NTNU

Norwegian University of
Science and Technology

Effects of Targeted Knockout Mutations on Translational Capacities in *Escherichia coli*

Establishing Methods to Increase and
Evaluate Translational Capacities in *E. coli*
BW25113

Mats Håkon Grøn Jønland

Chemical Engineering and Biotechnology
Submission date: June 2017
Supervisor: Trygve Brautaset, IBT

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

Summary

The overall aim of this thesis was to investigate how genome engineering might be used to generate *Escherichia coli* strains with increased capacities for recombinant protein production. As translation constitute a possible bottleneck in recombinant production processes [Mahalik et. al, 2014] this work focused on evaluating and increasing *translational capacities* in *E. coli*.

Two methods were used to evaluate translational capacities in *E. coli*:

1. Two plasmid reporter systems were established to investigate levels of ribosome expression. These plasmids carried red fluorescent protein genes (*mCherry*), under control of ribosomal promoters. Levels of expression of ribosomal constituents were evaluated by measuring fluorescence from strains carrying reporter plasmids.

2. Exponential phase growth rates were used to assess translational capacities. Protein synthesis is generally the growth rate limiting factor during exponential phases, and a positive linear correlation between growth rates and ribosome concentrations have been observed in earlier work [Forchhammer & Lindahl 1971; Bremer & Dennis 1996; Scott et al 2010; Klumpp et. al, 2013].

The plasmid based reporter systems were characterized and verified as reliable reporters for levels of ribosome expression in wild type *E. coli BW25113*. However, fluorescence readings were only deemed reliable when evaluated *qualitatively*.

Three genes were knocked out (*uspA*, *dps* and *ompA*) in an effort to increase translational capacities in *E. coli BW25113*. The general rationale of this approach was that by knocking out non-essential genes, cells can more efficiently allocate resources towards growth or recombinant protein production.

A set of *single gene* knockout mutants in which the target genes (*uspA*, *dps* and *ompA*) had been replaced with kanamycin resistance cassettes were received from the KEIO collection [Baba et. al, 2006]. None of this gene *substitution* mutants were deemed to have increased translational capacities, compared to wild type *E. coli BW25113*. This was explained by the notion that these strains did not have significantly reduced proteome sizes, as genes were *replaced* not *removed*.

A set of single gene *deletion* mutants were generated by removing the kanamycin resistance cassettes from the strains from the KEIO collection. *Deletion* of the *ompA*

gene resulted in enhanced levels of ribosome expression and a 6±4% increase in cell growth rates. Hence, it was concluded that *ompA deletion* mutants had increased translational capacities compared to wild type *E. coli BW25113.* OmpA is an abundant protein in fast growing *E. coli* cells. The observed increased in translational capacities in *ompA deletion* mutants were accredited to the notion that these strains had significantly smaller proteomes than wild type cells. Dps and UspA are stress related proteins, and constitute only small parts of the *E. coli* proteome during exponential phase growth. Gene deletions of *uspA* or *dps* did not result in increased translational capacities in *E. coli BW25113*.

The results obtained in this work suggest that introduction of targeted knockout mutations is a valid strategy for increasing translational capacities in *E. coli.* Genes that are highly expressed during exponential growth phases should be targeted for *deletion* mutations for this approach to be effective.

To generate strains with significantly increased capacities for recombinant protein production it is most likely necessary to introduce multiple targeted knockout mutations. In this work, CRISPR optimized MAGE recombineering (CRMAGE) [Ronda et. al, 2016] was evaluated as a possible method for generating multiple knockout mutations in the *E. coli* genome. A ~4% mutation efficiency was achieved when utilising this method. This was very low compared to earlier work (~98%) [Ronda et. al, 2016] and further optimization is probably necessary before this method can be efficiently used to knock out genes in our labs.

Sammendrag på norsk

Målet med denne oppgaven var å undersøke hvordan endring av genomet til *Escherichia coli* kan øke vertens kapasitet for rekombinant proteinproduksjon. Translasjon er en potensiell flaskehals i produksjon av rekombinante proteiner. Dette arbeidet var fokusert på å utvikle metoder for å evaluere og øke den *translasjonelle kapasiteten* in *E. coli.*

To metoder ble brukt til å evaluere translasjonelle kapasiteter i E. coli:

1. To plasmid-baserte reportersystemer ble utviklet for å undersøke nivåer av ekspresjon av ribosomer. Disse plasmidene inneholdt et gen for et rødt fluorescerende protein, under kontroll av ribosomal promotorer. Ekspresjon av ribosomer ble evaluert ved å måle fluorescens fra stammer med disse reporterplasmidene.

2. Vekstrater i eksponentielle vekstfaser ble brukt til å evaluere translasjonelle kapasiteter. Proteinsyntese er vanligvis den begrensende faktoren i eksponentielle faser, og en positiv lineær korrelasjon mellom vekstrater og ribosomkonsentrasjoner har blitt observert i tidligere arbeid [Forchhammer & Lindahl 1971; Bremer & Dennis 1996; Scott et al 2010; Klumpp et. al, 2013].

De plasmid-baserte reportersystemene ble karakterisert og verifisert som gode reportere for ekspresjon av ribosomer i *E. coli BW25113*. Fluorescensmålinger ble kun dømt til å være pålitelige når de ble tolket *kvalitativt*.

I et forsøk på å øke den translasjonelle kapasiteten i *E. coli BW25113* ble tre gener slått ut (*uspA*, *dps* og *ompA*). Tanken bak å slå ut disse genene var at ved å slå ut ikke-essensielle gener vil verten kunne bruke ressurser mer effektivt.

Et sett med enkelt-gen utslags mutanter ble mottatt fra KEIO samlingen [Baba et. al, 2006]. I disse stammene hadde genene blitt erstattet av en kanamycin resistens kassett. Ingen av disse stammene hadde økt translasjonell kapasitet sammenlignet med villtype *E. coli BW25113*. Det var foreslått at disse stammene ikke hadde signifikant reduserte proteomstørrelser, og at de derfor ikke fikk økt translasjonelle kapasiteter.

Et nytt sett med enkelt-gen utslags mutanter ble generert ved å fjerne kanamycin resistens kassettene fra stammene fra KEIO samlingen. Utslag av *ompA* genet resulterte i økte nivåer av ekspresjon av ribosomer og en 6±4% økning i vekstrater. Det

var derfor foreslått at utslag av dette genet førte til økt translasjonell kapasitet i *E. coli BW25113*. OmpA er et høyt uttrykt protein i *E. coli* under eksponentiell vekst. Det var foreslått at utslag av *ompA* førte til økt translasjonell kapasitet fordi disse stammene fikk reduserte proteomstørrelser.

Resultatene fra dette arbeidet viser at å generere utslagninger av gener kan være en gyldig strategi for å øke translasjonelle kapasiteter i *E. coli.* Gener som er sterkt uttrykt i eksponentielle faser bør bli markert for fjerning for at denne metoden skal være effektiv.

For å lage stammer med signifikant økt kapasitet for rekombinant protein produksjon, er det mest sannsynlig nødvendig å slå ut flere gener i samme vert. CRISPR optimized MAGE recombineering (MAGE) ble evaluert som en potensiell metode for å oppnå dette. En mutasjonseffektivitet på 4% ble oppnådd. Dette var svært lavt i forhold til resultater publisert tidligere (98%) [Ronda et. al, 2016]. Dette viste at mer arbeid må bli lagt i å optimalisere denne metoden før den kan bli brukt effektiv i vår lab.

Table of Contents

1 Background

1.1 Recombinant protein production

In 1982, the first recombinant therapeutic protein was approved and launched to the commercial market [Huang et. al, 2012]. The protein, *humilin*, was a synthetic human insulin, produced in *Escherichia coli*. This drug provided a cheaper and better treatment for diabetes, compared to insulin derived from animals. Since that time, the recombinant protein industry has experienced dramatic growth, and more than 150 unique recombinant therapeutics have been approved by the Food and Drug Administration (FDA). Due to the upcoming expiration of many blockbuster drug patents, and the advent of biosimilar drugs, the recombinant therapeutics market is increasingly competitive. The worldwide recombinant protein market is expected to be worth about $300 billions by 2021[Mordorintelligence].

The central dogma of biotechnology is depicted in Figure 1.1. This dogma describes the process of synthesizing proteins based on DNA sequences (expressing genes). The word "recombinant" means that recombinant DNA technologies have be applied to make the host express the protein in question. Recombinant DNA technologies refers to methods to isolate, manipulate and combine different genes.

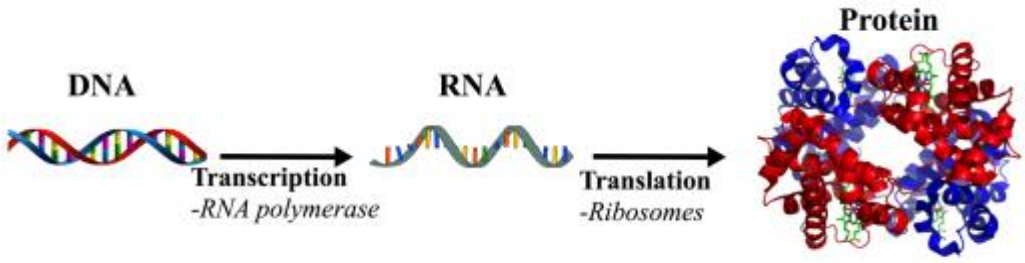


Figure 1.1: *Gene expression - the central dogma of biotechnology. DNA is transcribed to RNA, which can be translated into proteins.*

The process outlined in Figure 1.1 is a simplification of a complex and highly regulated biological process. The workflow of recombinant protein production is far from trivial, and a myriad of problems can decrease yields of functional proteins. A fast-growing demand for recombinant proteins has driven an extensive research effort to increase recombinant expression yields. This research has mainly focused on vector systems design and improvement of transcription rates. Today, effective vector systems ensure that transcription can be conducted in such high levels that translation often becomes the rate limiting step in recombinant protein production [Mahalik et. al, 2014].

The focus of this thesis is on the second step of protein synthesis – translation. The dynamics of transcription and regulation thereof will not be discussed in this work, nor will the mechanisms of protein folding and post translational modifications.

Translation is the process of synthesizing polypeptides with specific amino acid sequences, based on RNA templates. The process is catalysed by large, complex, enzymes called *ribosomes*. Ribosomes join amino acids to a growing polypeptide chain in a highly specific manner, based on the codons in a RNA template. Three different RNA species are involved in translation:

- **Messenger RNAs (mRNAs)** – Contain codons corresponding to specific amino acids. Provides the template for the amino acid sequences in proteins.

- **Transport RNAs (tRNAs)** – Act as adaptors between amino acids and ribosomes. Responsible for supplying ribosomes with amino acids, and presenting them in steric configurations that facilitate peptide bond formation.

- **Ribosomal RNAs (rRNAs)** – Constitute structural, and catalytic, parts of ribosomes.

Ribosomes and RNAs constitute the machinery for protein synthesis. By upscaling this machinery, or increasing its effectiveness, it may be possible to increase hosts' capacities towards recombinant protein production. The discussion provided in the following sections aim to uncover how one might utilise host genome engineering approaches to increase translational capacities.

Although *E. coli* remains the most widely used host organism, a wide range of prokaryotic and eukaryotic cell lines is used for commercial recombinant protein production. Every host organism brings its own advantages and challenges, but *E. coli* generally provides the easiest and cheapest platform [Demain & Vaishnav, 2009]. All further discussion will focus on expression in *E. coli.*

1.2 Translation – a possible rate limiting step in recombinant protein production

In nature, transcription is generally the rate limiting step in protein expression [Nelson & Cox, 2013]. Regulation of translation offers fine tuning towards environmental changes. However, when expressing a recombinant gene, rates of transcription often supersede the cell's translation capacity, resulting in a translational bottleneck [Mahalik et. al, 2014]. When translation is the rate limiting step in a recombinant protein production process, high rates of transcription can be offset by high levels of mRNA degradation. Hence, efforts should be made to increase rates of translation, to increase recombinant protein yields. To do this, it is important to identify factors that might affect rates of translation.

1.2.1 Ribosome concentrations is generally the main host specific determinant for translational capacities

To effectively synthesize proteins, ribosomes need to be provided with sufficient amounts of amino acids and energy. For translation of a gene to be initiated, ribosomes must be able to locate and bind to the gene's mRNA transcript. Translation is a complex process, and a lot of factors can potentially influence reaction rates. Some of the most likely rate decreasing factors in translation of recombinant genes are listed below [Mahalik et. al, 2014]:

1. Low level of ribosome binding to the recombinant mRNAs
2. Shortage of amino acids or transfer RNAs (tRNAs)
3. High global levels of RNA degradation
4. Low number of actively translating ribosomes

None of these factors should be viewed as a universal rate limiting factor. What the most influential rate decreasing factor(s) is, might vary between individual recombinant production systems.

Factor 1 and 2 in the aforementioned list are largely gene specific. Problems related to these factors can often be alleviated by engineering the sequence of the recombinant gene. How efficiently a given mRNA is translated is largely dependent on the sequence of the Shine-Dalgarno (SD) site, upstream of the initiation codon. e.g. the SD-sequence UAAGGAGG initiate translation roughly four times more efficiently than the shorter AAGGA sequence, in *E. coli* [Ringquist et. al, 1992]. The secondary structure of mRNA can greatly affect production of recombinant proteins [Ivanovski et. al, 2002].

Ivanovski and colleagues showed that a hairpin structure in a mRNA template greatly decreased the level of translation of the gene.

Limitations related to shortage of amino acids or tRNAs can be alleviated by optimizing the gene sequence to fit with the relevant host´s codon bias [Rosano & Ceccarelli, 2014]. Alternatively, specific amino acids can be supplemented to the growth medium.

Some translational limitations may be alleviated by engineering the sequence in the recombinant gene. However, overcoming other restrictions might require engineering of the host cell itself. The global level of RNA degradation and intracellular ribosome concentrations are examples of factors that are host specific. Earlier work has isolated ribosome concentrations as the main determinant for translational capacities in *E. coli* [Scott et. al, 2014; Bosdriesz et. al, 2015]. However, more ribosomes do not directly equate to increased translational capacities, as each ribosome need to be supplied with enough energy and amino acids to synthesise proteins at optimal rates. Hence, synthesis and degradation of ribosomes are strictly regulated. Sections 1.2.2 - 1.2.4 cover the dynamics of RNA degradation and synthesis in *E. coli*.

1.2.2 Ribosome synthesis

Prokaryotic ribosomes contain two major subunits: the 50S, - and 30S subunit [Snustad, 2012]. “S” is an abbreviation of “Svedbergs”, which is a size measurement, based on rates of sedimentation. 31 ribosomal proteins and 2 rRNA species (5S rRNA and 23S rRNA) constitute the 50S ribosomal subunit. The 30S subunit contains 21 ribosomal proteins and one rRNA specie (16S rRNA). During initiation of translation, the 30S and 50S subunits assemble around a mRNA initiation codon (usually ATG), to form the 70S ribosome. When translation is terminated, the 70S ribosome is dissociated and the 30S and 50S subunits are recycled.

The energy demanding process of ribosome synthesis is tightly regulated. During exponential growth, ribosomes make up about a quarter of cellular dry weight in *E. coli* [Nelson & Cox, 2013]. The dynamics of ribosome synthesis and degradation are heavily correlated to growth rates [Scott et. al, 2014]. In general, concentrations of ribosomes increase linearly with growth rates, as more ribosomes are needed to maintain higher rates of protein synthesis.

1.2.2.1 Ribosome synthesis is regulated by controlling transcription of rRNA genes

Transcription of rRNA genes are generally the rate limiting step in ribosome synthesis [Maeda et. al, 2015]. At high growth rates, rRNA promoters account for more than half of the global transcription in *E. coli* [Bremer & Dennis, 1987]. 7 different rRNA encoding operons (rrn operons) are distributed throughout the *E. coli* genome. Slight heterogeneity has been observed between these operons, but the general organization is conserved. All rrn operons carry genes for all 3 rRNA species necessary to synthesize ribosomes. An illustration of the general rrn operon organization is depicted in Figure 1.2.

Figure 1.2: *General organization of rrn operons in E. coli [Maeda et. al, 2015]. P1 and P2 are promoter regions. Red arrows annotate rRNA encoding regions and the green arrow annotate a tRNA encoding spacer region. The two black blocks to the right symbolize terminator regions.*

Two promoter regions control transcription of rrn operons. The P1 promoter regulates high level of transcription during exponential growth. The P2 promoter regulates low level, basal, expression, and is largely inactive while P1 is active [Zhang & Bremer, 1996]. Transcription from the P1 promoter is regulated by both *trans* acting transcription factors, and *cis* acting elements. A scheme of a rrn operon P1 promoter is depicted in Figure 1.3.

