



Norwegian University of  
Science and Technology

# Effects of Sec-Pathway Modifications on Periplasmatic Translocation of Recombinant Proteins in *Escherichia coli*

**Andrea Iselin Elvheim**

Biotechnology (5 year)

Submission date: May 2018

Supervisor: Trygve Brautaset, IBT

Co-supervisor: Agnieszka Gawin, IBT  
Jostein Malmo, Vectron Biosolutions

Norwegian University of Science and Technology  
Department of Biotechnology and Food Science



## Abstract

Recombinant proteins are expressed for use in both research, industry and medicine. The Gram-negative bacteria *Escherichia coli* (*E. coli*) is often used as an expression host. It has several properties making it a suitable host, such as fast growth rates and simple protocols for genetic modifications. In addition, it is very well characterized, and has safely been used for a long time as an expression host, and as a laboratory strain. Expressing human or other eukaryotic proteins in *E. coli* can in many cases be challenging. Expression hosts of human, mammalian, or eukaryotic origin must in some cases be used instead. However, using *E. coli* or other prokaryotic hosts tend to be cheaper, simpler, and more efficient. As these proteins often are target proteins, especially as pharmaceuticals, overcoming these challenges are of interest.

One strategy for overcoming some of the challenges of using *E. coli* is translocating the recombinant proteins to periplasm, the space between the outer and inner membrane of Gram-negative bacteria. The environment of the periplasm is different from cytoplasm. This gives some advantages, such as improved folding and solubility. Downward processing is also simplified by periplasmic translocation. In *E. coli*, proteins can be translocated either post-translationally or co-translationally through the Sec-pathway, or through the Tat-pathway. In this work, components of the Sec-pathway were genetically modified, to increasing the levels of periplasmic translocation for recombinant proteins. The target genes were *secM*, *secG* and *ffh*. The expression of *secG* and *ffh* were supposed to be downregulated by changing the ribosomal binding site and the start codon of the genes. The signal peptide of *secM* were changed to increase the expression of SecA, an important component of the Sec-pathway.

CRMAGE, a combination of  $\lambda$ -Red recombineering based MAGE and CRISPR/Cas9, were used to generate the genomic modifications. Single mutants of *secM* and *secG* were successfully constructed, along with a double mutant harbouring both these mutations. A fusion protein of IgG Fc fragment and sfGFP, with a signal peptide for translocation through the Sec-pathway, were constructed as a reporter. The protein should be in an unfolded state until it has been translocated, thus the fluorescence of sfGFP could serve as a reporter of periplasmic translocation levels. The periplasmic translocation seemed to have been lowered by the introduced mutations. Slightly in the *secM* mutant strain, and more in the *secG* mutant strain and the mutant strain harbouring both mutations.

## Sammendrag

Rekombinante proteiner er uttrykt for bruk i forskning, industri og medisin. Den Gram-negative bakterien *Escherichia coli* (*E. coli*) blir ofte bruk som vertsorganisme for uttrykk av proteiner. Den har mange egenskaper som gjør den til en god vertsorganisme, blant annet høy vekstrate og enkle protokoller for genetisk modifikasjon. Den er i tillegg veldig godt karakterisert og har lenge, trygt blitt brukt som vertsorganisme for uttrykk av proteiner og som laboratoriestamme. Uttrykk av menneskelige eller andre eukaryote proteiner i *E. coli* kan i mange tilfeller være utfordrende. I slike tilfeller kan menneskelige, pattedyr-, eller eukaryotiske vertsorganismer brukes. Samtidig har bruk av *E. coli* eller andre prokaryotiske vertsorganismer en tendens til å være billigere, enklere, og mer effektivt. Fordi slike proteiner ofte er uttrykt for bruk, spesielt i farmasi, er det av interesse å løse disse problemene.

En strategi for å løse noen av problemene med å bruke *E. coli* er translokasjon av rekombinante proteiner til periplasma, rommet mellom den indre og ytre cellemembranen hos Gram-negative bakterier. Miljøet i periplasma er annerledes enn i cytoplasma. Dette gir noen fordeler, slik som forbedret folding og løselighet. Videre prosessering er også forenklet ved translokasjon til periplasma. I *E. coli* kan proteiner translokteres post-translasjonelt eller co-translasjonelt gjennom Sec-reaksjonsveien, eller gjennom Tat-reaksjonsveien. I dette arbeidet ble komponenter av Sec-reaksjonsveien genetisk modifisert for å øke mengden rekombinant protein translokert til periplasma. Målgenene var *secM*, *secG*, og *ffh*. Uttrykket av *secG* og *ffh* skulle nedreguleres ved å endre bindingssetet for ribosomer og startkodonet for genene. Signalpeptidet for *secM* ble endret for å øke uttrykket av SecA, en viktig komponent i Sec-reaksjonsveien.

CRMAGE, en kombinasjon av MAGE basert på  $\lambda$ -Red-rekombinering og CRISPR/Cas9, ble brukt for å generere de genomiske modifikasjonene. Enkeltmutanter med mutasjoner i *secM* og *secG* lyktes å bli laget, i tillegg til en dobbelmutant med begge mutasjonene. Et fusjonsprotein av IgG Fc og sfGFP, med et signalpeptid for translokasjon gjennom Sec-reaksjonsveien, ble konstruert som et reporterprotein. Proteinet burde være i en ikke-foldet tilstand til etter translokasjon, dermed bør mengden fluorescens fra sfGFP kunne brukes for å målet nivået av translokasjon til periplasma. Translokasjonen til periplasma så ut til å ha blitt senket av mutasjonene. Noe i *secM*-mutantstammen, og mer i *secG*-mutantstammen og dobbelmutanten.

# Table of Contents

1	Introduction .....	1
1.1	Periplasmic Translocation of Recombinant Proteins in <i>Escherichia coli</i> .....	2
1.1.1	Pathways for Periplasmic Translocation in <i>Escherichia coli</i> .....	3
1.1.2	Sec-Pathway Targets for Genomic Modification .....	5
1.2	The CRMAGE System for Genome Engineering .....	7
1.2.1	$\lambda$ Red Recombineering Based MAGE .....	7
1.2.2	CRISPR/Cas9 .....	9
1.2.3	CRMAGE .....	11
1.3	Reporter Protein for Measuring Translocation Levels of Recombinant Proteins .....	13
1.3.1	IgG2 Fc Fragment .....	13
1.3.2	Superfolder Green Fluorescent Protein .....	15
1.3.3	XylS/ <i>Pm</i> Promoter System .....	15
1.4	Aims of This Thesis .....	16
2	Materials and Methods .....	17
2.1	Bacterial Strains .....	17
2.2	Plasmids .....	17
2.3	Antibiotics .....	18
2.4	Media and Solutions .....	18
2.4.1	Lysogeny Broth (LB) .....	18
2.4.2	Lysogeny Agar (LA) .....	19
2.4.3	SOC-Medium .....	19
2.4.4	Phosphate Buffered Saline (PBS) .....	19
2.5	Heat-Shock Transformation .....	20
2.5.1	Preparation of RbCl Competent Cells .....	20
2.5.2	Heat-Shock Transformation of RbCl Competent Cells .....	21
2.6	Electroporation .....	21

2.7	Polymerase Chain Reaction.....	22
2.7.1	Q5 PCR .....	22
2.7.2	Taq and Colony PCR.....	23
2.7.3	CloneAmp HiFi PCR .....	24
2.8	Gel Electrophoresis.....	25
2.9	DNA Sequencing .....	25
2.10	SDS-PAGE .....	26
2.11	Western Blot .....	26
2.12	USER Cloning .....	27
2.13	Gibson Enzymatic Assembly of DNA Molecules .....	27
2.14	Designing MAGE Oligonucleotides and gRNA Sequences .....	29
2.14.1	Designing MAGE Oligonucleotides .....	29
2.14.2	Designing gRNA Sequences .....	32
2.15	Construction of pMAZ-SK with gRNA Insert.....	33
2.15.1	Annealing of gRNA .....	33
2.15.2	Amplification of pMAZ-SK Backbone.....	33
2.15.3	Assembly and cloning of pMAZ-SK with gRNA Insert.....	34
2.16	Transformation of MG1655 with pMA7CR_2.0 and pZS4Int_tetR .....	34
2.17	CRMAGE .....	34
2.17.1	Curing the pMAZ-SK Plasmid.....	35
2.17.2	Curing of pMA7CR_2.0 and pZS4Int_tetR .....	36
2.18	SDS-PAGE for Confirmation of Cas9 .....	36
2.19	Construction and Transformation of Vector with Reporter Protein .....	36
2.19.1	Design of Reporter Protein.....	36
2.19.2	Assembly of pVB1-251-kan with OmpA-IgG Fc-linker-sfGFP Insert.....	37
2.19.3	Fluorescence Test .....	37
2.20	Protein Analysis of Reporter.....	38

2.20.1	Protein Isolation .....	38
2.20.2	SDS PAGE Before Western Blot .....	38
2.20.3	Western Blot for Identification of Reporter Protein.....	39
2.21	Growth Studies.....	39
2.22	Comparison of Mutant Strains .....	39
3	Results .....	41
3.1	Establishing and Optimizing CRMAGE .....	41
3.1.1	SDS-PAGE for the Verification of Cas9 Expression.....	41
3.1.2	Tagged glyA as a Positive Control for CRMAGE.....	42
3.2	Assembly of pMAZ-SK with gRNA Targeting the Sec-Pathway.....	46
3.3	Construction of Sec-Pathway Mutant Strains.....	47
3.3.1	Mutant Strains with Single Mutations.....	47
3.3.2	Mutant Strain with Double Mutation .....	50
3.4	Assembly of Vector Expressing Reporter Protein.....	51
3.4.1	Western Blot and SDS-PAGE for Identification of the Reporter Protein .....	52
3.4.2	Growth Experiment with MG1655 Expressing the Reporter Protein .....	56
3.5	Expression of the Reporter Protein in the Mutant Strains .....	59
4	Discussion .....	62
4.1	Establishing and Optimizing CRMAGE .....	62
4.2	Frame Shift Mutations in <i>secM</i> Mutants .....	64
4.3	Higher Mutation Efficiency when Introducing Second Mutation .....	65
4.4	No <i>ffh</i> Mutants Were Created .....	66
4.5	Reliability of the Reporter Protein.....	67
4.6	Effects of the Introduced Mutations .....	68
5	Conclusion.....	71
6	References .....	72
	Appendix 1 – DNA Sequences.....	i

Primers .....	i
DNA Sequence of OmpA – IgG2 Fc – sfGFP Fusion Protein.....	ii
Appendix 2 – Sequencing Results.....	iv
Sequencing of SecG Mutant Strains.....	iv
Sequencing of SecM Mutant Strains .....	v
Sequencing of OmpA – IgG Fc – sfGFP.....	vi

# 1 Introduction

Recombinant proteins are proteins expressed from recombinant DNA, DNA that has been recombined or engineered, and thus cannot be found in nature. Chemical synthesis of proteins is often expensive, complex, and difficult. Therefore, the preferred method is within host cells. An expression vector carrying the DNA sequence for the target protein is introduced into the host, and the target protein produced (1). Several different hosts can be used, human cells, other mammalian cells, yeasts, bacteria, and plant cells (2). The hosts differ in their abilities to modify proteins post-translationally, and the environments within different cells can vary and affect the ability to properly fold and solubilize proteins. The vector, the host organism, and growth conditions must all be optimized for the most efficient production of each specific protein (1, 3). The expression of recombinant proteins in bacteria has greatly simplified and improved the production and purification of proteins needed in both research, industry and medicine (3). However, there are still many challenges to overcome in order to efficiently express every kind of proteins, and there are room for improvements and progress within the field of recombinant proteins (3).

There are many different recombinant proteins used as pharmaceuticals. Examples are hormones, growth factors, vaccines, blood factors, monoclonal antibodies, and more. Many of these have human origin. Production of mammalian proteins in prokaryotes can be problematic, as prokaryotic hosts are unable to post-translationally modify proteins the same way as human cells. Because of this, mammalian expression hosts, or in some cases yeast, must sometimes be chosen. However, when applicable, bacterial systems tend to be simple and cheap (1).

The most widely used host for expression of recombinant proteins has been the bacteria *Escherichia coli* (*E. coli*) (4). Between 2010-2014, 29 % of the biopharmaceutical products approved were expressed in *E. coli* (2). *E. coli* grows fast under laboratory conditions and can easily be genetically modified. It is simple to grow to dense cultures, and appropriate culture media is simple and inexpensive to make. It is also capable of expressing foreign proteins as a large percentage of total protein expressed (3, 4). *E. coli* has been used as a laboratory strain for a long time, and it is well characterized. Several drugs are produced in *E. coli*. It has safely been used for a long time, and it is recognized by drug regulatory authorities. The use of *E. coli* is also cost effective (1).

## **1.1 Periplasmic Translocation of Recombinant Proteins in *Escherichia coli***

Gram negative bacteria, like *E. coli*, have two cell membranes, the inner, cytoplasmic membrane, and the outer membrane. The space between these membranes is called the periplasm (5). Many of the problems of expressing recombinant proteins in the cytoplasm can be solved by translocating the proteins into the periplasm or out of the cell. Translocation to the periplasm and secretion out of the cell of recombinant proteins can both improve their biological activity and simplify downstream processing. As the proteins are located either in periplasm or the culture media, they can be more easily recovered and purified. If the target proteins are located in the culture media, they can be extracted without disrupting the cells at all. After translocation to the periplasm, the proteins can either be further secreted out to the media by the cell itself, or the outer membrane can be disrupted to facilitate release. Either way, the cytoplasmic membrane will remain intact. As *E. coli* secrete very few proteins to the culture media, and as the amount of proteins is much higher in cytoplasm than periplasm, secretion and translocation can greatly reduce contamination from host proteins. This means less effort has to be put into purifying the protein after extraction (5).

The cytoplasm of *E. coli* is a more reducing environment than the cytoplasm of eukaryotic cells. This can be a problem when expressing eukaryotic recombinant proteins, as it can prevent their correct folding. Especially the formation of disulphide bonds is affected. The *E. coli* periplasm is less reducing, and there are enzymes present which can assist the proteins in forming disulphide bonds. Unfolded proteins are also prone to aggregation. The tendency of some of the proteins to aggregate in the cytosol is also a problem that can be reduced by translocation. The protein concentration is lower in periplasm and the culture media than in cytosol. This can be a factor for improving the solubility of the proteins, thus reducing aggregation. In addition, translocation makes it possible to improve folding and reduce aggregation by methods like co-translocation of chaperones, proteins that assist in folding of other proteins. The half-life of recombinant proteins can also be extended by secretion and translocation as there are less proteases present in periplasm and media than in cytoplasm (5).

In addition, the process of translocation can be used as a method to remove the initial methionine from the N-terminal of recombinant proteins where it is normally not present (5). The initial methionine is added for the most common start codon, AUG. The initial methionine

is removed in many proteins. However, this does not always happen when a protein is expressed recombinantly (6). The addition of the initial methionine can change the properties of the protein, for example reducing the activity and stability. If the protein is used therapeutically and is supposed to be introduced into a human or animal, the additional methionine may make the protein sufficiently different from the native form that it binds antibodies. Hence, the immune system is activated and breaks down the protein. After translocation, the N-terminal signal sequence are cleaved off by a membrane associated protease, thus the initial methionine is removed as well (5).

### 1.1.1 Pathways for Periplasmic Translocation in *Escherichia coli*

The general secretion (Sec) pathway and the twin arginine translocation (Tat) pathway are the two major pathways for translocation of proteins across the cytoplasmic membrane in bacteria (7). Thus, in *E. coli*, proteins translocated through these pathways are released in periplasm (5). The Sec-pathway translocates unfolded proteins, either co-translational or post-translational, through two different mechanisms. The Tat-pathway translocates folded proteins (7). Proteins are targeted for translocation by the different translocation pathways based on signal peptides on the N-terminal end of the proteins (7). After translocation through the Sec-translocase the signal peptide can be cleaved off. Translocation without cleavage can happen, but the protein will then be bound to the membrane through the signal peptide (8).

In the post-translational mechanism of the Sec-pathway translated proteins are kept unfolded and translocation-competent by the specific chaperone SecB or general chaperones (7). SecB is only found in proteobacteria, such as *E. coli* (9). The protein is delivered to SecA which drive the protein through a protein conducting channel, the SecYEG complex. SecA is an ATPase and it is the motor protein of the Sec system. The SecYEG complex consist of SecY, SecE and SecG. SecY and SecE forms the main part of the channel, while secG is only peripherally connected (7).

In the co-translational mechanism, or signal recognition particle (SRP) pathway, SRP binds the protein during translation. SRP in *E. coli* consist of two components a 4.5S RNA and a polypeptide called fifty-four homologue (Ffh). Ffh is homologous to SRP54, which is a

component of SRP in the endoplasmic reticulum (8). The nascent protein chain is transferred by SRP to the SecYEG complex with help from a SRP receptor (SR or FtsY). It is then translocated, simultaneously to being translated. The SRP mechanism is mainly used for integration of membrane proteins, but it is also used for translocation of proteins into periplasm (7). A representation of the SecA- and SRP pathway are shown in Figure 1.1.

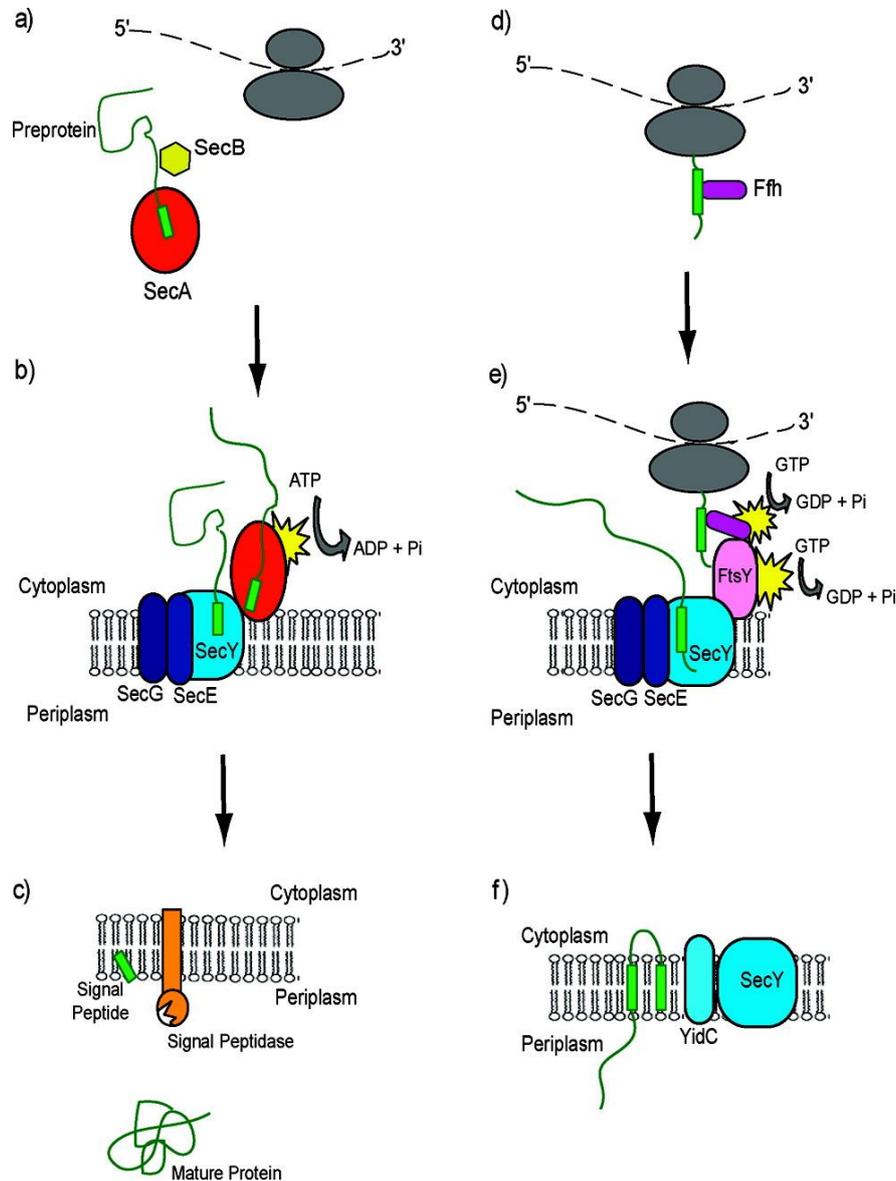


Figure 1.1: The post-translational pathway of the Sec-pathway are shown in a, b and c. The preprotein is kept unfolded by SecB after translation and targeted to SecA (a). The protein is translocated by the SecYEG complex (b). In periplasm the signal peptide is cleaved off by signal peptidase and the protein is allowed to fold into mature protein (c). The co-translational pathway, or SRP pathway of the Sec-pathway are shown in d, e, f. The SRP particle, containing Ffh, recognises the nascent peptide chain (d). SRP binds the SRP receptor FtsY, and the protein is translocated by the SecYEG complex (e). Most proteins using the SRP pathway are incorporated into the membrane. YidC is involved in incorporating the proteins (f) (8).

### 1.1.2 Sec-Pathway Targets for Genomic Modification

As the co-translational pathway is mainly used for integration of membrane proteins, the majority of recombinant proteins are translocated through the post-translational pathway (5). Thus, making the post-translational pathway more efficient could be used to make the translocation of many recombinantly expressed proteins more efficient. In this thesis, components of the Sec-pathway were genetically modified, with the goal of affecting the translocation of proteins, to make the post-translational pathway more effective. The target genes were *ffh*, *secM*, and *secG*. As previously mentioned, Ffh is the peptide component of SRP in *E. coli*. The SRP initiates translocation through the SRP pathway, by binding to the nascent protein chain (8). Ffh is vital for the cell, and *ffh* knockout mutants have been shown to non-viable (10).

SecG is a non-essential component of the Sec-translocon, and knock-out mutations have been shown not to cause severe growth or transport defects in several wild type (WT) strains of *E. coli*, under standard laboratory conditions. However, in combination with other mutations that compromise the translocation machinery *secG* knock-outs could be deleterious (11). Mutations in *secG* has also been shown to affect the transport of mutant signal sequences more than the transport of wild type signal sequences (12). SecG is thought to be involved in stimulation of translocation of proteins by regulating the membrane cycling of *secA* (13), and it is thought to be important for the stability of SecYE (12).

The expression of the ATPase motor SecA is regulated by *secM* (secretion monitor). SecA is a motor protein for translocation through the Sec-pathway. It powers the translocation of proteins, using ATP (5). The genes *secM* and *secA* are in the same operon and are transcribed together as one single mRNA. When *secM* are translated, there is a translational pause. This pause prevents the formation of secondary structures that blocks translation of *secA*. (14). A stem-loop structure can be formed in the region between the *secM* and *secA* gene on the RNA, blocking the ribosome from binding to the Shine-Dalgarno sequence of *secA*, and thus blocking translation of this gene. During the translational pause, the Shine Dalgarno sequence is exposed and SecA mRNA can be translated (15). The translation of *secM* is continued when the protein chain interacts with the translocation machinery of the Sec-pathway. The polypeptide chain is likely released due to mechanical force (16).

By these mechanisms, the expression of SecA is under feedback regulation. If the translocation of proteins, among them SecM, is inefficient, more SecA is expressed to increase the efficiency. When the translocation of secM is efficient, less secA is expressed. The expression of SecA is thus fine-tuned, depending on the needs of the cell (14). As translation of SecA is only possible during this translational pause, an extended pause could possibly increase the translation of SecA. After translocation into periplasm, SecM is rapidly degraded by proteases (14). The release of SecA from the membrane has been reported as a rate-limiting step for periplasmic translocation. Therefore, increased expression could increase the levels of translocation.

The *ffh* and *secG* genes were supposed to be downregulated by changing the ribosomal binding sites (RBS) and the start codons. The *secG* gene has been proven to be non-essential (11). It was therefore chosen as a target, to study the effect of its lowered availability on the performance of the Sec-pathway. *Ffh* was chosen as Ffh is a vital part of the SRP-pathway. The co-translational and the post-translational pathway converge at one translocon, comprised of the motorprotein SecA and the SecYEG complex. As the number of translocons are limited, downregulating the SRP pathway might make more translocons available to the post-translational pathway, making it more effective. As Ffh is essential to the cell, the aim was to only downregulate the expression. Subsequently, one region of the signal sequence of *secM* was substituted with a regular signal sequence directing SecM to the SRP pathway. The goal was to upregulate SecA synthesis by slightly increasing the translational pause of SecM translation.

After creating single mutants, double mutants and a triple mutant were supposed to be created. The combined effect of all three mutations were considered likely to be greater than each single mutation. Especially, the double mutation with downregulation of the SRP pathway, combined with the substituted signal peptide of SecM, were expected to give cumulative effects. The genomic modifications made are summarized in Table 1.1.

