Department Chemie

Lehrstuhl für Biochemie

Establishing a genetic code expansion system in *Bacillus subtilis*

Christopher Markus Scheidler

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat) genehmigten Dissertation.

Vorsitzende(r): Univ.-Prof. Dr. Matthias J. Feige

Prüfer der Dissertation:

1. TUM Junior Fellow Dr. Sabine Schneider
2. Univ.-Prof. Dr. Kathrin Lang

Die Dissertation wurde am 11.03.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 20.04.2020 angenommen.

Für Papa

Zusammenfassung

Methoden zur Erweiterung des genetischen Codes über die 20 kanonischen Aminosäuren hinaus sind sowohl in der Grundlagenforschung als auch für medizinische Anwendung von besonderem Interesse. Seit der Einführung des Amber Suppression Systems im Jahr 2000, mit welchem unnatürliche Aminosäuren mit neuen chemischen Funktionalisierungen positionsspezifisch in einem Zielprotein eingebaut werden können, wurde das System in fast allen Modellorganismen, bis hinauf zu Säugetieren etabliert – jedoch nicht im biotechnologisch höchst relevanten Bakterium *Bacillus subtilis*. Der positionsspezifische Einbau ermöglicht zusätzlich eine maßgeschneiderte Veränderung des Zielproteins für Anwendungen sowohl in der Grundlagen- als auch in der Arzneimittelforschung. Dieses grampositive Bakterium zeichnet sich durch eine Reihe von Vorteilen wie dem „generally recognized as safe“ Status der Food and Drug Administration (FDA), der vielfältigen genomischen Manipulierbarkeit und dem leistungsfähigen Sekretionssystem gegenüber anderen bakteriellen Expressionssystemen aus und wird vor allem zur Herstellung von Enzymen für die Industrie genutzt. Im Rahmen dieser Dissertation wurde die Amber Suppressionsmethode zum ersten Mal in *B. subtilis* etabliert. Um die verschiedenen benötigten Bestandteile dieses Systems in den richtigen Mengen und Verhältnissen zu erzeugen wurden die Expressionsplasmide zunächst auf Flexibilität und Gesamtausbeute hin optimiert. Im ersten Schritt wurde gezeigt, dass die Amber Suppressionsmethode mit Einbauraten von knapp 30% durchaus konkurrenzfähig zu bestehenden Expressionssystemen mit ähnliche Einbauraten ist, und aufgrund der genannten Vorteile von *B. subtilis* eine echte Alternative zu diesen bietet. Der größte Vorteil von *B. subtilis*, das extrem leistungsfähige SEC-Sekretionssystem, sollte im nächsten Schritt genutzt werden um die pharmazeutisch relevante Proteinklasse der Antikörperfragmente zu produzieren und in den Überstand zu sekretieren. Hierfür wurde das SEC-System genutzt und mittels CRISPR/Cas9 verbessert, um auch in Anwesenheit der unnatürlichen Aminosäure voll funktionsfähig zu sein. Durch die weitere Co-Expression eines extrazellulären Chaperons konnte das Wildtyp Fragment der leichten Kette des Modellantikörpers MAK33 mit Ausbeuten von ca. 20 mg l^{-1} sowie das mittels der Amber Suppressionsmethode modifizierte Fragment mit Ausbeuten von ca. 2 mg l^{-1} erfolgreich produziert werden. Zur Verifizierung der korrekten Faltung und Funktionalität wurde das bioorthogonal modifizierte Antikörperfragment aus dem Überstand der Expressionskultur gewonnen und die unnatürliche Aminosäure mit einem fluoreszenzmarkierten Tetrazinliganden markiert. Mittels zweier angepasster „Enzyme-linked Immunosorbent Assays“ konnte die erfolgreiche Bindung, welche nur bei korrekter Faltung des Fragments möglich ist, an das Antigen schlussendlich nachgewiesen werden. Dieses neu etablierte System zur Herstellung bioorthogonal modifizierter Proteine zeichnet sich zusammenfassend durch eine hohe Flexibilität und Ausbeute bei gleichzeitiger Reduktion der benötigten Arbeitsschritte zum Erhalt des Zielproteins aus.

Abstract

Methods to expand the genetic code beyond the 20 canonical amino acids are of great interest for basic research as well as medical applications. The establishment of the amber suppression method in 2000, enabled the site-specific incorporation of unnatural amino acids with novel chemical functionalities in a target protein. The site-specific incorporation enables tailored manipulation of the target protein for further studies. Throughout the years, this method was introduced in nearly all model organisms, including mammals, but not in the biotechnological highly relevant bacterium *Bacillus subtilis*. This gram-positive bacterium is characterized by several advantages over other bacterial expression systems like the "generally recognized as safe" status granted by the Food and Drug Administration (FDA), the diverse available tools for genetic manipulations and the efficient secretions system. However, its use is mainly for the large-scale production of industrial enzymes. In this thesis, the amber suppression method was established in *B. subtilis* for the first time. To fine-tune the individual components of this method, the expression plasmids were thoroughly optimized for the overall flexibility and yields. In the first step, incorporation rates of about 30% could be achieved, already illustrating that this bacterium is an alternative to existing expression systems. The most significant advantage of *B. subtilis* is its extremely productive secretion system. This system is used to produce and secrete the pharmaceutical relevant protein class of antibody fragments into the supernatant. Therefore, the SEC system was used and modified using CRISPR/Cas9 to resist the influence of the unnatural amino acid on the SEC system. By co-expression of an extracellular chaperon, the variable light chain of the model antibody MAK33 could be secreted with yields of about 20 mg l^{-1} whereas the bioorthogonally modified antibody fragment showed yields of approximately 2 mg l^{-1}. To verify the correct folding and functionality, the modified antibody fragment was purified from the supernatant and linked with a fluorescently labelled tetrazine ligand. Using a modified "Enzyme-linked Immunosorbent Assay", successful binding of the antibody fragment to its antigen could be shown, which only occurs if the protein is folded correctly. In conclusion, this new established system to produce and secrete bioorthogonal-modified proteins is characterized by its high flexibility and overall yields while reducing the required steps to receive the target protein.

Inhaltsverzeichnis

Titelpage ..... a

Zusammenfassung 1

Abstract 2

1. Introduction 7

1.1. History of *Bacillus subtilis* 7

1.2. Applications and advantages of *B. subtilis* 8

1.2.1. Secretion systems of *B. subtilis* 9

1.2.2. Industrial use and strain engineering of *B. subtilis* 12

1.3. Protein engineering and genetic code expansion 14

1.4. Objective of this work 21

2. Results and discussion 22

2.1. Optimizing the available expression system 22

2.1.1. Replicative plasmid increases the protein expression 23

2.1.2. Stabilization of mRNA through secondary structure elements 25

2.1.3. Evaluation of *B. subtilis* promoters 26

2.1.4. Cultivation dependent promoter activity 28

2.1.5. Promoter-inducibility using isopropyl-β-D-thiogalactopyranosid 30

2.1.6. Summary 32

2.2. Establishing amber suppression in *B. subtilis* 33

2.2.1. General considerations and expression setup 33

2.2.2. Expression of the amino acyl tRNA synthetase $PylS^{Norb}$ 35

2.2.3. Expression of the $tRNA^{CUA}$ *pylT* 36

2.2.4. Expression of GFP^{Norb26} and verification using mass-spectrometry 38

2.2.4.1. Fluorescence experiments 38

2.2.4.2. Purification and quantification of GFP^{Norb26} 39

2.2.4.3. Verification using mass-spectrometry 40

2.2.5. Summary 41

2.3. Protein secretion in *B. subtilis* 42

2.3.1. Expression setup for the secretion of proteins 42

2.3.2. Expression and secretion of MAK33-VL 42

2.3.2.1. MAK33-VL expression tests 44

2.3.2.2. Verification of MAK33-VL expression 44

2.3.2.2.1. Western blotting for MAK33-VL-StrepII 44

2.3.2.2.2. Verification using mass spectrometry 45

2.3.3. Secretion of D1.3scFv 46

2.3.4. Secretion of Am3-114 ... 50
2.3.5. Summary ... 53
2.4. GFP fusion constructs to assess ncAA incorporation ... 54
2.4.1. State of research on the genomic context of target sites ... 54
2.4.2. C-terminal GFP fusion constructs and analysis of ncAA incorporation at different target sites ... 55
2.4.3. Expression of MAK33-VL^{Norb33} in K07^{Amber} ... 56
2.5. Engineering the SEC system to reconstitute the secretion ... 59
2.5.1. Stop codon exchange via CRISPR/Cas9 ... 60
2.5.2. Single-base exchange of *secA* and *secG* ... 62
2.5.3. Restoration of SecY expression ... 63
2.5.4. Summary ... 65
2.6. Expression and secretion of bioorthogonal modified proteins ... 66
2.6.1. Expression and purification of MAK33-VL^{Norb33} ... 66
2.6.2. Mass spectrometry analysis and click modification ... 68
2.6.3. Expression of D1.3scFv^{Norb74} ... 69
2.6.4. Summary ... 71
2.7. Functional characterization of MAK33-VL^{Norb33} ... 72
2.7.1. Summary ... 73
3. Conclusion and outlook ... 74
4. Material and methods ... 75
4.1. Materials ... 75
4.1.1. Devices ... 75
4.1.2. Chemicals ... 76
4.1.3. Cultivation media ... 76
4.1.4. Enzymes, standards and kits ... 78
4.1.4.1. DNA-modifying enzymes ... 78
4.1.4.2. DNA/protein markers ... 79
4.1.4.3. Kit systems ... 79
4.1.5. Oligonucleotides ... 79
4.1.6. Plasmids ... 83
4.1.7. Bacterial strains ... 84
4.1.8. Buffers ... 87
4.2. Software, databases and tools ... 89
4.3. Molecular biology methods ... 90
4.3.1. *E. coli* cultivation and transformation ... 90
4.3.2. *B. subtilis* cultivation and transformation ... 90

4.3.3. Agarose gel electrophoresis ... 91
4.3.4. Polymerase chain reactions ... 91
4.3.4.1. Standard PCR ... 91
4.3.4.2. Colony PCR ... 91
4.3.4.3. QuikChange PCR ... 91
4.3.4.4. Overlapping PCR ... 91
4.3.5. Cloning strategies ... 92
4.3.5.1. Promoter screening ... 92
4.3.5.2. mRNA stabilization ... 92
4.3.5.3. C-terminal MAK33-VL-GFP-fusion constructs ... 93
4.3.5.4. Strain engineering by CRISPR/Cas9 ... 93
4.3.5.5. Generation of the expression plasmid ... 94
4.3.5.6. Amber suppression plasmid ... 94
4.3.5.7. Secretion plasmid ... 95
4.4. Protein chemical methods ... 95
4.4.1. SDS-polyacrylamide gel electrophoresis (SDS-PAGE) ... 95
4.4.2. Protein expression and purification ... 95
4.4.2.1. Intracellular protein expression ... 95
4.4.2.2. Extracellular protein expression ... 96
4.4.2.2.1. Expression of MAK33-VL ... 96
4.4.2.2.2. Expression of D1.3scFv ... 96
4.4.2.2.3. Expression of Am3-114 ... 96
4.4.2.3. Protein purification ... 96
4.4.2.4. Western blotting ... 97
4.5. Fluorescence measurements ... 97
4.6. Verification of functionality ... 98
4.7. Mass spectrometry ... 98
Bibliography ... 99
Abbreviations ... 109
Appendix ... 111
List of tables ... 116
List of figures ... 117
List of publications ... 120
Acknowledgments ... 121
Declaration ... 122

1. Introduction

Research in every domain of life relies on establishing model organisms to bundle research efforts and improve exchange in the scientific community. A model organism should therefore be cheap, easy to use and non-pathogenic. Throughout the last decades representative model organisms were specified step-by-step for each kingdom of life. One of the kingdoms, the bacteria, can separated through different classification schemes, with the Gram staining, already developed by the Danish microbiologist Hans Christian Gram 1883 is still the most used method to classify bacteria by the properties of the cell wall.(1) Especially easy and reproducible cultivation conditions were a central argument for early microbiologists whereby *Escherichia coli* became the model organism for gram-negative bacteria and *Bacillus subtilis* for gram-positive bacteria.

1.1. History of *Bacillus subtilis*

B. subtilis is one of the oldest known microorganisms. It was discovered as "*Vibrio subtilis*" by Christian Gottfried Ehrenberg in 1835 and renamed by Ferdinand Julius Cohn in 1872 to "*Bacillus subtilis"*.(2, 3) However, as the storage of bacterial strains like in today's strain databases was not possible at that time, the *B. subtilis* strains as they are known today are most likely genetically different from the one described by Cohn in 1872.(2)

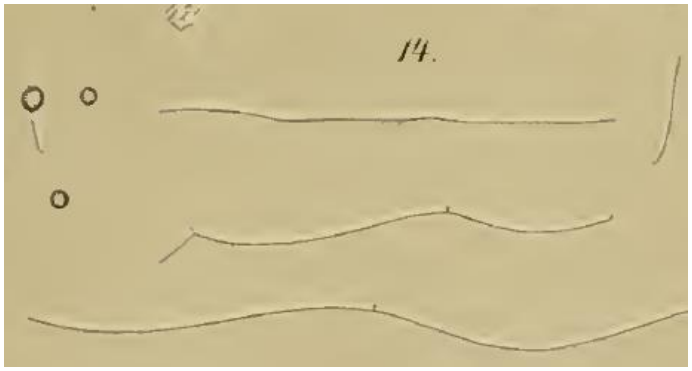

Figure 1: Historic drawing of *B. subtilis*, partially in the dividing state by Cohn from 1872(2)

Even modern research since the 1930s with the use of safe storage and reproducible culture conditions has not been able to protect *B. subtilis* from the mixing of different sibling strains.(4, 5) To analyze early history and relationships in the *B. subtilis* legacy strains, Zeigler et al. partially reconstructed a family tree (Figure 2). The central strain which the modern microbiology of *B. subtilis* since the 1940s is based upon the strain 168 which itself is a domesticated, so adapted to the laboratory conditions, version of the Marburg strain. This strain, together with its sibling strains "23", "120", "160" "166", were created by irradiation of the unknown parent strain using X-ray or UV by Burkholder & Giles in 1947.(6) This group at Yale University abandoned its *Bacillus* research and therefore, only some of the strains created could be recovered.

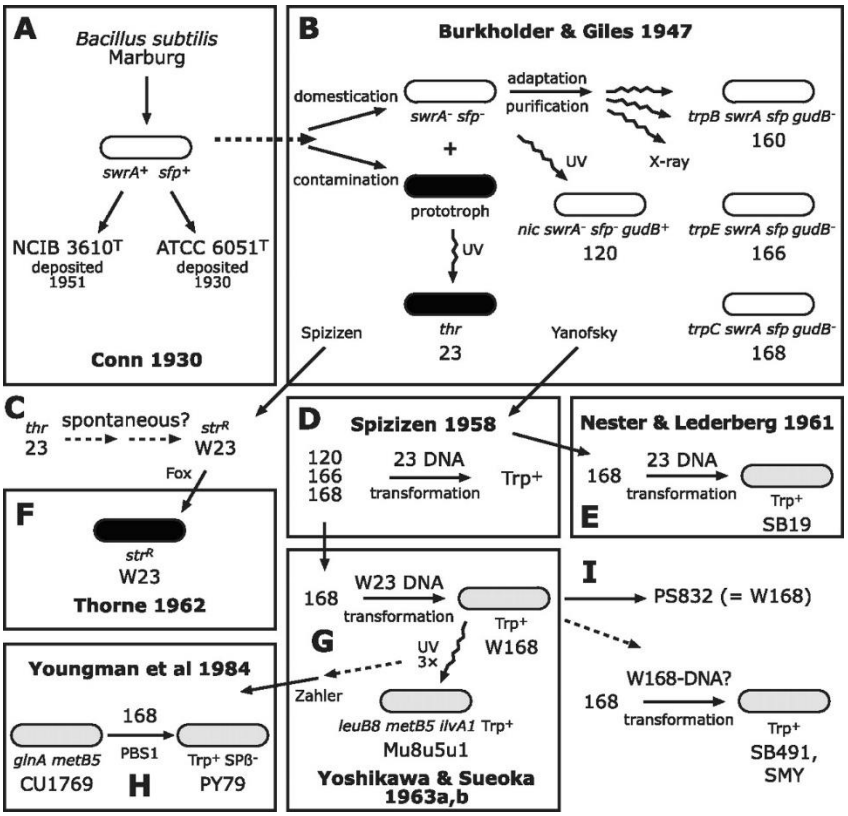

Figure 2: Model describing the history and relationships between the early *B. subtilis* strains
Each box represents a publication in which the illustrated modifications and genetic manipulations were conducted to the *B. subtilis* strains. Figure taken from (5).

Since then, *B. subtilis* has been one of the mainly used organism in modern microbiology due to several advantages. In contrast to other microorganisms, *B. subtilis* is naturally competent and can take up genetic material by its own. The natural competence is a direct consequence of its habitat in the soil as it is beneficial to take up foreign DNA and gain a selective advantage when competing with billions of other microorganisms. The regulation of competence is complex and, can be induced by growth in a glucose minimal salts-based medium, whereby a low percentage of the *B. subtilis* cells become competent, taking up DNA actively and integrate it into their genome.(7) The natural competence enabled researchers to study single genes, create first targeted gene knockouts and draw first genomic maps before sequencing methods became available in the 1970s.

1.2. Applications and advantages of *B. subtilis*

In parallel to the use of *B. subtilis* in fundamental research, it became one of the main organisms to produce natural products in the biotechnological industry. Here, the GRAS (generally recognized as safe) status granted by the Food and Drug Administration (FDA) is highly important.(8) This status marks the host as save for the production, drastically reducing purification costs for the production of natural products, as *B. subtilis* does not contain any endotoxins. In contrast, when using gram-negative bacteria like *E. coli*, toxic lipopolysaccharides (LPS) are a big problem as concentrations in the µg kg^{-1}

range can already induce a septic shock in humans.(9) Besides to the absence of LPS, several benefits like its modest nutrition requirements, but especially the sophisticated and powerful secretion system, pathed its way into the industry.

1.2.1. Secretion systems of *B. subtilis*

The secretion of enzymes and proteins into its surroundings is far spread among microorganisms and addresses various tasks from nutrient acquisition over cell-cell communication to defence mechanisms. Throughout evolution several different bacterial secretion systems were developed by gram-positive) and gram-negative bacteria.(10, 11) The biggest difference between gram-positive and gram-negative secretion systems is its structural complexity. Gram-negative bacteria possess very sophisticated and specialized systems as multiple membranes have to be crossed, which however can reduce the amount of secreted product or can be hard to target to secrete heterologous proteins for industrial applications.(10-14) In contrast, the secretion systems of gram-positive bacteria, like *B. subtilis*, are much simpler as only one membrane has to be crossed but due to their natural habitat in the soil nutrients from various sources like (poly-) saccharides, lipids and proteins have to be broken down bevor nutrient uptake can take place.

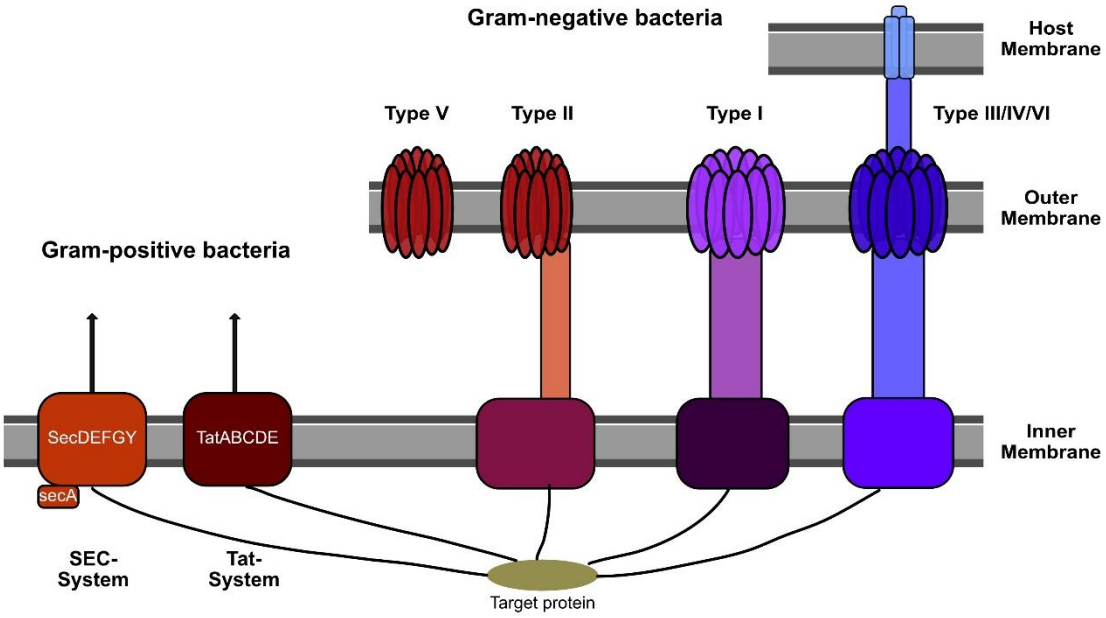

Figure 3: Overview of several bacterial secretion systems
Gram-positive bacteria possess relatively simple secretion systems, as only one membrane has to be crossed, like the "secretory protein translocation" (SEC) or the "twin-arginine translocation" (Tat) system. Gram-negative bacteria, additionally to the secretion through the SEC and Tat system, feature very complex systems, crossing at least two membranes to secrete into its surroundings. Type I and Type II systems consist of two transmembrane domains, each spanning over two membranes and enable the secretion of proteins into the supernatant. Type III/IV/VI are secretion systems, which are involved in in the secretion of toxins in pathogenic bacteria like *Yersinia, Salmonella, Vibrio, Legionella* or *Pseudomonas*. These systems secrete the toxins through their inner and outer membrane and can even insert them directly into the target organism. Type V systems are autotransporter system which form a barrel structure in the membrane by themselves and are involved in cell-adhesion processes.

The wild type *B. subtilis* secretes not less than 113 different proteins into its surroundings, not only for the acquisition of nutrients, but also for intercellular communication and chemotaxis using at least four different pathways, namely the "secretory protein translocation" (SEC) pathway, the "twin-arginine translocation" (Tat) pathway, ABC transporters and the pseudopilin pathway.(15) The pseudopilin pathway is important for the DNA uptake during competence and secretes required proteins for forming the dsDNA binding pseudopilus.(16) ABC transporters are on the hand used for the uptake of nutrients or ions and on the other hand for the export/efflux of antibiotics or small peptides.(17) However, the secretion of larger proteins is performed by the other two pathways the SEC pathway, which secretes most of the proteins into the supernatant and the Tat pathway. have been studied extensively throughout the years and are well characterized.(11-14, 18-20)

Figure 4-A, shows the SEC system, consisting of the SecYEG transmembrane channel, the ATPase SecA and the accessory protein SecDF. Targeting SecA and the SecYEG channel is performed by an N-terminal signal peptide. After the signal peptide is produced by the ribosome it is directly recognised by the signal recognition particle FtsY. The nascent peptide chain is directed to the SecYEG membrane channel, where it is actively transported through the membrane by the ATP driven motor-protein SecA. After the secretion, the signal peptide (SP) is cleaved off by membrane-bound signal peptidases (SPases) and the protein folds extracellularly with help of chaperones like PrsA.(14, 21, 22)

The mechanism of the Tat system shown in Figure 4-B is however completely different. Instead of an unfolded peptide chain, the target protein carrying an N-terminal signal peptide, is translocated through the membrane already in its folded state. In the first step shown in Figure 4-B on the left, the signal peptide of the target protein is recognized by TatB, being associated to the docking complex consisting of TatC and TatB. If this association is successful, the TatB/C-cargo protein complex can interact with the pore-forming complex built by TatA and the protein is translocated through the membrane.(12, 13)

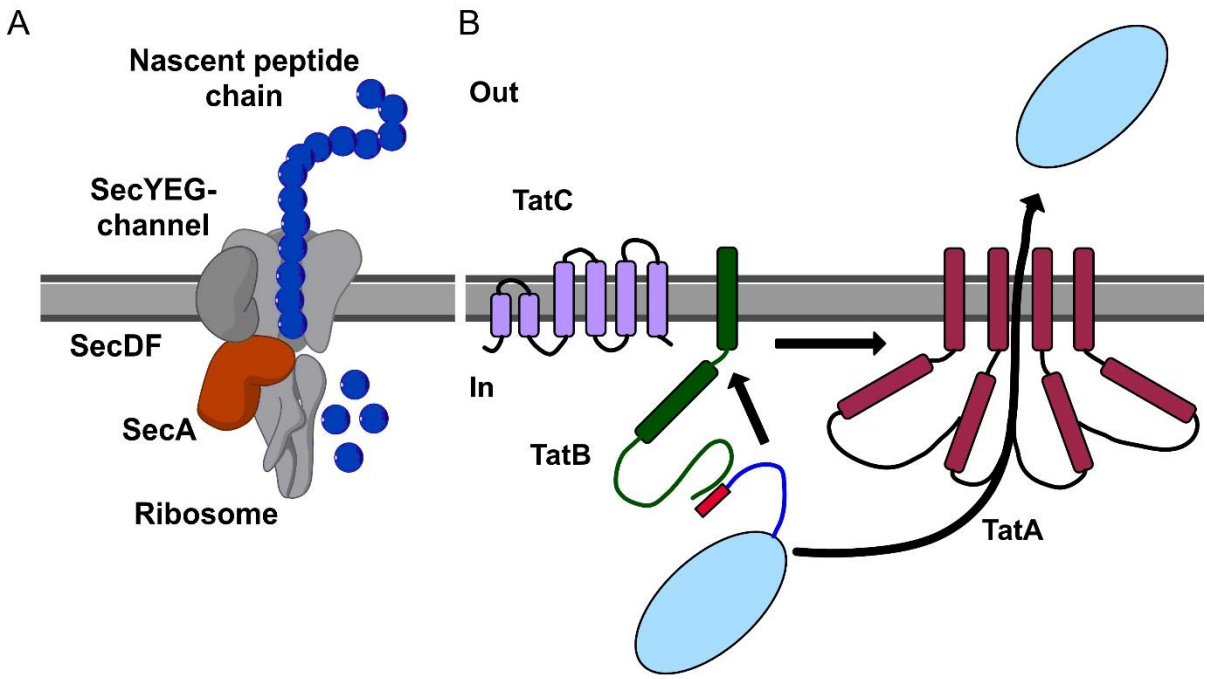

Figure 4: Scheme of the SEC and Tat system of *B. subtilis*
A: A nascent peptide chain is translocated through the SecYEG channel into the extracellular space. Translocation occurs via ATP-hydrolysis by SecA on which the ribosome has docked on. The protein is synthesized and directly transported through the membrane B: Tat system of *B. subtilis*: TatB recognizes the target protein carrying an N-terminal signal peptide. The pore-forming complex of TatA subsequently translocates the protein in its folded state.

Although both described secretion systems of *B. subtilis* greatly differ in their mechanisms, signal peptides play an important role and show some common properties. The signal peptide itself consists of three regions: A positively charged region at the N-terminus, a hydrophobic core region and a cleavage site for signal peptidases. However, a mode of action, which could predict a suitable signal peptide for a target protein to be secreted, could not be established yet, making the search for a signal peptide a bottleneck, which has to be addressed by screenings.

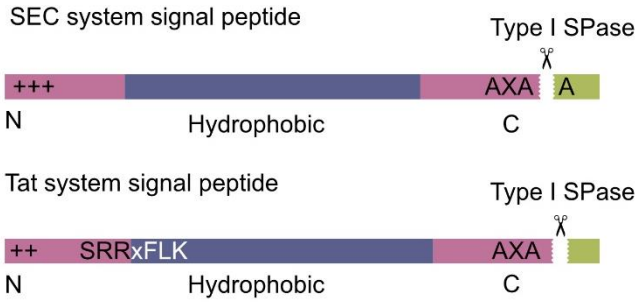

Figure 5: Structure of signal peptides targeting the SEC and the Tat system
Both signal peptides contain a positive charged head group followed by a hydrophobic core and a signal peptide cleavage site (AXA). The Tat signal peptide additionally contains a characteristic twin-arginine motif. The signal peptide is cleaved off by signal peptidases (SPases).

Especially the SEC system exhibits a vast potential for the secretion of (heterologous) proteins due to its simple structure and easy addressability is therefore of great importance in research and industry.

1.2.2. Industrial use and strain engineering of *B. subtilis*

The main use of *B. subtilis* in the industry is for the high-yield production of industrial enzymes. Through excessive strain engineering and selection, a broad range of enzymes classes like proteases, lipases or glycosidases are produced in the g l^{-1} scale. Just for laundry detergents, hundreds of tons of pure enzymes are produced each year by *B. subtilis*.(23, 24) Examples are here α-Amylase (up to 3 g l^{-1}), Proinsulin (1 g l^{-1}), Lipase A (0.6 g l^{-1}) or Aminopeptidase (1.7 g l^{-1}). (23, 25) A further accelerator is the increasing use of enzymes as an alternative to classical organic chemistry, which in turn favours the use of organisms like *B. subtilis* as expression hosts.(24) On top of that, the spotlight on sustainability has driven even more work towards enzymes as an alternative catalyst source, of course requiring a cheap and high yielding expression host like *B. subtilis*. (26)

Although *B. subtilis* has been excessively used in the industry for decades, no commercial expression strains as known for *E. coli* like "BL21" or "Rosetta (DE3)" are available yet. Moreover, the existing "engineered" strains are closely related to the wild type strains as they were only selected to produce a homologous enzyme. However, in recent years efforts for a more structured strain engineering approach were made.(27)

Here, the creation of the Δ6 strain by Westers in 2003, a strain where six large dispensable regions, mainly phage islands, were deleted from the genome marked the first systematic approach towards an engineered *B. subtilis* expression strain.(28) These deletions had no effect on the growth performance, motility or energy metabolism – however, no huge improvements towards a microbial cell factory were achieved either.(28) In parallel, *B. subtilis* strains deficient of several extracellular proteases were created, which showed improved overall secretion yields due to increased proteins stability in the supernatant.(23, 29) Nevertheless, the genetic background of the strains is still very similar to the wild type, which makes the strain very flexible and adaptive to new conditions and energy sources. Nonetheless, this is not required for the cultivation in the laboratory or an industrial setting, where defined parameters are crucial in minimizing unwanted effects from silenced genes or disruptive gene clusters. An example for a disruptive by-product is the pulcherrimin pathway. Pulcherrimin is a cyclic di-peptide consisting of two leucines which is important for iron uptake.(30, 31) In *B. subtilis* wild type strains, this dipeptide is produced and secreted into its surroundings to scavenge iron from possible competitors. However, when purifying proteins from the supernatant using a His_6-tag in the laboratory, this cyclic dipeptide also binds to the Ni^{2+}-NTA resin and compromises the purification. Although the *B. subtilis* genome was already sequenced in 1997, unwanted artefacts like the example of pulcherrimin underline the need to fully understand the genomic repertoire of *B. subtilis* as many genes and their products are still unknown.(32) Here, it is very desirable to gain

more insights into the core genome and learn about all essential/non-essential genes for basic research as well as industrial and pharmaceutical applications.

To further elucidate this, the "Minibacillus" project was started in 2013 aiming to reduce the *B. subtilis* genome and create a strain which is equipped with the minimal core genome.(27) Minimizing a bacterial genome can target two important points: a) The identification of the minimal set of genes to sustain life, and b) the design of a controllable microbial cell factory for sustainable heterologous protein expression.(33) In a first step of the "Minibacillus" project, each gene of the entire *B. subtilis* genome was categorized into the following six categories: essential, reasonably important, dispensable, (functional-) paralogues or competence. Based on this categorization 642 genes were marked as necessary for the design of the "Minibacillus" genome and a strategy to stepwise delete the unnecessary genes was designed. By using marker-free deletion plasmids, no residual components remain after each deletion step, but the plasmids have to be applied in a consecutive fashion. Several intermediates during this project were analysed in detail and are shown in Figure 6. The "Minibacillus" project goes hand in hand with "Subtiwiki", a database where all genes, their interaction partners, pathways and regulatory network, identified in *B. subtilis,* are collected and accessible together with the relevant literature. Since 2009, this database has become the most convenient way to receive information about certain genes or phenotypes.(34-36) Through several years of work in the "Minibacillus" project, about 1.5 mega bases or approximately 1500 genes were deleted from the *B. subtilis* 168 genome. As a drawback, the generation time slightly increased and the genetic competence decreased, which however is crucial for the strain engineering process. This could be overcome by introducing an inducible competence cassette. In the status of the project, two strains PS38 and PG10 were created and evaluated more closely for their behaviour and properties.(37)

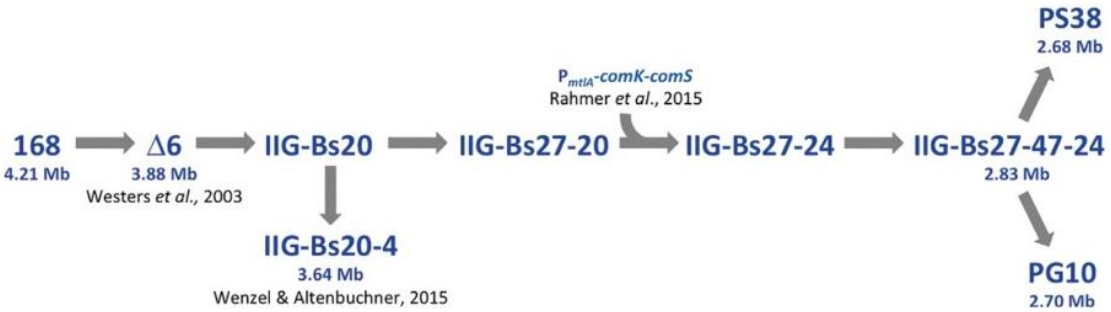

Figure 6: Intermediates from the creation of PG10 and PS38 from the *B. subtilis* wild type 168
The described strains were evaluated more closely. Figure taken from Reuss et al. 2017.(37)

In these two strains, changes in the overall protein expression profile as well as the energy metabolism, like for example a drastic upregulation of the arginine/ornithine metabolism, were found when compared to the wild type strain.(37) Furthermore, several new targets for further deletion steps could be identified. As a side effect, the strain PG10 shows drastically increased secretion yields and improved secretion of proteins, even of such, which could not be produced before.(38) This new, genome reduced strain is a huge step to promote *B. subtilis* as a cheap and effective expression organism with several advantages compared to other expression hosts like *E. coli,* which are at the moment most often used for the expression of heterologous proteins. These proteins, especially small, pharmaceutically relevant proteins/peptides, are also excellent targets for protein engineering to increase their stability or specificity for their use in medical applications or in research.

1.3. Protein engineering and genetic code expansion

The term protein engineering describes the enhancement, modification and optimization of proteins for pharmaceutical tasks or their use in research. The broad range of methods which were developed since 1970s, pathed the way to not only modify proteins but also to study their function in great detail.(39) Interesting parameters are in general the optimization of kinetic properties, like turnover numbers or the affinity to a certain target. In addition, the stability, solubility and protein half-life are very relevant for industrial or medical applications. One of the central methods is hereby the site-directed mutagenesis. This important method was established by 1978 by Smith and colleagues, leading to the Nobel Prize in 1993.(40) Using this technology, vast improvements in protein properties and in understanding their functions, could be achieved. Other methods to alter the overall properties of the target protein include taking influence on the solubility, charge or plasma half-life through for example PEGylation (conjugation with polyethylene glycol PEG), PASylation (P = proline, A = alanine and S = serine) or protein fusions. (41-44) For protein fusions, the target protein is genetically fused to a protein with a long plasma half-life like transferrin or albumin, increasing the molecular weight and the plasma half-life.(45, 46) Alternatively to the fusion to a full-length protein, PASylation can be performed.(44) With this method, the target protein is genetically fused to a polypeptide chain consisting of proline, alanine and serine, increasing the size of the protein and therefore also the plasma half-life.(44) PEGylation differs from the first two methods. Here, PEG chains, which are equipped with a reactive head group, are mixed with the expressed protein. Depending on the used head group, several functional groups of the protein, like amines, histidine-imidazoles, cysteins, hydroxyl or carboxylic groups can be targeted.(42, 43) The broad availability of most functional groups leads to a mixture of protein-PEG-conjugates, exhibiting a longer plasma half-life and increased protein stability. However, a mixture of protein-PEG-conjugates is not very favorable for the production of pharmaceuticals.

A better alternative for site-specific labelling is the thiol group of cysteine. Cysteine is the amino acid with the lowest abundance in proteins, minimizing the variety of possible side products if it is used for a site-specific reaction. The thiol group can be targeted by various reactive reagents, but in the last years, maleimide-based reagents are most often utilized, for example for the creation of antibody-drug conjugates (ADC). (47, 48) Since the first ADC for medical use was developed in 2001, six further ADCs have been approved with several others in the pipeline. ADCs are created by coupling a linker and a drug to cysteins of the antibody, resulting in a covalent link. This combination shows good specificities with reduced side effects when compared to existing therapies of for example small molecules.(49, 50) As a drawback, these ADCs are costly in the overall development process. In an initial step, a functional antibody with high target affinity has to be found, requiring exhausting B-cell selection in a living animal and subsequent screenings for the best antibody. After the successful evaluation, the expression can be carried out in mammalian cell lines, which requires substantial higher monetary input compared to bacterial expression systems.(51) The purified antibody is coupled with the drug in the next step forming the final ADC. Although the used maleimides are reliable tools, there can be stability issues in the environment in the plasma for medical applications. The ADC can be attacked by cysteins from other plasma proteins or glutathione, leading to a loss of the coupled drug throughout time. Therefore, it would be desirable to uncouple the labelling/linking reactions from the existing reactive groups of a protein via new functionalities, increasing the overall stability, specificity and reactivity for the coupling. However, as proteins are limited to the 20 canonical amino acids a new method had to be developed to expand the functional groups in proteins beyond the preexisting. This can be achieved by using the amber suppression system.

The amber suppression system is based on recoding of the amber stop codon to a “coding codon”. In all domains of life, 61 of the 64 codons code for amino acids while three codons define the chain termination of the produced peptide/protein chain from the ribosome. The amber stop codon is the least used codon to terminate the production of proteins. But already in the 1960s, it was recognized that in certain *E. coli* strains, the amber stop codon can be suppressed, allowing the incorporation of an amino acid into the peptide chain instead of a chain termination .(52) Today it is known, that amber, or in general, stop codon suppression exists in all domains of life and is especially a common tool for viruses to produce their required proteins for assembly.(53) With this basis, the Schultz group established the amber suppression system as a method to expand the genetic code beyond the 20 canonical amino acids. This opened up a completely new research field with nearly unlimited novel, bioorthogonal amino acids carrying various new chemical functionalities not found in organisms.(54, 55) In their first work in 2000, the tyrosine amino acyl tRNA synthetase(aaRS)/tRNA pair from *Methanocaldococcus jannaschii* was used.(55) This archaeal aaRS/tRNA pair is bioorthogonal to the used host organism *E. coli,* loading its amino acid specifically onto its corresponding tRNA.

Bioorthogonal in general describes the term that something, in this case an aaRS/tRNA pair, which is introduced into an organism, does not interfere with any biological processes of the host.(56) In the work by Schultz, the tRNA was mutated to contain CUA in its anticodon loop, now recognizing the amber stop codon, and was introduced together with its corresponding aaRS. Thus, a tRNA, loaded with an amino acid, recognizing the amber stop codon as a coding codon is now present, next to the 61 other endogenous tRNAs. The loaded $tRNA^{CUA}$ identifies the amber stop codon and is compatible with the *E. coli* ribosome, but is not subjected to *E. coli* proofreading enzymes, which would possibly edit the loaded $tRNA^{CUA}$ if the tRNA is not bioorthogonal in the organism. By inserting the amber stop codon in the DNA sequence of the target protein, the codon is now recognized during the translation by the $tRNA^{CUA}$ and the amino acid loaded on the tRNA is incorporated into the target protein instead of a chain termination mediated by the Peptide chain release factor 1 (RF1). The final product, the target protein, now carries a non-canonical amino acid (ncAA) in its sequence.