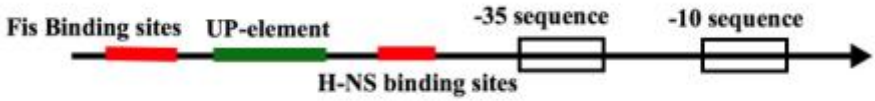


Figure 1.3: *rrn P1 promoter organization [Maeda et. al, 2015]. Transcription factor binding sites are annotated with red markers. The cis acting UP-element is highlighted with a green marker. The number of FIS and H-NS binding sites varies between the different rrn operons. All rrn operons contain some FIS binding sites, but not all contain H-NS binding sites.*

The UP-element (Figure 1.3) is an AT-rich region upstream of the -35 box. The alpha subunit of *RNA polymerase* (RNAP) interacts with this region, increasing transcription from the P1 promoter. Each of the seven rrnP1 promoters in *E. coli* contains 3-8 Fis binding sites. Fis is a global transcription regulator, that is present in large intracellular concentrations during exponential growth phases [Hirvonen et. al, 2001]. Binding of

Fis increases rates of transcription from all rrn P1 promoters. Due to slight heterogeneity in sequence and number of Fis binding sites, the Fis-mediated inducibility is different for all seven rrnP1 promoters in *E. coli* [Hirvonen et. al, 2001; Maeda et. al, 2015].

H-NS is a global silencer that inhibits transcription from most rrnP1 promoters [Maeda et. al, 2015]. Not all rrnP1 promoters have H-NS binding sites, and the level of H-NS mediated repression is different for all seven rrnP1 promoters in *E. coli.*

Transcription from all rrnP1 promoters is downregulated by the cellular alarmone molecule (p)ppGpp [Lemke et. al, 2011]. (p)ppGpp is produced in various stress responses [Carneiro et. al, 2011]. This molecule inhibits transcription from rrnP1 promoters by altering RNA polymerase's (RNAP) affinity, and by downregulating *Fis* expression. When an uncharged tRNA (tRNA without amino acid) is loaded into a translating ribosome, ppGpp synthesis is induced [Bosdriesz et. al, 2015]. This mechanism synchronizes cellular concentration of ribosomes and the availability of amino acids, making sure that each ribosome can be used as efficiently as possible. A schematic representation of regulation of ribosome synthesis is depicted in Figure 1.4.

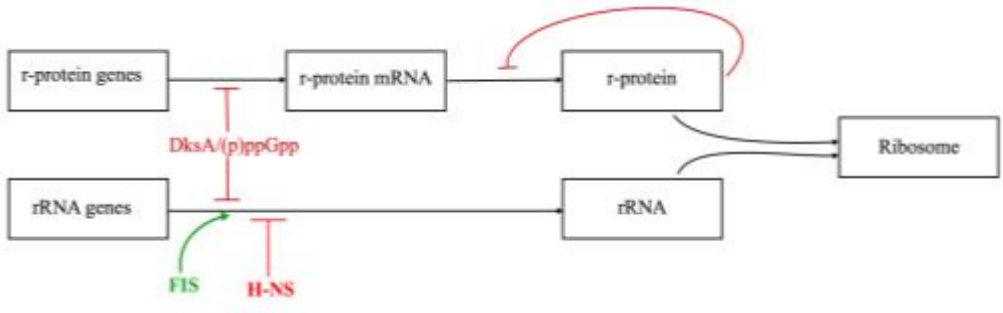


Figure 1.4: *Schematic representation of regulation of ribosome synthesis in E. coli [Lemke et. al, 2011; Maeda et al, 2015]. Green arrows annotate factors that induce expression. Red arrows annotate factors that inhibit expression. The red arrow from the "r-protein" box symbolize a translational feedback inhibition (Figure 1.5).*

1.2.2.2 Synthesis of ribosomal proteins is synchronized to rRNA concentrations through translational feedback inhibition

The expression of ribosomal proteins (rProteins) is regulated by a translational feedback inhibition mechanism [Lemke et. al, 2011]. There more than 50 rProtein genes spread out through the *E. coli* genome. All of these genes are organized in operons with at least one other rProtein gene. The translational feedback inhibition is driven by some rProteins' binding affinity towards mRNA transcripts from their own operon. Free rProteins can bind to their own transcript, inhibiting translation of parts of, - or their entire operon. However, rProteins have higher binding affinity towards rRNAs, compared to their own transcripts. Translational feedback inhibition will only occur at considerable levels when all available rRNAs are saturated with rProtein. This mechanism ensures that production of rProteins is synchronized to rRNA transcription. A schematic overview of the rRNA dependent translational feedback inhibition is depicted in Figure 1.5.

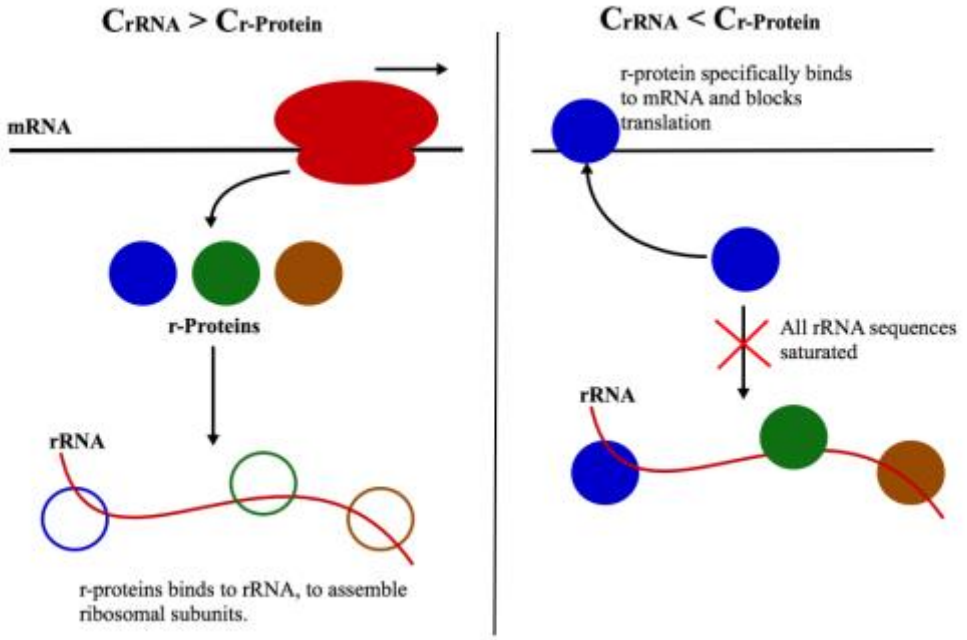


Figure 1.5: *Schematic representation of rRNA dependent feedback inhibition of ribosomal proteins (rProteins) synthesis [Lemke et. al, 2011]."C" refers to intracellular concentrations. rProteins bind to rRNA with higher affinity than its own mRNA, and inhibits translation only when all rRNA species are saturated.*

1.2.3 Ribosome degradation

Ribosomes represents about a quarter of the total cellular dry weight in *E. coli* during exponential growth [Nelson & Cox, 2013]. The constituents of ribosomes can be an important storehouse for energy and nutrients during times of stress and starvation. Hence, degradation of ribosomes is generally upregulated during stress responses and periods of decreased growth [Deutscher, 2006; Piir et. al, 2011]. rRNAs constitute the structural backbones of ribosomes, and ribosome degradation is heavily dependent on the global cellular level of RNA digestion.

1.2.3.1 RNA digestion affects rates of translation by decreasing concentrations of mRNA transcripts, and by initiating ribosome degradation.

RNA is degraded by a group of enzymes called *Ribonucleases (RNase)*. *RNases* are not strictly sequence specific, like DNA restriction enzymes, but are known to recognize certain patterns and motifs in RNA sequences [Jain, 2002]. The *in vivo* stabilities of different RNAs are highly varying. rRNAs and tRNAs have relatively long half-lifes, and are referred to as *stable RNAs* [Mahalik et. al, 2014]. mRNAs have significantly lower stability, compared to stable RNAs [Deutscher, 2006]. The stabilities of different mRNAs typically range between 10min and 10hours.

The stability of a given RNA is determined by its susceptibility for attack from *RNases*. rRNAs are generally buried in ribosomes, and are unavailable for *RNase* attack. mRNAs are typically not organized in larger molecules, and are, in that regard, open to *RNase* degradation. However, being translated is a way for mRNAs to avoid degradation. Ribosomes can "hide" parts of the sequence from *RNase* as they move along the mRNA molecule [Deutscher, 2006]. Highly translated transcripts are significantly more stable, compared to mRNAs with low levels of translation.

The most important determinant for the global level of RNA degradation is the intracellular concentration of *RNases* [Deutscher, 2006]. The building blocks of RNA can provide essential nutrients in times of stress or starvation, and the intracellular level of *RNase* generally increases during stress responses, or entry into stationary phase [Chen & Deutscher, 2010; Piir et. al, 2011]. In general, the level of RNA degradation is much higher in phases with slow (or no) growth, compared to exponential growth phases [Cairrão et. al, 2001; Chen & Deutscher, 2010].

The level of ribosome degradation in *E. coli* is largely dependent on the intracellular concentration and activity of different *RNases* [Zundel et. al, 2009]. rRNAs provide the main structural element in ribosomes, and digestion of rRNA is the rate determining step of ribosome degradation. *RNases* initiate degradation of ribosomes by degrading rRNA. Moreover, a decreased mRNA population makes ribosomal subunits more susceptible for *RNase* attack.

During retarded growth, intracellular levels of *RNase* are upregulated [Deutscher, 2006]. In turn, the global concentration of mRNAs drops. If decreased sufficiently, low mRNA levels will inhibit 70S ribosome formation, as there will be a shortage of transcripts for ribosomal subunits to join around. Hence, a decrease in mRNA levels leads to an increased concentration of unassembled ribosomal subunits [Zundel et. al, 2009]. The majority of exposed rRNA is found at the surface of ribosomal subunits [Schuwirth et. al, 2005; Steitz, 2008]. Hence, the 50S and 30S subunits are much less stable towards *RNase* degradation, compared to the assembled 70S ribosome. *RNases* catalyse ribosome degradation both by indirectly increasing the level of unassembled subunits, and by initiating degradation of the subunits. A scheme of this mechanism is provided in Figure 1.6.

The notion that levels of RNA degradation and ribosome degradation is tightly coupled was confirmed by Piir and colleagues [Piir et. al, 2011]. They discovered that intracellular concentrations of both RNA and ribosomes decrease dramatically as *E. coli* enters stationary phase. Some key results from Piir and colleagues' study are presented in Figure 1.7 and 1.8.

Based on published data, it is reasonable to expect that by repressing genes that are upregulated during stress responses, one can decrease the global level of RNA degradation in *E. coli* [Piir et. al, 2011; Mahalik et. al, 2014]. By decreasing levels of RNA digestion, it is possible that one can generate strains with increased translational capacities.

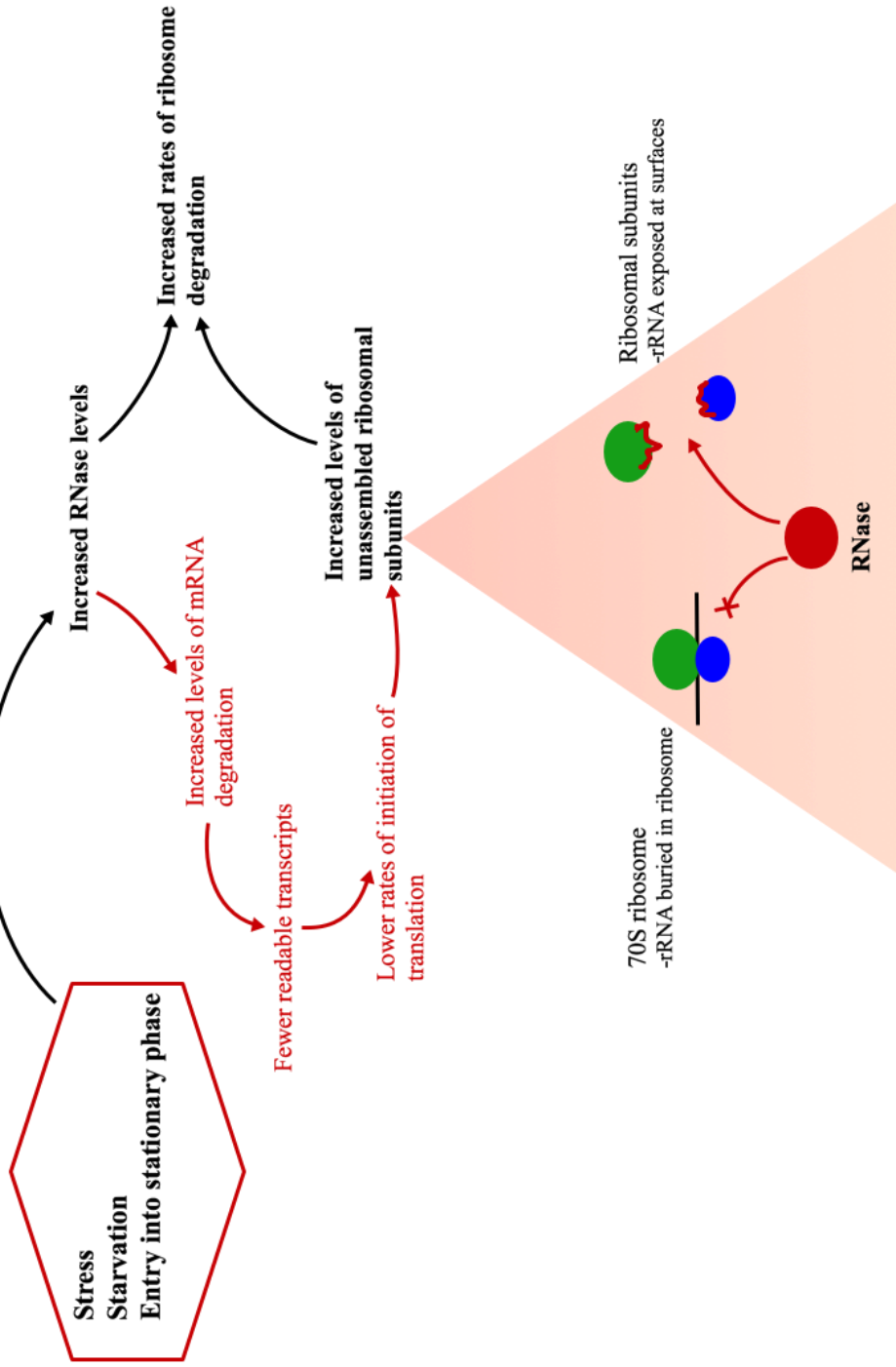


Figure 1.6: *Mechanism of RNase catalysed degradation of ribosomes [Zundel et al, 2009; Steitz, 2008; Schwurtz et. al, 2005]. RNase catalyse ribosome degradation indirectly by increasing levels of unassembled ribosomal subunits, and directly by degrading exposed rRNA.*

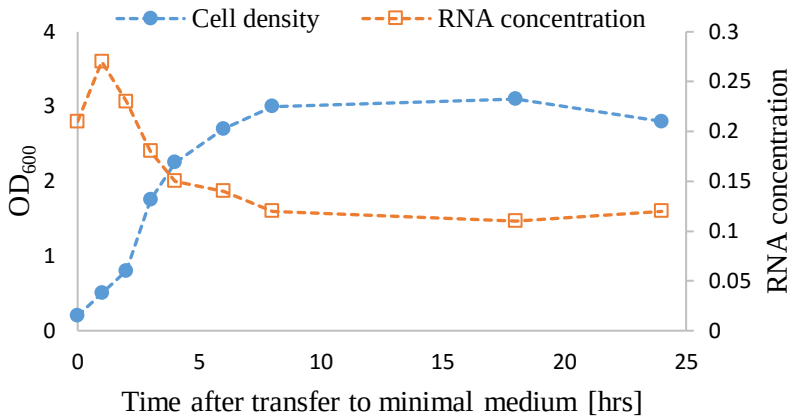


Figure 1.7: *Dynamics of RNA concentration in growing E. coli cells - Obtained from [Piir et. al, 2011]. Cells were grown until OD_{600} reached 0,2, before being transferred to glucose containing MOPS medium. RNA concentrations were calculated as the ratio of absorbance of DNaseI treated cell lysates at 260nm, to the optical density of the culture at 460nm.*

