGTTTATGCCGGATGGTATGCCATCCGC CGCGCATGAATTACTTCTGGCACTGCCGACAATAAAACGTTGCCCGCTGCGCATGTTTAGTCG CCACAATCGGCGTACCGCACACCCGACACGGCTCACCTTTTCGCCCGTAAACCTGCAATTCCT GAGCGAAATAGCAAGCTTAGATCTATTACCCTGTTATCCCTACTCGACCCGG	1)
wild-type *stx1B* (gBlock)	GCGGATAACAATTTCACACAGAATTCAGATCT**AAAGAGGAGAAA**AGAGCCATGGCGACGCCTG ATTGTGTAACTGGAAAGGTGGAGTATACAAAATATAATGATGACGATACCTTTACAGTTAAAG TGGGTGATAAAGAATTATTTACCAACAGATGGAATCTTCAGTCTCTTCTTCTCAGTGGCGCAAA TTACGGGGATGACTGTAACCATTAAAACTAATGCCTGTCATAATGGAGGGGGATTCAGCGAAG TTATTTTTCGTGGTTCTGGCCATCACCATCACCATCACTAAACTAGTAGATCTAATTCGACTC CTGTTGATAGATCCAG	1)
stx1B-K8[am] (gBlock)	GCGGATAACAATTTCACACAGAATTCAGATCT**AAAGAGGAGAAA**AGAGCCATGGCGACGCCTG ATTGTGTAACTGGATAGGTGGAGTATACAAAATATAATGATGACGATACCTTTACAGTTAAAG TGGGTGATAAAGAATTATTTACCAACAGATGGAATCTTCAGTCTCTTCTTCTCAGTGCGCAAA TTACGGGGATGACTGTAACCATTAAAACTAATGCCTGTCATAATGGAGGGGGATTCAGCGAAG TTATTTTTCGTGGTTCTGGCCATCACCATCACCATCACTAAACTAGTAGATCTAATTCGACTC CTGTTGATAGATCCAG	1)
super folder*gfp-ssrA* (gBlock)	TTAGCGACGTCTTTACACTTTATGCTTCCGGCTCGTATGTTGGGTACC**AAAGAGGAGAAA**TTA ACTATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGT GATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGAAAA CTCACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACT ACTCTGACCTATGGTGTTCAATGCTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTT TTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGG ACCTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTA AAGGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAACTCGAGTACAACTTTAAC TCACACAATGTATACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATT CGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATT GGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAA GATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACA CATGGCATGGATGAGCTCTACAAAGCAGCGAACGATGAAAACTATTCTGAAAACTATGCGGAT GCCTCTTAAAAGCTT**AAAGAGGAGAAA**TTAACTATGGTGTC	1)
integration fragment $P_{L\mathit{lacO}-1}$ (gBlock)	CATGATGACTGCACCCAAAATTACATTTATCGGCGCTGGTTCGACGATTTTCGTTAAAAATAT TCTTGGTGATGTGTTCCATCGCGAGGCGCTGAAAACGGCGCATATTGCCCTGATGGACATTGA TCCCACCCGCCTGGAAGGTACCCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCG TTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTT GAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCA TCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTCCTGT CGATAAATGTGAGCGGATAACATTGACATTGTGAGCGGATAACAAGATACTGAGCACATCAGC AGGACGCACTGACCTGCGTCTGGGCCGGAGTTAATCTGTATGGATTTGTCACAGCTAACACCA CGTCGTCCCTATCTGCTGCGTGCATTCTATGAGTGGTTGCTGGATAACCAGCTCACGCCGCAC CTGGTGGTGGATGTGACGCTCCCTGGCGTGCAGGTTCCTATGGAATATGCGCGTGACGGGCAA ATCGTACTCAACATTGCGCCGCGTGCTGTCGGCAATCTGGAACTGGCGAATGATGAGGTGCGC TTTAACGCGCGCTTTGGTGGCATTCCGCGTCAGGTTTCTGTGCCGCTGGCTGCCGTGCTGGCT ATCTACGCCCGTGAAAATGGCGCAGGCACGATGTTTGAGCCTGAAGCTGCCTACGATGAAGAT ACCAGCATCATGAATGATGAAGAGGCATCGGCAGACAACGAAACCGTTATGTCGGTTATTGAT GGCGACAAGCCAGATCACGATGATGACACTCATCCTGACGATGAACCTCCGCAGCCACCACGC GGTGGTCGACCGGCATTACGCGTTGTGAAGTAATGACCGGCTTATCGGTCAGTTTCACCTGAT TTACGTAAAAACCCGCTTCGGCGGGTTTTTGCTTTTGGAGGGGCAGAAAGATGAATGACTGTC CACGACGCTATACCCAAAAGAAAGACACCATCGAATGGTGCAAAACCTTTCGCGGTATGGCAT GATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATG TCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACG TTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACC GCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGG CCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCA GCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATC TTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTG CTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCA TCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCAT TGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTC TGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCG ACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCA CTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCG GGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTT ATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACC GCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGG TGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATT CATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGATAACTTTCAGCAA AAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGGCATGCCCAGGCAT	1)