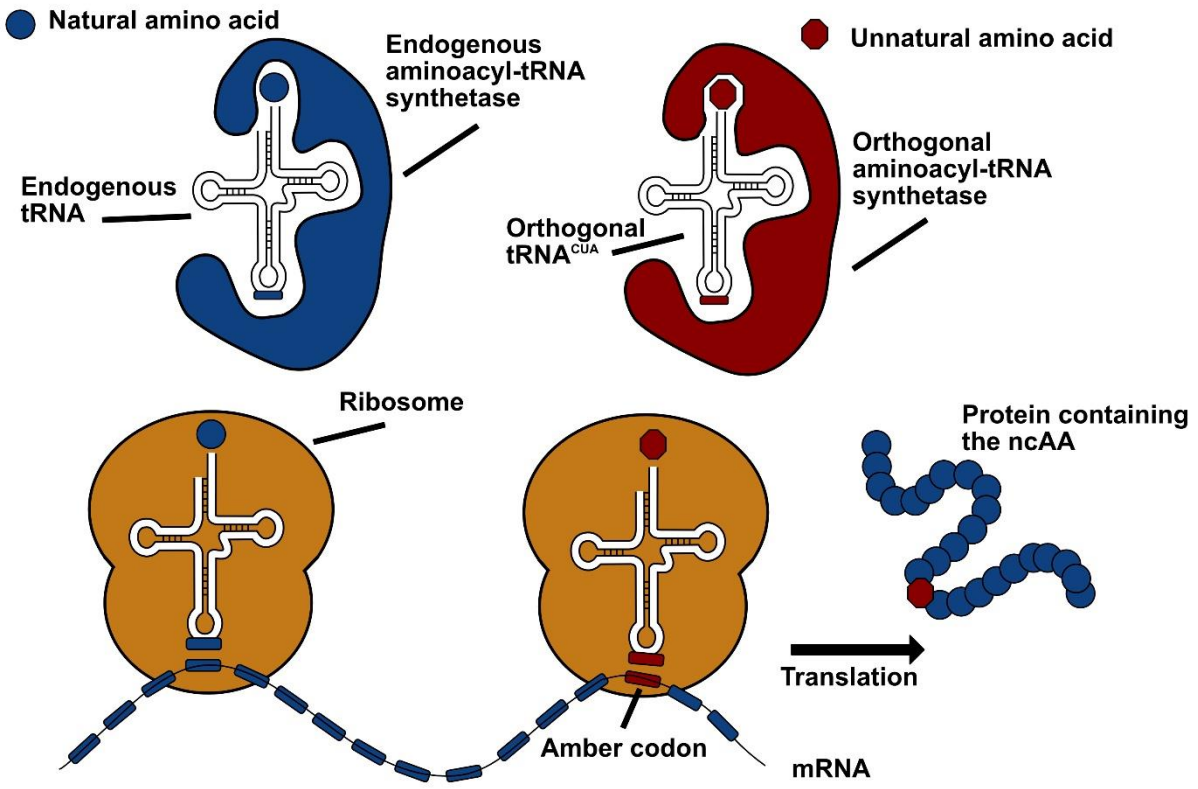

Figure 7: Genetic code expansion by amber codon suppression
A bioorthogonal aaRS/$tRNA^{CUA}$ pair is introduced to the target organism. The bioorthogonal aaRS charges the unnatural amino acid onto the corresponding $tRNA^{CUA}$. The amber stop codon shown in red in the mRNA sequence is recognized by the charged $tRNA^{CUA}$ and the unnatural amino acid is introduced into the target protein instead of a chain termination, yielding the full-length protein.

With the discovery of pyrrolysine, the 22nd proteinogenic amino acid in 2002, the amino acid repertoire was expanded further.(57) This amino acid is found in the active site of the monomethylamine

methyltransferase playing a crucial role in the methane metabolism of methanogenic organisms like *Methanosarcina barkeri/mazei* but more importantly, it is incorporated endogenously by amber codon suppression. Its corresponding tRNA $pylT^{CUA}$ already recognizes the amber stop codon as a coding codon and can produce the proteins for methane metabolism.(58) A further advantage of these archaeal aaRS/tRNA pairs from *M. mazei/barkeri* or from *M. jannaschii* is that these pairs are bioorthogonal in all expression systems, ranging from bacteria to yeast, cell lines and nematodes up to mammals.(59-61) But the amber suppression system always competes with the native chain termination by RF1, reducing the overall yields. Therefore, different *E. coli* strains have been created without endogenous amber stop codons and through a RF1 knockout, the counter player of the amber suppression system was removed.(62) This enables incorporation rates of 100%, but changing these many stop codons in the genome comes with several drawbacks. Although the suppression and incorporation rates are at an optimum, this strain suffers from drastic growth deficiencies in fact even lowering the overall yields of the target protein carrying a novel reactive group.

In order to be able to incorporate amino acids with novel chemical functionalities, the $aaRS^{PylS}$ and $aaRS^{TyrS}$ have been further engineered to encode hundreds of different unnatural amino acids. The engineering process is based on random mutagenesis of the active site of the aaRS and subsequent selection in presence of the new ncAA. In the first step, a positive selection is carried out. For this, an amber stop codon is inserted into an antibiotic resistance gene and the cells are grown in presence of the ncAA and the antibiotic. Thus, cell growth can only occur if the aaRS loads the amino acid onto the $tRNA^{CUA}$ and the ncAA is incorporated into the resistance gene. For $aaRS^{TyrS}$, a further selection, because of its higher miss-incorporation rate, has to be carried out, eliminating aaRS, which charge the $tRNA^{CUA}$ with the endogenous tyrosine. Therefore, an active site tyrosine from a toxic gene, for example the Barnase which randomly cleaves DNA, is mutated to an amber stop codon. Cells are than grown in the absence of the ncAA and will only survive if the endogenous amino acid is not incorporated.(63) To increase the overall efficiency of the aaRS, this selection process can be carried out multiple times. With this engineering process, over 100 different ncAAs can be incorporated of which some of the more recent ones together with their modes of action and possible applications in fundamental research, protein engineering or drug development are shown in Figure 8.

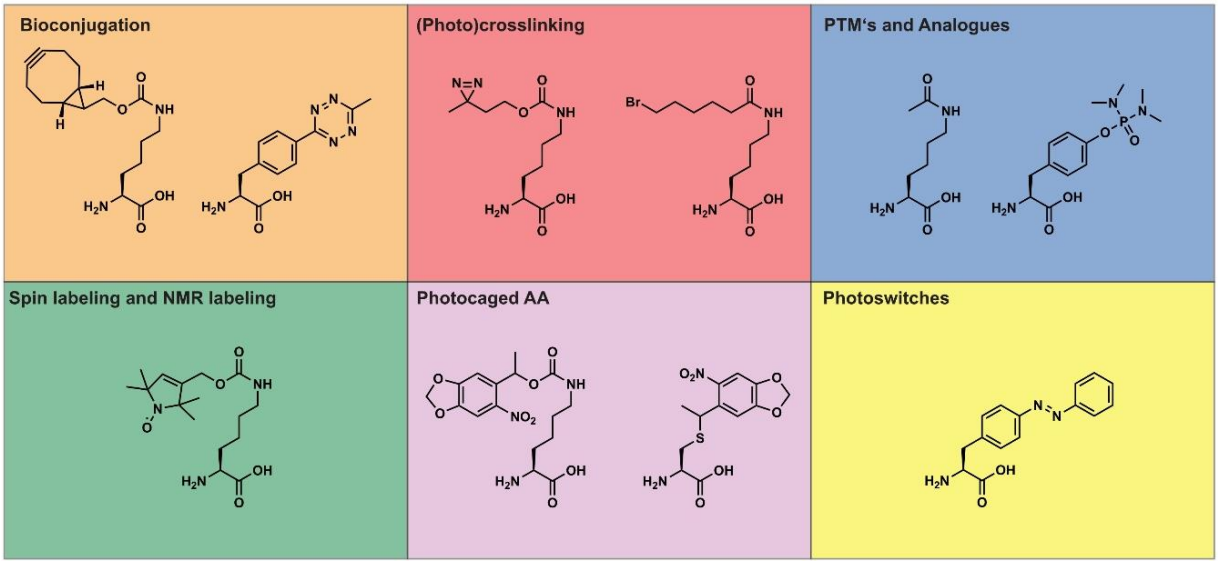

Figure 8: Variety of different ncAAs and their biotechnological applications
(Courtesy: Maximilian Fottner, TUM)

Next to the ncAAs shown in Figure 8, a broad range of reactive group like, alkynes, azides, borates or ncAAs, which alter chemical properties, can be introduced into the protein of choice. (63-72) The vast variety of reactive groups and the possible applications arising from them require, reaction partners which are fast, biocompatible/nontoxic, and function in aqueous environment, especially when working with living cells.

The first coupling reaction meeting nearly all of the criteria was the Cu(I)-catalyzed Huisgen 1,3 dipolar cycloaddition, based on the Huisgen cycloaddition from 1964, and introduced by Sharpless and co-workers in 2002.(73, 74) This reaction is fast and shows nearly perfect regioselectivity, but Cu(I) salts as the catalyst limit the use in cellular applications. This reaction, also called CuAAC (copper-catalyzed azide-alkyne cycloaddition) is one of the most prominent representatives of the "Click Chemistry".(75) Reactions featured under the term "Click Chemistry" are generally characterized by showing high yields of a single reaction product with minimal by-products. The reactions is under a high thermodynamic driving force and additionally compatible to water or oxygen. As Cu(I) is cytotoxic for *in vivo* applications, copper-free Huisgen cycloadditions like the strain-promoted azide-alkyne cycloaddition (Figure 9-A) were developed.(76) Instead of a catalyst, which would activate the alkyne for the reaction, the alkyne is present in a strained eight-membered ring (cyclooctyne). By adding ligands like fluorine to the ring, the reactivity can be further increased, in turn decreasing the reaction time.(77) Another type of bioorthogonal reactions with biocompatibility are inverse-electron demand Diels-Alder reactions. The Diels-Alder reaction is a [4+2] cycloaddition and was developed by Otto Diels and Kurt Alder in 1928.(78) In the recent development, strained cyclooctenes or other activated alkenes like the endo-norbornene lysine, react with tetrazines and form a conjugate with gaseous

nitrogen as the only by-product.(79) Figure 9-B demonstrates an example of an inverse-electron demand Diels-Alder cycloaddition.

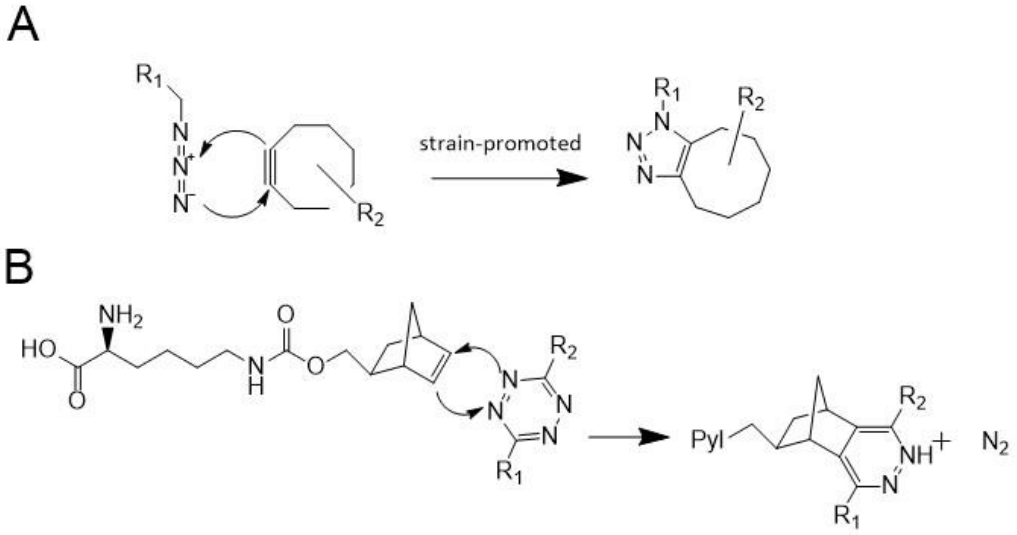

Figure 9: State of the art bioorthogonal reactions
A: Reaction of a strained cyclooctyne with an azide ligand forming the conjugate between both components. B: Endo-norbornene lysine reacts with a tetrazine ligand forming the conjugate with gaseous nitrogen as the only by-product.

By now incorporating ncAAs carrying the reactive groups shown above via amber suppression into the protein of choice, (*in vivo*) protein labelling can be easily performed. Here, live cell imaging to study protein-protein interactions or the search for possible reaction partners, is highly desirable. By bioorthogonally labelling a target protein with a fluorophore attached through a biocompatible click reaction, the behaviour of the target can be studied in its native state much more easily, in contrast to genetic fusions with heavy fluorescent protein like GFP. (70, 80) Alternatively, methods like BONCAT (bioorthogonal non-canonical amino acid tagging) were developed which allow time resolved monitoring of changes in the protein expression behaviour.(80, 81)

Next to the advantages in research, site specific labelling is also highly important in the pharmaceutical industry as since the first antibody-drug conjugates, the benefits of these combinations have been on hand. However, as the production of antibodies is quite cost expensive, new protein classes, derived from and inspired by antibodies have been developed since the 1980s.(82) The affinity of antibodies to its antigen results from the variable domains, which together constitute the antigen-binding site at the top auf both arms of the antibody.(83) These antibody-derived proteins include “single chain variable fragments” or “single chain antibodies” as shown in Figure 10.

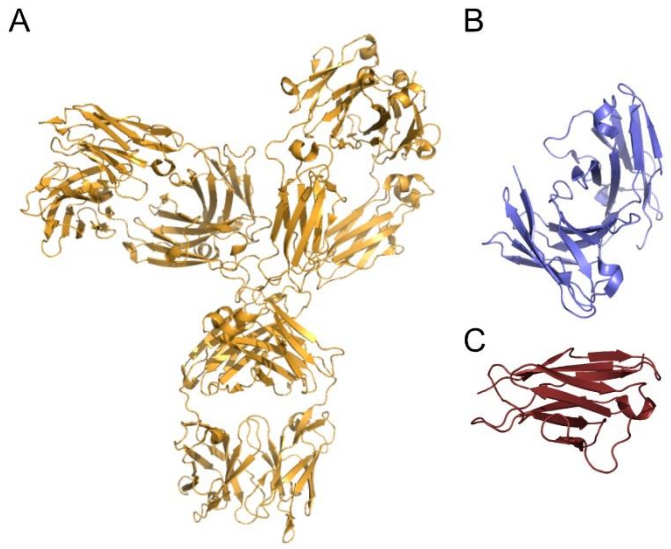

Figure 10: Structural examples of an antibody and antibody derived single-chain fragments
A: Structure of an antibody (PDB 5DK3), B: Structure of a single chain variable fragment (PDB 1P4K), C: Structure of a single chain antibody

The low weight of these domains (about 13 kDa each) compared to a full-length antibody (approx. 150 kDa) as well as their high stability makes them ideal for bacterial expression. In the last decades, various combinations of antibody fragments and their expression in bacterial hosts like *E. coli* and *B. subtilis* with yields up to the gram per litre scale have been reported.(23, 84, 85) To date, several antibody domain based pharmaceuticals have already been approved for human treatments by the FDA or the European medicines agency (EMA) or are in clinical trials.(86, 87) By using engineered *B. subtilis* strains for the expression of these pharmaceuticals, costs could further be reduced. In addition, modifications of these small proteins, for example by using the amber suppression method to increase their plasma half-life or overall stability, could increase the number of applications. However, the amber suppression method as todays most used method for site-specific protein modifications by expanding the genetic code beyond the 20 canonical amino acids has not been established in *B. subtilis* yet.

1.4. Objective of this work

Within this thesis, several different topics towards establishing *B. subtilis* as an expression host for bioorthogonally-modified proteins are being addressed.

The aim of the first part of the thesis is to engineer the protein expression of *B. subtilis* for high yields of the target protein while simultaneously controlling the system. This should be accomplished by evaluating different promoters and optimizations of the overall expression construct. The main goal of this work is the implementation of a genetic code expansion system in *B. subtilis*. Therefore, the amber suppression method as the most used method for bioorthogonal modifications of proteins was selected. The further advantages of *B. subtilis*, namely the highly efficient secretion system should be used to secrete antibody-derived proteins into the supernatant, easing their overall purification by reducing the number of purification steps. By combining the amber suppression system and the secretion of the antibody-derived proteins, bioorthogonally modified antibody fragments should be secreted into the supernatant.

2. Results and discussion

2.1. Optimizing the available expression system

Although *B. subtilis* is one of the best-studied microorganisms and a key player in the biotechnological industry for the large-scale production of enzymes, commercial plasmids or expression systems are quite rare and not as vastly engineered compared to *E. coli* plasmids. In the first part of this work, the overall expression of a target protein was optimized to assure suitable high expression levels and gain more flexibility for later cloning tasks. In general, several steps from the DNA to the protein level can be addressed to tune the expression as visualized in the figure below. The first step is hereby the consideration if the expression should be carried out via a genomically integrated expression construct or via a replicative plasmid, which is present in multiple copies. On the DNA-level of the expression construct, the promoter and a general control of the expression are of great importance as they regulate the amount transcribed mRNA. The mRNA itself bears several factors, notably the ribosome-binding site (RBS) and more general the overall mRNA stability that influences the mRNA half-life and therefore the availability for the ribosome to produce the target protein.

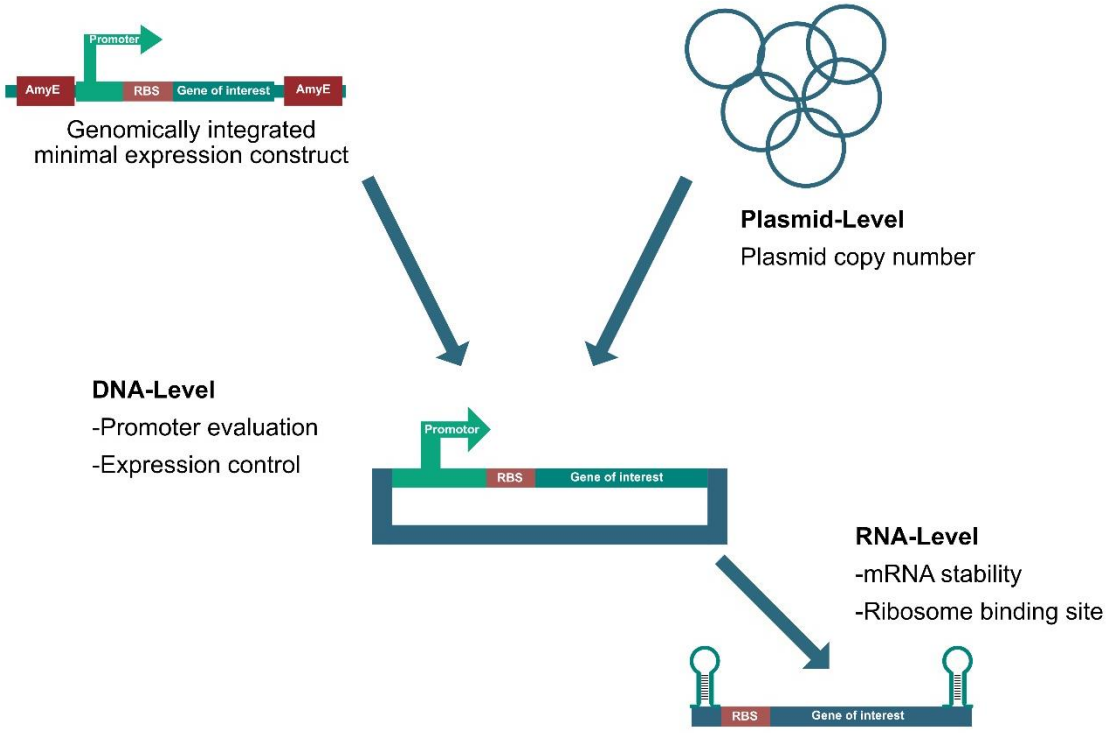

Figure 11: Overview of different optimization targets on DNA and mRNA level
The different target can potentially be optimized in order to achieve higher expression yields of a target protein.

When optimizing the overall expression of a target protein, different factors have to be considered and adjusted to meet the needs for each protein of interest. The first part of this work shows the step-by-step optimization and its impact on the expression rates.

2.1.1. Replicative plasmid increases the protein expression

In *B. subtilis*, there are two different options for a plasmid based protein expression: Integrating and replicative plasmids. Both plasmid types rely on *E. coli* as the shuttle vehicle for cloning and plasmid amplification.

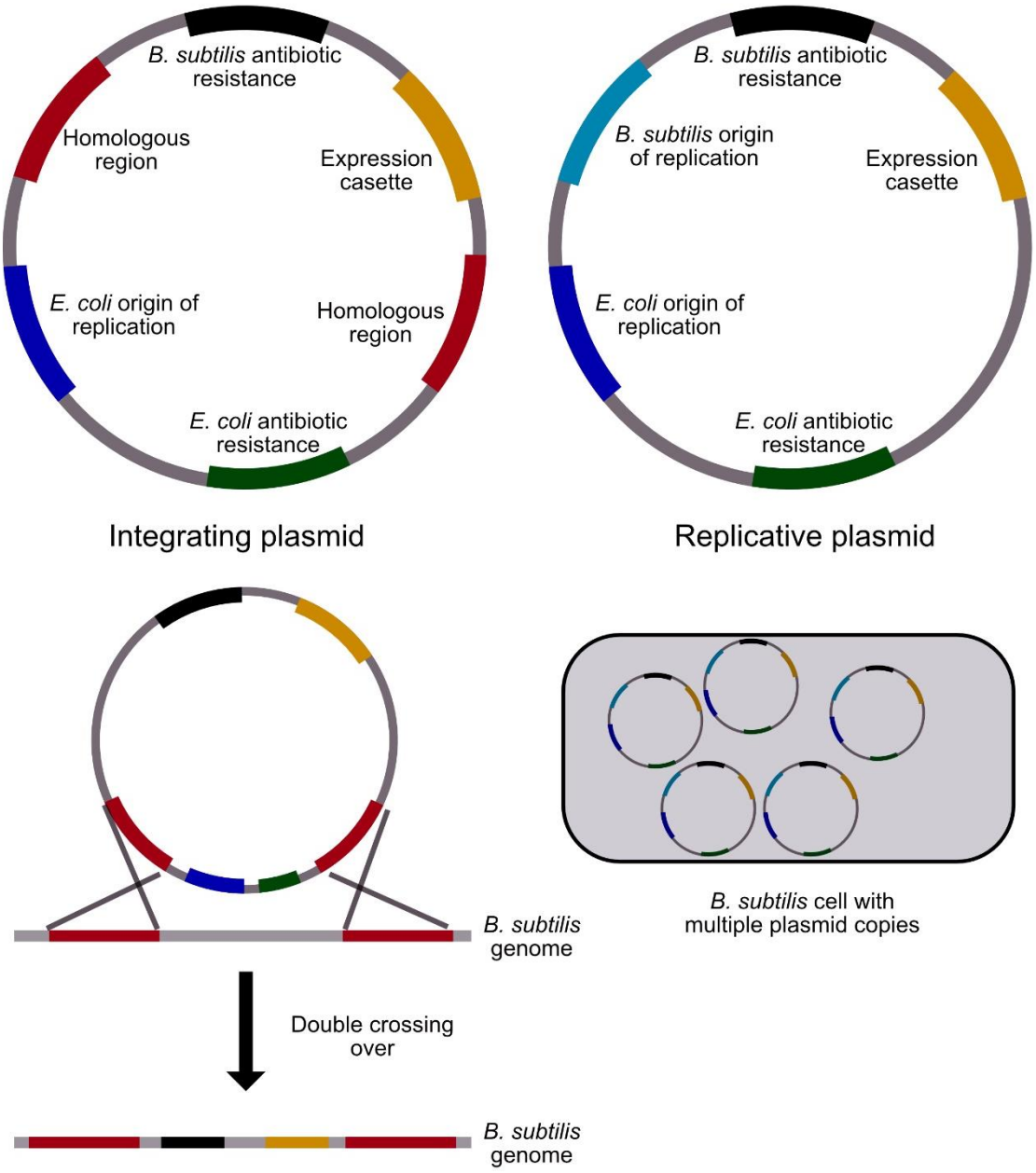

Figure 12: Structural overview of integrating and replicative plasmids in *B. subtilis*
Both plasmid types require an *E. coli* antibiotic resistance and origin of replication for cloning and plasmid construction in *E. coli*. For selection in *B. subtilis*, an additional antibiotic resistance is present. The integrating plasmid can perform a double crossing over, integrating the *B. subtilis* antibiotic resistance gene and the expression cassette into the genome. The *E. coli* part is lost. The replicative plasmid is found in multiple copies in the cell, depending on the used origin of replication.

Irreversible integrating plasmids carry two homologous regions of about 1000 bp in their sequence, which are homologous to the *B. subtilis* genome. The area in between, the antibiotic resistance as well as the expression cassette can be integrated into the genome under antibiotic selection. By

homologous recombination a double crossing over between the homologous regions, as shown in Figure 12 occurs, the *B. subtilis* part of the plasmid is integrated irreversibly and the *E. coli* part of the plasmid is lost. After this no further selection with antibiotics is required.(88) These integrating plasmids are suited for a constant, well-controlled gene expression without overproduction of the target protein. In contrast, replicative plasmids, like the pLIKE plasmid, do not carry homologous region but contain a *B. subtilis* origin of replication. In this plasmid, the Ori1030 from the *Bacillus thuringiensis* plasmid pHT1030 is used and about 15 ±-5 plasmid copies per cell are present.(89, 90) However, constant selection pressure by antibiotics is required otherwise, the plasmid is lost overtime.

To evaluate the difference between the integrated and the replicative plasmid on the target gene expression, *gfp* controlled by the constitutive promoter P_{veg} was cloned into the integrating plasmid psBBs1c, which integrates into the *amyE*-locus and the replicative pLIKE plasmid. The two plasmids were transformed into the protease deficient *B. subtilis* strain K07 and positive clones were pooled to mitigate potential differences in the expression levels in different clones. The strains K07-1 (pSBBs1c-P_{veg}-GFP) and K07-2 (pLIKE-P_{veg}-GFP) were inoculated in LB medium containing the respective required antibiotics overnight and subsequently grown in the defined MCSE medium in the plate reader for 24h.

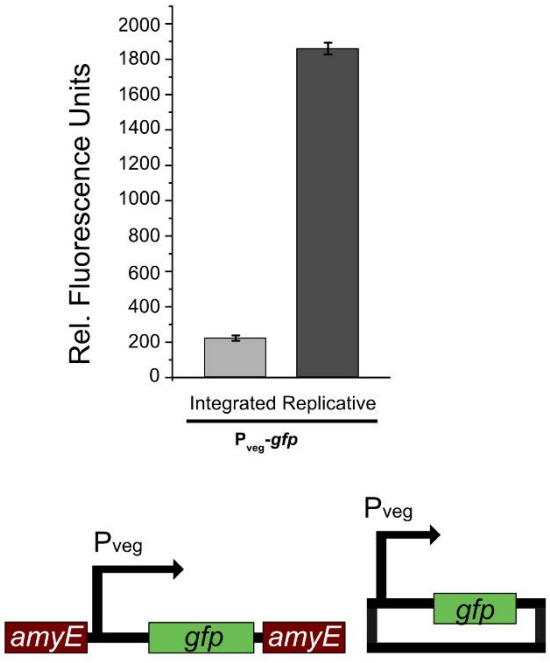

Figure 13: Difference between an integrated and replicative expression construct
The two *B. subtilis* strains K07-1 and K07-2, left expressing GFP from a genomically integrated construct, right from the replicative pLIKE plasmid were grown in 96-well plates in the plate reader. Both constructs control *gfp* via the same promoter P_{veg}. Relative fluorescence measured after 24h of cultivation. The measured fluorescence values were normalized to an OD of 1

The fluorescence signal shown by K07-2, expressing GFP with the replicative expression plasmid shows a ten times higher GFP signal than the strain K07-1 carrying the genomically integrated construct,

which is in agreement with the reported plasmid copy number.(89) However, during cultivation of the cells, no green color as usually with GFP-expressing bacterial strains was visible, indicating a further potential for optimization.

2.1.2. Stabilization of mRNA through secondary structure elements

In contrast to DNA, mRNA is in general not very stable and is additionally subjected to a rapid degradation by various RNases, which are digesting the mRNA from both ends or cleaving it internally.(91) The so-called mRNA half-life time is also a further regulation tool for gene expression in all domains of life.(92) In previous studies, stabilizing the mRNA via structural elements has been used to increase the mRNA half-life.(93) Based on this work, the plasmid pLIKE was equipped with either a 5' stem-loop, a 3' terminator or both elements flanking the cloning site of the gene of interest using a SLIM-PCR (Site-directed, ligase independent polymerase chain reaction).(94) The possible three combinations of the secondary structure elements were transformed into *B. subtilis* strain K07, resulting in the strains K07-3 (pLIKE-P_{veg}-5'Stem-GFP-His_6), K07-4 (pLIKE-P_{veg}-GFP-His_6-3'-Term) and K07-5 (pLIKE-P_{veg}-5'Stem-GFP-His_6-3'-Term). The strains were tested together with the maternal expression construct K07-2 and grown in LB medium overnight. After inoculation in MCSE medium, the strains were grown in 96-well plates in the plate reader. Figure 14 shows the influence of the different secondary structure elements on the protein expression of GFP.

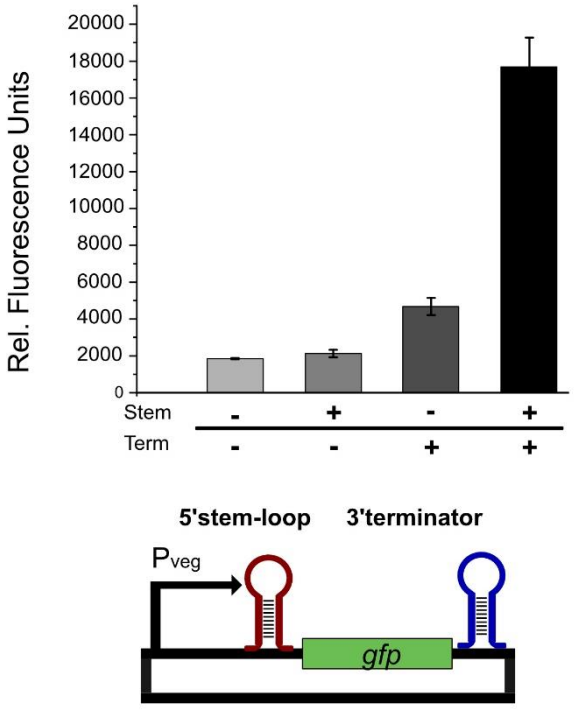

Figure 14: Influence of mRNA stabilizing elements on the protein expression
Relative Fluorescence signal after 24h cultivation in MCSE medium. Values are normalized to an OD of 1. All strains are under control of the strong promoter P_{veg} ensuring a constant protein expression. Ten wells for each of the four strains were set up to decrease influences by the plate layout. The strains contain no, one or both mRNA stabilizing secondary structure elements.

The maternal strain K07-2 without any stabilizing elements exhibits the same fluorescence intensity as in the previous experiment (Figure 13). The addition of the 5' stem-loop shows only a slight increase in fluorescence. Although the 5' end is now protected from the degradation by RNases – the 3' end is still a target of degrading enzymes. The introduction of the 3' terminator to the sequence, however, comes with two advantages: The 3' end of the mRNA is protected from cleavage and additionally, the terminator secondary structure leads to the release of the mRNA from the DNA/RNA duplex after transcription.(95) This makes the RNA polymerase available for further transcription and the production of more mRNA templates for protein translation. With the combination of both secondary structure elements, a drastic increase of the fluorescence signal is observed due to the protection of the mRNA from cleavage by exoribonucleases and increase of the mRNA half-life thus more mRNAs are available for translation into the target protein, in this case GFP. Bye introduction of the two secondary structure elements, higher amounts of the protein of interest can be produced.

2.1.3. Evaluation of *B. subtilis* promoters

To increase the flexibility of the protein expression levels, a set of promoters, previously analyzed in a promoter screening, were closely re-evaluated.(96) Therefore, the four strongest reported promoters, $P_{trnQ'}$ P_{sigx}, P_{groES} and P_{43}, were chosen and the promoter region was cloned upstream of the 5' stem-loop into the GFP expression plasmid #1. The four strongest promoters in this screening were P_{trnQ}, the promoter of an arginine-tRNA, P_{sigx} the promoter of the sigma factor X, responsible for regulation of the cell wall stress response (97), P_{43} a previously described constitutive promoter (98, 99) and P_{groES}, the promoter of the GroES chaperone. In contrast to the previous analysis, only the genomic region upstream from the transcription start site "+1" to the end of the upstream gene was amplified from the *B. subtilis* 168 genome. For testing the promoter strength, the *B. subtilis* K07 was transformed with the resulting five expression plasmids #1-5 and positive clones were inoculated in LB medium overnight. After inoculation in MCSE medium, the strains K07-5 to K07-9 (P_{veg}/P_{groES}/P_{43}/P_{sigx}/P_{trnQ} - 5'*gfp*-His_6-3') were grown in the plate reader and the GFP fluorescence determined over time. In contrast to the previous promoter comparison, the strongest promoter is P_{groES} of instead P_{trnQ}.

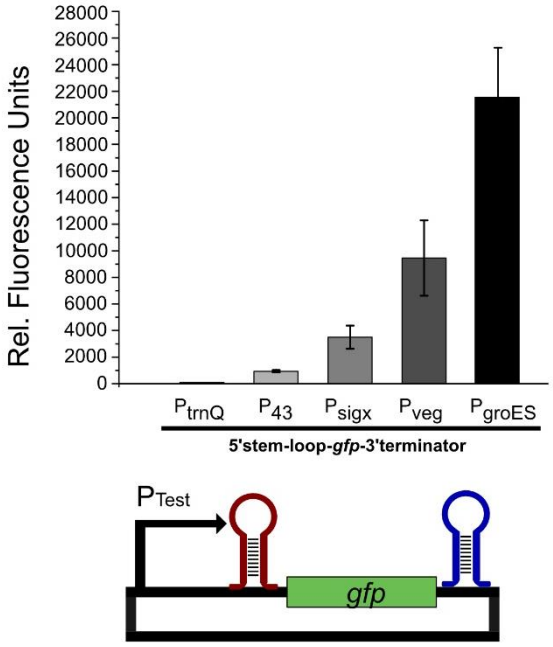

Figure 15: Expression profile of five different *B. subtilis* promoters
Relative GFP fluorescence of four strong promoters and P_{veg} after 24h cultivation in the plate reader. The expression construct for the promoter "P_{Test}" to be evaluated controls the double-stabilized *gfp* expression construct. Ten biological replicates are measured in this setup to minimize differences from the plate layout.

This is most likely due to the fact that in the study of Nikoloff and co-workers, 2000 bp upstream of the respective gene were included, which possibly contains additional ORFs or regulatory elements impacting on the overall gene expression. For P_{trnQ} for example, several strongly expressed proteins lie upstream of the promoter, which certainly influence the promoter activity.

Moreover, during cloning of the different promoters, a *Bam*HI site was introduced directly upstream of the 5' stem-loop in the P_{veg}-5'*gfp*-His_6 3' construct which is also transcribed. The addition of six nucleic bases of the restriction site decreases the GFP expression by about 40%. These six bases in the mRNA are not part of the 5' stem-loop, and could hereby allow RNases to attack the protected mRNA via the free 5' end, opening up the stem-loop, which would alternatively not be accessible.

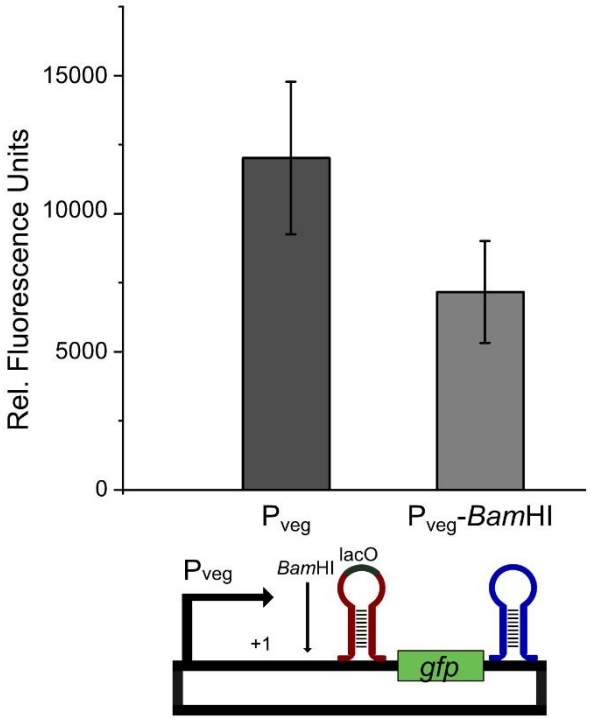

Figure 16: Influence of the *Bam*HI site directly downstream of the transcription start site on the overall fluorescence
Six additional bases in front of the 5'stem-loop of the P_{veg}-5'*gfp*3' construct decrease the fluorescence signal by 40%. The fluorescence was normalized to an OD of 1 and 10 wells were measured for each condition.

As in some cases a constitutive but moderate expression is desired, the weak promoter P_{sigx} is chosen to be analyzed together with P_{groES}. As these two native promoters are not characterized and engineered, their expression behavior in different media is investigated.

2.1.4. Cultivation dependent promoter activity

To examine a possible change in the promoter activity of P_{groES} and P_{sigx} under different cultivation conditions, the strains K07-6 and K07-8 are grown in different expression media with varying the carbon sources.

For this, three complex media, LB, 2YT, TB and two defined media, MCSE and MN-CSE were chosen. The three complex media contain yeast extract and peptone, TB is equipped with glycerol as an additional carbon source. Both defined media are buffered and contain all required nutrients for *B. subtilis,* with casamino acids and fructose as carbon sources in both media, while MN-CSE additionally contains glycerol. The *B. subtilis* K07-6 strain expressing *gfp* under control of P_{groES} was inoculated in LB medium overnight and grown in the five target media in the plate reader. The left part of Figure 17 presents the different growth curves in all five used media. In general, all of them show OD values above 1.0, except LB, and stay at that level throughout the 24h of cultivation. As expected, the lower nutrient concentration of LB leads to a lower overall cell density. The fluorescence signal on the right half of Figure 17 is normalized to an OD density of 1.0. Here, all five different media show a

strong expression of GFP. Bacteria grown in LB and 2YT medium, which are both only based on peptone and yeast extract, exhibit the same expression behavior depicted as a constitutive expression of GFP throughout the 24h cultivation time. However, as 2YT displays a higher cell density, the yields per culture volume are higher. The strain grown in TB medium, which contains next to peptone and yeast extract also glycerol, exhibits a bi-phasic fluorescence signal and growth, both indicating a switch in the used carbon source for the metabolism. Bacteria grown in the defined media MCSE and MN-CSE also exhibit a weak bi-phasic expression corresponding to their growth behavior but without a strong expression stop as with TB-medium. Both defined media reach higher OD values than the complex media, but the relative fluorescence signal does not achieve the strong expression in LB or 2YT.

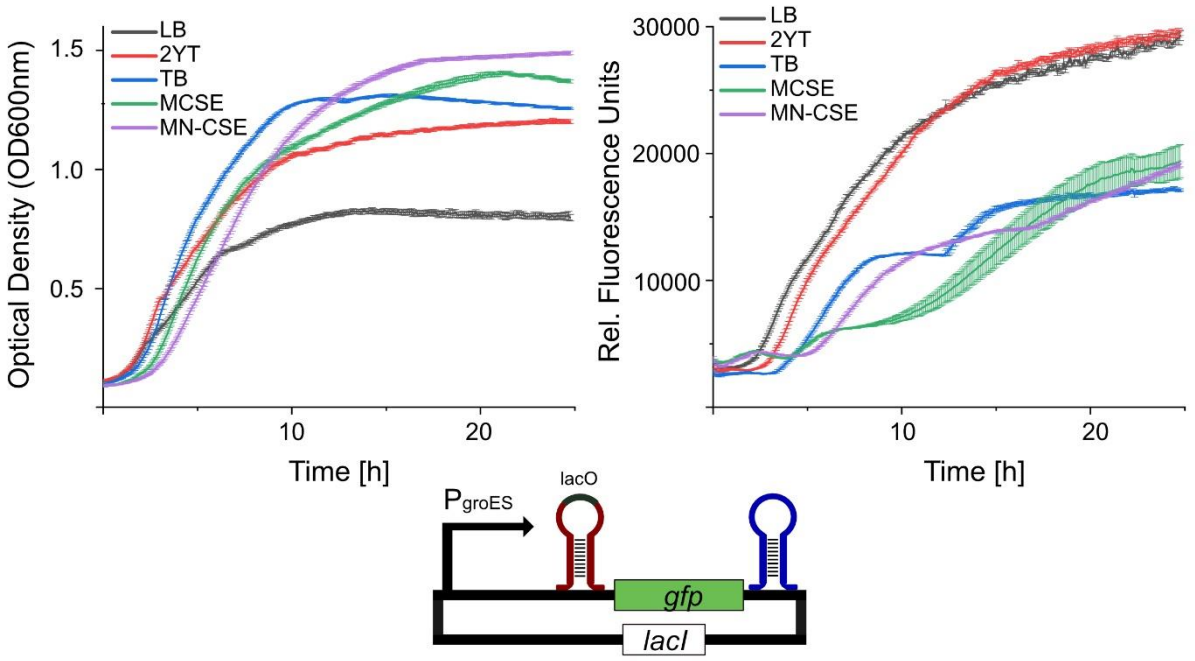

Figure 17: Growth curve and expression profile of *B. subtilis* K07-6 expressing *gfp* under control of P_{groES} in different media
The strains were inoculated to an OD of 0.1 and grown at 37°C under shaking. The strains were cultivated in the different expression media over a time course of 24h

Taken together, in 2YT and LB media, the strongest overall expression yield is found, which is in addition characterized by a constitutive increase and without any bi-phasic growth or lag phases. However, it has to be taken in account that the normalization of the fluorescence leads to a superelevation of the fluorescence in LB medium where the yields per volume are lower due to the lower cell density. The other three media do not attain the yields of GFP although showing partially stronger cell growth. Therefore, 2YT medium can be considered as a suitable medium for the expression of a target protein controlled by P_{groES}.