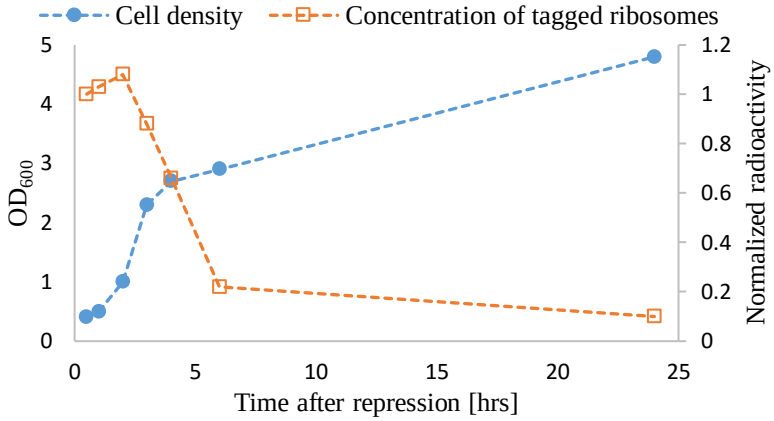


Figure 1.8: *Degradation of tagged ribosomes during batch cultivation of E. coli – Obtained from [Piir et. al, 2011]. A rrn operon was cloned into a plasmid vector and fused with a radioactive marker. The rrn operon carrying vector was transformed into E. coli cells, so that they would express radioactively labelled ribosomes. Transcription of rRNA from the plasmid constructs was repressed by glucose. By transferring growing cells to glucose containing MOPS medium, Piir and colleagues monitored rates of degradation of radioactively labelled ribosomes.*

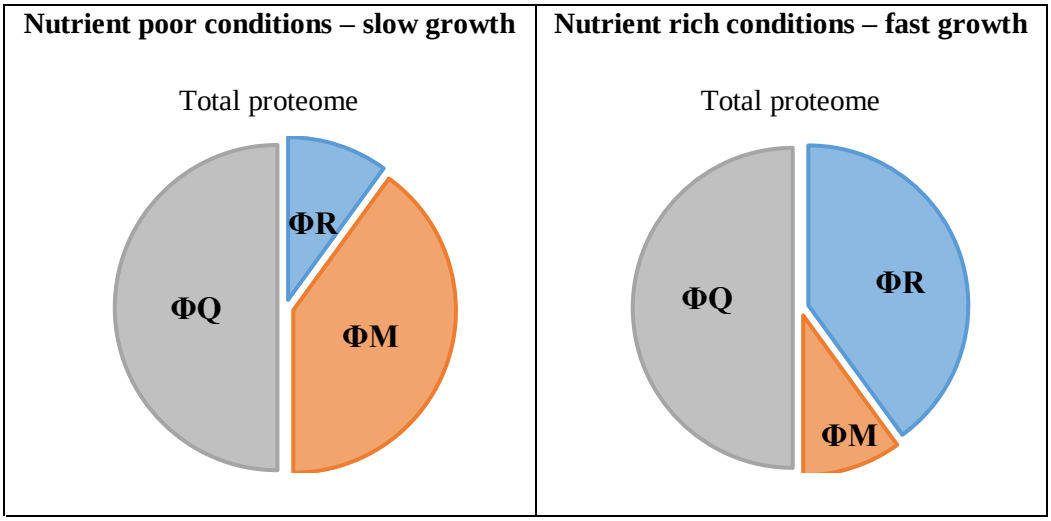
The results obtained from Piir and colleagues (Figure 1.7 and 1.8) [Piir et. al, 2011] clearly show that both intracellular RNA, - and ribosome concentrations decrease dramatically during the transition from exponential phase to stationary phase. However, the remaining ribosomes seemed to be quite stable once cells had entered the stationary phase. Piir and colleagues reported that radioactively labelled ribosomes remained stable for up to 48hrs after cells had reached stationary phase.

1.2.4 Ribosome concentrations are optimized to maximise cell growth rates

The underlying regulatory networks, that influence ribosome levels in *E. coli*, are complex and not yet fully understood [Hui et. al, 2015]. However, several simple models have predicted the dynamics of ribosome regulation, with high correlations to empirical data [Scott et. al, 2014; Bosdriesz et. al, 2015; Hui et. al, 2015]. These models use coarse-grained proteome partitioning. An example of one such proteome partitioning is depicted in Figure 1.9.

Ribosome synthesis and degradation are heavily correlated with growth rate. It has been postulated that ribosome concentrations in *E. coli* are continuously regulated to maximise growth rates in different conditions [Scott et. al, 2014; Bosdriesz et. al, 2015]. Maximization of growth rates is an important fitness strategy for bacteria [Bosdriesz et. al, 2015]. In environments that support growth, the growth rate of a bacterium equates directly to fitness because it allows it to outcompete other bacteria. During irregular feast-famine fluctuations, growth rate maximization remains important because it gives rise to more offspring and spread of genetic diversity. The growth rate of a cell is determined by the rate at which it is able to duplicate its proteome [Molenaar et. al, 2009; Scott et. al, 2014]. This, in turn, depends on the cellular ribosome concentration, and how efficiently each ribosome is used.

For ribosomes to be used efficiently, they need to be supplied with enough amino acids. Because the size of a bacterium's proteome hardly changes across different conditions, synthesizing more of certain proteins goes at the expense of others [Bremer & Dennis, 1996; Scott et. al, 2014]. Synthesis of more ribosomal proteins reduces the concentration of metabolic proteins, which produce constituents for protein synthesis. This indicates that an optimal ribosome concentration exists, matching the supply of amino acids to the demand for protein synthesis. In nutrient rich environments, metabolic enzymes are used more efficiently because each enzyme is more saturated with substrate. Hence, a smaller pool of metabolic proteins is necessary to provide amino acids for protein synthesis. In such conditions, the pool of ribosomal proteins can be upregulated. This will increase rates of protein synthesis, which in turn will increase growth rates. This logic is highlighted in Figure 1.9.



ΦR – Fraction of ribosomal proteins

ΦM – Fraction of metabolic proteins

ΦQ – Fixed part of proteome

Figure 1.9: *Coarse grained proteome partitioning, used to model dynamics of growth and cellular ribosome concentration in E. coli [Scott et. al, 2014; Bosdriesz et. al, 2015; Hui et. al, 2015]. Fraction of ribosomal proteins refers to proteins that are part of ribosomes, and initiation factors and other proteins that are essential for translation. Fraction of metabolic proteins refers to proteins that are active in substrate metabolism and amino acid synthesis. The fixed part of proteome partition (ΦQ) includes proteins that are neither directly catabolic or involved in translation.*

1.2.4.1 Proteome partitioning models reveal possible strategies to increase translational capacities in bacteria

In models using proteome partitioning (Figure 1.9), ΦR and ΦM (fractions of ribosomes and metabolic proteins) dictates the cell's translational capacity. The fixed fraction of the proteome (ΦQ) includes proteins without any direct catabolic or anabolic function (structural proteins, receptors, flagellar proteins etc.). ΦQ needs to be duplicated before a cell division event. Hence, altering the size of ΦQ is predicted to affect the cell's generation time.

Models like this provide an explanation towards why generation times generally decline when bacteria produce large amounts of heterologous proteins. Production of a recombinant protein will add to the size of ΦQ, on the expense of ΦR and ΦM. Hence, recombinant protein production is predicted to decrease translation capacities, leading to decreased growth rates.

Reducing the size of ΦQ is predicted to increase the translation capacity in bacteria. This might be of interest towards recombinant protein production, where the host´s translational capacity can limit yields of functional proteins [Mahalik et. al, 2014]. ΦQ can be reduced by knocking out genes that constitutes a large part of the proteome. Alternatively, one can aim to limit the various stress responses that can occur during recombinant protein expression. Increased synthesis of stress response related proteins will increase the size of ΦQ, and subsequently decrease translational capacity. Moreover, stress responses trigger ribosome degradation which might affect the proportion between ribosomes and metabolic proteins (ΦR and ΦM). Stress responses that drive the proportion between ΦR and ΦM towards sub-optimal levels will probably have detrimental effects on the cell's translational capacity. Limiting stress responses might serve to maintain recombinant protein yields at higher levels.

Strategies to increase translational capacity in *E. coli* are further discussed in section 1.3.

1.3 Genome engineering strategies for increased translational capacity in *E. coli*

Two possible strategies to increase rates of translation in *E. coli* were evaluated in this work. Following the discussion in section 1.2.4.1, both methods aimed to increase translational capacities by decreasing the fixed portions of cells' proteomes (ΦQ). The two strategies considered in this work were:

1. Limiting the host cell's stress response mechanisms. By knocking out genes involved in the complex stress response regulatory networks, it is possible that one can limit stress triggered increases in ΦQ, and avoid unnecessary degradation of ribosomes.

2. Knockout of non-essential genes with large translational burdens. By knocking out genes that are not essential for survival, one can decrease ΦQ and, hence, increase translational capacities.

The rationale behind both strategies is discussed further in the following two sections (1.3.1 and 1.3.2).

Both strategies were executed by generating knockout mutations in the *E. coli* chromosome. There are several possible approaches to knock out a gene, and three different methods were evaluated in this work (Section 1.4). To limit stress responses, it is possibly sufficient to knock out stress response related genes in such a way that the genes yield non-functional protein products. Alternatively, target genes can be replaced by e.g. resistance markers. To create a leaner proteome (decrease ΦQ), translation of the targeted genes must be suppressed. Expression of non-functional proteins will still require energy and amino acids. To create a leaner proteome, knockouts should be generated in such a way that the knocked-out gene is not expressed at all. The differences, between the two strategies investigated in this work, are highlighted in Figure 1.10.

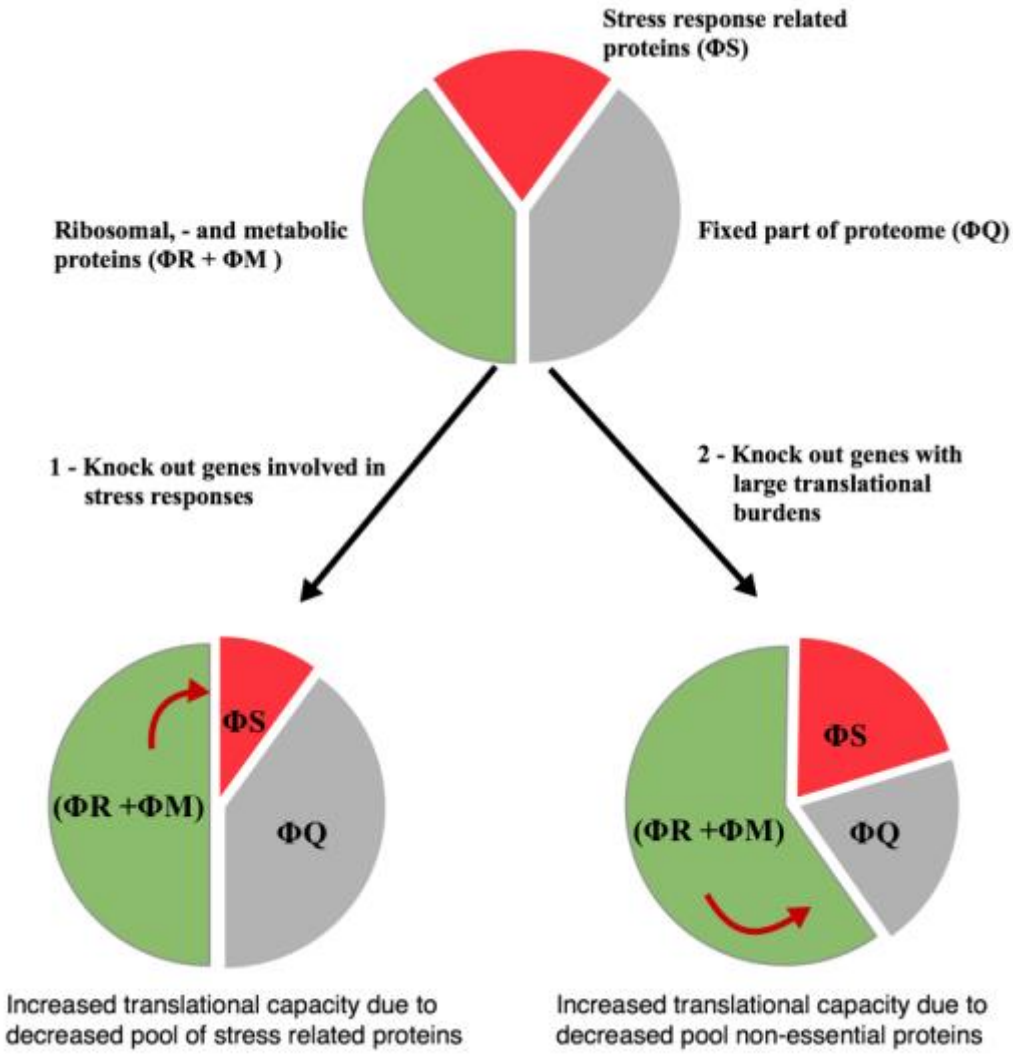


Figure 1.10: *Differences between the two strategies investigated in this work. Both strategies aim to increase the translational capacity in E. coli. However, the two strategies might require different types of knockout mutations. Due to the complexity of stress response regulatory networks, a loss-of-function mutation in stress response genes might cascade to affect the expression of other stress response genes. To decrease the size of the pool of non-essential proteins, knock out mutations that stops the expression of genes are necessary*.

1.3.1 Knock out genes that might cause unwarranted, detrimental stress responses

Nature has not optimized *E. coli*, or any other organism, as producers of recombinant proteins. High levels of recombinant protein expression generally induce a large range of stress responses in the host cell [Hoffman & Rinas, 2004]. These stress responses retard growth and product formation, and can mediate degradation of the recombinant proteins. Conditions in industrial scale bioreactors are rarely as homogenous as in lab-scale experiments. Oxygen and substrate gradients can induce stress responses, potentially yielding a large range of scale-up problems.

Possible stress response triggers in an industrial recombinant protein production include:

- Substrate starvation (C, N, P…)
- Low level of aminoacylated tRNAs (Amino acid shortage)
- Oxygen, - or substrate gradients in the bioreactor, causing starvation or osmotic stress
- Temperature gradients in the bioreactor
- Accumulation of misfolded proteins
- Accumulation of toxic by-products

In nature, stress responses are vital for survival, as conditions can be highly fluctuating. In the environment of a bioreactor, it is possible that some stress responses are unwarranted, as conditions generally can be optimised for growth. By eliminating genes that are upregulated during certain stress responses, it is possible to prevent cells from reacting prematurely to certain triggers. In that way, high levels of product formation can be maintained for longer durations. Earlier work has shown that *E. coli* cells lacking the cellular alarmone molecule (p)ppGpp produced more of a recombinant *chloramphenicol acetyltransferase*, compared to cells with (p)ppGpp [Dedhia et. al, 1997]. In another study, it was concluded that *E. coli* strains lacking in expression of the general stress response sigma factor (rpoS) produced more human *lectine*, compared to strains expressing rpoS [Jeong et al, 2004].

It is important to notice that some stress responses might be vital for cells' survival in conventional bioreactor conditions. Mutants lacking combinations of stress response genes might become sickly, and require high levels of process control to stay alive. Different stress responses are linked through complex cascades, and we do not yet fully understand the complete stress response regulatory network in *E. coli* [Wick & Egli, 2004]. Hence, stress responses are hard to model, and knockout of stress response related genes might yield unexpected phenotypical effects.

1.3.2 Knock out non-essential genes with large translational burdens

In exponential growth phases, and certainly when transcribing recombinant genes, there are a large excess of mRNA for ribosomes to translate [Mahalik et. al, 2014]. Each native mRNA transcript is, in theory, competing with recombinant mRNAs to be translated. Some native genes are probably unnecessary for survival in a bioreactor environment. Eliminating non-essential genes might increase the host's translational capacity, by decreasing the fixed part of the proteome (ΦQ – Figure 1.10). Expression of non-essential genes lead to a sub-efficient use of energy, that could otherwise support recombinant protein expression.

Price and colleagues found that 22% of all protein production in *E. coli K-12* was non-essential, during cultivation in a minimal glucose medium [Price et. al, 2016]. It was proposed that the unnecessary proteins were on "standby" in case conditions change. As conditions can be tightly regulated in the bioreactor environment, typical "standby" proteins could be possible targets for knockout mutations. Price and colleagues estimated that the burden of preparing for other conditions could reduce *E. coli's* growth rate by more than 13%.

The effect of recombinant expression on maximum specific growth rates in *E. coli* has been investigated [Scott et. al, 2010; Bienick et. al, 2014]. Both studies show, that increasing recombinant protein expression has a linear negative effect on the cells specific growth rates. In a review, Valgepea and colleagues concluded that the data from the aforementioned studies suggest that increasing expression of a recombinant protein by 1% yields a ~3% decrease in specific growth rate [Valgepea et. al, 2015]. Based on these results, it might be sensible to assume that by decreasing a cells proteome by ~1% the cells specific growth rate would increase by ~3%. This notion has been strengthened by two other studies [Fischer & Sauer, 2005; Muntel et. al, 2014]. Both showed that the specific growth rate in *Bacillus subtilis* was increased by ~30% when the total proteome was reduced by ~9%. The proteome was reduced by eliminating the flagellar regulator gene *sigD*. In another recent experiment, it was observed that deleting genes with higher protein expression cost led to a greater growth advantage in *E. coli* cells [D'Souza et. al, 2014]. Combined, these results suggest that host proteome optimization can be considered a powerful tool to engineer superior host strains for recombinant protein production. These results are readily explained by the proteome partitioning approach discussed in section 1.2.4.1. Decreasing the fixed part of the proteome frees up translational capacity, that can be utilised to increase growth rates or recombinant protein yields.