[1] this sequence was designed in this project

[2] genomic sequence of *E. coli* MG1655 or BL21

A.3 Materials and instruments

A.3.1 Enzymes and reagents

Supplementary Table 5: List of all enzymes and reagents.

name	cat. nr.	supplier
5x Phusion HF buffer	M0530S	New England Biolabs, Ipswich, MA
α-D-glucose monohydrate	6780.2	Roth, Karlsruhe, Germany
agarose LE	840004	Biozyme, Hessisch-Oldendorf, Germany
albumin Fraction V (BSA)	T844.2	Roth
ammonium chloride	K298.2	Roth
ampicillin	A0166	Sigma-Aldrich, St.Louis, MO
BamHI (FastDigest)	FD0054	Thermo Fisher Scientific Inc., Waltham, MA
BglII (FastDigest)	FD0084	Thermo Fisher Scientific
Boc-Lys-OH (BocK)	E-1610.0025	Bachem Holding, Bubendorf, Switzerland
Comassie Blue-250R	3862.2	Roth
copper(II) sulfate pentahydrate	31293-M	Sigma Aldrich
D(+)-biotin	3822.1	Roth
ddH_2O		Fresenius Kabi, Graz, Austria
dNTP's	R0181	Thermo Fisher Scientific
di-potassium hydrogen phosphate	T875.2	Roth
di-sodium hydrogen phosphate (Na_2HPO_4)	T876.2	Roth
EDTA	CN06.1	Roth
ethanol	20821.330	VWR International, Pennsylvania, USA
ethidium bromide	46066	Fluka, St. Louis, MO
GeneRuler 1 kb Plus DNA Ladder	SM1331	Thermo Fisher Scientific
glycerol	3908.3	Roth
glycin	3187.3	Roth
HindIII (FastDigest)	FD0504	Thermo Fisher Scientific
H-L-Lys(EO-N3)-OH (AzK)	HAA2080.0025	Iris Biotech GmbH, Marktredwitz, Germany
hydrochloric acid 32 %	4625.2	Roth
hydrochloric acid, fuming	4625.1	Roth
imidazole	3899.3	Roth
isopropyl β-D-1-thiogalactopyranoside (IPTG)	CN03.3	Roth
KpnI (FastDigest)	FD0524	Thermo Fisher Scientific
L-alanine	3076.1	Roth
L-arabinose	5118.1	Roth
L-arginine	3144.1	Roth
L-asparagine monohydrate	HN23.1	Roth
L-aspartic acid	T202.1	Roth