To assess the promoter development of P_{sigx}, the five media were tested with *B. subtilis* strain K07-8. The left part of the Figure 18 represents the growth of the bacterial strain in the five media. Again all

media except LB show a constant or increasing OD throughout the 24 h of measurement. Growth in LB medium however begins to decline after about 10 hours, indicating that the nutrients are depleted. The relative fluorescence intensity exhibits a completely different expression behavior compared to P_{groES}. The expression in the complex media 2YT and TB show a strong increase in the fluorescence, with a nutrient switch in TB medium as observed with P_{groES} (Figure 17). The strong fluorescence signal in LB medium is in this case an artifact from the normalization. As the OD itself declines in presence of a constant fluorescence signal, the relative fluorescence increases. The defined media MCSE/MN-CSE only show a very weak overall expression that starts to increase after 24 hours of cultivation.

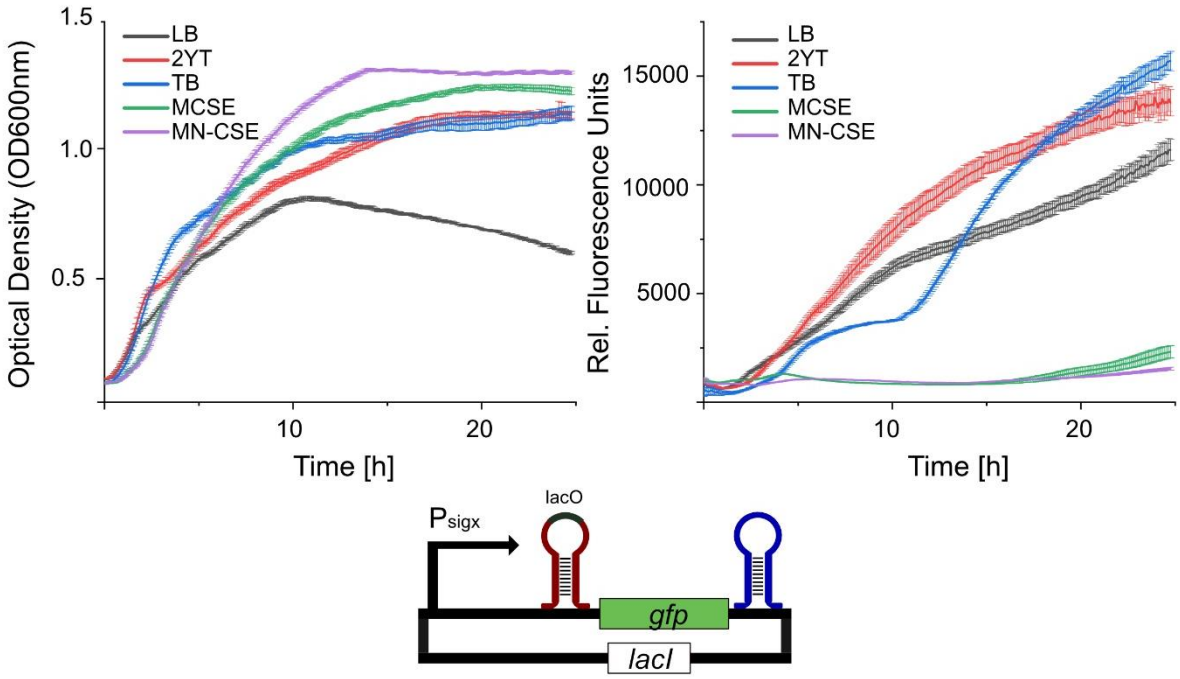

Figure 18: Growth curve and expression profile of *B. subtilis* K07 expressing *gfp* under control of P_{sigx} in different media
The strains were inoculated to an OD of 0.1 and grown at 37°C under shaking. The strains were cultivated in the different expression media over a time course of 24h.

The detailed investigation of both promoters revealed differences in the expression behavior depending on the used cultivation medium. Based on these results, 2YT was chosen for further experiments, since a constant promoter activity can be observed for both strains, which is desirable for the overall protein expression.

2.1.5. Promoter-inducibility using isopropyl-β-D-thiogalactopyranosid

Although a strong promoter with high expression levels is for most applications a desired goal, a timed expression of the target protein is equally important. A continuous or too early expression of the target protein, especially if the target protein is toxic for the host organism, can stress the cells, resulting in lower expression levels or growth of bacteria, which suppress the protein expression.

As native promoters are often subjected to strict regulations, a specific repressor counterpart exists which can be used for a controlled protein expression in the laboratory. Repressors can either act on the DNA, where they block the binding of the RNA polymerase and inhibit the transcription or on the mRNA where they block the binding of the ribosome, preventing the translation. An example is the pBAD promoter, where the transcription is induced in the presence of arabinose.(100) In *B. subtilis* sugar-based promoter/repressor systems using xylose, mannose or mannitol as inductors have been established in the recent years.(101-103) However, these three systems are the endogenous promoter, which are not optimized for the overexpression of a target protein.

An alternative are expression systems based on anhydrotetracycline or isopropyl-β-D-thiogalactopyranosid (IPTG) which are well engineered and used in many expression organism including *B. subtilis*.(104, 105) The anhydrotetracycline system relies on the binding of anhydrotetracycline to rtTA (reverse tetracycline transactivator), which can now activate the expression of the target gene.(106, 107) The most prominent one, the IPTG induction system originates from the Lac operon, which consists of a promoter, three operators (LacO1/2/3) and three proteins, the β-galactosidase LacZ, the transport protein LacY and a β-galactoside transacetylse LacA. This operon is negatively regulated by the Lac inhibitor LacI. In this operon, the repressor LacI forms a tetramer and can bind to the Lac operators. Upon binding, the transcription by the RNA polymerase is blocked and no mRNA is produced. Already in 1984, this system was engineered for the use in *B. subtilis* by introducing the Lac inhibitor and the Lac operator to the expression plasmid.(105) An advantage of this system is the well-studied modularity as directly adding the Lac operator to an expression construct makes the promoter IPTG inducible.(93, 105) In the work by Phan et al. (93), the Lac operator was directly fused to the 5' end of the mRNA as mentioned in the previous.

The plasmid #9 containing the strong promoter P_{groES} and the 5'/3'-end stabilized *gfp*, was equipped with *lacI* and transformed into *B. subtilis*. Positive clones were pooled and grown overnight in LB medium. After growth in MCSE medium, the strains were split into two samples and one of the two samples was induced with 1 mM IPTG.

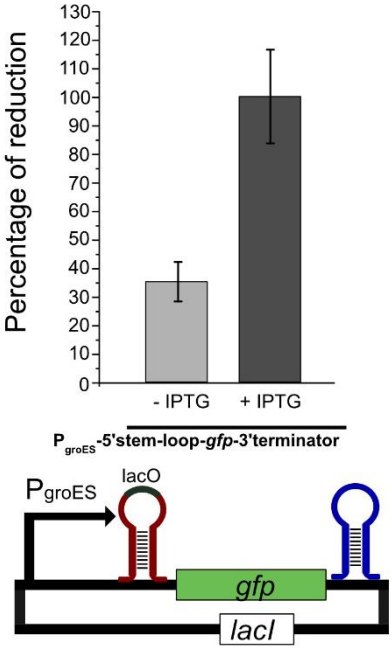

Figure 19: Effect of the addition *lacI* on the GFP expression
LacI was added to the expression construct P_{groES}-5' *gfp* 3' construct. Without IPTG, a residual promoter activity can be detected. The addition of 1 mM IPTG activates the full expression level.

The fluorescence signal was normalized to the IPTG-induced strain. Here, the strain without induction shows an overall signal of about 30%, implying that the expression setup is not completely tight and shows a slight read through. Compared to existing IPTG inducible promoters like P_{spac}, a certain leakiness is observed.(105) Additionally, without an further screening and promoter optimization as later performed by Phat et al. (2015), a basal promoter activity is also found.(108) However, the amount of promoter read through is an acceptable range and should not be of concern for the next steps in this work as the used proteins are not very toxic.

2.1.6. Summary

By addressing various parts of an expression system as shown in Figure 11, the overall protein production of GFP in *B. subtilis* was thoroughly optimized. The created expression system is a good start for further optimizations but cannot compete with highly engineered commercially available expression plasmids for *E. coli* like the pET- plasmid series yet. However, the screened and evaluated promoters are very useful to tune the expression levels of a target protein, as sometimes a moderate expression is favoured over an overexpression. Interestingly, the two promoters P_{groES} and P_{sigx} showed very different expression behaviours depending on the used cultivation medium. Stabilizing the mRNA with secondary structure elements greatly increased the mRNA half-life, in turn increasing the protein yields, and simultaneously made the promoters IPTG inducible.

2.2. Establishing amber suppression in *B. subtilis*

To perform amber suppression as a method for genetic code expansion and to equip a protein with a new, non-coding amino acid, three points have to be addressed:

- Introducing an amber stop codon into the protein of choice at the desired site
- Expression of a bioorthogonal aminoacyl tRNA synthetase
- Expression of the corresponding tRNA targeting the amber stop codon

In the following paragraphs, bottlenecks associated with genetic code expansion by amber suppression such as the overall expression setup, the aaRS/tRNA ratio and expression levels and the genomic context of the incorporation site are addressed.(109-112)

2.2.1. General considerations and expression setup

In initial experiments, a single plasmid system, as used for amber suppression in *E. coli,* containing both the target protein and the aaRS, based on the pLIKE plasmid was designed. Here, the strong promoter P_{groES} controlled the target protein *gfp* and the moderate promoter P_{sigx} controlled the pylS^{Norb}, a triple mutant (Y306G, Y384F, I405R) with both elements present on the replicative plasmid pLIKE.(71, 113)

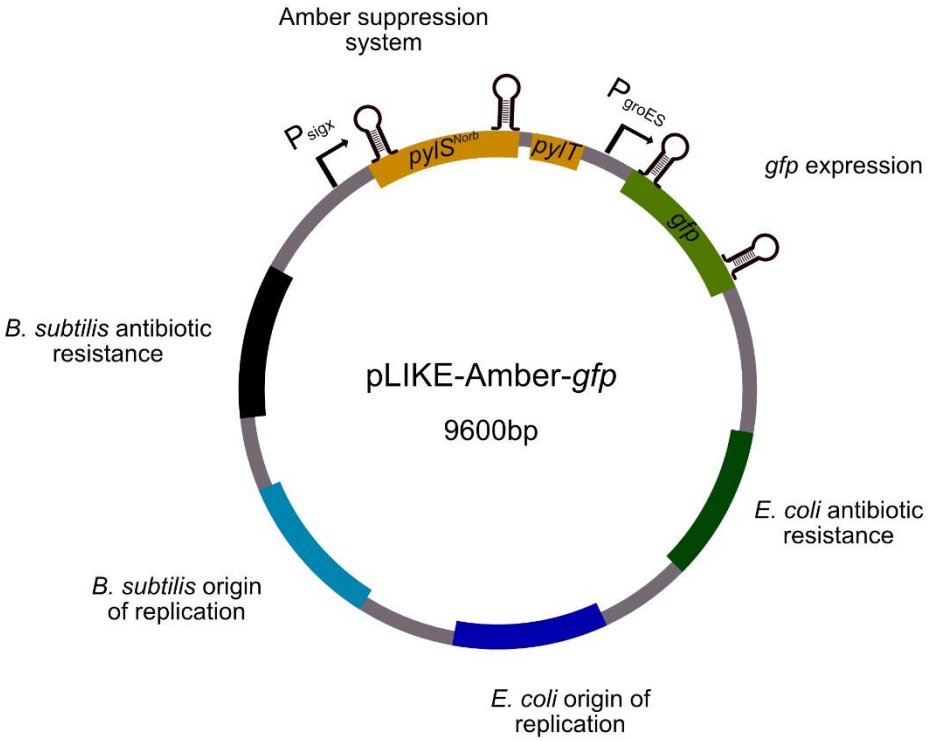

Figure 20: One-plasmid expression setup for amber suppression based on the pLIKE plasmid
The aaRS, the pylS triple mutant pylS^{Norb} is controlled by the promoter P_{sigx}, *gfp* by the promoter P_{groES}, both expression constructs are present on one plasmid.

This combination of a strong and weak promoter was stable in *E. coli* but showed stability issues when transformed in *B. subtilis* as shown in Figure 21, which is most likely due to the large plasmid size of 9600 bp and the expression of two proteins on the replicative plasmid. Other combinations with two strong promoters like P_{veg}/P_{veg} or P_{groES}/P_{veg} already showed plasmid fragmentations or mutations in the tightly regulated cloning strains of *E. coli* like XL1 blue or XL10 gold.

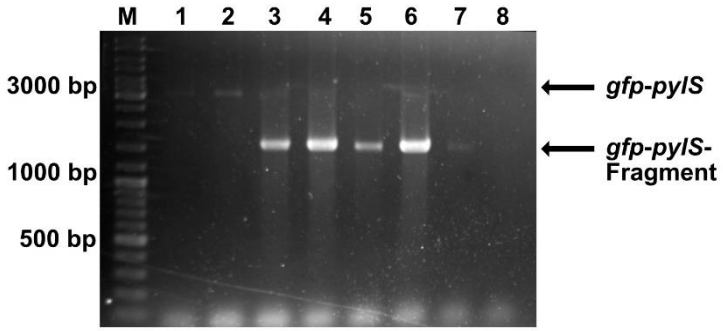

Figure 21: Instability of the one-plasmid system in *B. subtilis*
Agarose gel of a single colony PCR on eight *B. subtilis* clones after a transformation of the pLIKE one-plasmid system shown in **Figure 20** containing *gfp*, *pylS* and *pylT*. The top arrow indicates the expected size of the PCR product, the lower arrow the detected fragment.

The agarose gel above shows the deletion of about 1000 bp detected by the single colony PCR. Sanger sequencing of several plasmids revealed that the deletion is repeatedly found in $pylS^{Norb}$, indicating that the DNA sequence is unstable in this area. To overcome this, both the target protein as well as the synthetase were separated onto two individual plasmids. The replicative plasmid pLIKE is used to express the target protein, in this case GFP with a C-terminal His_6-tag.(89) This allows a strong overexpression as described in the previous section and a purification using Ni^{2+}-NTA affinity chromatography. The synthetase $pylS^{Norb}$ and its corresponding tRNA *pylT* are placed on the pSBBS1c plasmid.(88) This plasmid integrates irreversibly into the *amyE* locus via a double crossover under selection with chloramphenicol. After the integration, a further antibiotic selection is not required.

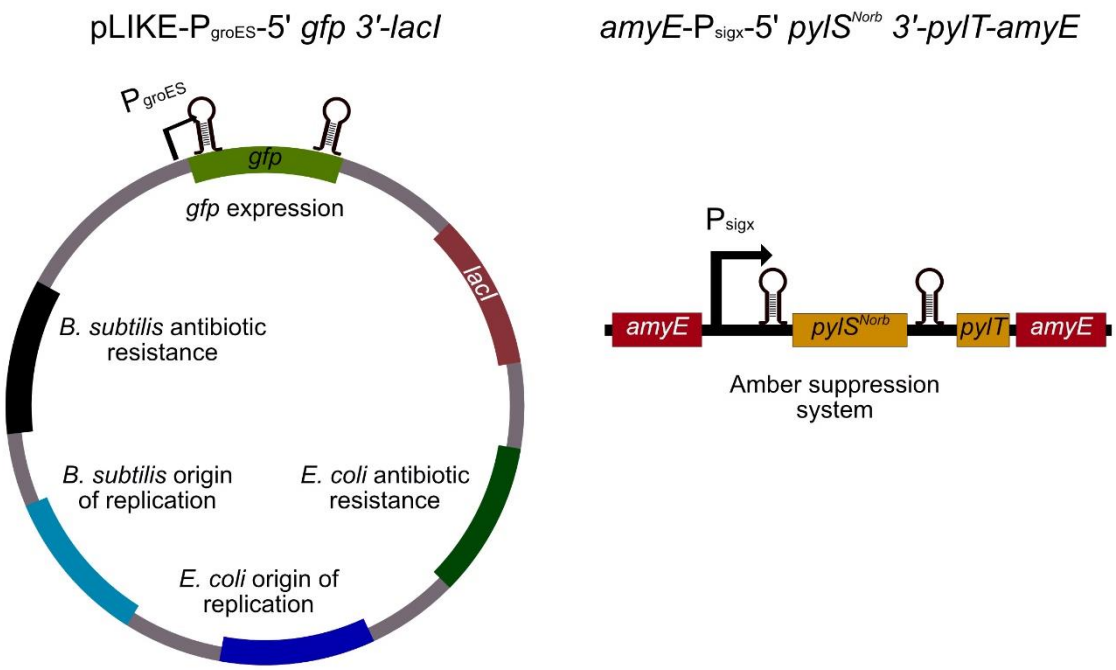

Figure 22: Two plasmid expression system
Gfp is expressed under control of P_{groES} on the replicative pLIKE plasmid. *PylS*^{Norb} is expressed under control of P_{sigx} together with *pylT* while genomically integrated into the *amyE* locus.

Although two consecutive transformation steps are now necessary, there are several advantages: The expression of the synthetase is now reduced, lowering the overall expression stress for the bacteria. Additionally, the plasmid stability greatly increased, visualized by a stable expression and no detected mutations/deletions when a single colony PCR is performed.

2.2.2. Expression of the amino acyl tRNA synthetase PylS^{Norb}

The expression of PylS^{Norb} was further investigated as an absent fluorescence signal from the target protein GFP for the incorporation of unnatural amino acids can have different reasons, ranging from a non-efficient incorporation site, a .(71) To test if the PylS^{Norb} is expressed, it was equipped with an N-terminal FLAG-tag and verified using western blotting against the FLAG tag.

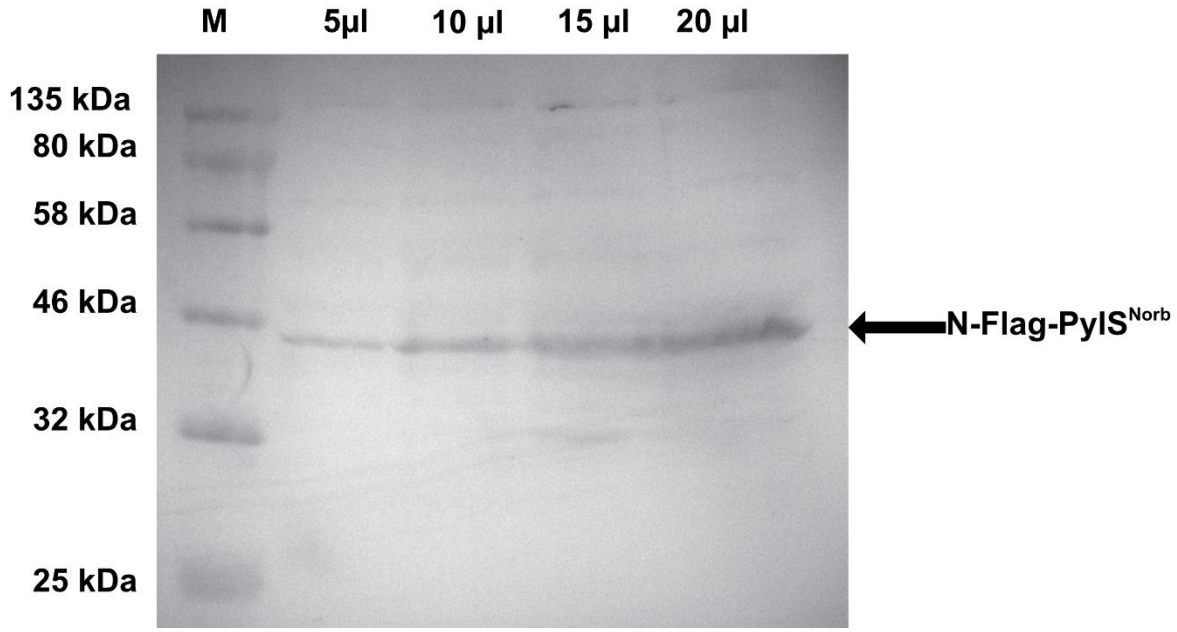

Figure 23: Western Blot against the N-terminal FLAG tag of PylS^{Norb}
The lysate was loaded in increasing amounts to detect even small amounts of the target protein. The band corresponding to PylS is marked with an arrow.

The western blot against the FLAG tag shows strong bands at around 50 kDa weight which corresponds to the size of PylS. These findings confirm that the synthetase PylS^{Norb} is successfully expressed in *B. subtilis*

2.2.3. Expression of the tRNA^{CUA} *pylT*

One of the three important parts to perform amber suppression is the expression of the tRNA *pylT*. In every organism, tRNAs are essential for the protein production and are therefore regulated and processed, leading to the mature tRNA (Figure 24).(91, 114-116) Both ends of the transcribed pre-tRNA are processed by different RNases. In *B. subtilis* and in most known organisms, the 5' end is processed by the universally conserved endonucleolytic ribozyme RNase P. The processing of the 3' end is more complex in *B. subtilis* than compared to *E. coli*. In *E. coli*, all tRNAs are encoded with at 3' CCA tail. However, in *B. subtilis*, 27 of 86 tRNAs lack this CCA tail, which has to be attached. If the CCA tail is missing, RNase Z and exonucleases process the 3' before a CCase attaches the tail.(114) If the CCA tail is present at the 3' end, cleavage by RNase Z is prevented and the 3' end is processed by RNases like RNase PH.

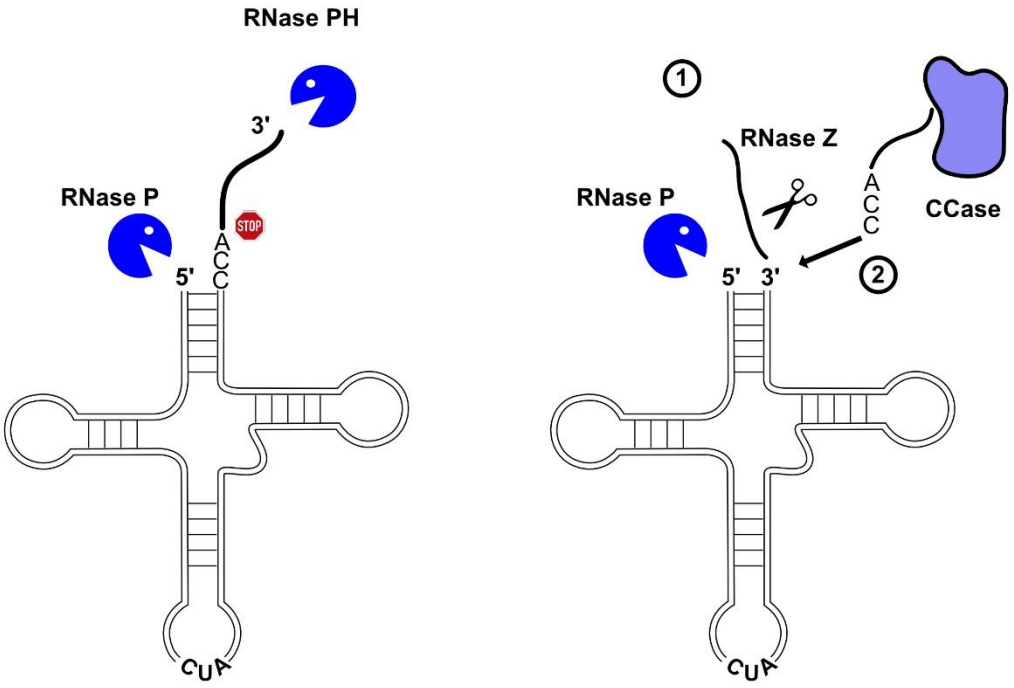

Figure 24: tRNA Maturation of most tRNAs in *B. subtilis*
All tRNAs are processed by RNase P at the 5' end. The 3' end of tRNAs carrying a 3'CCA are processed by RNase PH. tRNAs missing the CCA motif are first cleaved by RNase Z and subsequently a CCA tail is attached by a CCase.

To rule out any problems with the processing of the pre-tRNA and/or the transcription from the DNA, *pylT* is placed into the genomic region surrounding an endogenous *B. subtilis* serine tRNA as shown in Figure 25-A, ensuring the expression and processing of the bioorthogonal tRNA.

As previously reported, low incorporation rates can be caused by low abundance of *pylT*.(109, 117) Therefore a second tRNA construct, containing four additional copies of *pylT,* was designed and placed in the endogenous *B. subtilis* tRNA regions as shown in Figure 25-B.

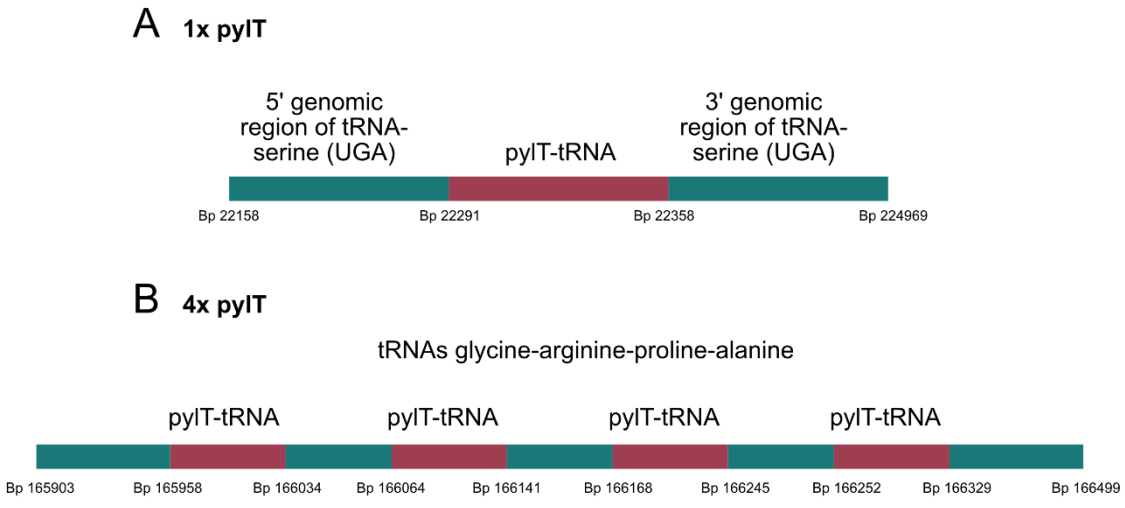

Figure 25: Structure of the 1x *pylT* and 4x *pylT* constructs
A: 1x *pylT* construct, *pylT* is placed in the genomic region of the serine tRNA. B: 4x *pylT* construct, four *pylT* copies are placed in the genomic region of four endogenous tRNAs. The location in the *B. subtilis* 168 genome are listed.

The 4x *pylT* construct was cloned together with a further copy of *lacI* into the second integrating plasmid pXT (#15).(118) This second integrating plasmid integrates via a double crossing over into the *thrC* locus disrupting the threonine synthesis in the target strain. The confirmation of the expression is carried out by testing the incorporation of endo-norbornene-lysine into the target protein GFP as a fluorescence signal is only visible upon successful incorporation.

2.2.4. Expression of GFP^{Norb26} and verification using mass-spectrometry

2.2.4.1. Fluorescence experiments

The advantage of using GFP as a reporter for the amber suppression system is the direct detectable GFP fluorescence. Moreover, since the experiments carried out in 96 well plates in the plate reader require low volumes, only very little of the ncAA must be used. For this, three strains based on *B. subtilis* K07 were created. The first strain K07-10 expresses the wild type GFP and is therefore the reference for the expression. The second strain K07-11 expresses GFP with an amber stop codon and contains one copy of *pylT*. The third strain K07-12 is equipped with the additional pXT plasmid and four more copies of *pylT* and is used to assess the impact of the *pylT* amounts. These strains were all cultivated in MCSE medium and grown with and without the ncAA. After 20 h, the fluorescence in the was analyzed.

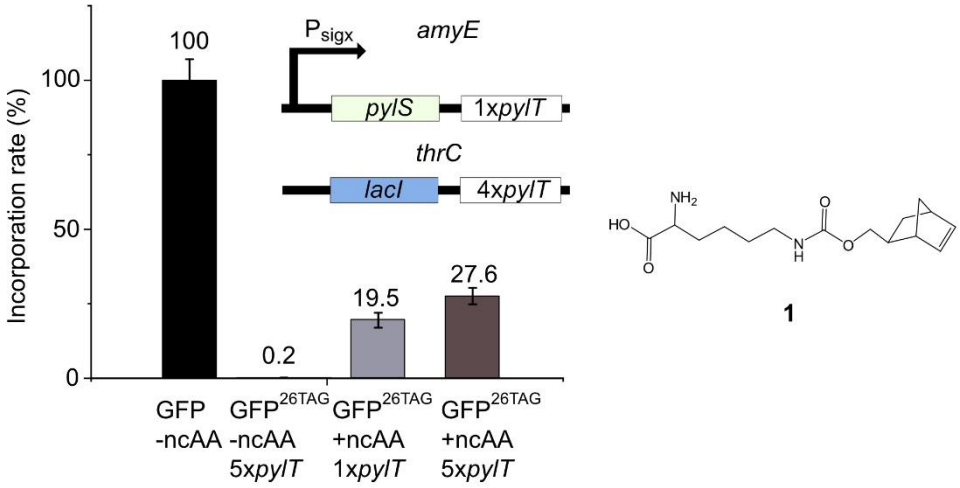

Figure 26: Incorporation rates of 1x *pylT* and 5x *pylT* compared to the wild type GFP
Compound **1**, the ncAA endo-norbornene lysine was incorporated into *gfp*^{26TAG}. The expression was carried out in the plate reader for 20h of cultivation time. The genomic background of the strain K07-11/12 with the genomically integrated elements is shown.

As expected, the positive control, expressing full-length GFP yielded the highest fluorescence signal, which is set to 100%. Upon addition of the ncAA **1** in the strains K07-11/12, an increase in the

fluorescence is visible, suggesting a successful incorporation. In the absence of the ncAA only very weak fluorescence of less than 1 % of the wild type control is observed, indicating that the translational read through is negligible. The construct with five tRNA copies exhibits a further increase in the incorporation rate compared to the strain with only one tRNA copy, resulting in a final incorporation rate of nearly 30%.

These results demonstrate for the first time that amber suppression in *B. subtilis* works with incorporation rates comparable to *E. coli*.

2.2.4.2. Purification and quantification of GFP^{Norb26}

In order to proof the incorporation of endo-norbornene lysine into GFP, the bioorthogonally modified GFP^{Norb26} was expressed and purified from *B. subtilis*.

For this, *B. subtilis* strain $K07^{Amber}$-1 containing the $GFP^{K26Amber}$ expression plasmid #18 was cultivated overnight in 2YT medium at 25°C in presence of 1 mM ncAA **1** (see Figure 26) and induced with 1 mM IPTG. After cell lysis using lysozyme and sonification, the cleared lysate was purified using His SpinTrap columns and the respective fractions were analyzed on an SDS-PAGE gel (Figure 27). Clear elution bands with a high enrichment of GFP^{Norb26} using the His_6-tag could be obtained. Here, already in the lysate fraction a prominent band between 20 kDa and 30 kDa, corresponding to the size of GFP, is visible. The elution fraction shows some minor impurities but the most band corresponds most likely to GFP^{Norb26} with about 2mg g^{-1} cell weight or 20 mg l^{-1} culture volume, which is comparable to yields achieved for the expression of GFP^{ncAA} using amber suppression in expression hosts like *E. coli*.

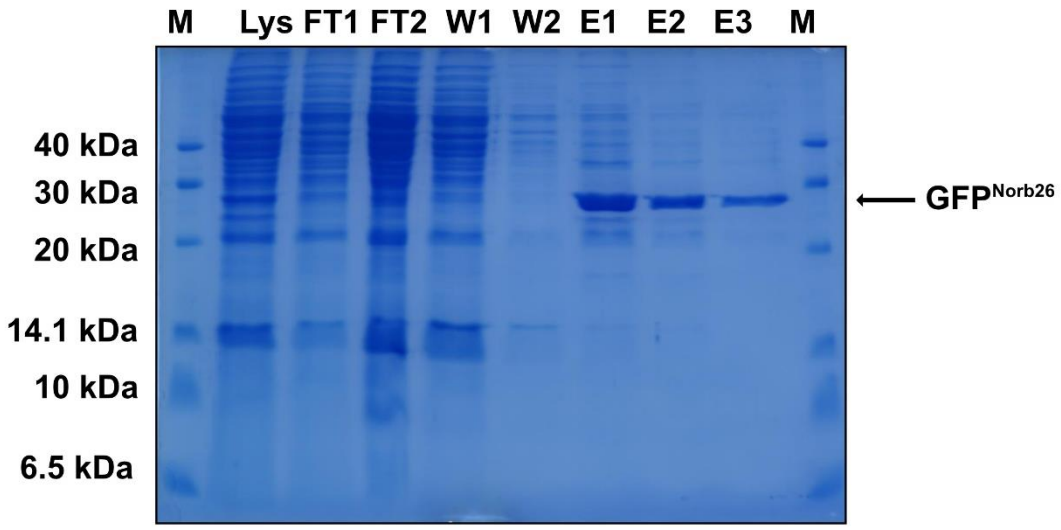

Figure 27: Purification of GFP^{Norb26} from *B. subtilis* strain $K07^{Amber}$-1
Purification was performed using His SpinTrap columns. (M = Marker, Lys = Lysate, FT = Flow Through, W = Wash, E= Elution) The lysate already shows a defined GFP-lane, which was purified using the C-terminal His_6-tag.

2.2.4.3. Verification using mass-spectrometry

The fluorescent measurement showed that the GFP fluorescence is only detectable in the amber codon carrying GFP^{K26TAG} strain if the ncAA is added. To proof the incorporation of the ncAA into GFP electron spray ionization mass-spectrometry (ESI-MS) was carried out. Therefore, the elution fraction E1 from the purification from Figure 27 was mixed with a 10-fold excess of tet-Atto647N **2** and incubated at room temperature for two hours. Both mass-spectra, with and without compound **2** were analyzed. The top of the two mass spectra (Figure 28) represents the GFP^{Norb26} with a mass of 27792 Da (calc. 27796 Da) and the N-terminal cleavage species, where the N-terminal methionine was cleaved off. This is a common observation occurring during the purification from intracellular proteins expressed in bacteria.(119) By incubating the sample with the tetrazine ligand **2**, an increase in the mass by 909 Da is clearly visible, confirming that the ncAA reacted completely, forming the conjugate between the bioorthogonal modified GFP^{Norb26} and the tetrazine ligand **2**. This mass spectrometry measurement verifies the correct incorporation of the ncAA as well as its functionality and solvent accessibility for further reactions.

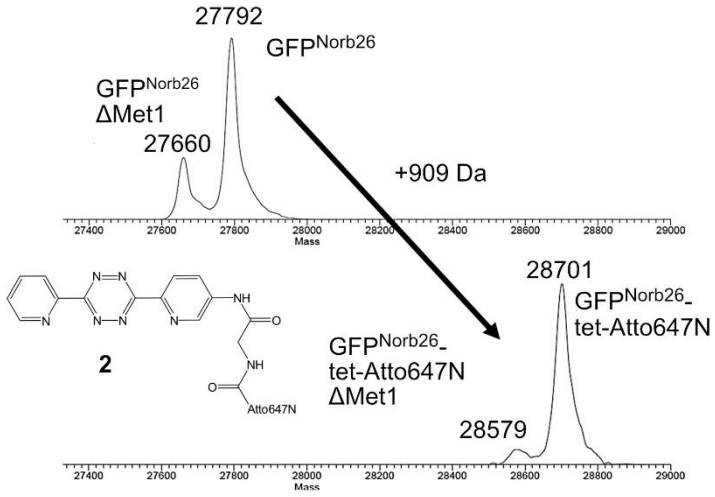

Figure 28: Mass spectra of GFP before and after the reaction with a click reagent
Reaction of GFP containing the ncAA endo-norbornene-lysine **1** at position 26 with the tetrazine ligand tet-Atto647N **2**. Top: Mass spectrum of GFP^{Norb26}, two peaks are visible corresponding to GFP^{Norb26} and GFP^{Norb26} without the N-terminal methionine. Bottom: Mass spectrum of GFP^{Norb26} tet-Atto647N. The mass spectrum shows a clear shift from the unreacted product to the reacted product.

2.2.5. Summary

Taken together, by genomically integrating *pylS*^{Norb}/*pylT* and encoding *gfp* on the replicative plasmid, the plasmid instability, due to the size and the overall expression load, could be overcome. The successful incorporation rates of endo-norbornene lysine into GFP could be further increased from 20% to 30% by the addition of four additional tRNA copies to the expression strain, yielding about 20 mg l^{-1}. Here, one has to keep in mind that the sequence context of the amber codon has also an impact on the suppression rates, which for GFP was investigated. The purified GFP^{Norb26} was verified using ESI-MS revealing that the ncAA is incorporated correctly. Furthermore, the functionality could be proved through the reaction with the tetrazine ligand **2** visualized by a shift in the MS.

2.3. Protein secretion in *B. subtilis*

One of the most important features of *B. subtilis* is its ability to secrete fully functional proteins into its surroundings. For this work, the SEC system as the most effective system with the highest secretion yields was chosen, because the target proteins are all of small size and generally of stable nature. As target proteins, three antibody derived proteins, MAK33-VL a variable light chain against the human creatine kinase, D1.3scFv a single chain variable fragment against lysozyme and Am3-114 an i-Body against the human CXCR4 receptor, were chosen because of the pharmaceutical relevance of this protein group as well as their favorable protein-chemical properties.(120-122)

2.3.1. Expression setup for the secretion of proteins

The expression setup consists of the previously used promoter P_{groES}, followed by the 5' and 3' double stabilized region for the gene of interest. The N-terminal signal peptide is exchangeable by an *Xba*I/*Not*I digest and can therefore be altered. An advantage of the *Not*I restriction site is that its translated protein sequence is translated in this reading frame to a triple alanine AAA that is simultaneously the signal peptidase cleavage site. The extracellular membrane bound signal peptidase recognizes the cleavage site AXA (alanine – random AA– alanine). The target protein is downstream of the cleavage site and can be exchanged with a *Not*I/*Xho*I restriction digest.

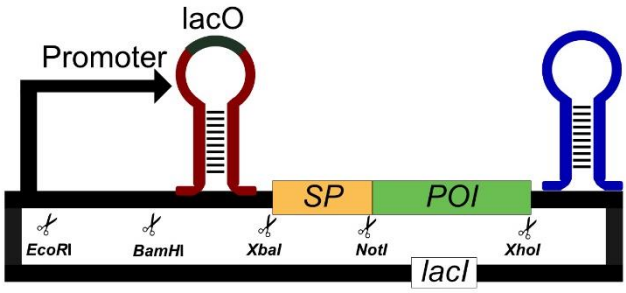

Figure 29: Optimized modular expression plasmid based on the double stabilized pLIKE plasmid #2
Each single component can be exchanged via a distinct restriction endonuclease digest, enabling the screening of different target proteins.

2.3.2. Expression and secretion of MAK33-VL

The target protein chosen to be tested is the variable light chain of the murine antibody MAK33 targeting the human creatine kinase (hmCK, subtype MM).(120, 123) This monoclonal antibody inhibits the hmCK by about 80% and was designed as a diagnostic tool to optimize the diagnostic of myocardial defects in the clinic. The crystal structure of the Fab fragment could be elucidated in 2001 and is shown in Figure 30.(124) The Fab fragment is the antigen binding part of an antibody and can be generated by cleaving an antibody with papaine. Furthermore, this antibody has been thoroughly studied by

several groups throughout the last decade and is known to be a very stable protein, but is expressed as inclusion bodies in *E. coli*.(125-128)

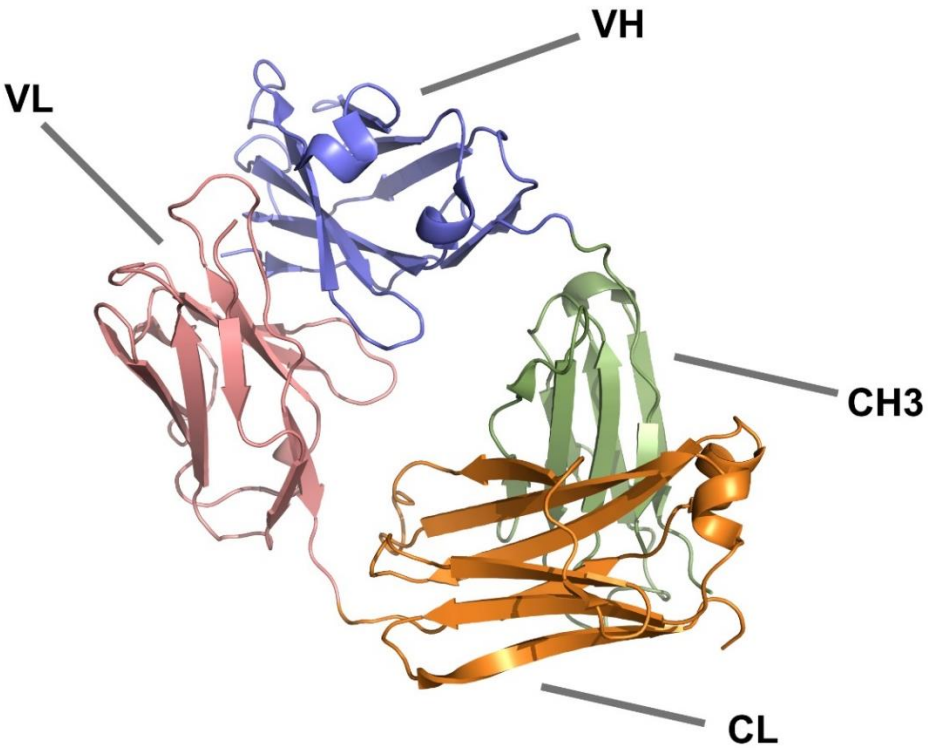

Figure 30: Protein crystal structure of the MAK33-Fab fragment (PDB 1FH5)
The variable light chain (VL) used in this work is colored in salmon.