With this strategy, it is important that the induced knockout mutations silences expression of the gene. Knockout mutations that only generate a non-functional protein product, will not decrease the fixed part of the host's proteome. A multitude of different approaches might be used to induce knockout mutations in *E. coli*. Some methods for genome engineering in bacteria are discussed in section 1.4.

1.4 Methods for genome engineering in bacteria

A large range of methods for genome engineering in bacteria have been described in the literature. In this section, some techniques for generating targeted knockout mutations in bacterial chromosomes are discussed.

1.4.1 Traditional lambda Red recombination

In 2000, Datsenko and Wanner proposed a method for precise genome engineering in *E. coli*, using the homologous recombination system of bacteriophage λ [Datsenko & Wanner, 2000]. The recombination system, called "Red", consists primarily of three proteins [Poteete, 2001]:

- λ exoribonuclease, which processively digests the 5´ end of double stranded DNA
- β protein, that stabilizes single stranded DNA and promotes strand annealing
- γ protein, that binds to the cell's native DNA repair enzyme, RecBCD, and inhibits its activities.

Expression of these proteins induce a 'hyper recombineering' state in *E. coli*, in which recombination events between DNA species occur at high frequencies.

By transiently expressing the proteins in the λ Red system, Datsenko and Wanner inserted specific oligonucleotides (oligos) in the *E. coli* chromosome. By using oligos flanked by homology sequences to a target gene, they were able to generate knock out mutations with a very high precision. The method is further explained in Figure 1.11. To be able to select successful mutants, an antibiotic resistance marker is usually inserted in the place of the knocked-out gene.

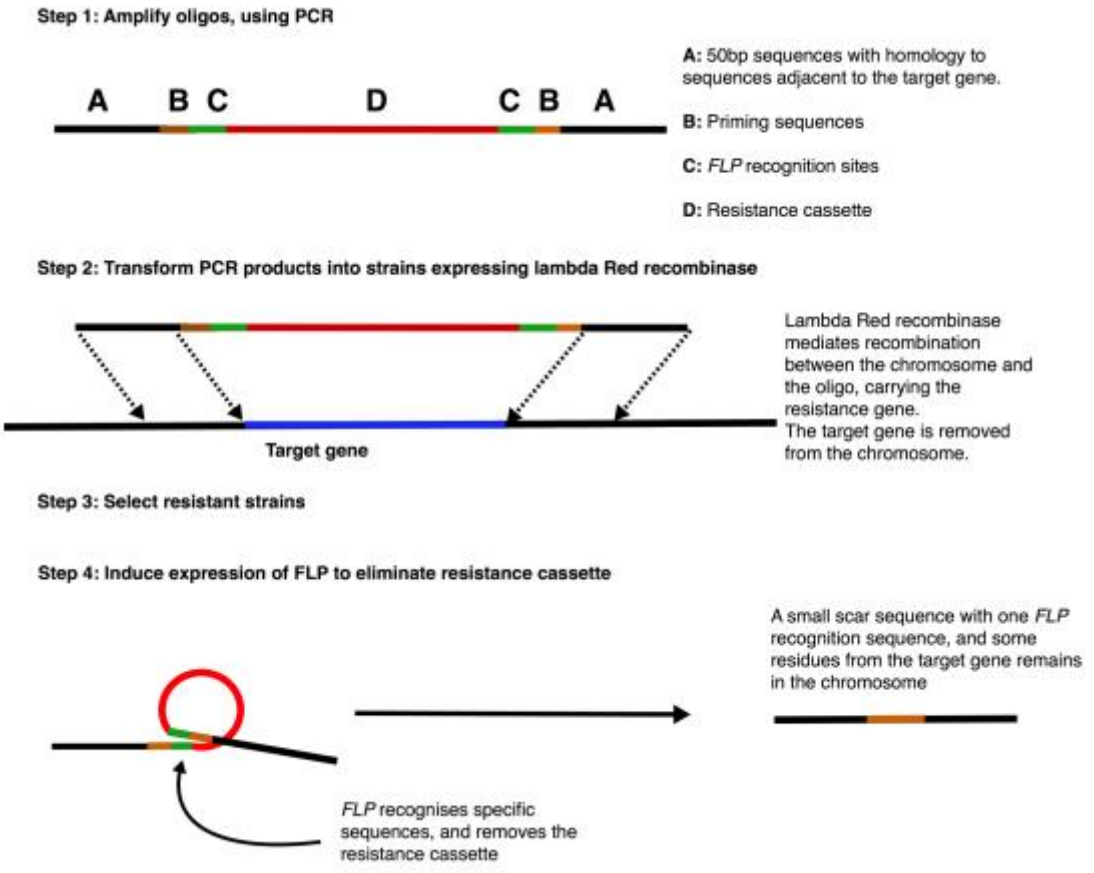


Figure 1.11: *Inactivation of chromosomal genes using oligos and λ Red recombination [Datsenko & Wanner, 2000]. Chromosomal genes are targeted by using PCR products containing a resistance cassette, flanked by FLP recombinase (*FLP*) recognition sites and homology regions to chromosomal sequences adjacent to the target gene. FLP mediated removal of resistance cassette yields a scar in the chromosome.*

By using the method depicted in Figure 1.11, a collection of nearly 4000 single gene knock out mutants of *E. coli BW25113* has been made [Baba et. al, 2006]. The library is called "the KEIO collection", and is believed to comprise single-gene deletions of all non-essential genes in *E. coli BW25113*. The last step in Figure 1.11 (Using FLP to remove resistance cassette) was not done when making the KEIO collection. All KEIO mutants carry a kanamycin resistance gene, in the place of the single gene knock out.

1.4.2 Multiplex Automated Genome Engineering (MAGE)

MAGE is a high throughput method for genome engineering. The method uses λ Red recombineering to simultaneously include multiple single stranded DNA (ssDNA) oligonucleotides (oligos) into bacterial genomes. By engineering specific oligos, MAGE can introduce new genes or targeted knockouts in bacteria.

During the MAGE procedure, mutagenic ssDNA oligos are transformed into cells. The cells are subsequently inoculated for about 2 hours. During cell division, ssDNA oligos anneal to specific sequences in the lagging strand of the replication fork, inducing mutations. (Figure 1.12) [Gallagher et. al, 2014]. During the procedure, cells are induced to express β-proteins. β-proteins protects ssDNAs from the cells native DNA degrading machinery.

One "MAGE cycle" include three main steps: cell growth, induction of β-protein expression, and transformation. As only a portion of the cells will carry the desired mutations after one cycle (Figure 1.13), it is necessary to run several subsequent MAGE cycles to introduce mutations in an appreciable number of cells. The general workflow in each MAGE cycle is summarized in Figure 1.12. The mechanism of ssDNA-genome recombineering is explained in Figure 1.13.

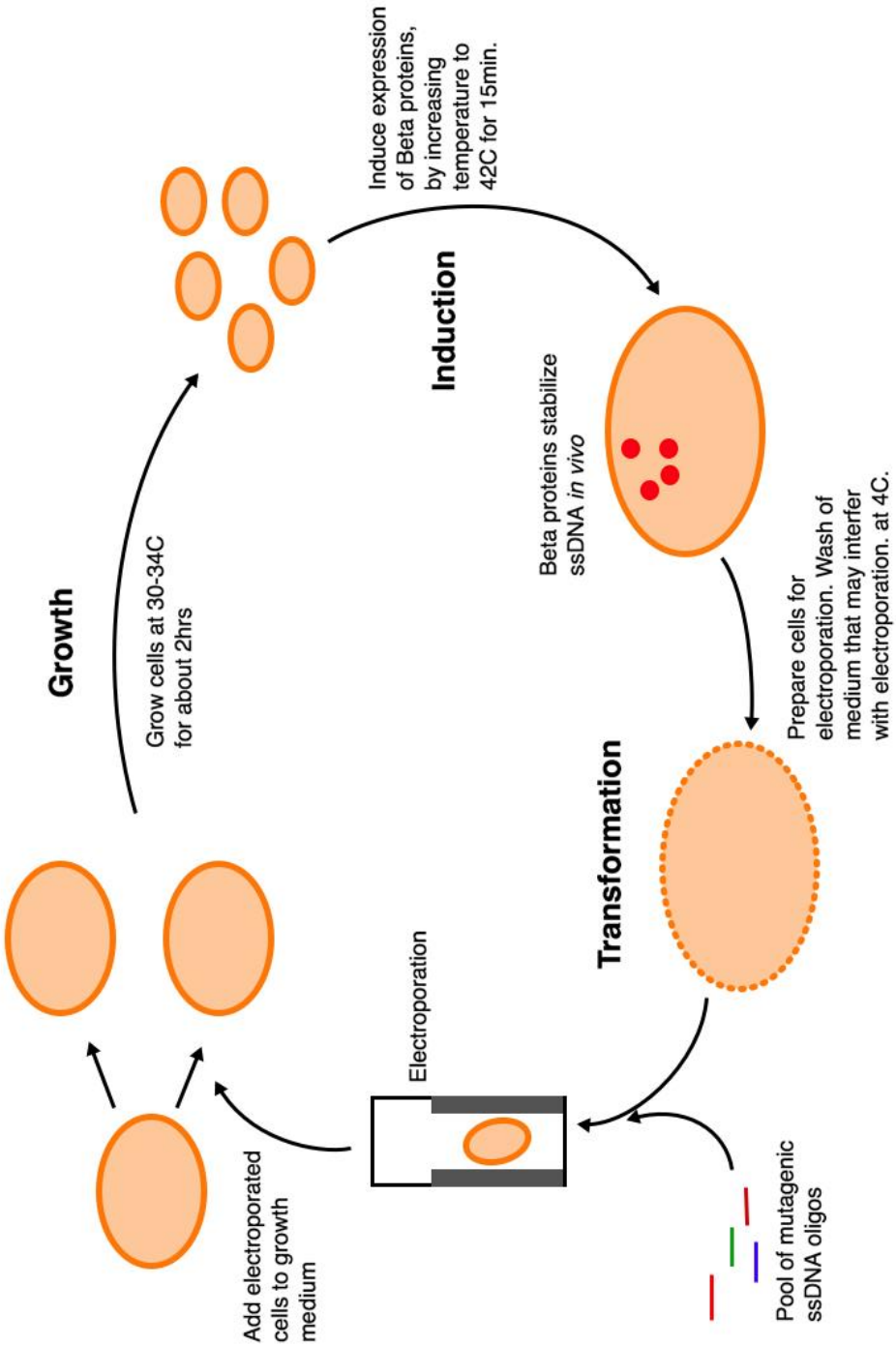


Figure 1.12: *Workflow in MAGE cycles, for genome engineering in E. coli [Gallagher et. al, 2014]. ssDNA = Single Stranded DNA. Oligos = oligonucleotides.*

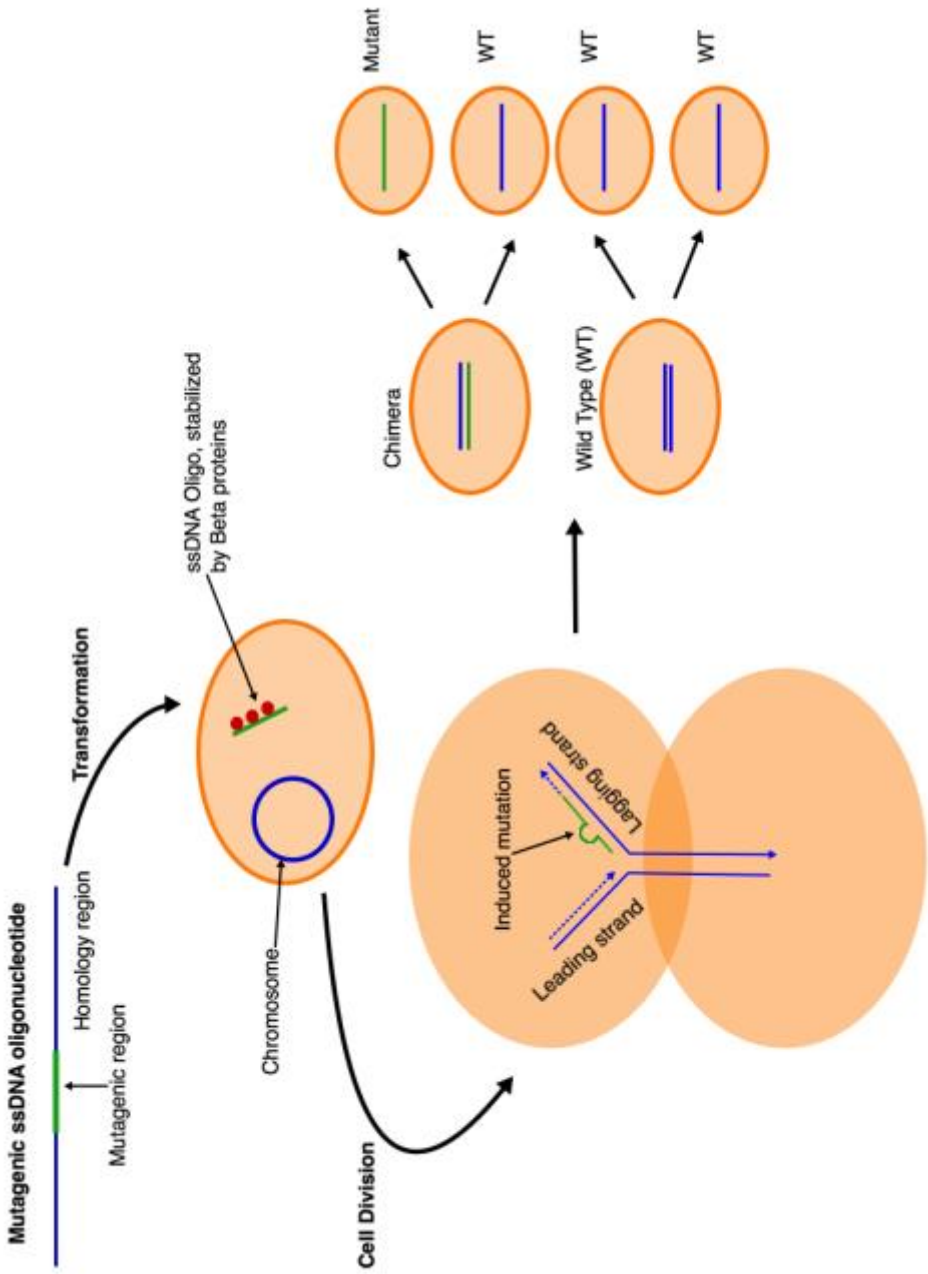


Figure 1.13: *Mechanism of MAGE mutagenesis [Gallagher et. al, 2014]. The mutagenic ssDNA oligonucleotide is stabilized by Beta proteins, and anneals to the lagging strand of the replication fork during cell division. This induces insertions, mismatches or deletions in the genome of 1/4 of daughter cells.*

To achieve effective mutagenesis with MAGE, the native DNA repairing methyl-directed mismatch repair system must be inactivated. This leads to an increased rate of random mutations in the cell's genome. To precisely evaluate the effect of specific mutations, the level of off-site mutations should be kept to a minimum.

A group of researchers have made an improved MAGE system [Nyerges et. al, 2015]. The new system, called "pORTMAGE", is based on a single plasmid, coding for *λ Red recombinase*, and a dominant mutant allele of the *MutL* protein. *MutL* is a key factor in the methyl-directed mismatch repair system, and expression of the dominant mutant allele of *MutL* was shown to essentially knock out this DNA repair system [Nyerges et. al, 2015]. By putting transcription of the dominant *MutL* allele under control of a heat inducible repressor, transient suppression of DNA repair was achieved in *E. coli*. Inactivating the cell's DNA repair mechanism only in periods of oligo integration yielded a significantly lower level of off-site mutations, compared to the regular MAGE protocol.

1.4.3 CRISPR/Cas9

Development of the CRISPR/Cas9 system has given researchers a powerful tool for genome engineering. Since it was discovered that CRISPR/Cas9 could be used for genome editing, in 2012, the number of publications based on this technology has virtually exploded [Doudna & Charpentier, 2014; crisprupdate.com].