name	cat. nr.	supplier
L-cysteine	1693.2	Roth
L-glutamine	3772.1	Roth
L-glutamic acid	A3712	BioChemica, Billingham, UK
L-histidine	97062-598	VWR
L-isoleucine	3922.2	Roth
L-leucine	A3460	BioChemica
L-lysin hydrochlorate	9357.3	Roth
L-phenylalanine	A3442	Biochemica
L-proline	T205.2	Roth
L-serine	A1708	BioChemica
L-threonine	T206.2	Roth
L-tryptophane	4858.4	Roth
L-tyrosine	A3437	BioChemica
L-valine	4879.3	Roth
kanamycin sulfate	T832.2	Roth
LB-agar (Lennox)	X65.3	Roth
LB-medium (Lennox)	X964.2	Roth
magensium sulfate heptahydrate	A537.4	Roth
methionine	9359.3	Roth
MES-buffer	NP0002	Thermo Fisher Scientific
milk powder	T145.2	Roth
Ni-NTA agarose	30210	Qiagen, Hilden, Germany
OneTaq Quick-Load 2x Master Mix	M0486S	NEB
PaeI (FastDigest)	FD0604	Thermo Fisher Scientific
PageRuler prestained protein ladder	SM0671	Thermo Fisher Scientific
Phusion® High-Fidelity DNA Polymerase	M0530S	NEB
potassium chloride (KCl)	60128	Fluka
potassium dihydrogen phosphate (KH_2PO_4)	P018.2	Roth
SOB-medium	AE27.1	Roth
sodium ascorbate	A7631	Sigma Aldrich
sodium chloride	3957.4	Roth
sodium dihydrogen phosphate monohydrate ($NaH_2PO_4 * H_2O$)	T879.2	Roth
sodium dodecyl sulfate (SDS)	2326.1	Roth
sodium hydroxide	P031.2	Roth
spectinomycin hydrochloride pentahydrate	ab141968	abcam, Cambridge, UK
SYPRO Orange Gel stain	S6650	Thermo Fisher Scientific

name	cat. nr.	supplier
T4 DNA ligase	EL0011	Thermo Fisher Scientific
T5 exonuclease	162340	Biozyme
Taq ligase	M0208S	NEB
thiamine hydrochloride	T911.1	Roth
tris (3-hydroxypropyltriazolyl methyl)amine (THPTA)	762342	Sigma Aldrich
tris base	4855.3	Roth
triton-X100	3051.3	Roth
urea	3941.1	Roth

A.3.2 Media composition and solutions

Supplementary Table 6: List of all prepared buffers and solutions.

buffer and solutions	composition
1 x PBS	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4
5x M9 salt solution	33.9 g/L Na_2HPO_4, 15 g/L KH_2PO_4, 2.5 g/L NaCl, 5 g/L NH_4Cl
10 mM sodium phosphate (NaPi)	10 mM Na_2HPO_4, 10 mM NaH_2PO_4, pH 8.0
cleaning buffer	50mM NaPi, 150 mM NaCl, 500 mM imidazole, pH 7.2
elution buffer	50mM NaPi, 150 mM NaCl, 300 mM imidazole, pH 7.2
Gibson assembly master mix	for 1.2 mL: 1x ISO reaction buffer, 6.4 U T5 exonuclease, 20 U master mix Phusion® High-Fidelity DNA Polymerase, 6.4 kU Taq DNA ligase
lysis buffer	50mM NaPi, 150 mM NaCl, 10 mM imidazole, pH 7.2
trace element stock	40 g/L $FeSO_4 * 7H_2O$, 10 g/L $AlCl_3 * 6H_2O$, 7.3 g/L $CoCl_2 * 6H_2O$, 2 g/L $ZnSO_4 * 7H_2O$, 2 g/L $Na_2MoO_4 * 2H_2O$; 1 g/L $CuCl_2 * 2H_2O$, 0.5 g/L H_3BO_4, 10 g/L $MnSO_4 * H_2O$, 414 mL/L HCl conc. (37 %, fuming)
wash buffer	50mM NaPi, 150 mM NaCl, 30 mM imidazole, pH 7.2

Supplementary Table 7: List of all prepared media.

medium	composition
SOC-medium	5 g/L yeast extract, 20 g/L tryptone, 0.6 g/L NaCl, 0.2 g/L KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose
LB-agar (Lennox)	10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 20 g/L agar
LB-medium (Lennox)	10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl
M9-medium	1x M9 salt solution, 20 mM glucose, 1 mM MgSO4, 1 µg Ca^{2+}, 0.001x trace elements, 0.1 mg thiamine, 0.1 mg biotin
M9aa-medium	1x M9 salt solution, 20 mM glucose, 1 mM MgSO4, 1 µg Ca^{2+}, 0.001x trace elements, 0.1 mg thiamine, 0.1 mg biotin, 0.05 g amino acids (dissolved in 0.1 M HCl)