Table 1.1: Summary of the genomic modifications planned for this thesis, and the goal behind them.

Target	Modification	Goal
<i>ffh</i>	Downregulation: Change of ribosomal binding site and start codon.	Downregulate the SRP pathway, to promote translocation through the post-translational pathway
<i>secG</i>	Downregulation: Change of ribosomal binding site and start codon.	Downregulate a non-vital component of the SecYEG translocon, to assess the effect
<i>secM</i>	Change of signal sequence.	Upregulating the expression of <i>secA</i> , by downregulating the translocation of <i>secM</i> , and/or increasing the length of its translational pause

## 1.2 The CRMAGE System for Genome Engineering

The target mutations were introduced using CRISPR/Cas9 optimized MAGE recombineering (CRMAGE). The method combines  $\lambda$  Red recombineering based MAGE and CRISPR/Cas9 for a fast and simple way of genome editing (17).

### 1.2.1 $\lambda$ Red Recombineering Based MAGE

The  $\lambda$  Red system is a system for homologous recombineering, adapted from the  $\lambda$ -bacteriophage. It allows genomic modifications, both mismatches, insertions, or deletions, to be made, using an oligo of double-stranded DNA (dsDNA) or single-stranded DNA (ssDNA) in combination with  $\lambda$  Red proteins (Figure 1.2). When ssDNA oligos are used, only one protein is needed: the  $\lambda$  Red  $\beta$ -protein.  $\lambda$  Red  $\beta$ -protein binds ssDNA and hybridises it to complementary DNA, thus facilitating homologous recombineering (18).

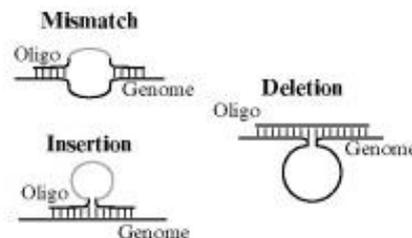


Figure 1.2: Illustration of how different genomic modifications can be created using the  $\lambda$  Red system. Adapted from (18)

For recombineering using ssDNA oligos, the most efficient approach is designing the oligos to hybridize with the lagging strand during DNA replication. The DNA polymerase can only synthesize DNA in the 5' to 3' direction, thus one strand is synthesized coherently, called the leading strand, while the other is synthesized as many small fragments, called the lagging strand. These small fragments are called Okazaki fragments (19). When the ssDNA hybridize to the lagging strand it is able to be incorporated into the genome in a way resembling the ligation of the Okazaki fragments. Oligos hybridizing to the leading strand can also be incorporated into the genome, but with a significantly lower efficiency (18, 20). In *E. coli* DNA replication happens in two directions, creating two replichores, thus extra care had to be taken when designing oligos, in order to target the lagging strand (Figure 1.3) (18).

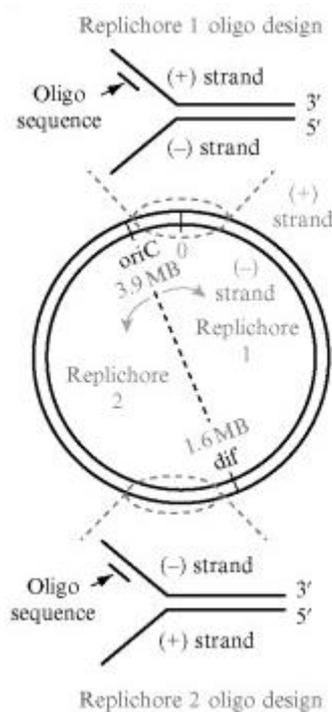


Figure 1.3: Genome map of *E. coli* showing the origin of replication (*oriC*) and the two replichores, with illustrations of how oligos must be designed to hybridize to the lagging strand, depending on the location of the target. Adapted from (18)

The efficiency of  $\lambda$  Red recombineering can be low and is dependent on several factors. Larger mutations and smaller homology regions give lower efficiency, while at the same time larger oligos can give lower efficiency as they more often form secondary structures which can inhibit their binding to DNA, and thus, recombineering. The efficiency can be increased by running several rounds of recombineering on the same cell population. The number of mutants in the

population is then increased with each cycle. Also, introduction of phenotypic markers can be done, like antibiotic resistance, which allows for effective selection of mutant strains. However, such markers will often have to be removed, which can leave scars in the genome. Additionally, such markers can often not be used if several mutation is introduced in the same strain (18).

Another factor affecting the efficiency of the method, is repair mechanisms of the cell. The methyl-directed mismatch repair (MMR) system is a system in found in *E. coli* for proofreading of the DNA sequence after DNA replication. The MMR system is capable of repairing mutations introduced by  $\lambda$  Red recombineering, thus making the method less efficient. It has been shown that knock out mutations of components in the MMR system can decrease its abilities to make such repairs. Such knock-outs can therefore be combined with  $\lambda$ -red recombineering in order to increase its efficiency (20).

The  $\lambda$  Red system has been used for Multiplex Automated Genome Engineering (MAGE), where several genomic modifications can be introduced simultaneously. In MAGE a mix of different oligos are transformed into the cell at the same time, and to increase the efficiency, several cycles of recombineering is run. Thus the abundance of mutants are increased every round (18, 21). However, running several rounds can be laborious, time consuming, and expensive. An improved version of MAGE, employing CRISPR/Cas9 for negative selection, has the potential to enhance  $\lambda$  Red recombineering, and allow efficient multiplexing, of at least two mutations, in one single round of electroporation, without leaving any scars in the genome (17).

### 1.2.2 CRISPR/Cas9

Clustered regularly interspaced short palindromic repeats (CRISPR) are genomic regions found in many archaea and bacteria. They consist of several palindromic repeated sequences which separates short variable regions, called spacers (22). The spacers in the CRISPR locus have been found to originate from for example bacteriophages or foreign plasmids, and the CRISPR locus, in combination with CRISPR-associated (Cas) genes, have been found to be systems for adaptive immune responses in prokaryotes. The cell can incorporate phage DNA, and other foreign DNA it encounters, into one end of its CRISPR locus. Later, this DNA sequence can be

used for defence if the bacteria faces the same threat again (23). The CRISPR/Cas-systems protects the cell by causing degradation of foreign nucleic acids, resulting from nuclease activity in some Cas-proteins (24).

There are three stages of the CRISPR defence system. First, spacer acquisition, or adaptation, where the target sequences, called protospacer, are incorporated into the CRISPR locus. The second stage is CRISPR RNA (crRNA) biogenesis, where pre-crRNA is transcribed and processed into mature crRNA. The last stage is interference where Cas-proteins targets protospacers, guided by the crRNA, and then causes interference of the phage or plasmid (24).

There are so far 6 classified types of CRISPR-systems: I-VI, distinguished by signature Cas genes associated with each system. There are also several subtypes of each type where the Cas genes are slightly varied. The 6 types are sorted into two classes. Class 1 contains type I, III and IV, and are characterized by having multi-subunit effectors, made up of several Cas proteins. Class 2 contains type II, V and VI, and are characterized by having single unit effectors, consisting of only one protein (24).

CRISPR/Cas9 is a type II CRISPR system with Cas9 as the effector. Cas9 is dependent on two different RNAs: crRNA and trans-activating crRNA (tracrRNA), which is necessary for both maturation of crRNA and for cleavage of DNA. The crRNA and the tracrRNA are bound together by pairing of the tracrRNA to the repeat sequence of the crRNA. The RNAs binds to Cas9, guide it to the correct sequence and facilitates a double stranded break of the DNA sequence. Both RNAs are essential for nuclease activity (25). Cas9 has two subunits with nuclease activity, HNH and RuvC, which cut one DNA strand each (24) For laboratory use, the crRNA and the tracrRNA can be designed as one chimeric single sequence, a synthetic single guide RNA (sgRNA) (Figure 1.4). The sgRNA contains the tracrRNA sequence linked to a crRNA sequence (gRNA) of choice with a linker that allows base pairing between the two sequences and secondary structures as in the natural system. The sequences of the crRNA and tracrRNA could also be shortened in the sgRNA without the complex losing the ability to make double stranded breaks (25).

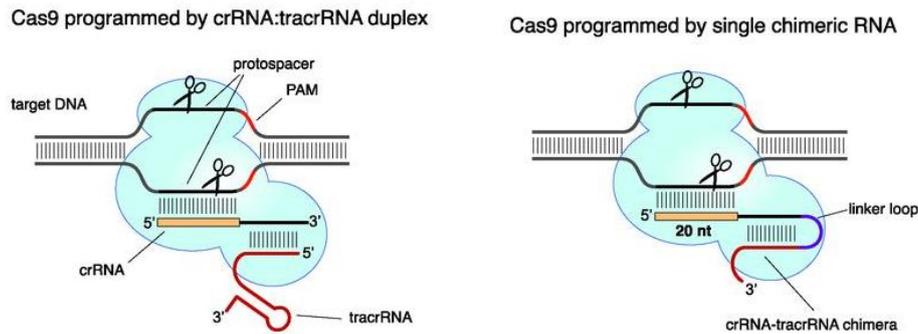


Figure 1.4: Representation of Cas9 with crRNA:tracrRNA duplex and with the crRNA-tracrRNA chimera, or sgRNA. The two RNAs are linked together with a linker loop that allows them to keep the natural conformation. Cas9 is guided by the RNAs and the PAM sequence and make a double stranded break at the protospacer (25).

In addition to a sequence complimentary to the crRNA, to make a double stranded break Cas9 needs a protospacer adjacent motif (PAM site). The PAM site is not present in the CRISPR locus. Thus, Cas9 does not target the CRISPR locus and leaves the genome of the cell uncut. The PAM site seems to be important both for binding of the Cas9-tracrRNA:crRNA complex and also for allowing Cas9 to cut after binding. The PAM site of the CRISPR/Cas9 system of *Streptococcus pyogenes* (*S. pyogenes*), the system used in this thesis, is NGG (25). The CRISPR/Cas9 system can be used to make double stranded breaks in any specific location with a PAM site, thus CRISPR/Cas9 can be used to target almost any location (26).

### 1.2.3 CRMAGE

CRMAGE is a method of genome engineering in which CRISPR/Cas9 and  $\lambda$  Red recombineering are combined. The genomic modification is introduced using  $\lambda$  Red recombineering, and the CRISPR/Cas9 system is used for negative selection against the wild type sequences. The MAGE oligos are designed to incorporate the desired mutation, while simultaneously disrupt the PAM-site targeted by the CRISPR/Cas9 system. Thus, mutants are not targeted by Cas9, while a DSB (double stranded break) should be mediated in wild type cells, causing cell death (17). A representation of the CRMAGE system is shown in Figure 1.5.

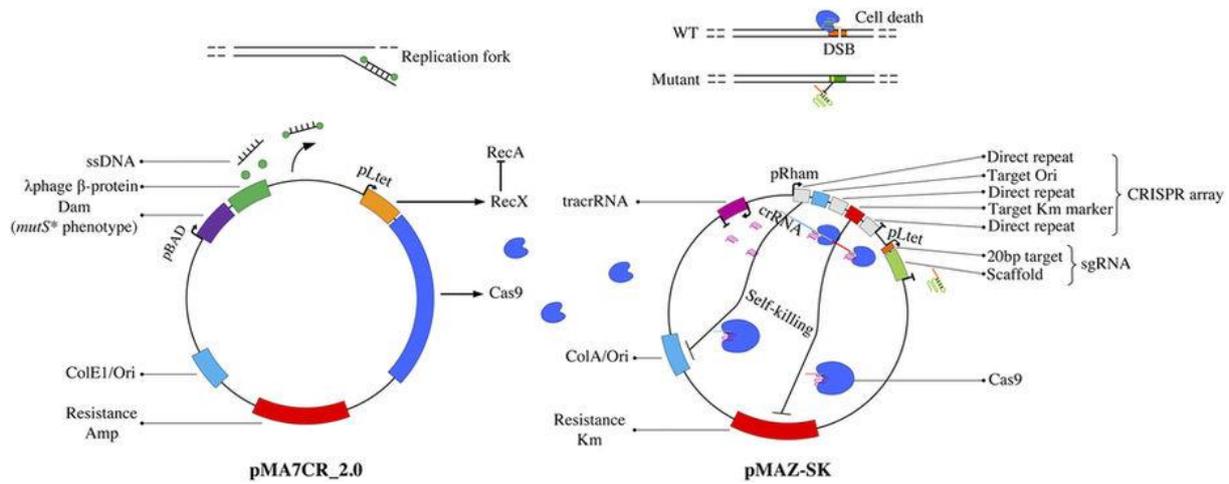


Figure 1.5: The main plasmids of the CRMAGE method, shown together with a representation of genomic modification by  $\lambda$  Red recombineering, and negative selection by CRISPR/Cas9. The promoter pLtet is induced by aTet, pBAD is induced by L-Arabinose, and pRham is induced by L-Rhamnose. ColE1 is the origin of replication in pMA7CR\_2.0 and ColA is the origin of replication in pMAZ-SK (17).

The CRMAGE system consist of two main plasmids, pMA7CR\_2.0 and pMAZ-SK. The  $\lambda$  Red  $\beta$ -proteins and Cas9 are expressed by pMA7CR\_2.0, and an sgRNA are expressed by pMAZ-SK. The expression of  $\lambda$  Red  $\beta$ -proteins are induced by L-Arabinose, and the expression of Cas9 and the sgRNA are induced by aTetracycline (aTet). L-Arabinose also induces the overexpression of Dam methyltransferase, in order to increase the efficiency of the  $\lambda$  Red recombineering. Overexpression of the *Dam* gene has been shown to result in a mutator (*mutS*<sup>-</sup>) phenotype (17). MutS is one of the components of the MMR system, and knock out mutations has previously been shown to increase the efficiency of  $\lambda$  Red recombineering (20). The efficiency of the negative selection by CRISPR/Cas9 is increased by the expression of recX, induced by aTetracycline. RecX inhibits the activity of recA, a major component of the DNA repairsystem. Inhibition of recA should therefore make the cells more vulnerable to damage by DSBs (17).

The pMAZ-SK plasmid is created with a system for self-destruction, based on CRISPR/Cas9. DSBs are mediated in two positions of the plasmid backbone, the origin of replication and the kanamycin antibiotic marker, and the plasmid is thereby destroyed. In the self-killing system Cas9 is guided by a crRNA:tracrRNA duplex. The tracrRNA is expressed constitutively while

the expression of crRNA is induced by L-Rhamnose, and aTetracycline induce expression of Cas9 (17).

In addition to the two main plasmids of the CRMAGE system, a third plasmid is needed; pZS4Int-tetR. This plasmid expresses TetR repressor, which is important for the control of the expression of Cas9 and sgRNA. The pLtet promoter provides constitutive expression, but is repressed by the TetR repressor. The TetR repressor is inhibited by aTet. Thus, aTet induces the expression of Cas9 and sgRNA.(17).

### **1.3 Reporter Protein for Measuring Translocation Levels of Recombinant Proteins**

In order to measure the effect of the introduced mutations a reporter protein had to be constructed. The reporter protein was designed as a fusion protein between an industry relevant protein and an easily detectable fluorescent protein. Superfolder green fluorescent protein (sfGFP) were fused to human IgG2 Fc fragment. IgG2 Fc was chosen as it has industrial value to Vectron Biosolutions (personal communication with Jostein Malmo, Vectron Biosolutions). A linker was placed between the coding sequences of the proteins in order to retain the natural folding of the protein components, and the signal sequence of *ompA* was placed at the N-terminal. The *ompA* signal sequence mediates protein translocation through the post-translational mechanism of the Sec-pathway (27). Thus, the reporter protein should be transported into periplasm. The sfGFP was expected to be kept unfolded before translocation by the SecB chaperone. As GFP do not fluoresce unless it is properly folded (28), the amount of fluorescence was proposed to work as an indication of the amount of translocated reporter protein.

#### **1.3.1 IgG2 Fc Fragment**

Antibodies consist of four polypeptide chains, two light chains and two heavy chains. Each light chain is bound to one heavy chain, and the two heavy chains are bound together, forming the characteristic y-shaped form of antibodies (Figure 1.6). All these bonds are disulphide bonds. The tips of the Y are formed by one light chain and one heavy chain together. This is called the variable region, and is responsible for binding to antigens. Thus, this region varies for each specific antigen. The base of the y is formed by the two heavy chains. This region is

called the constant region, and it is identical for all antibodies of the same class. The region is also called Fc region, because the fragment is crystallisable. The Fc region can be bound by Fc receptors and are responsible for the effector activity of the antibodies (29).

In the reporter protein, the Fc fragment from the human IgG2 were used (Figure 1.6). The sfGFP sequence were fused to the sequence of the Fc region of one heavy chain. The heavy chains of IgG2 are bound together with four disulphide bonds (29). As the Fc region is responsible for the dimerization of antibodies the reporter protein is likely to form dimers, if it can form disulphide bonds. IgG antibodies can induce several functions of the immune system. Examples are binding Fc receptors on macrophages to induce phagocytosis, fixation of the complement system and inducing antibody dependent cell-mediated cytotoxicity (29).

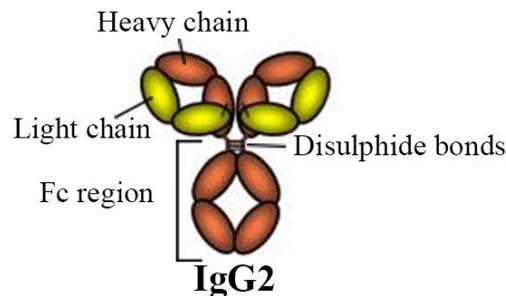


Figure 1.6: Illustration of an IgG2 antibody, showing how one light chains binds to one heavy chain, through one disulphide bond, and how two heavy chains are bound together with four disulphate bonds. The Fc region, responsible for inducing effector functions is marked. Adapted from (30).

Fc fragments used in fusion proteins has therapeutic applications. Fc fragments can be fused to relevant therapeutic proteins, such as receptors, ligands, enzymes, cytokines, and others (31). Fusion with an Fc fragment can increase the serum half-life of protein drugs, due to binding to the neonatal Fc receptor. It can also improve the stability and solubility of the drug. The Fc fragment can also be used to control the immunogenicity of the fusion partner, both by activating and inhibiting the immune system (32). Bound in a fusion protein, the Fc fragment can still mediate its effector functions(31). Fc fusion proteins have been used in treatment of diseases like HIV, cancer, and rheumatoid arthritis (31, 33). Fc fusions has been shown capable of drug delivery through the placenta to the fetus, and for transport across mucosal surfaces (32).

### 1.3.2 Superfolder Green Fluorescent Protein

Green fluorescent protein (GFP) is a fluorescent protein originally discovered in the jellyfish *Aequorea*. GFP does not require any enzymes specific to the jellyfish in order to fluoresce. Therefore, the *gfp* gene alone can be cloned into other species and will give fluorescence as long as it is able to properly fold (34). Wild type GFP misfolds in *E. coli*, and even variants with improved folding capacities tends to misfold and exhibit reduced fluorescence when expressed in fusion with other proteins instead of alone. The superfolder GFP (sfGFP) is mutated to increase the folding of GFP when it is part of a fusion protein. It was created by Pédelacq et al. by choosing the GFP variants with strongest fluorescence out of a library of GFP variants fused to a poorly folded polypeptide. The folding and fluorescence of sfGFP is not affected by misfolding of its fused protein. Thus, the fluorescence is proportional to the expression of the fusion protein, not the folding of its fusion partner (28). The theoretical molecular weight of sfGFP is 26.9 kDa. The excitation of sfGFP peaks at 488 nm, and the emission peaks at 510 nm (28).

### 1.3.3 XylS/*Pm* Promoter System

The the IgG-sfGFP fusion protein was expressed using pVB1-251, kindly provided by Vectron Biosolutions. The pVB1-251 plasmid contains an XylS/*Pm* expression cassette, which consist of the *Pm* promoter, the *xylS* gene, and *Ps2* promoter. The fusion protein was put under the regulation of the *Pm* promoter, which is regulated by the XylS transcription factor. XylS is constitutively expressed at low levels from the promoter *Ps2*. When benzoic acid derivatives, like m-toluic acid, are added XylS forms dimers and activates the *Pm* promoter. Benzoic acid derivatives that can be used as inducers are typically cheap, non-toxic, and diffuse passively into the cells. The level of expression from *Pm* is correlated with the concentration of inducer (35). The properties of the XylS/*Pm* system make it a good system for industrial production of recombinant proteins. The system was chosen in order to test the properties of the mutant under industry relevant conditions.

The pVB1-251 plasmid also contains *kan*, giving resistance to the antibiotic kanamycin, and *trfA-251*, regulating the copy number of the plasmid. In *trfA-251* (cop251M) amino acid 251 is switched with a methionine. The gene *trfA* is important for replication of the plasmid, and this mutation gives the plasmid a higher copy number (36). A high copy number (~40 copies per

cell) was used to have a higher gene expression of the reporter. For industrial purposes regulating the copy number is important to be able to optimize the level of gene expression.

#### **1.4 Aims of This Thesis**

The aims of this thesis have been mentioned, but can be summarized to:

1. Establishing and optimizing the CRMAGE method at our laboratory.
2. Using CRMAGE to generate mutant strains of *E. coli* MG1655, with modified Sec-pathway components. The mutations were intended to increase the levels of periplasmic translocation of recombinant proteins, by improving the capacity of the Sec translocon. The genomic targets were *secM*, *secG*, and *ffh*.
3. Constructing a reporter for measuring how the mutations affect periplasmic translocation in MG1655. The reporter was designed as a fusion protein of IgG2 Fc and sfGFP. It was designed with the signal peptide of OmpA in order to be translocated through the post-translational pathway.

## 2 Materials and Methods

### 2.1 Bacterial Strains

In this work, *E. coli* DH5 $\alpha$ , was used for cloning purposes. DH5 $\alpha$  has several properties making it suitable for cloning. Some of them are: 1) highly efficient transformation can be achieved; 2) it does not produce nonspecific endonuclease (*endA1*), causing it to yield plasmid DNA of high quality; 3) it has low levels of homologous recombination (*recA1*), allowing it to maintain plasmids stable (37). *E. coli* MG1655 was the strain mainly used in this work. It is often used as a *E. coli* K-12 WT-strain, as it has very few genetic modifications. The F plasmid and the  $\lambda$  prophage has been removed, and other than that it has only obtained a few additional mutations (38). BW25113 was used for parts of the optimization of CRMAGE, as the method will be used on this strain later at our laboratory in other works. All three, DH5 $\alpha$ , BW25113, and MG1655 are *E. coli* K-12 derivatives (37-39).

### 2.2 Plasmids

The plasmids used in the work with this thesis are presented in Table 2.1.

Table 2.1: The plasmids used, with their properties and references.

Plasmid	Properties	References
pMA7CR_2.0	Expresses Cas9 and RecX, induced by aTet, expresses $\lambda$ Red $\beta$ -proteins and Dam methyltransferase, induced by L-Arabinose. Gives resistance to Amp.	(17)
pZS4Int_tetR	Represses the expression of Cas9 and sgRNA, by expressing TetR repressor. Gives resistance to Spc.	(40)
pMAZ-SK	Expresses sgRNA, induced by aTet, and contains a self-killing system, induced by L-Rhamnose. Gives resistance to Kan.	(17)
pMAZ-SK_glyA	pMAZ-SK with gRNA targeting <i>glyA</i>	This work
pMAZ-SK_ffh	pMAZ-SK with gRNA targeting <i>ffh</i>	This work
pMAZ-SK_secM	pMAZ-SK with gRNA targeting <i>secM</i>	This work
pMAZ-SK_secG	pMAZ-SK with gRNA targeting <i>secG</i>	This work

pVB1-251-kan	Expression vector with the XylS/Pm system and Kan resistance	Vectron Biosolutions (unpublished)
pAE001	pVB1-251-kan expressing the ompA-IgG Fc-linker-sfGFP fusion protein, induced by m-toluic acid. Gives resistance to Kan	This work

### 2.3 Antibiotics

Antibiotics was used in media and agar plates for selection of bacteria with the required plasmids. Table 2.2 shows the antibiotics used and their concentration.

Table 2.2: The antibiotics used, their concentration in the stock solutions made, and the final concentration used in media and agar plates.

Antibiotic	Stock concentration	Final concentration
Ampicillin (Amp)	50 mg/mL	100 µg/mL
Kanamycin (Kan)	50 mg/mL	50 µg/mL
Streptomycin (Spc)	50 mg/mL	100 µg/mL

### 2.4 Media and Solutions

Components of media and solutions made and used for several purposes are described in this section.

