Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

Sine Alise Hartvigsen Hansen

Engineering the genome of *Escherichia coli* BL21 for improved secretory production of recombinant proteins

Master's thesis in Chemical Engineering and Biotechnology
Supervisor: Trygve Brautaset and Agnieszka Gawin
July 2019

NTNU
Norwegian University of
Science and Technology

Sine Alise Hartvigsen Hansen

Engineering the genome of *Escherichia coli* BL21 for improved secretory production of recombinant proteins

Master's thesis in Chemical Engineering and Biotechnology
Supervisor: Trygve Brautaset and Agnieszka Gawin
July 2019

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

NTNU
Norwegian University of
Science and Technology

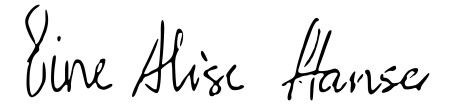
Preface

This thesis is submitted as part of the Master's Thesis course, TBT4900, for students in the study program "Chemical Engineering and Biotechnology" at the Department of Biotechnology and Food Science, Norwegian University of Science and Technology, July 2019. The project accounts for 30 credits and was conducted in collaboration with Vectron Biosolutions. The work is a contribution to a long-term project aiming at optimizing the rate of periplasmic secretion of recombinant proteins in *E. coli* by a CRISPR/Cas9-based mutagenesis approach.

I would like to thank Agnieszka Gawin once again for excellent supervision, and for following me on every step along the way. I am thankful for all the work you have done to facilitate and help me through this project, and for providing the background knowledge necessary. I would also like to thank my supervisor, professor Trygve Brautaset and all of the competent people in his research group who are always helpful and ready to answer questions.

All of the people at Vectron Biosolutions has provided good input along the way. Thank you all for welcoming me to be part of the student group, and for all the good advice.

Trondheim July 2, 2019



Sine Alise H. Hansen

Abstract

The pharmaceutical industry is currently seeing an increasing focus on recombinant pharmaceuticals. Most of the recombinant pharmaceutical proteins have eukaryotic origin that require posttranslational modifications such as disulfide bond formation. To facilitate better folding of such proteins in microbial host cells, they are often translocated to the periplasm. As the periplasmic production yields still remain low in comparison with the cytoplasmic production, the process requires optimization. Current approaches for optimization of recombinant protein production and secretion are to a great extent based on the co-expression of chaperons and factors enhancing folding and stability of a target protein, or knock-out mutations of proteases.

In this study, a novel approach for regulating the expression of native genes in the genome is used as an approach to facilitate better folding and stability of recombinant proteins in the periplasm. Seven strains of *Escherichia coli* were constructed by introducing mutations of three genes encoding periplasmic proteins, and combinations of these mutations. Up-regulation of the signal peptide peptidase SppA and the periplasmic chaperone Skp, and down-regulation of the periplasmic protease DegP was obtained by changing the respective RBS sequences in the genome of *E. coli* BL21 by CRMAGE recombineering. A reporter plasmid expressing alkaline phosphatase fused to the industrially relevant single-chain antibody scFv173 was used to compare the created mutants strains. The periplasmic recombinant protein production was measured by colorimetric assay after release of protein by osmotic shock.

The *degP* down-regulation resulted in a significant decrease in production of alkaline phosphatase compared to the wildtype BL21 strain. The mutation of *sppA* gave significantly higher expression rates compared to the other constructed strains, but not high enough to state a significance compared to the wildtype BL21 strain. No other significant differences were observed between the units of phosphatase activity among studied strains and the wildtype BL21 strain. The potential of the up-regulation of *sppA* should be further investigated.

Sammendrag

Et økt fokus på rekombinante legemidler er sett i den farmasøytiske industrien. De fleste rekombinante legemidler stammer fra eukaryoter, og krever posttranslasjonsmodifikasjoner slik som dannelse av disulfid-bindinger. For å fasilitere bedre folding av slike proteiner i mikrobielle vert-celler, blir de ofte translokert til periplasma. Siden utbyttet av rekombinante proteiner produsert i periplasma er lave sammenlignet med utbyttet av cytoplasmisk ekspresjon, har denne prosessen et potensiale for optimalisering. Nåværende tilnærminger for optimalisering av rekombinant proteinproduksjon og sekresjon er i stor grad basert på ko-ekspresjon av chaperoner og faktorer som forbedrer folding og stabilitet av mål-proteinet, eller null-mutasjoner av proteaser.

I denne studien har en ny metode for å regulere ekspresjon av villtype-gener i genomet vært benyttet med mål å fasilitere bedre folding og stabilitet av rekombinante proteiner i periplasma. Sju stammer av *Escherichia coli* ble konstruert ved å introdusere mutasjoner i tre gener som koder for periplasmiske proteiner og kombinasjoner av disse mutasjonene. Økt translasjonsrate av signal-peptid peptidasen SppA og det periplasmiske chaperonet skp, og lavere translasjonsrate av den periplasmiske proteasen DegP ble oppnådd ved å endre de respekterende RBS-sekvensene i genomet til *E. coli* BL21 ved CRMAGE rekombinering. De mutante stammene ble sammenlignet ved bruk av et reporter-plasmid som uttrykker alkalisk fosfatase fusjonert med antistoffet scFv173 som er av industriell interresse. Rekombinant proteinproduksjon ble målt ved bruk av kolorimetri etter frigjøring av proteiner i periplasma ved osmotisk sjokk.

Ned-regulering av *degP* førte til signifikant avtagelse i produksjon av alkalisk fosfatase sammenlignet med BL21-stammen. Mutasjon av *sppA* førte til signifikant høyere ekspresjonsrate sammenlignet med de andre konstruerte stammene, men ikke nok til å konstatere en signifikans i forskjellen fra villtype BL21. Potensialet av opp-regulering av *sppA* bør undersøkes videre.

Abbreviations

A summary of the abbreviations used in this thesis is presented below.

Amp - Ampicillin
AP - Alkaline Phosphatase
aTet - anhydrous Tetracycline
ATP - adenosine triphosphate
bp - base pairs
Cas - CRISPR-associated
Cas9 - nickase with paired guide RNA
CRISPR - Clustered Regularly Interspaced Short Palindromic repeats
CRMAGE - CRISPR-optimized MAGE recombineering
crRNA - CRISPR RNA dH_2O - distilled water
DNA - Deoxy-ribonucleic acid
Kan - Kanamycin
MAGE - Multiplex Automated Genome Engineering
nt - nucleotides
OD - Optical Density
PAM - Protospacer Adjascent Motif
PCR - polymerase chain reaction
RBS - Ribosome Binding Site
RNA - Ribonucleic acid
r.t. - room temperature
rpm - rounds per minute
SDOM - standard deviation of the mean
Spc - Spectinomycin
ssDNA - single stranded DNA
TIR - Translation Initiation Rate
U - units

Contents

CHAPTER 1

Introduction

1.1 Recombinant protein production

Recombinant protein production is a field in constant development, and has been a field of research since the first recombinant plasmid was constructed and expressed in *E. coli* in the 1973 [1]. Human insulin was the first commercial recombinant pharmaceutical produced, also in *E. coli* [2], and revealed a range of opportunities for recombinant protein production. The field of recombinant pharmaceutics has continued growing, and is still a major field in recombinant protein production, involving pharmaceuticals such as hormones, vaccines and human antibodies [3, 4]. Complex proteins of therapeutic interest constitute a challenge because their high specificity might be easily affected by the post-translational changes [5].

The pharmaceutical industry is currently seeing an increasing focus on recombinant pharmaceuticals. The market for recombinant protein production is growing with a forecasted compound annual growth rate of 6.5% between 2019 and 2024 according to Mordor Intelligence [6]. The highest market share is seen for the pharmaceutical industry, where almost half of the market revenue share is for therapeutic use and drug discovery. Other dominant players are basic research and the biotech industry [6]. Pharmaceutical products require high quality, and there are strict requirements for purity and safety of drugs. This adds extra challenges for recombinant protein expression of pharmaceutical products [5]. The process of recombinant protein production is therefore a subject of ongoing optimization [3].

Bacteria are popular hosts for recombinant protein production due to rapid growth at high cell densities [7], and the ability to grow on inexpensive media. There is also a broad understanding of bacterial gene expression, and the bacterial fermentation technology is well-established [5]. However, the recombinant production of complex molecules such as antibodies require post-translational modifications, and mammalian hosts such as the chinese hamster ovary (CHO) cells has gained popularity for production of biopharmaceuticals [4, 8]. These cells however, require com-

plex media including vitamins, hormones, growth factors and antibodies. Moreover, the cells are more fragile, requiring specially designed bioreactors, and grow slowly compared to bacteria [5]. The main challenges in the biopharmaceutical industry is, in fact, the optimization of productivity, cost, and potency [8]. The rapid growth, low costs [5], readily available methods and recombinant strains [9] therefore makes *E. coli* a host of interest for recombinant protein expression.

The enterobacterium *E. coli* is a gram-negative bacterium found naturally in the colon of mammals [10, 11]. During the first years when recombinant expression technology was emerging, *E. coli* was the main platform for recombinant production of biopharmaceuticals [4]. The popularity of *E. coli* as an expression host has led to the development of a high number of tools and methods for high-level expression. A variety of attempts to optimize the bacterium as an expression host has generated numerous engineered recombinant strains of *E. coli* [9]. *E. coli* is therefore the first-choice microorganism for the production of recombinant proteins [12]. Due to challenges of heterologous protein expression such as product toxicity, incorrect folding of target protein and inclusion body formation, not all proteins may be expressed in *E. coli* as of today [13], and there is therefore a need for optimization of this host. One approach is directing the recombinant protein to the periplasmic space, harbouring enzymes facilitating correct folding [11].

1.2 Secretory production of recombinant proteins in *E. coli*

High-level expression of recombinant proteins may cause aggregation, incorrect protein folding, product toxicity and protein degradation, that in turn can negatively affect the protein yield [14]. Periplasmic translocation of heterologous proteins offers a solution to some of these challenges, including incorrect recombinant protein folding, product toxicity and cytoplasmic degradation [11, 14]. Here, secretory production refers to the translocation through the cytoplasmic membrane into the periplasmic space of gram-negative bacteria.

1.2.1 Advantages and challenges of secretory production of recombinant proteins in *E. coli*

One drawback of using *E. coli* as a production host is its inability to easily secrete proteins to the periplam, and significant reserch interest is therefore seen in the optimization of recombinant protein secretion [14]. Periplasmic secretion may be beneficial due to the possibility of extracting solely the periplasmic fraction, lowering amount of native protein contamination in the protein extract [11, 14, 15].

One of the challenges of cytoplasmic recombinant protein production in *E. coli* is

the limited capacity of creating disulfide bonds due to reducing conditions of the cytoplasm [7, 16]. This may be a challenge when producing pharmaceutical proteins, or working with complex proteins that require correct disulfide bonding for proper function. In the periplasm of *E. coli*, however, a set of proteins involved in the formation of disulfide bonds, the Dsb system (from **d**i**s**ulfide **b**onds) facilitates and/or catalyzes the formation of disulfide bonds [16] as presented in figure 1.1.

Incorrect folding due to the lack of disulfide bond formation in the cytoplasm may in turn lead to aggregation and inclusion body formation. This may also be a challenge due to the requirement of more complicated downstream processing such as isolation, re-naturation and re-folding [15]. In contrast to the cytoplasm, the periplasm has a lower protein content. This may lead to an increased protein solubility and stability [14]. Better folding properties of the periplasmic space may therefore help overcome the challenge of cytoplasmic inclusion body formation.

There are, however some challenges also with periplasmic protein production. Foreign proteins are susceptible to protease degradation [17]. The production of periplasmically secreted recombinant proteins is also limited by the physical volume of the periplasm. This normally results in reduced cell growth and cellular burden [15]. A common approach for enhanced protein secretion is co-expression of proteins facilitating periplasmic protein production from additional helper plasmids [15, 16, 18–20]. The co-expression of different proteins and selection markers may cause metabolic burden to the host because the production utilizes a significant amount of the host cell's resources, removing those resources away from host cell metabolism [21].

The limited capacity of the translocation machinery is also a challenge causing limited yields of secreted recombinant proteins [11, 15]. There are two main pathways for periplasmic secretion of recombinant proteins; the Sec-dependent pathways and the twin-arginine translocation (TAT) pathway. The mechanisms of Sec-dependent periplasmic translocation are presented in figure 1.1 from Schegel et al. [22]. The Sec pathways depend on the bacterial translocon that consists of a heterotrimeric protein complex, SecYEG [23]. Targeting a protein to the Sec-dependent pathway requires a signal peptide at the N-terminal of the recombinant protein [24]. The signal sequence is characterized by a positively charged N-terminal segment, a hydrophobic trans-membrane segment, and a cleavage site where the signal sequence is removed from the mature protein after secretion to the periplasm [24, 25]. Depending on the hydrophobicity of the signal sequence as well as the binding of a trigger factor, the protein is translocated either post-translationally, SecB-dependently or co-translationally, SRP-dependently [14].

The TAT secretion pathway in *E. coli* is Sec-independent and allows folded proteins to be secreted to the periplasm independent of ATP [11]. As for the Sec dependent translocation, a signal sequence is needed to direct the protein to the TAT-pathway. This sequence has a conserved twin-arginine motif, and is longer than the signal sequence for Sec-dependent translocation [26].

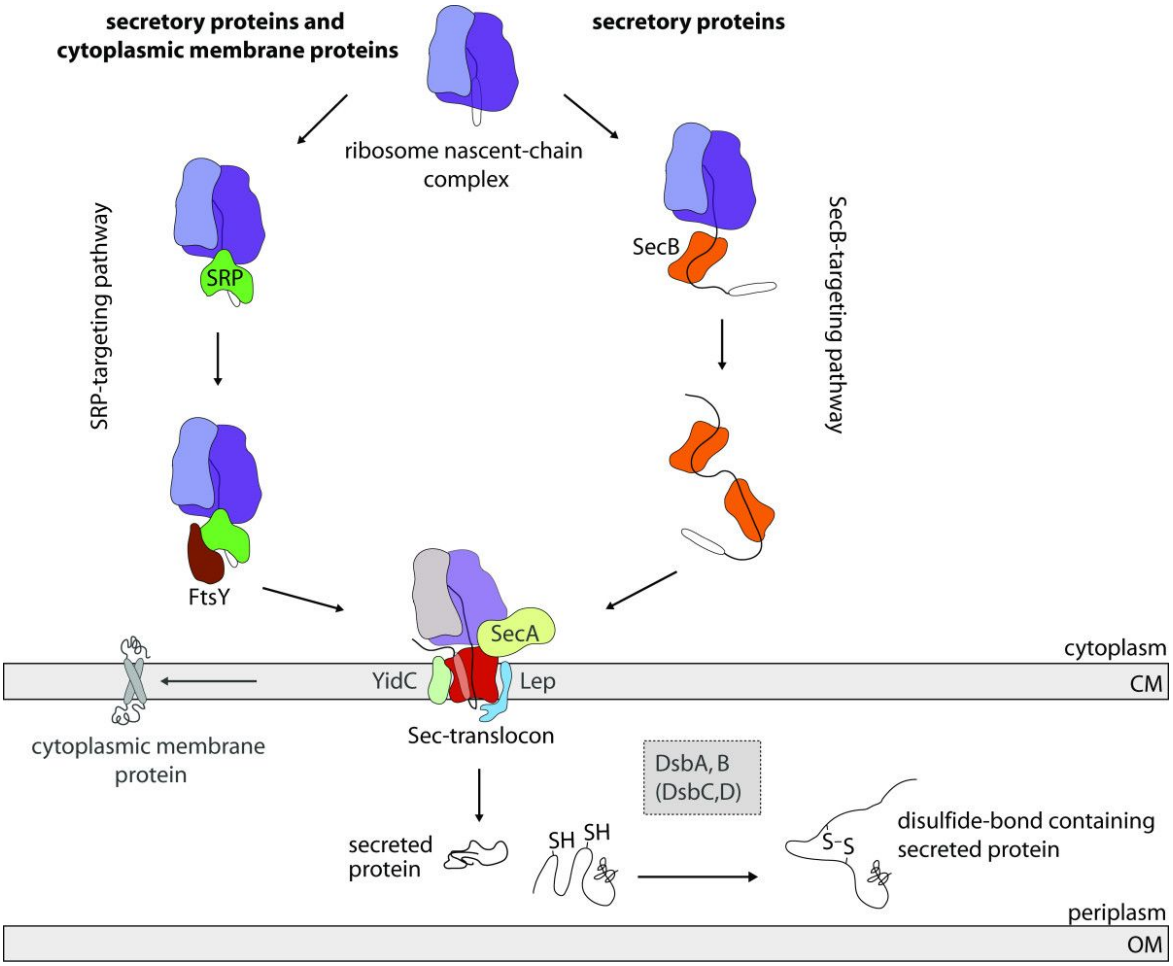


Figure 1.1: Periplasmic secretion through the Sec-dependent pathways. Figure from Schegel et al. [22]. In the SecB-dependent pathway, the ribosome nascent-chain complex is targeted by the chaperone SecB chaperones in the cytoplasm that keep the polypeptide in an unfolded state before it is transferred to the translocon and translocated to the periplasm. In the Signal Rcognition Particle (SRP)-targeting pathway, the ribosome nascent chain-complex is targeted by SRP and its receptor FtsY that targets the complex to the Sec-translocon where the protein is translocated co-translationally. This pathway is also used for the integration of cytoplasmic membrane proteins [22]. In the periplasm, disulfide bond formation is facilitated by the Dsb system [16]. SecA: ATPase associated with the Sec-translocon [23]; OM: outer membrane; YidC: cytoplasmic membrane protein translocase/insertase [23].

When the signal sequence (either for the Sec- or Tat-pathway) of the recombinant protein is cleaved off, it also removes the N-terminal methionine, potentially avoiding negative impacts on stability and activity of the product [14]. One of the main challenges of protein secretion is the limitation in the capacity of the secretion machinery limiting the yields of secreted proteins [11, 15]. The secretion machinery is therefore one of the main targets for optimizing periplasmic production of recom-

binant proteins.

1.3 Approaches for optimizing production of heterologous proteins in the periplasm

Periplasmic secretion offers a potential solution to common challenges of recombinant protein production, but it still requires optimization. Several things must be taken into account when attempting to produce a recombinant protein in *E. coli*, depending on the requirements of the final product. The approach for optimizing the production of a specific product may depend on the quality requirement, challenges with folding, protein toxicity, inclusion body formation etc. For some products, the final structure and quality are critical, while in other cases inclusion body formation is a greater challenge. This section gives an overview of current achievements in optimizing production of heterologous proteins in the periplasm. More details on the subject can be found in the reviews by Kleiner et al. [14] and Yoon et al. [15].

One of the first things that must be decided is the choice of a **signal sequence** directing the protein of interest to the desired translocation pathway. As mentioned earlier, the signal sequence has a positively charged N-region, a hydrophobic H-region and a conserved region (C-region) that mediate the removal of the signal sequence from the mature protein [24]. For signal sequences targeting the Sec-dependent pathway, the charge of the N-terminal may enhance the translocation by increasing the interaction with SecA [14]. The length and hydrophobicity of the H-region also affects the translocation efficiency [27]. Choice of the appropriate signal sequence therefore constitutes an important step of optimizing translocation of a recombinant protein.

Co-expression of proteins from a helper plasmid together with the target protein is a common approach for assisting translocation and folding of secreted recombinant proteins [15, 16, 18–20]. To overcome the challenge of inclusion body formation and incorrect folding, co-expression of molecular chaperones [20] and enzymes involved in disulfide bond formation [19] has been successfully used. Examples are the co-expression of periplasmic chaperone Caf1M and usher protein Caf1A from *Yersinia pestis* increasing the folding and secretion of human Interleukin-1β in *E. coli* [20]. Disulfide bond formation in *E. coli* may be improved by co-expression of proteins involved in disulfide bond formation (DsbA, DsbB, DsbC and DsbD), and helper plasmids has been constructed for this purpose [19].

As mentioned, the limited capacity of the secretion machinery may lead to lower yields of the target protein [11, 15]. In cases where the secretion is the rate-limiting step, an approach is to co-express proteins involved in protein secretion such as the co-expression of SecY and/or SecE of the bacterial translocon [15].

Using plasmid vectors for co-expression of proteins assisting in the secretion of a tar-

get protein is, however, associated with a need of using additional selection markers and inducing agents. The addition of antibiotics to the medium induce the expression of antibiotic resistance genes, taking up some of the protein producing capacity of the cell [21].

An alternative approach is the fusion with a highly soluble protein such as a maltose-binding protein (MBP) for increased solubility. The fused MBP may be cleaved off by a specific protease after production [28]. Another approach for making *E. coli* more efficient for secretory production is making **strains deficient in proteases** or genes regulating proteases [15]. The combination of the mutations in the three genes *degP* and *prc* encoding proteases, and *spr*, a suppressor of growth defects in Prc-deficient strains allowed the high-level accumulation of an antibody fragment in the periplasm of *E. coli* [17]. Also the commonly used strain of *E. coli* BL21 is deficient in *lon* and *ompT* proteases [29].

1.3.1 Potential genomic targets for improved secretory production of recombinant proteins

As illustrated in the previous section, most methods for improved secretory production are based on the co-expression of chaperons or other proteins assisting in secretion from plasmids or the complete knock-out of proteases. Only few attempts have been made to change the expression of native genes by genome editing. Changing the expression levels of native proteins of *E. coli* BL21 therefore constitutes an under-represented approach to improving protein secretion, and gives the opportunity to fine-tune the expression levels of proteins [30]. By using this approach, the need for antibiotic selection markers that may affect growth is not needed as in the case of plasmids. Below, a selection of the potential native genomic targets for enhanced recombinant protein production is presented.