For the expression construct and initial tests, six signal peptides were chosen which were used in several previous works and showed good secretion efficiencies.(18, 129-131) Thorough PCR optimization had to be performed, as these small signal peptides are in fact too short to be efficiently produced by the high-performance polymerases like Q5 or Phusion and were therefore produced with the slower Taq polymerase. The antibody fragment was cloned into the *Xba*I/*Xho*I digested double stabilized expression vector #2 together with the six signal peptides. The two different affinity tags (His_6- and StrepII-tag) allow different verification methods for western blotting and purification. Here, only positive clones for "AmyE-MAK33-VL" construct as shown in Figure 31 could be obtained.

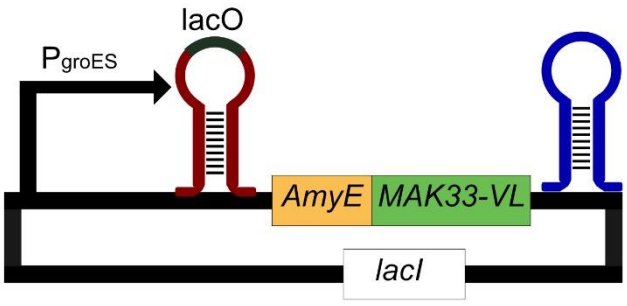

Figure 31: pLIKE-PgroES-5'AmyE-MAK33-VL-3' expression plasmid with a C-terminal His_6- or StrepII-tag
AmyE-MAK33-VL is expressed under control of the strong promoter P_{groES}. The transcribed mRNA is stabilized by secondary structure RNA elements making the expression IPTG inducible.

2.3.2.1. MAK33-VL expression tests

To express MAK33-VL the expression plasmids #19 (StrepII-tag) and #21 (His_6-tag) were transformed into *B. subtilis* K07 creating the strain K07-13 (StrepII-tag) and K07-14 (His_6-tag). Positive clones were verified by single colony PCR and positive clones were pooled. After overnight expression in 2YT medium at 25°C, the supernatant was loaded onto an SDS-PAGE.

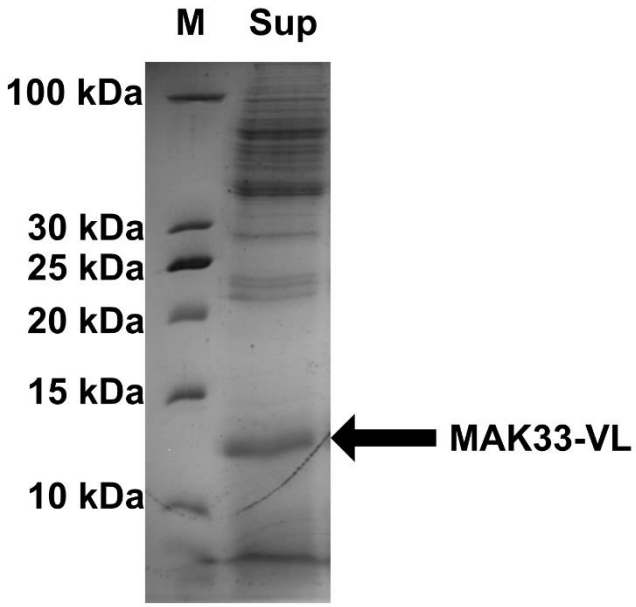

Figure 32: SDS-PAGE analysis of the total supernatant of *B. subtilis* strain K07-14 expressing MAK33-VL-His_6.
A clear band between 10 kDa and 15 kDa indicating a successful secretion of MAK33-VL into the supernatant.

Figure 32 represents the total secreted protein in the supernatant. Most secreted proteins of *B. subtilis* are larger than 20 kDa and only one single band is visible between the marker bands 10 kDa and 15 kDa corresponding most likely to the secreted MAK33-VL protein.

2.3.2.2. Verification of MAK33-VL expression

The visible band in the supernatant can be enriched using affinity chromatography (see Figure 33). To verify is the secreted protein is the target protein and is present with or without the cleaved signal peptide, western blotting and mass spectrometry was performed.

2.3.2.2.1. Western blotting for MAK33-VL-StrepII

For western blotting, the *B. subtilis* strain expressing MAK33-VL with a C-terminal StrepII-tag was expressed at 25°C overnight and purified using StrepTactin Spin columns.

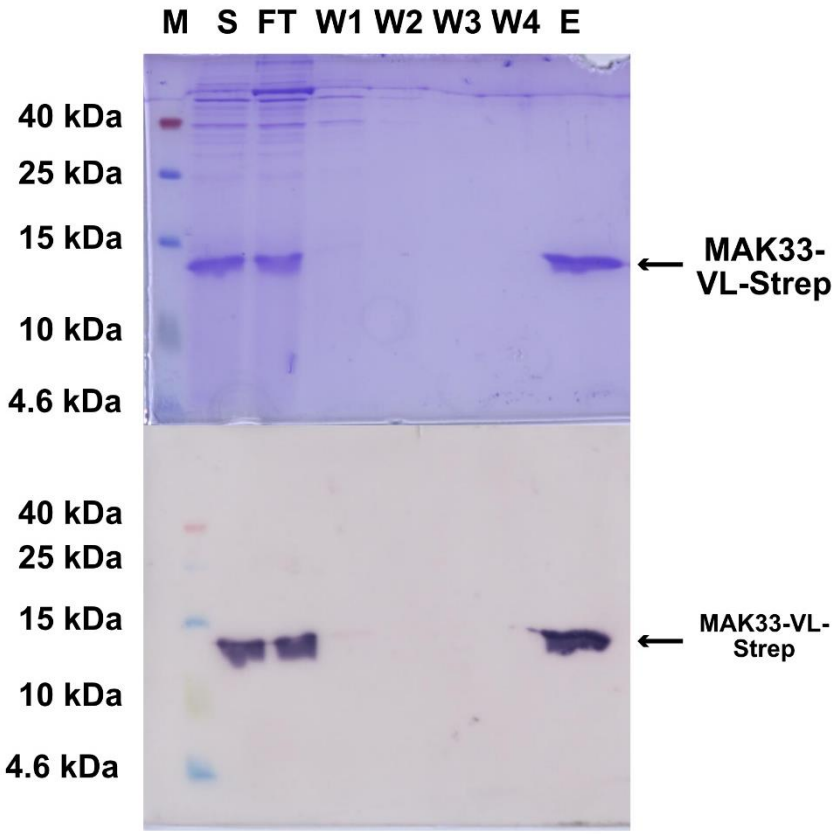

Figure 33: SDS-PAGE and Western Blot analysis of K07-13 expressing MAK33-VL-Strep
Top: SDS-PAGE analysis of the purification of MAK33-VL from the supernatant using StrepTactin Spin columns. (S: Supernatant, FT: Flow through, W: Wash, E: Elution) Bottom: Western Blot using StrepTactin-AP conjugate as the immuno-detection reagent. The corresponding bands of MAK33-VL are marked with an arrow.

The SDS-PAGE analysis already reveals a very strong band in the supernatant, which completely saturates the StrepTactin spin columns. Therefore, a strong signal is still detectable in the flow through. The wash fractions 1-4 only show minor impurities, which are successfully eliminated from the column. In the elution fraction, one clear band of MAK33-VL can be found. By performing a Western Blot against the StrepII-tag using StrepTactin-AP the bands highlighted in the SDS-PAGE show a positive signal and are therefore confirm the presence of a StrepII-tag as present in the expressed MAK33-VL.

2.3.2.2.2. Verification using mass spectrometry

In parallel, the *B. subtilis* strain expressing the His_6-tagged variant of MAK33-VL was cultivated at 25°C overnight. MAK33-VL was purified from the supernatant using His SpinTrap columns and the elution fraction is analyzed by ESI-MS. Interestingly, two different peaks appeared in the mass-spectrometry measurement. The peaks perfectly correspond to MAK33-VL (calc. 12615 Da) and Ala-MAK33-VL (Calc. 12686 Da) with a further N-terminal alanine residue (see Figure 34).

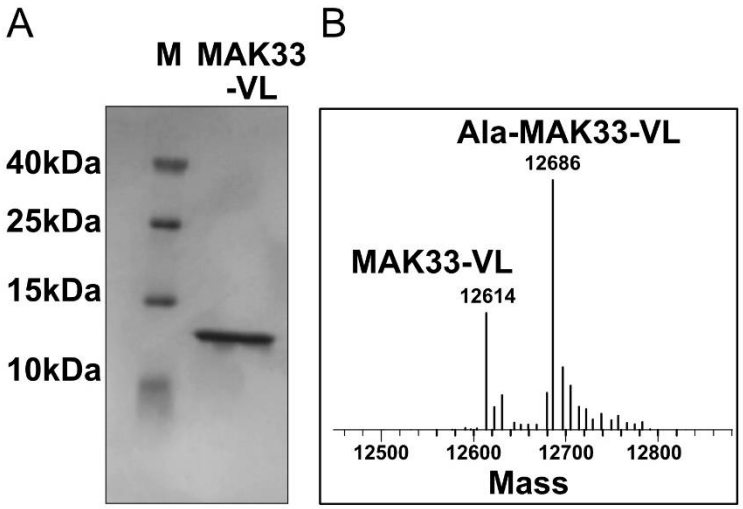

Figure 34: SDS-PAGE analysis and mass spectrum of MAK33-VL
A: Wild type MAK33-VL purified via Ni^{2+}-NTA-affinity chromatography. B: ESI mass spectrometry results of purified MAK33-VL-His_6. MAK33-VL is found in two species, with and without an N-terminal alanine residue.

These two species are the product of the variability of the signal peptidase, which can cleave of the signal peptide directly before or after the last alanine residue of the signal peptide cleavage site as shown in Figure 35.

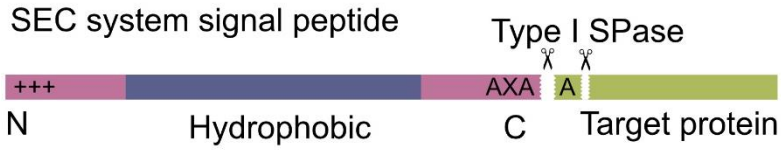

Figure 35: Scheme of a SEC system signal peptide with possible cleavage sites
The two alternative cleavage sites of the Type I signal peptidase in front and after the last alanine of the signal peptide are illustrated through scissors.

Taken together, the variable light chain of MAK33 can be expressed in good yields of approx. 20 mg l^{-} and easily purified from the supernatant using either the StrepII-tag or the His_6-tag. The used AmyE-signal peptide mediates a clean secretion into the supernatant without any residual protein left in the cell pellet. As a further advantage, after only one purification step, the target protein can be found without visible impurities.

2.3.3. Secretion of D1.3scFv

As a second protein to test the secretion system, the single chain variable fragment D1.3scFv, derived from the antibody D1.3 was chosen. This protein is a lysozyme-specific antibody fragment and has previously been successfully produced in *B. subtilis.* (122, 132-134) To verify the protein expression system, D1.3scFv was cloned together with the N-terminal signal peptide *lipA* found in the study into the pLIKE expression plasmid, generating plasmid #23. The plasmid was transformed into *B. subtilis* strain K07 creating K07-15. In a first expression test, the *B. subtilis* strain was cultivated in MCSE

medium and purified the protein from the supernatant using His SpinTrap columns. The SDS-PAGE analysis of the expression culture revealed a successful expression of D1.3scFv. The elution fraction shows two bands, one higher contaminant, which is strongly expressed, and the target protein slightly above the 25 kDa protein marker band. The corresponding Western Blot against the His_6-tag confirmed the exhibits a signal for a His_6-tag at the size of the target protein D1.3scFv, however the overall yields are very low.

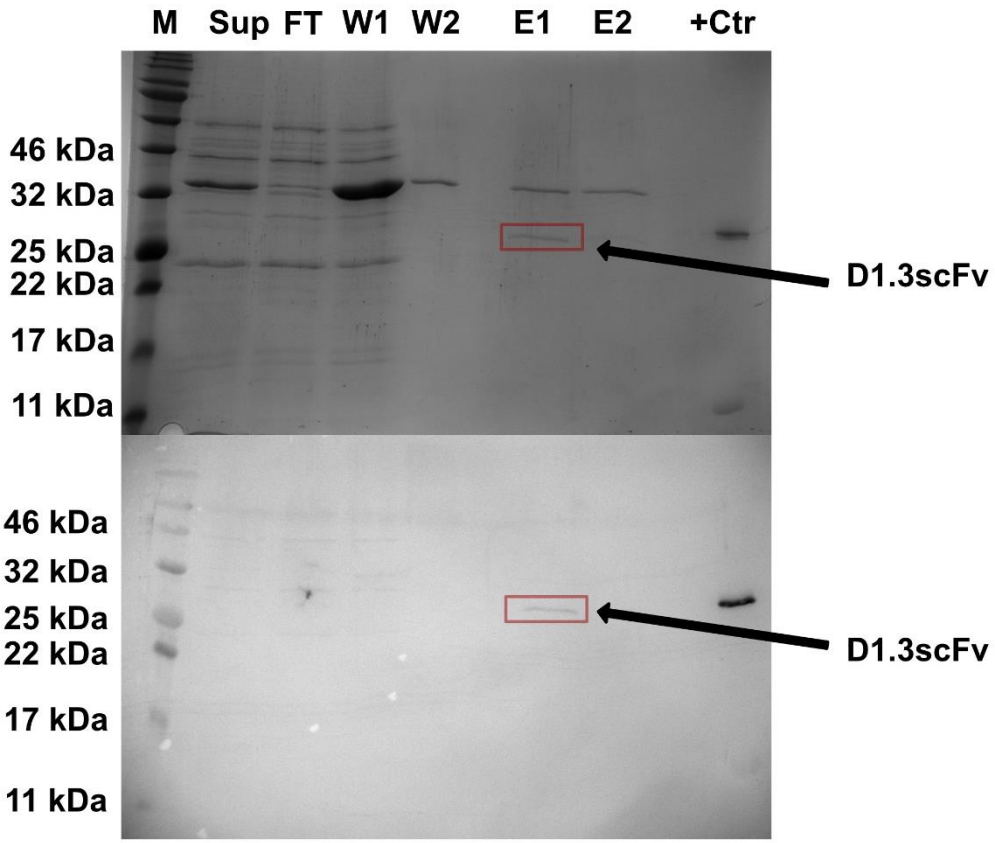

Figure 36: SDS-PAGE and Western Blot of the initial purification of D1.3scFv
Top Image: SDS-PAGE analysis from the purification of D1.3scFv using His SpinTrap columns (Sup: Supernatant, FT: Flow through, W: Wash, E: Elution). Bottom: Corresponding Western Blot against the His_6-tag. His_6-tagged GFP is used as a control.

As the expression media and the temperature can both have a large impact onto the overall protein yields the secretion of the target protein was tested in four different media, MCSE as a defined media, and LB, 2YT and TB as complex media, at three different temperatures (37°C, 30°C and 25°C) overnight. To further evaluate the secretion efficiency, both supernatant and cell pellet were analyzed by an SDS-PAGE and checked by western blotting against the His_6-tag.

Figure 37 represents the SDS-PAGE analysis (top) of the total 12 expression tests. The MCSE medium exhibits generally a lower variety of secreted proteins with one very prominent but unknown band at about 32 kDa. The complex media generally expresses a broad range of different proteins, which is comprehensible due to the different nutrients present in the media. The Western Blot against the His_6-tag clear visible bands at slightly above the 25 kDa protein marker band corresponding to our target protein detected.

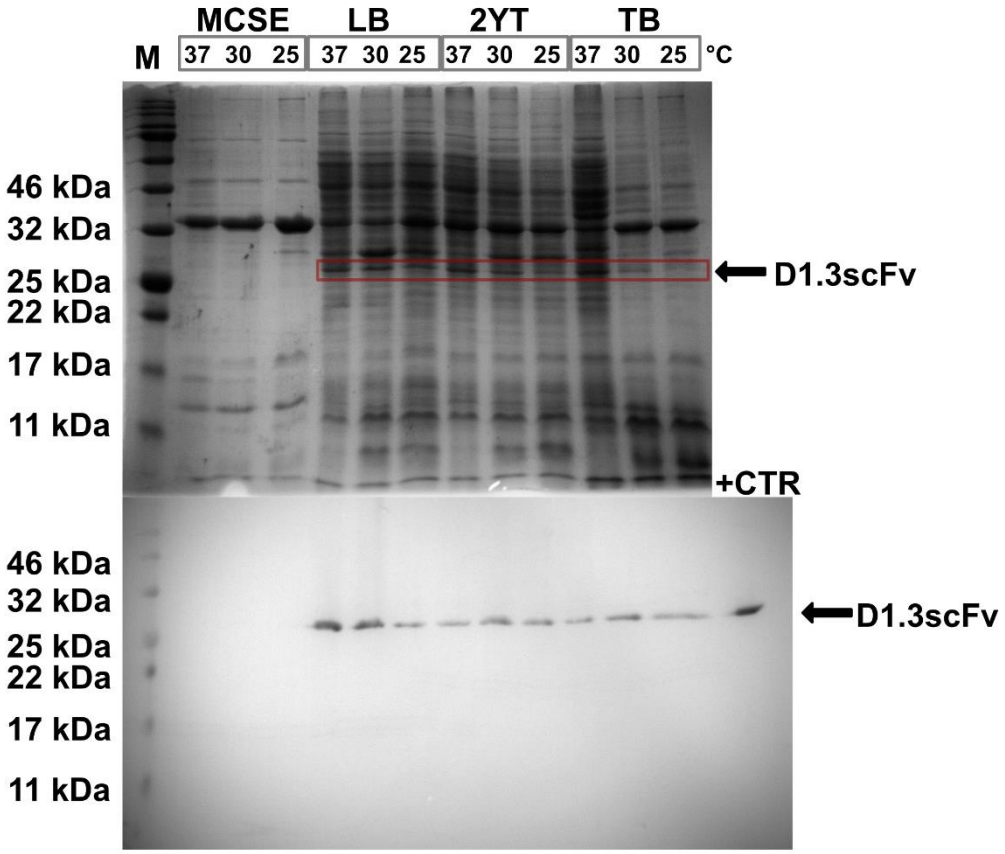

Figure 37: Supernatant from a screening for suitable expression conditions for D1.3scFv
D1.3scFv was expressed in different culture media with varied expression temperatures. Top: SDS-PAGE analysis of the total supernatant secreted by *B. subtilis* strain K07-D1.3scFv. The secreted band corresponding to D1.3scFv is marked with a red box. Bottom: Western Blot against the His_6-tag of the secreted D1.3scFv.

To assess the secretion efficiency, the pellet of the expression cultures were analyzed in parallel (see Figure 38). Therefore, the cell pellet corresponding to 2 ml culture volume was boiled in SDS loading buffer loaded onto an SDS-PAGE. Interestingly, the visible intensities of the proteins greatly vary depending on the media and the expression temperatures implying that the overall cell densities already differ. However, the Western Blot against the His_6-tag revealed only very weak bands in the cell pellet, indicating that the target protein D1.3scFv is secreted predominantly into the supernatant, with residual protein left within the cytosol.

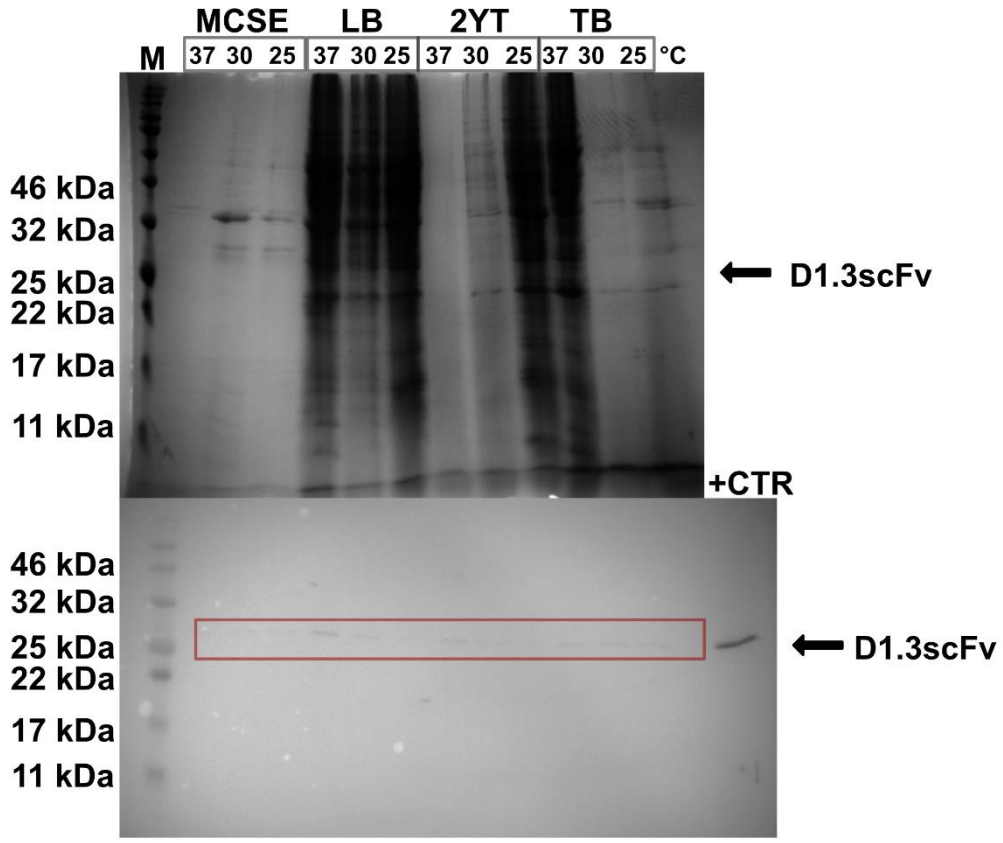

Figure 38: Cell pellet fraction from screening for suitable expression conditions for D1.3scFv
D1.3scFv was expressed in different culture media with varied expression temperatures testing for the secretion efficiency. Top: SDS-PAGE analysis of the corresponding cell pellet. Bottom: Western Blot against unscreted D1.3scFv. The unsecreted fraction is marked with a red box.

In addition, K07-15 expressing D1.3scFv-His_6 was cultivated in 2YT medium at 25°C overnight and the target protein was purified from the supernatant using Ni^2-NTA affinity chromatography (Figure 39 A). The verification using ESI-MS was successful and shows as previously for MAK33-VL two possible SPase cleavage products. In this case, the cleavage occurs before (calc. 26309 Da, found 26313 Da) or after the second alanine (calc. 26380 Da, found 26376 Da) upstream of the target protein and not as seen before with MAK33-VL up/downstream of the last alanine residue. However, compared to the MS of MAK33-VL, the overall spectrum of D1.3scFv exhibited a generally low signal, which could also explain the other peaks found in the spectrum.

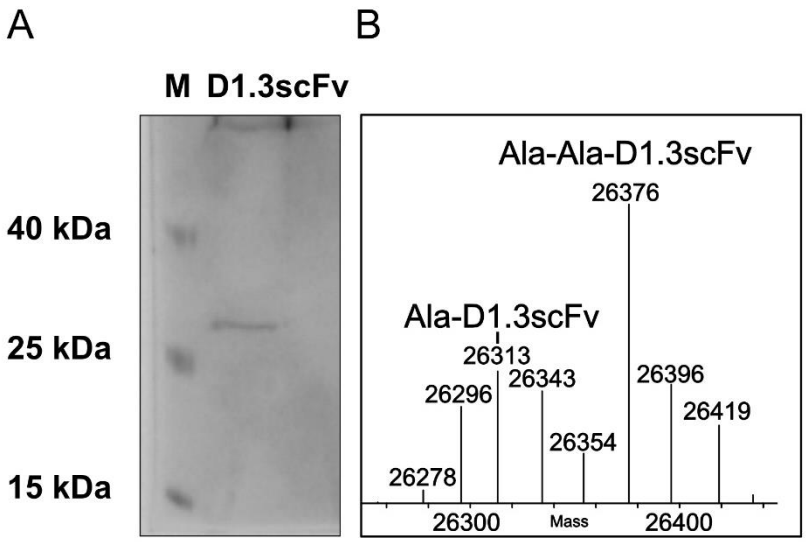

Figure 39: SDS-PAGE and mass spectrum of purified D1.3scFv
A: Wild type D1.3scFv-His_6-tag was purified using Ni^{2+}-NTA affinity chromatography analyzed by an SDS-PAGE. B: ESI-mass spectrum of the secreted D1.3scFv. D1.3scFv is secreted with one or two N-terminal alanine residues.

With this, it can be concluded that D1.3scFv is secreted successfully into the supernatant and can be purified using its His_6-tag.

In general, one can conclude that the complex media are again better for the overall secretion of the target protein as shown before with MAK33-VL. The amount of the secreted D1.3scFv fragment is in a fair range and demonstrates the great flexibility of the expression system.

2.3.4. Secretion of Am3-114

As a third target protein, the single chain antibody Am3-114 was chosen. Am3-114 is an i-body based on the human NCAM-1 (neural cell adhesion molecule 1), an Ig-domain as shown in Figure 40. An i-body is a fully functional single-domain antibody.(121) The target of Am3-114 is the human C-X-C chemokine receptor type 4 (CXCR4), which is a target candidate for the treatment of idiopathic pulmonary fibrosis. CXCR4 is a human G-protein coupled receptor and is expressed in many different cell types where it is involved in the mobilization of cells, for example for wound healing.(135) However, it is also present in tumor cells, enabling them to migrate into new tissues.(136) Additionally, it is also an important receptor for the docking and internalization of the human immunodeficiency virus (HIV) in human cells.(121, 137, 138)

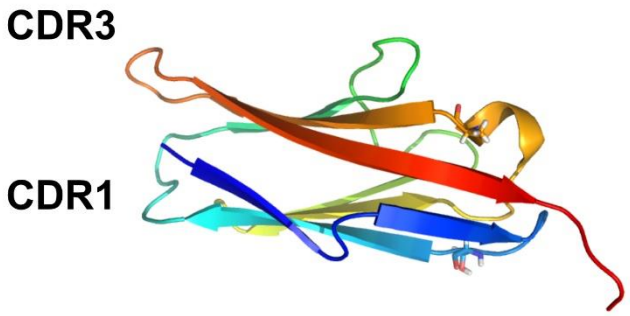

Figure 40: Structure of the NCAM I domain (PDB 5AEA), the scaffold for the development of Am3-114.

In the described publications by Griffiths et al., the complementary-determining regions 3 loop (CDR3) was replaced by a larger loop and affinity maturation was carried out to fit in the deep ligand pocket of the CXCR4 receptor.(121, 137) After several rounds of maturation, Am3-114 was derived with the highest affinity for CXCR4 with low off-target effects. To express and secrete the target protein in *B. subtilis*, the sequence of MAK33-VL in plasmid #19 was replaced with the sequence for Am3-114 with a*myE* as the N-terminal signal peptide. The created plasmid #28 was transformed into the *B. subtilis* PG10, yielding strain PG10-1. PG10 is a genome-reduced strain where 35% of the genome were deleted and which shows improved yields for the secretion of (difficult) proteins.(37, 38) Three positive clones were grown individually at 25°C in 2YT medium overnight and their protein fractions were analyzed by SDS-PAGE and western blotting. In the supernatant of all three positive clones, a rather diffuse band is detectable, which also exhibits a positive signal in the Western Blot against the StrepII-tag. However, the three pellet fractions show a strong band at the same height as in the secretion fractions indicating that the secretion is not as efficient as previously found with MAK33-VL or D1.3scFv.

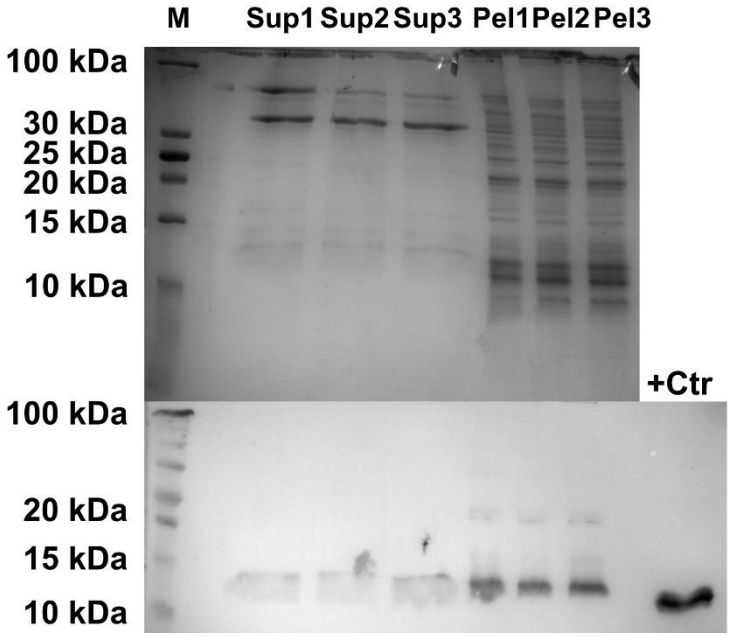

Figure 41: SDS-PAGE and Western Blot on three clones expression Am3-114
SDS-PAGE analysis (top) of three individual positive clones expressed in 2YT medium. The supernatant (Sup) as well as the cell pellet (Pel) of each clone were analyzed. The bottom image shows the Western Blot against the StrepII-tag of Am3-114. MAK33-VL was used as a positive control.

To ensure that the protein found in the supernatant is the secreted target protein without the signal peptide, the supernatant was purified using StrepTactin Spin columns and ESI mass spectrometry was carried out.

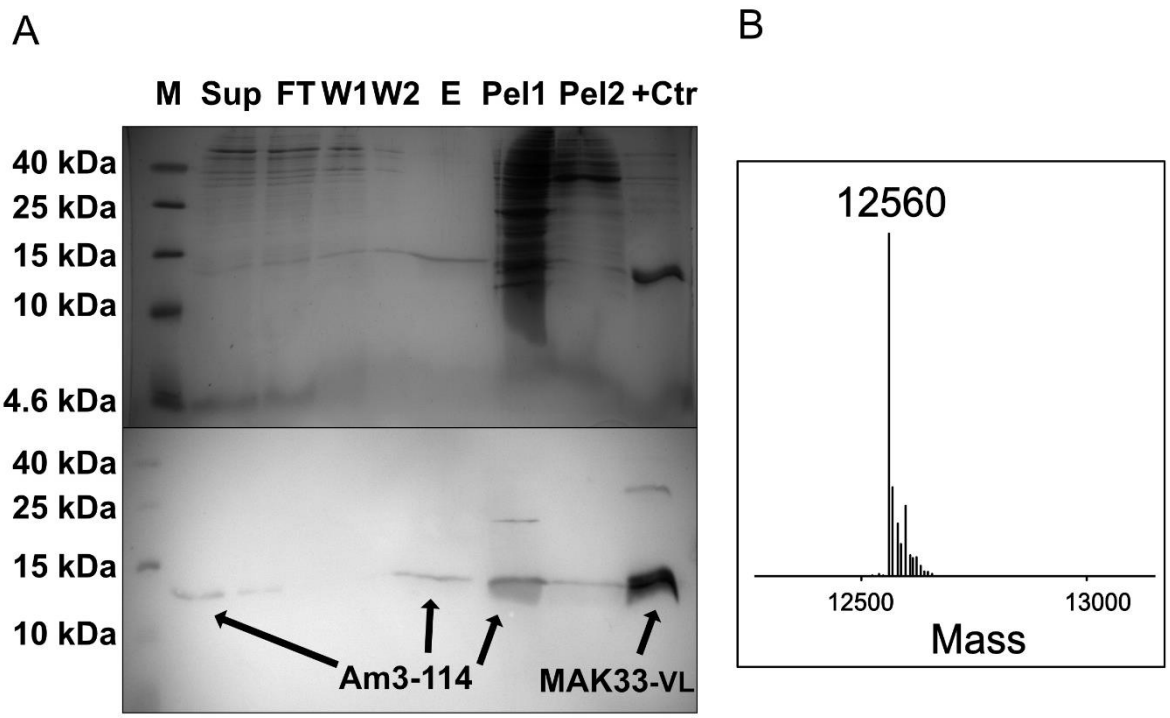

Figure 42: SDS-PAGE and Western Blot analysis of purified Am3-114 with corresponding mass spectrum
A: SDS-PAGE analysis of the purification of Am3-114 using StrepTactin Spin columns. (M: Marker, Sup: Supernatant, W: Wash, E: Elution, Pel: Pellet) MAK33-VL is used as a positive control. Bottom: Western Blot against the StrepII-tag using Strep-Tactin-AP. B: ESI mass spectrum of the purified Am3-114 WT protein.

Figure 42-A shows the successful purification of Am3-114 from the supernatant using the StrepII-tag. In the supernatant and elution fraction from the Spin columns, a positive signal for the StrepII-tag can be detected. The cell pellet fraction is the pellet left after separation from the supernatant. Before loading on the SDS-PAGE, the pellet was boiled for 10 minutes at 90°C. In this fraction, most of the protein detectable, with no visible shift in the size indicating that the signal peptide is not present. Otherwise, a strong shift by approximately 3 kDa would be expected. The elution fraction from the supernatant was subsequently used for the verification by ESI-MS. The mass spectrum (Figure 42-B) of the purified Am3-114 wild type protein shows the desired mass of 12560 Da (calc. 12565 Da), verifying that the protein is secreted in the supernatant with the cleaved signal peptide. However, the overall yields are not as high as with both previously secreted proteins, making it not possible to calculate total yields of the target protein as most of the AM3-114 is still found in the cell pellet.

To increase the secretion efficiency, several alternative signal peptides were produced and tested but despite multiple trials, no positive transformants in *E. coli* could be obtained. The signal peptides itself were cloned into the expression vector and replaced the previous signal peptide *amyE* but in all sequenced plasmids premature stop codons in the ten first base pairs after the translation start point were found. In addition, more stringent *E. coli* strains like XL10 gold could not resolve this issue, indication a more general problem with the protein expression of Am3-114 in *E. coli*, which could result from the fact that the promoter of the expression construct P_{groES} is slightly active in *E.coli,* but is already too toxic for the cells.

Taken together, although Am3-114 seemed to be an interesting target protein, there is a more general issue with the expression of this protein, which would require substantial effort which in turn is not the overall goal of this work as MAK33-VL can be expressed and secreted in sufficient yields.

2.3.5. Summary

The designed expression setup allows the exchange of all relevant parts such as the protein of choice the signal peptide or the desired promoter. This flexibility is evident from the three different secreted proteins, two single chain antibodies and a single chain variable fragment. Here, the yields ranged from about 20 mg l^{-1} (MAK33-VL) to about 10 mg l^{-1} (D1.3scFv). The yield for Am3-114 was too low and could only be detected by western blotting. As from all three proteins, the variable light chain of MAK33, secreted using the AmyE signal peptide, showed the best overall results regarding the expression yields and the secretion efficiency. Hence, this protein was chosen for the next step, the combination of the efficient amber suppression system and the secretion of proteins into the supernatant.

2.4. GFP fusion constructs to assess ncAA incorporation

2.4.1. State of research on the genomic context of target sites

As described in the introduction on amber suppression, the genomic context upstream and downstream of the introduced amber stop codon is important for the incorporation rates and the selection of the target site. Until now, three groups worked on the effect of the codon context neighboring the amber stop codon, Pott et al. (2014)(111), Xu et al. (2016)(110) and Schwark et al. (2018)(109). In the first investigation by Pott and co-workers, both the tyrosine-based system from *M. jannaschii* as well as the pyrrolysine-based system from *M. mazei* were tested. Their setup consisted of a randomized sequence (6bp each) upstream and downstream of the amber stop codon, followed by a downstream reporter gene – *gfp* or the *chloramphenicol acetyltransferase*. Their findings showed that both the tyrosine as well as the pyrrolysine based systems exhibit a high preference for purine bases upstream and downstream of the amber stop codon. In contrast, Xu and co-workers focused on three bases upstream and downstream of the amber stop codon using the pyrrolysine based amber suppression system. As a reporter they used the β-Galactosidase in *E. coli*, which is generally not as precise as a fluorescence readout. Here, a strong preference for adenosine at the +4 position (with TAG as +1 to +3) was observed, while a variability in the subsequent positions occurred. In the recent work by Schwark and co-workers from 2018, they again looked at the sequence context but directly evaluated the natural sequence context in *gfp* while mutating each single tyrosine to an amber stop codon and incorporated tyrosine again via the tyrosine system from *M. jannaschii*. Their findings also showed that +4 sites containing an adenosine (or more general a purine base) exhibit the best incorporation rates.

Taken together, according to these reports, a target site with high amber suppression rates is characterized by the presence of many neighboring purine residues and most importantly by the presence of an adenosine at the +4 position.

Therefore, the sequence of MAK33-VL was analyzed for positions, which are purine rich, for example with a lysine or asparagine upstream or downstream of the amber stop codon, but at the same time do not interfere with the binding of the protein to its target, the hmCK. This search revealed Ans33, Leu41 and Tyr51 as potentially suitable sites due to their sequence context and their localization in unstructured loop regions.

2.4.2. C-terminal GFP fusion constructs and analysis of ncAA incorporation at different target sites

In order to analyze amber suppression at the identified target sites Asn33, Leu41 and Tyr51, MAK33-VL C-terminally fused to GFP was used. This has on the one hand the advantage that the amount of the ncAA needed to evaluate if a certain incorporation site is functional is reduced to about 1 mg as only small culture volumes are required. But more importantly, as GFP is located at the C-terminus of the construct, a direct fluorescence readout appears upon incorporation of the ncAA. This direct readout requires no further protein purification using affinity chromatography, detection via SDS-PAGE analysis or western blotting to confirm but only a plate reader to perform the assay overnight. A schematic of the construct is shown in Figure 43. Both proteins are connected via a 15 amino acid glycine-serine $(GGGGS)_3$ linker providing the flexibility for both domains to fold correctly. The three target sites are marked in orange, all facing lateral of the protein and located in loop regions where binding to the antigen should not be influenced.(127)

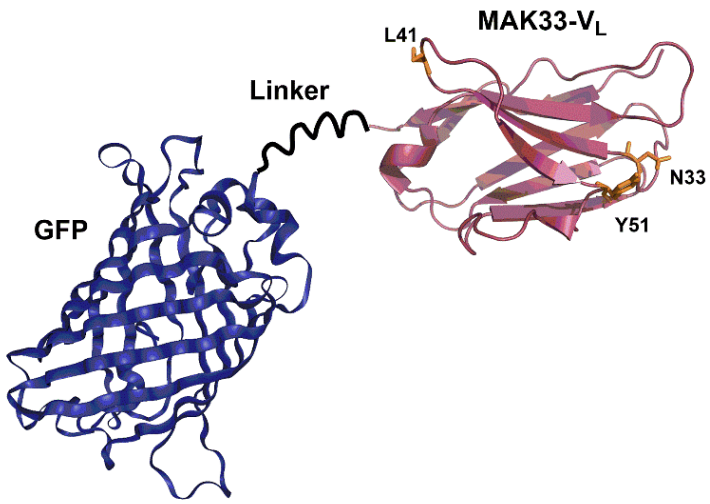

Figure 43: Scheme of the GFP-fusion construct with the possible incorporation sites
The sites Asn33, Leu41 and Tyr51 are marked in the figure. MAK33-VL is located at the N-terminus of the construct and connected to GFP via a glycine-serine linker. (PDB GFP 1EMA; PDB MAK33-VL 1FH5)

The expression plasmids #11-13 were created as described in section 4.3.4.4. and transformed into $K07^{Amber}$. The three *B. subtilis* strains $K07^{Amber}$-2/3/4 containing the target sites $Asn33^{TAG}$, $Leu41^{TAG}$ and $Tyr51^{TAG}$ were grown in MCSE medium and, after reaching an OD of 1.0, induced with 1mM IPTG and optionally supplemented with 1 mM ncAA. The fluorescence measurement was carried out in the plate reader at 37°C under shaking overnight. Figure 44-B shows the fluorescence after 18 hours of measurements.