#### 2.4.1 Lysogeny Broth (LB)

LB was used as growth media for all strains used in this work. The compounds shown in Table 2.3 were solved in water and the medium was sterilized by autoclaving.

Table 2.3: The components of LB, and their concentration.

Compound	Concentration
Tryptone	10 g/L
Yeast extract	5 g/L
NaCl	5 g/L

#### 2.4.2 Lysogeny Agar (LA)

For the preparation of agar plates, 15 g/L agar was added to LB before autoclaving. Then the appropriate antibiotics were added, and the LA was poured into petri dishes before it solidified.

#### 2.4.3 SOC-Medium

SOC-medium was used for recovery of the cells after heat shock transformation and electroporation. The compounds in Table 2.4 were solved in water, the medium was sterilized by filtration, and stored in 1.5 mL tubes at -20°C.

Table 2.4: The components of SOC-medium, and their concentration.

<b>Compound</b>	<b>Concentration</b>
Tryptone	20 g/L
Yeast extract	5 g/L
NaCl	0.5 g/L
KCl	2.5 mM
Glucose	3.6 g/L
MgCl <sub>2</sub>	5.08 g/L

#### 2.4.4 Phosphate Buffered Saline (PBS)

PBS was used for resuspension of cells before fluorescence measurements. The compounds in Table 2.5 were solved in water and the pH was adjusted to 7.4 using HCl. The buffer was adjusted to the correct volume, and then sterilized by autoclaving.

Table 2.5: The components of PBS, and their concentration.

<b>Compound</b>	<b>Concentration</b>
NaCl	8 g/L
KCl	0.2 g/L
Na <sub>2</sub> HPO <sub>4</sub>	1,44 g/L
KH <sub>2</sub> PO <sub>4</sub>	0.24 g/L

## 2.5 Heat-Shock Transformation

Transformation is the uptake of free DNA by a cell. Cells that are able to be transformed are competent. Cells can be made competent by treatment with  $\text{Ca}^+$ , followed by cooling (41), and then transformed by incubating the cells with free DNA, on ice, followed by a heat shock. Heat shock transformation was used for transformation of DH5 $\alpha$  after Gibson assembly.

### 2.5.1 Preparation of RbCl Competent Cells

First, 1% were inoculated in 50 mL Psi-medium (Table 2.6) from an overnight culture of DH5 $\alpha$ . The cells were grown until the optical density at 600 nm ( $\text{OD}_{600}$ ) was approximately 0.4. Then, the cells were harvested by centrifugation (6500 rcf, 5 minutes, 4 °C) and resuspended in 20 mL cold TFB1 (Table 2.7). After 5 minutes of incubation on ice, the cells were again harvested by centrifugation (6500 rcf, 5 minutes, 4 °C), before they were resuspended in 1.5 mL of cold TFB2 (Table 2.8). The cell suspension was immediately freezed in liquid nitrogen as 100  $\mu\text{L}$  aliquots in Eppendorf tubes, and stored at -80 °C.

Table 2.6: The components of Psi-medium. The compounds were solved in water. The medium was pH adjusted to pH 7.6 using KOH, and sterilized by autoclaving.

<b>Compound</b>	<b>Concentration</b>
Tryptone	20 g/L
Yeast extract	5 g/L
MgSO <sub>4</sub>	5 g/L

Table 2.7: The components of TFB1. The compounds were solved in water. The buffer was pH adjusted to pH 5.8 using dilute acetic acid, and sterilized by filtration.

<b>Compound</b>	<b>Concentration</b>
KAc	30 mM
RbCl	100 mM
CaCl <sub>2</sub> ·2H <sub>2</sub> O	10 mM
MnCl <sub>2</sub> ·4H <sub>2</sub> O	50 mM
Glycerol	15 % (v/v)

Table 2.8: The components of TFB2. The compounds were solved in water. The buffer was pH adjusted to pH 6.5 using dilute NaOH, and sterilized by filtration.

<b>Compound</b>	<b>Concentration</b>
MOPS	10 mM
CaCl <sub>2</sub> ·2H <sub>2</sub> O	75 mM
RbCl	10 mM
Glycerol	15 % (v/v)

### 2.5.2 Heat-Shock Transformation of RbCl Competent Cells

The cells were first thawed on ice. Up to 10  $\mu$ L of DNA (50-100 ng) was added to 100  $\mu$ L of competent cells, and the mix were incubated on ice for 30-60 minutes. The cells were heat-shocked at 37°C for 1-2 minutes. Then, after 2 minutes of incubation on ice, 900  $\mu$ L of SOC at 37°C were added. After 1-2 h incubation at 37°C with shaking, the cells were plated on LA plates appropriate antibiotics.

## 2.6 Electroporation

Electroporation is a method for introducing DNA into cells. After being made electrocompetent, the cells are given a brief electric shock which makes the cell membrane permeable and allows free DNA to enter the cells (19). Electroporation was used for transformation of strains other than DH5 $\alpha$ .

First, electrocompetent cells were made. A 1% culture was inoculated, from an overnight culture, in 10 mL LB. It was incubated at 37 °C until OD<sub>600</sub> was approximately 0.45. After incubation, 1.5 mL of the cell suspension was aliquoted into Eppendorf tubes. The cells were harvested by centrifugation (12 000 rcf, 1 minute, 4 °C) and resuspended in 1 mL cold sucrose (300 mM). The cells were harvested by centrifugation again (12 000 rcf, 1 minute, 4 °C), and then resuspended in 50  $\mu$ L cold sucrose (300 mM). The cells were kept on ice during the whole procedure, and where used for electroporation within 15 minutes.

Maximum 3  $\mu$ L plasmid DNA (100-500 ng) was added to the cells. The mixture was added to electroporation cuvettes and electroporated using a Gene Pulser Xcell from Bio-Rad. The pre-

set program for electroporation of *E. coli* was used; 1.8 kV for 1 mm cuvettes or 2.5 kV for 2 mm cuvettes. Immediately after pulsing, 1 mL of 37 °C SOC was added. The cells were incubated at 37 °C with shaking at 225 rpm for 1 h, and then plated on LA plates with appropriate antibiotics.

## **2.7 Polymerase Chain Reaction**

Polymerase Chain Reaction (PCR) is a method for amplifying DNA sequences. PCR is a DNA replication reaction performed in vitro. A DNA template is added to a reaction mixture together with a DNA polymerase and primers. The polymerase then replicates the DNA sequence specified by the added primers. PCR reactions are run as repeated cycles with three different temperature steps, heating for separation of DNA strands, cooling for primers to anneal, and then the optimum temperature for DNA synthesis. The annealing temperature is decided based on the annealing temperature of the primer pair. The primers bind easier in lower temperatures, thus lower temperatures gives a higher chance for unspecific product, and higher temperatures gives a higher chance of no product. For each cycle the number of DNA molecules is doubled, DNA sequence between the primers is thus amplified exponentially (19). PCR was used to amplify DNA both for cloning of a specific sequence, and for analysing DNA sequences. Different DNA polymerases were used to amplify DNA for different purposes.

### **2.7.1 Q5 PCR**

Q5® High-Fidelity DNA Polymerase from New England BioLabs (NEB) was used for amplification of DNA for cloning purposes. A mastermix was made according to Table 2.9, and 50 µL was aliquoted into each PCR tube. The reaction was run in a thermo cycler according to the program in Table 2.10. The recommendations from the producer was followed regarding temperatures and concentrations.

Table 2.9: The compounds for Q5 PCR.

<b>Compound</b>	<b>Volume per reaction</b>
Q5® Reaction Buffer from NEB	10 µL
10 mM dNTPs	1 µL
10 mM F primer	1.25 µL
10 mM R primer	1.25 µL
Q5® High-Fidelity DNA Polymerase from NEB	0.5 µL
Nuclease free water	35 µL
DNA template	1 µL

Table 2.10: Cycling conditions for Q5 PCR. The lines mark the start and end of the cycles, 35 cycles were run.

<b>Temperature</b>	<b>Time</b>
98 °C	1 min
98 °C	10 sek
Annealing temperature	15 sek
72 °C	1 min 30 sek
72 °C	2 min

### 2.7.2 Taq and Colony PCR

Colony PCR were used to screen for successful transformants or mutants among colonies grown on a selective plate after transformation or CRMAGE. Taq DNA Polymerase from NEB was used for colony PCR. A mastermix was created according to Table 2.11, and 10 µL was aliquoted into PCR tubes. Colonies were picked, replated, and transferred to the PCR tubes, one colony per tube. The samples were run in a thermo cycler according to the program in Table 2.12. In order release the DNA from the cells, the samples were kept at 95 °C for 10 minutes in the beginning of the program. The recommendations from the producer was followed regarding temperatures and concentrations.

Table 2.11: The compounds used for colony PCR.

<b>Compound</b>	<b>Volume per reaction</b>
10x Standard Taq Reaction Buffer from NEB	1 $\mu$ L
10 mM dNTPs	0.2 $\mu$ L
10 mM F primer	0.2 $\mu$ L
10 mM R primer	0.2 $\mu$ L
Taq DNA Polymerase from NEB	0.05 $\mu$ L
Nuclease free water	8.35 $\mu$ L

Table 2.12: Cycling conditions for colony PCR with Taq polymerase. The lines mark the start and end of the cycles, 30 cycles were run.

<b>Temperature</b>	<b>Time</b>
95 °C	10 min 30 sec
95 °C	20 sec
Annealing temperature	40 sec
68 °C	1 min
68 °C	5 min

### 2.7.3 CloneAmp HiFi PCR

CloneAmp HiFi Polymerase, from TaKaRa, was used once, after problems using Q5 PCR. TaKaRa states that CloneAmp HiFi Polymerase is accurate, efficient, has high sensitivity, and high priming efficiency, thus it was a good candidate for solving the problems. The CloneAmp HiFi polymerase comes in a premix. The premix was mixed with DNA and primers according to Table 2.13. PCR tubes were run in a thermo cycler with the cycling conditions in Table 2.14. The recommendations from the producer was followed regarding temperatures and concentrations.

Table 2.13: Compounds for CloneAmp HiFi PCR.

<b>Compound</b>	<b>Volume per reaction</b>
CloneAmp HiFi PCR Premix from TaKaRa	12.5 $\mu$ L
10 mM F primer	0.6 $\mu$ L
10 mM R primer	0.6 $\mu$ L
DNA template	<100 ng
Sterilized distilled water	Up to 25 $\mu$ L

Table 2.14: Cycling conditions for CloneAmp HiFi PCR. The whole program was cycled, 30 cycles were run.

<b>Temperature</b>	<b>Time</b>
98 °C	10 sec
55 °C	15 sec
72 °C	5 sec/kb

## 2.8 Gel Electrophoresis

Gel electrophoresis is a method for separating DNA molecules based on their size. DNA is applied to wells in an agarose gel, and electric currents are used to migrate the negatively charged DNA molecules towards the positive electrode. Larger DNA molecules will move more slowly through the gel matrix than smaller DNA molecules, thus the molecules are separated by size (19). Gel electrophoresis was used for both analysing PCR products and for purification of DNA fragments. The gels were made using 0.8 % agarose, TAE buffer was used as running buffer, and the gels were run with 80 V. For DNA being cut and purified, 0.005 % GelGreen from Biotium were added to the gel as stain. Otherwise, 0,005 % GelRed from Biotium was used.

## 2.9 DNA Sequencing

Correct assembly of plasmids, and correct insertion of genomic modifications were confirmed with sequencing. Samples containing DNA and a sequencing primer were sent to GATC Biotech, where they were sequenced by Sanger sequencing. The samples contained preferably 400-500 ng plasmid DNA or 100-400 ng purified PCR product, together with 25 pmol of the

primer. The data from the sequencing was analysed using Clone Manager Professional 9, Chromas, and Benchling.

## 2.10 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide-gel electrophoresis (SDS-PAGE) is a gel electrophoresis method for separating proteins after molecular weight. SDS is added to unfold the proteins and give them a negative charge, proportional to the size of the protein. The protein samples are loaded on a gel and an electrical field is applied to the gel and the running buffer. The proteins are then separated on the gel as the larger proteins will move more slowly through the gel towards the positive pole (19).

SDS PAGE was used to analyse the protein composition of cell cultures, and to separate proteins before Western blot. Protein samples were mixed with loading dye (Table 2.15) before loaded on the gel. ClearPAGE SDS Gel 12% from C.B.S Scientific were used, together with Teo-Tricine SDS Running Buffer from C.B.S Scientific were used. After running, the gel was rinsed in water, stained with Bio-Safe™ Coomassie G-250 Stain from Bio Rad overnight, rinsed in water, and then imaged using ChemiCoc™ XRS+ from Bio-Rad.

Table 2.15: The compounds used for SDS loading buffer, and their concentrations. DTT (1,4-dithiotreitol) was only added to loading dye for reducing conditions.

<b>Compound</b>	<b>Concentration</b>
Tris-HCl (pH 6.8)	50 mM
SDS	2 % (w/v)
Bromophenol blue	0,1 % (v/v)
Glycerol	10 % (w/v)
DTT*	100 mM

## 2.11 Western Blot

Western blotting is a method for detecting specific proteins, using antibodies, after an SDS PAGE. The proteins are transferred onto a membrane from the gel by an electric current. The membrane is then soaked in a solution with antibodies binding to the specific protein. Then, the antibodies are bound by secondary antibodies that can be detected (19).

After SDS PAGE, for separation of proteins, the gel was placed on top of a membrane, between transfer stacks. The proteins were transferred onto the membrane using the Trans-Blot® Turbo™ transfer system from Bio-Rad. The turbo program for mini gels was used. The proteins were detected using iBind™ Flex Western System from Invitrogen by Thermo Fisher Scientific. The primary and secondary antibody were diluted with iBind™ Flex solution from the iBind™ Flex Solution Kit. A iBind™ Flex Card were inserted into the iBind™ Western Device, and wetted with iBind™ Flex Solution. The membrane was laid on the card, and the device were loaded with the solution and the antibodies. The membrane is then automatically blotted. The system is based on capillary forces, causing a sequential flow of the different solutions, with primary antibody, secondary antibody, and without, across the membrane. After blotting, the membrane was transferred to a clean plate, covered in 3,3',5,5'-Tetramethylbenzidine (TMB) from Sigma Aldrich, and incubated for 1-10 minutes before the reaction was stopped with water. The membrane was imaged using ChemiCoc™ XRS+ from Bio-Rad.

## **2.12 USER Cloning**

USER (uracil-specific excision reagent) cloning was used by the authors of the CRMAGE protocol to insert the gRNA sequences into pMAS-SK (17). USER cloning is a method for DNA assembly based on the annealing of overlapping single stranded regions. Overlaps are created by using PCR primers containing an uracil at the position where the single stranded region should start. USER enzymes are then used to excise the uracil residues, thus creating the overlaps. The overlaps can then anneal to homologous regions, and the product can be used to transform *E. coli* (42). For CRMAGE, this method was used to create single stranded overlaps on the plasmid backbone. The overlaps on the gRNAs was designed to create overlaps when two single-stranded oligos were annealed together, as described in section 2.15.1. DNA assembly using USER cloning was not applied successfully in the initial stages of the presented project, thus the method for assembly of pMAZ-SK were adapted to Gibson assembly.

## **2.13 Gibson Enzymatic Assembly of DNA Molecules**

New sequences were inserted into plasmids using enzymatic assembly of DNA molecules as described by Gibson et al (43). DNA sequences with overlapping regions at the ends are added to a reaction mixture. T5 exonuclease degrades the 5' ends, thus exposing the overlapping 3' ends which are free to anneal. Phusion polymerase repairs the sequence degraded by the

exonuclease and *Taq* ligase ligates the two fragments together. The reaction is carried out at 50 °C. T5 exonuclease is degraded during the incubation period, and does not interfere with the repair process (43). The process is shown in Figure 2.1.

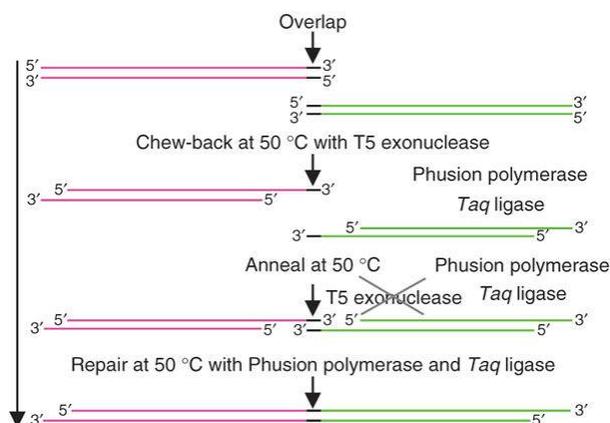


Figure 2.1: The reaction steps of Gibson assembly. First degradation of the 5' ends by T5 exonuclease, annealing of the 3' ends, and repair by Phusion polymerase and *Taq* ligase (43).

Isothermal Reaction Buffer (IRB) and Assembly Master-Mix (A-MM) was prepared as described in Table 2.16 and Table 2.17. The A-MM was stored in 15  $\mu$ L aliquots, and 5  $\mu$ L of DNA fragments for assembling was added. The DNA fragments was added in equimolar masses. Of the biggest fragment,  $\leq 100$  ng was added, and the amount of the larger fragment was used as a reference for the smaller fragment. The A-MM with added DNA was incubated at 50 °C for 1 h. After incubation the mix was directly used for transformation of DH5 $\alpha$ .

Table 2.16: Compounds and their concentrations in the isothermal reaction buffer (IRB).

Compound	Final concentration
Tris-HCl (pH 7,5)	500 mM
MgCl <sub>2</sub>	50 mM
DTT	50 mM
NAD	5 mM
dNTPs	1 mM
PEG-8000	25%

Table 2.17: The volume used of compounds in the Assembly Master-Mix

Compound	Used volume ( $\mu\text{L}$ )
IRB (frozen aliquot)	320
T5 Exonuclease (NEB; 10 U/ $\mu\text{L}$ )	0.64
Phusion® High-Fidelity DNA Polymerase (NEB; 2 U/ $\mu\text{L}$ )	20
Taq DNA Ligase (NEB; 40 U/ $\mu\text{L}$ )	160
MilliQ water	700

## 2.14 Designing MAGE Oligonucleotides and gRNA Sequences

The sequences of the oligos for  $\lambda$  Red recombineering and of the gRNAs for CRISPR/Cas9 was designed in silico before the oligos were ordered. The software Benchling ([www.benchling.com](http://www.benchling.com)) was used for managing and visualizing the genome of MG1655. The genome of MG1655 was downloaded from the NCBI database (GenBank accession: U00096.3).

### 2.14.1 Designing MAGE Oligonucleotides

The MAGE oligos for generating *secG* and *ffh* mutants were designed to downregulate the expression of the genes by changing the complete ribosomal binding site (RBS) sequence and changing the start codon. First, the EcoCyc database ([www.ecocyc.org](http://www.ecocyc.org)) was used to find the translational start site. The RBS calculator by Salis Lab (<https://salislab.net/software/forward>) (44, 45) were used to calculate the translation initiation rate of the genes, using a proportional scale from 0 to over 100 000. The sequence of the mRNA was inputted, together with which target organism was used. Then, a new ribosomal binding site was designed using another feature of the same RBS calculator. The coding sequence of the gene, and 20 bp of the sequence in front of the ribosomal binding site were input, along with the target translation initiation rate. The target translation initiation rate was selected based on the calculated translation initiation rate. As *ffh* is essential, the aim was to not downregulate it too much. As *secG* is non-essential, it was downregulated much more. Both the calculated original translation initiation rates and the new ones are shown in Table 2.18. Only one RBS sequence was constructed for each gene as an initial test. Ideally, different RBS sequences would have been designed and tested, with different target translational initiation rates.

Table 2.18: The calculated translation initiation rates (TIR) of *secG* and *ffh*, and the new TIR with the new designed RBS sequences.

<b>Gene</b>	<b>TIR</b>	<b>New TIR</b>
<i>secG</i>	1794.51	102.26
<i>ffh</i>	308.40	104.84

In Benchling, the parts of the mRNA sequence upstream of the coding sequence were replaced by the designed RBS sequence. At the same time the start codons of the genes were changed from AUG to CUG. AUG is the most common start codon, but translation can also be initiated at other codons. However, the strength of other codons as start codons are lower, thus changing the start codon can be used to lower expression rates (46).

For *secM* the N-region of its signal peptide was replaced by an alternative signal peptide, DsbA, which target the protein to the SRP-pathway (27). Like for *secG* and *ffh*, the change was first made *in silico*, in Benchling. All three oligos were designed with 100 bp overlaps on each side of the substitution. The oligos were designed to hybridize to the lagging strand. The locations of the target genes in relation to *oriC* and *dif*, with the illustration of how to target the lagging strand is shown in Figure 2.2. The *secM* gene is located on replicore 1, therefore the oligo was designed to hybridize to the plus strand. Thus, the overlaps were identical to the minus strand, and as the gene is located on the plus strand the insert on the oligos were the complimentary sequence of the wanted mutation. The *ffh* and *secG* gene are located on replicore 2, thus the oligos were designed to hybridize to the minus strand. Both genes are located on the minus strand, the inserts were therefore the complimentary sequence of the mutations.

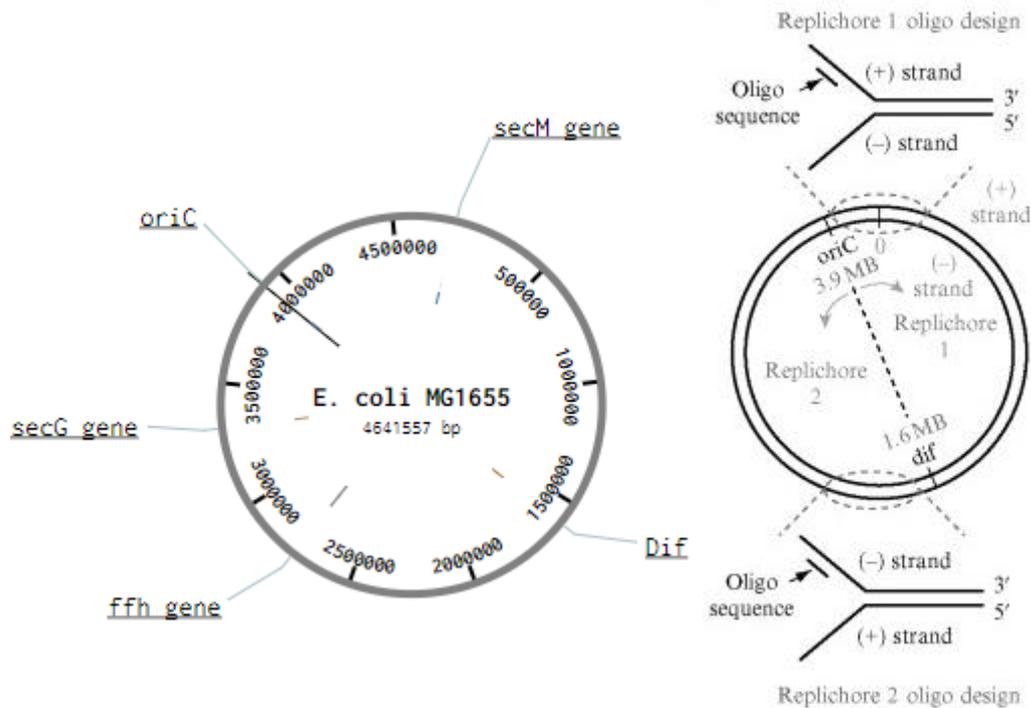


Figure 2.2: The locations of the target genes, *secM*, *secG*, and *ffh*, in relation to *oriC* and *dif* on the genome of *E. coli* MG1655 (GenBank accession: U00096.3), shown together with an illustration of how to design MAGE oligos targeting the lagging strand, adapted from (18). The illustration with the target genes are created in Benchling.

The mutations were deliberately made large enough to make screening for mutants by colony PCR possible. This demanded that the mutation was large enough to allow a primer sequence to bind specifically to the insertet sequence. The *secG* oligo was ordered as a Megamer™ ssGene Fragment from Integrated DNA Technologies (IDT). The oligos for *secM* and *ffh* was ordered from Biologio, as they could not be synthesized by IDT. The sequences of the oligos are shown in Table 2.19.

Table 2.19: The sequence of the MAGE oligos as they were ordered. The inserted sequences is shown in uppercase letters, and the homology regions are shown in lowercase letters.