SppA is a signal peptide peptidase, and assists in protein secretion by mediating the cleavage of the signal peptide from secreted proteins. It is encoded by *sppA*, a gene that is the only gene of the transcription unit as seen in figure 1.2. The over-expression of SppA in *Bacillus licheniformis* has been found to increase the extracellular protein yields of α-amylase and nattokinase with 67% and 30% respectively. This was performed by Cai et al. by integration of an additional copy of the gene into the genome [31]. The SppA derived from *B. licheniformis* ($SppA_{BL}$) however, hold different properties compared to the SppA derived from *E. coli* ($SppA_{EC}$). The $SppA_{EC}$ holds a narrower substrate preference compared to the $SppA_{BL}$ [32]. The research done for this project indicates that over-expression of *sppA* has never been performed in *E. coli*.

Figure 1.2: The transcription unit of *sppA* consists only of the *sppA* gene. Hairpin-structure symbolizes a rho-independent terminator. The figure is obtained from the EcoCyc database. sppAp6: promoter (no high-quality experimental evaluation); σ^{70}: σ factor 70.

DegP is a periplasmic protease required for survival at high temperatures. It functions by efficiently removing misfolded outer membrane proteins which might prevent normal cell growth [33]. It has been reported that the protein also functions as a chaperone at low temperatures [34]. However, later studies indicate that DegP do not function as a chaperone, but rather only as a protease [33]. *degP* is also the only gene of the transcription unit as seen in figure 1.3.

Figure 1.3: The transcription unit of *degP*. The figure is obtained from the EcoCyc database. CpxR: DNA-binding transcriptional dual regulator; degPp: promoter; σ^{24}: σ factor 24; H-NS: DNA-binding transcriptional dual regulator represses transcription from promoter.

Skp or the "**S**even **k**iloDalton **P**rotein" is a periplasmic chaperone. It has been shown to interact with a range of different outer membrane proteins [35], and improves folding of the outer membrane protein OmpA [36]. Proteins with poor propensity to form soluble products have been suggested to have a higher chance of solubility improvement by co-expression with *skp* [37]. This theory was strengthened as the solubility of a single-chained antibody (scAb) fragment was improved by co-expressing *skp* [38]. When co-expressed in the cytoplasm, expression of correctly folded Fab Antibody fragments in the cytoplasm is increased [39].

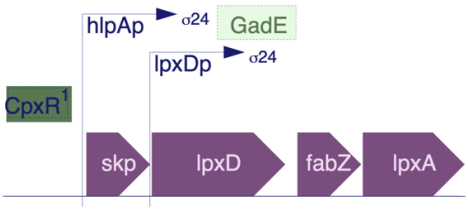


Figure 1.4: The *skp, lpxD, fabZ, lpxA* operon. The figure is obtained from the EcoCyc database [40] (not to scale). CpxR: DNA-binding transcriptional dual regulator; hlpAp: promoter; lpxDp: promoter; σ^{24}: σ factor 24; GadE: DNA-binding transcriptional activator.

The *skp* gene is ordered in an operon together with *lpxD*, *fabZ* and *lpxA* [41] as seen in figure 1.4. LpxD [42] (UDP-3-O-(3-hydroxy-myristoyl)glucosamine N-acetyltransferase) and LpxA (UDP-N-acetylglucosamine acyltransferase) are both enzymes involved in lipid A biosynthesis. Lipid A is a major component of the lipopolysaccharide (LPS) membrane of gram-negative bacteria, and is essential for cell viability [43]. FabZ (β-hydroxyacyl-acyl carrier (ACP) dehydratase) is involved in the type II, dissociated fatty acid biosynthesis [44].

σ factors are subunits of RNA polymerase complex enzymes, and functions as control units that guides the enzyme to the promoter [45]. Both transcription of the *skp* operon and *degP* are controlled by the sigma factor σ^{24}. The σ^{24} regulon is activated upon accumulation of misfolded polypeptides in the periplasm of *E. coli* [41]. *sppA* is controlled by the most common sigma factor, σ^{70} [45].

Alltogether, the native genes *sppA*, *skp* and *degP* constitutes three potential genomic targets for enhanced periplasmic protein production in *E. coli*. Based on the presented review, the following hypotheses are proposed for the genomic targets: The up-regulation of *sppA* and *skp* has a poteintial of enhancing periplasmic recombinant protein production by increased signal peptide peptidase and chaperone activity respectively. Down-regulation of *degP* could potentially reduce protease activity, inhibiting recombinant protein degradation.

1.3.2 Regulating native gene expression levels by RBS replacement

The ability to create knock-out mutations has for a long time been an important method for mapping the function of genes. An example was the characterization of DegP, also called HtrA (high temperature requirement). The knock-out mutation of this gene resulted in inability to grow at high temperatures [46]. Traditionally, up-regulation has been performed by co-expression on a plasmid vector [15, 16, 18–20], or by adding an additional copy of the gene to the genome [31]. With the possibility of genome editing techniques such as recombineering, it is now possible to change single base pairs or larger sequences in order to change the start codon or to create mutations in the ribosome binding site (RBS) [47] or the promoter [48].

In this thesis, regulation of native gene expression has been performed by changing the complete RBS and by changing the start codons of these genes. The RBS of a gene is part of the 5′ untranslated region (5′-UTR) harbouring the Shine-Dalgarno (SD) sequence that directs the mRNA to the right position of the ribosome [45]. The sequence of the RBS affects the translation initiation rate of the gene [30]. Salis Ribosome Binding Site Calculator is a tool that generates altered RBS sequences [49] based on a thermodynamic model developed by Salis et al. [30, 50]. This model allows to calculate the translation initiation rate (TIR) of an initial RBS. This step is referred to as reverse engineering. For generation of a new RBS sequence, called forward engineering, the thermodynamic model is used in an algorithm that generates a new RBS sequence with the desired TIR. The algorithm starts with a ran-

dom mRNA sequence. For each iteration, a random single nucleotide change is performed, and the TIR is calculated again. This is repeated until the calculated TIR of the generated RBS sequence is within a certain limit of the desired TIR[30].

This method for regulating the gene expression level allows to specify the expression level without the need for trial- and error optimization [30]. The RBS sequence is easily generated and can be inserted to the genome by recombineering (see section 1.4). RBS replacement may be combined with start codon modification to further manipulate the TIR. However, some challenges need to be considered when using this method, limiting the potential genomic targets. Altering the RBS sequence and/or the start codon will not only affect the TIR of the native gene, but also any other genes organized in the same operon. When selecting the genomic targets for regulation of expression, other genes in the same operon that will also be affected by the up-/down-regulation must be considered. Genomic targets that are the only gene of the transcription unit such as *sppA* and *degP* are therefore suited targets for regulation by this method. *skp* is in an operon with the *lpxD*, *fabZ*, and *lpxA* genes, but these three genes are under control of the separate lpxDp promoter [41].

1.4 Genome editing techniques

The CRISPR/Cas9 system has in the latest years been given a lot of attention due to the specificity of the method and the wide potentials. However, genome editing is not a completely new idea – already in the mid 1990's Zink-Finger Nucleases (ZFN's) were modified to target specific DNA sequences [51]. Later, λ phage proteins were found to facilitate recombination in *E. coli*, allowing to introduce specific changes in the genome [52]. In this chapter, different methods for genome editing will be addressed, with a focus on CRMAGE.

1.4.1 CRISPR/Cas9

CRISPR/Cas (Clustered Regularly Interspaced Palindromic Repeats and the CRISPR-associated proteins) are components of the adaptive immune systems of various bacteria and archaea [53, 54]. The native function of the system is to act against invading genetic elements such as plasmids and viral DNA by targeting the foreign DNA at specific sequences and creating blunt double-stranded breaks (DSBs) [53]. Three types of CRISPR systems have been identified [53, 54]. The type II system consisting of the Cas9 nuclease and two non-coding CRISPR RNAs (crRNAs): pre-crRNA and tracrRNA [55] is one of the best characterized. The CRISPR/Cas9 system has been modified, and is now widely used in various forms for genetic engineering [51].

The CRISPR/Cas adaptive immunity is mediated by a three-step process: **adaption**, **expression** and **interference**. In the **adaption** phase, short sequences of the for-

eign DNA (protospacers), chosen by recognition of the protospacer adjascent motif (PAM) sequence, are integrated to the CRISPR loci as spacer sequences. The PAM sequence varies between CRISPR systems, and consists of a few nucleotides [53]. For the Cas9 derived from *Streptococcus pyogenes* (type II system), the target DNA must be followed by a 5′-NGG PAM sequence [54]. The CRISPR locus in the genome contains direct repeats separated by the integrated spacers from foreign DNA. The CRISPR region is transcribed to pre-crRNA that is processed into discrete units of crRNA in the **expression** step. Each discrete unit contains a guide sequence (gRNA) that targets the Cas9 nuclease to the target DNA loci, and a direct repeat region [53]. In the type II system, a tracrRNA pairs with the direct repeat fragment, and the dual RNA sequence is loaded to a Cas9 nuclease [51]. In the **interference** phase, the spacer sequence interacts with the targeted DNA by Watson-Crick base pairing, and the nuclease creates a DSB in the targeted sequence $\sim$3 bp upstream of the PAM sequence [54].

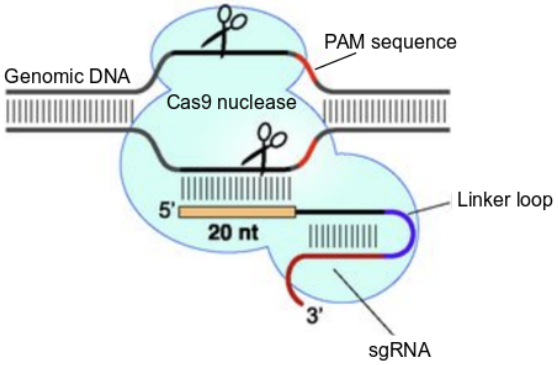


Figure 1.5: The Cas9 nuclease creates a specific double stranded break at the DNA locus containing the PAM sequence at the 3′ end. The linker loop connects the tracrRNA to the crRNA containing the gRNA (20 nt). Figure obtained from Jinek et al. with some adaptations [56].

This adaptive immune system may be utilized to create targeted DSB, and the system is under intensive research to find new applications and improvements [51]. The dual trackrRNA:crRNA has been engineered to a sgRNA complex, conserving the gRNA feature at the 5′ end and the double-stranded sequence at the 3′ end that binds to Cas9 (see figure 1.5) [56]. In contrast to the ZFNs and TALENs that require complex protein modifications, the Cas9 can be targeted to a specific DNA loci simply by changing the gRNA sequence [51].

1.4.2 MAGE recombineering

Multiplex Automated Genome Engineering is a method that allows efficient large-scale programming and evolution of cells. The method utilizes λ/Red β-proteins to simultaneously modify many locations in the chromosome of a cell or a population

of cells [57]. The bacteriophage λ/Red β-protein promotes annealing of complementary DNA, and allows the recombination event between a short ssDNA sequence and genomic or plasmid DNA [52]. Utilization of phage proteins for genetic engineering is called recombineering or homologous DNA recombineering [58].

Homologous DNA recombineering is achieved in *E. coli* by electroporation of an oligonucleotide, here called the MAGE oligo. The ends are homologous to the target loci of the genome (or other DNA segments) [52]. Desired changes are introduced by a non-homologous region in the MAGE oligo, or by deleting a region. The efficiency of MAGE recombineering depends on the number of changed nucleotides and the length of the homology region [57]. The MAGE oligo hybridizes to the target DNA loci close to the replication fork between discontinuous Okazaki fragments during replication as illustrated in figure 1.6 [59].

1.4.3 CRMAGE - CRISPR optimized MAGE recombineering

CRMAGE is a method that is based on λ/Red recombineering, and uses the CRISPR/-Cas9 system for negative selection. It is used to introduce small, specific genome modifications such as deletions, insertions or substitutions [47]. Like for MAGE recombineering, a MAGE oligo with homologous ends as shown in figure 1.6, and a heterologous region is designed to incorporate the desired mutation facilitated by λ/Red recombineering enzymes [47]. The MAGE oligo is directed to the replication fork during replication [59].

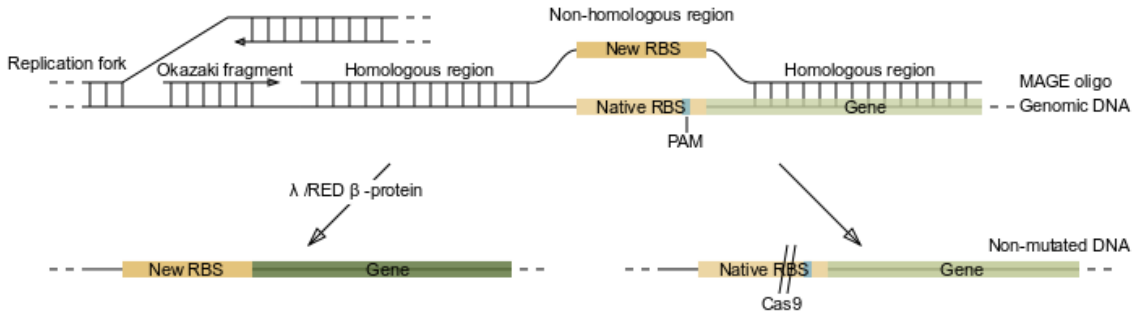


Figure 1.6: Illustration of CRMAGE based on theory from Ronda et al. [47]. The MAGE oligo binds between the Okazaki fragments close to the lagging strand of the replication fork [59]. λ/RED β-protein facilitate the recombination event [52], allowing the insertion of a new RBS sequence. Simultaneously, the PAM sequence is interrupted, and the Cas9 nuclease is targeted to create double-stranded breaks in non-mutated DNA [47].

Presented in figure 1.6, a new RBS sequence is inserted in place of the native RBS sequence by λ/Red β-proteins. A gRNA is constructed that specifically targets the Cas9 nuclease to create DSBs in non-mutated DNA as a function of negative selection [47]. The CRMAGE system consists of three plasmids; pMA7CR_2.0 and pMAZ-

SK presented in figure 1.7 and pZS4Int-tetR harbouring a Tet repressor inhibiting expression of Cas9 and sgRNA [47].

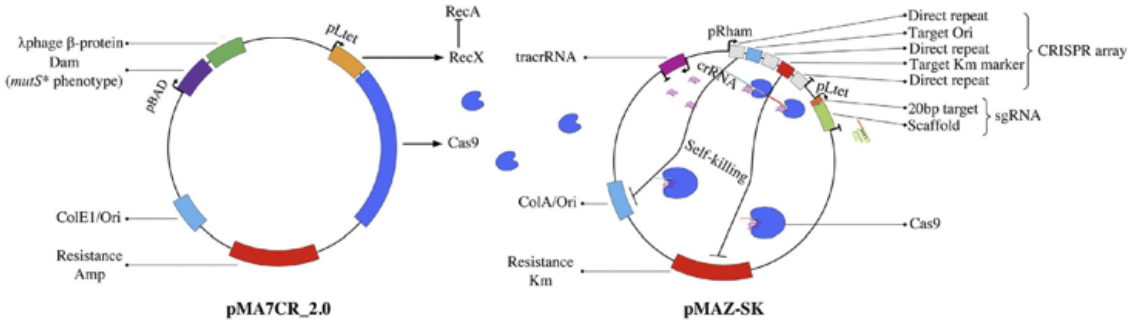


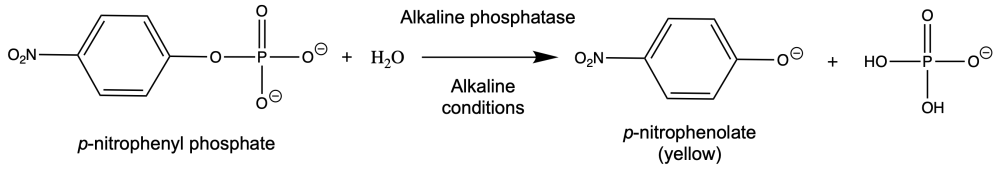
Figure 1.7: Schematic overview of the CRMAGE system with plasmid maps of pMA7CR_2.0 and pMAZ-SK. Figure from Ronda et al. [47]. The Cas9 nuclease, λ phage β proteins and RecX inhibiting the RecA function, blocking the repair of DSBs are expressed from pMA7CR_2.0 [47]. λ phage β proteins are expressed with Dam proteins, giving hypermutable cells [60]. pMAZ-SK harbours the CRISPR-array consisting of direct repeats and the spacer sequences targeting Cas9 for self-killing. The sgRNA contains the 20 bp target sequence for directing the Cas9 nuclease. A tracrRNA is combined with the crRNAs for targeting ColA/Ori and kanamycin resistance genes for self-destruction [47]. Amp: ampicillin; Km: kanamycin.

1.5 Reporter systems for monitoring periplasmic recombinant protein expression

Monitoring of periplasmic recombinant protein expression may be done by different reporter systems. Most are based on fusion of a protein of interest to a reporter protein such as an enzyme [61] or a fluorescent protein.

Fusion with fluorescent proteins is an important technique to analyze the sub-cellular localization of proteins. The fluorescing property allows direct visualization of protein localization by fluorescence microscopy. Fusion of the genes encoding the protein of interest and the fluorescent protein is often called a reporter construct. A popular and useful fluorescent marker is the green fluorescent protein (GFP) [45], but wild-type GFP tends to misfold when expressed in *E. coli* [62]. Superfolding GFP (sfGFP) is an optimized variant engineered to have better folding properties, also in fusions [63]. In previous work, the functionality of a reporter system of an antibody fragment fused to sfGFP was tested. The evaluation of periplasmic protein expression by fluorescence measurements was determined as unreliable because cytoplasmic reporter fusions were also fluorescent. It was therefore not possible to conclude whether the reporter was only periplasmically expressed [64].

Alkaline phosphatase is an enzyme that is natively produced in various forms in diverse organisms such as mammals [65] and bacteria [66]. The enzyme catalyze the de-phosphorylation of various organic compounds. De-phosphorylation of ρ-nitrophenyl phosphate is catalyzed by alkaline phosphatase and creates the yellow substance p-nitrophenolate as shown in equation 1.5.1 in alkaline conditions [65].



(1.5.1)

The alkaline phosphatase is inactive when produced in the cytoplasm. The gene for alkaline phosphatase, $phoA$ is suited for use in fusions because the alteration of the N-terminus does not seem to affect the enzymatic activity to any great extent [61].

As most enzyme reactions, the kinetics of the alkaline phosphatase reaction follow the Michaelis-Menten kinetics shown in equation 1.5.2.

$$v = \frac{v_{max} \cdot s}{\mathrm{K}_m + s} \tag{1.5.2}$$

where v is the volumetric rate of reaction, v_{max} is the maximum rate of reaction, s is the substrate concentration and K_m is the Michaelis constant [67]. For low substrate concentrations ($s << \mathrm{K}_m$), the reaction rate increases linearly with the substrate concentration, and follow first-order kinetics ($v \approx \frac{v_{max}}{K_m} s$) with rate constant $\frac{v_{max}}{K_m}$. When the substrate concentration is sufficiently high ($s >> \mathrm{K}_m$), the K_m is negligibly small compared to s, and the reaction rate will be approximately equal to the maximum rate of reaction ($v \approx v_{max}$). That means that the reaction rate approaches a constant value with high substrate concentrations, and follow zero-order kinetics [67]. For enzymatic assays, it is therefore critical that the substrate concentration is sufficiently high so that the reaction rate is constant.

The reporter system used in this thesis, pTA16-s173-AP, was constructed by Balzer et al. [68] and is shown in figure 1.8. The genetic elements of the reporter plasmid are described in table 1.1. It is based on the RK2 replicon [68], characterized by the minimal components oriV and *trfA* allowing self-replication [69]. Expression of the reporter protein in this plasmid is under control of the XylS/*Pm* regulator/promoter system [70]. The selection marker of pTA16-s173-AP is *kan* conferring kanamycin resistance. The reporter construct encodes the protein fusion of an industrially relevant single-chain antibody fragment of human origin, scFv173 [71] and alkaline phosphatase [61]. The signal sequence of PelB directs the reporter system to the periplasm [72].

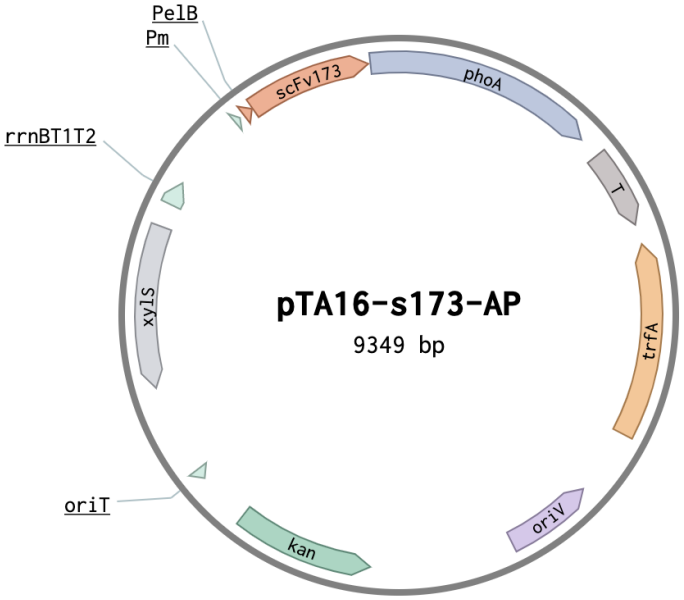


Figure 1.8: Plasmid map of the reporter plasmid pTA16-s173-AP harbouring the reporter gene construct created by Balzer et al. [68].