The site Asn33^{TAG} shows a clear increase in the fluorescence compared to the control without the ncAA. However, Leu41^{TAG} and Tyr51^{TAG} exhibit no increase in the fluorescence in presence of the ncAA at all, despite a suitable sequence context.(109-111)

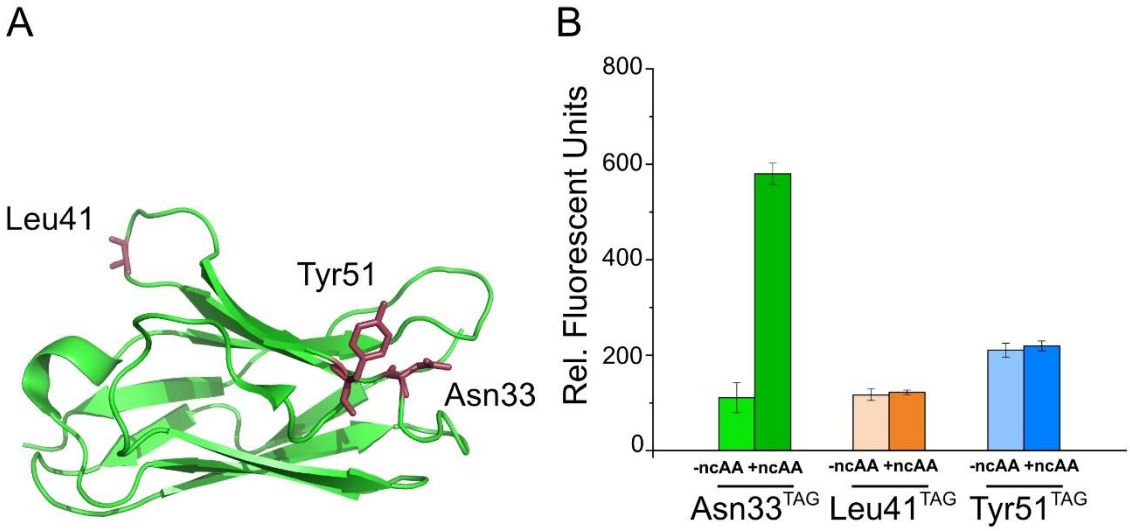

Figure 44: Location and successful identification of Asn33 as a target site
A: Location of the three target sites in MAK33-VL. The CDR loops are located at the right sind of the protein, distant to the tested incorporation sites. (PDB 1FH5). B: Normalized GFP fluorescence of the three strains in presence and absence of the ncAA after 18h of growth in the plate reader.

This is most likely due to the difference in the expression hosts used for the protein expression. Furthermore, the fusion construct is secreted in its active confirmation into the supernatant and exhibits its fluorescence there as the supernatant without the cells showed nearly the same fluorescence values. This is in fact only possible because the used GFP folds slower than super-fold GFP, which would actually be stuck in the membrane as it folds before successful translocation.

Since Asn33^{TAG} appears to be a suitable target site in the sequence of MAK33-VL, it was introduced in MAK33-VL lacking the C-terminal GFP-fusion, creating plasmid #22.

2.4.3. Expression of MAK33-VL^{Norb33} in K07^{Amber}

To express MAK33-VL^{Norb33}, plasmid #22 was transformed into the strain K07^{Amber} creating strain K07^{Amber}-5. The strain was grown in 2YT medium to an OD of 1.0. After addition of the ncAA and further growth for one hour at 37°C, the expression was carried out with induction by IPTG at 25°C overnight. However, after following the purification protocol as performed with the wild type MAK33-VL only very low yields of MAK33-VL^{Norb33} could be obtained.

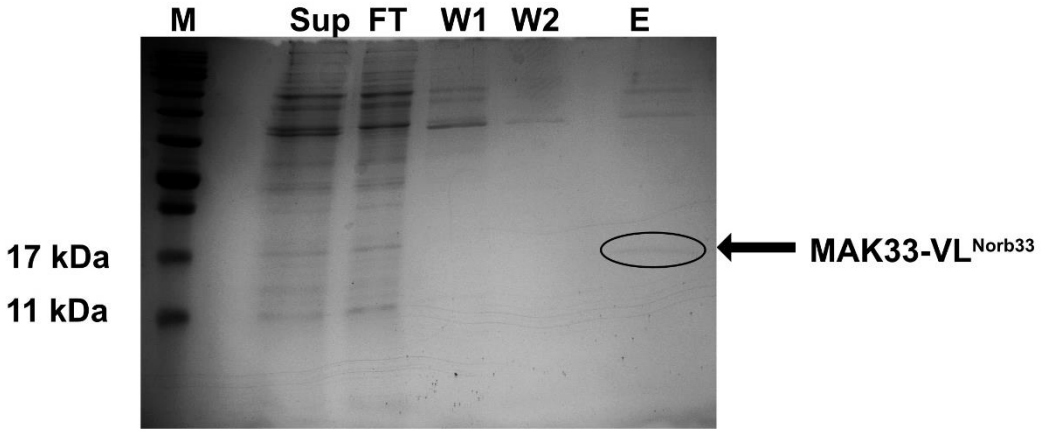

Figure 45: Successful but weak purification of MAK33-VL^{Norb33}
SDS-PAGE analysis of the purification of MAK33-VL^{Norb33} using His SpinTrap columns from the supernatant (Sup: Supernatant, FT: Flow through, W: Wash, E: Elution). Only a very weak band corresponding to MAK33-VL^{Norb33} is visible in the elution fraction and marked with an arrow.

Furthermore, compared to previous expression tests, the overall secretion as shown in the supernatant fraction was generally lower.

To test if the weak supernatant is an artefact from the solubilization of the precipitated supernatant, the *B. subtilis* strain K07^{Amber}-5, previously transformed with the expression plasmid #22, was grown overnight in LB medium. The expression was carried out in 2YT medium supplemented with the respective antibiotics and grown to an OD_{600} of 1.0. At this point, the culture was split into two expression cultures, one was induced with the ncAA and IPTG, the other one only with IPTG. After expression overnight at 25°C, the supernatant as well as the cell pellets were analyzed via an SDS-PAGE. The comparison of the secreted and intracellular proteome revealed that the secretion is significantly reduced in the presence of the ncAA, while no differences in the cytosolic fractions can be observed.

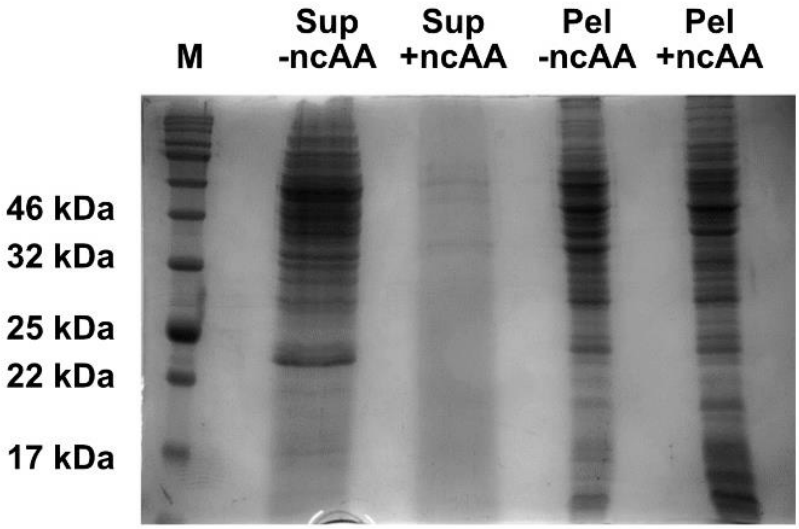

Figure 46: The addition of the ncAA decreases the total secreted proteins
SDS-PAGE analysis of the extracellular (Sup) and intracellular (Pel) proteins of *B. subtilis* strain K07^{Amber} with and without added ncAA in an overnight cultivation (2YT, 25°C).

The same effect on the secreted proteins can be seen in the PG10^{Amber} strain (see Figure 52), which partially rules out possible unknown effects by secondary metabolite pathways or from dispensable regions. This indicates that the SEC system might be impacted by the presence of the amber suppression system and the ncAA.

2.5. Engineering the SEC system to reconstitute the secretion

As shown in the previous section, when using the strain K07^{Amber} and PG10^{Amber}, the overall secretion is drastically reduced if the ncAA is added to the medium. A detailed analysis of the *B. subtilis* genomes of 168, PY79 (parent strain of K07) and PG10 revealed that 3 of the 5 genes (*secA*, *secG* and *secY*) of the SEC system carry an amber stop codon and are therefore also a target of the amber suppression system and translational read through. The products of the three genes are shown in Figure 47 below.

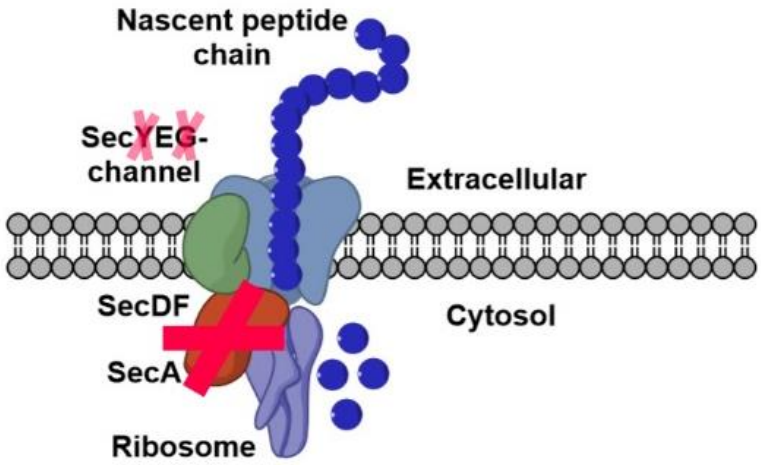

Figure 47: SEC system of *B. subtilis* with marked amber stop codons in the corresponding genes
The SEC system consists of a transmembrane channel formed by SecY, SecE and SecG, the accessory protein SecDF and the ATPase SecA. The ribosome is guided to the SecYEG channel by the signal recognition particle FtsY (not shown) and translates the nascent peptide chain directly through the channel.

Taking the incorporation rates of approximately 30%, which should also apply for each of the three expressed SEC system proteins, the secretion capacity should be reduced by at least 70% which is in line with the visible reduction in protein secretion in Figure 46. To tackle this and increase the yields of the modified target protein, the secretion capacity has to be reconstituted.

One possibility to overcome the induced translational read through is to encode the genes for *secA*, *secG* and *secY* with a different stop codon on a plasmid. However, this would require selection with a second antibiotics, which could impact on the growth rate. Moreover, it was previously shown that the overexpression of the SEC system proteins can even inhibit the overall secretion.(21) Therefore, ideally the protein expression should be regulated as in the native genomic context.

In order to exchange the amber stop codons of *secA, secG* and *secY* to the ochre stop codon, the genome of the *B. subtilis* strain needs to be edited. A genome engineering method revolutionizing the field of biology in recent years is the CRISPR/Cas9 system (Clustered Regularly Interspaced Short Palindromic Repeats).(139) The CRISPR/Cas9 system is a bacterial defense system against invading

nucleic acids and mobile genetic elements. The Cas9 endonuclease activity is here controlled through so-called crRNA and tracrRNA. Upon base pairing between the crRNA and the complementary target DNA, the Cas9 endonuclease cleaves the target DNA, resulting in a site-specific double strand break. Here, cleavage by Cas9 can only occur if the protospacer-adjacent motif (PAM) next to the target site consisting of three nucleotides with the sequence NGG is present. Since Cas9 is functioning not only in bacteria but also in eukaryotes and mammals, the system can be exploited for genome modification, taking advantage of homology directed repair (HDR) and non-homologous end joining (NHEJ). (140, 141)

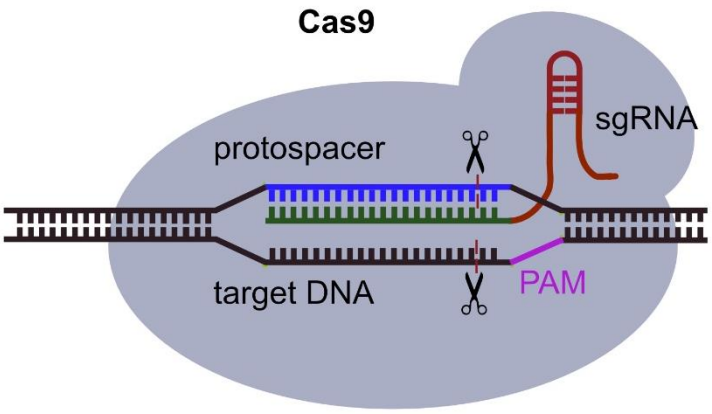

Figure 48: Mechanism of Cas9-mediated DNA cleavage.
crRNA and tracrRNA recognize the protospacer and the PAM of the target DNA. Upon correct recognition, the DNA is cleaved and a double strand break is produced.(140)

In the last years, this system was engineered to be easily used as a genetic manipulation tool, which enables the precise modification, knockdown or knockout of a target gene.(140, 141) The engineering of CRISPR/Cas9 included modifying the targeting RNAs which were fused together to a single-guide RNA (sgRNA) whereby only one RNA has to be expressed. In this work, the system was used to target the genomic amber stop codons in *secA*/*Y*/*G* and modify them to an ochre stop codon to overcome the influence of the ncAA which is incorporated instead of an amber stop codon. To target an amber stop codon of the SEC system, a suitable sgRNA is designed as described in the publication. The sgRNA guides the Cas9 to the target site on the genome where the Cas9 performs a double strand break (DSB). This DSB can be repaired via two different pathways, Non-homologous end joining (NHEJ) and Homologous recombination (HR). NHEJ is mainly present in eukaryotic cells while HR is performed in bacteria. For HR, an intact DNA strand is required which functions as a template for repair of the DSB.

2.5.1. Stop codon exchange via CRISPR/Cas9

The plasmid pJOE8999 encoding for the CRISPR/Cas9 system was kindly provided by Joseph Altenbuchner from the University of Stuttgart and is a typical *E. coli*/*B. subtilis* shuttle plasmid. For

amplification and cloning in *E. coli*, it contains a pUC origin of replication and a kanamycin resistance.(142) The *B. subtilis* part of the plasmid consists of a temperature sensitive origin of replication, rep pE194^{ts}.(143) This origin of replication is derived from the plasmid pE194, an isolate from an antibiotic-resistant *Staphylococcus aureus* strain, that can only function properly at temperatures below 37°C. Additionally, the plasmid provides a mannose inducible *cas9* gene from *Streptococcus pyogenes* and a synthetic *B. subtilis* promoter P_{vanp} that controls the expression of the sgRNA. Next to the sgRNA, a DNA template for subsequent repair after the DSB has to be provided. In *B. subtilis*, this DNA template needs to be at least 70 bp long, but longer DNA templates increase the repair efficiency, therefore about 1000 bp are taken as homologous regions as it was previously performed using genome-integrating plasmids.(144) This intact DNA strand is also provided on the pJOE8999 plasmid. As the goal of this part of the work is to exchange the amber stop codon to an ochre stop codon (TAG→ TAA) only a single base change would be enough for an effect. To be able to perform a "pre-screening" for positive clones, the sequence was not only altered from TAG to TAA but a *Hind*III cleavage site was also introduced as part of the ochre stop codon as shown below (Figure 49).

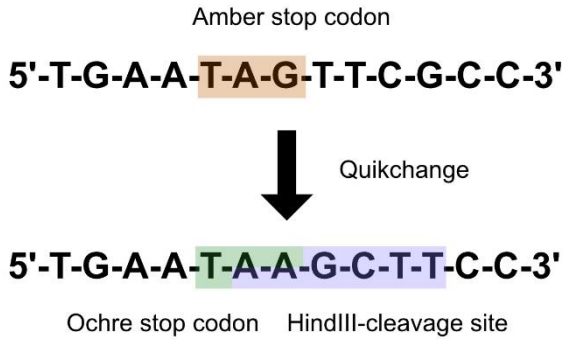

Figure 49: Scheme of the mutations introduced via a QuikChange PCR.
The *Hind*III cleavage site enables the fast screening for positive clones after the modification by CRISPR/Cas9.

The introduction of the *Hind*III cleavage site has several advantages. First, by altering several DNA bases, the plasmid itself is not targeted by the sgRNA as the number of mismatches is too high. Second, when screening the colonies via a single colony PCR, the PCR product can be cleaved with *Hind*III, which yields two DNA fragments upon positive mutation from TAG to TAA. Third, *Hind*III-HF is active in the Onetaq DNA polymerase buffer, hence *Hind*III can directly cleave the PCR product from the single colony PCR in the reaction buffer without a previous purification step.

A further advantage of this whole system is that the pJOE8999 plasmid can be efficiently cured from the cells. After the mutation was performed, the cells are incubated overnight at 50°C (without antibiotic selection). At this temperature, the ori pE194^{ts} is not active and the plasmid be lost during cell division. After this, the pJOE8999 plasmid can be transformed again whereby a further gene can be targeted.

2.5.2. Single-base exchange of *secA* and *secG*

To target the *secA* gene, the plasmid #29 (pJOE+secA_TAA+sgRNA_secA) was transformed into the $PG10^{Amber}$ strain. Positive clones were selected with $LB^{Kan+Manose}$ plates to induce the mannose-dependent promoter for *cas9* and incubated at 30°C where the temperature sensitive origin of replication is active. The grown clones were screened via a single colony PCR using the Onetaq DNA polymerase. One-half of the PCR product was digested with *Hind*III-HF and the corresponding bands analyzed by an agarose gel. Positive clones, visible by a cleavage of the PCR product into half as shown in Figure 50 were selected and plated onto LB^{0} plates at 50°C overnight and additionally at 42°C for one more day. To test if the plasmid is lost, the cells are plated onto LB^{Kan5} plates at 30°C overnight where growth should only happen if the plasmid is still present.

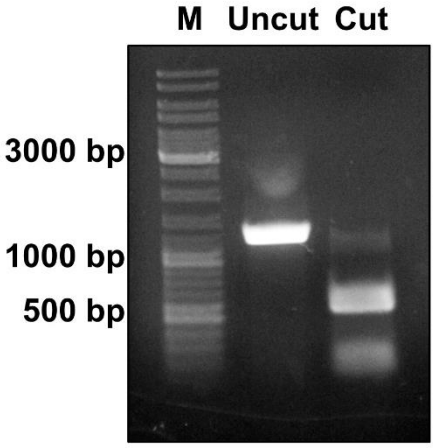

Figure 50: Single colony PCR on the strain PG10 after transformation with plasmid pJOE+secA_TAA_sgRNA_secA (#29). After plasmid curing, the genomic region of *secA* was amplified, and the PCR product optionally digested with *Hind*III. The "Cut" lane shows the desired cleavage of the PCR product by the introduced *Hind*III cleavage site.

Figure 49 represents the successful mutation of the amber stop codon of *secA* to the ochre stop codon TAA. The addition of *Hind*III, as shown in Figure 50, cleaves the PCR product from the single colony PCR in half, verifying in mutation in the genome. This efficient cloning scheme makes it possible to save not only time but also reduced the sequencing costs to modify a single amber stop codon since fewer clones need to be sequenced.

To mutate the second amber stop codon in *secG*, this procedure was repeated with the plasmid #30 (pJOE+secG_TAA+sgRNA_secG) on the $PG10^{Amber\text{-}secA}$ strain. After transformation and selection, positive colonies were again verified using colony PCR and subsequently digested with *Hind*III.

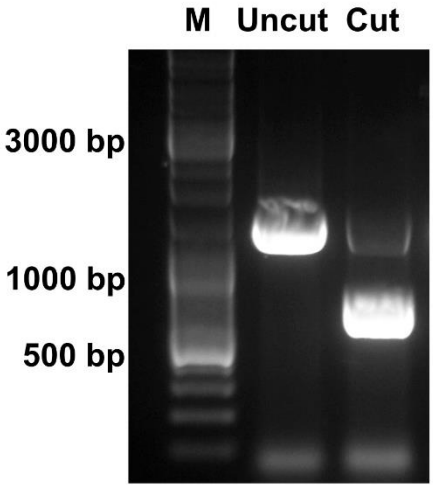

Figure 51: Single colony PCR of the strain PG10$^{Amber\text{-}secA}$ after transformation with plasmid pJOE-secG_TAA (#30).
The genomic region of *secG* was amplified and digested with HindIII, positively verifying the correct mutation of the stop codon, creating the strain PG10^{secGA} was created.

Again, the desired cleavage of the PCR product is visible. A faint residual left in the "Cut" lane is probably due to the large amount of produced PCR product, which requires more time to be cut completely. A longer incubation period would have certainly resulted in only cut PCR product. A mixture of two single colonies was ruled out by a final sequencing of the genomic regions of *secA* and *secG*, which both showed the distinct modification of the amber stop codon to the ochre stop codon.

However, the third target gene *secY* could not be mutated using CRISPR/Cas9. Despite trying all four possible sgRNAs, which are in range of the stop codon and adjacent to a PAM site, no positive clones could be obtained. The sequencing of the four plasmids showed the desired, correct sequence but no mutation of the amber codon of *secY* in the *B. subtilis* genome was detected, despite several transformation tries with each of the four sgRNA variants. A possible reason for this could be that SecY is an essential protein where already the DSB introduced by the Cas9 could lead to cell death before the DNA strand is repaired. In contrast, SecA is also an essential protein but the modification worked directly in the first try. Alternatively, a "protected" DNA sequence is also possible, where the sgRNA cannot detect the target DNA because of DNA-binding proteins occupying to the target site.

To solve this problem, *secY*^{TAA} was provided on an additional plasmid as described in the following paragraph.

2.5.3. Restoration of SecY expression

To restore the SEC system completely, the construct "P_{secG}-*secG*^{TAA}-*secY*^{TAA}" was designed. By analyzing the transcription levels of the SEC system proteins in different conditions using SubtiWiki, P_{secG} the promoter of *secG* was chosen. In complex media, both genes show the same transcription levels, which should in turn ensure correct protein expression levels, as SecG is required in the same amounts as

SecY for forming the SecYEG channel. By transforming the created integrating "Secretion plasmid" #17 (pXT_P_{secG}-secG^{TAA}-secY^{TAA}-4xpylT_P_{sigx}-5'prsA3') into the PG10$^{Amber\text{-}secGA}$ strain, the strain PG10$^{Amber\text{-}Sec}$ was created.

This strain now contains all required components to perform amber suppression and is also optimized to secrete proteins into the supernatant if the ncAA is present. To verify if the rescue of the secretion capacity by exchanging the amber stop codons of the SEC system worked, the PG10^{Amber} and the PG10$^{Amber\text{-}Sec}$ strain were both cultivated with and without the ncAA and the total secreted protein was analyzed by an SDS-PAGE. In Figure 52, it is clearly visible that die restoration of the SEC system via CRISPR/Cas9 and a genomic co-expression was successful and a huge impact on the overall secretion. The optimized secretion system shows nearly the same secretion levels compared to the strain without added ncAA.

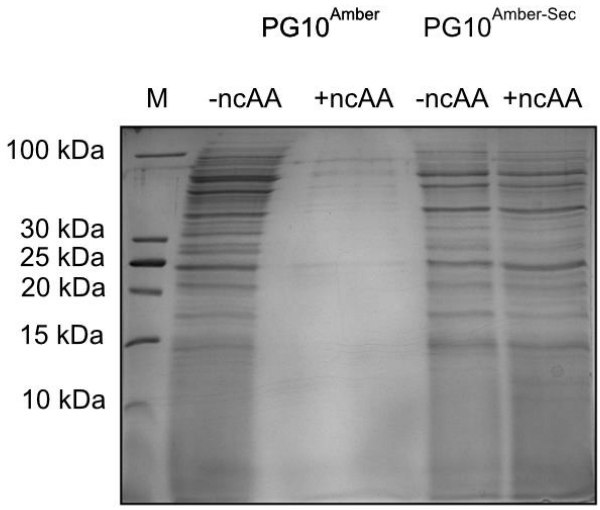

Figure 52: SDS-PAGE analysis of PG10^{Amber} and PG10$^{Amber\text{-}Sec}$.
Both strains were grown overnight in 2YT with and without the ncAA. The total supernatant was precipitated using ammonium sulfate and loaded onto an 18% SDS-gel after resuspension.

Furthermore, the overall growth rates of the three strains PG10, PG10^{Amber} and PG10$^{Amber\text{-}Sec}$ was analyzed in the plate reader. Therefore, the three strains were grown in presence and absence of IPTG and the ncAA and the growth behavior was monitored (Figure 53). Here, doubling times of about 70 minutes were recorded for all strains, indicating that the presence of the ncAA does not influence the overall growth of the different strains.

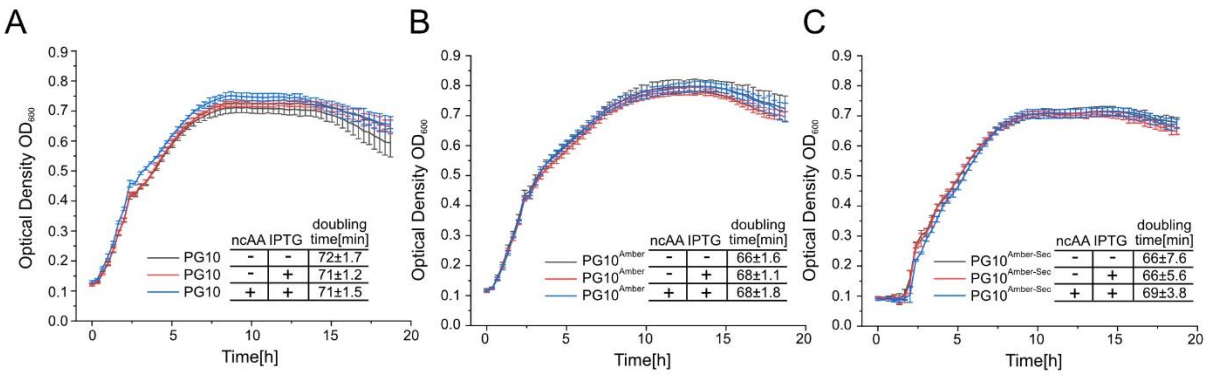

Figure 53: Growth curves of A) PG10, B) PG10^{Amber} and C) PG10$^{Amber-Sec}$.
The strains were grown in 2YT in presence and absence of 1 mM ncAA and 1 mM IPTG and the optical density (OD_{600}) was measured every 20 minutes.

The doubling times of 70 minutes are however much slower than the reported times from the literature (K07 approx. 20 minutes, PG10 approx. 30 minutes) which is most likely due to its cultivation in the plate reader. Here, the ncAA does not show an influence on the overall growth rates. These findings further underline that the reduction of the yields is an issue of the SEC system.

2.5.4. Summary

By using the single-plasmid system designed by Joseph Altenbuchner (142), the amber stop codons of *secA* and *secG* were successfully exchanged to ochre stop codons. The screening for positive colonies by inserting a *Hind*III cleavage site additionally saves time through a visible effect on the agarose gel and minimizes the sequencing efforts. However, the amber stop codon of *secY* could not be modified despite testing all available sgRNAs. By co-expressing *secY*^{TAA} on the "Secretion plasmid", a near complete restoration of the secreted proteins in the supernatant was achieved as clearly visualized by the SDS-PAGE analysis in Figure 52. Taken together, the created PG10$^{Amber-Sec}$ strain now drastically improves the yields of all secreted proteins if an ncAA is present in the medium which should certainly improve the overall yields for the expression of MAK33-VL^{Norb33}.

2.6. Expression and secretion of bioorthogonal modified proteins

Due to the rescue of the overall secretion shown in Figure 52 and a reduced effect of the ncAA on the growth rates in the created strain PG10$^{Amber\text{-}Sec}$ strain in the presence of the ncAA, an increase of the amount of the secreted antibody fragment is expected.

2.6.1. Expression and purification of MAK33-VL^{Norb33}

The strain PG10$^{Amber\text{-}Sec}$ was transformed using plasmids #20 (pLIKE-P_{groES}-5'AmyE-MAK33-N33TAG-VL-StrepII-3'-LacI) and #22 (pLIKE-P_{groES}-5'AmyE-MAK33-N33TAG-VL-His_6-3'-LacI) respectively to express the C-terminal His_6-tagged and StrepII-tagged protein variants. The positive clones were pooled to level out individual expression differences between the single colonies. Expression and purification was carried out as described previously for the wild type protein MAK33-VL. After purification via affinity chromatography, the fractions were analyzed on an SDS-PAGE and protein bands were verified using western blotting against the StrepII-tag.

In the SDS-PAGE analysis, a faint band in the supernatant and flow through fractions corresponding to the secreted MAK33-VL^{Norb33} is visible. The supernatant itself was not as concentrated as shown in the previous experiment (Figure 52) as the precipitated protein was resuspended in a larger volume to ensure sufficient protein solubilization. The StrepTactin columns exhibit signs of overloading as the four flow through fractions FT1-4 still show a positive signal in the western blot. In the wash fractions W1 and W2, only minor contaminants are visible which is due to the high specificity of the StrepTactin columns as well as the secretion into the supernatant, reducing the overall amount of protein, which could binding unspecifically to the columns. In the elution fraction E only one band at the height of the wild type control is visible which is characterized by a positive signal in the Western Blot, underlining most likely the production of the bioorthogonally modified protein.

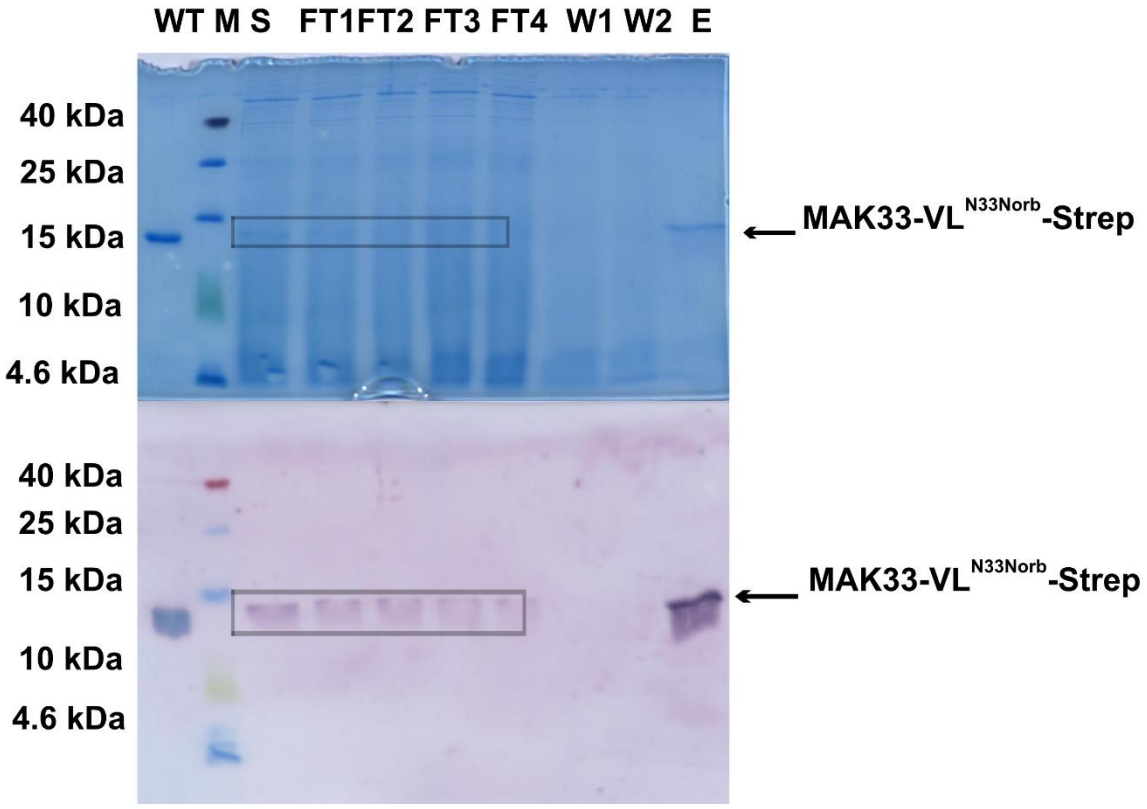

Figure 54: SDS-PAGE analysis and western blotting against the StrepII-tag.
The SDS-PAGE analysis (top) shows a clear elution fraction E after purification of the supernatant (S) using StrepTactin Spin columns. The Western Blot (bottom) shows an additional signal of the target protein in the flow through fractions (FT). The control fraction on the left, the MAK33-VL wild type protein, runs on the same height as the purified MAK33-VL^{Norb33}.

A further advantage of secreting the protein into the supernatant is the high purity of the elution fraction after only one purification step, not only for the highly specific StrepII-tag but also for the His_6-tag (Figure 55).

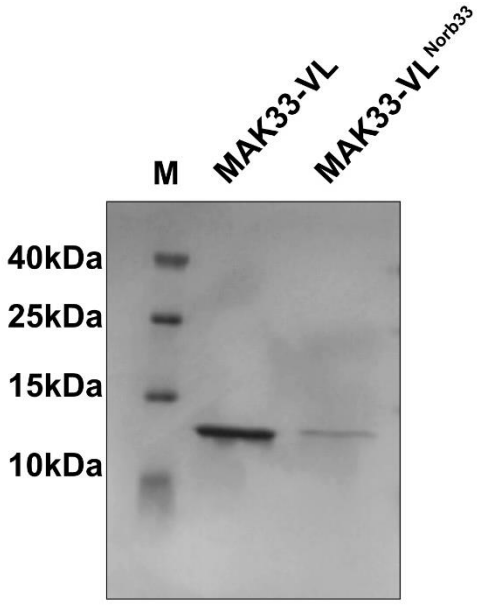

Figure 55: SDS-PAGE analysis of the elution fractions of MAK33-VL and MAK33-VL^{Norb33} purified via their His_6-tag.
The amount of protein loaded in each lane corresponds to the same culture volumes.

With overall yields of approximately 2 mg l^{-}, the expression and purification of the antibody fragment shows an incorporation rate of about 15%. This is slightly lower than the incorporation rate observed for GFP. However, one has to take into account that GFP^{Norb33} is purified from the cytosol, while $MAK33\text{-}VL^{Norb33}$ in addition has to be secreted. A comparison to reports utilizing *B. subtilis* for the production of single chain antibodies reveals yields for the wild type proteins in the range of about 5 to 15 mg l^{-1}, which are also achieved for the secretion of the MAK33-VL wild type protein.(85, 145)

2.6.2. Mass spectrometry analysis and click modification

For further verification of the identity of the correct incorporated ncAA in the target protein, the StrepII-tagged and the His_6-tagged protein variants from $MAK33\text{-}VL^{Norb33}$, shown in the previous paragraph, were analyzed by ESI-MS. In addition, $MAK33\text{-}VL^{Norb33}$-Strep and $MAK33\text{-}VL^{Norb33}\text{-}His_6$ were reacted with compounds **3** and **4** respectively. The SPase cleavage pattern of the wild type MAK33-VL can again be detected. The two peaks found in the untreated fraction correspond to $MAK33\text{-}VL^{Norb33}$ (calc. 12996 Da) and $Ala\text{-}MAK33\text{-}VL^{Norb33}$ (calc. 13067 Da), verifying the correct incorporation of the ncAA. Upon addition of the compound **3**, a mass-shift by 159 Da is clearly visible in the spectrum, confirming the reaction and forming the $MAK33\text{-}VL^{Norb\text{-}Tet}$ conjugate.

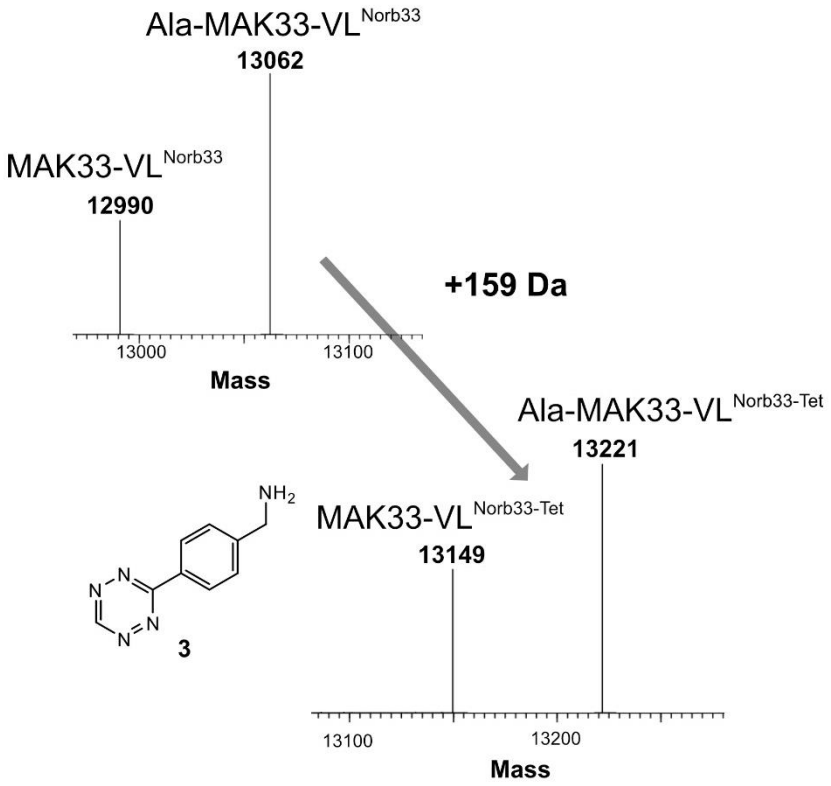

Figure 56: ESI-mass spectra of the purified StrepII-tagged $MAK33\text{-}VL^{Norb33}$
Top: The unreacted protein is found in two variants, the full-length antibody fragment and variant with an N-terminal alanine. Bottom: By adding the tetrazine ligand **3,** endo-norbornene lysine reacts with the ligand and forms the protein-ligand conjugate.

The His_6-tagged protein was reacted with compound **4** and analyzed in the same fashion. Again, both possible cleavage products are detected in the elution fraction. However, upon addition of the tetrazine ligand, only the fraction corresponding to Ala-MAK33-$VL^{Norb33\text{-}Fluoro}$ is detectable. This could be due to differences in the ionization properties of both protein variants, whereby the smaller peak of MAK33-$VL^{Norb33\text{-}Fluoro}$ is below the detection limit. However, the mass shift by 805 Da is found as expected verifying the functionality of the His_6-tagged protein variant.

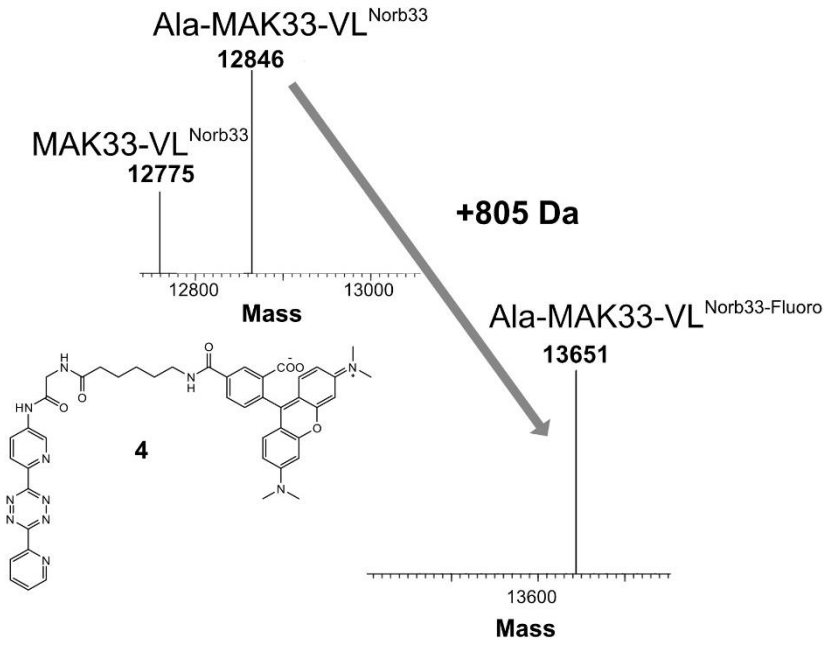

Figure 57: ESI-mass spectrum of the His_6-tagged MAK33-VL^{Norb33}
Top: The unreacted protein is present with two species, with and without an optional N-terminal alanine. Bottom: By adding the TAMRA containing tetrazine ligand **4**, a mass shift of 805 Da is visible.

In general, both protein variants show a high conversion upon addition of the respective tetrazine ligands with reaction times less than one hour. This indicates that the incorporated ncAA endo-norbornene lysine is exposed to solvent and can react with the added ligands to the conjugate as expected.

2.6.3. Expression of D1.3scFv^{Norb74}

Since the second target protein, the single chain variable fragment of the D1.3scFv, could be expressed and secreted in sufficient amounts, four different incorporation sites ($Ser61^{TAG}$, $Ser74^{TAG}$, $Lys75^{TAG}$ and $Ser84^{TAG}$) were tested. Again, sites were chosen according to their location in unstructured loop regions and sequence context, based on the previous studies.(109-111) Therefore, the D1.3scFv wild type plasmids #23 (pLIKE-P_{groES}-5'LipA-D1.3scFv-His_6-3'-) and the four plasmids #24-27 (pLIKE-P_{groES}-5'LipA-D1.3scFv-K75TAG/S61TAG/S74TAG/S84TAG-His_6-3') containing D1.3scFv with one of the four possible

incorporation sites, were transformed into PG10$^{Amber\text{-}Sec}$, creating the strains PG10$^{Amber\text{-}Sec}$-3-7 Grown colonies were tested by single-colony PCR and positive clones pooled. Protein expression was carried out as evaluated in the previously in 2YT at 25°C overnight. The proteins were purified from the supernatant using His SpinTrap columns. As previous experiments showed that the target protein can be purified in such purity with only one affinity chromatography step, only the supernatant and corresponding elution fractions are analyzed via an SDS-PAGE.