CRISPRs (Clustered Regularly Interspaced Palindromic Repeats) were first observed in bacterial genomes in 1987 [Ishino et. al, 1987]. The biological function of these repeats was largely a mystery, until 2006, when it was proposed that CRISPRs are parts of a bacterial adaptive immune system [Makarova et. al, 2006]. Indeed, in 2010, CRISPRs were identified to work together with a endonucleotic DNA cleaving protein called Cas, in an adaptive bacterial immune response [Garneau et. al, 2010].

The CRISPR repeats, found in bacterial genomes, are sequences of foreign DNA, deriving from earlier viral infections [Doudna & Charpentier, 2014]. These sequences can be transcribed to form CRISPR RNAs (crRNAs). These RNAs interacts with another RNA specie (tracrRNAs), to recruit Cas protein. Once a Cas protein is recruited, crRNA guides the enzyme to digest DNA at very specific locations. In nature, this system allows bacteria to store foreign genetic information, to quickly degrade infectious DNA during a viral infection. The mechanism of CRISPR/Cas9 mediated adaptive immune response is depicted in Figure 1.14.

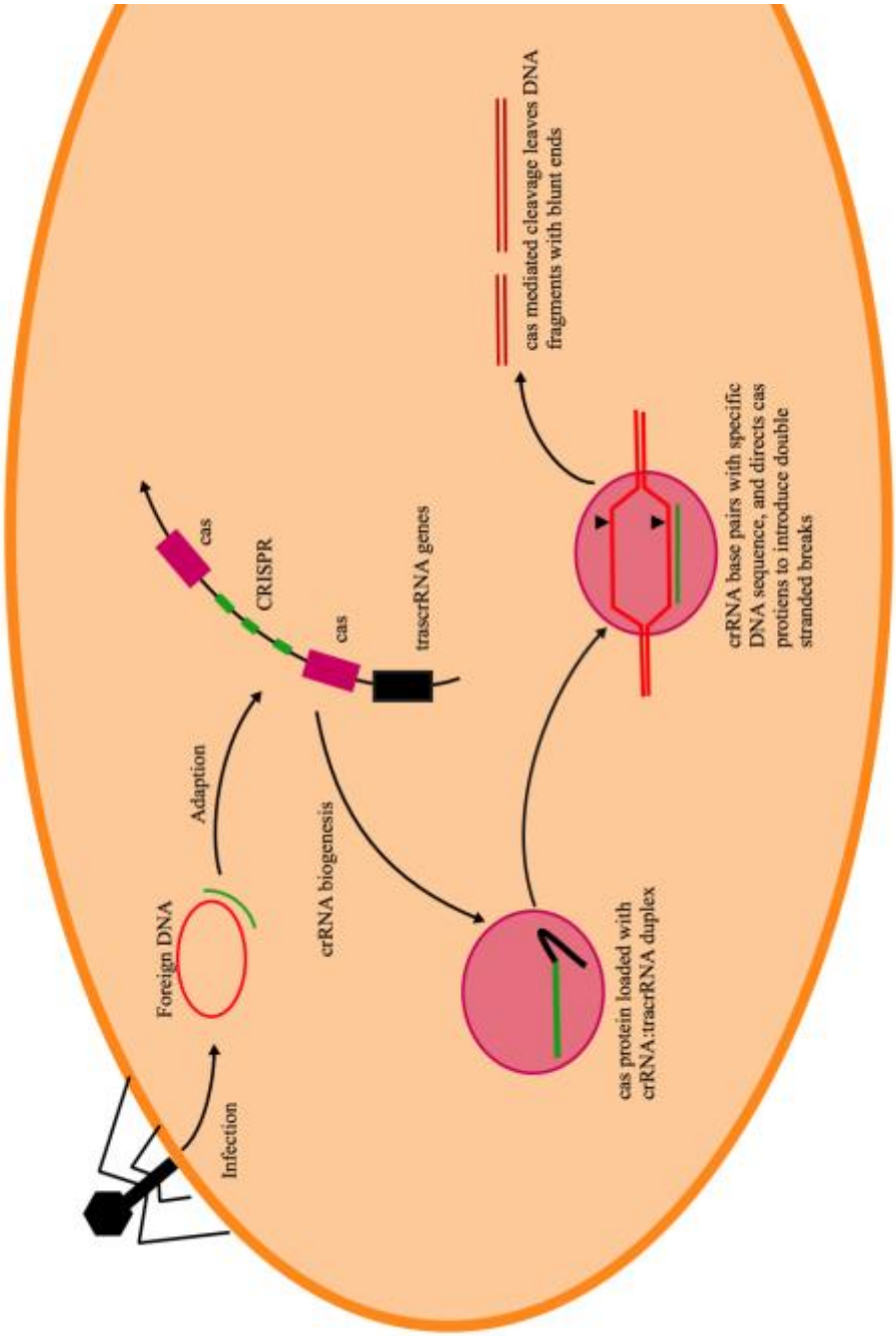


Figure 1.14: *CRISPR/Cas9 mediated adaptive immune response in bacteria [Doudna & Charpentier, 2014].*

In nature, three main components constitute the CRISPR/Cas system:

- Cas proteins, which unwind and digest DNA
- tracrRNAs, which recruit Cas proteins
- crRNAs that directs DNA cleavage, by base pairing with DNA sequences.

Researchers have successfully combined tracrRNAs, and crRNAs in a single guide RNA molecule denoted "sgRNA" [Doudna and Charpentier, 2014]. By altering the sequence in the DNA binding part of sgRNA, it is possible to program Cas9 proteins to cleave virtually any gene. The differences between natural CRISPR/Cas9 systems and the engineered system are highlighted in Figure 1.15.

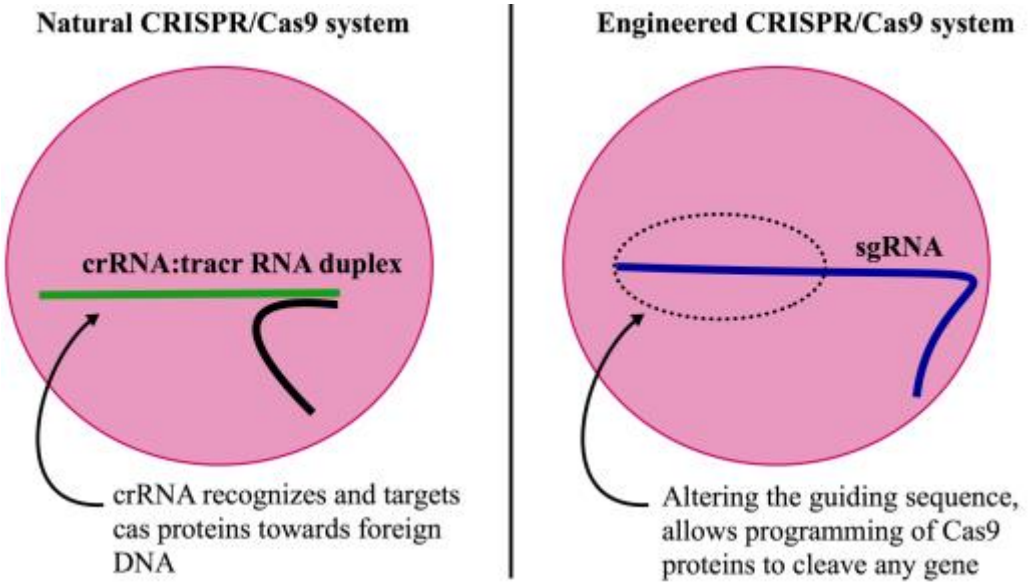


Figure 1.15: *Principal differences between natural and engineered CRISPR/Cas9 systems. [Doudna & Charpentier, 2014]*

Cas9 proteins cleave double stranded DNA, and leave blunt ends. If Cas proteins are targeted towards chromosomal genes, DNA breaks will trigger the endogenous cellular DNA repair machinery. DNA breaks can be repaired by non-homologous end joining of DNA. Alternatively, a donor DNA can be inserted at the cleavage sites, in a homologous recombination event. Non-homologous end joining will yield a knockout mutation, while a donor DNA of choice can be inserted with high precision during homologous recombination events. Hence, engineered CRISPR-cas systems allows researchers to alter, or insert genes with very high levels of precision.

It is important to notice that Cas9 will not successfully bind to and digest a DNA sequence if it is not followed by a Protospacer Adjacent Motif (PAM). PAM is a 3bp sequence (NGG), located directly downstream of the sequence targeted by Cas9. In nature, the PAM recognition ability of Cas9 provides a level of control, to ensure that it digests only foreign DNA, and not the host's DNA.

For CRISPR/Cas9 mediated genome editing in bacteria, a simple one-plasmid method has recently been developed [Zhao et. al, 2016]. In this system, Cas9, sgRNA and donor DNA is placed on the same plasmid, together with a DNA recombinase gene (*recA*). A schematic map of the principal plasmid composition is depicted in Figure 1.16. To control the timing of chromosome digestion, Cas9 genes are put under the control of an arabinose inducible promoter, with low level of basal expression.

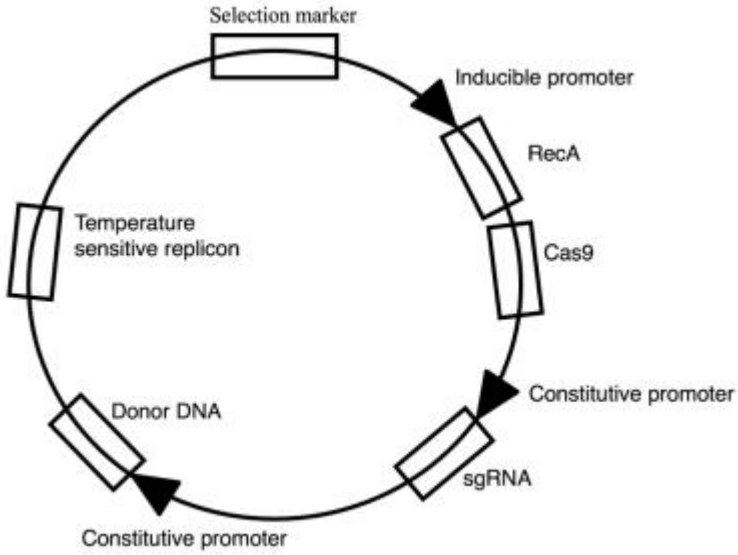


Figure 1.16: *Principal plasmid composition of an one-plasmid system for CRISPR/Cas9 mediated genome editing in bacteria [Zhao et. al, 2016]. The donor DNA constitutes a template for homologous recombination. A temperature sensitive replicon allows for easy plasmid curing after genome editing.*

1.4.4 CRISPR optimized MAGE recombineering (CRMAGE)

CRMAGE (CRISPR optimized MAGE recombineering) is a method based on MAGE recombineering, using CRISPR/Cas9 for negative selection against non-mutated cells [Ronda et. al, 2016]. MAGE oligos are used to induce point specific mutations (section 1.4.2), while simultaneously changing a nearby PAM sequence, recognised by Cas9 (section 1.4.3). Synthetic sgRNAs are used to guide Cas9 to create double stranded breaks, and thus kill cells that have not had the PAM sequence altered by the MAGE recombineering event. In that way CRISPR/Cas9 provides a negative selection against cells, that have not been mutated in the MAGE cycle. The mechanism of this method is further explained in Figure 1.17.

Ronda and colleagues induced a single point mutation in *E. coli* with a 98% efficiency, using a single round of CRMAGE recombineering, versus a 5% efficiency using standard MAGE recombineering [Ronda et. al, 2016]. The CRMAGE system allows for multiplex genome engineering, with high efficiencies, and can significantly decrease the laborious workload of genome engineering in bacteria.

Three genes were targeted for knockout in this work. Both CRMAGE and traditional lambda red recombineering were used to generate knockouts. The CRMAGE system used is accounted for in detail in section 5.7. The biological functions of the genes that were targeted for knockouts are described in section 1.6.

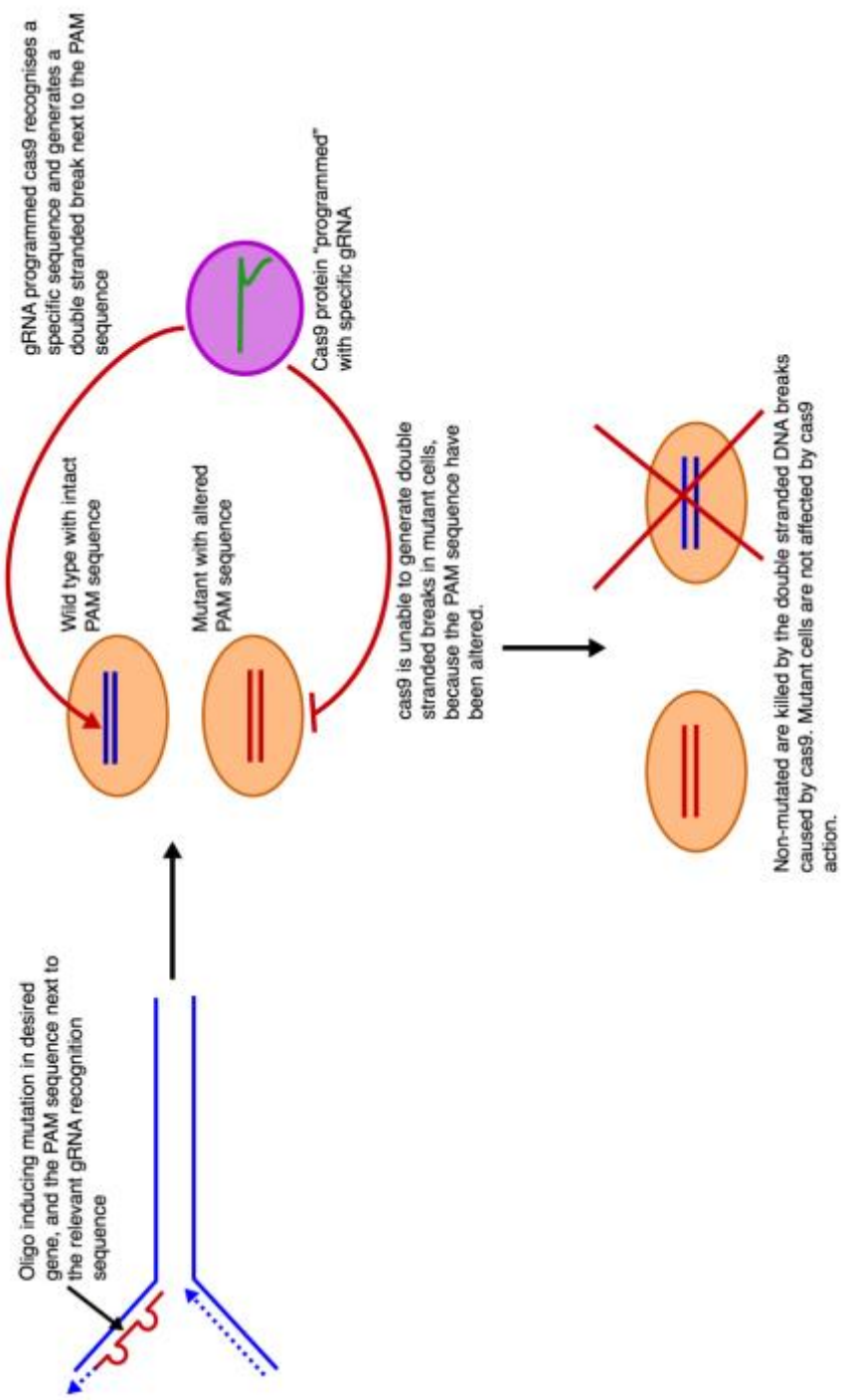


Figure 1.17: *General mechanism of CRISPR optimized MAGE (CRMAGE) [Ronda et. al, 2016].*

1.5 Aims and objectives of this thesis

The work presented in this thesis is done as a part of an international research project called LEANPROT. This thesis serves as a preparatory study for the continued work in the LEANPROT project. The aim of the project is to remove unnecessary and resource exhaustive proteins from *E. coli*, to free up resources for recombinant protein production. By reducing proteome sizes, it may be possible to increase translational capacities in *E. coli*. This can improve production capacities, as translation can be the rate limiting step in many recombinant protein production processes [Mahalik et. al, 2014]. The main objectives of this thesis were to:

1. Establish and characterize reporter systems to evaluate translational capacities in *E. coli*.

2. Investigate what kinds of genes should be knocked out, to increase translational capacities in *E. coli*.

3. Test genome engineering techniques that may be used to knock out genes from the *E. coli* chromosome. And evaluate if knocking out the same genes with different techniques yield different phenotypes.

To investigate strategies for increasing translational capacities, a few selected genes were knocked out, using different approaches. The relevant genome engineering techniques are described in section 1.4. The natural functions of the targeted genes are accounted for in section 1.6. Two approaches were implemented to evaluate translational capacities in the different knockout strains. These methods are described in section 1.7.