Table 1.1: The genetic elements of pVB1-251 and reporter segments. AB: antibody.

Segment	**Description**	**Reference**
xylS	Regulator	[70]
Pm	Promoter	[70]
oriT	Origin of conjugative plasmid transfer from RK2	[69]
oriV	Origin of vegetative replication from RK2 replicon	[69]
kan	Gene conferring kanamycin resistance	
trfA	Replication initiation from RK2.	[73]
T	Transcription terminator	
phoA	Alkaline phosphatase	[61]
scFv173	Industrially relevant single-chain AB fragment	[71]
PelB	Signal sequence of PelB from *Erwinia carotovara*	[72]
rrnBT1T2	Transcriptional terminator	[74]

1.6 Aim of this thesis

The aim of this thesis is to optimize periplasmic production of recombinant proteins in *E. coli* by introducing genomic mutations using CRMAGE recombineering. The hypothesis is that the up-regulation of the translation of signal peptide peptidase SppA and the periplasmic chaperone Skp, and down-regulation of the periplasmic protease DegP respectively will lead to improved secretion of the reporter protein.

The aim can be divided into six parts:

- Constructing a synthetic RBS sequence and start codon for the desired regulation of the native genes *sppA*, *skp* and *degP*
- Construction of the pMAZ-SK helper plasmid including *in silico* gRNA design
- Up-regulate the translation initiation rates (TIRs) of *sppA* and *skp* and down-regulate *degP* TIR by changing the respective RBS sequences in the genome of *E. coli* BL21 by CRMAGE recombineering
- Plasmid curing - elimination of CRMAGE helper plasmids from constructed strains
- Establishment of a fuctioning protocol for osmotic shock and alkaline phosphatase assay
- Analysis of constructed strains by evaluation of the periplasmic expression efficiency of the reporter system by alkaline phosphatase assay

CHAPTER 2

Materials and methods

Media and solutions are described in appendix A.

2.1 Bacterial strains

E. coli DH5α was used for cloning purposes in this project to allow more convenient DNA isolation as this strain has been found to provide higher yields of DNA for double-stranded DNA sequencing [75]. For the rest of the experiments, including CRMAGE, *E. coli* BL21 was used due to the features favoured in protein expression such as the ability to grow to higher cell densities [76]. BL21 is deficient in Lon and OmpT proteases making the strain suitable for recombinant protein production [13]. The strains constructed in this thesis were based on BL21, and are presented in the results chapter 3.1, table 3.1.

2.2 Plasmid construction and oligonucleotide design

The plasmids used in this project are presented in table 2.1. The pMAZ-SK plasmid was modified to contain gRNA targeting the native sequence of the RBS for each of the targeted genes (*sppA, skp and degP*).

Plasmids used for CRMAGE

The pZS4Int-tetR plasmid express the TetR repressor inhibiting expression of Cas9 and sgRNA [47]. The plasmid carries a Spectinomycin resistance gene as selection marker and is a low copy-number plasmid with pSC101 origin of replication [78].

Table 2.1: Plasmids used in this project. AP: alkaline phosphatase

Plasmid	Remarks	Source
pMA7CR_2.0	Amp^r; Cas9 (aTet induced); λ phage β protein; RecX; Dam	[47]
pZS4Int-tetR	Spc^r; TetR repressor controlling Cas9 and sgRNA expression. Low copy nr. plasmid.	[77]
pMAZ-SK	Kan^r; sgRNA (aTet induced), CRISPR-array, Self-Killing (L-rhamnose induced); tracrRNA	[47]
pMAZ-SK_sppA	pMAZ-SK with gRNA targeting native RBS region of *sppA*	This project
pMAZ-SK_skp	pMAZ-SK with gRNA targeting native RBS region of *skp*	This project
pMAZ-SK_degP	pMAZ-SK with gRNA targeting native RBS region of *degP*	This project
pTA16-s173-AP	Kan^r ; expressing scFv173-AP fusion (m-toluate induced)	[68]

The gRNA is cloned into the pMAZ-SK plasmid that expresses the aTetracycline inducible sgRNA. The plasmid carries a Kanamycin resistance gene as selection marker. It also carries the sgRNA targeting the origin of replication and the kan^r gene, resulting in a plasmid self-killing function inducible by L-rhamnose [47].

2.2.1 *In silico* oligonucleotide design for CRMAGE

When designing the CRMAGE system, two oligos are needed. The MAGE oligo including the new RBS sequence, and the gRNA that is needed to target the Cas9 nuclease to non-mutated sequences for negative selection.

The MAGE recombineering oligo is needed to be able to change the desired DNA sequence, in this case the RBS corresponding to the *sppA*, *skp* and *degP* genes. It is a single stranded oligo consisting of the new RBS sequence and a long overlap that is complementary to the flanking sequences to ensure correct binding of the oligo to the specific target sequence (see section 1.4 for details).

The synthetic RBS sequences were obtained as follows:

1. The EcoCyc database [40] was used to find the -35 and -10 promoter regions and transcription start site of each target gene. The complete genome sequence of BL21(DE3) (Genbank: CP001509.3) was downloaded to the online DNA editing tool, Benchling (`benchling.com`)

2. Initial rate of translation was found for each native RBS by using the Reverse Engineer tool [49]. The mRNA sequence was entered from the transcription

start site. BL21(DE3)(ACCTCCTTA) was chosen as the organism, and the version 2.1 Free Energy Model was run.

3. A new RBS sequence was designed by using the Forward Engineer tool [49]. The desired TIR and the protein coding sequence (from the start codon) was entered. The length of the new RBS and the new TIR was noted and the equivalent base pairs marked in Benchling. The new RBS sequence was also used to replace one of the PAM sequences recognized by Cas9 (NGG). If no PAM is found within the native RBS sequence, a single nucleotide substitution can be introduced mediated by the MAGE oligo to interrupt the PAM sequence.

4. The new TIR was finally found by using the Reverse Engineer tool on the newly obtained sequence.

When constructing the MAGE oligo, the efficiency is critically increased when the oligo is added between the Okazaki fragments of the "lagging" strand. Therefore, the position of the gene in either replichore 1 or 2 and the location of the gene on the forward or reverse strand decides whether the MAGE oligo should be designed as a forward or reverse sequence as illustrated by Gallagher et al. in figure 2.1. *sppA* is located in replichore 2, and *skp* and *degP* is located in replichore 1 – all on the forward (+) strand. The MAGE oligo for *sppA* was therefore designed as a forward sequence, while the MAGE oligos for *skp* and *degP* were constructed as reverse sequences. 90 bp overlaps were used for the construction of the MAGE oligos, and the final oligos are shown in table D.2 (Appendix D). The native RBS of *degP* did not contain a PAM sequence, and the MAGE oligo therefore also introduces a single nucleotide substitution to interrupt a PAM sequence outside the native RBS.

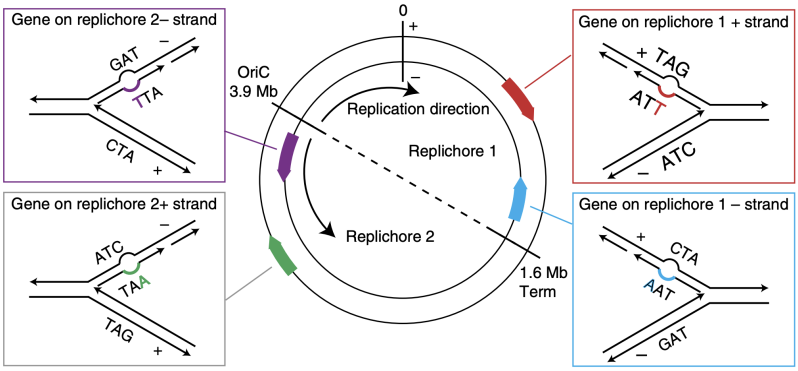


Figure 2.1: The placement of a gene in either replichore 1 or 2 of *E. coli* and on the forward or reverse strand determines whether the MAGE oligo should be designed as a forward or reverse sequence. Illustration adapted from Gallagher et al. [59].

The gRNA sequences are needed to target the Cas9 nuclease to create DSBs in sequences that has not been changed by MAGE recombineering as a method of negative selection. The gRNA was designed using Benchling's built-in CRISPR tool. This tool searches for the PAM-sequence (-NGG) in a marked region, creates a ~20

bp gRNA, and calculates the off-target score [79] for this sequence. This score is a specificity score between 0 and 100, and gives an indication of the probability for incorrect binding of the Cas9 protein to non-targeted DNA in the genome, where higher score is better.

2.2.2 *In vitro* plasmid construction

pMAZ-SK, pMA7CR_2.0 and pZS4Int-tetR plasmids were extracted from DH5α cells using the Zymo Research plasmid MiniPrep kit (catalog # D4016). Polymerase chain reaction (PCR) was used to amplify the pMAZ-SK backbone. pMAZ-SK_sppA, pMAZ-SK_skp and pMAZ-SK_degP plasmids were constructed by assembly of the pMAZ-SK backbone to gRNA sequences for the respective genes.

pMAZ-SK backbone amplification

CloneAmp™ HiFi PCR Premix (Cat. No. 639298) was used to amplify the pMAZ-SK backbone. 0.5 µL of each primer from 10 µM stock (Gibson_pMAZ_backbone_F and Gibson_pMAZ_backbone_R, see Appendix D), $<$ 100 ng template (pMAZ-SK) and 11.2 µL dH_2O to one 12.5 µL aliquot of CloneAmp™ HiFi PCR Premix. All PCR reactions were run in a Mastercycler® nexus X2 PCR machine with cycler conditions shown in table 2.2.

Table 2.2: Thermal cycler conditions used for PCR with CloneAmp™ HiFi PCR.

Step	Temperature	Time
Denaturation	98 °C	10 s
Annealing	55 °C	15 s
Elongation	72 °C	2.5 min

Gel electrophoresis

Gel electrophoresis was used for size-based separation of DNA fragments of PCR products and digested plasmids. DNA samples were loaded to a GelRed/GelGreen 0.8% agarose gel with Purple Gel Loading Dye (6×, #B7024S) and run at 100 V, 400 mA in TAE running buffer. The gels were analyzed in a Gel Doc™ Gel Documentation System from BioRad. GelRed was used for analyzing purposes, while GelGreen was used in cases where the DNA fragments were to be extracted from the gel. 2-Log DNA Ladder from New England Biolabs (catalog #NS3200S) was used for all gel electrophoresis steps in this thesis, and is shown in figure 2.2.

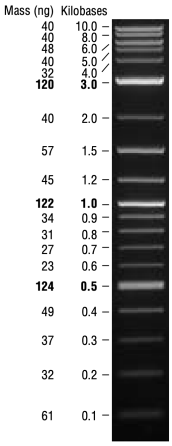


Figure 2.2: 2-log DNA ladder (NS3200S), New England Biolabs.

Annealing

Single stranded gRNA sequences provided by Sigma-Aldrich were annealed to double stranded DNA. The gRNA sequences were ordered as complementary single-stranded fragments (see D) and annealed by preparing a mix of 10 µL forward oligo, 10 µL reverse oligo (100 µM), 10 µL NeB Buffer 4 and 70 µL water. The annealing mix was boiled at 95 °C for 5 min, and the temperature was gradually decreased by 1 °C/min.

Gibson assembly

Gibson assembly [80] was used to assemble the pMAZ-SK backbone to the gRNA fragments. 5 µL DNA mixture with equimolar masses of backbone and gRNA insert was added to 15 µL Assembly Master Mix (see Appendix A) and incubated for 1 h at 50 °C. The master mix was used directly for transformation into *E. coli* DH5-α (see section 2.3).

Restriction digestion

Restriction digestion was used to verify that the correct plasmids were extracted. The online DNA editing tool, Benchling was used to find restriction sites and appropriate restriction enzymes. 1 µL DNA, 1 µL buffer (see table 2.3), 5 µL nuclease free water and 0,25 µL restriction enzyme was incubated in a water bath at 37 °C for approximately 1 h. The digestion mix was separated by gel electrophoresis. Expected fragment sizes after digestion with the restriction enzymes and buffers used are presented in table 2.3. pMAZ-SK plasmids were also sent for sequencing for verification of the correct gRNA insert.

Table 2.3: Restriction digestion. All reactions were run at 37 °C, 1h.

Plasmid	Restriction enzyme	Expected fragm. sizes (bp)	Buffer
pMA7CR_2.0	DraIII-HF	10934	3.1
pZS4Int-tetR	EcoRI	6000	2.1
pMAZ-SK	EcoRI	2779	2.1
pMA7CR_2.0	NheI	7132, 3802	2.1
pZS4Int-tetR	NdeI	4492, 741, 581, 186	2.1
pZS4Int-tetR	DraIII	5380, 620	3.1
pMAZ-SK	NheI	1783, 973, 23	2.1

2.3 Preparation of competent cells and transformation

E. coli DH5α was made competent by RbCl for transformation of the constructed pMAZ-SK plasmids, while electroporation was used for transformation of *E. coli* BL21. Preparation of competent cells and transformations were performed as previously described [64].

2.3.1 RbCl competent cells

Previously prepared competent cells with a competency of $5.4 \cdot 10^7$ cfu/μg DNA[64] were used for RbCl transformation. Transformation was performed by adding 1-50 ng DNA to competent (100 μL alliquot) cells tawed on ice, and incubated on ice (20 - 30 min). Cells were heat-shocked (42 °C, 45 seconds), incubated on ice (2 min), and SOC media (1 mL, 37 °C) was added before incubation (1 h, 37 °C, 225 rpm). Cells were plated on LA medium with appropriate selective antibiotics.

2.3.2 Electrocompetent cells

Electroporation was used for transformation of pMA7CR_2.0, pZS4Int-tetR and the reporter system pTA16-s173-AP. Electrocompetent *E. coli* BL21 was prepared fresh before each transformation by incubating 1% overnight culture in LB medium (10 mL) to OD_{600} = 0.45±10% (37 °C, 225 rpm). Cell suspension was alliquoted (1.5 mL) on ice and cells were harvested (12000 g, 1 min, 4 °C). Pellet was resuspended in cold sucrose (300 μM, 1 mL) and harvested (12000 g, 1 min, 4 °C). Cells were resuspended in glycerol (cold, 10%) and kept on ice until used for electroporation (within 15 min).

Transformation was conducted by adding 100-500 ng high quality (no trace salt) plasmid DNA to electrocompetent cells and electroporated in a cold 0.2 cm cuvette using a BioRad Gene Pulser Xcell™ Electroporation System at 2500 V (Pre-set protocol for *E. coli*). SOC medium (1 mL, 37 °C) was added immediately after electroporation, and cells were incubated (37 °C, 225 rpm, 1 h) before plating on LA medium

with appropriate selective antibiotics.

2.4 CRMAGE

2.4.1 CRMAGE protocol

The CRMAGE (CRISPR-optimized MAGE recombineering) protocol is adapted from Ronda et al. [47] with a few changes.

15 mL LB lennox (100 μg/mL ampicillin, 100 μg/mL spectinomycin) was inoculated with 150 μL ON culture of BL21;pMA7CR_2.0;pZS4Int-tetR and incubated (37 °C, 225 rpm) to OD_{600nm}=0.5. Expression of λ Red proteins was induced by adding L-arabinose (0.2 g/mL) to a final concentration of 0.2 %. Incubation was continued (10-15 min), and the culture was put directly to an ice-water bath and left to cool for 20 min to avoid degradation of λ Red proteins.

Cells were washed by centrifugation (6500×g, 4 °C, 5 min) and resuspension in ice-cold MilliQ water three times with 35 mL, 15 mL and 1 mL water respectively. The cell suspension was finally centrifuged (7000×g, 4°C, 6 min), and the supernatant was removed completely before resuspending in ≈400 μL ice-cold MilliQ water. 50 μL cell suspension was added to 0.5 μL MAGE oligo (10 pmol/μL) and 1 - 3 μL pMAZ-SK in water (≈250 ng). The cells were electroporated at 2.5 kV in a 2 mm gap cuvette and added 950 mL LB lennox (100 μg/mL ampicillin, 100 μg/mL spectinomycin) immediately after electroporation. The culture was recovered in a 13 mL tube (37 °C, 225 rpm, 1 h) before selection for pMAZ-SK transformants by adding kanamycin to a concentration of 50 μg/mL.

Expression of Cas9 and transcription of sgRNA for negative selection was induced by addition of aTetracycline (0.001 g/mL) to a final concentration of 200 ng/mL. Tubes containing aTet were covered due to light sensitivity of aTetracycline. The culture was left to recover overnight (≈18 h) and plated on selective media (LA with 100 μg/mL ampicillin, 100 μg/mL spectinomycin and 50 μg/mL kanamycin)

2.4.2 Verification of insert

Colony PCR was used to verify whether the new ribosome binding site sequences were inserted into the genome. The forward primers (table B.1, Appendix D) were designed to bind specifically to the insert, while the reverse primer was designed to bind about 400 nucleotides downstream of the insert. Consequently, a PCR product in the size range of 400 nt will only appear if the insert was successfully inserted.

NEB *Taq* DNA polymerase with Standard Taq Buffer (catalog #M0273) was used for

colony PCR. Colonies were picked, replated, and a small part was transferred to 10 µL alliquots of mastermix (table A.1 in Appendix A). The samples were denatured (10 min, 95 °C) before proceeding with the normal *Taq* DNA Polymerase protocol shown in table 2.4 with 30 rounds in a Mastercycler® nexus X2 PCR machine. The annealing temperatures were adapted for the different primer pairs.

Table 2.4: Thermal cycler conditions used for PCR with *Taq* DNA Polymerase.

Step	**Temperature**	**Time**
Denaturation	95 °C	20 s
Annealing	45 - 68 °C	40 s
Elongation	68 °C	1 min/kb

The PCR product was analyzed by gel electrophoresis as described in section 2.2.2. Sequencing conducted by Lightrun, GATC Biotech AG (`www.gatc-biotech.com/lightrun`) was used to verify the inserts after colony PCR. Colony PCR with sequencing primers (see D.1) covering the sequence upstream and downstream of the insert was used, and the product was run on GelGreen (0.8%) electrophoresis gel. The band conferring to the amplified DNA sequence was cut out and purified by Zymoclean Gel DNA Recovery Kit (catalog #D4007/D4008). The resulting concentration of DNA was measured with a NanoDrop™ from ThermoFisher Scientific, and sent for sequencing with one of the sequencing primers.

Q5 DNA polymerase was used in cases where the resulting DNA concentration from the gel purification was too low to send for sequencing directly (< 20 ng/µL). Q5 was used because Taq polymerase is more error-prone, but can be used for colony PCR due to low template quality requirements. A larger reaction volume was used in order to achieve a higher concentration of DNA by loading a larger volume to the gel. The master mix (see A.2 in Appendix A) was added 1 µL low-concentration DNA and the reaction was run with thermocycle conditions shown in table 2.5, 30 rounds.

Table 2.5: Thermal cycler conditions used for PCR with Q5 DNA Polymerase. An initial 30 s denaturation at 95 °C for the first round and an additional 2 min elongation for the last round was used.

Step	**Temperature**	**Time**
Denaturation	95 °C	15 s
Annealing	55 °C	15 s
Elongation	72 °C	∼3 min/kb

2.5 Plasmid curing

Plasmid curing is the elimination of a plasmid from a cell. The method for curing the self-killing pMAZ-SK plasmids was adapted from Ronda et al. [47] with some adaptations. An overnight culture of the correct strain harbouring pMAZ-SK, pMA7CR_2.0 and pZS4Int-tetR was directly induced with L-rhamnose (100 μg/mL) and aTetracycline (100 μg/mL), incubated >4 h and plated on LA plates supplemented with Spectinomycin (100 mg/mL) and Ampicillin (100 mg/mL). Verification of pMAZ-SK curing was performed by re-plating colonies on LA supplemented with kanamycin (50 μg/mL).

2.5.1 Curing by high-temperature growth

pMA7CR_2.0 and pZS4Int-tetR was described by Elvheim to be curable by repeated growth at 42 °C on non-selective LA plates [81]. This method was successful for curing pMA7CR_2.0, but unsuccessful in this project for pZS4Int-tetR curing, and alternative curing methods were tried.

2.5.2 Curing by electroporation

A method by Heery et al. for curing by electroporation [82] was tested in order to cure the constructed strains from pZS4Int-tetR. A 10 mL overnight culture of BL21;pZS4Int-tetR in LB was chilled in ice and harvested (4 °C, 9000 g). The cells were washed by resuspension in dH_2O, and washing was repeated three times. Finally, the pellet was resuspended in dH_2O of equal amount as pellet ($\sim$ 400 μL), and 40 μ cell suspension was added to a cold 2 mm electroporation cuvette. The cells were electroporated at 2,5 kV, 25 μF, and 960 μL SOC-medium was added. The cells were left to recover for 1 h (37 °C, 225 rpm), and plated on non-selective LA.

2.5.3 Curing with acridine orange and penicillin G

Curing by acridine orange treatment was performed based on the methods described by Tomás and Kay [83]. Different working concentrations of acridine orange (75 μg/mL and 100 μg/mL) and penicillin G (1000 U/mL and 10^4 U/mL) were tested.

5 mL LB supplemented with 75 μg/mL or 100 μg/mL acridine orange (7,5 μL or 10 μL from 50 mg/mL acridine orange stock solution) was inoculated with 1% overnight culture (BL21;pZS4Int-tetR in 5 mL LB). The acridine orange culture was incubated overnight (37 °C, 225 rpm) without light exposure. The following day, cells were plated on LA and LA supplemented with spectinomycin (100 μg/mL) to check for curing efficiency by acridine orange alone.