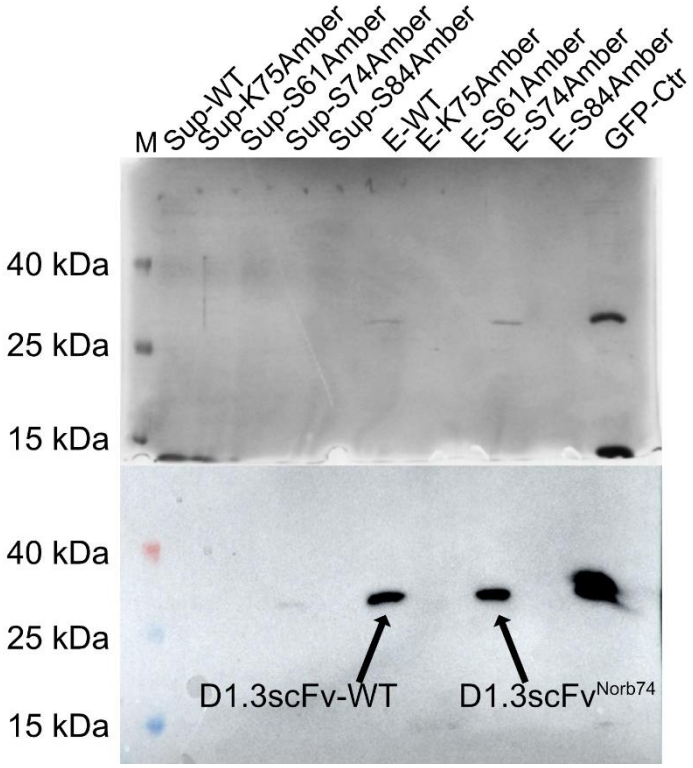

Figure 58: Incorporation site evaluation of D1.3scFv
Top: SDS-PAGE analysis of the supernatant and elution fractions from a purification using His SpinTrap columns of the five D1.3scFv (+/-TAG) expressing strains. Bottom: Western Blot against the His_6-tag of D1.3scFv. GFP-His_6 is used as a control.

The SDS-PAGE analysis exhibits two visible bands in the elution fractions. One of the wild type protein D1.3scFv and a second band corresponding to the elution fraction of D1.3scFv^{Norb74}. However, the supernatant fractions show hardly any visible bands, which is most likely from the low amount of sample applied to the SDS-PAGE paired with a larger volume for solubilization. The verification via western blotting against the His_6-tag reveals two signals, one for the wild type protein and one for the D1.3scFv^{Norb74} protein. However, only a mass spectrum of the wild type protein could be obtained which did not differ from the mass spectrum previously obtained (Figure 39) during the initial purification of the wild type protein. The modified protein D1.3scFv^{Norb74}did not show a clear mass spectrum, as the ionization for ESI-MS has been a general issue with this protein.

2.6.4. Summary

By using the engineered PG10$^{Amber\text{-}Sec}$ strain, the target proteins MAK33-VL^{Norb33} and D1.3scFv^{Norb74} could be expressed and purified from the spent medium. In contrast to D1.3scFv^{Norb74} where only low quantities could be obtained, the yields for MAK33-VL^{Norb33} were approximately 2 mg l^{-1} or 10 g cells. Moreover, endo-norbornene lysine incorporation and functionalization with tetrazine derivatives was verified via ESI-MS. Additionally the incorporation rate of about 15% compared to the wild type MAK33-VL is more than twice as high as in a previously reported genetic code expansion system established in *Bacillus cereus*.(146) For both proteins, a higher incorporation rate would be desirable but for MAK33-VL it is in an acceptable range in respect to all different system, which are involved in the production and secretion of the antibody fragment. The high purity of the elution fractions after only one-step of affinity chromatography is clearly superior to the purification from an intracellular expression, which often requires several further purification steps.

2.7. Functional characterization of MAK33-VL^{Norb33}

Although it is known that MAK33-VL is a very stable protein, requiring harsh conditions to be denatured (126), correct folding and function upon ncAA incorporation and secretion are necessary. Therefore, a previously for MAK33-VL developed enzyme-linked immunosorbent assay (ELISA) was modified to analyze the binding of MAK33-VL$^{Norb33\text{-}Fluoro}$ to hmCK.(127) For this, biotinylated hmCK was first bound to streptavidin coated 96 well plates, than increasing amounts of MAK33-VL$^{Norb33\text{-}Fluoro}$ (Figure 59) were added and the fluorescence recorded.

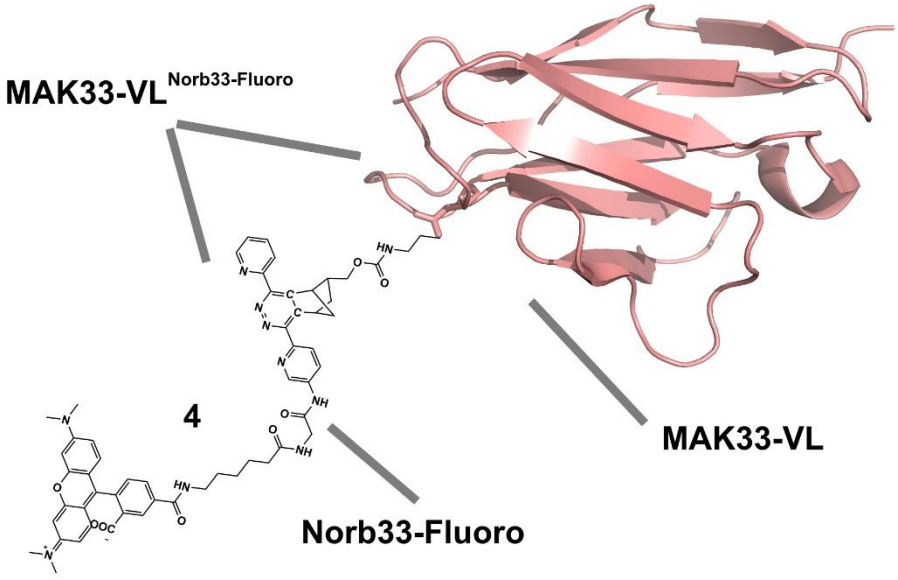

Figure 59: Scheme of MAK33-VL$^{Norb33\text{-}Fluoro}$ conjugate (PDB 1FH5)
The tetrazine compound **4** containing the TAMRA fluorophore is attached via the ncAA of MAK33-VL^{Norb33}, forming the conjugate.

Here, with rising amounts of MAK33-VL$^{Norb33\text{-}Fluoro}$, an increase in the detected fluorescence signal is visible, indicating that the protein binds to its target (Figure 60-A). In addition, the slope decreases with larger quantities of MAK33-VL$^{Norb33\text{-}Fluoro}$, showing a saturation of the binding sites of hmCK. The rise in the fluorescence with increasing amounts of the labelled antibody fragment visualizes that the POI binds to its target protein and is therefore present in its correctly folded state. The standard deviations of up to 15 % are in range for a sensitive assay, where binding to the wells in the pMol range occurs.

In the second ELISA assay, wild type MAK33-VL and MAK33-VL$^{Norb33\text{-}Fluoro}$ are used simultaneously to rival for the binding sites of hmCK (Figure 60-B). The amounts of used MAK33-VL$^{Norb33\text{-}Fluoro}$ are held constant and compete with increasing quantities of the wild type MAK33-VL for the binding sites of hmCK. The detected fluorescence exhibits an exponential decline, indicating that with a higher fold excess of MAK33-VL, MAK33-VL$^{Norb33\text{-}Fluoro}$ cannot bind to the hmCK, thus leading to a lower fluorescence signal.

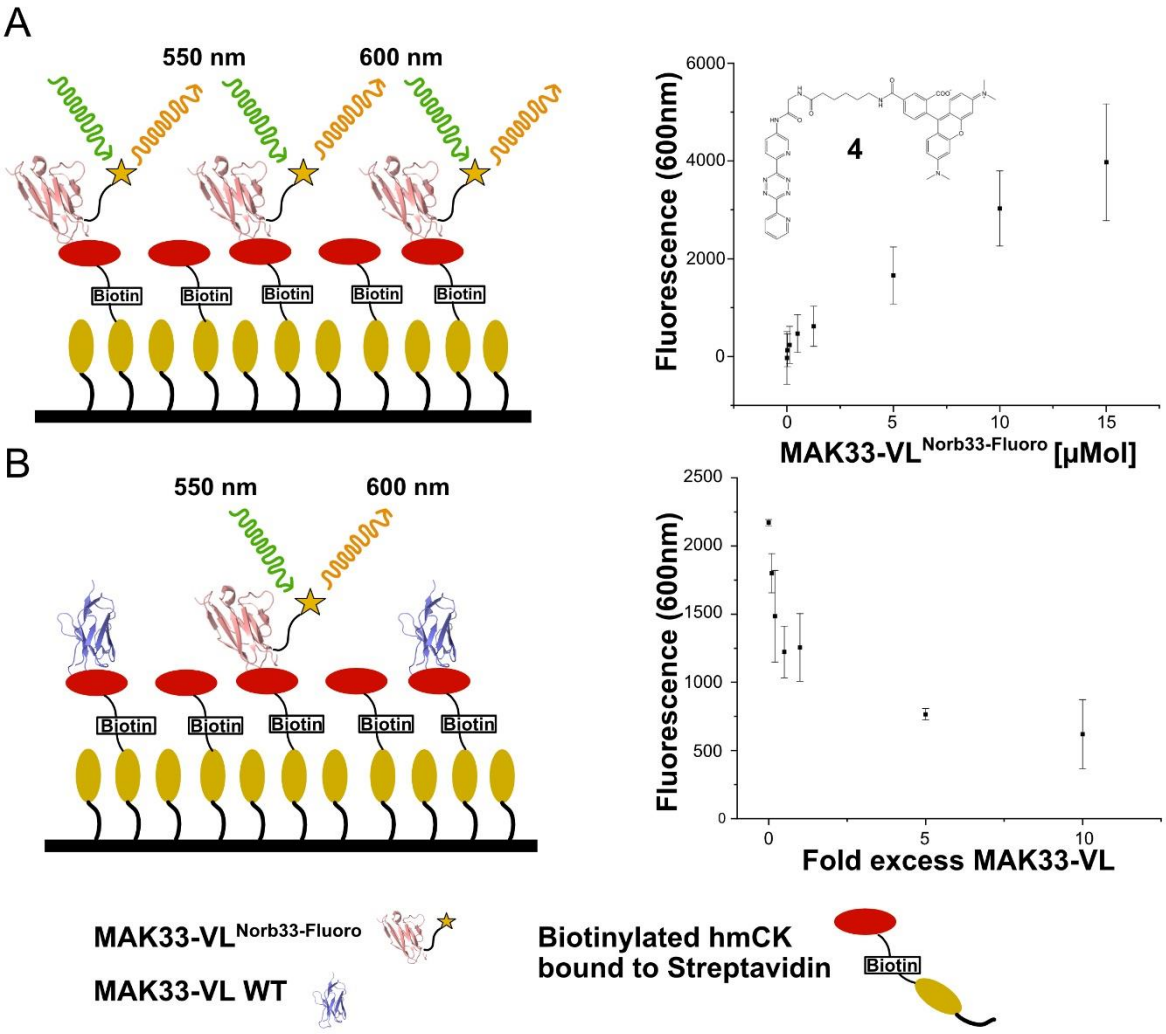

Figure 60: Experimental setup and verification oft he MAK33-VL functionality
Streptavidin coated 96 well plates are loaded with biotinylated hmCK A: Titration of hmCK (5pM/well) with increasing amounts of MAK33-VL$^{Norb33\text{-}Fluoro}$. B: 1000 pMol of MAK33-VL$^{Norb33\text{-}Fluoro}$ are mixed with increasing excess of MAK33-VL in presence of hmCK^{Biotin}.

Both assays clearly support that the secreted and labelled antibody fragment MAK33-VL$^{Norb33\text{-}Fluoro}$ is produced in its correct protein fold and binds to the target protein hmCK. Although the steep decrease in the first measuring points of the competition assay could indicate that the overall affinity of the modified antibody fragment is decreased slightly compared to the wild type as the introduced ncAA endo-norbornene lysine and the attached fluorophore are rather large.

2.7.1. Summary

By using fluorescently labelled MAK33-VL^{Norb33} in an ELISA against hmCK, it could be shown that the POI binds to its antigen. Moreover, binding of MAK33-VL$^{Norb33\text{-}Fluoro}$ could be competed off by adding wild type MAK33-VL. This shows that in *B. subtilis* expressed and secreted MAK33-VL, which contains the ncAA endo-norbornene lysine is fully functional.

3. Conclusion and outlook

The aim of this thesis was to establish the amber suppression system as a method to expand the genetic code beyond the 20 canonical amino acids and secrete a pharmaceutically relevant protein into the supernatant.

Therefore, initially the used expression plasmid was optimized to express the target protein in suitable amounts. Here, the stabilization of the mRNA of the POI by secondary RNA structures gave a tremendous increase in the overall yields. By further evaluating several promoters for their expression behaviour in different growth conditions, suitable promoters for different expression levels were found. To be able to express heterologous proteins in *B. subtilis* containing ncAA, the amber suppression system was introduced. Here, by the stable genomic integration of the aaRS^{Norb}/*pylT* while overexpressing *gfp*^{K26TAG}, the ncAA could be incorporated with rates of about 30 %. ESI-MS hereby revealed a successful incorporation and functionalization with a tetrazine ligand.

For the pharmaceutically relevant protein class of antibody-derived proteins, two single chain antibodies, MAK33-VL and Am3-114, and the single chain variable fragment D1.3scFv were successfully expressed and secreted into the spent medium. The direct purification from the supernatant enabled very pure elution fractions from affinity chromatography. By engineering the SEC system of the genomically reduced *B. subtilis* strain PG10, the overall secretion in presence of the ncAA could be recovered, making it possible to secrete the MAK33-VL^{Norb33} and D1.3scFv^{Norb74} into the supernatant. By applying two modified ELISAs, the correct functionality and protein fold of MAK33-VL^{Norb33} could be verified, thus illustrating the overall success of the established genetic code expansion system in *B. subtilis*.

However, the incorporation rates of about 15 % as achieved for the expression of MAK33-VL^{Norb33} bears unexploited potential for greater yields. As the target site evaluation using C-terminal fusion constructs revealed differences for possible incorporation sites compared to *E. coli*, a screening as performed by Pott and coworkers could give new insights for efficient target sites in *B. subtilis*. Additionally, two points during the translation can be addressed. First, it was shown for *E. coli* that engineering the Elongation Factor TU, which delivers the charged ncAA-tRNA conjugate to the ribosome, can elevate the incorporation rates.(147) Secondly, by downregulating the expression of RF1 as the counterpart for the amber suppression system, the incorporation rates could also be increased.

4. Material and methods

4.1. Materials

4.1.1. Devices

All devices in this work were used as stated by the manufacturer if not stated otherwise.

Table 1: Used devices

Purpose	Product	Manufacturer
Centrifuges	1-14	Sigma Zentrifugen
	1-14K	Sigma Zentrifugen
	Minispin	Sigma Zentrifugen
	8K	Sigma Zentrifugen
	6-16K	Sigma Zentrifugen
	3-16K	Sigma Zentrifugen
	3-30K	Sigma Zentrifugen
	4K15	Sigma Zentrifugen
	3XR	Thermo Fisher Scientific
Rotors	11805	Sigma Zentrifugen
	13845	Sigma Zentrifugen
	12500-H	Sigma Zentrifugen
	11133	Sigma Zentrifugen
	12150-H	Sigma Zentrifugen
	11150	Sigma Zentrifugen
	13220	Sigma Zentrifugen
	90920	Sigma Zentrifugen
	F15-8x50cy	Thermo Fisher Scientific
	TX-1000	Thermo Fisher Scientifc
Gel electrophoresis	Might Small™ II system	Hoefer Inc.
	Semi-Dry Blotter	Biometra
	power supply EPS-600	Pharmacia Biotech
	peqlab 40-0911	Peqlab Biotechnologie
	peqPOWER 300	Peqlab Biotechnologie

	gel imaging system G:Box	Syngene
Miscellaneous & Consumables	Spark10M	Tecan
	Nanodrop 1000 Spectrophotometer	Thermo Scientific
	T1000 Thermal Cycler	Biorad
	MJ Mini Thermal Cycler	Biorad
	inoLab pH720	WTW
	MicroPulser electroporation apparatus	Biorad
	MR Hei-Standard heating stirrer	Heidolph Instruments
	Dry oven DRY-Line 53	VWR
	Autoclave Classic Media	Prestige Medical
	Varioclave autoclave	HP Medizintechnik
	96-well plates 655097	Greiner Bio-One International GmbH

4.1.2. Chemicals

All chemicals were ordered from common suppliers such as Biozym, Carl Roth, Fluka, Merck, Sigma-Aldrich and VWR.

4.1.3. Cultivation media

Table 2: Bacterial growth media

Name	Ingredients
Lysogeny Broth (LB)	1 % (w/v) NaCl 1 % (w/v) Tryptone 0.5 % (w/v) Yeast extract
2YT	1 % (w/v) NaCl 1.6 % (w/v) Tryptone 1.0 % (w/v) Yeast extract
TB	17 mM KH_2PO_4 72 mM K_2HPO_4 1.2 % (w/v) Tryptone 2.4 % (w/v) Yeast extract 0.5 % (w/v) Glycerol

SOC	10 mM NaCl 2.5 mM KCl 10 mM $MgCl_2$ 10 mM $MgSO_4$ 2 % (w/v) Tryptone 0.5 % (w/v) Yeast extract 20 mM Glucose
LB-Agar	1 % (w/v) NaCl 1 % (w/v) Tryptone 0.5 % (w/v) Yeast extract 1.5 % (w/v) Agar-agar
MNGE	59 mM K_2HPO_4 44 mM KH_2PO_4 3.4 mM Sodium citrate 2 % (w/v) Glucose 0.2 % (w/v) Potassium glutamate 0.11 g l^{-1} NH_4^+-Fe(III)-citrate 50 mg l^{-1} Tryptophan 3 mM $MgSO_4$
Expression Mix	2.5 % (w/v) Yeast extract 2.5 % (w/v) Casamino acids 250 mg l^{-1} Tryptophan
MCSE	40 mM MOPS 3.3 g l^{-1} $(NH_4)_2SO_4$ 0.385 mM KH_2PO_4 0.615 mM K_2HPO_4 50 mg l^{-1} Tryptophan 50 mg l^{-1} Threonine 22 mg l^{-1} Ammonium ferric citrate, 2.3 mg l^{-1} $MnSO_4$ x 4 H_2O 123 mg l^{-1} $MgSO_4$ x $7H_2O$ 0.8 % (w/v) K-glutamate, 0.6 % (w/v) Na-succinate 1.0 % (w/v) CAA 2.5 % (w/v) Fructose

MN-CSE	59 mM K_2HPO_4 44 mM KH_2PO_4 3.4 mM Sodium citrate 50 mg l^{-1} Tryptophan 50 mg l^{-1} Threonine 22 mg l^{-1}Ammonium ferric citrate, 2.3 mg l^{-1} $MnSO_4$ x 4 H_2O 123 mg l^{-1} $MgSO_4$ x 7 H_2O 0.8 % (w/v) K-glutamate, 0.6 % (w/v) Na-succinate 1.0 % (w/v) CAA 1.0 % (w/v) Glycerol 2.5 %(w/v) Fructose

Complex media from Table 2 were autoclaved before use. The single components for the defined media were autoclaved or sterile filtered.

4.1.4. Enzymes, standards and kits

4.1.4.1. DNA-modifying enzymes

DNA modifying enzymes were purchased from New England Biolabs (NEB Biolabs) and used following the manufacturers guidelines. The amount of required enzyme was calculated considering the enzyme stock concentration and individual parameters depending on each enzyme as stated by the manufacturer.

Table 3: Used enzymes

Product	Source
Phusion High-Fidelity DNA Polymerase	Finnzymes
GoTaq® G2 DNA Polymerase	Promega
Q5 High-Fidelity DNA Polymerase	NEB Biolabs
Pfu DNA Polymerase	Promega
OneTaq DNA Polymerase	NEB Biolabs
Antarctic Phosphatase	NEB Biolabs
T4 DNA Ligase	NEB Biolabs
T4 Polynucleotide Kinase	NEB Biolabs

Lysozyme, lyophilized	Carl Roth
DNase I (bovine pancreas), lyophilized	Sigma

4.1.4.2. DNA/protein markers

Table 4: DNA and protein markers

Ladder	Source
peqGOLD DNA ladder Mix	peqLab
PageRuler™ Unstained Low Range Protein Ladder	ThermoFisher Scientific
Spectra™ Multicolor Low Range Protein Ladder	ThermoFisher Scientific
Color Prestained Protein Standard, Broad Range	NEB Biolabs

4.1.4.3. Kit systems

Table 5: Used kit systems

Product	Supplier
Wizard® SV Gel and PCR clean-up system	Promega
Monarch® DNA Gel extraction kit	NEB Biolabs
Monarch® Plasmid Miniprep kit	NEB Biolabs
His SpinTrap Columns	GE Life Sciences
StrepTactin®XT Spin Columns	IBA Life Sciences

4.1.5. Oligonucleotides

Table 6: Used DNA oligonucleotides
Cleavage sites are marked in italic

Number	Name	Sequence	Project
1	PS_P_{trnQ}_*Eco*RI_fwd	GATC *GAATTC* CTCCTTCACCTTTTAGGTA	Promoter screening
2	PS_P_{trnQ}_*Bam*HI_rev	GATC *GGATCC* GCACAACTAATTATACGATAATT	
3	PS_P_{groES}_*Eco*RI_fwd	GATC *GAATTC* ATGGTATGTACTCCTTTGTTAAGT	
4	PS_P_{groES}_*Bam*HI_rev	GATC *GGATCC* ACAATTCTTATAATAAAGAATCTCCCTTC	
5	PS_P_{43}_*Eco*RI_fwd	GATC *GAATTC* CTGTTCTCATTTGCGGC	

6	PS_P_{43}_*Bam*HI_rev	GATC *GGATCC* TTTATATTTTACATAATCGCGCGC	
7	PS_P_{sigX}_*Eco*RI_fwd	GATC *GAATTC* AATCGAGTCTGAATTTGCC	
8	PS_P_{sigX}_*Bam*HI_Bind_rev	GATC *GGATCC* TTGAAACCCCTCCGTT	
9	SLIM_Stemloop_P_{veg}_rev short	ACATTTATTGTACAACACGAGC	mRNA stabilization
10	SLIM_Stemloop_P_{veg}_rev tailed	AATTGGTGTTGGTTGTTGTTGTGGAATT GTGATCCGCTCACAATCCACAACATTTA TTGTACAACACGAGCCCA	
11	SLIM_Stemloop_P_{veg}_fwd short	AGTTAAGGAGGATTCTAGATGA	
12	SLIM_Stemloop_P_{veg}_fwd tailed	TGTGGATTGTGAGCGGATCACAATTCCA CAACAACAACCAACACCAATTAGTTAAG GAGGATTCTAGATGA	
13	SLIM_Term_RepGen_rev short	TTATTAATGGTGATGGTGATGGTGTTTG	
14	SLIM_Term_RepGen_rev tailed	GAAAAAGGGGCAGCCCGCTCATTAGGC GGGCTGCCCGGGGACGTCGCTAGCTTA TTAATGGTGATGGTGAT	
15	SLIM_Term_RepGen_fwd short	ATGAAACTAGTTCGTCGACTGCC	
16	SLIM_Term_RepGen_fwd tailed	GCTAGCGACGTCCCCGGGCAGCCCGCC TAATGAGCGGGCTGCCCCTTTTTCATGA AACTAGTTCGTCGACTGCC	
17	QC_GFP_K26Amber_fwd	TGGTGACGTTAATGGGCACTAGTTTTCT GTCAGTGGAGAGG	ncAA incorporation in GFP
18	QC_GFP_K26amber_rev	CCTCTCCACTGACAGAAAACTAGTGCCC ATTAACGTCACCA	
19	MAK33_Sec_*Xba*I_*Not*I_ fwd	GATC *TCTAG A*TGGGGGCGGCC GCAGATATTGTGCTAACTC AGTCTC	Secretion of MAK33-VL
20	MAK33-VL-StrepTag-*Xho*I-rev	GATC *CTCGAG* TTAATGGTGATGGTGATGGTG TCTTTTCAGCTCCAGCTTGG	
21	MAK33_His6_*Xho*I_rev	GATC *CTCGAG* TTATTTTTCGAACTGCGGGTGGCTCCAT C TTTTCAGCTCCAGCTTGG	
22	AmyE_fwd	GATC *TCTAG A*TGTTTGCAAAACGATTCAAA	
23	AmyE_rev	GATC T*GC GGC CGC* AGC CGC CGG TCC TG	
24	*Xba*I-MAK33VL-fwd	GATC *TCTAG A*TGTTTGCAAAACGATTCA	C-terminal GFP fusion constructs

25	Linker-mak33vl-rev	GCTTCCGCCACCACCTGAGCCTCCGCCACCTGATCCGCCTCCGCCTCTTTTCAGCTCCAGCTTGG	C-terminal GFP fusion constructs
26	Linker- GFP -fwd	GGCGGAGGCGGATCAGGTGGCGGAGGCTCAGGTGGTGGCGGAAGCAGTAAAGGAGAAGAACTTTTCACTGG	
27	GFP-*Xho*I-rev	GATC CTCGAG TTATTAATGGTGATGGTG	
28	MAK33-VL-N33TAG-fwd	GCAGTTTTTGTTGATACCAGTGGATCTAATTGCTAATACTTTGGCTGGCCCTGCAGG	Incorporation sites for MAK33-VL
29	MAK33-VL-N33TAG-rev	CCTGCAGGGCCAGCCAAAGTATTAGCAATTAGATCCACTGGTATCAACAAAAACTGC	
30	MAK33-VL-L41TAG-fwd	GAGAAGCCTTGGAGACTCAGTCTATTTTTGTTGATACCAGTGTAGGTTGTTGAATACT	
31	MAK33-VL-L41TAG-rev	AGTATTCAACAACCTACACTGGTATCAACAAAAATAGACTGAGTCTCCAAGGCTTCTC	
32	MAK33-VL-Y51TAG-fwd	GATGGACTGGGAAGTCTATTTGATGAGAAGCCTTGGAGACT	
33	MAK33-VL-Y51TAG-rev	AGTCTCCAAGGCTTCTCATCAAATAGACTTCCCAGTCCATC	
34	*Xba*I-LipA-fwd	GATC *TCTAG A*TGAAAAAAGTCCTGATG	Expression and secretion of D1.3scFv
35	*Xho*I-His6-D1.3scFv-rev	GATC *CTCGAG* TTAATGATGATGGT	
36	QC_D13_S61Amber_fwd	TGAAATGCTCAGACGGCTTTTCAGAGTCTAGTTATAATCTGTATTGCCATCTCCC	Incorporation sites D1.3scFv
37	QC_D13_S61Amber_rev	GGGAGATGGCAATACAGATTATAACTAGACTCTGAAAAGCCGTCTGAGCATTTCA	
38	QC_D13_S74Amber_fwd	CTTCAGAAAGACCTGGCTTTTCTAGTTATCTTTTGAAATGCTCAGACGGCTTTTCAGT	
39	QC_D13_S74Amber_rev	ACTGAAAAGCCGTCTGAGCATTTCAAAAGATAACTAGAAAAGCCAGGTCTTTCTGAAG	
40	QC_D13_S84Amber _fwd	GCAATAATATCTTGCCGTATCATCTGTATGGATCTAGTTCATCTTCAGAAAGACCTGGCTTTTGCTA	
41	QC_D13_S84Amber_rev	TAGCAAAAGCCAGGTCTTTCTGAAGATGAACTAGATCCATACAGATGATACGGCAAGATATTATTGC	
42	*Not*I-Am3-114	GATC *GCG GCC GC*C CTG CAA GTT GAT	Expression and secretion of Am3-114
43	*Xho*I-Am3-114-Strep	GATC *CTCGAG* TTATTTCTCGAACTGCGGA	

44	P_{secG}-secG-SD-secY_fwd	GATC *TTAATTAA* CGCTTTTCTCCTTT	Expression construct *secY*^{TAA}
45	P_{secG}-secG-SD-secY_rev	GATC *GCTCAG* CGAAAAAGGG	
46	sgRNA-secA-fwd	TACG CCGTACTGAATAGTTCGCCC	CRISPR/Cas9 engineering of *secA*^{TAA}
47	sgRNA-secA-rev	AAAC GGGCGAACTATTCAGTACGG	
48	secA-Region-fwd	GATC *GTCGAC* CTCCCCATCTTGTATACATTCTAAG	
49	secA-Region-rev	GATC *TCTAGA* GAGGTCATTTATAAGCAGCG	
50	QC-secA-fwd	ATTGCTGCGGCCGTACTGAATAAGCTTC CCCGGCAAGTTTACTGACCGC	
51	QC-secA-rev	GCGGTCAGTAAACTTGCCGGGGAAGCT TATTCAGTACGGCCGCAGCAAT	
52	fwd-*Sal*I-secG	GATC *GTCGAC* ATAAAAATGATAGGGATTCCCTGTC	CRISPR/Cas9 engineering of *secG*^{TAA}
53	rev-*Xba*I-secG	GATC *TCTAGA* CCCTTTATGGGTACAGTGTAACC	
54	QC-secG-fwd	CACATCAGACCTTATACAAACATAAGCT TATAGGATATAAGCAAGCGCAATCGTTA ACAC	
55	QC-secG-rev	GTGTTAACGATTGCGCTTGCTTATATCCT ATAAGCTTATGTTTGTATAAGGTCTGAT GTG	
56	sgRNA-secG-fwd	TACG TTATACAAACATTGCCCTAT	
57	sgRNA-secG-rev	AAAC ATAGGGCAATGTTTGTATAA	
58	*Sal*I-secY-fwd	GATC *GTCGAC* CGTTTTTGATAACGCCGC	CRISPR/Cas9 engineering of *secY*^{TAA}
59	*Xba*I-secY-rev	GATC *TCTAGA* TTAATGACATAATCAATCGGCTTG	
60	QC-secY-fwd	GTGAAACGAAACTACCGTGGATTTATGA AAAACTAAGCTTAATGGATTTATCCATT CCCTCTTAATAAAGAGA	
61	QC-secY-rev	TCTCTTTATTAAGAGGGAATGGATAAAT CCATTAAGCTTAGTTTTTCATAAATCCAC GGTAGTTTCGTTTCAC	
62	sgRNA-secY-fwd	TACG TTATGAAAAACTAGAGGAAA	
63	sgRNA-secY-rev	AAAC TTTCCTCTAGTTTTTCATAA	
64	sgRNA2-secY-fwd	TACG GTGGATTTATGAAAAACTAG	
65	sgRNA2-secY-rev	AAAC CTAGTTTTTCATAAATCCAC	
66	sgRNA_secY3_fwd	TACG TTGGTGAAACGAAACTACCG	
67	sgRNA_secY3_rev	AAAC CGGTAGTTTCGTTTCACCAA	
68	sgRNA_secY4_fwd	TACG GGTAGTTTCGTTTCACCAAC	
69	sgRNA_secY4_rev	AAAC GTTGGTGAAACGAAACTACC	
70	pLIKE-fwd	CAGGGTTTTCCCAGTCACGAC	ColonyPCR pLIKE
71	pLIKE-rev	CAATTTCACACAGGAAACAGCTATGAC	

4.1.6. Plasmids

Table 7: Used and created plasmids
The pLIKE-plasmids are based on pLIKE-rep from (89), pSBbs1c from (88), pJOE8999 from (142) and pXT from (118)

Number	Name
1	pLIKE-P_{veg}-5'-GFP-His_6-3'
2	pLIKE-P_{groES}-5'-GFP-His_6-3'
3	pLIKE-P_{43}-5'-GFP-His_6-3'
4	pLIKE-P_{sigx}-5'-GFP-His_6-3'
5	pLIKE-P_{trnQ}-5'-GFP-His_6-3'
6	pLIKE-P_{veg}-GFP-His_6
7	pLIKE-P_{veg}-5'Stem-GFP-His_6
8	pLIKE-P_{veg}-GFP-$His_6$3'-Term
9	pLIKE-P_{veg}-5'Stem-GFP-His_6-3'-Term
10	pLIKE-P_{groES}-5'AmyE-MAK33-VL-GFP-His_6-3'-LacI
11	pLIKE-P_{groES}-5'AmyE-MAK33-VL-N33TAG -GFP-His_6-3'-LacI
12	pLIKE-P_{groES}-5'AmyE-MAK33-VL-L41TAG- GFP-His_6-3'-LacI
13	pLIKE-P_{groES}-5'AmyE-MAK33-VL-Y51TAG-GFP-His_6-3'-LacI
14	pSBbs1c-P_{sigX}-5'pylS-3´-1xpylT
15	pXT-4xpylT_LacI
16	pXT-4xpylT-P_{sigX}-5'prsA3'
17	pXT_P_{secG}-$secG^{TAA}$-$secY^{TAA}$-4xpylT_P_{sigx}-5'prsA3'
18	pLIKE-P_{groES}-5'-GFP_K26TAG-His_6-3'
19	pLIKE-P_{groES}-5'AmyE-MAK33-VL-StrepII-3'-LacI
20	pLIKE-P_{groES}-5'AmyE-MAK33-N33TAG-VL-StrepII-3'-LacI
21	pLIKE-P_{groES}-5'AmyE-MAK33-VL-His_6-3'-LacI
22	pLIKE-P_{groES}-5'AmyE-MAK33-VL-N33TAG-His_6-3'-LacI
23	pLIKE-P_{groES}-5'LipA-D1.3scFv-His_6-3'
24	pLIKE-P_{groES}-5'LipA-D1.3scFv-K75TAG-His_6-3'
25	pLIKE-P_{groES}-5'LipA-D1.3scFV-S61TAG-His_6'
26	pLIKE-P_{groES}-5'LipA-D1.3scFV-S74TAG-His_6'
27	pLIKE-P_{groES}-5'LipA-D1.3scFV-S84TAG-His_6'
28	pLIKE-P_{groES}-5'AmyE-Am3-114-StrepII-3'
29	pJOE+secA_TAA+sgRNA_secA

30	pJOE+secG_TAA+sgRNA_secG
31	pJOE+secY_TAA+sgRNA_secY_I
32	pJOE+secY_TAA+sgRNA_secY_II
33	pJOE+secY_TAA+sgRNA_secY_III
34	pJOE+secY_TAA+sgRNA_secY_IV

4.1.7. Bacterial strains

Table 8: *E. coli* K12 strains

Strain	Genotype	Usage
DH5α	F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG _80d*lacZ*ΔM15 Δ(*lacZYAargF*) U169, hsdR17(rK- mK+), *λ*–	Cloning & plasmid amplification
Xl1-Blue	endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB$^+$ lacI^q Δ(lacZ)M15] hsdR17(r_K^- m_K^+)	Cloning & plasmid amplification
Xl10-Gold	endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tet^R F'[proAB lacI^{q}ZΔM15 Tn10(Tet^R Amy Cm^R)]	Cloning & plasmid amplification

Table 9: Used and created *B. subtilis* strains

Strain	Genotype	Reference/Usage
168		PCR amplification
K07	*ΔnprE ΔaprE Δepr Δmpr ΔnprB Δvpr Δbpr*	Protease deficient strain of *B. subtilis* WT PY79 BGSC ID 1A1133
K07^{Amber}	K07 + *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* *thrC*::4x*pylT-lacI*	Amber suppression strain; incorporation of *endo*-norbornene lysine
PG10	Genome reduced strain of 168	(37, 38)
PG10^{secA}	PG10_*secA*^{TAA}	Engineering the SEC-system
PG10^{secGA}	PG10_*secA*^{TAA}_*secG*^{TAA}	
PG10$^{Amber\text{-}Sec}$	PG10_secGA_TAA + *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* *thrC*::P_{secG}-*secG*^{TAA}-*secY*^{TAA}-4x*pylT*-P_{sigx}-5'*prsA*3'	Amber suppression strain; incorporation of *endo*-norbornene lysine, SEC system rescued

K07-1	K07 *amyE::*pSBBs1c-P_{veg}-GFP	Optimization of the expression system
K07-2	K07 + pLIKE-P_{veg}-GFP	
K07-3	K07 + pLIKE-P_{veg}-5'Stem-GFP-His_6	
K07-4	K07 + pLIKE-P_{veg}-GFP-His_6-3'-Term	
K07-5	K07 + pLIKE-P_{veg}-5'Stem-GFP-His_6-3'-Term	
K07-6	K07 + pLIKE-P_{groES}-5'-GFP-His_6-3'	
K07-7	K07 + pLIKE-P_{43}-5'-GFP-His_6-3'	
K07-8	K07 + pLIKE-P_{sigx}-5'-GFP-His_6-3'	
K07-9	K07 + pLIKE-P_{trnQ}-5'-GFP-His_6-3'	
K07-10	K07 *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* + pLIKE-P_{groES}-5'-GFP-His_6-3'	Proof of principle of amber suppression in *B. subtilis*
K07-11	K07 *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* + pLIKE-P_{groES}-5'-GFP_K26TAG-His_6-3'	
K07-12 **= K07^{Amber}-1**	K07 *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* *thrC*::4x*pylT-lacI* + pLIKE-P_{groES}-5'-GFP_K26TAG-His_6-3'	
K07-13	K07+pLIKE-P_{groES}-5'AmyE-MAK33-VL-StrepII-3'-LacI	Expression of pharmaceutical relevant proteins
K07-14	K07+ pLIKE-P_{groES}-5'AmyE-MAK33-VL-His_6-3'-LacI	
K07-15	K07+ pLIKE-P_{groES}-5'LipA-D1.3scFv-His_6-3'	
K07^{Amber}-2	K07 *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* *thrC*::4x*pylT-lacI* + pLIKE-P_{groES}-5'AmyE-MAK33-VL-N33TAG -GFP-His_6-3'-LacI	Target site evaluation for MAK33-VL
K07^{Amber}-3	K07 *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* *thrC*::4x*pylT-lacI* + pLIKE-P_{groES}-5'AmyE-MAK33-VL-L41TAG -GFP-His_6-3'-LacI	
K07^{Amber}-4	K07 *amyE*::P_{sigx}-5'*pylS*-3'- *pylT*	

	thrC::4x*pylT-lacI* + pLIKE-P_{groES}-5'AmyE-MAK33-VL-Y51TAG -GFP-His_6-3'-LacI	
K07$^{Amber-5}$	K07 *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* *thrC*::4x*pylT-lacI* + pLIKE-P_{groES}-5'AmyE-MAK33-N33TAG-VL-His_6-3'-LacI	Expression of MAK33-VL^{Norb33}
K07$^{Amber-6}$	K07 *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* *thrC*::4x*pylT-lacI* + pLIKE-P_{groES}-5'AmyE-MAK33-N33TAG-VL-Strep-3'-LacI	
PG10	PG10 + pLIKE-P_{groES}-5'AmyE-Am3-114-StrepII-3'	Expression of pharmaceutical relevant proteins
PG10$^{Amber-Sec}$-1	PG10_*secGA*^{TAA} + *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* *thrC*::P_{secG}-*secG*^{TAA}-*secY*^{TAA}-4x*pylT*-P_{sigx}-5'*prsA*3' + pLIKE-P_{groES}-5'AmyE-MAK33-N33TAG-VL-His_6-3'-LacI	Expression of MAK33-VL^{Norb33}
PG10$^{Amber-Sec}$-2	PG10_*secGA*^{TAA} + *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* *thrC*::P_{secG}-*secG*^{TAA}-*secY*^{TAA}-4x*pylT*-P_{sigx}-5'*prsA*3' + pLIKE-P_{groES}-5'AmyE-MAK33-N33TAG-VL-Strep-3'-LacI	
PG10$^{Amber-Sec}$-3	PG10_*secGA*^{TAA} + *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* *thrC*::P_{secG}-*secG*^{TAA}-*secY*^{TAA}-4x*pylT*-P_{sigx}-5'*prsA*3' + pLIKE-P_{groES}-5'LipA-D1.3scFv-His_6-3'	Expression of D1.3scFv with possible incorporation sites
PG10$^{Amber-Sec}$-4	PG10_*secGA*^{TAA} + *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* *thrC*::P_{secG}-*secG*^{TAA}-*secY*^{TAA}-4x*pylT*-P_{sigx}-5'*prsA*3' + pLIKE-P_{groES}-5'LipA-D1.3scFv-K75TAG-His_6-3'	

PG10$^{Amber\text{-}Sec}$-5	PG10_*secGA*^{TAA} + *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* *thrC*::P_{secG}-*secG*^{TAA}-*secY*^{TAA}-4x*pylT*-P_{sigx}-5'*prsA*3' + pLIKE-P_{groES}-5'LipA-D1.3scFV-S61TAG-His_6'	
PG10$^{Amber\text{-}Sec}$-6	PG10_*secGA*^{TAA} + *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* *thrC*::P_{secG}-*secG*^{TAA}-*secY*^{TAA}-4x*pylT*-P_{sigx}-5'*prsA*3' + pLIKE-P_{groES}-5'LipA-D1.3scFV-S74TAG-His_6'	Expression of D1.3scFv with possible incorporation sites
PG10$^{Amber\text{-}Sec}$-7	PG10_*secGA*^{TAA} + *amyE*::P_{sigx}-5'*pylS*-3'- *pylT* *thrC*::P_{secG}-*secG*^{TAA}-*secY*^{TAA}-4x*pylT*-P_{sigx}-5'*prsA*3' + pLIKE-P_{groES}-5'LipA-D1.3scFV-S84TAG-His_6'	

4.1.8. Buffers

Table 10: Buffers for protein purification, DNA manipulation or western blotting

Buffer	Components
HisLysis Buffer	50 mM NaH_2PO_4 500 mM NaCl 10 mM Imidazole pH 8.0
HisElution Buffer	50 mM NaH_2PO_4 500 mM NaCl 500 mM Imidazole pH 8.0
Strep Buffer W	100 mM Tris/HCl 150 mM NaCl 1 mM EDTA pH 8.0
Strep Elution Buffer	100 mM Tris/HCl 150 mM NaCl 1 mM EDTA 50 mM Biotin pH 8.0

TBS	50 mM Tris/HCl 150 mM NaCl pH 7.2
TBST	50 mM Tris/HCl 150 mM NaCl 0.05 % Tween20 pH 7.2
PBS	10 mM Na_2HPO_4 2 mM KH_2PO_4 137 mM NaCl 2.7 mM KCl pH 7.4
PBST	10 mM Na_2HPO_4 2 mM KH_2PO_4 137 mM NaCl 2.7 mM KCl 0.05 % Tween20 pH 7.4
5x SLIM H-buffer	125 mM Tris/HCl pH 9.0 750 mM NaCl 100 mM EDTA pH 8.0
5xSDS-sample buffer	312.5 mM Tris/HCl pH 6.8 10 % (w/v) SDS 0.01 % (w/v) Bromphenol blue 50 % (v/v) Glycerol 100 mM DTT
SDS-running buffer	25 mM Tris/HCl 192 mM Glycine 0.1 % (w/v) SDS
SDS-PAA stacking gel	125 mM Tris/HCl pH 8.6 5 % (w/v) Acrylamide/Bis-acrylamide mix (29:1) 0.1 % (w/v) SDS 0.1 % (w/v) APS 0.1 % (v/v) TEMED

SDS-PAA resolving gel	375 mM Tris/HCl pH 8.6
	15-18 % (w/v) Acrylamide/Bis-acrylamide mix (29:1)
	0.1 % (w/v) SDS
	0.1 % (w/v) APS
	0.1 % (v/v) TEMED
Coomassie staining solution	250 mg l^{-1} Coomassie R/G250
	40 % (v/v) Ethanol
	10 % (v/v) Acetic acid
Western Blot transfer buffer	80 % (v/v) SDS-running buffer
	20 %(v/v) Methanol
TAE Buffer	40 mM Tris/HCl pH 8.0
	40 mM Acetic acid
	1 mM EDTA

4.2. Software, databases and tools

Software

Table 11: Used software and online databases

Product	**Purpose**
Serial Cloner (http://serialbasics.free.fr/)	Creation of plasmid maps
ChemDraw 16.0 (Perkin Elmer)	Chemical structures
PyMOL (DeLano Scientific LLC, Schrödinger)	Protein structures
Origin 2018b (OriginLab Corp.)	Data analysis
Endnote (Clarivate Analytics)	Reference management
Affinity Designer (Serif Europe Ltd)	Graphic design

Online databases

Expasy (https://www.expasy.org/)	Protein related calculations
Subtiwiki (http://subtiwiki.uni-goettingen.de/)	*B. subtilis* database

4.3. Molecular biology methods

4.3.1. *E. coli* cultivation and transformation

E. coli strains DH5α, XL1-Blue and XL10-Gold were all cultivated in LB medium containing the respective antibiotics. Ampicillin was used at a final concentration of 100µg ml^{-1} and kanamycin was used at a final concentration of 30µg ml^{-1}.