Gene	MAGE oligo
<i>secG</i>	cgaaggaggctcccatatcagcgcctttaccttgctgcagcatgatcagaccaacaaggccaattgccacaat aaggaaaactactaaaagagcttcatacagGCGATGTATGAACAagggtgtaactaactaaccaag cgaatgaccttcgcaagggcaatTTTatcgattgtatcaactgcggaaaaaacagcaaacccgatgtgttcgct gag
<i>ffh</i>	gcacttcgcgcagcgtatcttttacgttgcttcagtgaggcgtccacggcactgatattgcgcagcgtgcgcga caaacgatcggtaaattatcaaacagATGTGGCAAAGATGTGTTCTTGTTTTTATGTG gccgcagtataacatgaaggcgtctttgttattgcaacgggtggagcagcgttcacctgacgtatactgcttc ttcttattgctcaaactgtcg
<i>secM</i>	gctcgtggttcggggttcgcttttgcgggcgcgttggttcggcggcgttgctgagcgcaggcaaacctaaact cgccgcaaccatccctaataagagCGCCGATGCGCTAAACGCTAAAACCTAAACCAG CCAGCGCCAGCCAAATCTTTTTTCATgttattgccatcccgatcatattcaaacgtatttagcgc acaaatcttatcattaaagcctgcagtcacagcacgaatgaagataagttgtgcggaaaa

#### 2.14.2 Designing gRNA Sequences

The gRNA sequences were designed in Benchling using their built-in tool. The wanted length of the sequence, the target organism, and the PAM sequence was inserted. The length used was 20 nucleotides, the target organism was *E. coli* MG1655, and the PAM sequence was NGG. The gRNAs were designed to have their PAM site and target sequence inside the region that would be removed during the recombineering. Thus, only non-mutants could be targeted, as mutants would no longer have neither the target sequence or the PAM site. The gRNAs were also chosen to have a low rate of off-target effect, meaning a low possibility to target other regions in the genome. Therefore, the target organism had to be entered (MG1655), in order for Benchling to calculate the off-target score. The gRNAs were ordered from Sigma-Aldrich, as two separate single stranded oligos, which were later annealed. They had overhangs at the ends for Gibson assembly. The sequences are shown in Table 2.20.

Table 2.20: The gRNA oligos ordered for each target gene, with the PAM sequence and off-target score. The gRNA sequence is shown in uppercase letters, and the overhang regions for Gibson assembly is in lowercase. The off-target score is calculated by Benchling, it ranges from 0-100, where higher is better.

Gene	gRNA F sequence	gRNA R sequence	PAM	Off-target score
<i>secG</i>	gagcacAATGCTTCAAC	ctaaacGCTTTATTGGT	GGG	100
	CAATAAAGCgttttagagct	TGAAGCATTgtgctcagtat		
	agaaat	ctct		
<i>ffh</i>	gagcacTCTCTCGCCTG	ctaaacTTTCCACCCCA	CGG	100
	GGGTGGAAAgtttagagct	GGCGAGAGAgtgctcagta		
	agaaat	tctct		
<i>secM</i>	gagcacTGAGTGGAATA	ctaaacGCGCGTCAGTA	TGG	100
	CTGACGCGCgttttagagct	TTCCACTCAgtgctcagtat		
	agaaat	ctct		

## 2.15 Construction of pMAZ-SK with gRNA Insert

### 2.15.1 Annealing of gRNA

A mix for annealing of the two single-stranded gRNA oligos was prepared as described in Table 2.21. The mix were incubated in a thermocycler at 95 °C for 5 minutes, and cooled by 0.1 °C/sec. Then, 10 µL of the mixture were run on a gel electrophoresis gel (gelgreen, 80V, 20 min). The product was cut out and purified from the gel using Zymoclean™ Gel DNA Recover Kit (D4002) from Zymo Research.

Table 2.21: The components in the mix for annealing of gRNA for inserting into pMAZ-SK

Component	Volume
Forward gRNA oligo (100 µM)	10 µL
Reverse gRNA oligo (100 µM)	10 µL
NEBuffer 4	10 µL
Water	70 µL

### 2.15.2 Amplification of pMAZ-SK Backbone

The backbone of pMAZ-SK was amplified using Q5 PCR as described in section 2.7.1, with 64 °C as annealing temperature. The primers were designed to leave overhangs for Gibson assembly, and to amplify the whole plasmid except the gRNA sequence used by the authors of

the CRMAGE protocol. The primers, pMAZ-SK Backbone Gibson F and R, are shown in Appendix 1. Then, gel electrophoresis (gelgreen, 80V, 55 min) was used to analyse the PCR product. The gel was run with 25  $\mu$ L of PCR product, and the band of the backbone was cut out and purified from the gel using Zymoclean™ Gel DNA Recover Kit (D4002) from Zymo Research.

### 2.15.3 Assembly and cloning of pMAZ-SK with gRNA Insert.

The plasmid was assembled using Gibson assembly as described in section 2.13. After Gibson assembly, the Gibson mix was used for heat-shock transformation of DH5 $\alpha$ , as described in section 2.5. The cells were plated on LA plates with Kan. Colony PCR was used to screen for correctly assembled plasmids, as described in section 2.7.2. The primer pairs were designed with one primer specific for the gRNA sequence, binding inside the insert and one universal primer binding outside the insert. The sequences of all gRNA check primers are shown in Appendix 1. The primer pairs should give a PCR product of approximately 800 bp for pMAZ-SK with the correct insert, and no specific PCR product for pMAZ-SK without the correct insert. Correctly assembled plasmids were purified using ZR Plasmid Miniprep™-Classic (D4016) from Zymo Research and the sent for sequencing as described in section 2.9.

## 2.16 Transformation of MG1655 with pMA7CR\_2.0 and pZS4Int\_tetR

MG1655 was transformed with pMA7CR\_2.0 and pZS4Int\_tetR using electroporation, as described in section 2.6, in order to be ready for genomic modification by CRMAGE. The electroporation was performed with both plasmids simultaneously, with a total of 3  $\mu$ L DNA added. After electroporation the cells were grown on LA plates with Amp and Spc.

## 2.17 CRMAGE

A 1% culture was inoculated, from an overnight culture of cells with pMA7CR\_2.0 and pZS4Int-tetR, in 15 mL LB with Amp and Spc. The culture was grown at 37 °C, with shaking, until OD<sub>600</sub> was 0.4-0.5. L-arabinose was added to a final concentration of 0.2%, using a stock solution of 0.2 g/mL, before the culture was incubated for another 10-15 minutes. L-arabinose induced the expression of  $\lambda$  Red  $\beta$ -proteins. The culture was cooled in an ice-water bath for at least 15-20 minutes. Then, the cells were harvested using centrifugation (6500 rcf, 5 min, 4 °C),

and resuspended in 35 mL cold water. The harvesting was repeated three more times, and the cells were resuspended in 15 mL, 1 mL, and finally 400  $\mu$ L.

Next, 50  $\mu$ L of the cells were added to a tube containing 0.5  $\mu$ L MAGE oligos at 10  $\mu$ M concentration and approximately 250 ng of pMAZ-SK. The cells were electroporated at 1.8 kV in a 1 mm gap cuvette or at 2.5 kV in a 2-mm gap cuvette. After electroporation the cells recovered in 950 $\mu$ L of LB, with Amp and Spc to select for pMA7CR\_2.0 and pZS4Int-tetR. The cells were transferred to a culture tube and incubated for 1 h at 37 °C with shaking. At this point, Kan was added to select for pMAZ-SK. The culture was incubated for another 2 h. Then aTetracycline was added to a concentration of 200 ng/mL, to induce expression of Cas9 and sgRNA. The cells were then incubated overnight, before they were plated on LA plates with Amp, Spc and Kan.

The colonies were screened using colony PCR, as described in section 2.7.2. The primer pairs were designed to have one primer binding inside the inserted sequence and the other outside, resulting in specific PCR products only for mutants. The DNA sequence of all genome check primers are shown in Appendix 1. The PCR product were analysed using gel electrophoresis (gelgreen, 80V, 40 min), and the product of positives were cut out, purified using Zymoclean™ Gel DNA Recover Kit (D4002) from Zymo Research, and sent for sequencing as described in section 2.9.

#### 2.17.1 Curing the pMAZ-SK Plasmid

The pMAZ-SK plasmid was removed from the cells after successful mutation using the self-killing system of the plasmid. An overnight culture was grown and a new culture of 10 mL LB with Amp and Spc was inoculated with 1% from the overnight culture. The culture was grown for 4 h and then induced with aTet, to a final concentration of 200 ng/mL, and L-rhamnose, to a final concentration of 0,2 %, to induce expression of Cas9 and the crRNA. The induced culture was incubated overnight and plated out on LA plates with Amp and spc. The colonies were checked for resistance towards Kan by replating colonies on a plate with amp, spc, and kan. No growth would indicate that the cells had lost their resistance to kan, and the plasmid.

### 2.17.2 Curing of pMA7CR\_2.0 and pZS4Int\_tetR

Before the properties of the constructed mutants were tested, the remaining plasmids of the CRMAGE system was removed. Cells were streaked out and grown on LA plates without antibiotics, at 42 °C for one or two days. Then, cell material was picked from the edge of the streak and replated on a new LA plate without antibiotics. The process was repeated until the cells had lost both plasmids. The loss of plasmids was checked by plating on plates with antibiotics. In the end, the cells were streaked out to get single colonies, and one colony without both plasmids were chosen for further experiments.

### 2.18 SDS-PAGE for Confirmation of Cas9

SDS-PAGE was used to verify the presence of Cas9 in cells with pMA7CR\_2.0. Overnight cultures were grown of MG1655 pMA7CR\_2.0 pMAZ-SK, MG1655 pMA7CR\_2.0 pZS4Int\_tetR pMAZ-SK (induced and non-induced), and WT MG1655. The cells were harvested from 8 mL and 10 mL of culture by centrifugation (5 min, 7000 rcf), and the pellet were resuspended in 8 mL and 4 mL 50 mM Tris-HCl. The cells were then lysed by sonification, using using Bronson Sonifier 250 with the flat tip for 3 minutes. After sonification, 1 mL of the solution was centrifuged for 15 minutes. The supernatant (soluble fraction) was transferred to a new tube, and the pellet (insoluble fraction) was resuspended in 100 µL of water. SDS PAGE was performed as described in section 2.10. 5 µL of each sample was transferred to new tubes with 5 µL of SDS loading buffer (Table 2.15) with DTT (1,4-dithiothreitol). The samples were heated at 100 °C for 10 minutes, and then the whole volume was loaded on the gel (12 wells). The SDS-PAGE was run at 100V for 1 h, before the gel was stained and imaged.

### 2.19 Construction and Transformation of Vector with Reporter Protein

A fusion protein with IgG fc-fragment and sfGFP was designed and constructed to be used as a reporter for levels of periplasmic translocation of recombinant proteins.

#### 2.19.1 Design of Reporter Protein

The reporter protein was designed by combining the sequences of OmpA signal peptide (27), IgG2 Fc fragment (47), linker (28), and sfGFP (28). An illustration of the fusion protein sequence is shown in Figure 2.3, and the full sequence of the reporter is shown in Appendix 1.

The linker was used to allow correct folding of the fused proteins. The start codon from sfGFP were changed to make sure no translation of only sfGFP were initiated. The sequence of the IgG2-sfGFP construct were ordered as two double stranded gene fragments with overlapping regions designed for Gibson assembly. The fragments were ordered from Integrated DNA Technologies (IDT) as gBlocks® Gene Fragments.

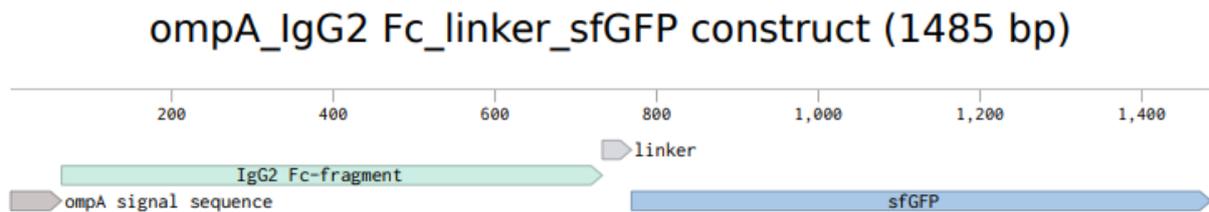


Figure 2.3: Illustration of the sequence of IgG2 Fc – sfGFP fusion protein used as a reporter. The OmpA signal sequence is at the beginning of the protein, and a linker between the fusion partners.

### 2.19.2 Assembly of pVB1-251-kan with OmpA-IgG Fc-linker-sfGFP Insert

After failed attempts at amplifying the pVB1-251-kan backbone using Q5 polymerase and Phusion polymerase, the backbone was finally successfully amplified using CloneAmp HiFi PCR, as described in section 2.7.3. The primers were designed to create overlaps for Gibson assembly. The fragment was cut out and purified from the gel after gel electrophoresis, using Zymoclean™ Gel DNA Recover Kit (D4002) from Zymo Research. The backbone and the two gene fragments was assembled into one plasmid using Gibson assembly, as described in section 2.13. Then, 15 colonies were screened by measuring levels of fluorescence, as will be described below. The plasmid was purified using ZR Plasmid Miniprep™-Classic (D4016) from Zymo Research. One sample was sent for sequencing for a final confirmation of correct assembly. The pVB1-251-kan plasmid with the OmpA-IgG Fc-sfGFP insert was renamed pAE001.

### 2.19.3 Fluorescence Test

First, the colonies were inoculated in LB with Kan, and incubated at 37 °C, with shaking, overnight. Then, 1 % was inoculated in 5 mL fresh LB with Kan and grown under the same conditions for 2 h. The inducer, m-toluic acid, were added to a concentration of 0.001 M, before the cells were incubated for another 2 h. For each colony, one induced and one non-induced culture was grown. Next, 1 mL from each culture were transferred to Eppendorf tubes and the cells were harvested by centrifugation (8 rcm, 5 min). The cells were resuspended in 1 mL PBS,

and 200  $\mu\text{L}$  of each sample was transferred to a black 96-well optical-bottom plate from ThermoFischer Scientific. The plate was read using Infinite 2000 PRO plate reader from Tecan. The excitation wavelength was 488 nm, and the emission was measured at 530 nm.

## **2.20 Protein Analysis of Reporter**

The reporter protein was analysed using SDS PAGE and western blotting. The proteins of MG1655 WT and MG1655 pAE001, both induced and non-induced were compared, in both reducing and non-reducing conditions. On the SDS-PAGE gel and the western blot membrane of IgG2 Fc, a standard of IgG Fc was also added, in reduced and non-reduced form. The standard used were Native Human IgG Fc fragment protein (ab90285) from Abcam.

### **2.20.1 Protein Isolation**

A cell pellet was created by first inoculating 1% of an overnight culture in 35 mL LB with Kan. The culture was grown until  $\text{OD}_{600}$  was approximately 2 at 37°C, with shaking, before m-toluic acid was added to a concentration of 0.001M. The culture was incubated for 18 h at 16 °C. The protocol for protein expression was obtained from Vectron Biosolutions. After incubation, the cells were kept on ice until lysis. Cells were harvested by centrifugation (6500 rcf, 5 min, 4 °C) from 30 mL of the culture. The cells were resuspended in 0.9% NaCl, 1 mL per 100 mg pellet. Next, 1 mL was transferred to an Eppendorf tube and centrifuged (6500 rcf, 5 min, 4 °C). The pellet was resuspended in 500  $\mu\text{L}$  of Cell Lytic Buffer B from Sigma, with 0,1  $\mu\text{L}$  Benzonase® Nuclease from Sigma-Aldrich. The tubes were incubated in room temperature, at 100 rpm shaking for 30 minutes. After incubation the tubes were centrifuged (6500 rcf, 5 min, 4 °C). The supernatant (soluble fraction) were transferred to another tube. The pellet (insoluble fraction) was resuspended in Teo-Tricine SDS Running Buffer from C.B.S Scientific. Both fractions were diluted 1:5 using the same SDS running buffer.

### **2.20.2 SDS PAGE Before Western Blot**

Three times concentrated loading buffer (Table 2.15) was added to the samples, 15  $\mu\text{L}$  3x loading buffer to 30  $\mu\text{L}$  sample, both with and without DTT for each sample. The samples with DTT was boiled at 95 °C for 5 minutes. SDS PAGE was performed as described in section 2.10.

The gel (17 wells) was loaded with 5 $\mu$ L of each sample. The gel was run at 130V for 90 minutes, before being stained and imaged.

### 2.20.3 Western Blot for Identification of Reporter Protein

The SDS PAGE gels was used to run western blot as described in section 2.11. The primary antibody used for detection of IgG2 Fc fragment were Mouse Anti-Human IgG Fc-UNLB (9042-01) from Southern Biotech. The secondary antibody used were Polyclonal Rabbit Anti-Mouse Immunoglobulins/HRP (P0260) from Dako. The primary antibody used for detection of sfGFP were Goat Anti-GFP Antibody (600-101-215) from Rockland. The secondary antibody used were Donkey Anti-Goat IgG H&L (HRP) (ab97110) from Abcam.

## 2.21 Growth Studies

MG1655 was transformed with pAE001, using electroporation as described in 2.6. A growth experiment was performed to compare the growth of MG1655 pAE001, both induced and none induced, to WT MG1655. A culture was grown overnight, and 1% inoculated in fresh LB with appropriate antibiotics. The OD<sub>600</sub> was measured at the point of inoculation, and then every 1-2 h. The fluorescence was measured at the point of induction and then every 2 h. Fluorescence measurement was done by harvesting cells from 1 mL culture and resuspending them in 1 mL PBS. Then, 200  $\mu$ L was transferred to black 96-well plates from Thermo Scientific, and the fluorescence measured with excitation wavelength 488 nm, and emission wavelength 530 nm. Growth studies was performed on WT MG1655 and MG1644 pAE001, both induced and non-induced, in order to investigate how pAE001 and expression of the reporter affected the growth characteristics of the cells.

## 2.22 Comparison of Mutant Strains

The mutant strains created was transformed with pAE001, using electroporation as described in section 2.6. In order to characterize translocation abilities of the mutants, the expression of the reporter protein was measured and compared to the expression of WT MG1655 pAE001. A 1% culture was induced from an overnight culture, and incubated at 37 °C, with shaking, until OD<sub>600</sub> was approximately 2. The cultures were induced with m-toluic acid to a final concentration of 0.001M. The culture flasks were moved to an incubator with 16 °C and

shaking, and incubated for 16h. The fluorescence was measured using Infinite 2000 PRO plate reader from Tecan, and growth was measured by measuring OD<sub>600</sub>.

## 3 Results

### 3.1 Establishing and Optimizing CRMAGE

There had been attempts to establish the CRMAGE method in our laboratory previously, but these had not been successful. The method had been very unreliable, and when it worked it had been very inefficient. As the success rate was very low, yet still existent, the initial hypothesis was that the MAGE component of the system worked, but not the CRISPR/Cas9 component. A CRMAGE experiment was run without MAGE oligos, to test the killing efficiency of CRISPR. Theoretically, all or most cells should have been killed, yet, the plates were overgrown. Further, prolonged exposure to the aTet inducer was tested, by adding the inducer to the LA plates. The plates were still overgrown. An SDS-PAGE was run to find out whether the reason was lack of Cas9 expression.

#### 3.1.1 SDS-PAGE for the Verification of Cas9 Expression

SDS-PAGE was run to confirm the expression of Cas9 in MG1655. Lysates were made from MG1655 with different combinations of CRMAGE plasmids, as described in Table 3.1. One gel was run for the insoluble proteins (Figure 3.1), and one for the soluble proteins (Figure 3.2). Cas9 was expected to be expressed in MG1655 pMA7CR\_2.0 pMAZ-SK and in induced MG1655 pMA7CR\_2.0 pZS4Int-tetR pMAZ-SK. The expected weight of *S. pyogenes* Cas9 is 158.4 kDa. On the gel of soluble proteins, one faint band, with the correct weight, could be seen for samples expected to express Cas9. No other differences between the samples could be seen. It was concluded that the cells were expressing Cas9.

Table 3.1: The strains studied by SDS PAGE to confirm the expression of Cas9, together with the number of the sample, and whether expression of Cas9 was expected.

<b>Nr.</b>	<b>Strain</b>	<b>Expression of Cas9</b>
1	MG1655 pMA7CR_2.0 pMAZ-SK	Yes
2	MG1655 pMA7CR_2.0 pZS4Int-tetR pMAZ-SK (Induced)	Yes
3	MG1655 pMA7CR_2.0 pZS4Int-tetR pMAZ-SK(Non-induced)	No
4	MG1655 WT	No

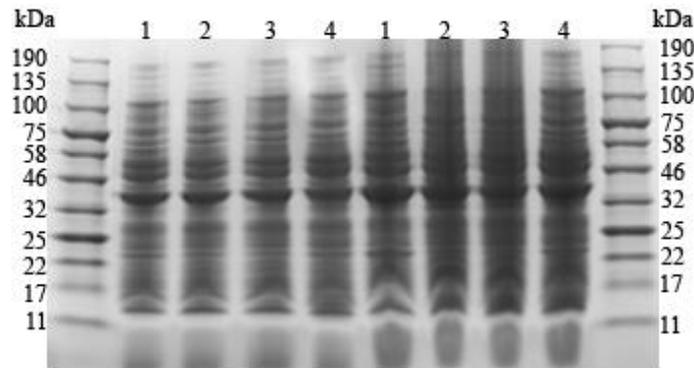


Figure 3.1: Insoluble proteins from SDS PAGE of MG1655 WT (4), MG1655 pMA7CR\_2.0 pMAZ-SK (1), and MG1655 pMA7CR\_2.0 pZS4Int-tetR pMAZ-SK, induced (2) and non-induced (3), in two different concentrations.

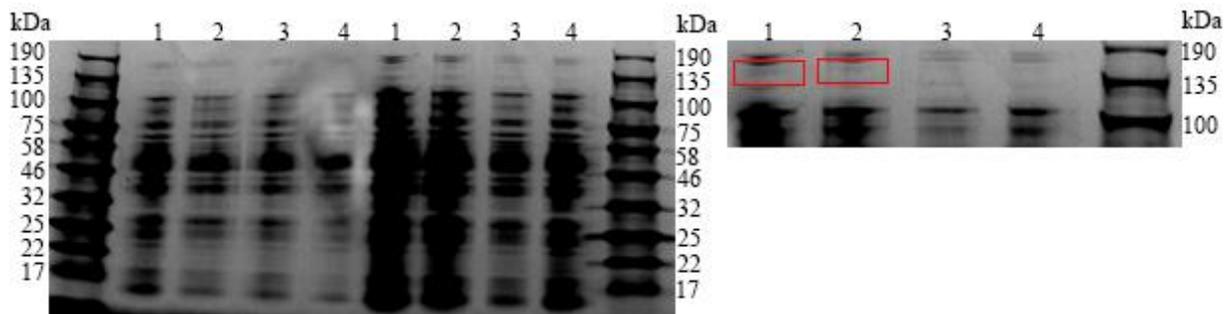


Figure 3.2: Soluble proteins from SDS PAGE of MG1655 WT (4), MG1655 pMA7CR\_2.0 pMAZ-SK (1), and MG1655 pMA7CR\_2.0 pZS4Int-tetR pMAZ-SK, induced (2) and non-induced (3), in two different concentrations. A small part of the samples with higher concentrations are shown enlarged, to show the faint bands with the molecular weight corresponding to Cas9 (158,4 kDa). The bands are marked with red rectangles.

### 3.1.2 Tagged *glyA* as a Positive Control for CRMAGE

With help from the group of Alex Toftgaard Nielsen, Technical University of Denmark, the original authors of CRMAGE a positive control for the CRMAGE method was obtained. A positive control would make it easier to find out whether the CRMAGE method had worked. Mutants with a tagged *glyA* gene were constructed. The tag was supposed to cause lower viability of the cells, therefore giving smaller colonies on the LA plate (48). The success and efficiency of the CRMAGE method could thus be seen, based on a phenotypic difference. The *glyA* gRNA insert was inserted into pMAZ-SK using USER cloning, the plasmids screened with colony PCR, and the CRMAGE protocol was run. There were no obvious phenotypic

differences on the plates, and colony PCR showed very low efficiency, from 1 to 3 of 30 colonies tested from each plate. The results indicated that the recombineering had worked, but still not the negative selection by CRISPR/Cas9.

To find out whether something was wrong with the sequence of the plasmids, both pMA7CR\_2.0 and pMAZ-SK were sent for sequencing. The sequencing results revealed that the sequences of both plasmids were correct, except that the *glyA* gRNA insert was not inserted. USER cloning was repeated, and plasmids from 10 different colonies were purified and sent for sequencing. The colonies were picked from both the first and the second round of cloning. The sequencing results showed that two colonies had the original insert of pMAZ-SK and the rest had reads that did not cover the gRNA sequence. Thus, it was concluded that the reason for not getting the negative selection working was problems with the cloning method.