5 mL LB supplemented with sublethal concentrations of spectinomycin (10 μg/mL) was 1% inoculated with the overnight acridine orange-treated culture, and incubated (37 °C, 225 rpm, 30-45 min) to ensure expression of antibiotic genes. Penicillin G (1000 U/mL or 10^4 U/mL) was added, and incubation continued for 90 min. The surviving cells were harvested by centrifugation (12 000 g, 5 min), supernatant was carefully removed, and the pellet was resuspended in 0.9% NaCl solution to a volume of 5 mL. 200 μL cell suspension was plated on non-selective LA and grown overnight in 37 °C. Curing efficiency was determined by re-plating single colonies on selective plates (LA supplemented with 100 μg/mL spectinomycin).

2.6 Protein expression and analysis

The constructed strains were analyzed by transformation of the reporter plasmid and expression of the reporter protein fusion scFv173-AP. Cell samples were prepared by inoculating 15 mL LB medium supplemented with kanamycin (50 μg/mL) with an ON culture to OD_{600nm} = 0.05. The cell culture was incubated at 37 °C to OD_{600nm}=0.5 and induced with m-toluate to a total concentration of 2 mM before 5 h expression at 30 °C. The OD_{600nm} was measured and the samples were frozen or used directly for further analysis.

2.6.1 Osmotic shock and alkaline phosphatase assay

Extraction of periplasmic proteins including the reporter protein was performed by cell exposure to a hypertonic solution followed by exposure to cold water (hypotonic solution), a method also called osmotic shock [84]. Osmotic shock was conducted as described by Hong et al. [85] with some modifications.

The culture sample (7.5 mL) from the protein expression was centrifuged (11700 g, 5 min, 4 °C). The pellet was resuspended in 1 mL 20% sucrose-0.03 M Tris-HCl (pH 8.0) and added 0.25 mL 5 mM disodium EDTA (pH 8.0). The sample was mixed (180 rpm, 10 min, r.t.) and centrifuged (13000 g, 10 min, 4 °C). The pellet was resuspended in 1.25 mL ice-cold dH_2O and mixed in an ice bath (180 rpm, 10 min). The periplasmic fraction (supernatant) and the cytoplasmic fraction were separated by centrifugation (13000 g, 10 min, 4 °C).

The alkaline phosphatase assay was conducted as described by Yang and Metcalf [86] with some modifications. 10 μL periplasmic fraction was added to a 96-well plate. Each sample was added 90 μL 1 M Tris-acetate (pH 8) with p-nitrophenyl phosphate (pNPP) (0.4%) simultaneously (0.36% total pNPP concentration), mixed in a plate reader and the 550 nm absorbance was measured before measuring the 420 nm absorbance every 20 seconds for 15 min. Absorbance measurement at 550 nm was repeated immediately after the last measurement at 420 nm.

Units of phosphatase were calculated as described in Appendix B, and analyzed using MATLAB® as described in Appendix C.

CHAPTER 3

Results

Seven strains (presented in table 3.1) were created by introducing mutations to the native genes *sppA* encoding a periplasmic signal peptide peptidase, *skp* encoding a periplasmic chaperone, and *degP* encoding a periplasmic protease. This involved the change of RBS sequences for each gene individually and in all combinations of double mutants and triple mutant in *E. coli* BL21. It was achieved by CRMAGE recombineering with the successfully assembled pMAZ-SK helper plasmid with gRNA insert and MAGE oligos constructed for all three target genes.

All constructed strains were cured for the CRMAGE helper plasmids pMA7CR_2.0 and pMAZ-SK, but the last CRMAGE plasmid pZS4Int-tetR, was not successfully cured. A protocol for analyzing the periplasmic recombinant expression efficiency by osmotic shock and alkaline phosphatase assay was tested and established in the lab. The constructed strains and the negative control (BL21) all harbouring the pZS4Int-tetR plasmid were transformed with the reporter plasmid, and the reporter protein was expressed and analyzed by colorimetric analysis.

In this chapter, the results from these experiments are presented. Further discussions and recommendations for further work are addressed in chapter 4.

Table 3.1: Constructed strains of *E. coli* BL21.

Strain	**Abbreviation**	**Target gene**
BL21_sppA	A	*sppA*
BL21_skp	s	*skp*
BL21_degP	P	*degP*
BL21_sppA+degP	A+P	*sppA* and *degP*
BL21_sppA+skp	A+s	*sppA* and *skp*
BL21_skp+degP	s+P	*skp* and *degP*
BL21_sppA+skp+degP	A+s+P	*sppA*, *skp* and *degP*

3.1 Preparation of the CRMAGE system

Preparation of the CRMAGE system included the *in silico* design of RBS sequences, gRNA sequences and MAGE oligos for each of the native genes targeted for regulation. The RBS sequences and start codons were designed for each native gene to give the desired up-/down regulation of the translation initiation rates (TIR).

The TIR of the native RBS sequences were calculated using Sali's RBS calculator for forward engineeering [49] and are presented in table 3.2 together with the TIRs of the genes with the constructed RBS sequences and altered start codons constructed by Sali's RBS calculator for reverse engineering [49]. The resulting up-/down-regulation was calculated by dividing the TIR of the constructed RBS by the TIR of the native RBS.

Table 3.2: Theoretical translation initiation rates (TIR) of the native and constructed RBS sequences, calculated by use of Sali's RBS calculator [49]. Underlined nucleotides are the start codons of *skp* and *degP* that were changed together with the respective RBS sequences.

Target gene	**Up-/down-regulation**	**TIR**	**RBS**
sppA	Native	2213.08	gacaggtgtgaccttaagttgggagaatac
sppA	$\times 9$	19838.22	caaacgtacccttaattatacctaacgaggagaaact
skp	Native	7363.31	tgcaaatgggatggtaaggagtttatt gtg
skp	$\times 11$	80140.81	cagtatatacgtctaacacttaggggaatatt atg
degP	Native	2139.09	attttgcgttatctgttaatcgagactgaaatac atg
degP	1/19	112.69	gagtaacgccactgatcgaaattgaggaa gtg

The constructed RBS sequences were used to create the MAGE oligos shown in Appendix D. These oligos were constructed so that they target the lagging strand during replication. The homology regions are 90 bp, and are complementary to the flanking regions of the inserted RBS sequence. The *sppA* mutation involves the up-regulation by changing the RBS sequence and a 25 bp deletion in the non-coding 5' region, both mediated by the same MAGE oligo. *skp* is up-regulated by the change in RBS sequence and the change of start codon (GTG $\rightarrow$ ATG), here referred to as the *skp* mutation. Down-regulation of *degP* is achieved by the change of the RBS sequence and the change of start codon (ATG $\rightarrow$ GTG), here referred to as the *degP* mutation.

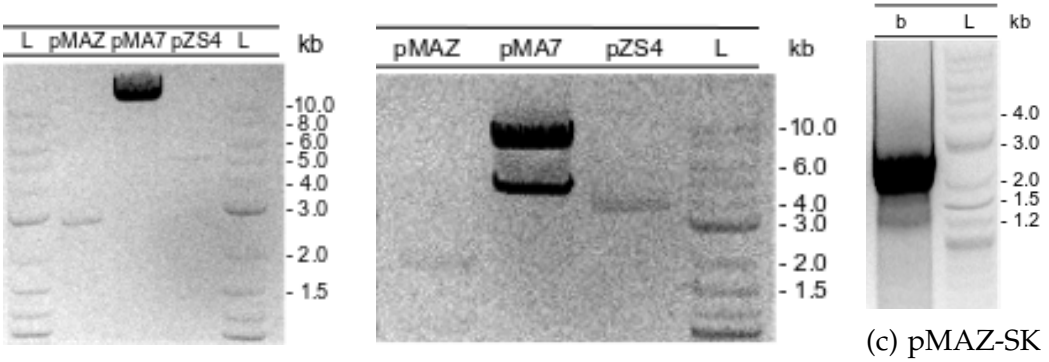
The gRNA oligos are presented in table 3.3. Single stranded gRNA sequences targeting the native RBS sequences of each of the native genes targeted for regulation (*sppA*, *skp* and *degP*) were created using Benchling. The off-target score was calculated to be 95 for the *sppA* gRNA and 50 for both the *skp* and the *degP* gRNAs. The off-target score ranges from 0 to 100, and gives an indication of the probability for incorrect binding of the Cas9 protein to non-targeted DNA in the genome, where

higher score is better.

Table 3.3: gRNA oligos that were cloned into pMAZ-SK (only the forward sequences are shown). The 20 nucleotide gRNAs are shown in capital letters. The lowercase nucleotides are the overlaps created for Gibson assembly, and are complementary to the pMAZ-SK backbone. The underlined bases of the sppA gRNA is not matching the nucleotide sequence of the native RBS.

Name	**Sequence (5′ → 3′)**	**PAM (-NGG)**
sppA_gRNA_F	gtgatagagatactgacacTAATCTATTG-CGCCTGTGACgttttagagctagaaatagc	AGG
skp_gRNA_F	gtgatagagatactgacacCCGGTGCAAA-TGGGATGGTAgttttagagctagaaatagc	AGG
degP_gRNA_F	gtgatagagatactgacacCACATTAGCA-CTGAGTGCACgttttagagctagaaatagc	TGG

The *in vitro* preparation for CRMAGE included verifying the CRMAGE helper plasmids (pMAZ-SK, pMA7CR_2.0 and pZS4Int-tetR) by restriction digestion and sequencing, and assembly of the constructed gRNA to the pMAZ-SK plasmid backbone. pMAZ-SK, pMA7CR_2.0 and pZS4Int-tetR plasmids were successfully extracted from DH5α cells. The plasmids were verified by restriction digestion with restriction enzymes creating single cuts (figure 3.1a) and multiple cuts (figure 3.1b). The DNA fragments were separated and visualized by gel electrophoresis.



(a) Single cut (b) Multiple cuts (c) pMAZ-SK backbone

Figure 3.1: Gel electrophoresis of (a),(b): digested plasmids pMAZ-SK (pMAZ), pMA7CR_2.0 (pMA7) and pZS4Int-tetR (pZS4). (c): pMAZ-SK backbone amplified by PCR. The expected fragment size was 2759 bp. L: ladder.

Expected sizes for digestion with the respective enzymes are shown in table 3.4 together with the fragment size estimated from the gel electrophoresis. All estimated fragment sizes are in agreement with the expected fragment sizes ($\pm$0.5 kb), except

the fragments of NdeI-restricted pMA7CR_2.0. This may be caused by excessive DNA loaded on the gel. Fragments smaller than 1 kb were not visible on the gel.

Table 3.4: Fragments of plasmids pMAZ-SK, pMA7CR_2.0 and pZS4Int-tetR after restriction digestion with specified enzymes. Single: single cut plasmids; Multi: Multiple cuts.

	Plasmid	Restriction enzyme	Expected fragment size(s) (bp)	Visible fragment(s) on gel (kb)
Single	pMAZ-SK	SexAI	2779	~ 3
	pMA7CR_2.0	DraIII-HF	10934	> 10
	pZS4Int-tetR	EcoRI	6000	~ 5.5
Multi	pMAZ-SK	NheI	1783, 973, 23	~ 2, ~ 1
	pMA7CR_2.0	NheI	7132, 3802	~ 8, ~ 5
	pZS4Int-tetR	NdeI	4492, 741, 581, 186	~ 4

Transformation of pMA7CR_2.0 and pZS4Int-tetR into *E. coli* BL21 was performed and verified by growth on selective antibiotic LA medium. The pMAZ-SK backbone was successfully amplified and run on a gel for purification and further gel extraction, and is shown in figure 3.1c. The amplified DNA fragments were within the correct size range (2-3 kb) (expected size: 2759 bp). The gRNA oligos provided by Sigma-Aldrich (table 3.3) were successfully annealed and assembled with the pMAZ-SK backbone, creating the three plasmids; pMAZ-SK_sppA, pMAZ-SK_skp and pMAZ-SK_degP. The plasmids were transformed into *E. coli* DH5α for plasmis amplification, and the transformation was verified by growth on kanamycin supplemented media. The plasmids were harvested, and correct inserts of the gRNA was confirmed with sequencing (see Appendix E).

3.2 Verifying the construction of seven mutant strains of *E. coli* BL21

New RBS sequences were introduced to *sppA*, *skp* and *degP* of BL21 using CRMAGE. Seven strains were successfully constructed: three single mutants, three double mutants and one triple mutant. Colony PCR with a forward primer specific to the inserted sequence and reverse primer downstream of the gene was used to verify the inserted sequence (primer sequences are shown in Appendix D, table D.1). An additional colony PCR with primers flanking the region upstream and downstream of the inserted sequence (sequencing primers) was performed, and the amplified DNA was sent for sequencing to verify the sequence of the insert and surrounding DNA.

3.2.1 Verifying construction of the *E. coli* BL21 single mutant strains

Three single mutant strains were created using the wildtype BL21 strain as a basis, and running CRMAGE with the constructed pMAZ-SK plasmid and MAGE oligo for each of the three native gene targets. Verification of introduced mutations was done by colony PCR.

Verifying construction of the *E. coli* BL21_sppA mutant strain. 18 colonies were screened by colony PCR for the insert corresponding to the new RBS of the *sppA* gene, and the result from the gel electrophoresis of the PCR products is shown in figure 3.2. Note that colony number 17 was not included, and number 18 is excluded due to a technical mistake. All other colonies screened gave PCR products of the correct size range. It appears that there are, however, two different kinds of products: 1, 3, 4, 7, 9, 13, 15, 16 and 19 appears more saturated and shorter than the PCR products from colony number 2, 5, 6, 8, 10, 11, 12, 14 and 20.

Amplification of the genomic fragments covering the mutated region of two colonies from each type of product (number 1,2,3 and 8) was conducted by colony PCR in two parallels with sequencing primers for *sppA*. Three colonies (number 1,3 and 8) gave a detectable PCR product, were sent for sequencing, and three out of three colonies showed the correct insert.

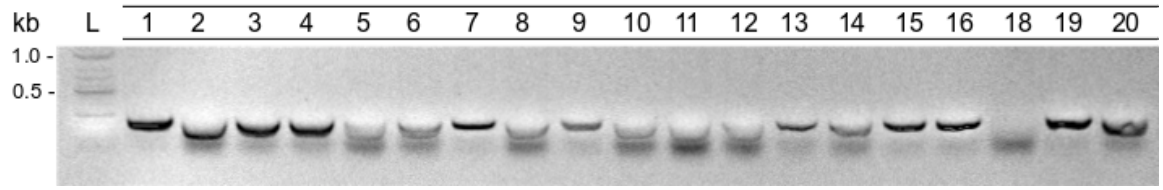


Figure 3.2: DNA fragments from colony PCR of BL21_sppA separated by gel electrophoresis. One of the primers used for colony PCR was specific to the inserted RBS sequence, and a product of 431 nt was expected only if the cells carried the correct insert. Colony nr. 17 and 18 excluded due to technical mistake.

Verifying construction of the *E. coli* BL21_skp mutant strain. In the first round of CRMAGE to create the BL21_skp strain, the colonies obtained were small, and phenotypically different from BL21. 10 colonies were screened, but no PCR product with the correct size was obtained. The cells were not viable when re-plated on a new plate with the same antibiotics. A new round of CRMAGE was run, and the resulting colonies were similar to the ones of BL21. 10 of the new colonies were screened by colony PCR for the insert corresponding to the new RBS of the *skp* gene. The result from the gel electrophoresis of the PCR products is shown in figure 3.3. Seven out of ten colonies (number 2, 4, 5, 6, 8, 9 and 10) screened gave PCR products of the correct size range, also here with a some variation in intensity and size.

Colony PCR with sequencing primers was performed in two parallels of four colonies (number 2, 4, 5 and 6), and colony number 4 and 5 gave detectable PCR products.

Sequencing of the amplified DNA verified correct DNA sequences from both colony 4 and 5, except one potential off-target substitution (G $\rightarrow$ A) outside the MAGE oligo region. To verify whether this mutation was introduced by the recombineering event, the *skp* region of wildtype BL21 was sequenced. The results revealed that the mutation occurs in the wildtype strain and was not introduced by CRMAGE.

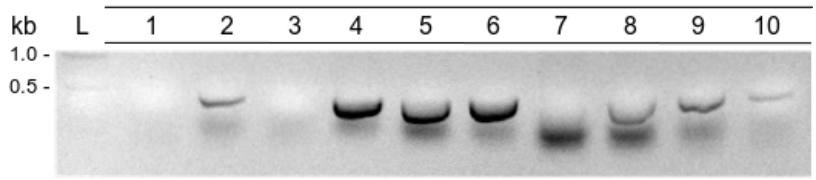


Figure 3.3: DNA fragments from colony PCR of BL21_skp separated by gel electrophoresis. One of the primers used for colony PCR was specific to the inserted RBS sequence, and a product of 458 nt was expected only if the cells carried the correct insert.

Verifying construction of the *E. coli* BL21_degP mutant strain. 10 colonies were screened by colony PCR for the insert corresponding to the new RBS of the *degP* gene, and the result from the gel electrophoresis of the PCR products is shown in figure 3.4. Three out of ten colonies (number 3, 8 and 10) gave PCR products within the correct size range, and some faint bands appeared on the electrophoresis gel for colonies 5, 7 and 9. Four colonies (number 3, 8, 9 and 10) were amplified by colony PCR in two parallels with sequencing primers for *degP*, and two colonies (number 3 and 8) gave a detectable PCR product. Sequencing of the amplified DNA verified the correct sequence from both colonies.

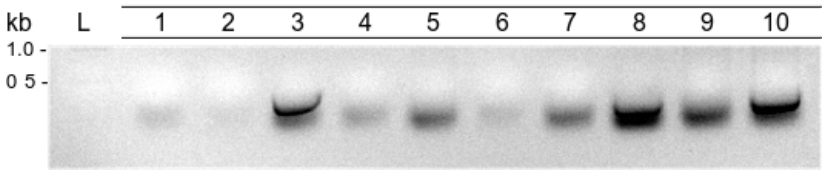


Figure 3.4: DNA fragments from colony PCR of BL21_degP separated by gel electrophoresis. One of the primers used for colony PCR was specific to the inserted RBS sequence, and a product of 391 nt was expected only if the cells carried the correct insert.

3.2.2 Verifying construction of the *E. coli* BL21 double- and triple mutant strains

Double mutants were constructed by curing one of the single mutants from the pMAZ-SK plasmid from the previous round of CRMAGE, adding an additional mutation with the single mutant as basis. BL21_sppA+degP and BL21_sppA+skp were constructed by mutating the *degP* and the *skp* genes respectively by CRMAGE

using BL21_sppA as a basis. BL21_degP+skp was attempted constructed by mutating *skp* using BL21_degP as a basis, but without success. However, BL21_skp+degP was successfully constructed by introducing the degP mutation to BL21_skp. The triple mutant, BL21_sppA+skp+degP was created by introducing the degP mutation to BL21_sppA+skp. In total, four double- and triple mutants were constructed and verified by colony PCR and sequencing.

Verifying construction of the *E. coli* BL21_sppA+degP double mutant strain. Seven colonies were obtained from this round of CRMAGE, and the result from the gel electrophoresis of the colony PCR products is shown in figure 3.5. Some of the colonies were run in two parallels. Three out of seven colonies (number 2, 3 and 4) gave PCR products within the expected size range (391 nt). All three colonies were amplified by colony PCR in two parallels with sequencing primers for *degP*, and one (number 4) was verified by sequencing and revealed the desired *degP* mutation.

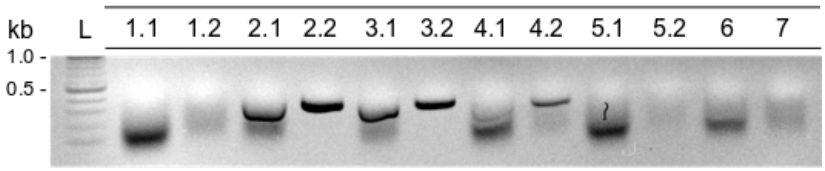


Figure 3.5: DNA fragments from colony PCR of BL21_sppA+degP separated by gel electrophoresis. One of the primers used for colony PCR was specific to the inserted RBS sequence for *degP*, and a product of 391 nt was expected only if the cells carried the correct insert.

Verifying construction of the *E. coli* BL21_sppA+skp double mutant strain. 16 colonies were screened by colony PCR for the insert corresponding to the new RBS of the *skp* gene, and the result from the gel electrophoresis of the PCR products is shown in figure 3.6. Five out of 16 colonies (number 8, 12, 14, 15 and 16) gave PCR products within the expected size range (458 nt). Colony PCR with sequencing primers was performed in two parallels on four colonies (number 8, 12, 14 and 15). Sequencing of the amplified DNA verified correct DNA sequence with the *skp* mutation for colony number 8.

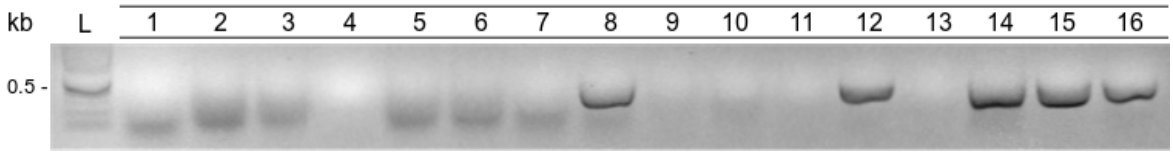


Figure 3.6: DNA fragments from colony PCR of BL21_sppA+skp separated by gel electrophoresis. One of the primers used for colony PCR was specific to the inserted RBS sequence, and a product of 458 nt was expected only if the cells carried the correct insert.