For transformation with plasmid DNA, 45 µl of electro competent *E. coli* cells were mixed with 1-2 µl of DNA to be transformed with and transferred into precooled electroporation cuvettes (2 mm slits, VWR). The bacteria were shocked with 2.5 kV and rescued directly with 750 µl prewarmed SOC medium. After incubation at 37°C for 45 minutes, the bacteria were plated on prewarmed LB agar plates containing the respective antibiotics and incubated at 37°C overnight.

4.3.2. *B. subtilis* cultivation and transformation

B. subtilis strains based on the K07 and PG10 strain were cultivated in the defined and complex media. Depending on the used plasmids, the following antibiotics were used: mls (erythromycin 1 µg ml^{-1}, lincomycin 25 µg ml^{-1}), chloramphenicol (5 µg ml^{-1}), spectinomycin (100 µg ml^{-1}), kanamycin (5 µg ml^{-1}).

Transformation of the naturally competent, protease deficient strain K07, which is derived from the wild type strain PY79 (BGSC ID 1A1133) was carried out as described in (148). In brief, 10 ml of MNGE medium are inoculated from a fresh over-night culture (starting OD_{600} = 0.1) and the cells are grown in a shaker at 37°C until an OD_{600} of 1.1-1.3 is reached. Following the addition of DNA (1-2 µg) to the cells and further incubation (1 hour, 37°C, shaking) the Expression Mix is added, and the cells are again grown for 1 hour at 37°C under shaking. Replicative plasmids are added directly to the cells, integrating plasmids are linearized using *Sca*I-HF to allow integration into the genome. To select for positive transformants, the cells are plated on LB-agar plates supplemented with the respective antibiotics and incubated overnight at 37°C. Positive transformants are verified by single colony PCR.

PG10-based strains were transformed according to a protocol provided by Jörg Stülke (University of Göttingen, Germany).(37, 38) During the genome reduction, the natural competence got partially lost, thus a mannitol-inducible competence cassette (*comK*/*comS*) was integrated into the genome.

5 ml of LB are inoculated with 500 µl from an overnight culture of *B. subtilis* strain PG10 and the cells are grown at 37°C in a shaker for 1.5 hours. To induce the competence, 5ml LB containing 0.5 % (w/v) mannitol and 5 mM $MgCl_2$ are added to the culture and grown for 1.5 hours. To transform the cell with the target plasmid, 1 ml from the culture is centrifuged at 7000 x g for 5 minutes and the

cell pellet is resuspended in 1 ml fresh LB supplemented with 1-2 µg of plasmid DNA. Following incubation under shaking for 1 hour at 37°C, 100 µl of the culture are plated onto selective agar plates and incubated overnight at 37°C. Positive transformants are verified by single colony PCR.

4.3.3. Agarose gel electrophoresis

DNA fragments and PCR products are separated by their size using 1 % - 2 % (w/v) agarose gels. DNA is stained using Roti®-GelStain Red (1:50 000) and the gels are analyzed using a G:Box from Syngene.

4.3.4. Polymerase chain reactions

4.3.4.1. Standard PCR

Standard PCRs were carried out using the Q5 DNA Polymerase (NEB Biolabs) under the manufacturer's instructions. The optimal annealing temperature was individually determined via a gradient PCR. $MgSO_4$ and Betaine were used as additives to optimize the product yields if necessary.

4.3.4.2. Colony PCR

Colony PCRs were carried out to verify the presence of the transformed plasmid in the respective *E. coli* or *B. subtilis* colony. The GoTaq G2 polymerase (Promega GmbH) was used in all cases except for the CRISPR/Cas9 experiments where the OneTaq DNA Polymerase (NEB Biolabs) was used. Clones were picked from the plate and added to the PCR mix to serve as a DNA template. Positive clones were cultivated overnight at 37°C in LB medium containing the respective antibiotics for further processing.

4.3.4.3. QuikChange PCR

To introduce point mutations into the DNA sequence, a QuikChange I PCR was carried out under the manufacturer's guidelines (Agilent Technologies Inc.). For this, forward and reverse primers were designed using the Agilent Technologies Inc. web tool, which carry the mutation of interest.

4.3.4.4. Overlapping PCR

The C-terminal MAK33-VL GFP-fusion expression constructs, including a serine-glycine linker sequence between the two proteins were cloned by overlapping primer-extension PCR. Both *MAK33-VL* and *gfp* were each amplified from the template plasmids (oligonucleotides 24-27). After purification, the two DNA fragments, containing the serine-glycine linker sequence as an overlap, were first fused together using Q5 DNA polymerase. In absence of oligonucleotides 24/27, the Q5 DNA polymerase uses the

complementary parts of the two DNA templates to create the fusion product, which in the next step is amplified using oligonucleotides 24/27. After purification, the fusion constructs are digested with *Xba*I/*Xho*I and cloned into the pLIKE expression vector, resulting in plasmid #10.

4.3.5. Cloning strategies

4.3.5.1. Promoter screening

From previous works, the strong promoter P_{veg} was already available in the pLIKE plasmid. From a promoter screening carried out in *B. subtilis* by Nikoloff *et al.* , several other strong promoters were reported.(96) To re-evaluate these findings, the promoters were cloned in the pLIKE expression plasmid.

For this, the promoter region was amplified from *B. subtilis* 168 genomic DNA using the oligonucleotides 1 to 8 using the Q5 DNA polymerase. The PCR products were digested with *Eco*RI and *Bam*HI and purified. The pLIKE-GFP plasmid #1 containing the P_{veg}-5'*gfp*-His_6-3' construct was digested with *Eco*RI and *Bam*HI and subsequently dephosphorylated using Antarctic phosphatase.(89) After purification via an agarose gel, the backbone and the inserts were ligated using T4 DNA Ligase (NEB Biolabs, 30 min, 22°C) and transformed into *E. coli* DH5α. Positive clones were identified via a single colony PCR using the respective oligonucleotides used for the promoter region amplification and the plasmid was extracted after overnight cultivation in LB^{Amp} medium. After verification by Sanger sequencing (GATC), the plasmids #1-5 were transformed into *B. subtilis* strain K07. Positive clones were verified using a single colony PCR (oligonucleotides 70/71) and pooled to level out differences between single colonies. The promoter strength and behavior of *B. subtilis* strains K07-5-9 were subsequently assessed by measuring the GFP fluorescence.

4.3.5.2. mRNA stabilization

Sequences coding for the 5' stem-loop as well as 3' terminator were previously reported by Phan T.T. *et al*.(93) These increase mRNA-stability against the degradation from RNases and in turn the overall protein expression. The sequences were inserted into the expression construct using the Site-directed, Ligase-Independent PCR (SLIM-PCR) method.(94) Using oligonucleotides 9-16, the plasmids #7 (3'-stabilised *gfp*), plasmid #8 (5' stabilized construct) and plasmid #9 (5' and 3', stabilized construct) were created and transformed into *E. coli* strain Xl1-blue. Positive colonies were identified by Sanger sequencing. The created expression plasmids #6-9 were than transformed into *B. subtilis* K07 , creating the strains K07 2-5, and the effect of the different secondary structure elements were assessed by measuring the GFP-fluorescence in the TECAN Spark 10M plate reader.

4.3.5.3. C-terminal MAK33-VL-GFP-fusion constructs

The amber stop codons at the positions Asn33, Leu41 and Tyr51 were inserted using QuikChange I PCR (Agilent Technologies Inc.) using oligonucleotides 28-34, creating the plasmids #11-13.

To access amber suppression rates of the chosen incorporation sites, plasmids #10-13 were transformed into the K07^{Amber} strain, creating the strains K07^{Amber}-2-4. MCSE medium was inoculated (1:100) from a LB overnight culture and grown for 3 hours to let the cells adapt to the new medium.(88) The bacterial cultures were supplemented with 1 mM IPTG and 1 mM of endo-norbornene lysine and transferred to a 96 well plate (Flat black µclear, Greiner Bio-One International GmbH). The cells are grown at 37°C under shaking in the plate reader and the OD_{600} and the fluorescence (480 nm excitation, 527 nm emission) recorded every 10 min. At least five independent wells were analyzed for each condition to minimize effects of the plate layout.

4.3.5.4. Strain engineering by CRISPR/Cas9

The single cloning steps for all three SEC genes are in likewise fashion: The genomic region for the repair of the double strand break is amplified from the *B. subtilis* 168 genome using the corresponding primers 48/49 (*secA*), 52/53 (*secG*) or 58/59 (*secY*). After digest with *Sal*I/*Xba*I, the fragment is ligated using T4-DNA ligase with the similarly digested and dephosphorylated plasmid backbone of pJOE8999 and subsequently transformed into *E. coli* by electroporation. Positive clones are confirmed by a single colony PCR, control digest and subsequent Sanger sequencing of the extracted plasmids. To mutate the amber stop codon and to introduce the *Hind*III cleavage site a QuikChange I PCR is carried out using oligonucleotides 50/51 (*secA*), 54/55 (*secG*) or 60/61 (*secY*). Clones from the QuikChange I PCR are screened using the *Hind*III cleavage site and the positive plasmid is extracted from an overnight culture. The plasmid is digested using *Bsa*I-HF and subsequently dephosphorylated. To insert the sgRNA, two oligonucleotides 46/47 (*secA*), 56/57 (*secG*) or 62-69 (*secY*) are designed as stated in the publication, phosphorylated using T4 polynucleotide kinase (NEB Biolabs) and ligated into the digested plasmid.(142) The created plasmid is transformed into *B. subtilis* PG10 and positive clones are verified via a single-colony PCR using OneTaq DNA polymerase and subsequent digest using *Hind*III-HF. Plasmid curing is carried out as stated in the publication.

This procedure was repeated for each of the three SEC genes.

4.3.5.5. Generation of the expression plasmid

The expression plasmid created in this work is based on the pLIKE plasmid. Via thorough design, a template plasmid, as shown in the figure below was created when MAK33-VL was chosen as a target protein for the protein secretion.

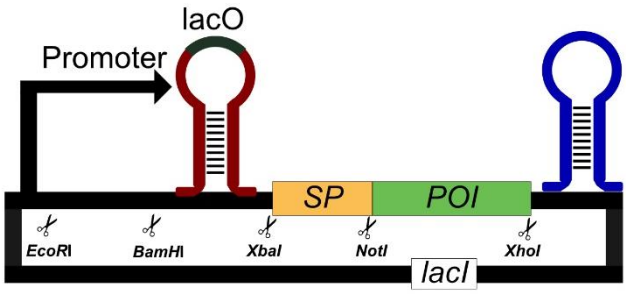

Figure 61: Overview of the pLIKE expression plasmid

The advantage of this setup is that each component can be easily exchanged, maximizing the flexibility for a new protein of interest.

The MAK33-VL gene was amplified using primers 19/21 and digested with *Not*I-HF and *Xho*I. The signalpeptide *amyE* was amplified from the *B. subtilis* 168 genome using primers 22/23 and digested with *Xba*I and *Not*I. *Not*I was chosen as the restriction endonuclease site between the signal peptide and the target protein because of its sequence "GCGGCGC". By adding an adenosine at the 9th position, the DNA is translated to AAA, which is in turn the signal peptide cleavage site for the signal peptidases.

The target proteins D1.3scFv and Am3-114 were cloned in a likewise fashion. Both genes were ordered from Invitrogen and amplified using the oligonucleotides 34/35 (D1.3scFv) and 42/43 (Am3-114). By digestion with the corresponding restriction enzymes, the genes were cloned into the expression plasmid.

4.3.5.6. Amber suppression plasmid

To perform amber suppression, a bioorthogonal aminoacyl tRNA synthetase and its corresponding tRNA must be expressed in the target organism. Here, an excessive overexpression is equally adverse as a non-sufficient expression of both components. Additionally, tRNA maturation greatly differs between *E. coli* and *B. subtilis*, making the correctly processed tRNA a further challenge. Therefore, the moderate promoter P_{sigX} is used to control *pylS*^{Norb} from *M. mazei*, which is flanked by the 5' and 3' mRNA stabilizing sequences. The corresponding tRNA *pylT* is flanked by sequences of the tRNA^{Ser} (see Figure 25) to ensure correct processing. Both constructs are cloned into the pSBBS1c plasmid, creating plasmid #14 which integrates into the *amyE*-locus after transformation in *B. subtilis*.(88)

4.3.5.7. Secretion plasmid

The plasmid pXT, integrating in the *thrC*-locus was used to further optimize the amber suppression efficiency and fully restore the protein secretion.

In a first step, the pXT plasmid was equipped with four additional copies of *pylT*. The four tRNA copies were placed in a genomic region as shown with 1x *pylT* in the amber suppression plasmid #14 (Figure 25). A copy of *lacI* was also integrated to further improve the regulation of the expressed proteins, creating plasmid #15.

In the second step to restore the protein secretion in *B. subtilis*, the 5' and 3' stabilized *prsA* gene under control of P_{sigx} (creating plasmid #16) and $secG^{TAA}$/$secY^{TAA}$ under control of P_{secG} were added to the plasmid, creating plasmid #17.

4.4. Protein chemical methods

4.4.1. SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

To separate proteins by their size, an SDS-PAGE is carried out. For the analysis of GFP and D1.3scFv, 15 % Polyacrylamide (PAA) gels are used, for MAK33-VL and Am3-114 18 % PAA. Protein samples are mixed with 5 x SDS loading buffer and heated to 90°C for 10 minutes. After cooling to RT, 5 µl to 20 µl depending on the sample concentration are loaded into the SDS gels. Gels are run at 35 mA per gel for approximately 35 minutes.

The gels are stained with Coomassie staining solution and destained using 10% acetic acid.

4.4.2. Protein expression and purification

4.4.2.1. Intracellular protein expression

GFP (plasmid #9) and GFP^{Norb26} (plasmid #18) were expressed in *B. subtilis* strain $K07^{Amber}$. Protein expression was carried out in 2YT medium supplemented with erythromycin (1 µg ml^{-1}) and lincomycine (25 µg ml^{-1}). After inoculation from an overnight culture to an OD_{600} of 0.1, the cells were grown at 37°C under shaking until OD_{600} = 1.0 was reached. *Endo*-norbornene lysine was added to a final concentration of 1 mM and bacteria grown for one hour at 37°C to allow adaption to the ncAA. After that, IPTG was added to a final concentration of 1 mM and protein expression was carried out overnight at 25°C.

4.4.2.2. Extracellular protein expression

4.4.2.2.1. Expression of MAK33-VL

For the expression of His_6- and StrepII-tagged MAK33-VL and MAK33-VL^{Norb33}, the respective plasmids #19-22 were transformed into the expressions strain $K07^{Amber}$, $PG10^{Amber}$ and $PG10^{Amber\text{-}Sec}$. 2YT medium supplemented with erythromycin (1 µg ml^{-1}) and lincomycine (25 µg ml^{-1}) was inoculated from overnight cultures to and OD_{600} of 0.1 and the cells grown until an OD_{600} of 1.0 was reached. The expression of the MAK33-VL was induced with 1 mM IPTG and the cultures grown overnight at 25°C. For the expression of the ncAA containing variant MAK33-VL^{Norb33}, 1 mM of *endo*-norbornene lysine was added and the bacteria grown for 1 hour at 37°C to adapt to the new condition. After that, IPTG is added to a final concentration of 1 mM and protein expression carried out overnight at 25°C.

4.4.2.2.2. Expression of D1.3scFv

To express the wild type protein, the respective plasmid #23 was transformed into the expression strain K07 and $PG10^{Amber\text{-}Sec}$. Initially, the strain K07 was used to test different expression media. Therefore, D1.3scFv was expressed and secreted in the defined medium MCSE and three complex media (LB, 2YT and TB). Following this, D1.3scFv was expressed in 2YT medium at 25°C overnight.

To incorporate the ncAA, the plasmids #24-27 were transformed into $PG10^{Amber\text{-}Sec}$, creating the strains $PG10^{Amber\text{-}Sec}$-3-7, and protein expression was carried using the established protocol from MAK33-VL^{Norb33}.

4.4.2.2.3. Expression of Am3-114

As both previously tested proteins MAK33-VL and D1.3scFv showed the best yields in 2YT medium, Am3-114 was expressed under the same conditions as well. The respective plasmid #28 was transformed into *B. subtilis* strain PG10. Positive clones were verified using single colony PCR and pooled to level out differences between clones. Expression was carried out in 2YT medium overnight at 25°C.

4.4.2.3. Protein purification

Purification of GFP and GFP^{Norb26} from the cells was performed using His SpinTrap columns (GE Healthcare, USA). The cell pellet was resuspended in HisLysis Buffer (50 mM NaH_2PO_4, 500 mM NaCl, 10 mM Imidazole, pH 8.0) and lysed using sonification. The lysate was cleared by centrifugation (20

min, 16000 x g, 4°C) and loaded onto the His SpinTrap columns. The columns were washed four times and GFP^{Norb26} was eluted from the column using 600 µl HisElution Buffer (50 mM NaH_2PO_4, 500 mM NaCl, 500 mM Imidazole, pH 8.0). The protein was analyzed by SDS-PAGE and verified using mass spectrometry.

Secreted proteins (MAK33-VL, MAK33-VL^{Norb33}, D1.3scFv, D1.3scFv^{Norb75} and Am3-114) were purified from the medium by first removing the cells through centrifugation (7000 x g, 10 min, 4°C), followed by precipitating the proteins from the remaining supernatant by adding 0.44 g ammonium sulfate per ml culture volume and incubation for 1 h at 8°C under gentle agitation. The precipitate is recovered through centrifugation (9000 x g, 30 min, 4°C) and the pellet resuspended in 1/25 of the culture volume. Affinity purification using either Strep-Tactin®XT Spin Column Kit (IBA GmbH, Germany) or His SpinTrap (GE Healthcare, USA) was carried out according to the manufacturer's instructions. Proteins were analyzed by SDS-PAGE and Western botting using Strep-Tactin-AP Conjugate (IBA GmbH, Germany) for the detection of strep-tagged proteins.

4.4.2.4. Western blotting

For western blotting, the proteins separated by an SDS-PAGE were transferred to an Immobilon-E PVDF membrane (Merck KgaA) using Biometra Fastblot B43 (Analytik Jena GmbH) for semi-dry blotting at 2.5 mA/cm² for 20 minutes. After transfer, the membrane was rinsed with ddH_2O and unsaturated binding sites were blocked with 5 % bovine serum albumin in TBST buffer for 60 minutes. The α-His-antibody (produced in mouse) was added to the blocking solution at a concentration of 1:5000 and binding to the His_6-tag was allowed for 60 minutes. After washing with TBST (3 x 10 minutes), the α-mouse antibody containing the alkaline phosphatase was added in 5 % BSA in TBST at a concentration of 1:10000 and incubated for 60 minutes.For developing, the membrane was washed with TBST (3 x 5 minutes) and TBS (3 x 5 minutes) and incubated in BCIP®/NBT-Blue Liquid Substrate solution (Sigma Aldrich) until bands became visible. For western blotting against the StrepII-tag, Strep-Tactin-AP Conjugate (IBA GmbH, Germany) was used. After transfer as stated above, the membrane was treated as described in the manufacturer's instructions.

4.5. Fluorescence measurements

All fluorescence measurements were carried out in a TECAN Spark 10M plate reader. For measuring the GFP fluorescence (excitation wavelength 480 nm, emission wavelength 527 nm), the bacteria were pre-grown in the respective media and the OD_{600} adjusted to 0.1. After inoculation of the 96 well plates,

the bacteria were grown at 37°C under shaking and the fluorescence as well as the OD_{600} were measured every 10 minutes.

The fluorescence label TAMRA of the tetrazine-ligand **4** was excited at 550nm and the emission was measured at 600nm.

4.6. Verification of functionality

The functionality of the in *B. subtilis* expressed MAK33-$VL^{Norb33\text{-}Fluoro}$ is verified by testing the binding to its target protein, the human creatine kinase subtype MM (hmCK; Roche Diagnostics, Mannheim). Therefore, biotinylated hmCK (5 pM) is mixed with MAK33-VL-$His_6^{Norb33\text{-}Fluoro}$ (0-1500 pM) in a total volume of 200 µl PBS, supplemented with 1x protein-free blocking agent (Visual protein BlockPro Protein Free Blocking reagent, Energenesis Biomedical CO. LTD, Taiwan). Following incubation of 5 min at RT, the reactions are transferred to streptavidin-coated a 96 well plate (Greiner Bio-One International GmbH) and incubated for 1h at room temperature under constant shaking. Unbound protein is removed by washing with 1x BlockPro (2x 200 µl, 5 min) and water (2x 200 µl, 5 min). The amount of MAK33-$VL^{Norb33\text{-}Fluoro}$ bound to hmCK is determined by measuring the fluorescence of the label TAMRA (excitation 550 nm, emission 600 nm) in the TECAN Spark 10M plate reader.

In addition, hmCK (5 pM) were mixed with MAK33-VL-$His_6^{Norb33\text{-}Fluoro}$ (1000 pM) and varying amounts of unmodified MAK33-VL (0-10000 pM) in a total volume of 200 µl. Following incubation for 1 h at RT, the reaction is transferred to a streptavidin coated 96 well plate (Greiner Bio-One International GmbH) and incubated for 1 h at room temperature under constant shaking. Unbound protein is removed by washing with 1x BlockPro (2x 200 µl, 5 min) and water (2x 200 µl, 5 min). The amount of MAK33-$VL^{Norb33\text{-}Fluoro}$ bound to hmCK is determined by measuring the fluorescence of the label TAMRA (excitation 550 nm, emission 600 nm) in the TECAN Spark 10M plate reader.

4.7. Mass spectrometry

Liquid chromatography-mass spectrometry (LC-MS) was carried out on an Agilent Technologies 1260 Infinity LC–MS system with a 6310-quadrupole spectrometer. The solvent system consisted of 0.1 % formic acid in water (buffer A) and 0.1% formic acid in acetonitrile (buffer B).

Bibliography

1. Friedländer C. Die Mikrokokken der Pneumonie. Fortschr Med. **1883**; 1: 19.

2. Cohn F. J. Untersuchung über Bakterien. Beitr Bio Pflanz. **1872**; 1: 101.

3. Ehrenberg C. G. Dritter Beitrag zur Erkenntniss grosser Organisation in der Richtung des kleinsten Raumes. Druckerei der Königl. Akademie der Wissenschaften**1835**.

4. Conn H. J. The identity of *Bacillus subtilis*. J Infect Dis. **1930**; 46(4): 341-50.

5. Zeigler D. R., Prágai Z., Rodriguez S., Chevreux B., Muffler A., Albert T., et al. The origins of 168, W23, and other *Bacillus subtilis* legacy strains. J Bacteriol. **2008**; 190(21): 6983-95.

6. Burkholder P. R., Giles N. H. Induced biochemical mutations in *Bacillus subtilis*. Am J Bot. **1947**; 34: 4.

7. Dubnau D. Genetic competence in *Bacillus subtilis*. Microbiol Rev. **1991**; 55(3): 30.

8. Sewalt V., Shanahan D., Gregg L., La Marta J., Carrillo R. The generally recognized as safe (GRAS) process for industrial microbial enzymes. Ind Biotechnol. **2016**; 12(5): 295-302.

9. Raetz C. R., Whitfield C. Lipopolysaccharide endotoxins. Annu Rev Biochem. **2002**; 71: 635-700.

10. Tsirigotaki A., De Geyter J., Sostaric N., Economou A., Karamanou S. Protein export through the bacterial Sec pathway. Nat Rev Microbiol. **2017**; 15(1): 21-36.

11. Green E. R., Mecsas J. Bacterial secretion systems: an overview. Microbiol Spectr. **2016**; 4(1).

12. Goosens V. J., Monteferrante C. G., van Dijl J. M. The Tat system of gram-positive bacteria. Biochim Biophys Acta. **2014**; 1843(8): 1698-706.

13. Palmer T., Berks B. C. The twin-arginine translocation (Tat) protein export pathway. Nat Rev Microbiol. **2012**; 10: 14.

14. Harwood C. R., Cranenburgh R. *Bacillus* protein secretion: an unfolding story. Trends Microbiol. **2008**; 16(2): 73-9.

15. Voigt B., Antelmann H., Albrecht D., Ehrenreich A., Maurer K. H., Evers S., et al. Cell physiology and protein secretion of *Bacillus licheniformis* compared to *Bacillus subtilis*. J Mol Microbiol Biotechnol. **2009**; 16(1-2): 53-68.

16. Maier B. Competence and transformation in *Bacillus subtilis*. Curr Issues Mol Biol. **2020**; 37: 57-76.

17. Gebhard S. ABC transporters of antimicrobial peptides in Firmicutes bacteria - phylogeny, function and regulation. Mol Microbiol. **2012**; 86(6): 1295-317.

18. Kolkman M. A. B., van der Ploeg R., Bertels M., van Dijk M., van der Laan J., van Dijl J. M., et al. The Twin-arginine signal peptide of *Bacillus subtilis* YwbN can direct either Tat- or Sec-dependent secretion of different cargo proteins: Secretion of active Subtilisin via the *B. subtilis* Tat pathway. Appl Environ Microbiol. **2008**; 74(24): 7507-13.

19. van Dijl J. M., Hecker M. *Bacillus subtilis*: from soil bacterium to super-secreting cell factory. Microb Cell Fact. **2013**; 12: 3-.

20. Wang G., Chen H., Zhang H., Song Y., Chen W. The secretion of an intrinsically disordered protein with different secretion signals in *Bacillus subtilis*. Curr Microbiol. **2013**; 66(6): 566-72.

21. Chen J., Fu G., Gai Y., Zheng P., Zhang D., Wen J. Combinatorial Sec pathway analysis for improved heterologous protein secretion in *Bacillus subtilis*: identification of bottlenecks by systematic gene overexpression. Microb Cell Fact. **2015**; 14(1).

22. de Souza G. A., Leversen N. A., Målen H., Wiker H. G. Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway. J Proteom. **2011**; 75(2): 502-10.

23. Westers L., Westers H., Quax W. J. *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. Biochim Biophys Acta. **2004**; 1694(1-3): 299-310.

24. Singh R., Kumar M., Mittal A., Mehta P. K. Microbial enzymes: industrial progress in 21st century. 3 Biotech. **2016**; 6(2): 174.

25. Guan C., Cui W., Cheng J., Liu R., Liu Z., Zhou L., et al. Construction of a highly active secretory expression system via an engineered dual promoter and a highly efficient signal peptide in *Bacillus subtilis*. New Biotechnol. **2016**; 33(3): 372-9.

26. Strohmeier G. A., Pichler H., May O., Gruber-Khadjawi M. Application of designed enzymes in organic synthesis. Chem Rev. **2011**; 111(7): 4141-64.

27. Reuss D. R., Commichau F. M., Gundlach J., Zhu B., Stulke J. The blueprint of a minimal cell: MiniBacillus. Microbiol Mol Biol Rev. **2016**; 80(4): 955-87.

28. Westers H., Dorenbos R., van Dijl J. M., Kabel J., Flanagan T., Devine K. M., et al. Genome engineering reveals large dispensable regions in *Bacillus subtilis*. Mol Biol Evol. **2003**; 20(12): 2076-90.

29. Ye R., Yang L. P., Wong S. L. Construction of protease-deficient *Bacillus subtilis* strains for expression studies: inactivation of seven extracellular protease and the intracellular LonA protease,. Proc International Symposium on Recent Advances in Bioindustry. **1996**: 10.

30. Kupfer D. G., Uffen R. L., Canale-Parola E. The role of iron and molecular oxygen in pulcherrimin synthesis by bacteria. Archiv für Mikrobiologie. **1967**; 56: 13.

31. Kluyver A. J., van der Walt J. P., van Triet A. J. Pulcherrimin the pigment of candida. Proc Natl Acad Sci. **1953**; 39: 10.

32. Kunst F., Ogasawara N., Moszer I., Albertini A. M., Alloni G., Azevedo V., et al. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature. **1997**; 390(6657): 249-56.

33. Commichau F. M., Pietack N., Stulke J. Essential genes in *Bacillus subtilis*: a re-evaluation after ten years. Mol Biosyst. **2013**; 9(6): 1068-75.

34. Mader U., Schmeisky A. G., Florez L. A., Stulke J. SubtiWiki--a comprehensive community resource for the model organism *Bacillus subtilis*. Nucleic Acids Res. **2012**; 40(Database issue): D1278-87.

35. Michna R. H., Commichau F. M., Todter D., Zschiedrich C. P., Stulke J. SubtiWiki-a database for the model organism *Bacillus subtilis* that links pathway, interaction and expression information. Nucleic Acids Res. **2014**; 42: D692-8.

36. Zhu B., Stulke J. SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. Nucleic Acids Res. **2018**; 46(D1): D743-D8.

37. Reuss D. R., Altenbuchner J., Mader U., Rath H., Ischebeck T., Sappa P. K., et al. Large-scale reduction of the *Bacillus subtilis* genome: consequences for the transcriptional network, resource allocation, and metabolism. Genome Res. **2017**; 27(2): 289-99.

38. Aguilar Suarez R., Stulke J., van Dijl J. M. Less Is more: Toward a genome-reduced *Bacillus c*ell factory for "difficult proteins". ACS Synth Biol. **2019**; 8(1): 99-108.

39. Ulmer K. M. Protein engineering. Science. **1983**; 219: 7.

40. Hutchison C. A., Phillips S., Edgell M. H., Gillam S., Jahnke P., Smith M. Mutagenesis at a specific position in a DNA sequence. J Biol Chem. **1978**; 253(18): 10.

41. Abuchowski A., McCoy J. R., Palczuk N. C., van Es T., Davis F. F. Effect of covalent attachment of polyethylene glycol on immunogenicity and circulating life of bovine liver catalase. J Biol Chem. **1977**; 252(11): 5.

42. Strohl W. R. Fusion proteins for half-life extension of biologics as a strategy to make biobetters. BioDrugs. **2015**; 29(4): 215-39.

43. Veronese F. M., Mero A. The impact of PEGylation on biological therapies. BioDrugs. **2008**; 22(5): 15.

44. Schlapschy M., Binder U., Borger C., Theobald I., Wachinger K., Kisling S., et al. PASylation: a biological alternative to PEGylation for extending the plasma half-life of pharmaceutically active proteins. Protein Eng Des Sel. **2013**; 26(8): 489-501.

45. Keefe D., Heartlein M., Josiah S. Transferrin fusion protein therapies: acetylcholine receptor-transferrin fusion protein as a model: Hoboken:Wiley; **2013**. 12 p.

46. Weimer T., Metzner H. J., Schulte S. Recombinant albumin fusion

proteins: Hoboken: Wiley; **2013**. 27 p.

47. Ravasco J., Faustino H., Trindade A., Gois P. M. P. Bioconjugation with maleimides: a useful tool for chemical biology. Chemistry. **2019**; 25(1): 43-59.

48. Christie R. J., Fleming R., Bezabeh B., Woods R., Mao S., Harper J., et al. Stabilization of cysteine-linked antibody drug conjugates with N-aryl maleimides. J Control Release. **2015**; 220(Pt B): 660-70.

49. Datta-Mannan A., Choi H., Stokell D., Tang J., Murphy A., Wrobleski A., et al. The properties of cysteine-conjugated antibody-drug conjugates are impacted by the IgG subclass. AAPS J. **2018**; 20(6): 103.

50. Birrer M. J., Moore K. N., Betella I., Bates R. C. Antibody-drug conjugate-based therapeutics: state of the science. J Natl Cancer Inst. **2019**.

51. Vazquez-Lombardi R., Nevoltris D., Luthra A., Schofield P., Zimmermann C., Christ D. Transient expression of human antibodies in mammalian cells. Nat Protoc. **2018**; 13(1): 99-117.

52. Goodman H. M., Abelson J., Landy A., Brenner S., Smith J. D. Amber suppression: a nucleotide change in the anticodon of a tyrosine transfer RNA. Nature. **1968**; 217(5133): 1019-24.

53. Beier H., Grimm M. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. Nucleic Acids Res. 29(23): 6.

54. Schultz P. G., Herberich B., Brock A., Wang L. Expanding the genetic code of *Escherichia coli*. Science. **2001**; 292: 3.

55. Wang L., Magliery T. J., Liu D. R., Schultz P. G. A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins. J Am Chem Soc. **2000**; 122(20): 2.

56. Hang H. C., Yu C., Kato D. L., Bertozzi C. R. A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. Proc Natl Acad Sci. **2003**; 100: 6.

57. Srinivasan G., James C. M., Krzycki J. A. Pyrrolysine encoded by UAG in archaea: charging of a UAG-decoding specialized tRNA. Science. **2002**; 296(5572): 1459.

58. Hao B., Gong W., Ferguson T. K., James C. M., Krzycki J. A., Chan M. K. A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science. **2002**; 296(5572): 1462.

59. Chin J. W. Expanding and reprogramming the genetic code of cells and animals. Annu Rev Biochem. **2014**; 83(1): 379-408.

60. Uttamapinant C., Howe J. D., Lang K., Beranek V., Davis L., Mahesh M., et al. Genetic code expansion enables live-cell and super-resolution imaging of site-specifically labeled cellular proteins. J Am Chem Soc. **2015**; 137(14): 4602-5.

61. Han S., Yang A., Lee S., Lee H. W., Park C. B., Park H. S. Expanding the genetic code of *Mus musculus*. Nat Commun. **2017**; 8: 14568.

62. Johnson D. B., Xu J., Shen Z., Takimoto J. K., Schultz M. D., Schmitz R. J., et al. RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. Nat Chem Biol. **2011**; 7(11): 779-86.

63. Xie J., Schultz P. G. Adding amino acids to the genetic repertoire. Curr Opin Chem Biol. **2005**; 9(6): 548-54.

64. Liu C. C., Schultz P. G. Adding new chemistries to the genetic code. Annu Rev Biochem. **2010**; 79: 413-44.

65. Brustad E., Bushey M. L., Lee J. W., Groff D., Liu W., Schultz P. G. A genetically encoded boronate-containing amino acid. Angew Chem Int Ed Engl. **2008**; 47(43): 8220-3.

66. Summerer D., Chen S., Wu N., Deiters A., Chin J. W., Schultz P. G. A genetically encoded fluorescent amino acid. Proc Natl Acad Sci **2006**; 103(26): 9785-9.

67. Agard N. J., Prescher J. A., Bertozzi C. R. A strain-promoted [3+2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. J Am Chem Soc. **2004**; 126: 2.

68. Deiters A., Cropp T. A., Summerer D., Mukherji M., Schultz P. G. Site-specific PEGylation of proteins containing unnatural amino acids. Bioorg Med Chem Lett. **2004**; 14(23): 5743-5.

69. Lang K., Davis L., Torres-Kolbus J., Chou C., Deiters A., Chin J. W. Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. Nat Chem. **2012**; 4(4): 298-304.

70. Lang K., Davis L., Wallace S., Mahesh M., Cox D. J., Blackman M. L., et al. Genetic encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic Diels-Alder reactions. J Am Chem Soc. **2012**; 134(25): 10317-20.

71. Schneider S., Gattner M. J., Vrabel M., Flugel V., Lopez-Carrillo V., Prill S., et al. Structural insights into incorporation of norbornene amino acids for click modification of proteins. Chembiochem. **2013**; 14(16): 2114-8.

72. Vrabel M., Kolle P., Brunner K. M., Gattner M. J., Lopez-Carrillo V., de Vivie-Riedle R., et al. Norbornenes in inverse electron-demand Diels-Alder reactions. Chemistry. **2013**; 19(40): 13309-12.

73. Rostovtsev V. V., Green G. G., Fokin V. V., Sharpless K. B. A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes**. Angew Chem Int Ed Engl. **2002**; 41(14): 4.

74. R. Huisgen, R. Grashey, Sauer J. Chemistry of alkenes. Interscience, New York**1964**.

75. Kolb H. C., Finn M. G., Sharpless K. B. Click Chemistry: Diverse chemical function from a few good reactions. Angew Chem Int Ed Engl. **2001**; 40(11): 2004-21.

76. Agard N. J., Baskin J. M., Prescher J. A., Lo A., Bertozzi C. R. A comparative study of bioorthogonal reactions with azides. ACS Chem Biol. **2006**; 1(10): 5.

77. Codelli J. A., Baskin J. M., Agard N. J., Bertozzi C. R. Second-generation difluorinated cyclooctynes for copper-free click chemistry. J Am Chem Soc. **2008**; 130: 8.

78. Diels O., Alder K. Synthesen in der hydroaromatischen Reihe. Liebigs Ann Chem. **1928**; 460(1): 98-122.

79. Blackman M. L., Royzen M., Fox J. M. Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. J Am Chem Soc. **2008**; 130(41): 13518-9.

80. Beatty K. E., Liu J. C., Xie F., Dieterich D. C., Schuman E. M., Wang Q., et al. Fluorescence visualization of newly synthesized proteins in mammalian cells. Angew Chem Int Ed Engl. **2006**; 45(44): 7364-7.

81. Dieterich D. C., Link A. J., Graumann J., Tirrell D. A., Schuman E. M. Selective identification of newly synthesized proteinsin mammalian cells using bioorthogonal noncanonicalamino acid tagging (BONCAT). Proc Natl Acad Sci. **2006**; 103(25): 6.

82. Huston J. S., Levinson D., Mudgett-Hunter M., Tai M. S., Novotný J., Margolies M. N., et al. Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. Proc Natl Acad Sci. **1988**; 85(16): 5.

83. Eisen H. N. Affinity enhancement of antibodies: how low-affinity antibodies produced early in immune responses are followed by high-affinity antibodies later and in memory B-cell responses. Cancer Immunol Res. **2014**; 2(5): 381-92.

84. Frenzel A., Hust M., Schirrmann T. Expression of recombinant antibodies. Front Immunol. **2013**; 4: 217.

85. Wu S.-C., Ye R., Wu X.-C., Ng S.-C., Wong S.-L. Enhanced secretory production of a single-chain antibody fragment from *Bacillus subtilis* by coproduction of molecular chaperones. J Bacteriol. **1998**; 180(11): 6.

86. Cablivi [press release]. https://www.ema.europa.eu/en/medicines/human/EPAR/cablivi: EMA2018.

87. FDA approves first therapy for the treatment of adult patients with a rare blood clotting disorder [press release]. https://www.fda.gov/news-events/press-announcements/fda-approves-first-therapy-treatment-adult-patients-rare-blood-clotting-disorder, 06.02.2019 2019.

88. Radeck J., Kraft K., Bartels J., Cikovic T., Durr F., Emenegger J., et al. The Bacillus BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with *Bacillus subtilis*. J Biol Eng. **2013**; 7(1): 29.