#### *3.1.2.1 Construction of pMAZ-SK<sub>glyA</sub> Using Gibson Assembly of DNA Molecules*

After the construction of pMAZ-SK with *glyA* gRNA with USER cloning failed, the gRNA inserts and primers for amplifying pMAZ-SK backbone was redesigned for Gibson assembly. Unlike USER cloning, Gibson assembly was well established and optimized at our laboratory. The Gibson mix was transformed into *E. coli* DH5 $\alpha$  and colony PCR was used to screen for correctly assembled plasmids. The primer pair were designed with one primer binding outside the inserted gRNA and one primer binding directly to the insert. The PCR product were expected to be approximately 800 bp. The gel electrophoresis gel from the colony PCR is shown in Figure 3.3. At least 14 out of 15 colonies tested seemed to have correctly assembled plasmids. Sample 9 had a band of correct size, but also an unspecific band the same size as bands from the negative control. Possibly, something was wrong with the assembly of the plasmid in this sample. The sample was therefore rejected. Correct assembly was confirmed by sequencing before the plasmids were used for construction of mutants.

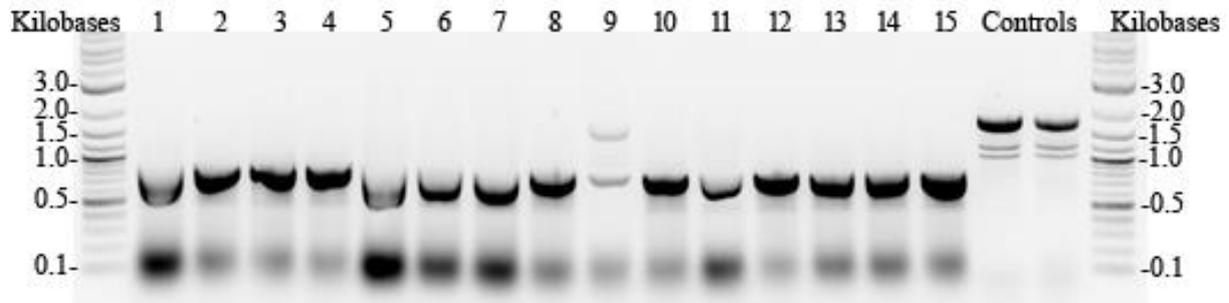


Figure 3.3: Gel from colony PCR after Gibson assembly of pMAZ-SK\_glyA. The colony PCR was run with 15 colonies and two negative controls. The negative controls were DH5 $\alpha$  pMAZ-SK without the correct insert. The expected size of the product was ~800 bp.

### 3.1.2.2 Construction of *glyA* mutants

Mutants of *E. coli* BW25113 with a tag added to *glyA* was constructed using CRMAGE. After overnight incubation the cells were plated out, the plates are shown in Figure 3.4. The phenotypic difference was not obvious; however, some colonies were identified as smaller than others. A possible explanation for the small difference is the density of cells on the plate. High density of cells can cause the cells to grow smaller colonies. Another possible explanation is that BW25113 were used instead of MG1655, which were originally used by Lauritsen et al. BW25113 was chosen as the strain will be used with the CRMAGE method later, in other projects. However, it is possible that the phenotypic difference is not as obvious in BW25113 as MG1655.

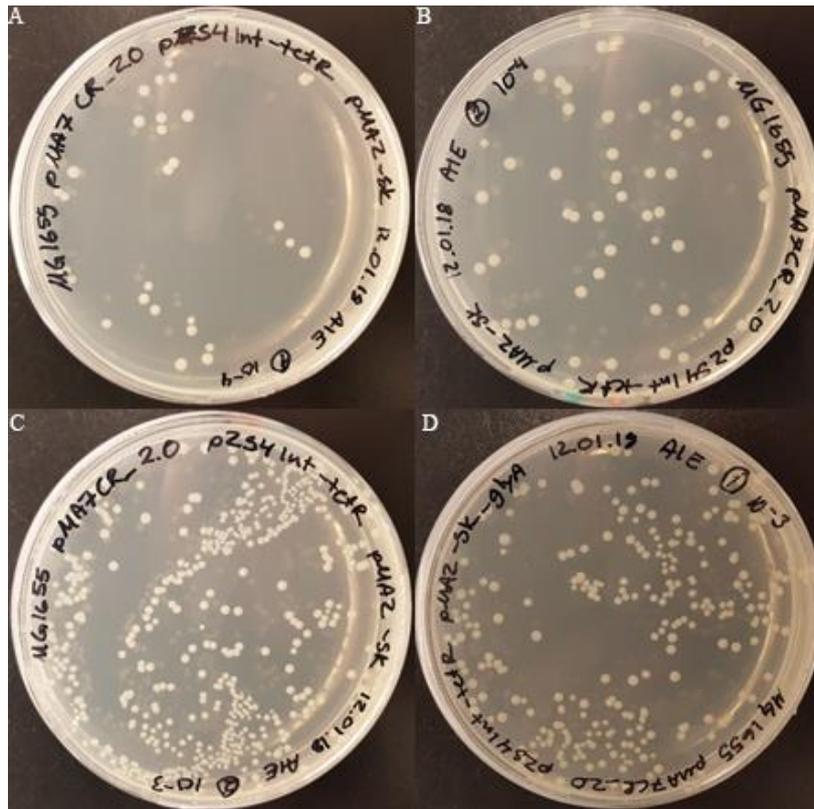


Figure 3.4: LA plates after generation of *glyA* mutants using CRMAGE. *GlyA* mutants were supposed to grow as smaller colonies. CRMAGE was run in two parallels, and each parallel was plated in two different concentrations, diluted  $10^3$  times (picture A and B) and  $10^4$  times (picture C and D). Some colonies were smaller than others, but the phenotypic difference was not as big as expected.

Mutants were screened for using colony PCR. The primer pair was designed with one primer binding inside the inserted *glyA* tag in the genome, and one binding outside. The PCR product was expected to be 953 bp. The gel picture from the gel electrophoresis is shown in Figure 3.5. Of the colonies considered small, 7 of 16 colonies screened seemed to be mutants. None of the 16 colonies considered large seemed to be mutants. The negative controls were BW25113 pMA7CR\_2.0 pZS4Int-tetR, without any mutations introduced. As the phenotypic difference was not as pronounced as expected, the efficiency of CRMAGE could not be determined only based on the growth on the plates. However, CRMAGE was finally working with higher efficiencies than expected from  $\lambda$  Red recombineering alone. The CRMAGE method was considered implemented, even though it was not fully optimized, and the construction of Sec-pathway mutants was begun.

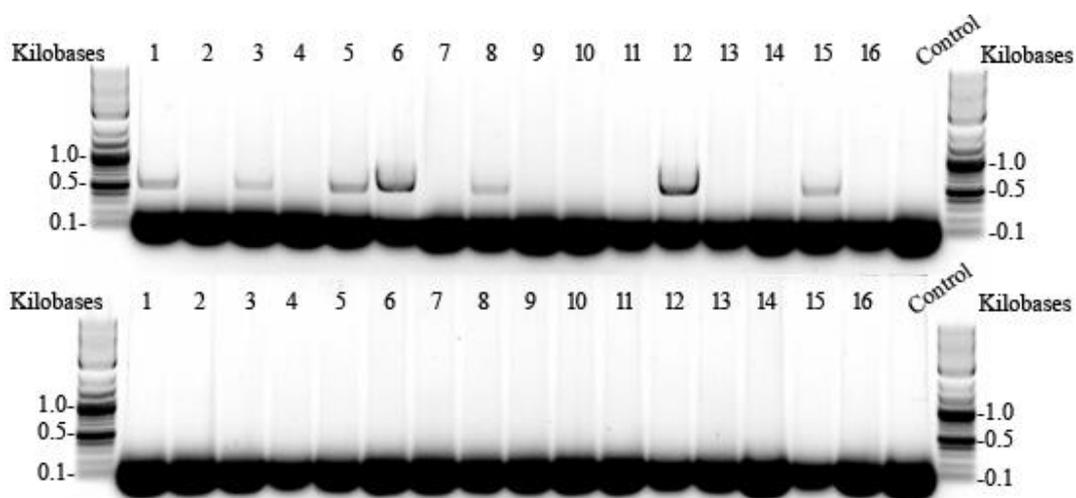


Figure 3.5: The gel after screening for *glyA* mutants using colony PCR. The upper row are colonies considered small, and the lower row are colonies considered large. Sixteen colonies of each size were screened and run with two negative controls. The lanes are overloaded with DNA, but the product seems to be the correct size, 953 bp (sample 1,3, 5, 6, 8, 12, and 15). The negative controls were MG1655 without mutations.

### 3.2 Assembly of pMAZ-SK with gRNA Targeting the Sec-Pathway

As the work with the positive control had revealed that USER cloning did not work, primers for amplification of the pMAZ-SK backbone and gRNA inserts were designed for Gibson assembly. The Gibson mix was transformed into *E. coli* DH5 $\alpha$  and correctly assembled plasmids were screened for using colony PCR. The primer pairs were designed with one universal primer binding outside the insert and one primer specific for each insert binding directly to the insert. The PCR product were expected to be approximately 800 bp for all the plasmids. The gel picture from the colony PCR are shown in Figure 3.6. For *ffh*, 10 of 11 colonies screened seemed correctly assembled, for *secM*, 11 out of 11, and for *secG*, 10 out of 11. Correct assembly of the plasmids were confirmed by sequencing before further use of the plasmid.

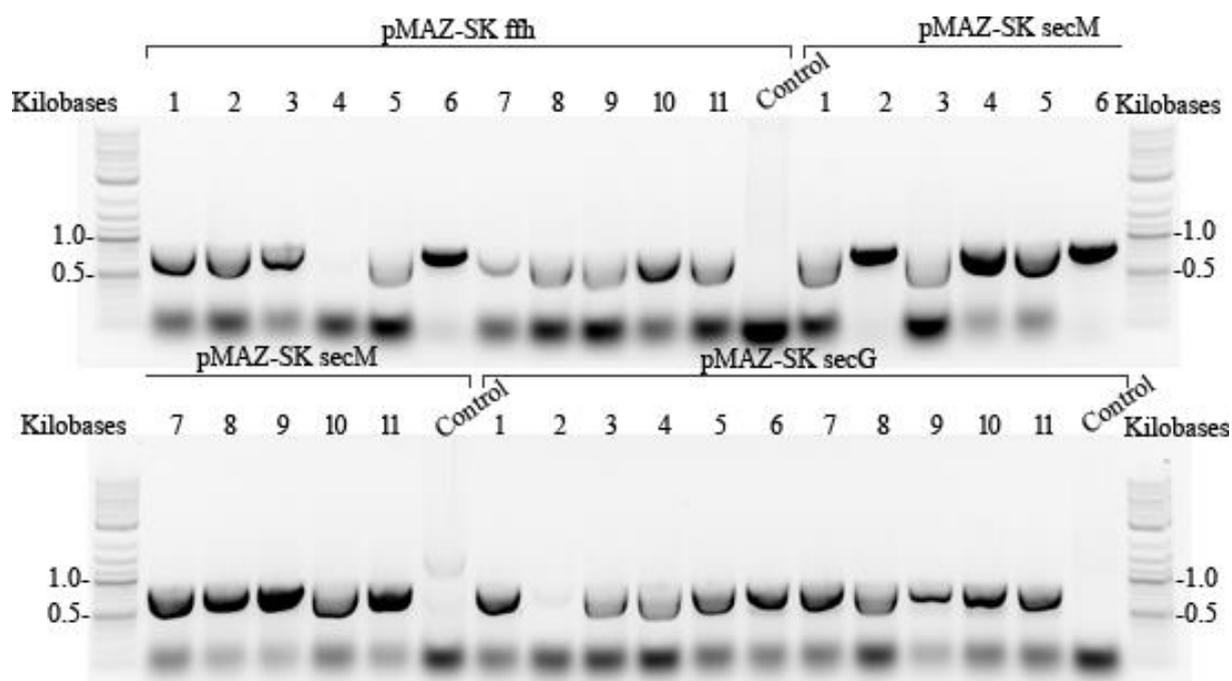


Figure 3.6: Gel from colony PCR after Gibson assembly of pMAZ-SK\_*ffh*, pMAZ-SK\_*secM*, and pMAZ-SK\_*secG*. Eleven samples and two negative controls were run for each plasmid. The negative controls were DH $\alpha$  pMAZ-SK without the correct insert. The expected size of the product were ~800 bp.

### 3.3 Construction of Sec-Pathway Mutant Strains

#### 3.3.1 Mutant Strains with Single Mutations

CRMAGE was used to introduce mutations in *ffh*, *secM*, and *secG* in MG1655, using the constructed pMAZ-SK plasmids and the ordered MAGE oligos. Screening for mutants were done using colony PCR. The three primer pairs were designed with one primer binding inside the insertion and one binding outside. The expected size of the PCR products for *ffh*, *secM*, and *secG* were, 494 bp, 479 bp, and 633 bp, respectively. The sizes of the PCR products were checked using gel electrophoresis. The gel pictures for *secG* mutants are shown in Figure 3.7. Of the colonies screened, at least 7 out of 20 contained a correct insert. Sample 1 and 3 seemed to have an unspecific PCR product in addition to the correct band. The results for *secM* are shown in Figure 3.8, 3 out of 20 mutants contained the correct insert. For *ffh*, 56 colonies were tested (Figure 3.9), and the colony PCR were further optimized by gradient PCR (Figure 3.10), yet, no mutant colonies were obtained. Another round of CRMAGE were run to create *ffh* mutants, with the same negative results.

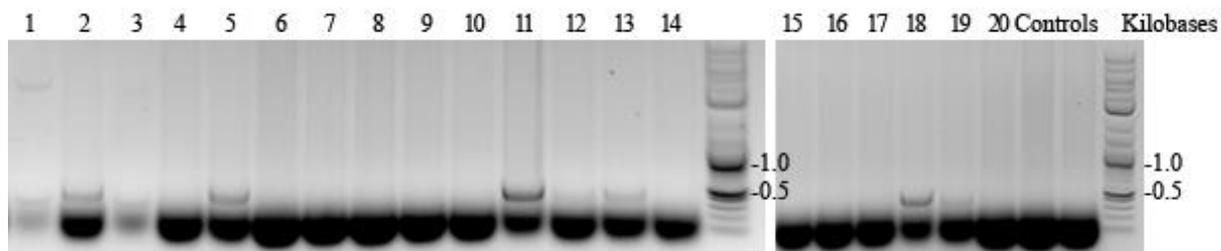


Figure 3.7: Gel pictures from screening for *secG* mutants, using colony PCR. Twenty colonies were screened, and run with two negative controls. Seven colonies had the expected band (sample 2, 5, 11, 12, 13, 18, and 19). Two additional colonies also had the expected band, but they had an unspecific band in addition (sample 1 and 3). The negative controls were MG1655 without mutations.

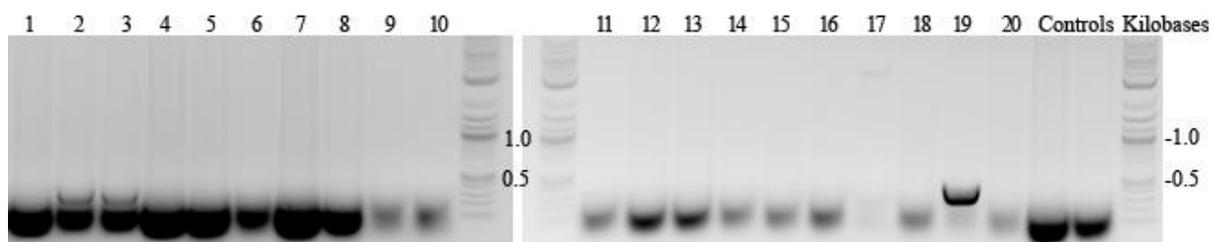


Figure 3.8: Gel pictures from screening for *secM* mutants, using colony PCR. Twenty colonies were screened, and run with two negative controls. Three colonies had the expected band (sample 1, 2, and 19). The negative controls were MG1655 without mutations.

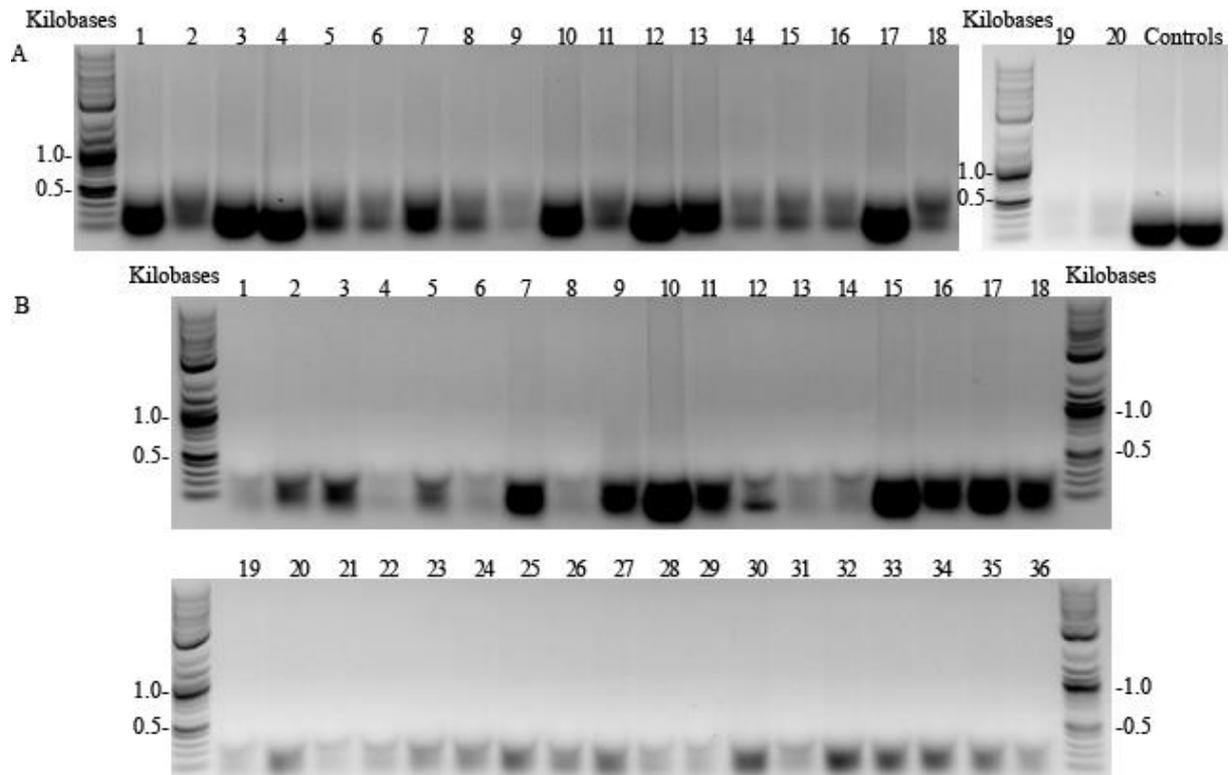


Figure 3.9: Gel pictures from screening for *ffh* mutants, using colony PCR. A: the first 20 colonies screened, and two MG1655 without mutations as negative controls. B: 36 colonies screened later. No colonies had the *ffh* mutation.

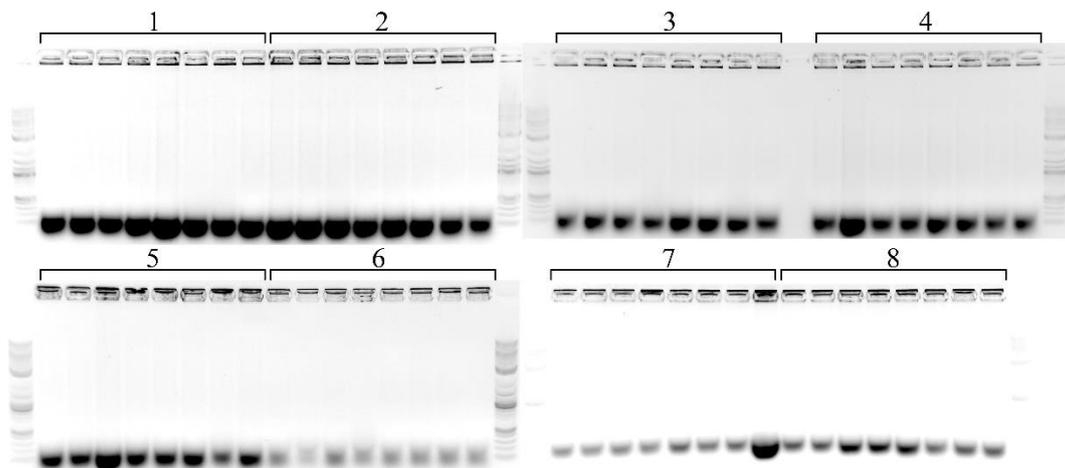


Figure 3.10: Gel pictures from gradient PCR for the optimization of screening for *ffh* mutants. Eight colonies were run, in eight different temperatures, from 55-65 °C. No colonies had the *ffh* mutation.

The PCR products of four *SecG* mutants and all three *secM* mutants were sequenced. The inserts of the *secG* mutants were correct, but all three *secM* mutants had off-target mutations causing a shift in the reading frame, in addition to the target mutation. The sequencing results is shown in Appendix 2. A summary of the off-target mutations in *secM* mutant strains is shown in Table 3.2. All mutant strains had an early stop codon. Two of the strains had the same reading frame from amino acid 46. All the strains were viable, even with the off-target mutation. Therefore, one strain was chosen for further study, of how the mutation affected the levels of periplasmic translocation. Sample 2 was chosen, as it had only one mutation, located outside the sequence of the target mutation. Even though the effects of the off-target mutation of sample 1 and 2 ultimately were the same, sample 1 had two mutations. Sample 3 had a mutation inside the inserted signal peptide, and had a stop codon, terminating translation earlier.

Table 3.2: A summary of the off-target mutations in *secM* mutant strains. The nature and position of the mutation is shown, along with a description of the effect.

Sample	Mutation	Effect
1	Insertion of G after bp 111	Frame shift from amino acid 38
	Deletion of GC 136 and 137	Frame shift from amino acid 46, termination of translation after 89 amino acids
2	Deletion of G 104	Frame shift from amino acid 35, termination of translation after 89 amino acids
3	Insertion of C after bp 47	Frame shift from amino acid 17, termination of translation after 57 amino acids

### 3.3.2 Mutant Strain with Double Mutation

The chosen *secM* mutant were cured of pMAZ-SK\_*secM*, to be ready for another round of CRMAGE. After curing of the plasmid, the *secM* mutant was transformed again with pMAZ-SK\_*secG*, for the creation of *secM* + *secG* double mutants. Colony PCR was used to screen for *secG* mutants, 18 out of 20 colonies screened contained the correct insert in the genome. The gel picture is shown in Figure 3.11. The first five positive samples and the two negative samples were also screened with primers targeting *secM*, to confirm that they all had the original mutation (Figure 3.12). It was confirmed that all seven colonies had the mutation they should have started out with, before the second round of CRMAGE. Thus, the construction of double

mutants was confirmed. The PCR products were sequenced, to confirm that the mutants had the correct insert.

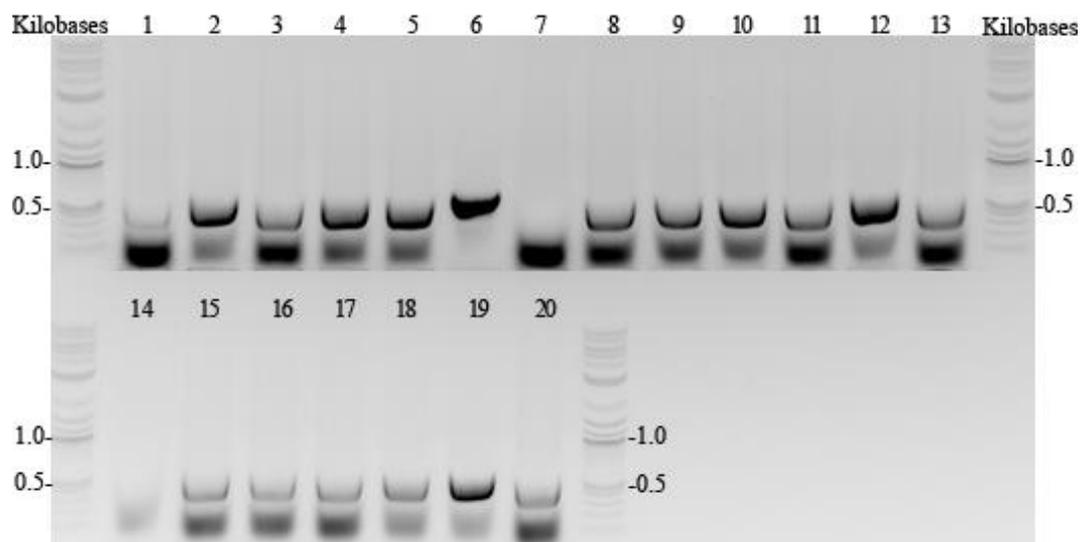


Figure 3.11: Gel picture from screening for introduced *secG* mutation in a *secM* mutant strain, using colony PCR. Twenty colonies were screened, 18 had the correct mutation.