Verifying construction of the *E. coli* BL21_skp+degP double mutant strain. The BL21_degP+skp mutant was first attempted to be constructed by mutating *skp* in

the BL21_degP mutant by CRMAGE. 15 colonies were screened by colony PCR with primer specific to the insert in *skp*, and the result from the gel electrophoresis of the PCR products is shown in figure 3.7 (upper figure). Only faint bands were detected for three of the colonies (number 1, 2 and 3), but colony PCR was still performed with sequencing primers in two parallels on four colonies (number 1, 2, 3 and 11). PCR products from colony number 1, 2 and 11 were sent for sequencing, but did not show the correct sequences.

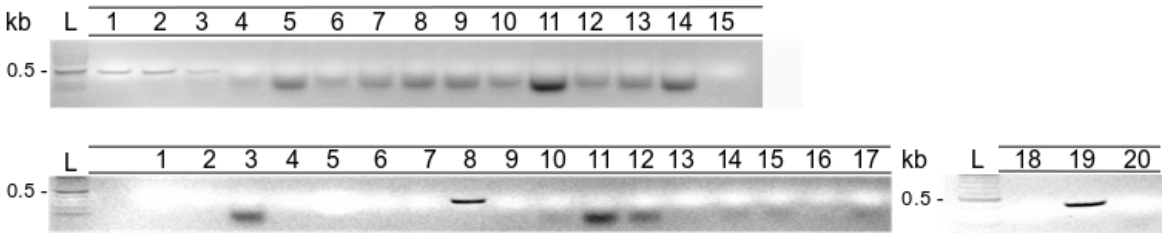


Figure 3.7: Colony PCR of BL21_skp+degP with primers specific to the inserted RBS sequence. The upper figure shows the gel from colony PCR with primers specific to a potential *skp* mutation after CRMAGE with BL21_degP as a basis. A product of 458 nt was expected only if the cells carried the correct *skp* insert. For the lower figure, primers specific to the inserted sequence of *degP* were used after a new round of CRMAGE with BL21_skp as a basis. A product of 391 nt was expected only if the cells carried the correct *degP* insert.

The DNA of an additional five colonies were later attempted amplified by colony PCR using sequencing primers (result from gel electrophoresis not shown). Only two of them gave PCR products and were sent for sequencing. One showed the non-mutated sequence, while the other had the correct insert, but with two off-target mutations. The CRMAGE protocol was therefore repeated, and screening by colony PCR was performed for 17 colonies, but no amplified DNA could be detected on the gel. Also, these cells were not viable when re-plating them, similar to the colonies obtained from the first round of CRMAGE for the *skp* mutant.

At this point, the only unsuccessful rounds of CRMAGE had been when attempting to introduce the *skp* mutation. When repeating the attempt of creating this double mutant, it was therefore decided that the BL21_skp mutant should be used as a basis, and the *degP* mutation should be introduced to create the BL21_skp+degP mutant. 20 colonies were screened by colony PCR for the insert corresponding to the new RBS of the *degP* gene, and the result from the gel electrophoresis of the PCR products is shown in figure 3.7 (lower figure). Two out of 20 colonies (number 8 and 19) gave PCR products within the expected size range (391 nt). Colony PCR with sequencing primers was performed in two parallels on both colonies, and the correct DNA sequences with the *degP* mutation was detected in both colonies by sequencing.

Verifying construction of the *E. coli* BL21_sppA+skp+degP triple mutant strain. 20 colonies were screened by colony PCR for the insert corresponding to the new

RBS of the *degP* gene, and the result from the gel electrophoresis of the PCR products is shown in figure 3.8. Five out of 20 colonies (number 4, 6, 17, 18 and 20) gave PCR products within the expected size range (391 nt). Colony PCR with sequencing primers was performed in two parallels on four colonies, and two of them gave detectable PCR products (colony 4 and 6). Both were sent for sequencing, and colony number four showed the correct DNA sequence with the *degP* mutation.

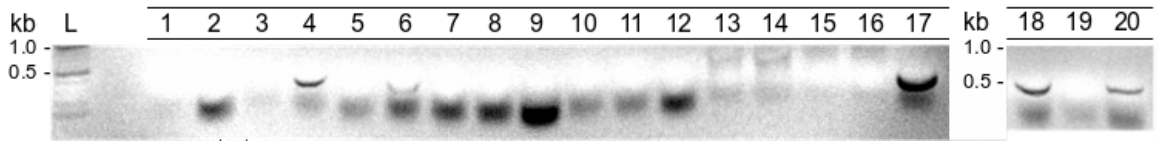


Figure 3.8: DNA fragments from colony PCR of BL21_sppA+skp+degP separated by gel electrophoresis. One of the primers used for colony PCR was specific to the inserted RBS sequence, and a product of 391 nt was expected only if the cells carried the correct insert.

The data from the verification process of the constructed strains of mutated BL21 may be summarized as shown in table 3.5. This data only includes the number of colonies checked from the last, successful round of CRMAGE for each of the constructed mutants. This data was used to calculate the CRMAGE recombineering efficiencies in table 3.6.

Table 3.5: Overview of the verification process of constructed strains using colony PCR and sequencing.

Mutation	**Colonies checked**	**Colonies that gave band**	**Samples sent for sequencing**	**Colonies w. correct insert**
sppA	18	18	3	3
skp	10	7	2	2
degP	10	3	2	2
sppA+degP	7	3	4	1
sppA+skp	16	5	4	1
skp+degP	20	2	2	2
sppA+skp+degP	20	5	2	1

3.2.3 CRMAGE recombineering efficiency

For each of the different mutants, between 7 and 20 colonies were screened by colony PCR, and 2-4 of them were sent for sequencing to confirm the inserted sequence and check for secondary mutations. This data is not sufficient to give a certain value of the recombination efficiency, but an overall estimation may be suggested in order to compare with other reported data of the recombineering efficiency of CRMAGE. The sequencing results may also give an estimation of the reliability of the method of verification by colony PCR screening.

The estimated recombineering efficiencies for each successful round of CRMAGE are shown in table 3.6. The recombineering efficiencies for each of the mutations are given by the percentage of the screened colonies giving rise to a band in the colony PCR. The overall recombineering efficiency for all successful rounds of CRMAGE in this project is 43%. Here, the unsuccessful rounds of CRMAGE giving non-viable colonies are not included in the calculation. In more than 60% of the cases, the method of screening by colony PCR was verified by sequencing.

Table 3.6: Recombineering efficiency for each of the mutations given by the percentage of the screened colonies giving rise to a band in the colony PCR. Values are rounded to the nearest 5 percentage points.

	degP, C			
	0		**1**	
	skp, B		**skp, B**	
sppA, A	**0**	**1**	**0**	**1**
0	-	70%	30%	10%
1	100%	30%	40%	25%

A difference in the recombineering efficiencies was seen between the CRMAGE systems for introducing each of the mutations in *sppA*, *skp* and *degP* as shown in table 3.7. Here, also the rounds of CRMAGE that resulted in non-viable colonies are included in the calculation.

Table 3.7: Recombineering efficiency for each of the mutations given by the percentage of the screened colonies giving rise to a band in the colony PCR. Values are rounded to the nearest 5 percentage points.

Target gene	**Recombineering efficiency**
sppA	100%
skp	20%
degP	20%

3.3 Elimination of CRMAGE plasmids from constructed strains

To be able to analyze the mutated strains of *E. coli* BL21 without having interruption of the results caused by the CRMAGE plasmids, the cells had to be cured of the helper plasmids pMAZ-SK, pMA7CR_2.0 and pZS4Int-tetR. The pMAZ-SK plasmid holds a self-killing function, making the plasmid easily cured by inducing with L-rhamnose. All constructed strains were successfully cured from the pMAZ-SK plasmids by this method.

Curing by growth in elevated temperatures (42 °C) (see section 2.5) was successful for curing pMA7CR_2.0 within 2-3 days, but curing pZS4Int-tetR was not successful by this method. Cells were re-plated every 1 - 2 days for over a month in parallel with other work without success. Acridine orange treatment alone, and acridine orange combined with penicillin G treatment was tested, without success. Curing by electroporation was also tested, but no viable cells were obtained on antibiotic media.

The loss of plasmid was checked by monitoring the ability to grow on selective media for that plasmid. After unsuccessful attempts to cure constructed strains from the pZS4Int-tetR plasmid, a plasmid extraction was performed to confirm whether two of the created strains (BL21_skp and BL21_sppA+degP) that showed antibiotic resistance still harboured the plasmid. A restriction digestion was performed on the plasmid extracts from the two strains and a negative control (BL21), in addition to a positive control (pZS5Int-tetR) using DraIII and NdeI. The products were run on a gel together with the non-digested plasmid extract (figure 3.9). Expected fragment sizes of pZS4Int-tetR after digestion and estimated sizes of DNA fragments from digestion are shown in table 3.8. No DNA fragments were detected for the wildtype BL21 (negative control).

Table 3.8: Expected fragemt sizes and the estimated size of visible fragments after digestion of plasmid extracts from BL21_skp (s), BL21_sppA+degP (AP) and BL21 (neg: negative control) before and after restriction digestion by specified enzymes. The visible fragments were of the same size for both s, AP and the positive control. *Non-digested DNA will have different properties than linear DNA, and travel at different rates through the gel.

Restriction enzyme	**Expected fragment size(s) (bp)**	**Visible fragment(s) on gel (kb)**
DraIII	5380, 620	~ 5.5
NdeI	4492, 741, 581, 186	~ 4, ~ 0.8
No digestion	6000*	~ 3.5*

The non-digested samples showed bands for BL21_skp and BL21_sppA+degP plasmid extracts and the positive control around 3-4 kb. Since the samples were non-digested and therefore circular, the fragments will move at different rates through the gel compared to linearized DNA. In all three cases, the extracts from BL21_skp and for BL21_sppA+degP showed the same bands as the positive control, verifying that the strains tested still harboured the pZS4Int-tetR plasmid.

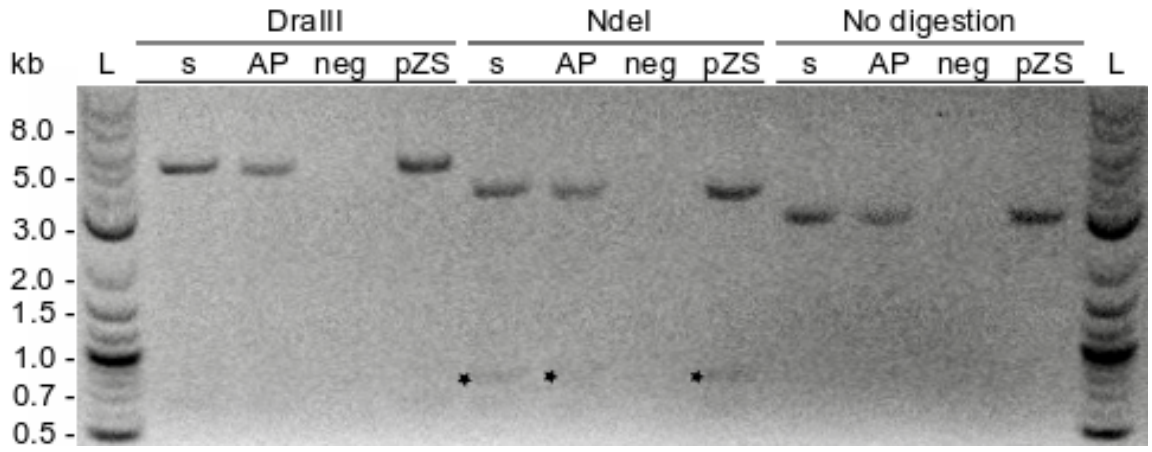


Figure 3.9: Gel electrophoresis after digestion of plasmid extracts from BL21_skp (s), BL21_sppA+degP (AP) and BL21 (neg: negative control) before and after restriction digestion by specified enzymes. The visible fragments were of the same size for both s, AP and the positive control, showing that the attempted cured strains still harboured the pZS4Int-tetR plasmid. Faint bands are marked with stars.

3.4 Evaluation of periplasmic recombinant protein production by alkaline phosphatase assay

The seven constructed, verified strains were successfully transformed with the reporter plasmid pTA16-s173-AP. The reporter system consisting of a single-chain antibody fused to alkaline phosphatase (AP) was expressed (5 h, 30 °C), and the periplasmic fraction was analyzed by AP assay as a method of comparing the periplasmic expression capacities of the constructed strains and the BL21 strain. Due to lack of success in pZS4Int-tetR curing, all strains were carrying this plasmid in addition to the reporter plasmid, and two antibiotic markers were used during expression. Units of phosphatase were calculated as described in Appendix B.

The optical density (OD) is proportional to the amount of product generated by the AP enzyme (see Appendix B). The OD of the reaction mixture was measured at 420 nm (OD_{420nm}) every 20 seconds for 15 minutes. Kinetic analysis of absorbance vs. time for each mutant strain and wildtype BL21, all harbouring the reporter plasmid pTA16-s173-AP and pZS4Int-tetR are shown in figure 3.10. Error bars are the standard deviations of the means (SDOM) calculated from the sample triplicates. The analysis shows that the OD_{420nm} is increasing linearly with time within the analyzed time frame. The last sample point (17 min) was used for further calculations of units of phosphatase. A linear regression analysis was conducted on the continuous measurement data for each strain. The calculated OD_{420nm} at t=0 was used as a blank for further calculations. Data from the linear regression can be found in Appendix B, table B.1.

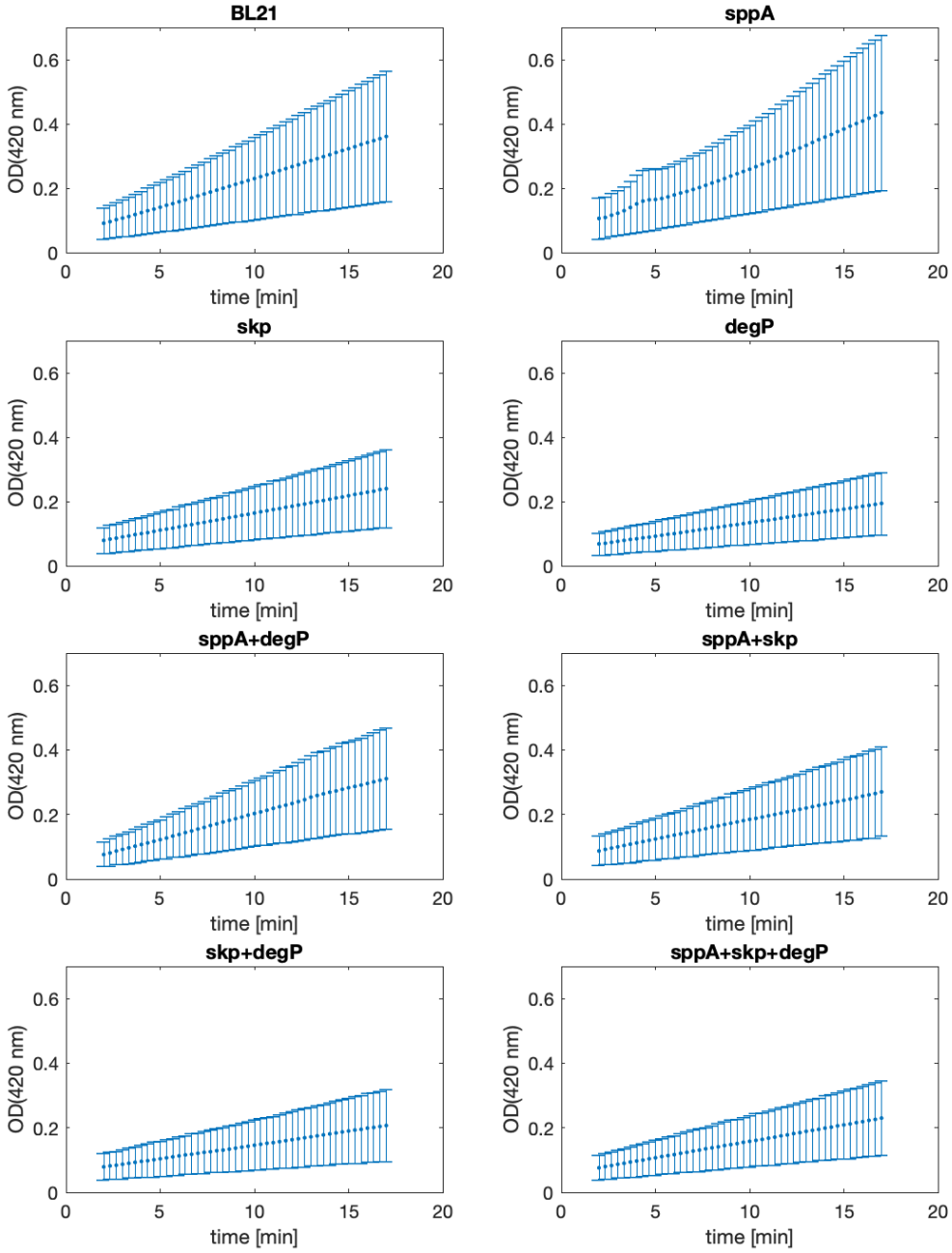


Figure 3.10: Continuous measurements of the OD_{420nm} in periplasmic samples added ρ-nitrophenyl shows that the absorbance is increasing linearly with time within the studied time frame. Error bars shows the standard deviation of the mean (SDOM).

Units of phosphatase for each constructed strain and the wildtype BL21 (wt) are

shown in figure 3.11. Units of phosphatase activity measured for DH5α without the reporter system was included as a negative control. As mentioned in section 1.5, wildtype *E. coli* produce some level of AP, explaining the low level of AP activity for the negative control. Error bars show the SDOM calculated from technical triplicates of each strain (see Appendix C).

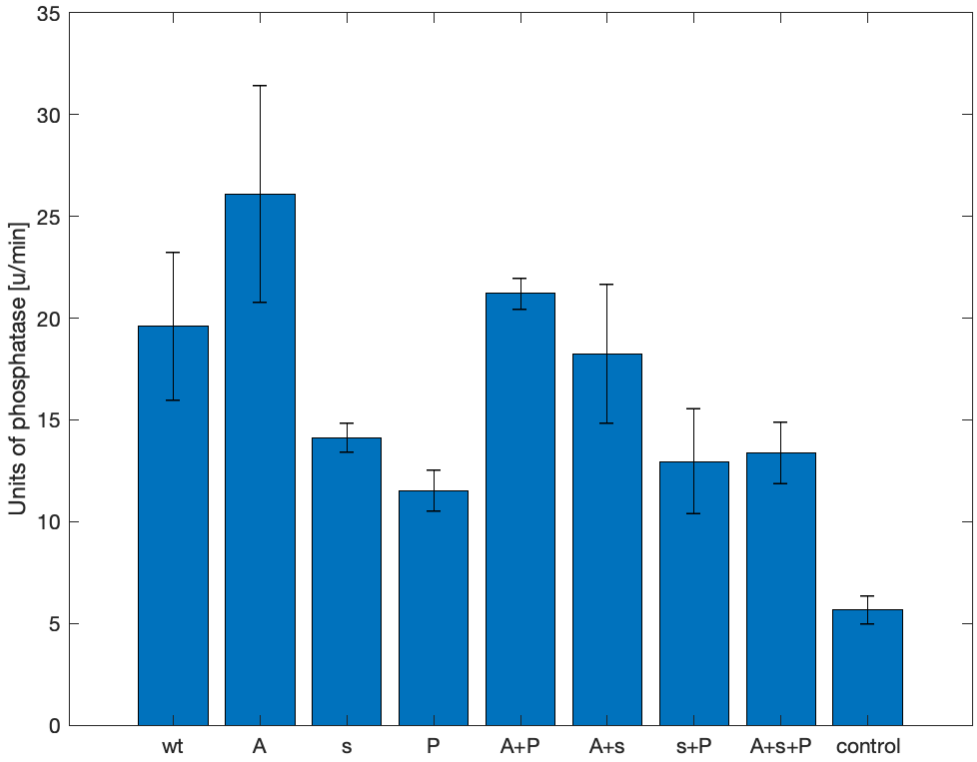


Figure 3.11: Units of phosphatase [u/min] for each strain of *E. coli* BL21 carrying the pTA16-s173-AP reporter system. wt: BL21;pTA16-s173-AP ; A: *sppA*; s: *skp*; P: *degP*; control: DH5α. Production of the reporter protein fusion with alkaline phosphatase was induced by 2 mM m-Toluate at OD=0.62 ±0.05 and expressed for 5h. The control was expressed overnight, 16 °C.

For the purposes of this project, it is of interest to know whether the individual mutations introduced, or the combination of mutations, has an effect on periplasmic protein expression. In order to test this hypothesis, a full factorial analysis with three replicates was conducted as presented in figure 3.9, where zero represents the non-mutated RBS sequence, and one represents the mutated strain.

An analysis of variance (ANOVA, see Appendix C) was used to test the zero hypothesis that each of the main effects are equal to zero against the hypothesis that at least one of them are different [87], resulting in 7 zero-hypotheses and alternative hypotheses included in Appendix C.