89. Toymentseva A. A., Schrecke K., Sharipova M. R., Mascher T. The LIKE system, a novel protein expression tollbox for *Bacillus subtilis* based on the *liaI* promoter. Microb Cell Fact. **2012**; 11(143): 13.

90. Arantes O., Lereclus D. Construction of cloning vectors for *Bacilus thuringiensis*. Gene. **1991**; 108: 5.

91. Condon C. RNA processing and degradation in *Bacillus subtilis*. Microbiol Mol Biol Rev. **2003**; 67(2): 157-74.

92. Haimovich G., Medina D. A., Causse S. Z., Garber M., Millan-Zambrano G., Barkai O., et al. Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. Cell. **2013**; 153(5): 1000-11.

93. Phan T. T., Nguyen H. D., Schumann W. Construction of a 5'-controllable stabilizing element (CoSE) for over-production of heterologous proteins at high levels in *Bacillus subtilis*. J Biotechnol. **2013**; 168(1): 32-9.

94. Chiu J., March P. E., Lee R., Tillett D. Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. Nucleic Acids Res. **2004**; 32(21).

95. Farnham P. J., Platt T. Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro. Nucleic Acids Res. **1981**; 9(3): 14.

96. Song Y., Nikoloff J. M., Fu G., Chen J., Li Q., Xie N., et al. Promoter screening from *Bacillus subtilis* in various conditions hunting for synthetic biology and industrial applications. PLoS One. **2016**; 11(7): 18.

97. Haldenwang W. G. The sigma factors of *Bacillus subtilis*. Microbiol Rev. **1995**; 59(1): 30.

98. Kim J.-H., Hwang B.-Y., Roh J., Lee J.-K., Kim K., Wong S.-L., et al. Comparison of P aprE , P amyE , and P P43 promoter strength for β-galactosidase and staphylokinase expression in *Bacillus subtilis*. Biotechnol Bioproc Eng. **2008**; 13(3): 313-8.

99. Wang P. Z., Doi R. H. Overlapping promoters transcribed by *Bacillus subtilis* σ^{55} and σ^{37} RNA polymerase holoenzymes during growth and stationary phases. J Biol Chem. **1984**; 259(13): 7.

100. Guzman L. M., Belin D., Carson M. J., Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol. **1995**; 177(14): 4121-30.

101. Wenzel M., Altenbuchner J. The *Bacillus subtilis* mannose regulator, ManR, a DNA-binding protein regulated by HPr and its cognate PTS transporter ManP. Mol Microbiol. **2013**; 88(3): 562-76.

102. Kreuzer P., Gärtner D., Allmansberger R., Hillen W. Identification and sequence analysis of the *Bacillus subtilis* W23 xylR gene and xyl operator. J Bacteriol. **1989**; 171(7): 3840.

103. Rahmer R., Morabbi Heravi K., Altenbuchner J. Construction of a super-competent *Bacillus subtilis* 168 using the P_{mtlA} -comKS inducible cassette. Front Microbiol. **2015**; 6: 1431.

104. Heravi K. M., Watzlawick H., Altenbuchner J. Development of an anhydrotetracycline-inducible expression system for expression of a neopullulanase in *B. subtilis*. Plasmid. **2015**; 82: 8.

105. Yansura D. G., Henner D. J. Use of the *Escherichia coli* lac repressor and operator to control gene expression in *Bacillus subtilis*. Proc Natl Acad Sci. **1984**; 81: 5.

106. Gossen M., Freundlieb S., Bender G., Müller G., Hillen W., Bujard H. Transcriptional activation by tetracyclines in mammalian cells. Science. **1995**; 268: 5.

107. Gossen M., Bujard H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci. **1992**; 89: 5.

108. Phan T. T., Tran L. T., Schumann W., Nguyen H. D. Development of Pgrac100-based expression vectors allowing high protein production levels in Bacillus subtilis and relatively low basal expression in Escherichia coli. Microb Cell Fact. **2015**; 14: 72.

109. Schwark D. G., Schmitt M. A., Fisk J. D. Dissecting the contribution of release factor interactions to amber stop codon reassignment efficiencies of the *Methanocaldococcus jannaschii* orthogonal pair. Genes (Basel). **2018**; 9(11).

110. Xu H., Wang Y., Lu J., Zhang B., Zhang Z., Si L., et al. Re-exploration of the codon context effect on amber codon-guided incorporation of noncanonical amino acids in *Escherichia coli* by the blue-white screening assay. Chembiochem. **2016**; 17(13): 1250-6.

111. Pott M., Schmidt M. J., Summerer D. Evolved sequence contexts for highly efficient amber suppression with noncanonical amino acids. ACS Chem Biol. **2014**; 9(12): 2815-22.

112. Schmied W. H., Elsasser S. J., Uttamapinant C., Chin J. W. Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. J Am Chem Soc. **2014**; 136(44): 15577-83.

113. Kaya E., Vrabel M., Deiml C., Prill S., Fluxa V. S., Carell T. A genetically encoded norbornene amino acid for the mild and selective modification of proteins in a copper-free click reaction. Angew Chem Int Ed Engl. **2012**; 51(18): 4466-9.

114. Wen T., Oussenko I. A., Pellegrini O., Bechhofer D. H., Condon C. Ribonuclease PH plays a major role in the exonucleolytic maturation of CCA-containing tRNA precursors in *Bacillus subtilis*. Nucleic Acids Res. **2005**; 33(11): 3636-43.

115. Zahler N. H. Recognition of the 5' leader of pre-tRNA substrates by the active site of ribonuclease P. RNA. **2003**; 9(6): 734-45.

116. Pellegrini O., Li de la Sierra-Gallay I., Piton J., Gilet L., Condon C. Activation of tRNA maturation by downstream uracil residues in *B. subtilis*. Structure. **2012**; 20(10): 1769-77.

117. Elena C., Ravasi P., Castelli M. E., Peiru S., Menzella H. G. Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. Front Microbiol. **2014**; 5: 21.

118. Derré I., Rapoport G., Msadek T. The CtsR regulator of stress response is active as a dimer and specifically degraded in vivo at 37°C. Mol Microbiol. **2000**; 38(2): 335-47.

119. Wingfield P. T. N-terminal methionine processing. Curr Protoc Protein Sci. **2017**; 88: 6 14 1-6 3.

120. Buckel P., Hübner-Parajsz C., Mattes R., Lenz H., Haug H., Beaucamp K. Cloning and nucleotide sequence of heavy- and light-chain cDNAs from a creatine-kinase-specific monoclonal antibody. Gene. **1987**; 51(1): 13-9.

121. Griffiths K., Dolezal O., Cao B., Nilsson S. K., See H. B., Pfleger K. D., et al. i-bodies, human single domain antibodies that antagonize chemokine receptor CXCR4. J Biol Chem. **2016**; 291(24): 12641-57.

122. Lakowitz A., Krull R., Biedendieck R. Recombinant production of the antibody fragment D1.3 scFv with different *Bacillus* strains. Microb Cell Fact. **2017**; 16(1): 14.

123. Abdel-Salam H. A., El-Khamissy T., Enan G. A., Hollenberg C. P. Expression of mouse anticreatine kinase (MAK33) monoclonal antibody in the yeast *Hansenula polymorpha*. Appl Microbiol Biotechnol. **2001**; 56(1-2): 157-64.

124. Augustine J. G., de La Calle A., Knarr G., Buchner J., Frederick C. A. The crystal structure of the fab fragment of the monoclonal antibody MAK33. Implications for folding and interaction with the chaperone bip. J Biol Chem. **2001**; 276(5): 3287-94.

125. Weber B., Hora M., Kazman P., Gobl C., Camilloni C., Reif B., et al. The antibody light-chain linker regulates domain orientation and amyloidogenicity. J Mol Biol. **2018**; 430(24): 4925-40.

126. Hora M., Sarkar R., Morris V., Xue K., Prade E., Harding E., et al. MAK33 antibody light chain amyloid fibrils are similar to oligomeric precursors. PLoS One. **2017**; 12(7): e0181799.

127. Herold E. M., John C., Weber B., Kremser S., Eras J., Berner C., et al. Determinants of the assembly and function of antibody variable domains. Sci Rep. **2017**; 7(1): 12276.

128. Nokwe C. N., Hora M., Zacharias M., Yagi H., John C., Reif B., et al. The antibody light-chain linker is important for domain stability and amyloid formation. J Mol Biol. **2015**; 427(22): 3572-86.

129. Brockmeier U., Caspers M., Freudl R., Jockwer A., Noll T., Eggert T. Systematic screening of all signal peptides from *Bacillus subtilis*: a powerful strategy in optimizing heterologous protein secretion in Gram-positive bacteria. J Mol Biol. **2006**; 362(3): 393-402.

130. Kakeshita H., Kageyama Y., Ara K., Ozaki K., Nakamura K. Propeptide of *Bacillus subtilis* amylase enhances extracellular production of human interferon-alpha in *Bacillus subtilis*. Appl Microbiol Biotechnol. **2011**; 89(5): 1509-17.

131. Liu R., Zuo Z., Xu Y., Song C., Jiang H., Qiao C., et al. Twin-arginine signal peptide of *Bacillus subtilis* YwbN can direct Tat-dependent secretion of methyl parathion hydrolase. J Agric Food Chem. **2014**; 62(13): 2913-8.

132. Skerra A. A general vector, pASK84, for cloning, bacterial production, and single-step purification of antibody F_{ab} fragments. Gene. **1994**; 141: 6.

133. Sally Ward E., Giissow D., Griffiths A. D., Jones P. T., Winter G. Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. Nature. **1989**; 341: 3.

134. Amit A., Mariuzza R., Phillips S., Poljak R. Three-dimensional structure of an antigen-antibody complex at 2.8 A resolution. Science. **1986**; 233(4765): 747-53.

135. Yellowley C. E., Toupadakis C. A., Vapniarsky N., Wong A. Circulating progenitor cells and the expression of Cxcl12, Cxcr4 and angiopoietin-like 4 during wound healing in the murine ear. PLoS One. **2019**; 14(9).

136. Tulotta C., Snaar-Jagalska B. E. CXCR4 signalling, metastasis and immunotherapy: zebrafish xenograft model as translational tool for anti-cancer discovery. J Cancer Met Treat. **2019**; 2019.

137. Griffiths K., Habiel D. M., Jaffar J., Binder U., Darby W. G., Hosking C. G., et al. Anti-fibrotic effects of CXCR4-targeting i-body AD-114 in preclinical models of pulmonary fibrosis. Sci Rep. **2018**; 8(1): 3212.

138. Arnolds K. L., Spencer J. V. CXCR4: a virus's best friend? Infect Genet Evol. **2014**; 25: 146-56.

139. Horvath P., Barragou R. CRISPR/Cas, the immune system of Bacteria and Archaea. Science. **2010**; 327(5962): 4.

140. Kick L., Kirchner M., Schneider S. CRISPR-Cas9: From a bacterial immune system to genome-edited human cells in clinical trials. Bioengineered. **2017**; 8(3): 280-6.

141. Pickar-Oliver A., Gersbach C. A. The next generation of CRISPR-Cas technologies and applications. Nat Rev Mol Cell Biol. **2019**; 20(8): 490-507.

142. Altenbuchner J. Editing of the *Bacillus subtilis* genome by the CRISPR-Cas9 system. Appl Environ Microbiol. **2016**; 82(17): 5421-7.

143. Villafane R. B., D.H.; Narayanan, C.S.; Dubnau, D.;. Replication control genes of plasmid pE194. J Bacteriol. **1987**; 168(10): 8.

144. Khasanov F. K., Zvingila D. J., Zainullin A. A., Prozorov A. A., Bashkirov V. I. Homologous recombination between plasmid and chromosomal DNA in *Bacillus subtils* requires approximately 70 bp of homology. Mol Gen Genet. **1992**; 234: 4.

145. Wu S. C., Yeung J. C., Duan Y., Ye R., Szarka S. J., Habibi H. R., et al. Functional production and characterization of a fibrin-specific single-chain antibody fragment from *Bacillus subtilis*: Effects of molecular chaperones and a wall-bound protease on antibody fragment production. Appl Environ Microbiol. **2002**; 68(7): 3261-9.

146. Luo X. Z., Zambaldo C., Liu T., Zhang Y. H., Xuan W. M., Wang C., et al. Recombinant thiopeptides containing noncanonical amino acids. Proc Natl Acad Sci. **2016**; 113(13): 3615-20.

147. Gan R., Perez J. G., Carlson E. D., Ntai I., Isaacs F. J., Kelleher N. L., et al. Translation system engineering in *Escherichia coli* enhances non-canonical amino acid incorporation into proteins. Biotechnol Bioeng. **2017**; 114(5): 1074-86.

148. Harwood C. R., Cutting S. M. Molecular Biological Methods for *Bacillus*. Chichester: John Wiley & Sons; **1990**.

Abbreviations

Abbreviation	Meaning
ADC	Antibody-drug conjugates
APS	Ammonium peroxosulfate
Asn	Asparagine
B. subtilis	*Bacillus subtilis*
bp	Base pair
DSB	Double strand break
EDTA	Ethylendiamintetraacetate
ELISA	Enzyme-linked immunosorbent assay
E. coli	*Escherichia coli*
g	Gravitational acceleration
GFP	Green fluorescent protein
His	Histidine
hmCK	Human creatine kinase
HR	Homologous recombination
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
kDa	Kilo dalton
M. mazei/bakeri	*Methanosarcina mazei/barkeri*
M. janashii	*Methanocaldococcus janashii*
ncAA	Non-canonical amino acid
NEB	New England Biolabs
NHEJ	Non-homologous end joining
Norb	Norbornene
NTA	Nitrilotriacetic acid
OD_{600}	Optical density at wavelength λ=600 nm
PAA	Polyacrylamide
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein database
POI	Protein of interest
RBS	Ribosome binding site

RT	Room temperature
SDS	Sodium dodecyl sulfate
SEC pathway	Secretory protein translocation pathway
SLIM	Site-directed, Ligase-Independent
SP	Signal peptide
TAT	Twin-arginine translocation pathway
v/v	Volume per volume
w/v	Weight per volume
WT	Wild type

Appendix

Protein sequences

Table 12: Amino acid sequence of expressed proteins
Residues mutated to amber stop codons are highlighted in bold.

Protein	Amino acid sequence
GFP	MSKGEELFTGVVPILVELDGDVNGH**K**FSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGV QCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGN ILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLSPDNHYLS TQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKHHHHHH
MAK33-VL	DIVLTQSPATLSVTPGDSVSLSCRASQSISN**N**LHWYQQK**L**HESPRLLIK**Y**ASQSISGIPSRFSGSGSGTD FTLSINSVETEDFGMYFCQQSNSWPLTFGAGTKLELKRHHHHHH
D1.3scFv	QVQLQESGPGLVAPSQSLSITCTVSGFSLTGYGVNWVRQPPGKGLEWLGMIWGDGNTDYNSALKS RLSISKDNSKSQVFLKMNSLHTDDTARYYCARERDYRLDYWGQGTTLTVSSGGGGSGGGGSGGGGS DIVLTQSPASLSASVGETVTITCRASGNIHNYLAWYQQKQGKSPQLLVYYTTTLADGVPSRFSGSGSGT QYSLKINSLQPEDFGSYYCQHFHSTPRTFGGGTKLEIKHHHHHH
Am3-114	LQVDIVPSQGEISVGESKFFLCQVAGSLSGIRISWFSPNGEKLTPNQQRISVVWNDDSSSTLTIYNANID DAGIYKCVVWRTGGYRHRYLVLGEATVNVKIFQAWSHPQFEK
PylS	MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARALRHHKYRKT CKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLENTEAAQAQPSGSKFS PAIPVSTQESVSVPASVSTSISSISTGATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPK DEISLNSGKPFRELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGID NDTELSKQIFRVDKNFCLRPMLAPNLGNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQ MGSGCTRENLESIITDFLNHLGIDFKIVGDSCMVFGDTLDVMHGDLELSSAVVGPRPLDREWGIDKP WIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL
PrsA	MKKIAIAAAITATSILALSACSSGDKEVIAKTDAGDVTKGELYTNMKKTAGASVLTQLVQEKVLDKKYKVS DKEIDNKLKEYKTQLGDQYTALEKQYGKDYLKEQVKYELLTQKAAKDNIKVTDADIKEYWEGLKGKIRA SHILVADKKTAEEVEKKLKKGEKFEDLAKEYSTDSSASKGGDLGWFAKEGQMDETFSKAAFKLKTGEVS DPVKTQYGYHIIKKTEERGKYDDMKKELKSEVLEQKLNDNAAVQEAVQKVMKKADIEVKDKDLKDTFN TSSTSNSTSSSSSNSK
AmyE-SP	MFAKRFKTSLLPLFAGFLLLFHLVLAGPAAAAA
LipA-SP	MKKVLMAFIICLSLILSVLAAPPSGAAAA

Expression constructs

The DNA and protein sequence for the expression constructs of the two main proteins in this work are shown below.

pLIKE-P_{groES}-5'-GFP-His_6-3'

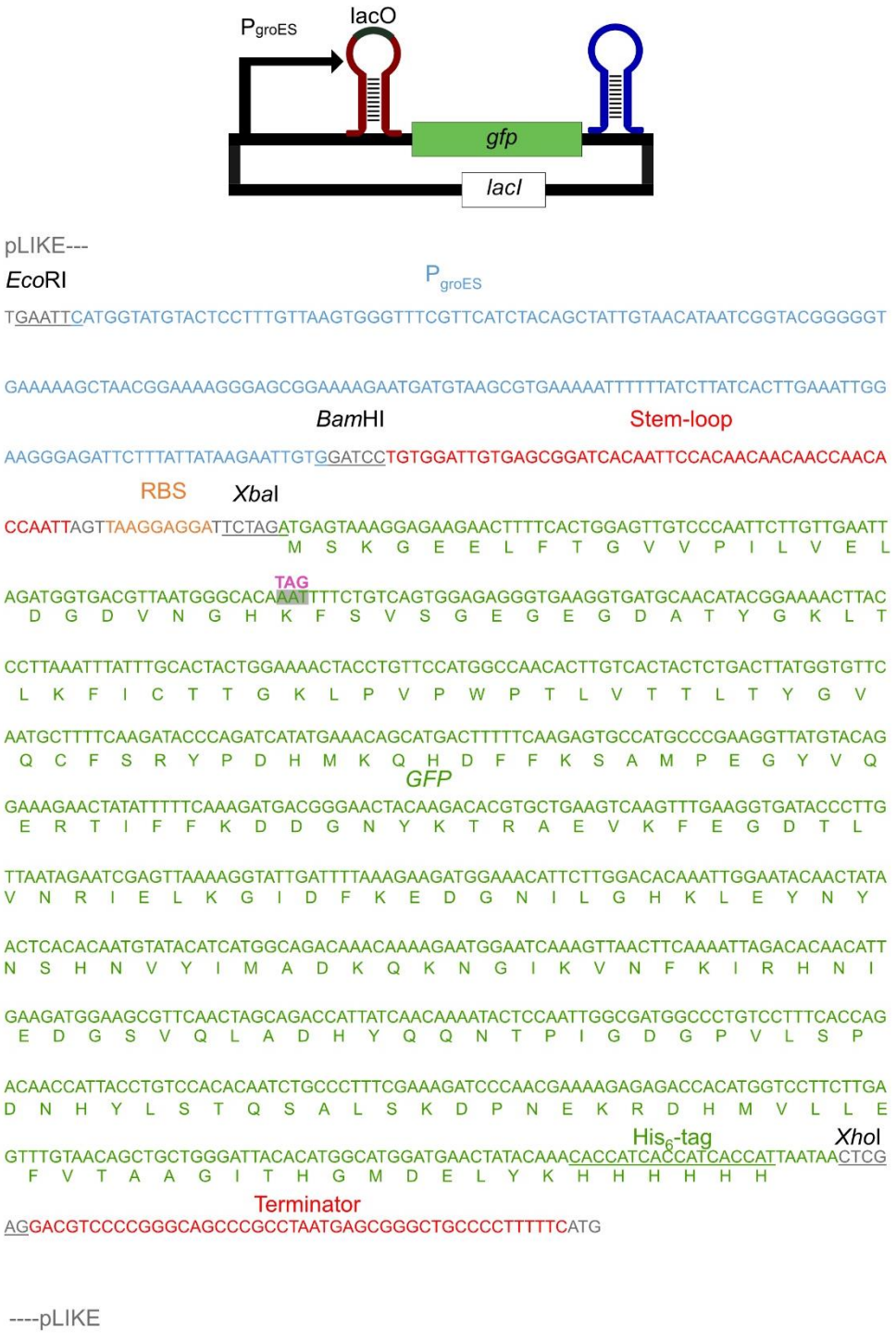

Figure 62: DNA sequence of the gfp-expression plasmid #9 pLIKE-P_{groES}-5'-GFP-His_6-3'
The position of the amber codon corresponding to Lys26, highlighted in pink. The promoter sequence is shown in blue, the structured RNA sequence elements in red and restriction sites used for cloning in grey. RBS=ribosome binding site.

pLIKE-P_{groES}-5'-AmyE-MAK33-VL-His_6-3'

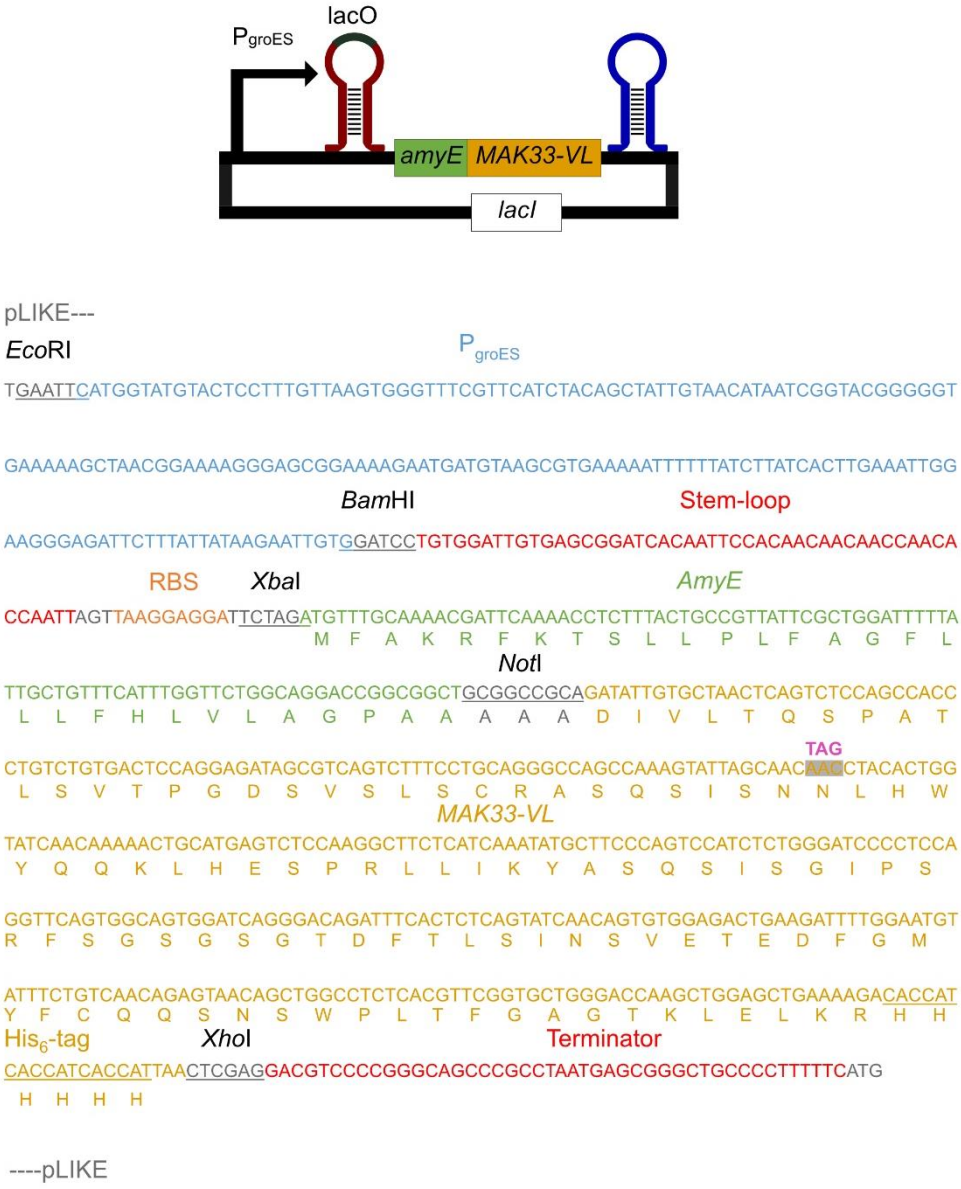

Figure 63: DNA sequence of the MAK33-VL expression plasmid #21 pLIKE-P_{groES}-5'-MAK33-VL-His_6-3'
The position of the amber codon corresponding to Asn33, highlighted in pink. The promoter sequence is shown in blue, the structured RNA sequence elements in red and restriction sites used for cloning in grey. RBS=ribosome binding site.

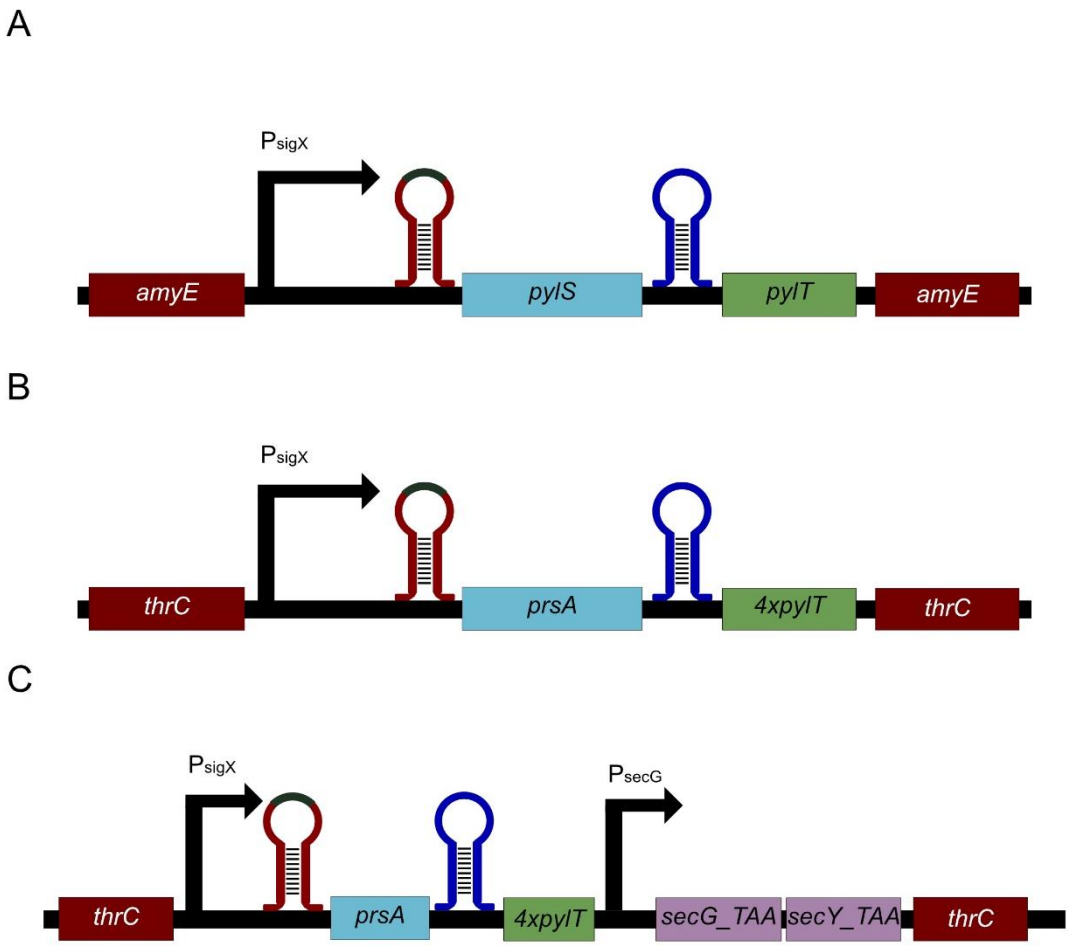

Figure 64: Representation of the different integrating plasmids used in this work
A) Amber suppression plasmid, the part of plasmid #14 integrated into the *amyE*-locus. B) pXT plasmid #16 integrated in the *thrC* locus. C) Secretion plasmid #17 integrated into the *thrC* locus carrying the components to rescue the SEC system.

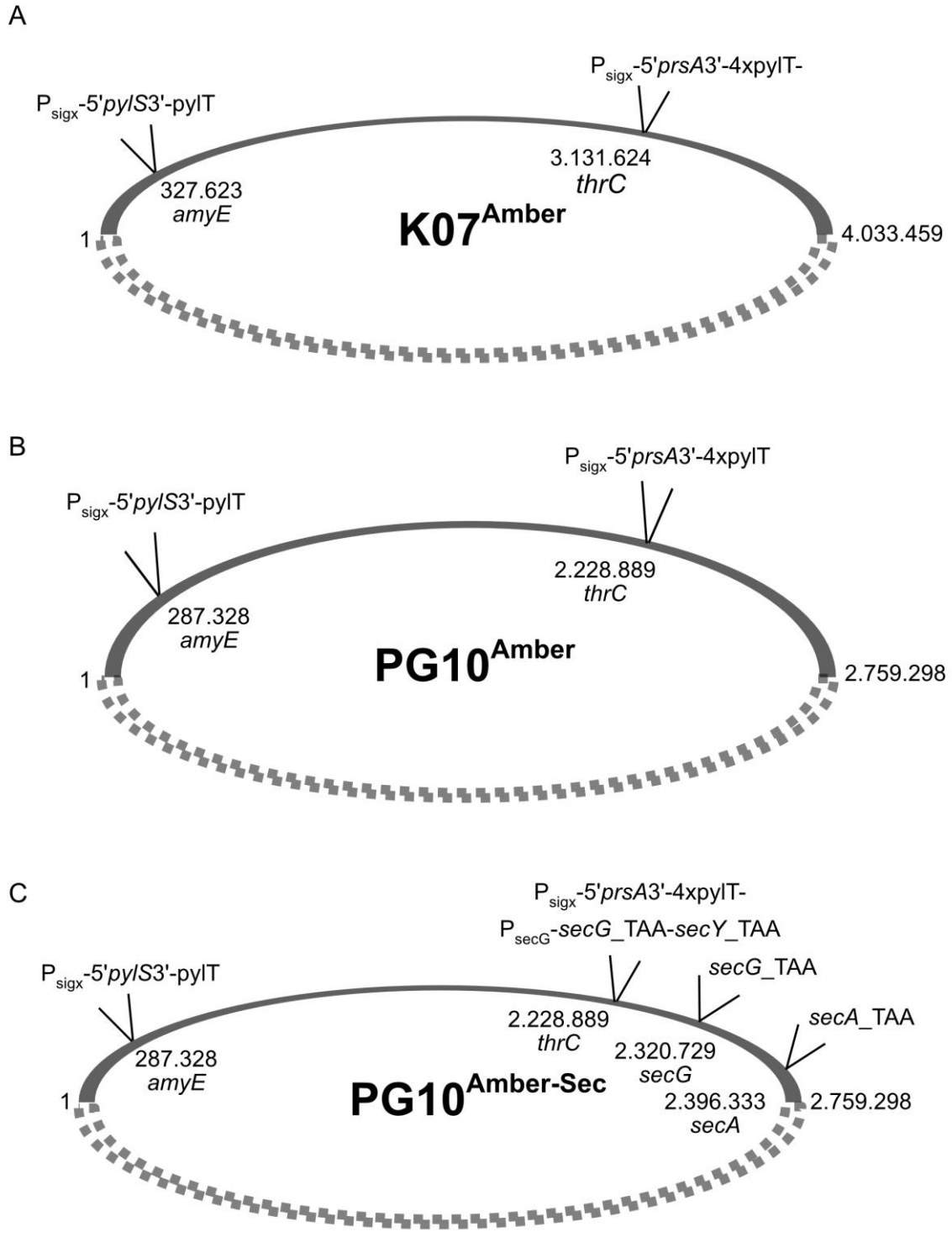

Figure 65: Genomic context of the strains A) KO7^{Amber}, B) PG10^{Amber} and C) PG10$^{Amber-Sec}$
The locations of the mutated sites or sites with an genomic integration are marked.

List of tables

Table 1: Used devices 75
Table 2: Bacterial growth media 76
Table 3: Used enzymes 78
Table 4: DNA and protein markers 79
Table 5: Used kit systems 79
Table 6: Used DNA oligonucleotides 79
Table 7: Used and created plasmids 83
Table 8: *E. coli* K12 strains 84
Table 9: Used and created *B. subtilis* strains 84
Table 10: Buffers for protein purification, DNA manipulation or western blotting 87
Table 11: Used software and online databases 89
Table 12: Amino acid sequence of expressed proteins 111

List of figures

Figure 1: Historic drawing of *B. subtilis*, partially in the dividing state by Cohn from 1872(2) 7
Figure 2: Model describing the history and relationships between the early *B. subtilis* strains 8
Figure 3: Overview of several bacterial secretion systems 9
Figure 4: Scheme of the SEC and Tat system of *B. subtilis* 11
Figure 5: Structure of signal peptides targeting the SEC and the Tat system 11
Figure 6: Intermediates from the creation of PG10 and PS38 from the *B. subtilis* wild type 168 13
Figure 7: Genetic code expansion by amber codon suppression 16
Figure 8: Variety of different ncAAs and their biotechnological applications 18
Figure 9: State of the art bioorthogonal reactions 19
Figure 10: Structural examples of an antibody and antibody derived single-chain fragments 20
Figure 11: Overview of different optimization targets on DNA and mRNA level 22
Figure 12: Structural overview of integrating and replicative plasmids in *B. subtilis* 23
Figure 13: Difference between an integrated and replicative expression construct 24
Figure 14: Influence of mRNA stabilizing elements on the protein expression 25
Figure 15: Expression profile of five different *B. subtilis* promoters 27
Figure 16: Influence of the BamHI site directly downstream of the transcription start site on the overall fluorescence 28
Figure 17: Growth curve and expression profile of *B. subtilis* K07-6 expressing gfp under control of P_{groES} in different media 29
Figure 18: Growth curve and expression profile of *B. subtilis* K07 expressing gfp under control of P_{sigx} in different media 30
Figure 19: Effect of the addition lacI on the GFP expression 32
Figure 20: One-plasmid expression setup for amber suppression based on the pLIKE plasmid 33
Figure 21: Instability of the one-plasmid system in *B. subtilis* 34
Figure 22: Two plasmid expression system 35
Figure 23: Western Blot against the N-terminal FLAG tag of $PylS^{Norb}$ 36
Figure 24: tRNA Maturation of most tRNAs in *B. subtilis* 37
Figure 25: Structure of the 1x *pylT* and 4x *pylT* constructs 37
Figure 26: Incorporation rates of 1x *pylT* and 5x *pylT* compared to the wild type GFP 38
Figure 27: Purification of GFP^{Norb26} from *B. subtilis* strain $K07^{Amber}$-1 39
Figure 28: Mass spectra of GFP before and after the reaction with a click reagent 40
Figure 29: Optimized modular expression plasmid based on the double stabilized pLIKE plasmid #2 42
Figure 30: Protein crystal structure of the MAK33-Fab fragment (PDB 1FH5) 43

Figure 31: pLIKE-PgroES-5'AmyE-MAK33-VL-3' expression plasmid with a C-terminal His_6- or StrepII-tag..... 43
Figure 32: SDS-PAGE analysis of the total supernatant of *B. subtilis* strain K07-14 expressing MAK33-VL-His_6. 44
Figure 33: SDS-PAGE and Western Blot analysis of K07-13 expressing MAK33-VL-Strep..... 45
Figure 34: SDS-PAGE analysis and mass spectrum of MAK33-VL..... 46
Figure 35: Scheme of a SEC system signal peptide with possible cleavage sites 46
Figure 36: SDS-PAGE and Western Blot of the initial purification of D1.3scFv 47
Figure 37: Supernatant from a screening for suitable expression conditions for D1.3scFv..... 48
Figure 38: Cell pellet fraction from screening for suitable expression conditions for D1.3scFv..... 49
Figure 39: SDS-PAGE and mass spectrum of purified D1.3scFv..... 50
Figure 40: Structure of the NCAM I domain (PDB 5AEA), the scaffold for the development of Am3-114...... 51
Figure 41: SDS-PAGE and Western Blot on three clones expression Am3-114 52
Figure 42: SDS-PAGE and Western Blot analysis of purified Am3-114 with corresponding mass spectrum 52
Figure 43: Scheme of the GFP-fusion construct with the possible incorporation sites 55
Figure 44: Location and successful identification of Asn33 as a target site 56
Figure 45: Successful but weak purification of MAK33-VL^{Norb33} 57
Figure 46: The addition of the ncAA decreases the total secreted proteins 58
Figure 47: SEC system of *B. subtilis* with marked amber stop codons in the corresponding genes.. 59
Figure 48: Mechanism of Cas9-mediated DNA cleavage...... 60
Figure 49: Scheme of the mutations introduced via a QuikChange PCR. 61
Figure 50: Single colony PCR on the strain PG10 after transformation with plasmid pJOE+secA_TAA_sgRNA_secA (#29). 62
Figure 51: Single colony PCR of the strain $PG10^{Amber\text{-}secA}$ after transformation with plasmid pJOE-secG_TAA (#30)...... 63
Figure 52: SDS-PAGE analysis of $PG10^{Amber}$ and $PG10^{Amber\text{-}Sec}$ 64
Figure 53: Growth curves of A) PG10, B) $PG10^{Amber}$ and C) $PG10^{Amber\text{-}Sec}$. 65
Figure 54: SDS-PAGE analysis and western blotting against the StrepII-tag. 67
Figure 55: SDS-PAGE analysis of the elution fractions of MAK33-VL and MAK33-VL^{Norb33} purified via their His_6-tag. 67
Figure 56: ESI-mass spectra of the purified StrepII-tagged MAK33-VL^{Norb33} 68
Figure 57: ESI-mass spectrum of the His_6-tagged MAK33-VL^{Norb33} 69
Figure 58: Incorporation site evaluation of D1.3scFv..... 70

Figure 59: Scheme of MAK33-VL$^{Norb33\text{-}Fluoro}$ conjugate (PDB 1FH5) 72
Figure 60: Experimental setup and verification oft he MAK33-VL functionality 73
Figure 61: Overview of the pLIKE expression plasmid 94
Figure 62: DNA sequence of the gfp-expression plasmid #9 pLIKE-P_{groES}-5'-GFP-His_6-3' 112
Figure 63: DNA sequence of the MAK33-VL expression plasmid #21 pLIKE-P_{groES}-5'-MAK33-VL-His_6-3' 113
Figure 64: Representation of the different integrating plasmids used in this work 114
Figure 65: Genomic context of the strains A) KO7^{Amber}, B) PG10^{Amber} and C) PG10$^{Amber\text{-}Sec}$ 115

List of publications

This thesis was conducted at the Chair of Biochemistry at the Technische Universität München from February 2016 to June 2019 and at the Ludwig-Maximilians Universität München from July 2019 to March 2020.

Parts of this thesis have been published:

Scheidler C.M., Vrabel M. and Schneider S. Genetic code expansion, protein expression and protein functionalisation in *Bacillus subtilis*, ACS Synth Biol, **2020**, doi: 10.1021/acssynbio.9b00458

Publications not present in this thesis

Scheidler C.M., Kick L.M, Schneider S.: 'Ribosomal Peptides and Small Proteins on the Rise', Chembiochem, **2019**, 20, 1479-1486, doi: 10.1002/cbic.201800715

Acknowledgments

After working for four years on this PhD thesis, it is time to thank several people for their support.

Dr. Sabine Schneider for the great opportunity to conduct this thesis in her lab, for patiently answering my crystallography questions, and supporting me during the work, especially when the bacteria did not behave as they should have.

Prof. Michael Groll and his whole group for providing their equipment during my time at the TUM.

The past and present members of the AG Schneider, Sabrina, Christine, Marion, Leo, Philipp and Felix for the great atmosphere and the continuous fun. Especially for enduring my sometimes questionable music taste. Moreover, a special “cheers” to Leo, for Whisky, Tennis, Bad Jokes and a friend who I found while working – without you the time in the lab would have been dull.

My students, Jannis, Van, Annika, Jan, Chrissi, Alejandro, Raphaela und Duygu for their work, in particular the UNITY crew!

The TUM Graduate School for their financial support for my research trips to Madrid and Nottingham.

The GRK2062 for financing my PhD thesis and the opportunity to attend several conferences and all the members of the GRK for the great discussions during our journal clubs.

Prof. Kathrin Lang and Prof. Fritz Simmel for their support and fruitful discussions in the TAC meetings.

The whole AK Lang, Suso, Max, Tuan-Anh, Marie-Lena and Marie for the great fun we had together and for the countless mass spec runs.

My friends at Moe! who endured this strange stories about “proteins” and “DNA”. §11

My family for their endless love and support and countless times when you cheered my up if the work did not go as planned.

Declaration

I, Christopher Markus Scheidler, hereby declare that I independently prepared the present thesis, using only the references and resources stated. This work has not been submitted to any examination board yet. Parts of this work have be published in scientific journals.