Section 2 describes all experimental results. Discussion of the experimental results is provided in section 3. How, and to what extent, each of the aforementioned goals were fulfilled is summarized in section 4.

1.6 Genes that were knocked out in this work

Three genes were targeted for knockouts in this work. The biological functions of these genes are accounted for in this section. Two general strategies were considered when selecting genes to knock out:

1. Knocking out stress related genes, to minimize detrimental stress responses that may occur during recombinant protein expression.

2. Knocking out genes that constitutes a large part of the global proteome.

Both strategies are discussed in further detail in section 1.3. Each gene being subject for the analysis in the presented work was knocked out in three different ways:

1. Single gene knockout mutants were received from the KEIO collection [Baba et. al, 2006]. In these strains, the target genes were replaced with kanamycin resistance cassettes, using traditional lambda red recombineering [Datsenko & Wanner, 2000].

2. A second set of single gene knockout mutants was created by removing the kanamycin resistance cassettes from the KEIO strains. This was done by inducing expression of FLP Recombinase in the knockout mutants from the KEIO collection [Datsenko & Wanner, 2000].

3. A third set of knockout mutants was created by using CRISPR optimized MAGE (CRMAGE) [Ronda et. al, 2016]. In the CRMAGE mutants, the genes' initiation codons were changed from ATG to AGG. This was expected to cause functional knockouts by decreasing levels of gene expression by a factor of roughly 10,000 [Hecht et. al, 2017].

Each set of mutants contained strains with single gene knockouts of the same three genes: *uspA*, *ompA* and *dps*. The cellular functions of these genes are discussed in section 1.6.1 - 1.6.3.

1.6.1 Universal stress protein A (UspA)

Universal stress protein A (UspA) is one of six proteins in the Usp family in *E. coli* [Kvint et. al, 2003]. As the name suggests, expression of UspA is induced by several stress triggers, including: entry to stationary phase, nutrient starvation, heat shock and exposure to ethanol, oxidants or antibiotics. UspA is one of the most abundant intracellular proteins in growth arrested *E. coli* cells.

It is hypothesized that UspA might have several functions *in vivo* [Kvint et. al, 2003]. It has been shown that the protein protects stationary phase cells from DNA damage [Gustavsson et. al, 2002]. Cells devoid of UspA die off quicker than wild type cells during periods of growth arrest [Nystrøm & Neidhart, 1993; Nystrøm et. al, 1994]. Overproducing UspA has been shown to lock cells in growth arrested states [Nystrøm & Neidhart, 1996].

UspA probably does not constitute a large fraction of the global proteome during exponential growth. However, UspA expression is quickly induced when growth rates fall below the maximal growth rate supported by the media, and reaches peak levels in ~20min [Nystrøm & Neidhart, 1993]. In periods where cells are experiencing some degree of stress (i.e. during recombinant protein expression) this protein will probably constitute a considerable part of the global proteome.

1.6.2 Outer membrane protein A (OmpA)

Outer membrane protein A (ompA) is an abundant protein in *E. coli*, occurring at a copy number of ~100,000 units per cell during exponential growth [Smith et. al, 2007]. It is a multifaceted molecule with many diverse roles. OmpA is extensively studied as a virulence factor in meningitis caused by *E. coli*. The protein has been shown to enhance outer membrane integrity, as *ompA* deficient mutants were more sensitive towards high osmolarity and acidic conditions compared to wild type cells [Wang, 2002].

Synthesis of OmpA is growth rate dependent, and increase proportionally with growth rate [Smith et. al, 2007]. In addition, expression of OmpA is altered in response to a wide range of stimuli. Some of the stimuli affecting expression are summarized in Table 1.1

Table 1.1: *The effect of different stimuli on ompA expression in E. coli [Smith et. al, 2007]*

Stimulus	Effect on level of expression
Nitrogen shortage	Increased
Acidic media	Increased
Adhesion to abiotic surfaces	Decreased
Entry into stationary phase	Decreased

As OmpA is a multifunctional protein the phenotypical effects of a knockout mutation in the *ompA* gene is hard to predict. OmpA is an abundant protein in exponential phase cells. Hence, it is possible that *ompA* deletion mutants can show increased growth rates, and higher translational capacity towards recombinant proteins. It is possible that *ompA* knockouts will have deleterious effects on cells growing in in acidic, - or high osmolarity conditions.

1.6.3 DNA binding protein from starved cells (*Dps*)

The DNA binding protein from starved cells (Dps) provide protection to cells during exposure to environmental stresses [Calhoun & Kwon, 2011]. The protein has three main properties: DNA binding, iron sequestration and ferroxidase activity. Dps protects cells from a large range of environmental assaults, including: high pressure, UV irradiation, oxidative stress, acidic stress, iron poisoning, ferroxidase poisoning and nutritional deprivation.

During exponential growth, Dps regulation is controlled by the redox-sensitive OxyR regulator. In stationary growth phases, the sigma factor σ^s controls expression. Dps constitute only a small part of the *E. coli* proteome during exponential phase (~6,000 units per cell), but during transition to stationary phase the intracellular Dps levels rise to about 180,000 units per cell [Calhoun & Kwon, 2011].

Dps protects cells from a variety of stresses. Hence, it is reasonable to assume that a wide range of stimuli may trigger Dps expression, through the complex stress response networks in *E. coli*. It is possible that *dps* knockout mutants will be less likely to induce unwarranted detrimental stress responses to the bioreactor environment, compared to wild type cells. On the other hand, it is likely that *dps* knockout strains will be more sensitive than wild type cells towards certain conditions. More careful regulation of the bioreactor environment might be necessary to run a process with *Dps* knockout mutants, compared to wild type cells.

Dps is only a small part of the *E. coli* proteome during unstressed exponential growth. Hence, *Dps* deletion mutants will probably not have significantly leaner proteomes during exponential phase, compared to wild type cells.

A summary of the genes knocked out in this work, are given in Table 1.2.

Table 1.2: *Genes knocked out in this work. The 'significance in this work' column relates the characteristics of each gene to the discussion in section 1.3.*

Gene	Cellular function	Regulation	Significance in this work
uspA	Protection from several different perturbations	Expression of UspA is triggered by various stresses and entry to stationary phase. UspA is one of the most abundant proteins in growth arrested *E. coli* cells.	*uspA* knockouts might increase translational capacities by decreasing the stress related fraction of the proteome (ΦS).
dps	Protection from several different perturbations	Expression of Dps is triggered by oxidative stress and entry to stationary phase. Dps is an abundant protein in *E. coli* during stationary growth phases.	*dps* knockouts might increase translational capacities by decreasing the stress related fraction of the proteome (ΦS).
ompA	Virulence Biofilm formation Outer membrane stabilization	OmpA is an abundant protein in *E. coli* during exponential phase growth. Expression of *ompA* increased proportionally with growth rates during exponential growth phases.	*ompA* knockouts might increase translational capacities by decreasing the fixed part of the proteome (ΦQ). Synthesis of OmpA might constitute a major translational burden during exponential growth.

1.7 Methods used to compare translational capacities of knockout mutants

Two different methods were used to compare the translational capacities in the knockout strains investigated in this work. A fluorescence reporter system was used to evaluate levels of ribosome expression, based on the activity from two ribosomal promoters (rrnBp1 and rpsJ), each present on different reporter plasmids. In addition, growth rate during exponential phase was used as a proxy to evaluate the strains' translational capacities.

1.7.1 Reporter system to monitor rRNA and ribosomal protein synthesis

In previous work (Biotechnology specialization course, TBT4505), reporters to measure the activities from two ribosomal promoters in *E. coli* were developed and tested by the author of this project. This was done by constructing two plasmids, in which the expression of a red fluorescent protein (mCherry) was controlled by a ribosomal promoter (one promoter on each plasmid). The two ribosomal promoters used were rrnBp1 and rpsJ.

rrnBp1 regulates transcription of one of the 7 rRNA operons in *E. coli* [Maeda et. al, 2005]. The promoter region contains 3 Fis binding sites, 3 H-NS binding sites and 1 Lrp (Leucine-responsive regulatory protein) binding site upstream of the -35 sequence [EcoCyc, rrnBp1]. rpsJ regulates transcription of the S10 operon in *E. coli* [Lemke et. al, 2011]. 11 ribosomal proteins (rProteins) are included in this operon. The rpsJ promoter is downregulated by (p)ppGpp/DksA.

Maps of the two plasmids, carrying the ribosomal promoters and red fluorescent reporter proteins, are depicted in Figure 1.18. Description of the genetic elements in the plasmids are given in Table 1.3.

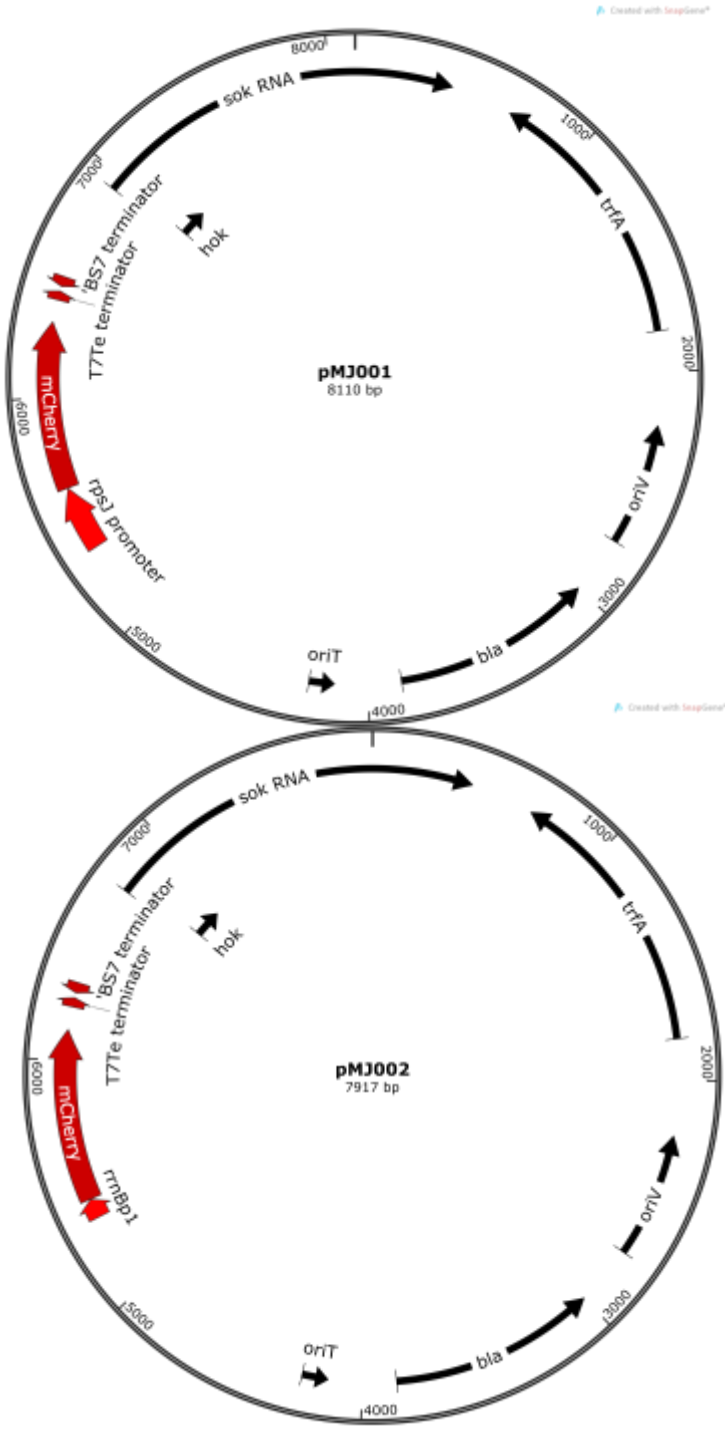


Figure 1.18: *The two reporter plasmids used to investigate activity from rrnBp1 and rpsJ. The ribosomal promoter region and the red fluorescent protein (mCherry) gene are highlighted with red colour. The plasmid containing the rpsJ promoter was named "pMJ001". The plasmid containing the rrnBp1 promoter was named "pMJ002". All genetic elements displayed in this figure are explained in Table 1.3.*

Table 1.3: *Description of genetic elements in Figure 1.18.*

Genetic element	Function
rpsJ/rrnBp1	Ribosomal promoters
mCherry	Red fluorescent protein
T7Te and BS7	Terminator sequences
Hok and sok RNA	RNAs that increase plasmid stability by killing plasmid-free daughter cells [Thisted & Gerdes, 1992]
trfA	Protein involved in initiation and control of replication from oriV
oriV and oriT	Origins of replication
bla	Ampicillin resistance gene (Selection marker)

As the level of fluorescence from a culture may be heavily dependent on the cell count in the culture, OD (optical density) adjusted levels of fluorescence (relative fluorescence / OD_{600}) should be used to evaluate activity from the promoters.

1.7.2 Growth rate as a proxy for translational capacity

The growth rates of exponentially growing cells were used to evaluate the cells' translational capacities. During exponential phases, growth rates depend on the rate at which cells can duplicate their own proteome [Molenaar et. al, 2009; Scott et. al, 2014]. This, in turn is dependent on the translational capacity of the cell. Hence, measuring strain's growth rates can provide an indication towards their translational capacities.

In theory, it is expected that strains with reduced proteome sizes will show increased growth rates [Fischer & Sauer, 2005; Muntel et. al, 2014]. These strains may grow faster because they have smaller proteomes, that need to be duplicated before a cell division event. Consequently, these strains will have more available translational capacity towards recombinant protein production.

When using growth rates as a proxy for translational capacities, strains without reporter plasmid inserts should be tested. Recombinant protein expression might vary unreliably between different strains, and corrupt growth rate comparisons data. A short case study can highlight this phenomenon:

Two knockout strains, with the same reporter plasmid inserts are grown in the same media. The growth rates of the two strains are approximately equal, but "strain A" produced much more reporter protein than "strain B". Strain A might have had a higher translational capacity than Strain B, because it maintained the same growth rate while producing more protein. However, this effect is hard to evaluate as the induced knockout might have affected the regulation of reporter protein transcription. Without additional information about rates of transcription of the reporter gene, it is hard to know how much of the apparent increase in production rates one can credit to an actual increase in translational capacities.

If one test strains without plasmid inserts, growth rates can be more reliably compared as recombinant protein expression will not be a source of variance. Different antibiotics might also retard growth unreliably, and cells should be cultured in medias without antibiotics.

Even when taking these precautions, one should be critical in directly attributing the increase (or lack thereof) in growth rate to the knockout of one gene. While some knockout mutations might indeed create a leaner proteome, they might also make cells less adaptable to some conditions, and in that way, decrease growth rates. Mutants with such knockouts might not show increased growth rates before they are introduced to more optimal conditions.

Growth rates should be used critically during evaluation of translational capacities, and should preferably be accompanied by other reporter systems. In this work, growth rates and activities from two ribosomal promoters were used together to evaluate translational capacities in *E. coli*.

2 Results

2.1 Strategy and experimental workflow

All experiments were designed to fulfil the aims of this thesis. The outlined objectives were to:

1. Implement and characterize reporter systems to evaluate translational capacities in *E. coli*.

2. Investigate what kinds of genes that should be knocked out, to increase translational capacities in *E. coli.*

3. Test genome engineering techniques that may be used to knock out genes from the *E. coli* chromosome. And to evaluate if knocking out the same genes, with different techniques, yield different phenotypes.

To fulfil the aforementioned objectives, three genes (*uspA, dps* and *ompA)* were knocked out in *E. coli BW25113*. The cellular functions of the knocked-out genes are summarized in Table 2.1. Three sets of single-gene knockout mutants, generated with different methods, were evaluated.

The first set of single gene knockout strains were received from the KEIO collection. The KEIO collection is a library of almost 4000 single gene knockout mutants of *E. coli BW25113* [Baba et al, 2006]. In these strains, the knocked-out genes were replaced by kanamycin resistance cassettes. FLP Recombinase (FLP) recombineering was used to remove the kanamycin resistance genes from the KEIO strains [Datsenko & Wanner, 2000], yielding a second set of single gene knockout mutants.

A third set of single gene knockout mutants was generated, using CRMAGE (CRISPR optimized MAGE recombineering – section 1.4.4). In these mutants, genes were not removed from the genome, but the sequence in the initiation codons were altered. Initiation codons were changed from AUG to AGG. This was expected to cause functional knockouts by decreasing levels of gene expression by a factor of roughly 10,000 [Hecht et. al, 2017]. The three different methods used to generate the mutations are summarized in Table 2.2.