Table 3.9: With three factors being the three different mutations of *sppA*, *skp* and *degP*, and three replicates, $2^3 \cdot 3 = 24$ samples are needed for a full factorial analysis with three replicates. 0 will in this case represent the non-mutated negative control.

	degP, P			
	0		**1**	
	skp, s		**skp, s**	
sppA, A	**0**	**1**	**0**	**1**
0	0	s	P	s+P
	0	s	P	s+P
	0	s	P	s+P
1	A	A+s	A+P	A+s+P
	A	A+s	A+P	A+s+P
	A	A+s	A+P	A+s+P

MATLAB was used for the ANOVA analysis (see Appendix C) of the units of phosphatase of the periplasmic extracts of the constructed strains. The results from the ANOVA analysis are shown in table 3.10. The Prob>F column contains the p-value for each effect. The low p-values for the single mutant effects (BL21_sppA, BL21_skp and BL21_degP), indicate that the zero-hypotheses, that the effect of the mutated version is the same as the non-mutated version, may be rejected. The interaction effect sppA*skp has a p-value of 0.02, showing that there is a significant interaction effect. There is no significant interaction between the *sppA* and *degP* mutants, neither between the *skp* and *degP* or for all three *sppA*, *skp* and *degP* mutants combined.

Table 3.10: Analysis of variance for the calculated units of phosphatase. Sum Sq. - sum of squares; d.f: degrees of freedom; Prob>F: p-value.

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
sppA	161.258	1	161.258	20.11	0.0004
skp	146.113	1	146.113	18.22	0.0006
degP	135.966	1	135.966	16.95	0.0008
sppA*skp	50.583	1	50.583	6.31	0.0231
sppA*degP	0.085	1	0.085	0.01	0.9195
skp*degP	18.082	1	18.082	2.25	0.1527
sppA*skp*degP	17.657	1	17.657	2.2	0.1573
Error	128.322	16	8.02		
Total	658.064	23			

Constrained (Type III) sums of squares.

An interaction plot was constructed using MATLAB (see Appendix C) and is shown in figure 3.12. The lines of the interaction plot between *skp* and *degP* are approximately parallel, while the interaction plot of *sppA* and *skp* have nonparallel lines, indicating an interaction effect. According to the results from the ANOVA, only the

interaction effect between *sppA* and *skp* is significant with a p-value of 0.02, consistent with the tendencies of the interaction plot. The steep slope in the line representing the wildtype *skp* (*skp*=0) in the middle left of figure 3.12 reflects the effect of the substituted RBS secuence of *sppA*. The line representing the mutated version of *skp* (*skp*=1) however, shows negligible change in protein expression with and without the mutation of *sppA*. For the upper middle figure, a steep slope is seen in the line representing the mutated variant of *sppA*, showing a decrease in expression level from the non-mutated *skp* to the mutated *skp*. When *sppA* is not mutated, there is a negligible difference between the mutated and wildtype *skp* variant.

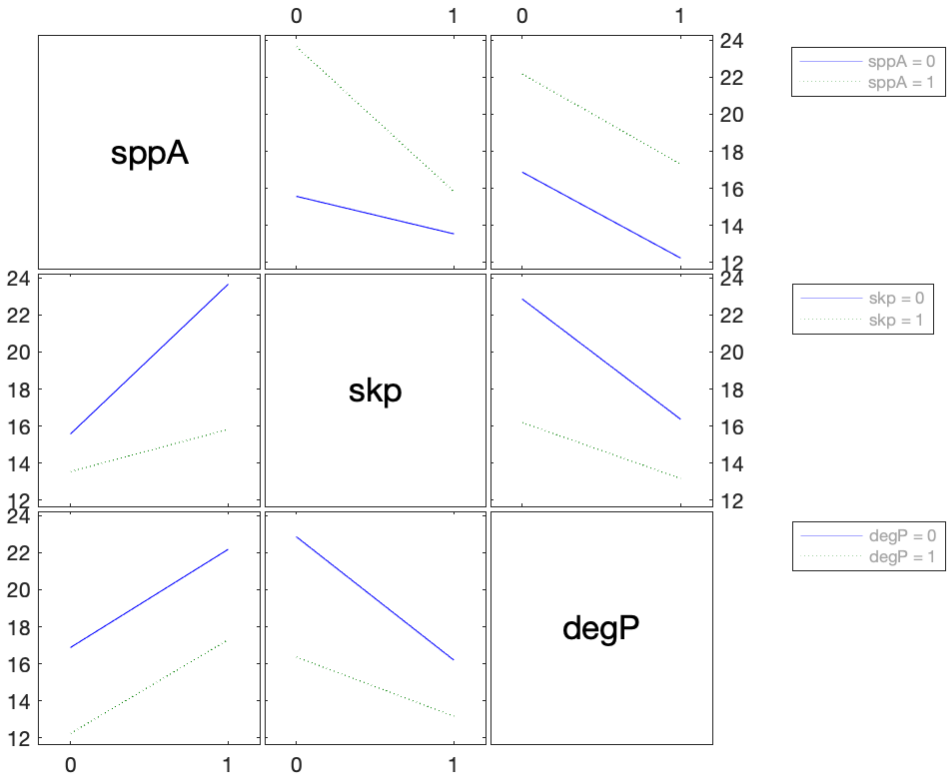


Figure 3.12: Interaction plot. 0 signifies the wildtype variant of the RBS sequence, 1 signifies the changed variant. Parallel lines indicate no interaction, such as for *sppA* and *degP*, while nonparallel lines indicates an interaction effect.

Since there is a significant interaction effect between *sppA* and *skp*, masking effects should be considered in the interpretation of the test results. However, since all main effects are significant, the results are still valid (see C). By ANOVA it has been indicated that the *sppA*-mutation has an effect on the AP activity. However, this analysis also depends on the responses of the double- and triple mutants. Additionally, the aim is to see if any of the mutants have a higher expression rate compared to BL21 since this is the current strain of choice for protein production.

The multcompare function in MATLAB for Tukey's honestly significant difference criterion (Tukey's hsd, see Appendix C) was used to analyze whether there is a significant difference between any of the strains protein expression levels compared to BL21. The result is shown in figure 3.13. Overlapping lines means that there is not a significant difference. The blue line represent BL21, and the only strain that does not overlap with BL21 is the *degP* mutated variant that has a significantly lower expression rate of the reporter protein compared to the wildtype BL21. Bl21_sppA which was the only strain with a higher value of units of phosphatase compared to the wildtype BL21 does not have a significant difference according to the multiple comparison test.

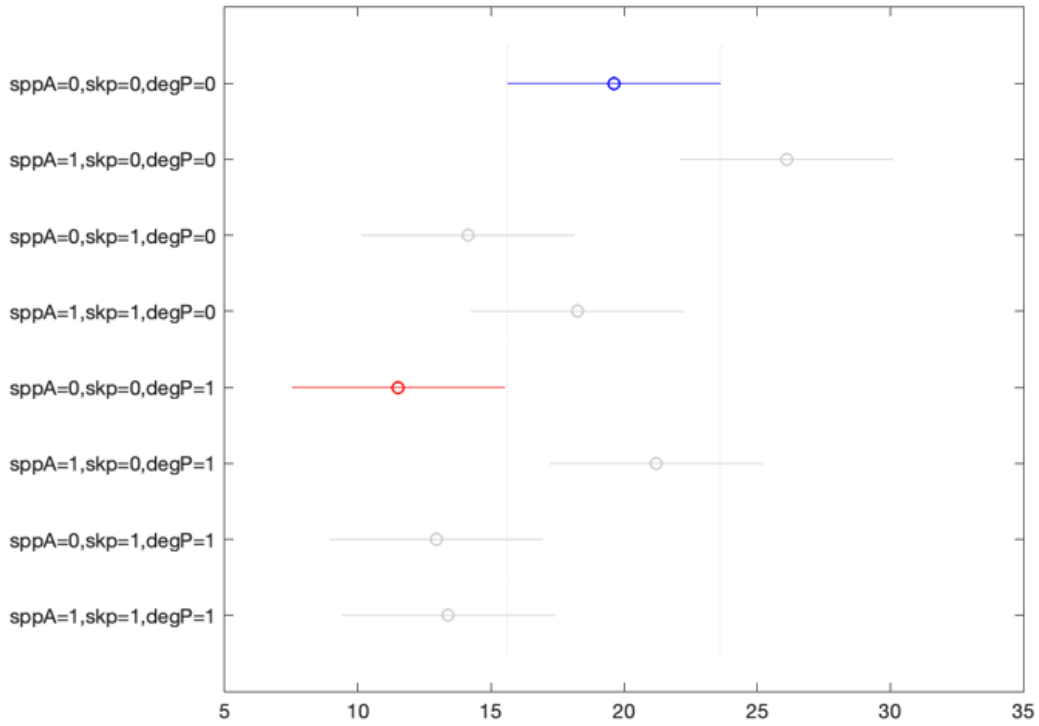


Figure 3.13: Multiple comparison for all mutant strains and BL21 (blue) by the Tukey Kramer method. Overlapping lines indicates not significant difference. 0=native, 1=mutated. BL21_degP is the only constructed strain that is significantly different from the wildtype BL21.

CHAPTER 4

Discussion

All seven strains constructed were successfully created using CRMAGE recombineering. The BL21_sppA strain with a higher translation initiation rate (TIR) of the periplasmic signal peptide peptidase SppA, showed an increase in alkaline phosphatase activity of the periplasmic extract compared to BL21, but not sufficiently high to state a significance in the difference. Lowering the TIR of the periplasmic protease DegP in BL21_degP gave significantly lower activity of the periplasmic protein extract compared to wildtype BL21. Increased periplasmic chaperone activity by the increased TIR of Skp in BL21_skp did not give a significant difference in alkaline phosphatase activity, but had a negative effect on *sppA*-mutated strains, illustrated by the significant interaction effect between the *sppA* mutation and the *skp* mutation. The other double- and triple mutants constructed did not show a significant difference in periplasmic expression reported by the alkaline phosphatase activity of the periplasmic protein extract. All constructed strains were cured for the CRMAGE helper plasmids pMA7CR_2.0 and pMAZ-SK, but the last CRMAGE plasmid pZS4Int-tetR, was not successfully cured. Here, these main results are discussed, and recommendations for further work are addressed.

4.1 CRMAGE recombineering showed a high efficiency of introducing the *sppA* mutation

The introduction of the sppA mutation showed a recombineering efficiency of up to 100%. A difference in the CRMAGE recombineering efficiency was seen for each of the rounds of CRMAGE, but an overall recombineering efficiency for all successful rounds of CRMAGE of 43% was obtained. Some attempts of CRMAGE for introducing the *skp* mutation resulted in non-viable colonies. The sequencing results showed a risk of false-positive results when verifying the correct insert by colony PCR.

The recombineering efficiencies vary between 10% for BL21_skp+degP and up to

100% for BL21_sppA. In comparison, Ronda et al. reported a recombineering efficiency for three different single nucleotide substitutions between 95.5% and 99.7%. For larger substitutions of six nucleotides, a recombineering efficiency of 62% was obtained by Ronda et al. [47]. In this project, the inserted sequences were larger than in [47], and the recombineering efficiency was therefore expected to be lower than for the six nucleotide substitution.

In section 1.4 of the introduction, it was explained that the efficiency of MAGE recombineering depends on the number of changed nucleotides [57]. Surprisingly, the highest recombineering efficiency was obtained when introducing the *sppA* mutation even though the sppA MAGE oligo was designed to delete 25 bp in addition to the change of RBS sequence. This illustrates the potential of multiple modifications of a DNA locus using the CRMAGE method. Several differences between the CRMAGE systems for introducing the mutations of *sppA* and *skp* or *degP* may explain the differences in recombineering efficiency.

As mentioned in section 3.2.2, there was a suspicion that the CRMAGE system for introducing the *skp* mutation might create off-target effects causing non-viable cells. The off-target score [79] for the constructed *skp* gRNA was 50, meaning that there is a probability that the Cas9 nuclease can create DSBs at other sequences than the native RBS that it is targeted to. Double-stranded breaks have a significant fitness cost, and may result in cell death [47].

A recombineering efficiency of 20% was seen for introducing the *degP* mutation. There is no PAM sequence within the native RBS sequence of *degP*, and a PAM sequence outside the insert had to be targeted when constructing the gRNA. The MAGE oligo was designed to interrupt this PAM sequence (TGG $\rightarrow$ TAG) by mismatch recombination at the same time as the change of RBS sequence so that the mutated DNA should not be targeted by Cas9 nuclease. However, Hsu et al. showed that also NAG PAM sequences may be targeted by Cas9, but with one fifth of the frequency of the NGG PAM sequence [79]. This leads to a low rate of nuclease activity also targeting the mutated sequence, and might explain the lower recombineering efficiency compared to that of the introduction of the *sppA* mutation. Additionally, the off-target score calculated for the constructed *degP* gRNA was also 50, like for the *skp* gRNA, giving a risk of off-target effects.

In addition to the variations in the constructed CRMAGE systems, the electroporation step in itself may play a role in the recombineering efficiency. Both the pMAZ-SK plasmid and the MAGE oligo must be electroporated in order to give a successful recombination. A higher number of cells transformed gives a higher probability of obtaining a correct mutant. Unsuccessful transformation would lead to non-viable colonies because the cells would not harbour the antibiotic resistance gene. However, this does not explain why only the introduction of the *skp* mutation led to non-viable cells.

In more than 60% of the cases, the method of screening by colony PCR was verified by sequencing. This indicates that there is a limited reliability of the colony PCR-

primers because they sometimes give false positive results. However, in 7 of the cases, the sequencing results were too short, and the result may have come from poor quality of the sample or other technical mistakes.

4.2 Alternative methods for elimination of CRMAGE helper plasmids from constructed strains

One of the most important advantages of regulating gene expression of native genes by genome editing rather than plasmid co-expression is that no antibiotic selection markers are needed. The elimination of the three CRMAGE helper plasmids (pMAZ-SK, pMA7CR_2.0 and pZS4Int-tetR) after recombineering was therefore desired in order to escape the need of using three antibiotic selection markers. The curing of pMAZ-SK by the self-killing function gave a high success-rate, and was achieved within two days. In previous work done at the lab [81], *E. coli* MG1655 were cured from pMA7CR_2.0 and pZS4Int-TetR by repeated growth at 42 °C. In the current project, this method was only effective for the elimination of the pMA7CR_2.0 plasmid, and different methods had to be tested for the curing of pZS4Int-tetR.

Traditional methods for plasmid curing, other than growth at high temperatures, include the use of interchalating agents such as ethidium bromide or acridine orange. Ethidium bromide has been extensively used for plasmid curing, and curing is usually seen at a high frequency [88], but was not tested due to toxicity and mutagenic properties. Curing by acridine orange treatment alone and in combination with penicillin G was tested at two different concentrations without success. Also electroporation has been reported to eliminate plasmids [82], but was not successful for elimination of the pZS4Int-tetR plasmid.

The possibility of curing the constructed strains at a higher temperature, for instance at 45 °C, could be investigated in the future. The *degP* mutated strains may show decreased growth at higher temperatures because DegP is essential for high-temperature growth (45 °C) [89]. However, no such indications were seen when growing the degP mutated strains at 42 °C, and the low amount of expressed DegP expressed in the *degP*-mutated strains may be sufficient for growth at high temperatures.

An alternative is to implement a similar method as the self-killing function of pMAZ-SK to the pZS4Int-tetR plasmid. Lauritsen et al. developed a plasmid (pFREE) for the purpose of curing common plasmid vectors, including pZS vectors. This method is also based on the Cas9 nuclease, and targets the ColE1 and pSC101 replicons. Like the pMAZ-SK, pFREE also has a self-killing function [90].

4.3 The potential of regulating the expression of Skp, DegP and SppA should be further inverstigated

For this thesis, one new RBS sequence was constructed for the regulation of each gene. The optimal approach would be to generate and test more RBS sequences with different TIR values estimated by Salis RBS design tool [49]. When only one expression level is tested, it gives an indication to whether this up-/down-regulation is beneficial. However, small changes are detected in this analysis, and it might be interpreted in two different ways: the insignificant changes compared to the wild-type BL21 might mean that the up-/down-regulations of these genes do not have a beneficial effect. Or, the slight increases/decreases in the expression levels might indicate a trend. The slight increase in expression level with the mutated *sppA* should therefore be further investigated by introducing mutations causing higher TIR to see if a higher expression rate further improves the expression of recombinant proteins.

Similarly, the slight negative effect of introducing the *skp* mutation could also be due to a too high TIR resulted from the introduced RBS sequence. An alternative approach could be to introduce an RBS giving a lower TIR for this gene. As indicated in the introduction, *skp* is located in an operon with in total four proteins: *lpxD, fabZ* and *lpxA* (see figure 1.4). Change of the TIR of *skp* may therefore to some degree affect the expression of these proteins allthough *lpxD, fabZ* and *lpxA* are under control by a separate promoter because there is no terminator separating the genes. The research of these proteins is mostly based on the knock-out [43] mutations of the genes and in vitro research [43, 44]. No literature found has indicated negative effects of the up-regulation of these genes, and the potential up-regulation of these genes from increasing the TIR of *skp* are not expected to affect the results for the *skp*-mutants.

The decrease in expression level for the *degP* mutated strain was unexpected, as the use of *degP*-deficient strains has been reported to allow high-level accumulation of an antibody fragment in the periplasm in combination with the knock-out mutations of *prc* and *spr*. However, also in the study by Chen et al., the degP mutation alone did not give a significant effect on the production levels [17]. BL21 is already deficient in two proteases [13]. The effect of down-regulating the TIR of *degP* in other strains of *E. coli* could therefore be investigated to see if there could be any effects of introducing this mutation in strains that are not deficient in proteases.

As mentioned in the introduction, the approach for optimization has to be evaluated in each specific case depending on the recombinant protein expressed. Ideally, several reporter systems should be tested in order to investigate whether the up-/down-regulations of selected genes affect the expression of recombinant proteins more when there are specific challenges, or when they are targeted to other secretion pathways by using different signal peptides. For example, the down-regulation of the protease DegP would be expected to give a more significant change when studying a reporter protein prone to degradation. Similarly, the up-regulation of the *skp*

gene would be expected to give a more significant effect in the case of proteins that are prone to be produced in an insoluble state. Additionally, the expected effect of the *skp* up-regulation was an increased solubility of the protein of interest, and this should have been analyzed specifically, for instance by western blot.

SppA was expected to give an increase in protein secretion, as the insertion of an additional copy of this protein was found to give an increased secretion in *Bacillus licheniformis* [31]. However, as mentioned in the introduction, there are differences in the SppA of *B. licheniformis* and the SppA of *E. coli*. The $SppA_{EC}$ holds a narrower substrate preference compared to the $SppA_{BL}$ [32]. Therefore, again, more reporter proteins should be analyzed in order to give a further conclusion on the effects of this mutation.

Metabolic burden due to recombinant protein over-expression and improper culture conditions may lead to cell lysis. The efficiency of secretion may be affected by cell lysis. One of the methods for evaluating cell lysis mentioned in the review by Kleiner et al. was to measure the native alkaline phosphatase in the fermentation supernatant [14]. Following this analogy, there is therefore reason to believe that some of the alkaline phosphatase would also leak into the media. The measured level of alkaline phosphatase in the periplasmic space would then be less than the total expressed protein. Measuring the levels of alkaline phosphatase in the cell-free medium would therefore give an indication of leakiness of the constructed strains.

CHAPTER 5

Conclusions

Seven strains of *E. coli* were constructed by introducing single gene mutations of *sppA*, *skp* and *degP*, and combinations of these mutations. Up-regulation of the translation initiation rate of *sppA* and *skp*, and down-regulation of *degP* was obtained by by changing the respective RBS sequences in the genome of *E. coli* BL21 by CRMAGE recombineering. This was achieved by the successful construction of the CRMAGE system for introducing all of the three target mutations. The pMAZ-SK helper plasmid with gRNA insert and MAGE oligos were successfully constructed for all three target genes.

All constructed strains were cured from the pMA7CR_2.0 and pMAZ-SK, but the pZS4Int-tetR plasmid was not successfully eliminated. Methods of curing by heat treatment, acridine orange, and acridine orange together with Penicillin G were not successful for elimination of the pZS4Int-tetR plasmid.

A protocol for analyzing the periplasmic recombinant expression efficiency by osmotic shock and alkaline phosphatase assay was tested and established in the lab. The constructed strains and the negative control (BL21) all harbouring the pZS4Int-tetR plasmid were successfully transformed with the reporter plasmid, and the reporter protein was expressed and analyzed by colorimetric analysis.

The *degP* down-regulation resulted in a significant decrease in production of alkaline phosphatase (AP) compared to the wildtype BL21. No other significant differences were observed between the units of phosphatase activity of the wildtype and the values of any of the constructed strains. The mutation of *sppA* resulted in higher levels of alkaline phosphatase compared to the other constructed strains, but not high enough to state a significance compared to BL21, and the up-regulation *sppA* in particular should be further investigated. The created strains should be further investigated by using other reporter systems in order to give a conclusion of the actual effects of the mutations.

Bibliography

[1] Stanley N Cohen et al. "Construction of biologically functional bacterial plasmids in vitro". In: *Proceedings of the National Academy of Sciences* 70.11 (1973), pp. 3240–3244.

[2] Irving S Johnson. "Human insulin from recombinant DNA technology". In: *Science* 219.4585 (1983), pp. 632–637.

[3] FR Schmidt. "Recombinant expression systems in the pharmaceutical industry". In: *Applied microbiology and biotechnology* 65.4 (2004), pp. 363–372.

[4] Laura Sanchez-Garcia et al. "Recombinant pharmaceuticals from microbial cells: a 2015 update". In: *Microbial cell factories* 15.1 (2016), p. 33.

[5] Rick Ng. *Drugs : from discovery to approval*. eng. New York, 2015.