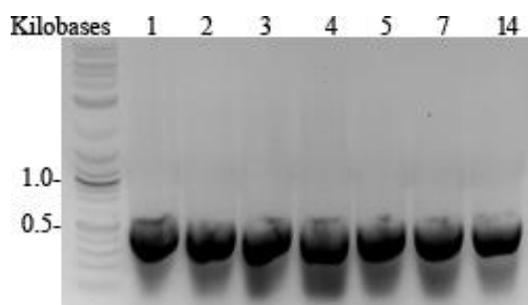


Figure 3.12: Gel picture confirming that the strain used for construction of double mutants carried the first mutation (*secM*). The sample numbers correspond to the sample numbers in Figure 3.11. Five samples with the *secG* mutation were run (sample 1-5), and the two samples without the *secG* mutation (sample 7 and 14).

### 3.4 Assembly of Vector Expressing Reporter Protein

Gibson cloning were used for the assembly of the pVB1-251 backbone and fragments containing the sequence of the ompA-IgG2 Fc-sfGFP fusion protein. Colonies were screened by testing the fluorescence. Inducible fluorescence would mean that sfGFP were able to fold correctly, and that its expression was controlled by the *Pm* promoter. Thus, it was an indication

that the plasmid was correctly assembled. A graph of the measured fluorescence is showed in Figure 3.13. Out of fifteen colonies initially obtained after Gibson assembly, one exhibited induced fluorescence. It also exhibited slightly higher fluorescence when not induced, due to leakiness of the promoter system. Correct assembly of the plasmid were confirmed by sequencing. The sequencing result are in Appendix 2.

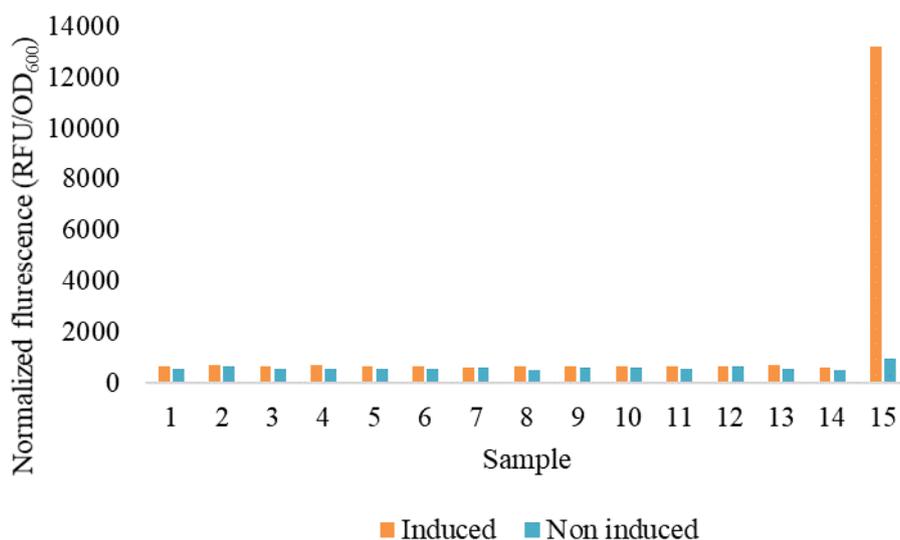


Figure 3.13: The results of the initial fluorescence test after Gibson assembly of pVB1-251 backbone and sequences for ompA-IgG2 Fc-sfGFP fusion protein. The strain used for analysis was DH5 $\alpha$ . The fluorescence is shown in relative fluorescence units (RFU), and it is normalized for OD<sub>600</sub>. The fluorescence was measured with excitation 488 nm and emission 530 nm.

### 3.4.1 Western Blot and SDS-PAGE for Identification of the Reporter Protein

Western blot and SDS-PAGE were used to identify the IgG2 Fc-sfGFP fusion protein, as described in section 2.20. An induced sample of MG1655 pAE001 was compared to a non-induced sample and WT MG1655, without pAE001. The fluorescence of the samples were measured to confirm the presence of sfGFP in the induced samples before running the gels. Three almost identical SDS-PAGE gels were run, one for SDS PAGE, and two for western blot. Western blots were run with primary antibodies against both IgG Fc fragment (Figure 3.14) and sfGFP (Figure 3.15). The SDS PAGE gel and the gel for western blot targeting IgG Fc were run with IgG Fc standards. The IgG Fc standards were used as positive controls to be able to define whether any potential problems are related to the antibody binding, the method or incorrect preparation of the samples. No standards were used for the sfGFP blot. The presence of GFP were confirmed by fluorescence, thus no bands would have meant problems with

antibody binding or methodology. The expected weights of fusion protein components are listed in Table 3.3.

There were no bands visible on the membrane with the primary IgG Fc antibody, except the IgG Fc standards. As the bands of the standard were visible, the western blot procedure had worked and the primary antibody seemed to bind IgG Fc. On the membranes with sfGFP, several bands were visible. There were bands of high concentrations corresponding to the size of both the fusion protein monomer and the fusion protein dimer, in reduced and non-reduced samples, respectively. There were some more bands which were likely to be degraded forms of the protein, and one band that was likely to be sfGFP alone. There were few visible differences between induced and non-induced MG1655 pAE001. On the blot with the lowest concentration of primary antibody, one band, that could be the reduced form of the reporter protein, is much stronger for the induced sample, than the non-induced, and also much stronger for insoluble than soluble.

The samples of MG1655 pAE001, both induced and non-induced, had dark smears. This is a known problem with IgG Fc, and is likely caused by its binding to other proteins (personal communication with Jostein Malmo, Vectron Biosolutions). Thus, the results from the sfGFP blot indicates that the complete fusion protein is present in the cells, even though there are no bands on the IgG Fc blot. The reason the antibody does not bind IgG could be that the sfGFP are blocking the binding site. The primary antibody used for binding to IgG Fc were monoclonal, thus all the antibodies bind the same binding site on IgG Fc. If then this binding site is blocked, none of the antibodies would be able to bind. However, on the sfGFP membrane there are bands that could be sfGFP alone. There are no bands corresponding to the size of IgG Fc on the IgG Fc membrane. This could indicate that the antibodies have trouble binding at all, but it could also mean that the concentration of IgG Fc alone is too low to be detected when using this concentration of antibodies.

Table 3.3: The expected molecular weights (kDa) of the protein components of the OmpA-IgG2 Fc-sfGFP fusion protein.

Protein component	Weight (kDa)
sfGFP	26.7
IgG2 Fc	25.1
IgG2 Fc-sfGFP	52.7
OmpA-IgG2 Fc-sfGFP	54.7

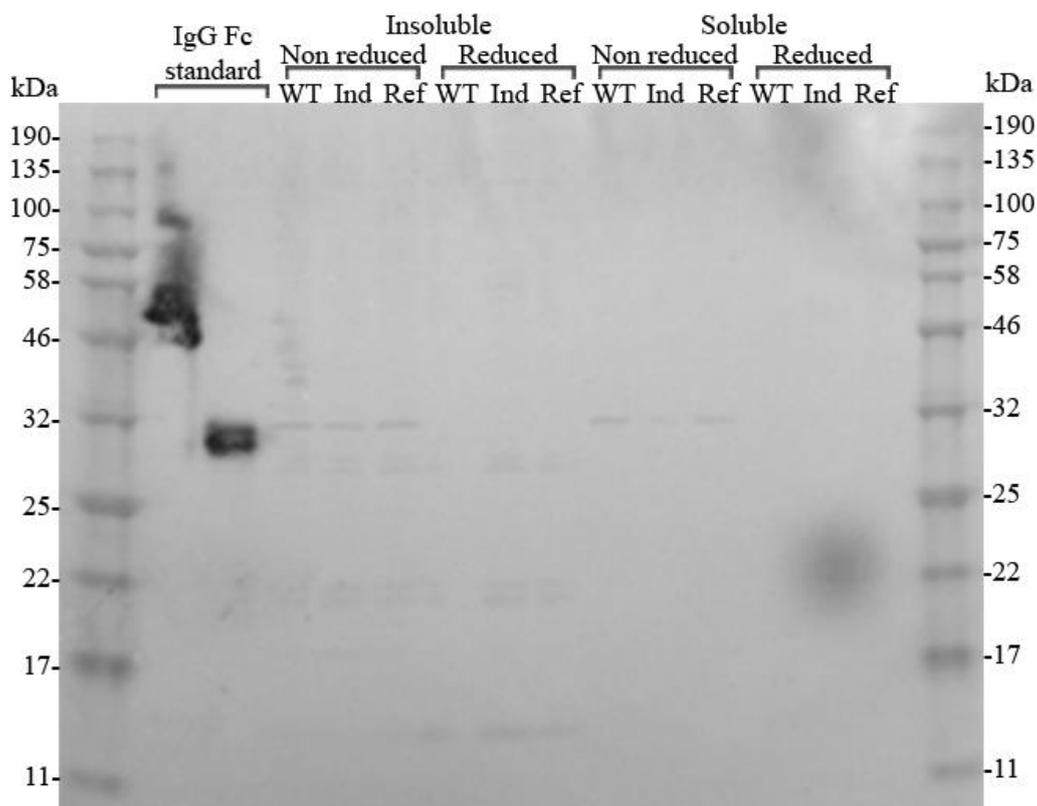


Figure 3.14: The membrane from the western blot with primary antibodies for IgG Fc. The induced sample of MG1655 pAE001 (Ind) was compared to non-induced (ref) and WT MG1655 (WT), without pAE001. Both the soluble and insoluble proteins were run, in reduced and non-reduced form. IgG Fc standard was run in the first two lanes. There were no differences between WT MG1655 and MG1655 pAE001.

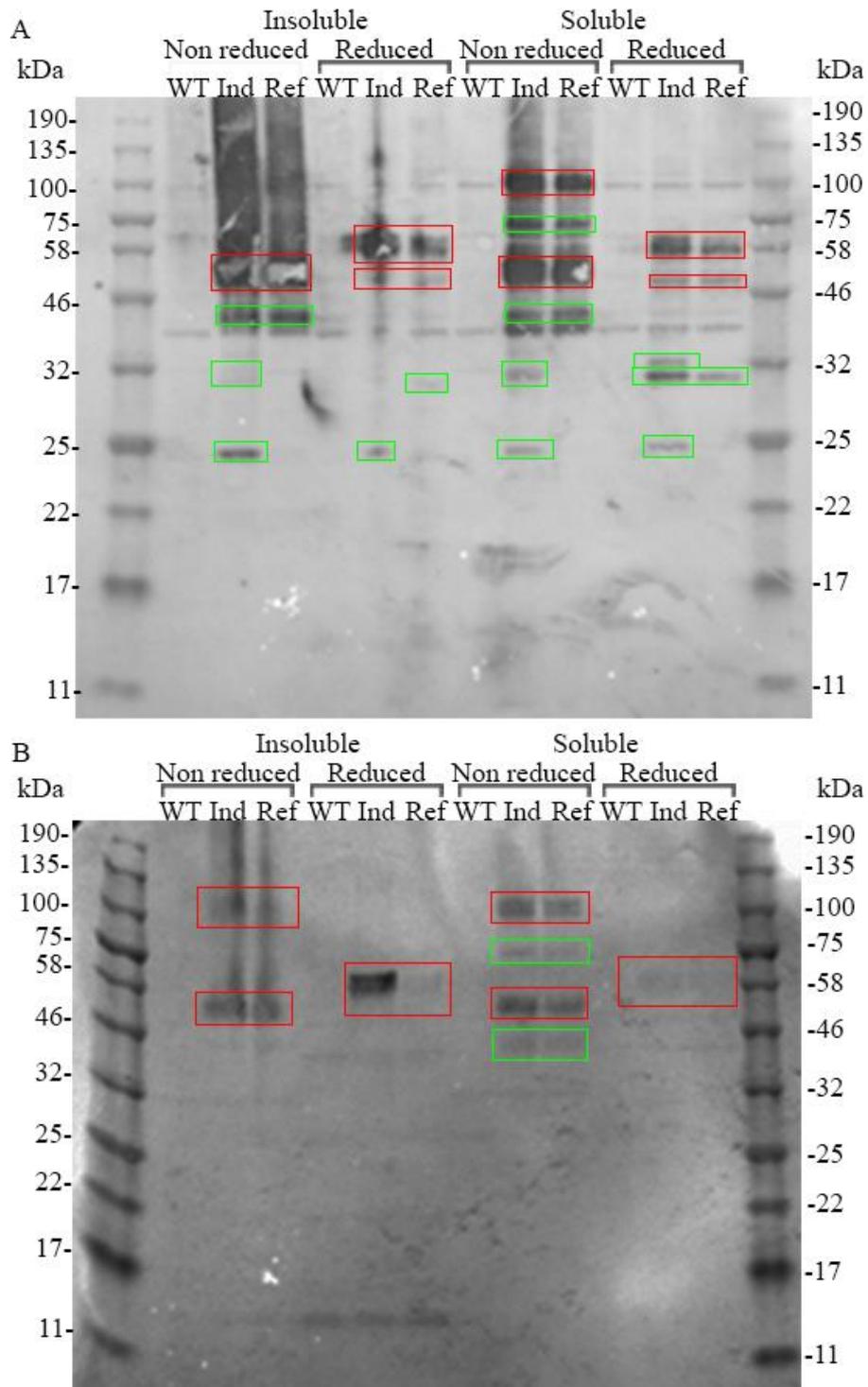


Figure 3.15: The membrane from the western blot with primary antibodies for sfGFP. The induced sample of MG1655 pAE001 (Ind) was compared to non-induced (Ref) and WT MG1655 (WT), without pAE001. Both the soluble and insoluble proteins were run, in reduced and non-reduced form. A: Primary antibody diluted 1:6000. B: Primary antibody diluted 1:600 000. The red rectangles mark bands that could correspond to the fusion protein, with or without OmpA, either as monomer (~57 kDa or ~53 kDa), or as dimer (~114 kDa or 106 kDa) The green bands mark bands that could correspond to sfGFP alone (~27 kDa) or other degraded products of the fusion protein containing sfGFP.

There were no bands clearly visible on the SDS PAGE gel corresponding to the bands identified on the sfGFP blot. However, there were some minor differences between MG1655 pAE001 induced and non-induced, and MG1655 WT. As no clear bands corresponding to the bands identified on the sfGFP blot can be identified, the concentration of the fusion protein seems to be low. The presence of sfGFP in the cultures used for protein analysis were confirmed by fluorescence. However, the relationship between fluorescence and concentration of sfGFP was not defined. Thus, it is possible that the strong fluorescence measured corresponds to a small amount of protein.

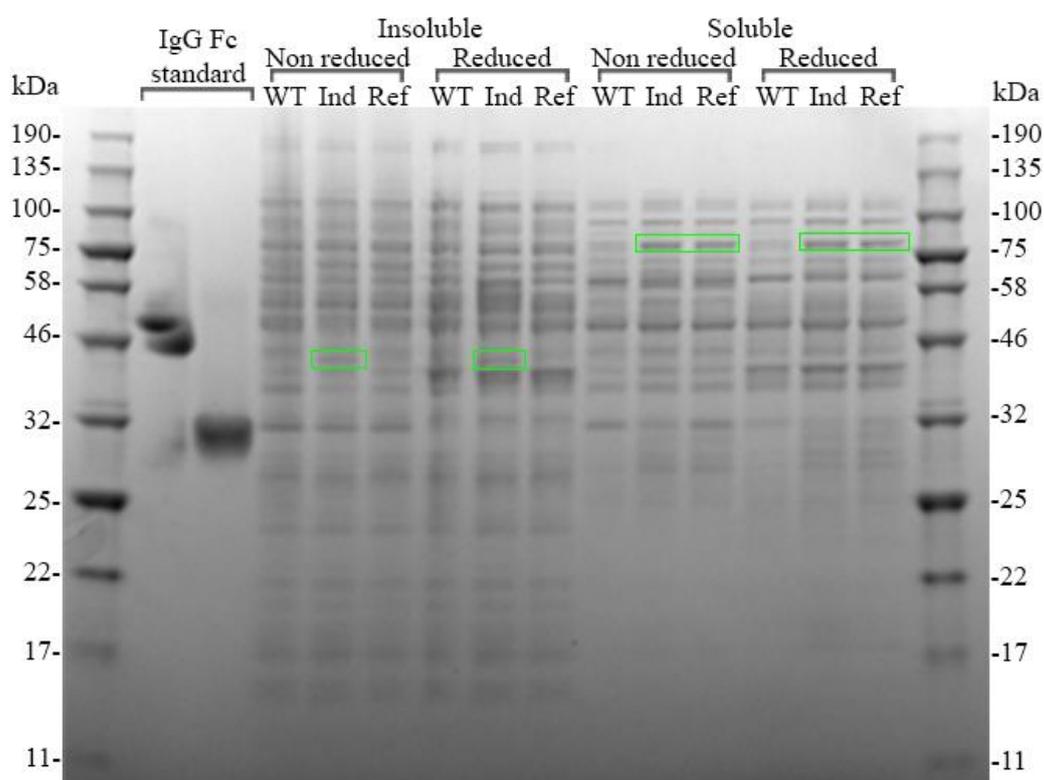


Figure 3.16: The SDS PAGE gel corresponding to the western blot membranes. The induced sample of MG1655 pAE001 (Ind) was compared to non-induced (Ref) and WT MG1655 (WT), without pAE001. Both the soluble and insoluble proteins were run, in reduced and non-reduced form. IgG Fc standard was run in the first two lanes. Bands differentiating the different samples are marked in green.

### 3.4.2 Growth Experiment with MG1655 Expressing the Reporter Protein

The growth of MG1655 pAE001 was measured, both from induced and non-induced samples, and compared to the growth of MG1655 WT. The growth was measured at 37 °C and 30 °C. The growth was measured by measuring OD, and the fluorescence was measured at the same time points. The growth curves are shown in Figure 3.17. As expected, the growth was

considerably lowered at 30 °C compared to 37 °C. The growth rates of *E. coli* MG1655 was clearly affected by pAE001, especially when expression of the reporter protein was induced.

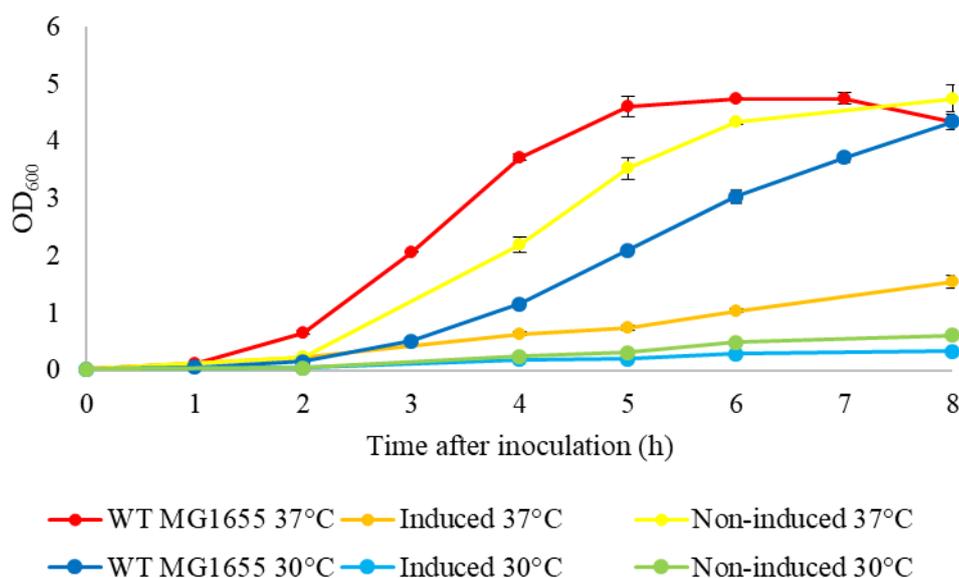


Figure 3.17: Growth curves for WT MG1655 and MG1655 pAE001, induced and non-induced, grown at 37 °C and 30 °C. The growth is measured by OD<sub>600</sub>, and the time after inoculation is measured in hours. The induced samples were induced 2 h after inoculation. Each value is an average of biological triplicates, and error bars are the standard deviation of mean.

The fluorescence measured at the same time points as the growth is shown in Figure 3.18, and the fluorescence normalized by OD<sub>600</sub> is shown in Figure 3.19. The fluorescence was highest in the induced samples were grown at 30 °C. This was especially evident when the fluorescence was normalized with OD. At 30 °C each cell is expressing more correctly folded sfGFP than at 37 °C. The normalized fluorescence increased for two hours after induction, then it began sinking. The total fluorescence increased for two hours at 37 °C and for three hours at 30 °C. Some leaky expression was seen, but the induced samples had many times higher fluorescence.

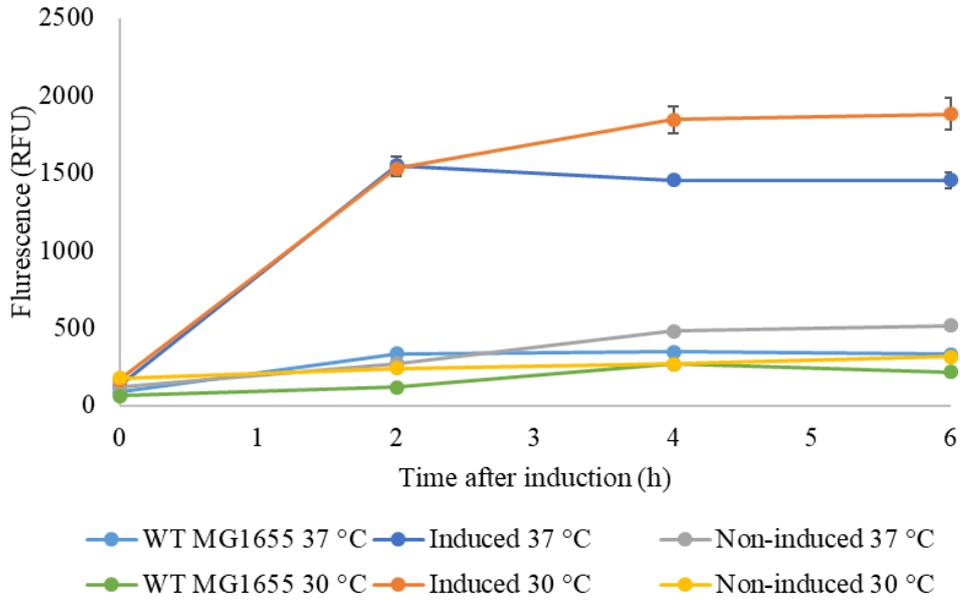


Figure 3.18: The fluorescence of WT MG1655 and MG1655 pAE001, induced and non-induced, grown at 37 °C and 30 °C, measured during growth experiment. Fluorescence is given in relative fluorescence units (RFU), and time after induction in hours. The point of induction was 2 h after inoculation. Each value is an average of biological triplicates, and error bars are the standard deviation of mean. The fluorescence was measured with excitation 488 nm and emission 530 nm.

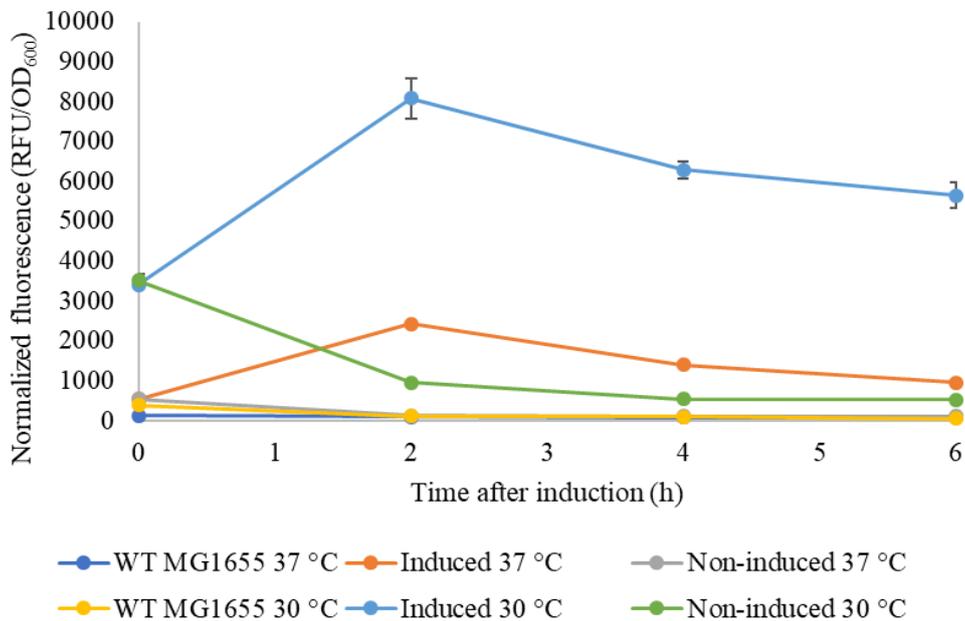


Figure 3.19: The fluorescence of WT MG1655 and MG1655 pAE001, induced and non-induced, grown at 37 °C and 30 °C, normalized for OD<sub>600</sub>. The normalized fluorescence is given in relative fluorescence units (RFU)/OD<sub>600</sub> and time after induction is given in hours. The point of induction was 2 h after inoculation. Each value is an average of biological triplicates, and error bars are the standard deviation of mean. The fluorescence was measured with excitation 488 nm and emission 530 nm.