***Table 2.1**: Function and significance of the genes knocked out in this work.*

Gene	Cellular function	Significance in this work
uspA	Protection from several perturbations	Expression of UspA is triggered by various stresses [1]. *uspA* knockouts might increase translational capacities by decreasing the stress related fraction of the proteome (ΦS).
dps	Protection from several perturbations	Expression of Dps is triggered by oxidative stress and entry to stationary phase [2]. *dps* knockouts might increase translational capacities by decreasing the stress related fraction of the proteome (ΦS).
ompA	Virulence Biofilm formation Outer membrane stabilization	OmpA is an abundant protein in *E. coli* during exponential growth phases [3]. *ompA* knockouts might increase translational capacities by decreasing the fixed part of the proteome (ΦQ).

[1]: Kvint et. al, 2003 [2]: Calhoun & Kwon, 2011 [3]: Smith et. al, 2007

***Table 2.2:** Methods used to generate knockouts. Each method is described in greater detail in section 1.4.*

Method name	KEIO collection	FLP recombination	CRMAGE
Knockout annotation	*Δgene::Kan*	*Δgene*	*Δgene_CR*
Explanation	Gene replaced with kanamycin resistance cassette.	Kanamycin resistance cassette removed from KEIO mutants.	Genes knocked out, by changing initiation codons from ATG to AGG.
Ref. to section with theory about method	1.4.1	1.4.1 5.6	1.4.4 5.7

The genes listed in table 2.1 were chosen as targets for knockout mutations following the discussion provided in section 1.3. The general rationale of this approach is that by knocking out non-essential genes, cells can allocate resources more effectively towards expression of recombinant proteins.

Two approaches were used to evaluate translational capacities of the different knockout mutants: a plasmid based, fluorescence reporter system was used as a reporter for ribosome expression. In addition, exponential phase growth rates were used as a proxy for translational capacities. Both methods are discussed in further detail in section 1.7. Characterization of the ribosome expression reporter system is described in section 2.2.

The general workflow of this thesis is summarized in Figure 2.1. Three genes were targeted for knockout, using three different techniques. Single gene knockout mutants were tested for increased translational capacities, using two different reporter systems.

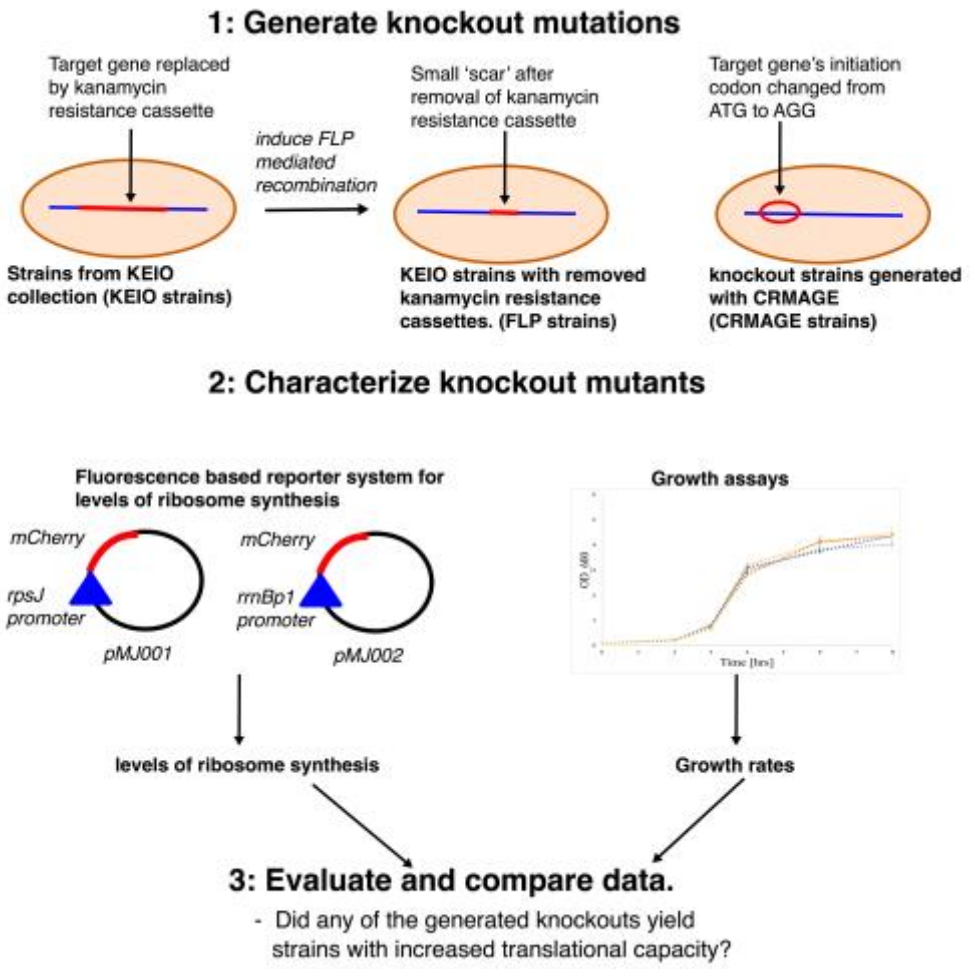


Figure 2.1: *The workflow of generating and characterising knockout mutants. Three different kinds of knockout mutations were investigated. Strains were compared with regard to their translational capacities. Translational capacities were estimated based on levels of ribosome expression, and growth rates during exponential growth phases.*

All data presented in this thesis are average values of at least three parallels. ANOVA analysis and t-tests were used to evaluate the significance of all results. The theory behind these statistical methods are described in appendix B. In general, a significance level of 5% was used. Error analysis are described in appendix B.

2.2 Characterization of the fluorescence based reporter systems for ribosome expression

Plasmid based fluorescence reporter systems were used to measure the activities from two ribosomal promoters in *E. coli.* Each plasmid carried a red fluorescence reporter gene (*mCherry*) under the control of a ribosomal promoter. The ribosomal promoters were rpsJ and rrnBp1. In wild type *E. coli*, rpsJ controls transcription of 11 rProtein genes, while rrnBp1 controls transcription of one of seven rRNA operons. Levels of expression from these promoters were evaluated by measuring fluorescence from strains carrying the reporter plasmids. The reporter plasmids were named pMJ001 (rpsJ promoter) and pMJ002 (rrnBp1 promoter). These plasmids were constructed by the author of this thesis, in earlier work (Biotechnology specialization course, TBT4505).

These reporters were used to evaluate the levels of expression of ribosomal constituents in the different knockout strains. Fluorescence readings from strains carrying reporter plasmids provided indications towards their levels of ribosome expression.

The reporter systems were characterized in wild type *E. coli BW25113.* Strains carrying reporter plasmids were cultured in LB medium with 100µg/ml ampicillin, at 37°C. Cell densities and relative fluorescence were measured at appropriate intervals for 8 hours. The resulting data are presented in Figures 2.2 and 2.3.

Fluorescence from strains carrying the rrnBp1 promoter was significantly higher, compared to strains carrying the rpsJ promoter (Figure 2.2), indicating that rrnBp1 is a stronger promoter than rpsJ in *E. coli BW25113*. The OD adjusted fluorescence (Relative fluorescence / OD_{600}) after 8 hours of inoculation was approximately 25% higher in cultures with rrnBp1 promoters, compared to cultures with rpsJ promoters.

Red fluorescent protein (Rfp) synthesis seemed to be heavily correlated with growth rate in both strains. This is highlighted in Figure 2.3, where it is visible that OD adjusted fluorescence remained approximately constant throughout the cultivations. The level of ribosome synthesis is one of the main determinants for growth rates in bacteria [Molenaar et. al, 2009; Scott et. al, 2014]. Hence, it is expected that the activity from ribosomal promoters should be growth rate dependent. The observation that levels of fluorescence indeed were correlated with growth rates indicated that these systems has merit as reporters for levels of ribosome expression.

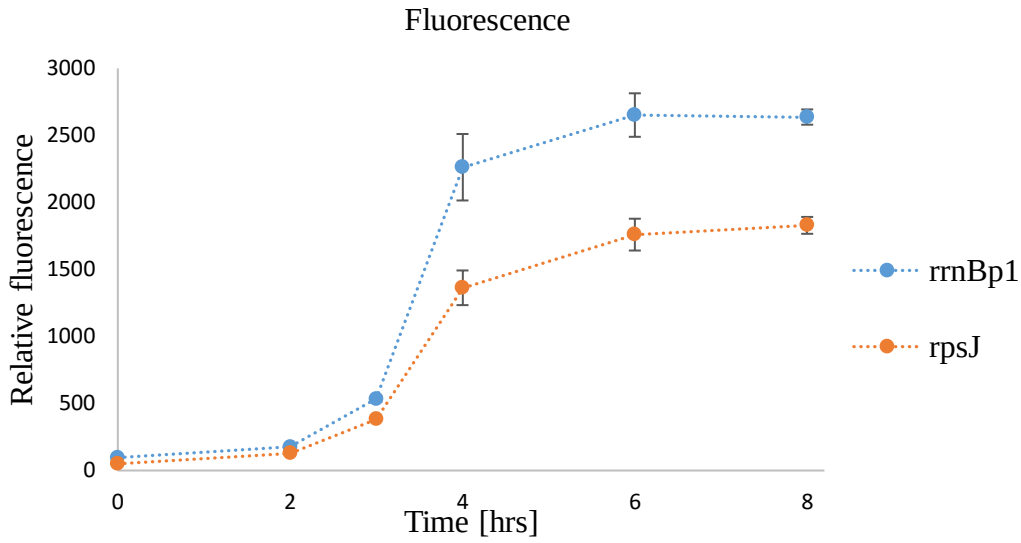


Figure 2.2: *Relative fluorescence from E. coli BW25113 strains expressing red fluorescent proteins under control of two different ribosomal promoters (rrnBp1 and rpsJ). Each data point is depicted as an average value of three parallels. Error bars are depicted as standard deviations of mean (SDOM).*

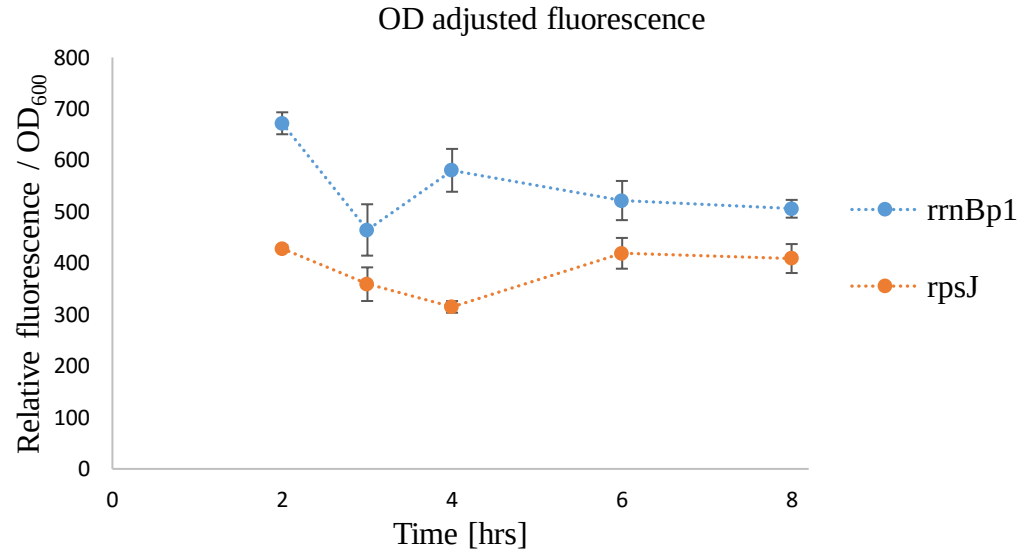


Figure 2.3*: OD adjusted fluorescence (relative fluorescence/OD_{600}) from E. coli BW25113 strains, expressing red fluorescent proteins under control of two different ribosomal promoters (rrnBp1 and rpsJ). Each data point is depicted as an average value of three parallels. Errors are calculated based on SDOM values of fluorescence and optical density readings. See appendix B for details about error analysis.*

The ribosome expression reporter systems were further characterized by treating *E. coli BW25113* with sub-lethal concentrations of chloramphenicol. Chloramphenicol inhibits protein synthesis by inhibiting the peptidyl transferase activity of ribosomes. Matthew Scott and colleagues [Scott et. al, 2010] discovered a negative correlation between growth rates and ribosome concentrations in *E. coli* growing in sub-lethal concentrations of chloramphenicol. This negative correlation was explained by the assumption that when translation is inhibited, the cell receives signals indicating that more ribosomes are needed to maintain protein expression. Hence, ribosome synthesis is increased, driving the proteome partitioning away from the growth rate maximizing optimum (Figure 2.4).

Scott and colleagues used intracellular RNA/protein ratio as a proxy for ribosome concentration. Similar experiments were done, using the fluorescence reporter systems investigated in this work. The results from these experiments are presented in Figure 2.5.

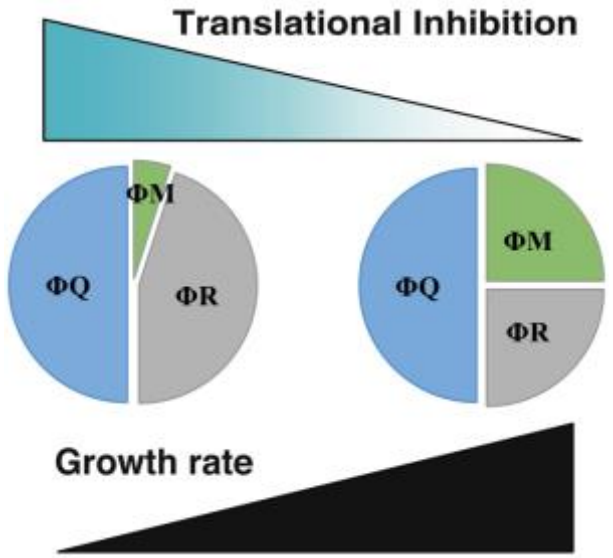


Figure 2.4: *Coarse grained proteome partitioning during inhibition of translation [Scott et. al, 2010]. ΦQ: fixed part of proteome, ΦM: metabolic proteins, ΦR: ribosomal proteins. When translation is inhibited (i.e. by treatment with antibiotics) growth rates decrease, while ribosome synthesis is upregulated.*

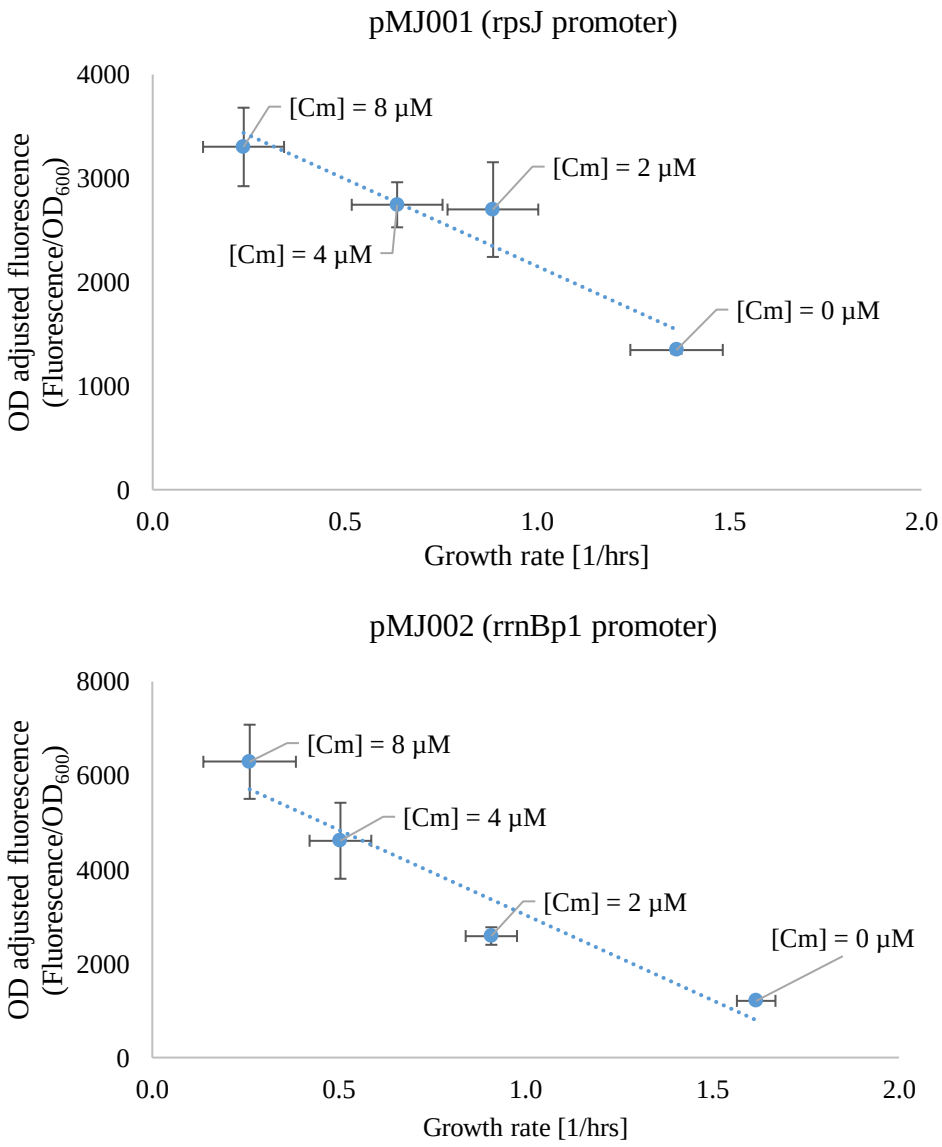


Figure 2.5: *Growth rate versus OD adjusted fluorescence for E. coli BW25113 carrying reporter plasmids. Cells were inoculated in LB with sub-lethal concentrations of chloramphenicol (Cm). Cm concentrations ranged from 0µM to 12µM. Growth rates were calculated from OD_{600} values measured at 30, - 60, - and 90 minutes after inoculation. OD adjusted fluorescence values were calculated from measurements made 2 hours after inoculation. Vertical error bars are presented as Standard deviations of mean (SDOM). Horizontal error bars were calculated from SDOM values of OD_{600} measurements (See appendix B for details about error analysis).*

Cells treated with sub-lethal concentrations of chloramphenicol showed increased activity from both ribosomal promoters (rpsJ and rrnBp1), while growth rates declined (Figure 2.5). This negative correlation is accordance with other results, obtained from similar experiments [Scott et. al, 2010]. These results confirmed that the fluorescence based reporter systems provide reliable information about ribosome expression in *E. coli.*

The OD adjusted fluorescence levels from strains carrying reporter plasmids remained approximately constant throughout 8 hours of inoculations in LB medium (Figure 2.3). Hence, OD adjusted fluorescence levels, measured at the end of cultivations, were deemed reliable to evaluate levels of ribosome expression throughout exponential and stationary growth phases.