[6] *RECOMBINANT PROTEIN MARKET - GROWTH, TRENDS, AND FORECAST*. URL: https://www.mordorintelligence.com/industry-reports/recombinant-protein-market (visited on 06/13/2019).

[7] François Baneyx. "Recombinant protein expression in Escherichia coli". In: *Current opinion in biotechnology* 10.5 (1999), pp. 411–421.

[8] Royston Jefferis. "Glycosylation of recombinant antibody therapeutics". In: *Biotechnology progress* 21.1 (2005), pp. 11–16.

[9] Germán L Rosano and Eduardo A Ceccarelli. "Recombinant protein expression in Escherichia coli: advances and challenges". In: *Frontiers in microbiology* 5 (2014), p. 172.

[10] Ben Lugtenberg and Loek Van Alphen. "Molecular architecture and functioning of the outer membrane of Escherichia coli and other gram-negative bacteria". In: *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* 737.1 (1983), pp. 51–115.

[11] FJM Mergulhao, David K Summers, and Gabriel A Monteiro. "Recombinant protein secretion in Escherichia coli". In: *Biotechnology advances* 23.3 (2005), pp. 177–202.

[12] Neus Ferrer-Miralles et al. "Microbial factories for recombinant pharmaceuticals". In: *Microbial cell factories* 8.1 (2009), p. 17.

[13] Kay Terpe. "Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems". In: *Applied microbiology and biotechnology* 72.2 (2006), p. 211.

[14] Gabriele RM Kleiner-Grote, Joe M Risse, and Karl Friehs. "Secretion of recombinant proteins from E. coli". In: *Engineering in Life Sciences* 18.8 (2018), pp. 532–550.

[15] Sung H Yoon, Seong Keun Kim, and Jihyun F Kim. "Secretory production of recombinant proteins in Escherichia coli". In: *Recent patents on biotechnology* 4.1 (2010), pp. 23–29.

[16] Mehmet Berkmen. "Production of disulfide-bonded proteins in Escherichia coli". In: *Protein expression and purification* 82.1 (2012), pp. 240–251.

[17] Christina Chen et al. "High-level accumulation of a recombinant antibody fragment in the periplasm of Escherichia coli requires a triple-mutant (degP prc spr) host strain". In: *Biotechnology and bioengineering* 85.5 (2004), pp. 463–474.

[18] JH Choi and SangYup Lee. "Secretory and extracellular production of recombinant proteins using Escherichia coli". In: *Applied microbiology and biotechnology* 64.5 (2004), pp. 625–635.

[19] Yoichi Kurokawa, Hideki Yanagi, and Takashi Yura. "Overexpression of protein disulfide isomerase DsbC stabilizes multiple-disulfide-bonded recombinant protein produced and transported to the periplasm in Escherichia coli". In: *Appl. Environ. Microbiol.* 66.9 (2000), pp. 3960–3965.

[20] Anton V Zavialov et al. "Secretion of recombinant proteins via the chaperone/usher pathway in Escherichia coli". In: *Appl. Environ. Microbiol.* 67.4 (2001), pp. 1805–1814.

[21] Bernard R Glick. "Metabolic load and heterologous gene expression". In: *Biotechnology advances* 13.2 (1995), pp. 247–261.

[22] Susan Schlegel et al. "Optimizing heterologous protein production in the periplasm of E. coli by regulating gene expression levels". In: *Microbial cell factories* 12.1 (2013), p. 24.

[23] Joen Luirink et al. "Biogenesis of inner membrane proteins in Escherichia coli". In: *Annu. Rev. Microbiol.* 59 (2005), pp. 329–355.

[24] Henrik Nielsen et al. "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites." In: *Protein engineering* 10.1 (1997), pp. 1–6.

[25] David JF Du Plessis, Nico Nouwen, and Arnold JM Driessen. "The sec translocase". In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1808.3 (2011), pp. 851–865.

[26] Tracy Palmer and Ben C Berks. "The twin-arginine translocation (Tat) protein export pathway". In: *Nature Reviews Microbiology* 10.7 (2012), p. 483.

[27] Yizhou Zhou et al. "Enhancing full-length antibody production by signal peptide engineering". In: *Microbial cell factories* 15.1 (2016), p. 47.

[28] Rachel B Kapust and David S Waugh. "Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused". In: *Protein Science* 8.8 (1999), pp. 1668–1674.

[29] Haeyoung Jeong, Hyun Ju Kim, and Sang Jun Lee. "Complete genome sequence of Escherichia coli strain BL21". In: *Genome announcements* 3.2 (2015), e00134–15.

[30] Howard M Salis, Ethan A Mirsky, and Christopher A Voigt. "Automated design of synthetic ribosome binding sites to control protein expression". In: *Nature biotechnology* 27.10 (2009), p. 946.

[31] Dongbo Cai et al. "A novel strategy to improve protein secretion via overexpression of the SppA signal peptide peptidase in Bacillus licheniformis". In: *Microbial cell factories* 16.1 (2017), p. 70.

[32] Sung-Eun Nam and Mark Paetzel. "Structure of signal peptide peptidase A with C-termini bound in the active sites: insights into specificity, self-processing, and regulation". In: *Biochemistry* 52.49 (2013), pp. 8811–8822.

[33] Zengyi Chang. "The function of the DegP (HtrA) protein: Protease versus chaperone". In: *IUBMB life* 68.11 (2016), pp. 904–907.

[34] Christoph Spiess, Alexandra Beil, and Michael Ehrmann. "A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein". In: *Cell* 97.3 (1999), pp. 339–347.

[35] Robert Chen and Ulf Henning. "Aperiplasmic protein (Skp) of Escherichia coli selectively binds a class of outer membrane proteins". In: *Molecular microbiology* 19.6 (1996), pp. 1287–1294.

[36] Paula V Bulieris et al. "Folding and insertion of the outer membrane protein OmpA is assisted by the chaperone Skp and by lipopolysaccharide". In: *Journal of Biological Chemistry* 278.11 (2003), pp. 9092–9099.

[37] Hendrick Bothmann and Andreas Plückthun. "Selection for a periplasmic factor improving phage display and functional periplasmic expression". In: *Nature biotechnology* 16.4 (1998), p. 376.

[38] Andrew Hayhurst and William J Harris. "Escherichia coliskp chaperone coexpression improves solubility and phage display of single-chain antibody fragments". In: *Protein expression and purification* 15.3 (1999), pp. 336–343.

[39] Raphael Levy et al. "Production of correctly folded Fab antibody fragment in the cytoplasm of Escherichia coli trxB gor mutants via the coexpression of molecular chaperones". In: *Protein expression and purification* 23.2 (2001), pp. 338–347.

[40] *EcoCyc E. coli Database*. URL: `https://ecocyc.org` (visited on 06/13/2019).

[41] Claire Dartigalongue, Dominique Missiakas, and Satish Raina. "Characterization of the Escherichia coliσE Regulon". In: *Journal of Biological Chemistry* 276.24 (2001), pp. 20866–20875.

[42] Craig M Bartling and Christian RH Raetz. "Steady-state kinetics and mechanism of LpxD, the N-acyltransferase of lipid A biosynthesis". In: *Biochemistry* 47.19 (2008), pp. 5290–5302.

[43] Akintunde Emiola, John George, and Steven S Andrews. "A complete pathway model for lipid A biosynthesis in Escherichia coli". In: *PloS one* 10.4 (2015), e0121216.

[44] Richard J Heath and Charles O Rock. "Roles of the FabA and FabZ β-hydroxyacyl-acyl carrier protein dehydratases in Escherichia coli fatty acid biosynthesis". In: *Journal of Biological Chemistry* 271.44 (1996), pp. 27795–27801.

[45] David L. Nelson. *Lehninger principles of biochemistry*. eng. 6th ed. New York: W.H. Freeman, 2013, p. 1198. ISBN: 9781464109621.

[46] Kathryn L Strauch, K Johnson, and J Beckwith. "Characterization of degP, a gene required for proteolysis in the cell envelope and essential for growth of Escherichia coli at high temperature." In: *Journal of Bacteriology* 171.5 (1989), pp. 2689–2696.

[47] Carlotta Ronda et al. "CRMAGE: CRISPR optimized mage recombineering". In: *Scientific reports* 6 (2016), p. 19452.

[48] Luke Z Yuan et al. "Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in E. coli". In: *Metabolic engineering* 8.1 (2006), pp. 79–90.

[49] URL: `https://salislab.net/software/` (visited on 01/21/2019).

[50] Amin Espah Borujeni, Anirudh S Channarasappa, and Howard M Salis. "Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites". In: *Nucleic acids research* 42.4 (2013), pp. 2646–2659.

[51] Jennifer A Doudna and Emmanuelle Charpentier. "The new frontier of genome engineering with CRISPR-Cas9". In: *Science* 346.6213 (2014), p. 1258096.

[52] Hilary M Ellis, Daiguan Yu, Tina DiTizio, et al. "High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides". In: *Proceedings of the National Academy of Sciences* 98.12 (2001), pp. 6742–6746.

[53] Kira S Makarova et al. "Evolution and classification of the CRISPR–Cas systems". In: *Nature Reviews Microbiology* 9.6 (2011), p. 467.

[54] F Ann Ran et al. "Genome engineering using the CRISPR-Cas9 system". In: *Nature protocols* 8.11 (2013), p. 2281.

[55] Le Cong et al. "Multiplex genome engineering using CRISPR/Cas systems". In: *Science* 339.6121 (2013), pp. 819–823.

[56] Martin Jinek et al. "A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity". In: *Science* 337.6096 (2012), pp. 816–821.

[57] Harris H Wang et al. "Programming cells by multiplex genome engineering and accelerated evolution". In: *Nature* 460.7257 (2009), p. 894.

[58] Neal G Copeland, Nancy A Jenkins, and Donald L Court. "Recombineering: a powerful new tool for mouse functional genomics". In: *Nature Reviews Genetics* 2.10 (2001), p. 769.

[59] Ryan R Gallagher et al. "Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA". In: *nature protocols* 9.10 (2014), p. 2301.

[60] GAIL E Herman and PAUL Modrich. "Escherichia coli K-12 clones that overproduce dam methylase are hypermutable." In: *Journal of bacteriology* 145.1 (1981), pp. 644–646.

[61] Charles S Hoffman and Andrew Wright. "Fusions of secreted proteins to alkaline phosphatase: an approach for studying protein secretion". In: *Proceedings of the National Academy of Sciences* 82.15 (1985), pp. 5107–5111.

[62] Sophie E Jackson, Timothy D Craggs, and Jie-rong Huang. "Understanding the folding of GFP using biophysical techniques". In: *Expert review of proteomics* 3.5 (2006), pp. 545–559.

[63] Jean-Denis Pédelacq et al. "Engineering and characterization of a superfolder green fluorescent protein". In: *Nature biotechnology* 24.1 (2006), p. 79.

[64] Sine Alise H Hansen. "Validation of functionality of the reporter system for monitoring periplasmic recombinant protein expression in *Echerichia coli*". In: (2018). Not published.

[65] Hans Ulrich Bergmeyer, Jürgen Bergmeyer, and Marianne Grassl. *Methods of enzymatic analysis*. Vol. 2. Elsevier, 2012, p. 682.

[66] Edith Brickman and Jon Beckwith. "Analysis of the regulation of Escherichia coli alkaline phosphatase synthesis using deletions and φ80 transducing phages". In: *Journal of molecular biology* 96.2 (1975), pp. 307–316.

[67] Pauline M. Doran. *Bioprocess Engineering Principles*. eng. 2nd ed. Vol. 9780080917702. Elsevier Science, 2012, p. 919. ISBN: 012220851X.

[68] Simone Balzer et al. "A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in Escherichia coli". In: *Microbial cell factories* 12.1 (2013), p. 26.

[69] Janet Martha Blatny et al. "Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon." In: *Applied and environmental microbiology* 63.2 (1997), pp. 370–379.

[70] Agnieszka Gawin, Svein Valla, and Trygve Brautaset. "The XylS/Pm regulator/promoter system and its use in fundamental studies of bacterial gene expression, recombinant protein production and metabolic engineering". In: *Microbial biotechnology* 10.4 (2017), pp. 702–718.

[71] Merete Thune Wiiger et al. "A novel human recombinant single-chain antibody targeting CD166/ALCAM inhibits cancer cell invasion in vitro and in vivo tumour growth". In: *Cancer immunology, immunotherapy* 59.11 (2010), pp. 1665–1674.

[72] Holger Thie et al. "SRP and Sec pathway leader peptides for antibody phage display and antibody fragment production in E. coli". In: *New biotechnology* 25.1 (2008), pp. 49–54.

[73] ROSS H Durland et al. "Mutations in the trfA replication gene of the broad-host-range plasmid RK2 result in elevated plasmid copy numbers." In: *Journal of bacteriology* 172.7 (1990), pp. 3859–3867.

[74] Jürgen Brosius et al. "Gene organization and primary structure of a ribosomal RNA operon from Escherichia coli". In: *Journal of molecular biology* 148.2 (1981), pp. 107–127.

[75] Robin G Taylor, David C Walker, and RR McInnes. "E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing." In: *Nucleic acids research* 21.7 (1993), p. 1677.

[76] Je-Nie Phue et al. "Modified Escherichia coli B (BL21), a superior producer of plasmid DNA compared with Escherichia coli K (DH5α)". In: *Biotechnology and bioengineering* 101.4 (2008), pp. 831–836.

[77] Rolf Lutz and Hermann Bujard. "Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements". In: *Nucleic acids research* 25.6 (1997), pp. 1203–1210.

[78] *EXPRESSYS: pZ vectors*. URL: `http://expressys.com/main_tools.html` (visited on 06/03/2019).

[79] Patrick D Hsu et al. "DNA targeting specificity of RNA-guided Cas9 nucleases". In: *Nature Biotechnology* 31 (2013).

[80] Daniel G Gibson et al. "Enzymatic assembly of DNA molecules up to several hundred kilobases". In: *Nature Methods* VOL.6 NO.5 (2009), pp. 343–347.

[81] Andrea Iselin Elvheim. "Effects of Sec-Pathway Modifications on Periplasmatic Translocation of Recombinant Proteins in Escherichia coli". MA thesis. NTNU, 2018.

[82] David M Heery et al. "Curing of a plasmid from E. coli using high-voltage electroporation." In: *Nucleic acids research* 17.23 (1989), p. 10131.

[83] Juan M Tomás and William W Kay. "A simple and rapid method for the elimination of R plasmids from enteric bacteria". In: *Current Microbiology* 11.3 (1984), pp. 155–157.

[84] Harold C Neu and James Chou. "Release of surface enzymes in Enterobacteriaceae by osmotic shock". In: *Journal of bacteriology* 94.6 (1967), pp. 1934–1945.

[85] Thomas Hong et al. "Periplasmic alkaline phosphatase activity and abundance in Escherichia coli B23 and C29 during Exponential and Stationary Phase". In: *Journal of Experimental Microbiology and Immunology (JEMI) Vol* 11 (2007), pp. 8–13.

[86] Kechao Yang and William W Metcalf. "A new activity for an old enzyme: Escherichia coli bacterial alkaline phosphatase is a phosphite-dependent hydrogenase". In: *Proceedings of the National Academy of Sciences* 101.21 (2004), pp. 7919–7924.

[87] Ronald E Walpole et al. *Probability and statistics for engineers and scientists*. 9th. Vol. 5. Macmillan New York, 2011, p. 791. ISBN: 978-0-321-62911-1.

[88] JT Trevors. "Plasmid curing in bacteria". In: *FEMS microbiology reviews* 1.3-4 (1986), pp. 149–157.

[89] Jae Hong Seol et al. "Protease Do is essential for survival of Escherichia coli at high temperatures: its identity with the htrA gene product". In: *Biochemical and biophysical research communications* 176.2 (1991), pp. 730–736.

[90] Ida Lauritsen et al. "A versatile one-step CRISPR-Cas9 based approach to plasmid-curing". In: *Microbial cell factories* 16.1 (2017), p. 135.

[91] Steven S Zumdahl. *Chemical principles*. eng. S.l., 2013.

[92] L Leon Campbell and F Marion Hulett-Cowling. "Purification and properties of an alkaline phosphatase of Bacillus licheniformis". In: *Biochemistry* 10.8 (1971), pp. 1364–1371.

[93] U. N. Baton. ""An introduction to error analysis" by John R. Taylor". In: *Strain* 33.4 (1997), pp. 133–134. ISSN: 0039-2103.

[94] MathWorks. *N-way ANOVA*. Version R2019a. URL: `https://se.mathworks.com/help/stats/n-way-anova.html` (visited on 06/04/2019).

[95] MathWorks. *Multiple Comparisons*. Version R2019a. URL: `https://se.mathworks.com/help/stats/multiple-comparisons.html#bum7ugv-1` (visited on 05/29/2019).

APPENDIX A

Media and solutions

Media

All media were autoclaved at liquid setting (20 min, 121 °C), and antibiotics were added after sufficiently cooling the medium.

Lysogeny Broth (LB)

1% (w/v)	Tryptone
0.5% (w/v)	Yeast extract
171 mM	NaCl
-	dH_2O

Lysogeny Agar (LA)

1% (w/v)	Tryptone
0.5% (w/v)	Yeast extract
171 mM	NaCl
1.5% (w/v)	Agar
-	dH_2O

LB Lennox (low salt concentration) was used for the CRMAGE protocol.

LB lennox

1% (w/v)	Tryptone
0.5% (w/v)	Yeast extract
85.5 mM	NaCl
-	dH_2O

After heat-shock transformations, the cells were recovered in SOC-medium.

SOC medium

0.5% (w/v)	Yeast extract
2% (w/v)	Tryptone
10 mM	NaCl
2.5 mM	KCl
20 mM	$MgSO_4 \cdot 7\,H_2O$
-	dH_2O

After cooling to less than 50 °C, add 2 mL filter sterilized glucose solution (20 mM). Aliquot in 1.5 mL tubes and store at -20 °C.

Psi medium

2% (w/v)	Tryptone
0.5% (w/v)	Yeast extract
0.5% (w/v)	$MgSO_4 \cdot 7\,H_2O$
-	dH_2O

pH adjusted to 7.6 using KOH.

Solutions

TFB1 and TFB2 was used to create RbCl competent *E. coli*.

TFB1

30 mM	KAc
100 mM	RbCl
10 mM	$CaCl_2 \cdot 2\,H_2O$
50 mM	$MnCl_2 \cdot 4\,H_2O$
15% (v/v)	Glycerol
-	dH_2O

pH adjusted to 5.8 using diluted acetic acid. Sterilize by filtration.

TFB2

10 mM	MOPS
10 mM	RbCl
75 mM	$CaCl_2 \cdot 2\,H_2O$
15 % (v/v)	Glycerol
-	dH_2O

pH adjusted to 6.5 using diluted NaOH. Sterilize by filtration.
TAE buffer was used as a running buffer for gel electrophoresis.

1× TAE buffer

4.84 g/L	Tris(hydroxymethyl)aminomethane
0.1142 % (v/v)	Acetic acid
0.05 M	Ethylenediaminetetraacetic acid (EDTA), pH 8
	dH_2O

Autoclave.
0.8 % agarose gel was used for separation of DNA fragments by gel electrophoresis.

Agarose gel

8 g/L	agarose
0.003 % (v/v)	GelRed or GelGreen
	1 × TAE buffer

Phosphate-buffered saline (PBS)

8 g/L	NaCl
0.2 g/L	KCl
1.44 g/L	Na_2HPO_4
0.24 g/L	KH_2PO_4
	dH_2O

Adjust pH to 7.4 with HCl (or NaOH) and autoclave.

20 % sucrose, 0.03 M Tris-HCl

20 %	Sucrose
0.03 M	Tris
	dH_2O

Adjust pH to 8.0 with HCl and sterilize by filtration.

5 mM disodium EDTA

5 mM	disodium EDTA
	dH_2O

Adjust pH to 8.0 with KOH and autoclave.

1M Tris-acetate buffer

1 M	Tris
	dH_2O

Adjust pH to 8.0 with glacial acetic acid.

PCR, assembly and annealing master mixes

The assembly master mix was used for Gibson assembly for plasmid assembly [80].

Assembly master mix (A-MM)
320 µL IRB
0.64 µL T5 exonuclease (NEB, 10 U/µL)
20 µL Phusion® High-Fidelity DNA Polymerase (NEB, 40 U/µL)
700 µL MilliQ

15 µL aliquots stored at -20 °C.
The PCR master mixes used for Taq PCR and Q5 PCR are shown in table A.1 and table A.2 respectively.

Table A.1: *Taq* DNA polymerase master mix used for colony PCR.

Reagent	**One reaction (µL)**
Nuclease free water	8.35
10× standard *Taq* Run Buffer	1.0
10 mM dNTPs	0.2
10µ M forward primer	0.2
10µ M reverse primer	0.2
Taq DNA Polymerase	0.05

Table A.2: Q5 DNA polymerase master mix used for amplifying DNA of low concentration.

Reagent	**One reaction (µL)**
Nuclease free water	17
Q5 Buffer	5
10 mM dNTPs	0.5
10µ M forward primer	0.625
10µ M reverse primer	0.625
Q5 DNA Polymerase	0.05
Template	1

APPENDIX B

Alkaline phosphatase assay calculations

The Beer-Lambert Law shown in equation B.0.1 relates the absorbance to the concentration of the absorbing species in a solution [91].

$$A = -\log_{10}\left(\frac{I}{I_0}\right) = \epsilon l c \tag{B.0.1}$$

where A is the absorbance, I_0 is the incident light intensity, and I is the detected light intensity after the light has passed through the sample. ϵ is the molar extinction coefficient, also called the molar absorptivity, and is constant with constant temperature and pH. l is the light path length (cm), and c is the concentration of the absorbing species [91].