### 3.5 Expression of the Reporter Protein in the Mutant Strains

The mutants were cured of all CRMAGE plasmids, then transformed with pAE001. The fluorescence from sfGFP of the mutant strains was compared to the fluorescence of MG1655 pAE001. The samples were grown until  $OD_{600}$  was approximately 2, then the samples were induced and incubated at 16 °C for 16h. The fluorescence was measured at the point of induction (Figure 3.20), after 2 h incubation (Figure 3.21), and after 16 h incubation (Figure 3.22). At 16 h the gain had to be slightly lowered to measure the fluorescence, due to limited capabilities of the apparatus, thus the fluorescence cannot accurately be compared between the timepoints. When normalized by OD, the fluorescence of the WT strain was highest. However, after 16 h there was very little difference between the WT and the *secM* mutant. The difference between the *secG*- and the *secM* + *secG* mutant was also very small.

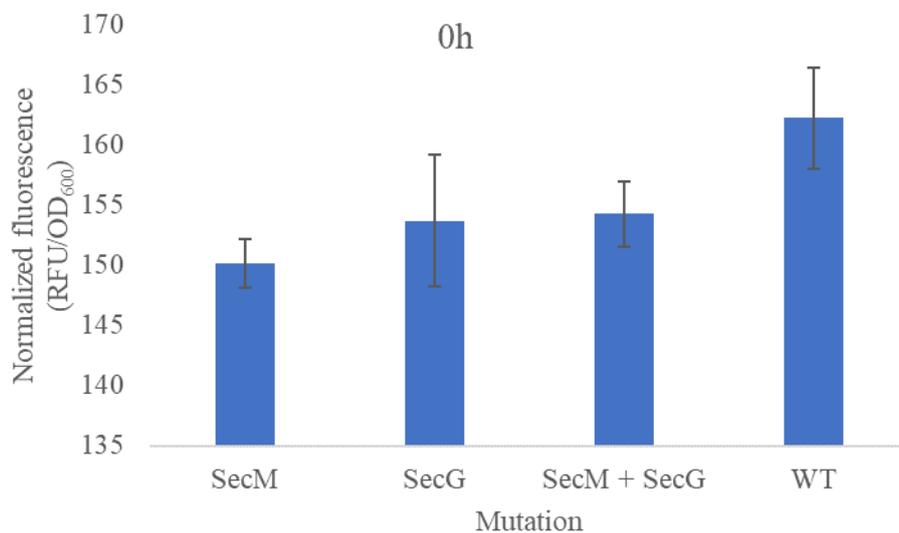


Figure 3.20: The fluorescence measured at the point of induction, when expressing the IgG Fc-sfGFP fusion protein from pAE001. The fluorescence is given in relative fluorescence units (RFU), normalized by  $OD_{600}$ , for each of the different mutants created, and compared to the WT strain. The values are the mean of biological triplicates, and the error bars are the standard deviation of the mean. Measured with gain 80.

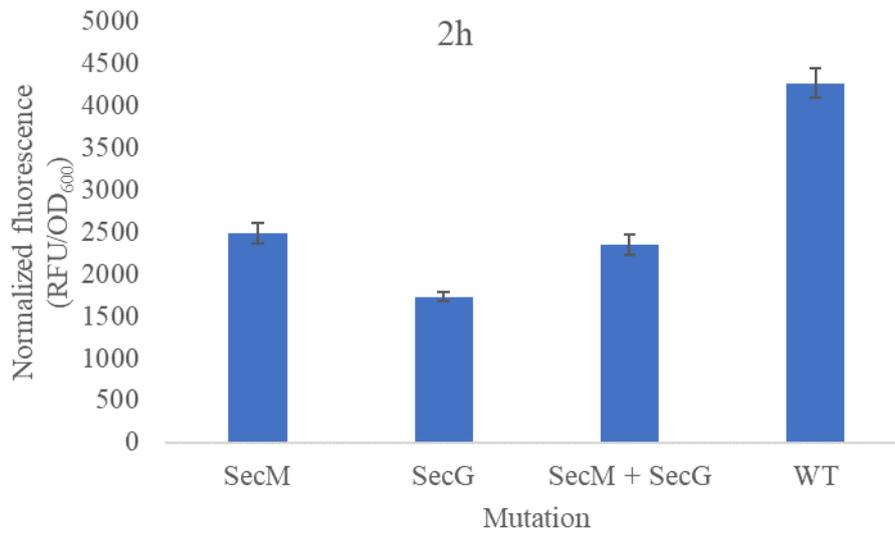


Figure 3.21: The fluorescence measured 2 h after induction, when expressing the IgG Fc-sfGFP fusion protein from pAE001. The fluorescence is given in relative fluorescence units (RFU), normalized by OD<sub>600</sub>, for each of the different mutants created, and compared to the WT strain. The values of the mutant strain are the mean of biological triplicates, and the value of the WT is the mean from a biological duplicate. The error bars are the standard deviation of the mean. Measured with gain 80.

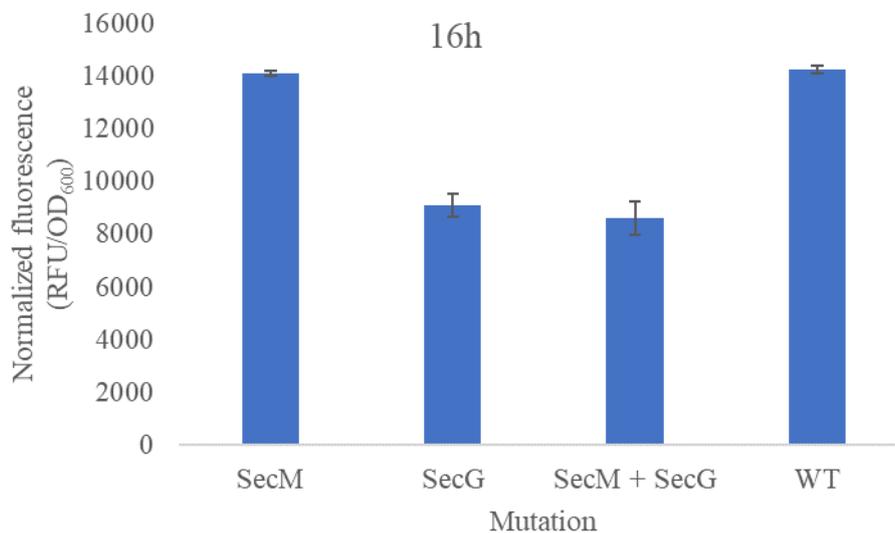


Figure 3.22: The fluorescence measured 16 h after induction, when expressing the IgG Fc-sfGFP fusion protein from pAE001. The fluorescence is given in relative fluorescence units (RFU), normalized by OD<sub>600</sub>, for each of the different mutants created, and compared to the WT strain. The values of the mutant strain are the mean of biological triplicates, and the value of the WT is the mean from a biological duplicate. The error bars are the standard deviation of the mean. Measured with gain 77.

The growth of the mutant strains, with pAE001, was studied by measuring OD<sub>600</sub> every other hour for 4 h after inoculation, before induction (Figure 3.23). The growth of the *secG* mutant was reduced. The growth of the *secM* mutant and the double mutant was slightly higher than the growth of the WT.

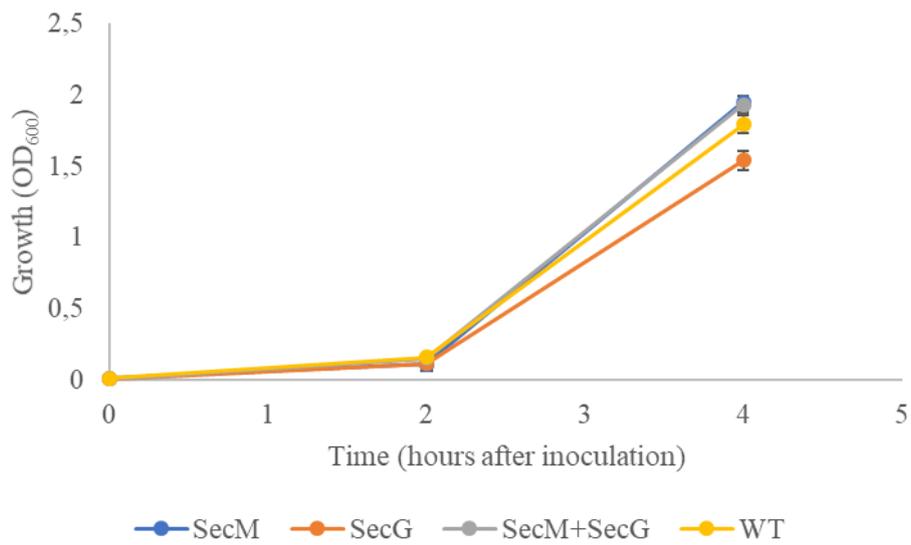


Figure 3.23: The growth of the mutant strains created and the WT, with pAE001. The growth was measured by OD<sub>600</sub> for 4 h after inoculation, before the cultures were induced for expression of the reporter.

## 4 Discussion

### 4.1 Establishing and Optimizing CRMAGE

The main problem of establishing CRMAGE at our laboratory was identified as achieving correct assembly of pMAZ-SK with a changed gRNA insert. Plasmids that seemed to be correctly assembled after colony PCR were revealed to contain the original gRNA when sequenced. Once the method of assembly was changed from USER cloning to Gibson assembly, correctly assembled plasmids were achieved with high efficiency and colony PCR seemed to be reliable for screening.

The authors of the CRMAGE protocol, Ronda et al., calculated the efficiency of CRMAGE by creating three knock out mutants, *galK*, *xylA*, and *lacZ* (17). The genes were knocked out by the substitution of 1 bp, resulting in the formation of a stop codon. The CRMAGE efficiencies were measured to 98 %, 99.7 %, and 96.5 %, respectively. They also measured the efficiencies without negative selection by CRISPR/Cas9, using MAGE alone. These were measured to 5 %, 0.6 %, and 3.6%, respectively (17). Thus, CRMAGE seems to be a great improvement of MAGE. The mutation efficiencies achieved in this work with the creations of *secM*, and *secG* mutants were much lower. The efficiencies achieved were 15 % for the *secM* mutation, and 35 % for *secG*. However, these mutations were also much larger than one single bp.

Ronda et al. tested CRMAGE for the introduction of larger mutation by introducing a 6 bp substitution (17). They changed the RBS sequence in front of GFP, in a strain with GFP inserted into the genome. For this longer mutation, they achieved an efficiency of 62 % for CRMAGE and 6 % for MAGE (17). In this work, the *secM* oligo introduced an insert of 57 bp, and the *secG* oligo introduced an insert of 17 bp. Thus, when both the sizes of the mutations and the achieved efficiencies are compared the results seems to be in agreement.

In addition to the mutations introduced in this work being larger, the MAGE oligos used were longer. Both these factors could explain the lowered efficiency of  $\lambda$  Red recombineering. Larger mutations lead to lower MAGE efficiencies. Longer oligos can increase the efficiencies, but it also increases the risk for formation of secondary structures which can lower the efficiencies (18). The design of MAGE oligos is therefore a difficult balance. Especially, as the formation

of secondary structures is dependent on the sequence, meaning the optimal length of the homology regions will vary between each mutation. Oligos with shorter overlaps could have been designed and tested. These could potentially be easier and cheaper to have synthesized. Also, if the oligos used forms secondary structures, shorter homology regions could have increased the efficiency.

There is still room for optimizing the protocol of the CRMAGE method in our laboratory, both regarding simplifying the protocol and increasing the efficiency. The protocol for electroporation used by Ronda et al. are slightly more laborious than the protocol for electroporation otherwise applied with the work of this thesis. As the creation of electrocompetent cells is the most work intensive part of the CRMAGE protocol, simplifying it would be of interest. However, this was not tested. If the protocols for electroporation could be exchanged without losing efficiency, this would be a great modification of the CRMAGE protocol. If not, reducing the amount of labour could sometimes be worth the cost of slightly lowering the efficiency.

In the original CRMAGE protocol the cells were incubated for two hours with aTet to induce CRISPR/Cas9 for negative selection. The duration of this incubation was a step that could potentially be optimized for increasing the mutation efficiency. After consultations with the group of prof. Alex Toftgaard Nielsen from Technical University of Denmark, the protocol was modified (17, personal communication with Ida Lauritsen and Alex Toftgaard Nielsen, Technical University of Denmark). The incubation time was extended to overnight incubation, approximately 18 hours. This change was kept in the protocol used for CRMAGE, but the effect of it was not studied in comparison to two hours incubation. The change decreased the length of the protocol on the first day, but increased the protocol by one day in total. It increased the time the cells were exposed to the inducer; therefore, it could have increased the efficiency of the negative selection. However, Ronda et al. reported almost 100 % killing of the WT after three hours incubation (17), thus it is possible that overnight incubation is longer than necessary. It is likely that the optimal incubation time would differ for different mutations. If the mutation causes reduced growth rate in mutants compared to the WT, it is likely that prolonged incubation increases the percent of WT cells in the population, thus lowering the efficiency.

## 4.2 Frame Shift Mutations in *secM* Mutants

The mutation frequency was lower for the *secM* mutation than the for the *secG* mutation. The homology regions of the two MAGE oligos were the same length, but the *secM* mutation were larger, than the mutation of *secG*. Larger mutations can lower the efficiency of  $\lambda$  Red recombineering (18). The larger mutation also means that the overall length of the *secM* MAGE oligo were longer. As mentioned, this increases the possibility that the oligo forms secondary structures, which can also lower the efficiency of the recombineering through less frequent binding to the homologous DNA sequence (18).

However, sequencing results revealed that all *secM* mutants had different off-target mutations. All three mutants had frame-shift mutations early in the sequence of the gene. Thus, the whole amino acid composition was changed, and translation was stopped by an early stop codon. This could indicate that the target mutation had deleterious effects on the cells, and that these off-target mutants were more viable than the target mutants. A deleterious mutation could create a selective pressure toward having an off-target mutation. Cells with a off-target mutation would outcompete those that does not, and the off-target mutants become more prevalent in the population. If the *secM* target mutation is deleterious this could be a possible explanation for why the mutation efficiency was lower than for *secG* mutants.

The native signal peptide of SecM is atypical. It is longer than usual, and also contain several atypical amino acids (14). The function of SecM is regulating the translation of *secA*. The mechanism of regulation is a translational pause during translation of SecM, which unfolds a secondary structure in the mRNA and exposes the Shine-Dalgarno sequence of *secA*. The translational pause is caused by the translation of an arrest peptide at the end of the gene, which halts the ribosome (49). The translation is then continued when the nascent polypeptide makes contact with the translocon. The signal sequence is important for making this contact (16). Overproduction of SecM, using *secM* with a defect signal sequence have been found lethal. This could possibly be due to a large number of ribosomes being bound to the mRNA in translational pause, affecting the translation of other proteins (50). As the target mutation was a change in the signal sequence of SecM, this is a likely explanation if the target mutation were in fact deleterious.

However, it is difficult to explain why the off-target mutation would cause higher viability. The off-target mutations frame-shifts during the translation of *secM*. This causes a shorter open reading frame, with a stop codon long before the arrest peptide. A simple conclusion would then be that the off-target mutation abolishes the translational pause, thus saving the cell. However, mutations in the sequence of the arrest peptide, with the effect of abolishing the translational pause has also been shown lethal. The translation of *SecM*, with the translational pause, is required for the cell to express the necessary levels of *SecA* (49). For the cells to be viable, the off-target mutation must allow a minimum expression of *SecA*, while also allowing release from the translational pause.

Another possible explanation for the off-target mutations are technical difficulties with the recombination. As all the off-target mutations are within the sequence of the MAGE oligo, this is a possibility. The off-target mutations are all different. Thus, there are nothing wrong with the sequence of the oligos. However, the insert is long, the homology regions are long, by extension making the MAGE oligo long. It is possible that this increases the possibility of mistakes happening during recombination, for example through the formation of secondary structures. This could be investigated by testing shorter homology regions. Other MAGE oligos with the same length of insert and homology regions could also be tested, to see whether they give the same effect.

### **4.3 Higher Mutation Efficiency when Introducing Second Mutation**

When introducing the *secG* mutation to one of the *secM* mutant strains, 18 of 20 colonies screened were double mutants. The mutation efficiency for creating double mutant were more than doubled compared to the mutation efficiency for creating *secG* single mutants. This is in agreement with the results of Ronda et al. (17). They tested CRMAGE for multiplexing by introducing two mutations simultaneously, the *galK* mutation and the GFP mutation previously mentioned. They used a pMAZ-SK with two sgRNAs. When screening colonies carrying one mutation, they found that all screened colonies also carried the other mutation. They had 100 % efficiency for simultaneously introducing both mutations. Thus, their population contained only, or almost only, non-mutants or double mutants. When they did the same experiment with MAGE alone, without negative selection by CRISPR/Cas9, they had only 10 % efficiency of simultaneously introducing both mutations. Meaning, only 10 % of colonies carrying one

mutation also carried the other (17). In this work the mutations were introduced sequentially, not simultaneously as by Ronda et al. (17). Yet, even though the methods for creating double mutants were different, it seems reasonable that the same mechanism is responsible for the effect in both works. Especially as the effect is so pronounced using both methods.

As the difference in efficiency for simultaneously introducing two mutations between CRMAGE and MAGE is so large, it is possible that the negative selection by CRISPR/Cas9 is the cause of the effect. Ronda et al. suggests that some cells are able to escape the negative selection by CRISPR/Cas9 due to mutations in their PAM sequence or a region 8 bp upstream of the PAM (17). The expression of Cas9 and sgRNA could be slightly leaky, creating a selection pressure towards having a such mutation, before the inducer is added. However, they also report that the abundance of such escapers is very low (17). The efficiency of generating *secG* mutants is higher when the mutation is introduced to *secM* mutants than when it is introduced to the WT. As the same protocol is used, it seems as if having already performed one round of CRMAGE somehow makes the target mutation more abundant or the escape mutation less abundant.

Knowing why the efficiency is increased for the construction of double mutants could be valuable for future use of the CRMAGE method, if introducing more than two mutations, or if colony PCR cannot be used for screening. It would also be interesting to study whether the effect applies to more than two mutations, both when introduced by multiplexing and when introduced in sequence.

#### **4.4 No *ffh* Mutants Were Created**

Mutant strains with the *secG* and *secM* mutations were successfully created using CRMAGE. However, after two rounds of CRMAGE and extensive screening and optimization of screening methods, no tested colonies had the *ffh* mutation. In previous studies, *ffh* has been put under the control of a promoter which is inducible by arabinose (10, 51). The strains were then shown to be dependent on arabinose, proving that the Ffh is essential for growth of *E. coli* (10). In another study Ffh were depleted by first growing the cells in the presence of arabinose, then after washing and diluting the cells were grown without arabinose. The depletion of Ffh was then

shown to disrupt the proper assembly of membrane proteins in the cytoplasmic membrane (51). Thus, if the combined downregulatory effect of the changed ribosomal binding site and the changed start codon were too high, the target mutants were likely to be unviable. This would explain why the generation of *ffh*-mutants were unsuccessful. The failure to properly assemble and integrate the cytoplasmic membrane proteins could lead to the loss of many functions, and affect the integrity of the membrane.

To find a mutation that would have caused a viable, preferably optimal, level of downregulation, several MAGE oligos could have been designed and tested. Ribosomal binding sites could have been created with different target translation initiation rates. These could have been tested in combination with different start codons, and without changing the start codon. As an insert larger than 10 bp is required for proper binding of an insert-specific primer during screening with colony PCR, only changing the start codon could not have been done. A single base pair change could not have been detected with PCR primers. All colonies would have had to be screened by sequencing, increasing both the cost and time. Then, mutants with varying degrees of *ffh* downregulation could have been tested to find a viable strain. If the strains had changed levels of translocation, different levels of downregulation could have been tested to find the optimum.

#### **4.5 Reliability of the Reporter Protein**

For sfGFP to fluoresce it needs to be properly folded (28). The IgG Fc-sfGFP fusion protein was expressed with the OmpA signal peptide in order to be translocated to periplasm through the post-translational pathway of the Sec-pathway. In *E. coli*, the chaperone SecB keeps proteins unfolded before translocation, and then helps the signal peptide bind SecA. It was therefore hypothesized that the chaperone SecB would keep the whole protein unfolded until after translocation. The sfGFP would then not be able to fluoresce in cytoplasm, but when translocated it would be able to fold properly and give fluorescence. Thus, the level of fluorescence, as measured by a plate reader, could serve as reporter. However, it is possible that the improved folding capacities of the sfGFP allowed it to fold prematurely in cytosol (28). The assumption that sfGFP would only be correctly folded in periplasm could be confirmed using fluorescence microscopy. This would allow the detection of the exact localization of folded

sfGFP. The assumption could also be confirmed by comparing the fluorescence of a lysate to the fluorescence of a periplasmic lysate.

If sfGFP folds prematurely, it would be unfit as a reporter for translocation levels through the Sec-pathway. Not only because it fluoresces in the cytoplasm and a plate reader would not be able to detect whether the fluorescence is from cytoplasm or periplasm. Also, as the translocons of the Sec-pathway are not able to translocate folded proteins (7, 9), the level of translocation for the reporter would be affected. Thus, fluorescence microscopy and periplasmic lysis would also be unreliable methods for studying translocation in the mutants.

If the IgG Fc-sfGFP fusion protein had been proven unfit as a reporter, other alternatives could have been tested. However, these reporters would also have had to be studied, and their fitness as reporter in this study assessed. The sfGFP variant of GFP was created to have improved folding capacities as a component of fusions protein (28). It is supposed to be able to fold correctly even when its fusion partner is not correctly folded. Other variants of GFP are more dependent on correct folding of their fused protein to fold correctly itself (28). Such GFP variants would be reporting correct folding of IgG Fc. If IgG Fc is consistently unfolded in cytoplasm and correctly folded in periplasm, other GFP variants could be better suited as reporters for periplasmic translocation. The reporter could also be changed completely, for example into alkaline phosphatase (PhoA). Correct folding and activity of the PhoA enzyme requires the formation of two disulphide bonds (52). Disulphide bonds are more easily formed in the oxidative environment of the periplasm(5), PhoA should therefore be more active after translocation (52). However, the activity in cytoplasm would have had to be measured. This could have been done by expressing IgG Fc fused to PhoA, without any signal peptide.

#### **4.6 Effects of the Introduced Mutations**

The IgG Fc sfGFP fusion protein was expressed, and the fluorescence measured. The results indicate that the introduced mutations affected the translocation of proteins, as expected. However, the effect was not positive. Lower levels of fluorescence were measured from all the mutant strains than from the wild type. The exception was after 16 hours incubation the difference between the *secM* mutant and the wild type was very low, and not significant, as the

standard deviation of the means overlap. Two hours after induction the fluorescence levels of all mutants were clearly lower, thus there could still be a difference between the *secM* mutant and the WT, even though it is not evident after 16 hours incubation.

The *secM* mutants did not seem strongly affected by the mutation. After 16 hours the level of fluorescence was very similar to the fluorescence level of the wild type, and the growth was not reduced. The levels of fluorescence measured for the *secG* mutant and the *secM* and *secG* double mutant were also very close. However, the *secM* mutation was designed to affect translocation also in combination with the *ffh* mutation, not only by itself. The *secM* mutation was supposed to change the atypical signal peptide of SecM, and ensure that SecM would use the SRP pathway for translocation. The *ffh* mutation was then supposed to downregulate the SRP pathway. Combined, the translocation of SecM was supposed to be lowered, thus increasing the expression of SecA. As *ffh* mutants were not successfully created, this hypothesis could not be tested. Also, as the *secM* mutants had off target mutations, it is difficult to say what effect the SecM mutation had, and what the potential combined effect of the two mutations would have been.