A.3.3 Instruments and materials

Supplementary Table 8: List of all used devices and materials in this study.

name	supplier
7500 Real-Time PCR System	Thermo Fisher Scientific
96-well microplates, black	
accuSpin Micro 17R centrifuge	Thermo Fisher Scientific
analytical scale	Sartorius, Göttingen, Germany
Avanti J-20XP centrifuge	Beckmann Coulter, Brea, CA
autoclave VX150	Systex, Wettenberg, Germany
Bradford reagent (Cat.No. 500-0006)	Bio-Rad Laboratories, Hercules, CA
Branson Sonifier 250	Emerson Electric Co., St. Louis, MO
Breathe Easy – Gas permeable sealing membrane for microtiter plates, size: 6x3.25 inches	Diversified Biotech, Dedham, MA
centrifuge 5415R	Eppendorf, Hamburg, Germany
centrifuge 5424	Eppendorf
centrifuge tubes, 50 ml and 500 ml	Thermo Fisher Scientific
CloneJET PCR Cloning Kit	Thermo Fisher Scientific
cryo vials	Thermo Fisher Scientific
culture flasks; 50, 250, 500, 1000 ml	Schott Duran, Wertheim/Main, Germany
desalting membrane	Merck, Darmstadt, Germany
Disposable 10 mL polypropylen columns	Thermo Fisher Scientific
EON plate reader	BioTek, Winooski, VT
G:Box HR	Syngene, Cambridge, UK
gel electrophoresis power supply (Powerease500)	Invitrogen, Carlsbad, CA
Gene Pulser®/Micropulser electroporation cuvettes	Bio-Rad
glas beads	Roth
Incubator HT Multitron II	InforsAG, Bottmingen, Switzerland
JA-10 rotor	Beckmann Coulter
JA-25.50 rotor	Beckmann Coulter
laboratory scale	Binder Rehab AG, Villmergen, Switzerland
laminar flow chamber	CleanAir, Woerden, Netherlands
MicroAmp Optical 96-well reaction plate	Applied Biosystems, Foster City, CA
MicroAmp Optical Adhesive Film	Applied Biosystems
micropulser	Bio-Rad
Nanodrop	Thermo Fisher Scientific

Supplementary Table 9: List of all used devices and materials in this study.

name	supplier
Ni-NTA resing	Qiagen
NuPAGE 4-12 % Bis-Tris Gel	Invitrogen
NuPAGE MES SDS Running Puffer	Invitrogen
OneTaq Quick-Load 2x Master Mix with Standard buffer	New England Biolabs
PCR machines (GeneAmp® PCR System 2720)	Thermo Fisher Scientific
PCR tubes	Greiner Bio-One International AG, Kremsmünster, Austria
petri dishes	Greiner
Photometer	Eppendorf
Pierce™ Disposable Columns, 10 mL	Thermo Fisher Scientific
pipettes; 2, 20, 200, 1000 µl	Deville, South Plainfiled, NJ
pipette tips	Greiner
PureYield™ Plasmid Miniprep System	Promega, Madison, WI
reaction tubes, 1.5 mL and 2 mL	Eppendorf
reaction tubes, 50 mL and 500 mL	Thermo Fisher Scientific
sterile inoculation loops	Roth
sterile syringe filters 0.22 µM	Roth
syringe 5, 10 mL	Braun, Kronberg im Taunus, Germany
SuperSignal West Dura Extended Duration Substrate	Thermo Fisher Scientific
Vivaspin 3000 MWCO tubes	GE Healthcare, Little Chalfont, UK
vortex	IKA-Werke GmbH & Co. KG, Staufen, Germany
Wizard SV gel and PCR clean-up system	Promega
Zeba Spin Desalting Columns 7K MWCO	Thermo Fisher Scientific