One unit of phosphatase is defined as "that amount of enzyme which liberates 1 µmol of p-nitrophenol per minute under the defined conditions" [92]. Using the Beer-Lambert equation, it is possible to calculate the change in concentration of p-nitrophenol and therefore also the units of phosphatase by using absorbance measurements within the wavelength range of absorbance for the compound. The spectrophotometer measures the optical density, OD, which is defined as the absorbance divided by path length. The Beer-Lambert equation may therefore be written as in equation B.0.2.

$$c = \frac{A}{\epsilon l} = \frac{\mathrm{OD}}{\epsilon} \tag{B.0.2}$$

To be able to compare samples of varying densities at harvest, the unit of phosphatase was normalized by dividing with the $\mathrm{OD}_{600\mathrm{nm}}$ at harvest. To correct for any absorbance caused by cells and cell debris, 1.75 $\mathrm{OD}_{550\mathrm{nm}}$ was subtracted from the $\mathrm{OD}_{420\mathrm{nm}}$ as described by Brickman and Beckwidth (1975). The equation used for calculating units of phosphatase is shown in equation B.0.3 [66].

$$U_{\mathrm{AP}} = 1000 \cdot \frac{\mathrm{OD}_{420} - 1.75 \cdot \mathrm{OD}_{550}}{\mathrm{OD}_{600} \cdot t} \tag{B.0.3}$$

The factor 1000 converts the unit from [u $\mathrm{min}^{-1}\mathrm{mL}^{-1}$] to [u $\mathrm{min}^{-1}\mathrm{L}^{-1}$]. Assuming that OD_{550} is constant through time and with a constant ϵ, the units of phosphatase,

U_{AP} is proportional to the concentration of p-nitrophenol, and may therefore be used for comparing the level of produced phosphatase between strains. Here, it is assumed that the density of cells at harvest is proportional to the amount of alkaline phosphatase produced.

The absorbance of the cultures at harvest were measured in 1 cm cuvettes, while the measurements of the $\mathrm{OD}_{550\mathrm{nm}}$ and $\mathrm{OD}_{420\mathrm{nm}}$ were done in 96-well plates. In the 96-well plate, the light path length depends on the volume and the meniscus of the solution, and a correction factor may be used to be able to compare the values. A correction factor was calculated by measuring the absorbance of the buffer used at 977 nm and 900 nm wavelengths in both the cuvette and the 96-well plate. The correction factor, k was calculated by equation B.0.4

$$k = \frac{\mathrm{OD}_{977}(\text{cuvette}) - \mathrm{OD}_{900}(\text{cuvette})}{\mathrm{OD}_{977}(\text{well}) - \mathrm{OD}_{900}(\text{well})} \tag{B.0.4}$$

Where OD_{977}(cuvette) and OD_{900}(cuvette) are the absorbances in the 1 cm cuvette at 977 nm and 900 nm respectively, and OD_{977}(well) and OD_{900}(well) are the absorbances in the 96-well plate at 977 nm and 900 nm respectively. The OD(cuvette) of the samples measured in the 96-well plates can then be calculated by equation B.0.5.

$$\mathrm{OD}(\text{cuvette}) = k \cdot \mathrm{OD}(\text{well}) \tag{B.0.5}$$

Data from linear regression

Linear regression was used to find the absorbance at the start of the reaction (t=0) by using the polyfit function in MATLAB. The extrapolated $\mathrm{OD}_{420\mathrm{nm}}$ at t=0 for each sample is shown in table B.1.

Table B.1: $\mathrm{OD}_{420\mathrm{nm}}$ at t=0 derived from linear regression analysis performed in MATLAB using the polyfit function.

Parallel	**1**	**2**	**3**
BL21	0,0435	0,0268	0,0538
BL21_sppA	0,0455	0,0402	0,0680
BL21_skp	0,0206	0,0227	0,0250
BL21_degP	0,0188	0,0177	0,0199
BL21_sppA+skp	0,0360	0,0400	0,0449
BL21_sppA+degP	0,0185	0,0291	0,0282
BL21_skp+degP	0,0224	0,0149	0,0122
BL21_sppA+skp+degP	0,0175	0,0241	0,0226
DH5α	0,0153	0,00945	0,0100

APPENDIX C

Statistics

Standard deviation

Standard deviation of the mean (SDOM), $\sigma_{\overline{x}}$, was used to calculate the error of the OD measurements. The SDOM is given by equation C.0.1 [93],

$$\sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{N}} \tag{C.0.1}$$

where N is the number of measurements and σ_x is the standard deviation given by equation C.0.2.

$$\sigma_x = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^{i} (x_i - \overline{x})^2} \tag{C.0.2}$$

x_i is the value of each data point, and $\overline{x}$ is the mean value.

Analysis of variance

The MATLAB function anovan was used to calculate the Analysis of Variance (ANOVA) on the dataset obtained from the alkaline phosphatase measurements (see section 2.6.1 and section 3.4). It is assumed that the populations are normally distributed with a common mean and variance [87]. Equation C.0.3 shows the model for the three factor experiment.

$$y_{ijkr} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkr} \tag{C.0.3}$$

y_{ijkr} is an observation of the response. μ is the overall mean. α_i, β_j and γ_k denotes the effects of factors A, B and C respectively at the ith, jth or kth level. $(\alpha\beta)_{ij}$ is the interaction effect between factors A and B. In this thesis, the factors are the three different mutations of *sppA*, *skp* and *degP* in the levels mutated=1 or wildtype=0.

Random disturbances, ϵ_{ijkr} are assumed to be independent, normally distributed, and have constant variance. The effects of all levels for each effects sum up to zero as shown in equation C.0.4 [87, 94].

$$\sum_{i=1}^{I} \alpha_i = 0, \quad \sum_{j=1}^{J} \beta_j = 0, \quad \sum_{k=1}^{K} \gamma_k = 0 \tag{C.0.4}$$

Similarly, the interaction effects sum up to zero for each coefficient as shown in equation C.0.5 [87, 94].

$$\begin{aligned} \sum_{i=1}^{I} (\alpha\beta)_{ij} = \sum_{j=1}^{J} (\alpha\beta)_{ij} = 0, \\ \sum_{i=1}^{I} (\alpha\gamma)_{ik} = \sum_{k=1}^{K} (\alpha\gamma)_{ik} = 0, \\ \sum_{j=1}^{J} (\beta\gamma)_{jk} = \sum_{k=1}^{K} (\beta\gamma)_{jk} = 0, \\ \sum_{i=1}^{I} (\alpha\beta\gamma)_i jk = \sum_{j=1}^{J} (\alpha\beta\gamma)_{ijk} = \sum_{k=1}^{K} (\alpha\beta\gamma)_{ijk} = 0 \end{aligned} \tag{C.0.5}$$

ANOVA test the zero hypothesis that each of the main effects are equal to zero against the hypothesis that at least one of them are different [87]. With three factors (A, s and P) of two levels (mutated=1 and non-mutated=0), the zero-hypotheses and the alternate hypotheses are:

1. H_0: $\alpha_0 = \alpha_1 = 0$
   H_1: $\alpha_0 \neq \alpha_1$

2. H_0: $\beta_0 = \beta_1 = 0$
   H_1: $\beta_0 \neq \beta_1$

3. H_0: $\gamma_0 = \gamma_1 = 0$
   H_1: $\gamma_0 \neq \gamma_1$

4. $H_0 : (\alpha\beta)_{ij} = 0$
   H_1: at least one $(\alpha\beta)_{ij} \neq 0$

5. $H_0 : (\alpha\gamma)_{ik} = 0$
   H_1: at least one $(\alpha\gamma)_{ik} \neq 0$

6. $H_0 : (\beta\gamma)_{jk} = 0$
   H_1: at least one $(\beta\gamma)_{jk} \neq 0$

7. $H_0 : (\alpha\beta\gamma)_{ijk} = 0$
   H_1: at least one $(\alpha\beta\gamma)_{ijk} \neq 0$

where α_i is the effect of factor A in the ith level, β_j is the effect of factor s in the jth level and γ_k is the effect of factor P in the kth level. $(\alpha\beta)_{ij}$ is the interaction effect for factors A and s, $(\alpha\gamma)_{ik}$ is the interaction effect for factors A and P, $(\beta\gamma)_{jk}$ is the interaction effect for factors s and P, and $(\alpha\beta\gamma)_{ijk}$ is the interaction effect for all factors A, s and P.

anovan displays an ANOVA-table (see 3.10), showing the calculated sum of squares, mean squared value, F-value and the p-value (Prob>F). A p-value $< \alpha$ where α is the significance level, here $\alpha = 0.05$, means that the H_0 hypothesis may be rejected in favour of the H_1 hypothesis. If the p-value is higher than the significance level, H_0 is not rejected. However, non-significant main effects may be a result of masking effects because of significant interaction effects. Therefore, if any of the interaction effects are significant, only those tests on the main effects that turn out to be significant are meaningful [87].

Multiple comparisons: Tukey-Kramer method

The MATLAB function multcompare uses the statistical data obtained from the ANOVA to do multiple pairwise comparisons of the treatment effects. In this thesis, the default method, Tukey's honestly significant difference criterion was used. This test is based on the studentized range distribution.

The zero-hypothesis H_0, is that the compared treatment effects are equal, H_0: $\alpha_i = \alpha_j$. In this thesis, that means that the effect of the mutation is the same as that of the wildtype. H_0 is rejected if

$$|t| = \frac{|\bar{y}_i - \bar{y}_j|}{\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} > \frac{1}{\sqrt{2}} q_{\alpha,k,N-k} \tag{C.0.6}$$

The upper percentile point, $q_{\alpha,k,N-k}$ is a function of α, the number of groups k, and degrees of freedom, $N-k$ where N is the total number of observations. This method is optimal for balanced one-way ANOVA with equal sample sizes [95]. This method was used on data with equal sample sizes.

Interaction plot

The interaction plot visualizes the effects of changing one factor as one moves from one level to another of a second factor [87]. The function interactionplot in MATLAB was used to display the interaction plot showing interaction effects between the mutants in section 3.4. In this thesis, two levels are displayed, mutated=1 or wildtype=0.

APPENDIX D

Oligonucleotides

Primers

Primers used for sequencing, colony PCR and PCR amplification are shown in table D.1.

Table D.1: Primers used in this project and their melting temperatures, T_m.

Primer name	**T_m [°C]**	**Sequence (5′ → 3′)**
Gibson_pMAZ_backbone_F	49.46	GTTTTAGAGCTAGAAATAGCAAG
Gibson_pMAZ_backbone_R	55.84	GTGCTCAGTATCTCTATCACTGATAGG
pMAZ-SK_seq_F	59.8	CGCCGAAGCGGGGTTTTTTG
FWD_seq_pTA16-s173_AP	53.4	AAGAAGCGGATACAGGAGTG
sppA_2_F	49.1	ATTATACCTAACGAGGAGAAAC
sppA_1_R	52.4	TTTGCCGATGTACTGCATAG
sppA_seq_F	54.8	CGGGACGTTAATCGACGTAG
sppA_seq_R	54.4	AGAGCTTTGCCGATGTACTG
skp_10x_F	52.1	TACGTCTAACACTTAGGGGG
skp_1_R	56.9	GTAAGCAACGGCGTTTGCAT
skp_seq_F	55.0	CGGTTCGTACTTCCTTCTTCTG
skp_seq_R	54.8	CAGATCGATATCCTGGCTGTTG
degP_10x_F	54.5	TTGAGTAACGCCACTGATCG
degP_10x_R	55.2	ATCAATGATGACGCCGGAAC
degP_seq_F	53.7	TTTCGCCGGAACGTTTAATC
degP_seq_R	53.8	AGCGCCATGAATTTCTGTTG

MAGE oligonucleotides

The MAGE oligonucleotide sequences as they were ordered are shown in table D.2. Uppercase letters are the inserted bases, and lowercase letters constitute the homology region.

Table D.2: MAGE recombineering oligos. Uppercase letters are the non-homologous region of the MAGE oligos, and includes the sequence of the constructed RBS sequences. The underlined sequence for *skp* shows the start codon. A single nucleotide substitution (G → A) was introduced in *degP* to interrupt the PAM sequence.

Gene	**Sequence (5′→ 3′)**
sppA	actcctgtcgtgatatttattcacaaaattaacacgagagtggattttgttacagcacagtccgcaa ttcctgctgacaagtaatctattCAAACGTACCCCTTAATTATACCTAACG AGGAGAAACTatgcgaacccttggcgatttattgccggatttttaaatggacgtg gcgtctgctgaatttcgtccgtgaaatggtacttaacctgttc
skp	gcccatgttgacgattgcaattttgtcagccgcctgagcagaagttgccagtgctaaaccgagac ctgcagctaataaccacttttt CATAATATTCCCCCTAAGTGTTAGACGTAT ATACTGccggagggtgcagttctttgcgtggcccggcgatcttatattgatcgcctaaagtc atcgctacactaccactacattcctttg
degP	agaagtctcagccgccgttgcagagagcggagataacgccaaacctaaactcagagcTagtg cactcagtgctaatgtggttttttttCACTTCCTCAATTTCGATCAGTGGCGT TACTCaaaattgctgtgttcttcagattcgttttatagcctgaagttccgaactaaagtttct gcaaaaggtaaaaatttattgttcgtctttac

APPENDIX E

Sequencing results

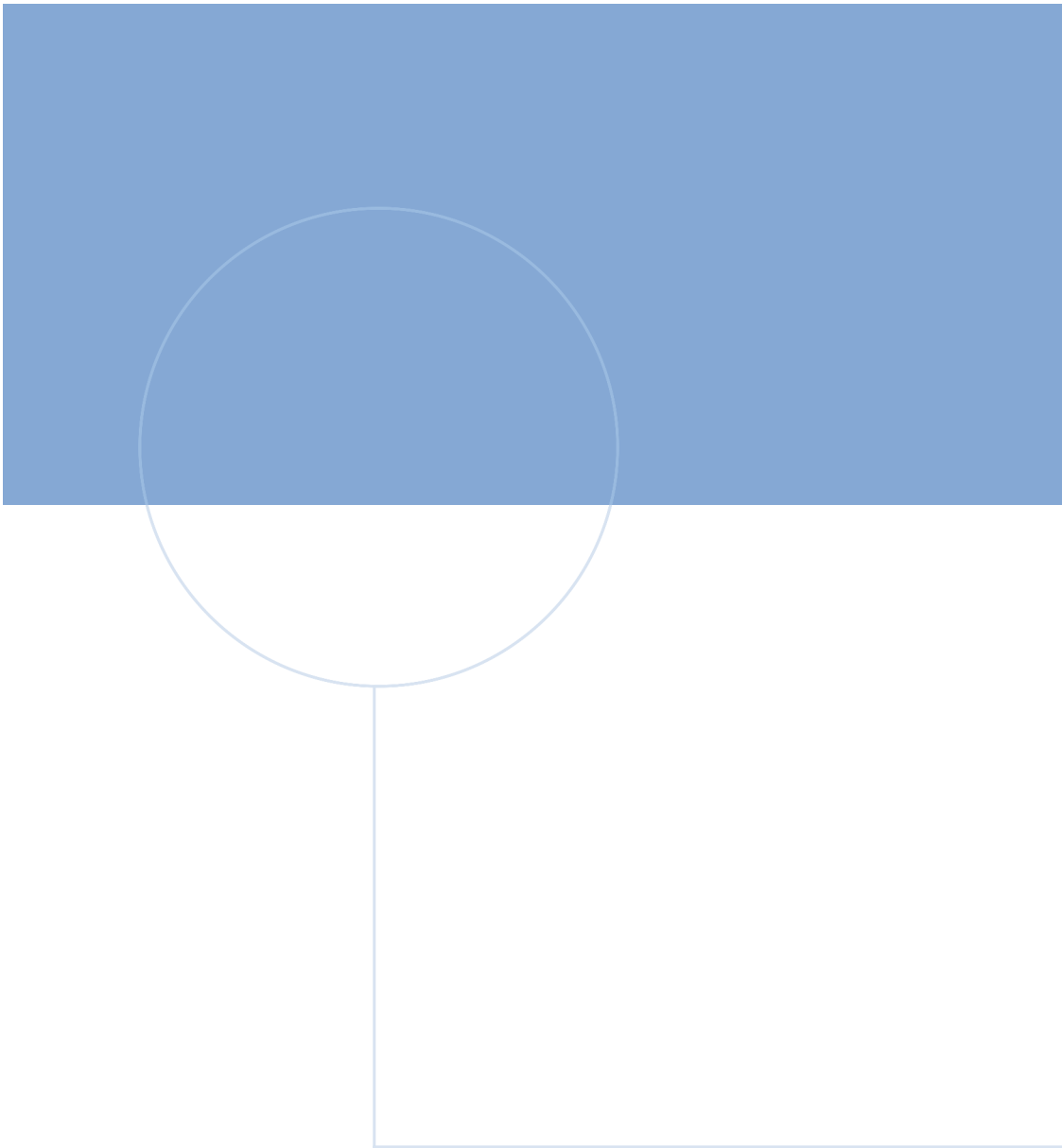
The complete genome sequence of BL21(DE3) (Genbank: CP001509.3) was used to check for secondary mutations by sequencing. A summary of the samples sent for sequencing, and the respective primers used are shown in table E.1.

Table E.1: Description of sequencing IDs. The description coloumn indicates either the plasmid that was sequenced of the strain that the sequence originates from. #: colony number. "correct" indicates that the expected sequence without mutations was obtained. Sequencing results that were too short to use are marked "short". Incorrect sequences are marked "-".

Seq. ID	Description	#	Primer	Seq.
70DI91	pTA16-s173-AP	-	pVB1_251_GFP_seq1_F	correct
70DI92	pTA16-s173-AP	-	pAG6026_backbone_trfA_R	correct
70DI87	pMAZ-SK_skp	1	pMAZ-SK_seq_F	-
70DI88	pMAZ-SK_skp	2	pMAZ-SK_seq_F	-
70DI89	pMAZ-SK_skp	3	pMAZ-SK_seq_F	correct
70DI90	pMAZ-SK_sppA	1	pMAZ-SK_seq_F	correct
70DJ53	pMAZ-SK_sppA	2	pMAZ-SK_seq_F	correct
70DJ52	pMAZ-SK_sppA	3	pMAZ-SK_seq_F	-
70DJ51	pMAZ-SK_degP	1	pMAZ-SK_seq_F	-
70DJ54	pMAZ-SK_degP	2	pMAZ-SK_seq_F	short
70DJ55	pMAZ-SK_degP	3	pMAZ-SK_seq_F	-
70DJ56	pMAZ-SK_degP	4	pMAZ-SK_seq_F	-
70DJ57	pMAZ-SK_degP	5	pMAZ-SK_seq_F	-
70DJ58	pMAZ-SK_degP	6	pMAZ-SK_seq_F	-
70DJ59	pMAZ-SK_degP	7	pMAZ-SK_seq_F	correct
70DJ60	pMAZ-SK_degP	8	pMAZ-SK_seq_F	short
70DJ61	pMAZ-SK_degP	9	pMAZ-SK_seq_F	-
70DJ03	pMAZ-SK_degP	10	pMAZ-SK_seq_F	-
70DJ04	pMAZ-SK_degP	11	pMAZ-SK_seq_F	-
70DJ05	pMAZ-SK_degP	12	pMAZ-SK_seq_F	-
70DJ06	BL21_sppA	1	sppA_seq_F	correct

Table E.1 continued

Seq. ID	Description	#	Primer	Details
70DJ07	BL21_sppA	1	sppA_seq_F	correct
70DJ08	BL21_sppA	3	sppA_seq_F	-
70DJ09	BL21_sppA	3	sppA_seq_F	correct
70DJ10	BL21_sppA	8	sppA_seq_F	correct
70DJ11	BL21_sppA	8	sppA_seq_F	correct
70DJ12	BL21_skp	4	skp_seq_F	correct
70DJ13	BL21_skp	5	skp_seq_F	correct
70DJ14	BL21_degP	3	degP_seq_F	correct
70DJ15	BL21_degP	8	degP_seq_F	correct
70DJ16	BL21_sppA+skp	8	skp_seq_F	correct
70DJ17	BL21_sppA+skp	12	skp_seq_F	short
70DJ18	BL21_sppA+skp	14	skp_seq_F	short
70DJ19	BL21_sppA+skp	15	skp_seq_F	short
70DJ20	BL21_degP+skp	1	skp_seq_R	short
70DJ21	BL21_degP+skp	2	skp_seq_R	-
70DJ22	BL21_degP+skp	11	skp_seq_R	-
70DJ30	BL21_degP+skp	17	skp_seq_F	original
70DJ31	BL21_degP+skp	18	skp_seq_F	-
70DJ23	BL21_sppA+degP	2	degP_seq_F	short
70DJ24	BL21_sppA+degP	3	degP_seq_F	short
70DJ25	BL21_sppA+degP	4	degP_seq_F	correct
70DJ26	BL21_sppA+degP	5	degP_seq_F	short
70DJ27	BL21_sppA	1	skp_seq_R	correct
70DJ28	BL21_sppA	8	skp_seq_R	correct
70DJ29	BL21	-	skp_seq_R	correct
70DJ32	BL21_skp+degP	8	degP_seq_F	correct
70DJ33	BL21_skp+degP	19	degP_seq_F	correct
70DJ34	BL21_sppA+skp+degP	8	degP_seq_F	correct
70DJ35	BL21_sppA+skp+degP	8	degP_seq_F	-



NTNU
Norwegian University of
Science and Technology