Only 20 colonies were screened for the *secM* mutation. Three of these had the target mutation, and all three also had off-target mutations. It is possible that mutants without off-target mutations would have been found if more colonies were screened. Then the effect of the target mutation could have been measured as well. The fluorescence of the reporter was only measured in one of the three mutants. As all three had different off-target mutations, it is possible that each strain was affected in a different way. The fluorescence of all three could have been measured and compared, along with any additional mutants found by screening.

The *secM* mutants seemed to grow slightly faster than the wild type before induction. The wildtype had a higher level of fluorescence at the point of induction, thus a possible explanation is that the wild type spent more energy and resources expressing the reporter. However, the *secM* and *secG* double mutant grew faster than the *secG* mutant before induction, while having the same or possibly higher level of fluorescence. Thus, it is possible that some aspect of the *secM* mutation allowed higher growth rates. Another possibility is that the effect of the *secM*

mutation slightly weakens the effect of the *secG* mutation. As SecG is thought to have a role in cycling binding of SecA to the SecYEG complex (8), it is possible that a slightly higher expression of SecA could ease the effects of SecG downregulation.

The measured fluorescence of sfGFP was lower from the mutant strains than from the WT strain. However, it is difficult to say whether the effect is caused by reduced translocation of the reporter to periplasm or other factors. The mutations could be affect the expression levels of proteins, lowering the amount of translated protein, thus also lowering the amount of translocated protein. Thus, the lowered levels of fluorescence could be caused by changes at the level of transcription, translation, translocation or folding.

## 5 Conclusion

After several tests and a lot of effort, the CRMAGE method were successfully established at our laboratory. Lower efficiencies were achieved than by the original authors of the CRMAGE protocol. A likely explanation is that the efficiencies were lowered due to larger mutations to be introduced. However, the large mutations make screening for mutants easier, reducing the need for high mutation efficiencies. For the purpose of future works, where higher efficiencies might be needed, such as the introduction of smaller mutations or multiplexing, there might still be some room for optimization. The design of MAGE oligos could be optimized, and possibly some aspects of the CRMAGE protocol.

CRMAGE was used to successfully generate two single-mutation strains, *secM* and *secG*. A mutant strain harbouring both mutations were also successfully generated. The two mutations were introduced sequentially, with the *secG* mutation being introduced to the *secM* mutant strain. The efficiency of introducing the *secG* mutation as a second mutation was remarkably higher than when introduced as the only mutation. Further exploration of this effect would be interesting in order to assess the potential of CRMAGE for multiplexing. Generation of a third mutant strain, with a mutation in *ffh* was not achieved. As the *ffh* gene is essential, the explanation is probably the nature of the mutation.

A fusion protein with IgG2 and sfGFP was successfully cloned into pVB1-251. It was successfully expressed and exhibited fluorescence. The fusion protein contained the signal peptide of OmpA, targeting it for translocation through the Sec-pathway. The fusion protein was used as a reporter for measuring the effect of the mutations on the mutant strains. The results indicated that the mutations successfully affected the translocation, lowering it. The translocation was lowered for all mutant strains, but the *secM* mutant strain was only slightly affected. However, it is possible that the decreased fluorescence is due to lowered expression, not only lowered levels of translocation.

## 6 References

1. Overton TW. Recombinant protein production in bacterial hosts. *Drug Discovery Today*. 2014;19(5):590-601.
2. Walsh G. Biopharmaceutical benchmarks 2014. *Nature Biotechnology*. 2014;32:992.
3. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology*. 2014;5(APR).
4. Yoon SH, Kim SK, Kim JF. Secretory production of recombinant proteins in *Escherichia coli*. *Recent Patents on Biotechnology*. 2010;4(1):23-9.
5. Mergulhão FJM, Summers DK, Monteiro GA. Recombinant protein secretion in *Escherichia coli*. *Biotechnology advances*. 2005;23(3):177-202.
6. Dalbøge H, Bayne S, Pedersen J. In vivo processing of N-terminal methionine in *E. coli*. *FEBS Letters*. 1990;266(1-2):1-3.
7. Natale P, Brüser T, Driessen AJM. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-Distinct translocases and mechanisms. *Biochimica et Biophysica Acta - Biomembranes*. 2008;1778(9):1735-56.
8. Rusch SL, Kendall DA. Interactions that drive Sec-dependent bacterial protein transport. *Biochemistry*. 2007;46(34):9665-73.
9. Kudva R, Denks K, Kuhn P, Vogt A, Müller M, Koch HG. Protein translocation across the inner membrane of Gram-negative bacteria: The Sec and Tat dependent protein transport pathways. *Research in Microbiology*. 2013;164(6):505-34.
10. Phillips GJ, Silhavy TJ. The *E. coli* *ffh* gene is necessary for viability and efficient protein export. *Nature*. 1992;359:744.
11. Flower AM, Hines LL, Pfennig PL. SecG is an auxiliary component of the protein export apparatus of *Escherichia coli*. *Molecular and General Genetics MGG*. 2000;263(1):131-6.
12. Belin D, Plaia G, Boulfekhar Y, Silva F. *Escherichia coli* SecG is required for residual export mediated by mutant signal sequences and for SecY-SecE complex stability. *Journal of Bacteriology*. 2015;197(3):542-52.
13. Nishiyama K-i, Suzuki T, Tokuda H. Inversion of the Membrane Topology of SecG Coupled with SecA-Dependent Preprotein Translocation. *Cell*. 1996;85(1):71-81.
14. Sarker S, Oliver D. Critical regions of *secM* that control its translation and secretion and promote secretion-specific *secA* regulation. *Journal of Bacteriology*. 2002;184(9):2360-9.
15. Zhang J, Pan X, Yan K, Sun S, Gao N, Sui SF. Mechanisms of ribosome stalling by SecM at multiple elongation steps. *eLife*. 2015;4(DECEMBER2015).
16. Goldman DH, Kaiser CM, Milin A, Righini M, Tinoco I, Bustamante C. Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo. *Science*. 2015;348(6233):457-60.
17. Ronda C, Pedersen LE, Sommer MOA, Nielsen AT. CRMAGE: CRISPR Optimized MAGE Recombineering. *Scientific Reports*. 2016;6.
18. Wang HH, Church GM. Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Methods in enzymology*. 2011;498:409-26.
19. Alberts B, Johnson A, Lewis J, Morgan D, Raff M, Roberts K, et al. *Molecular Biology of the Cell*. 6. ed. New York: Garland Science; 2015.
20. Costantino N, Court DL. Enhanced levels of  $\lambda$  Red-mediated recombinants in mismatch repair mutants. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(26):15748-53.
21. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*. 2009;460:894.

22. Terns MP, Terns RM. CRISPR-Based Adaptive Immune Systems. *Current opinion in microbiology*. 2011;14(3):321-7.
23. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. 2007;315(5819):1709-12.
24. Wright AV, Nuñez JK, Doudna JA. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell*. 2016;164(1-2):29-44.
25. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337(6096):816-21.
26. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences*. 2012;109(39):E2579-E86.
27. Thie H, Schirrmann T, Paschke M, Dübel S, Hust M. SRP and Sec pathway leader peptides for antibody phage display and antibody fragment production in *E. coli*. *New Biotechnology*. 2008;25(1):49-54.
28. Pédelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology*. 2006;24(1):79-88.
29. Owen JA, Punt J, Stranford SA. *Kuby Immunology*. 7th ed. Basingstoke: Macmillan Higher Education; 2013.
30. Vidarsson G, Dekkers G, Rispens T. IgG Subclasses and Allotypes: From Structure to Effector Functions. *Frontiers in Immunology*. 2014;5:520.
31. Yang C, Gao X, Gong R. Engineering of Fc Fragments with Optimized Physicochemical Properties Implying Improvement of Clinical Potentials for Fc-Based Therapeutics. *Frontiers in Immunology*. 2018;8(1860).
32. Levin D, Golding B, Strome SE, Sauna ZE. Fc fusion as a platform technology: potential for modulating immunogenicity. *Trends in Biotechnology*. 2015;33(1):27-34.
33. Beck A, Reichert JM. Therapeutic Fc-fusion proteins and peptides as successful alternatives to antibodies. *mAbs*. 2011;3(5):415-6.
34. Tsien RY. The green fluorescent protein. *Annual Review of Biochemistry* 1998. p. 509-44.
35. Gawin A, Valla S, Brautaset T. The XylS/Pm regulator/promoter system and its use in fundamental studies of bacterial gene expression, recombinant protein production and metabolic engineering. *Microbial Biotechnology*. 2017;10(4):702-18.
36. Durland RH, Toukdarian A, Fang F, Helinski DR. Mutations in the *trfA* replication gene of the broad-host-range plasmid RK2 result in elevated plasmid copy numbers. *Journal of Bacteriology*. 1990;172(7):3859-67.
37. Anton BP, Raleigh EA. Complete genome sequence of NEB 5-alpha, a derivative of *Escherichia coli* K-12 DH5α. *Genome Announcements*. 2016;4(6).
38. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, et al. The Complete Genome Sequence of *Escherichia coli* K-12. *Science*. 1997;277(5331):1453-62.
39. Grenier F, Matteau D, Baby V, Rodrigue S. Complete Genome Sequence of *Escherichia coli* BW25113. *Genome Announcements*. 2014;2(5):e01038-14.
40. Lutz R, Bujard H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research*. 1997;25(6):1203-10.
41. Madigan M, Martinko J, Bender K, Buckley D, Stahl D. *Brock Biology of Microorganisms*. Global ed. 14th ed. Essex: Pearson; 2015.

42. Nour-Eldin HH, Geu-Flores F, Halkier BA. USER cloning and USER fusion: The ideal cloning techniques for small and big laboratories. *Methods in Molecular Biology* 2010. p. 185-200.
43. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*. 2009;6(5):343-5.
44. Espah Borujeni A, Channarasappa AS, Salis HM. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research*. 2014;42(4):2646-59.
45. Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*. 2009;27:946.
46. Hecht A, Glasgow J, Jaschke PR, Bawazer LA, Munson MS, Cochran JR, et al. Measurements of translation initiation from all 64 codons in *E. coli*. *Nucleic Acids Res*. 2017;45(7):3615-26.
47. Teplyakov A, Zhao Y, Malia TJ, Obmolova G, Gilliland GL. IgG2 Fc structure and the dynamic features of the IgG CH2-CH3 interface. *Molecular Immunology*. 2013;56(1):131-9.
48. Martínez V, Lauritsen I, Hobel T, Li S, Nielsen AT, Nørholm MHH. CRISPR/Cas9-based genome editing for simultaneous interference with gene expression and protein stability. *Nucleic acids research*. 2017;45(20):e171.
49. Murakami A, Nakatogawa H, Ito K. Translation arrest of SecM is essential for the basal and regulated expression of SecA. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101(33):12330-5.
50. Nakatogawa H, Ito K. The Ribosomal Exit Tunnel Functions as a Discriminating Gate. *Cell*. 2002;108(5):629-36.
51. Yosef I, Bochkareva ES, Adler J, Bibi E. Membrane Protein Biogenesis in Ffh- or FtsY-Depleted *Escherichia coli*. *PLOS ONE*. 2010;5(2):e9130.
52. Manoil C, Mekalanos JJ, Beckwith J. Alkaline phosphatase fusions: Sensors of subcellular location. *Journal of Bacteriology*. 1990;172(2):515-8.

## Appendix 1 – DNA Sequences

### Primers

The primers used in this work, both for PCR reactions and for DNA sequencing for are listed in Table 1

Table 1: The primers used for PCR reactions and for DNA sequencing, with their sequence.

Name	Sequence
pMAZ-SK Backbone Gibson F	GTTTTAGAGCTAGAAATAGCAAG
pMAZ-SK Backbone Gibson R	GTGCTCAGTATCTCTACTGATAGG
glyA gRNA check F	TTAGCTGAGTCAGGAGAT
ffh gRNA check F	TCTCTCGCCTGGGGTGGA
secM gRNA check F	TGAGTGGAATACTGACGCGC
secG gRNA check F	AATGCTTCAACCAATAAAGC
rpos gRNA (universal) R	CGACCGCGTATTTTCGTCTC
glyA genome check F	CTTCCAGTTTCGTAGCAAAG
glyA genome check R	GCTCGAGGAACACTTCTACC
ffh genome check F	GAAATCAACGCCACCTG
ffh genome check R	ACACATCTTTGCCACATCTG
secM genome check F	AAGATTTGGCTGGCGCTGGCTG
secM genome check R	ATGCCTTGCGCCTGGCTTATCC
secG genome check F	ACAGGCGATGTATGAACAGG
secG genome checkR	TTGAACTGGCGCTGAAAC
pVB1-251 Backbone Gibson F	GCGGCCGCTGATAAGCTTG
pVB1-251 Backbone Gibson R	ATGTTCATGACTCCATTATTATT
pAE001 sequencing 1 F	AAGAAGCGGATACAGGAGTG
pAE001 sequencing 2 R	TGCGGTTTACCAGGGTATCG
pAE001 sequencing 3 F	GGACCATATGAAGCAGCATGAC

## DNA Sequence of OmpA – IgG2 Fc – sfGFP Fusion Protein

The whole sequence of the designed fusion protein is shown in Figure 1.



AACCGCATTGAGCTGAAAGGCATTGACTTTAAAGAAGACGGCAATATCCTGGGCCATAAGCTGGAATACAATTTT  
 sfgfp

AACAGCCACAATGTTTACATCACCGCCGATAAACAAAAAATGGCATTAAAGCGAATTTTAAAATTCGCCACAAC  
 sfgfp

GTGGAGGATGGCAGCGTGCAGCTGGCTGATCACTACCAGCAAAACACTCCAATCGGTGATGGTCCTGTTCTGCTG  
 sfgfp

CCAGACAATCACTATCTGAGCACGCAAAGCGTTCGTCTAAAGATCCGAACGAGAAACGGGATCATATGGTTCTG  
 sfgfp

CTGGAGTTCGTAACCGCAGCGGGCATCACGCATGGTATGGATGAACTGTACAAATGATGA  
 sfgfp

Figure 1: The whole sequence of the OmpA – IgG2 Fc – sfGFP fusion protein.

## Appendix 2 – Sequencing Results

### Sequencing of SecG Mutant Strains

The sequencing results of the *secG* mutant strain used for measuring the effect on periplasmic translocation are shown in Figure 2. All mutant strains sequenced seemed to have the correct insert, with no mutations, thus only one sequencing result is shown.

```
accgcaggtagcggcgatccacttgagcatgagtcagttaaagccgtgaccgcagaggttgaagctgcgctgggcaaccgtggacgcgtgttgctgcgtaaatccggcacc
template sequence SecG
|
ACCGCAGGTAGCGGCGATCCACTTGAGCATGAGTCAGTTAAAGCCGTGACCGCAGAGGTTGAAGCTGCGCTGGGCAACC GTGACGCGTGTGCTGCGTAAATCCGGCAC
aligned sequence 31

gaaccgttaattcgcgtgatggtggaaggcgaagacgaagcgcaggtgactgaattgcacaccgcatcgccgatgcagtaaaagccgtttaagcgttagataactggct
template sequence SecG
|
GAACCGTTAATTCGCGTGATGGTGAAGGCGAAGACGAAGCGCAGGTGACTGAATTTGCACACCGCATCGCCGATGCAGTAAAAGCCGTTTAAAGCGTTAGATAACTGGCT
aligned sequence 31

aaaaagcggcgattgtgcctttttctcagcgaacacatcggttttgctgttttttccgcagttgatacaatgcgataaaattgcccttgcaaggtcattcgctttg
template sequence SecG
|
AAAAGCGGCGATTGTGCCTTTTTCTCAGCGAACACATCGGTTTTGCTGTTTTTTCCGCAGTTGATACAATGCGATAAAATTGCCCTTGCGAAGGTCATTGCTTTG
aligned sequence 31

gtagtattcacaccTGTCATACATCGCctgtatgaagct
template sequence SecG
|
GTTAGTATTACACCCTGTTTCATACATCGCCTGTAAG---
aligned sequence 31
```

Figure 2: The sequencing results of one *secG* mutant strain. The sequenced fragment is aligned to the region of *secG*. The arrow shows the beginning of the *secG* gene. And the insert is marked.

## Sequencing of *secM* Mutant Strains

The sequencing results for *secM* mutant strains are shown in Figure 3. As all three mutants had off target mutations, all three are shown. Aligned sequence 2 was used for measuring levels of translocation, and for generating the double mutants.



Figure 3: Sequencing results of the *secM* mutant strains. The gene is shown from the beginning. The insert is marked

## Sequencing of OmpA – IgG Fc – sfGFP

The correct insertion of the fusion protein into the pVB1-251 plasmid was confirmed with sequenced using three different primers: pAE001 sequencing 1 F, pAE001 sequencing 2 R, pAE001 sequencing 3 F. Together the three reads cover the whole sequence, overlapping. The results confirm that the sequence is correct. The results are shown in Figure 4.

```
TCCGTGCTGACCGTTGTCATCAGGATTGGTTGAACGAAAAGAATACAAATGTAAGGTGAGCAATAAGGGTCTGCCAGCTCCCATTGAAAAGACAATTT
template sequence pVB1-251 with GFP construct

TCCGTGCTGACCGTTGTCATCAGGATTGGTTGAACGAAAAGAATACAAATGTAAGGTGAGCAATAAGGGTCTGCCAGCTCCCATTGAAAAGACAATTT
aligned sequence 12DC15 (12DC15.ab1)

TCCGTGCTGACCGTTGTCATCAGGATTGGTTGAACGAAAAGAATACAAATGTAAGGTGAGCAATAAGGGTCTGCCAGCTCCCATTGAAAAGACAATTT
aligned sequence 12DC14 (12DC14.ab1)

-----
aligned sequence 12DC16 (12DC16.ab1)

CAAAAACCAAAGGCCAACCCACGTGAACCGCAAGTATACACTTTACCCCCTTCACGTGAAGAAATGACAAAAATCAAGTATCGCTGACGTGTTTGGTCAA
template sequence pVB1-251 with GFP construct

CAAAAACCAAAGGCCAACCCACGTGAACCGCAAGTATACACTTTACCCCCTTCACGTGAAGAAATGACAAAAATCAAGTATCGCTGACGTGTTTGGTCAA
aligned sequence 12DC15 (12DC15.ab1)

CAAAAACCAAAGGCCAACCCACGTGAACCGCAAGTATACACTTTACCCCCTTCACGTGAAGAAATGACAAAAATCAAGTATCGCTGACGTGTTTGGTCAA
aligned sequence 12DC14 (12DC14.ab1)

-----
aligned sequence 12DC16 (12DC16.ab1)

AGGCTTCTACCCTTCAGACATTGCGGTAGAGTGGAATCGAACGGTCAACCAGAAAATAACTACAAAACCACTCCCCGATGTTAGACTCTGACGGCTCG
template sequence pVB1-251 with GFP construct

AGGCTTCTACCCTTCAGACATTGCGGTAGAGTGGAATCGAACGGTCAACCAGAAAATAACTACAAAACCACTCCCCGATGTTAGACTCTGACGGCTCG
aligned sequence 12DC15 (12DC15.ab1)

AGGCTTCTACCCTTCAGACATTGCGGTAGAGTGGAATCGAACGGTCAACCAGAAAATAACTACAAAACCACTCCCCGATGTTAGACTCTGACGGCTCG
aligned sequence 12DC14 (12DC14.ab1)

-----
aligned sequence 12DC16 (12DC16.ab1)
```

ATGAAAAAGACAGCTATCGCGATTGCAGTGGCACTGGCTGGTTTCGCTACCGTAGCGCAGGCCGTGGAATGCCCGCCTTGCCCAGCTCCTCCCGTTGCTG  
template sequence pVB1-251 with GFP construct

TTGAAAAAGACAGCTATCGCGATTGCAGTGGCACTGGCTGGTTTCGCTACCGTAGCGCAGGCCGTGGAATGCCCGCCTTGCCCAGCTCCTCCCGTTGCTG  
aligned sequence 120C15 (120C15.ab1)

ATGAAAAAGACAGCTATCGCGATTGCAGTGGCACTGGCTGGTTTCGCTACCGTAGCGCAGGCCGTGGAATGCCCGCCTTGCCCAGCTCCTCCCGTTGCTG  
aligned sequence 120C14 (120C14.ab1)

-----  
aligned sequence 120C16 (120C16.ab1)

GCCCGTCGGTCTTTTTATTCCTCCCAAACCTAAAGATACACTTATGATTAGTCGTAAGTCACTGTGTAGTTGTTGACGTTTCTCATGAGGA  
template sequence pVB1-251 with GFP construct

GCCCGTCGGTCTTTTTATTCCTCCCAAACCTAAAGATACACTTATGATTAGTCGTAAGTCACTGTGTAGTTGTTGACGTTTCTCATGAGGA  
aligned sequence 120C15 (120C15.ab1)

GCCCGTCGGTCTTTTTATTCCTCCCAAACCTAAAGATACACTTATGATTAGTCGTAAGTCACTGTGTAGTTGTTGACGTTTCTCATGAGGA  
aligned sequence 120C14 (120C14.ab1)

-----  
aligned sequence 120C16 (120C16.ab1)

TCCAGAGGTTCAATTTAATTGGTATGTAGATGGAGTAGAGGTTTATAACGCCAAGACAAAGCCTCGTGAAGAAGCAGTTTAATTCGACATTTTCGCGTTGTT  
template sequence pVB1-251 with GFP construct

TCCAGAGGTTCAATTTAATTGGTATGTAGATGGAGTAGAGGTTTATAACGCCAAGACAAAGCCTCGTGAAGAAGCAGTTTAATTCGACATTTTCGCGTTGTT  
aligned sequence 120C15 (120C15.ab1)

TCCAGAGGTTCAATTTAATTGGTATGTAGATGGAGTAGAGGTTTATAACGCCAAGACAAAGCCTCGTGAAGAAGCAGTTTAATTCGACATTTTCGCGTTGTT  
aligned sequence 120C14 (120C14.ab1)

-----  
aligned sequence 120C16 (120C16.ab1)

TTCTTTTTATACTCCAAATTGACGGTGGATAAAAAGTCGCTGGCAACAAGGCAATGTCTTCTCGTGCAGCGTAATGCACGAAGCTTTCATAACCATTACA  
template sequence pVB1-251 with GFP construct

TTCTTTTTATACTCCAAATTGACGGTGGATAAAAAGTCGCTGGCAACAAGGCAATGTCTTCTCGTGCAGCGTAATGCACGAAGCTTTCATAACCATTACA  
aligned sequence 120C15 (120C15.ab1)

TTCTTTTTATACTCCAAATTGACGGTGGATAAAAAGTCGCTGGCAACAAGGCAATGTCTTCTCGTGCAGCGTAATGCACGAAGCTTTCATAACCATTACA  
aligned sequence 120C14 (120C14.ab1)

-----  
aligned sequence 120C16 (120C16.ab1)

CCCAGAAGTCCTGTCTCTTTACCCGGGAAAGGGATCCGCTGGCTCCGCTGCTGTTCTGGCGAATTCGTAAGGCGAGGAGCTGTTCACTGGTGTGCT  
template sequence pVB1-251 with GFP construct

CCCAGAAGTCCTGTCTCTTTACCCGGGAAAGGGATCCGCTGGCTCCGCTGCTGTTCTGGCGAATTCGTAAGGCGAGGAGCTGTTCACTGGTGTGCT  
aligned sequence 120C15 (120C15.ab1)

CCCAGAAGTCCTGTCTCTTTACCCGGGAAAGGGATCCGCTGGCTCCGCTGCTGTTCTGGCGAATTCGTAAGGCGAGGAGCTGTTCACTGGTGTGCT  
aligned sequence 120C14 (120C14.ab1)

-----  
aligned sequence 120C16 (120C16.ab1)

CCCTATTCTGGTGGAAGTGGATGGTGTGATGCAACGGTCATAAGTTTTCCGTGCGTGGCAGGGTGAAGGTGACGCAACTAATGGTAACTGACGCTGAAG  
template sequence pVB1-251 with GFP construct

CCCTATTCTGGTGGAAGTGGATGGTGTGATGCAACGGTCATAAGTTTTCCGTGCGTGGCAGGGTGAAGGTGACGCAACTAATGGTAACTGACGCTGAAG  
aligned sequence 120C15 (120C15.ab1)

CCCTATTCTGGTGGAAGTGGATGGTGTGATGCAACGGTCATAAGTTTTCCGTGCGTGGCAAGGTTGAAGGTGACGCAACTAATGGTAACTGACGCTGAAG  
aligned sequence 120C14 (120C14.ab1)

-----  
aligned sequence 120C16 (120C16.ab1)