2.3 Characterization of translational capacities of knockout mutants from the KEIO collection

Three *single gene* knockout mutants of *E. coli BW25113* were received from the KEIO collection:

1. *uspA* knockout mutant (named "*ΔuspA::kan"*)
2. *ompA* knockout mutant (named "*ΔompA::kan"*)
3. *dps* knockout mutant (named "*Δdps::kan"*)

In these strains, the knocked-out genes had been replaced by a kanamycin resistance marker. To evaluate the translational capacities of the "KEIO strains", growth characteristics and levels of ribosome expression were investigated.

2.3.1 No significant difference in growth characteristics between KEIO strains and wild type *E. coli BW25113* was observed.

Strains without reporter plasmids were inoculated in LB medium at 37°C for 8 hours. The resulting growth curves are depicted in Figure 2.6.

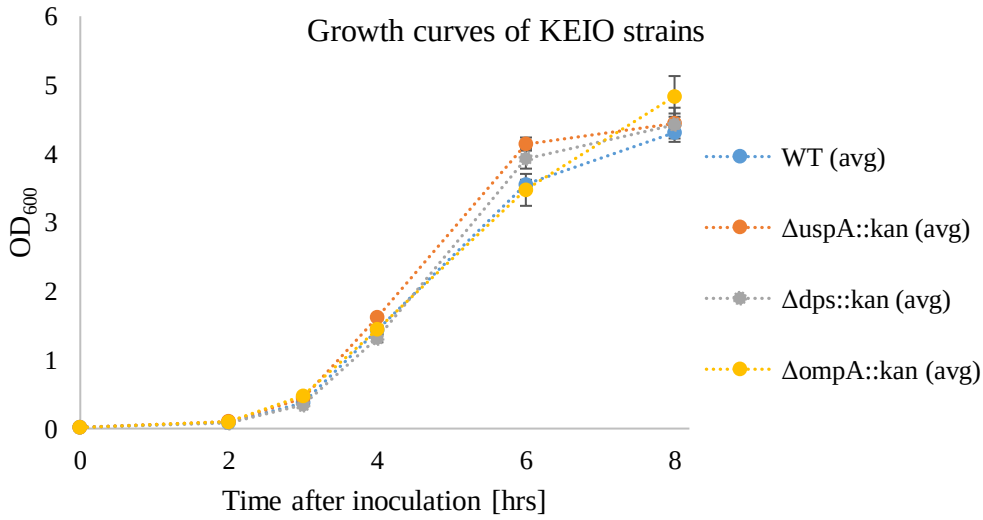
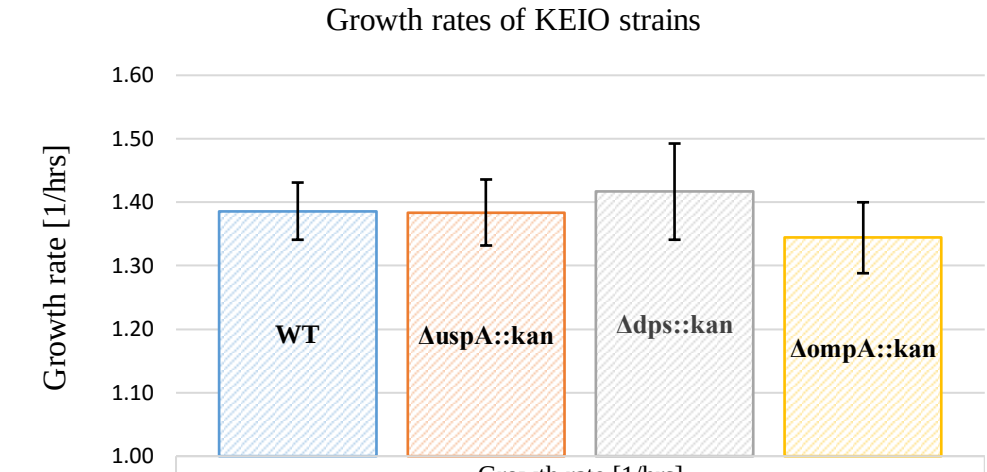


***Figure 2.6**: Growth curves for knockout strains from KEIO collection. Cells were grown in LB medium at 37℃, 225 RPM. Error bars are depicted as SDOM. WT = wild type E. coli BW25113.*

The growth rate in each culture was calculated based on the OD_{600} measurements at 2, 3 and 4 hours after inoculation (Figure 2.6). The growth rate of each strain is depicted in Figure 2.7.



	Growth rate [1/hrs]
WT	1.39
ΔuspA::kan	1.38
Δdps::kan	1.42
ΔompA::kan	1.34

Figure 2.7: *Growth rates of knockout strains from the KEIO collection. Errors were calculated based on the SDOM values of the optical density data, depicted in Figure 2.6. See appendix B for further details about error analysis.*

There was not observed any significant differences in growth characteristics between the KEIO strains and wild type *E. coli BW25113* (Figure 2.6 and 2.7). The growth rate of all strains was about $1.4hrs^{-1}$, corresponding to a doubling time of 30 minutes.

2.3.2 Expression from ribosomal promoters were increased in two single gene knockout strains

The levels of ribosome expression in the KEIO strains were evaluated by using the fluorescence based reporter systems. Both reporter plasmids (pMJ001-rpsJ and pMJ002–rrnBp1) were successfully transformed into all strains. Cells carrying the reporter plasmids were inoculated in LB medium, with ampicillin, for 8 hours. Relative fluorescence and cell densities were measured at the end of cultivations. The OD adjusted fluorescence values are depicted in Figure 2.8.

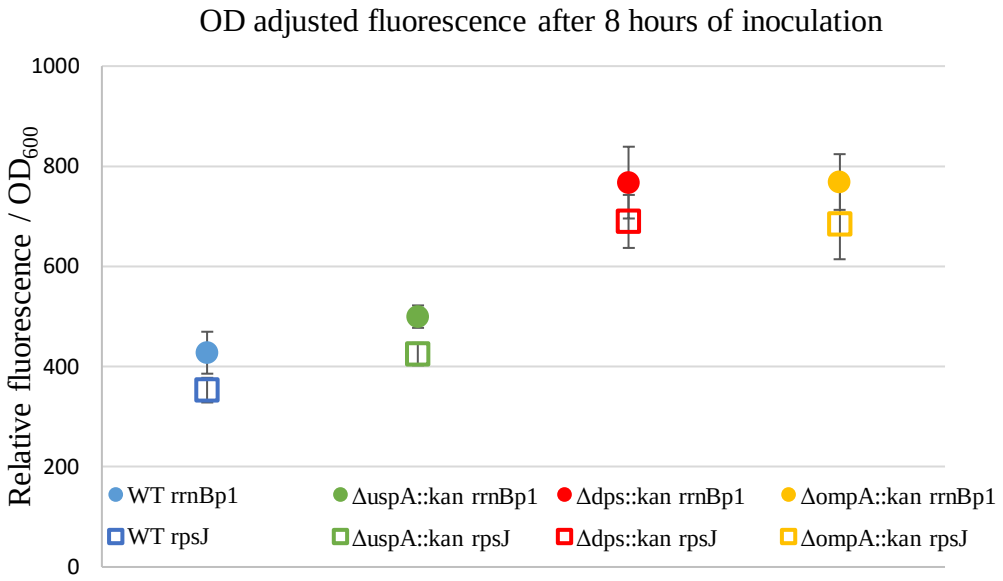


Figure 2.8: *levels of OD adjusted fluorescence from KEIO strains. Fluorescence from strains carrying the pMJ001 (rpsJ) plasmid are marked with squares, fluorescence from strains with the pMJ002 (rrnBp1) are marked with circles. Error bars are depicted as SDOM.*

The *Δdps::kan*, - and *ΔompA::kan* strains had significantly higher levels of OD adjusted fluorescence, compared wild type cells (Figure 2.8). For all strains, except *ΔuspA::kan*, there was no significant difference between the activities from the two ribosomal promoters (rrnBp1 and rpsJ). The statistical significance of these results was confirmed by doing an ANOVA analysis on the entire data set. The ANOVA table is depicted in Figure 2.9.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Strain	541806.6	1	541806.6	146.66	0
Reporter plasmid	36602.5	1	36602.5	9.91	0.0051
Strain*Reporter plasmid	114.8	1	114.8	0.03	0.8618
Error	73885.3	20	3694.3		
Total	652409.3	23			

Figure 2.9: *ANOVA analysis on OD adjusted fluorescence data from KEIO strains. The Prob>F factor is the main output factor, and describes the probability that the effect is random. Only effects with Prob>F values smaller than 0.05 were accepted as significant.*

As per the ANOVA analysis (Figure 2.9), the difference between the activities of the two ribosomal promoters (rrnBp1 and rpsJ) was deemed significant. By doing multiple t-tests, this effect was credited solely to the *ΔuspA::kan* strain, as the differences between the two promoters were deemed statistically insignificant in all other strains.

The ANOVA analysis, and multiple t-tests, confirmed that two of the knockout mutations (*Δdps::kan* and *ΔompA::kan*) had a significant effect on the fluorescence output from both reporter plasmids. Indicating that the level of ribosome expression was increased by replacing the *ompA*, - and *dps* gene with kanamycin resistance cassettes through lambda red recombination. No significant difference was observed between the *ΔuspA::kan* strain and wild type cells.

2.4 Characterization of translational capacities of KEIO strains with removed kanamycin resistance cassettes

The kanamycin resistance cassettes in the KEIO strains were flanked by two FLP recognition sites [Baba et. al 2006]. Following the protocol described by Datsenko and Wanner [Datsenko and Wanner, 2000], the kanamycin resistance cassettes were removed by inducing FLP Recombinase expression in the KEIO strains. This was done by transforming the strains with a plasmid (pCP20), carrying the yeast *flp recombinase* gene. The pCP20 plasmid contained a heat sensitive replicon that enabled curing of the plasmid by growing cells at 37°C.

Strains with successfully removed kanamycin resistance genes, and cured pCP20 plasmids, were screened by patching colonies on LB plates with ampicillin (resistance from pCP20), LB plates with kanamycin, and LB plates without antibiotics, in that order. Colonies that showed growth only on LB plates without antibiotics were assumed to constitute cells with successfully removed kanamycin resistance genes and cured pCP20 plasmids.

Mutants with removed kanamycin cassettes and cured pCP20 plasmid were successfully isolated for each of the KEIO strains, yielding a new set of single-gene knockout mutants (*ΔuspA, ΔompA* and *Δdps*). Experiments to evaluate translational capacities of these strains were conducted in the same way as with the KEIO strains (Section 2.3).

2.4.1 *ompA* deletion strains showed increased growth rates compared to wild type *E. coli BW25113*

ΔuspA, ΔompA and *Δdps* strains without reporter plasmids were inoculated in LB medium for 8 hours. The resulting growth curves are depicted in Figure 2.10. The growth rate in each culture was calculated based on the OD_{600} measurements at 2, 3 and 4 hours after inoculation. The growth rate of each strain is depicted in Figure 2.11.

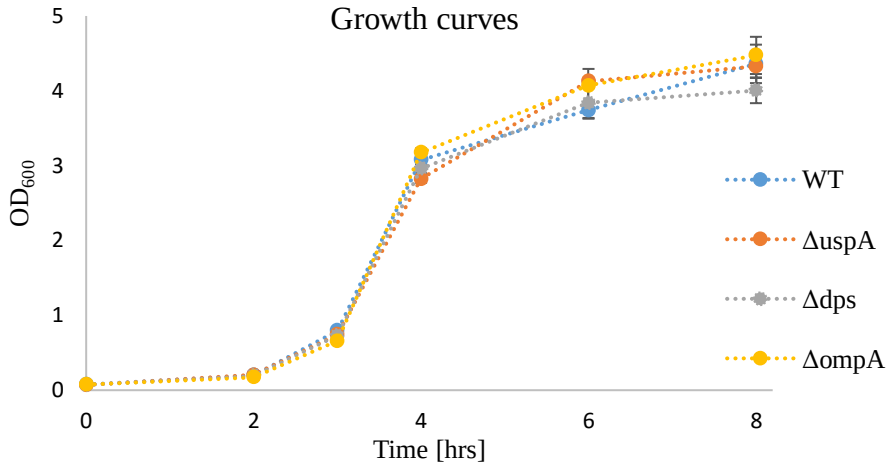
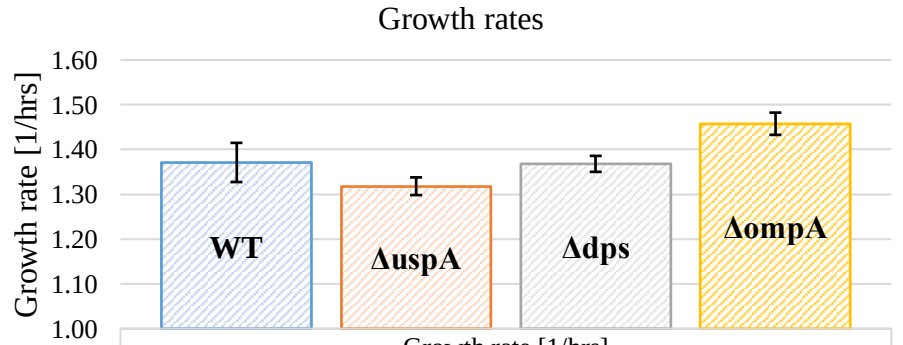


Figure 2.10: *Growth curves for KEIO strains with removed kanamycin resistance cassettes. Cells were grown in LB medium at 37 ℃, 225 RPM. Error bars are depicted as SDOM.*



	Growth rate [1/hrs]
WT	1.37
ΔuspA	1.32
Δdps	1.37
ΔompA	1.46

Figure 2.11: *Growth rates of KEIO strains with removed kanamycin resistance cassettes. Errors were calculated based on the SDOM values of the optical density data, depicted in Figure 2.10. See appendix B for further details about error analysis.*

One significant effect was observed in the data presented Figure 2.11: the *ΔompA* mutants showed significantly increased growth rates compared to wild type *E. coli BW25113*. The observed increase in growth rates was 6 ± 4 %. The statistical significance of this result was verified by conducting a two-way t-test with all data from wild type, - and *ΔompA* strains. In terms of growth characteristics, there was not observed any significant differences between *ΔuspA,* - *Δdps,* - and wild type cells.

2.4.2 *Δdps* and *ΔompA* strains showed increased expression from ribosomal promoters

The levels of ribosome expression in KEIO strains with removed kanamycin resistance genes were evaluated by using the fluorescence based reporter systems. Strains carrying the reporter plasmids were inoculated in LB medium, with ampicillin, for 8 hours. Relative fluorescence and cell densities were measured at the end of cultivations. The OD adjusted fluorescence data are depicted in Figure 2.12.

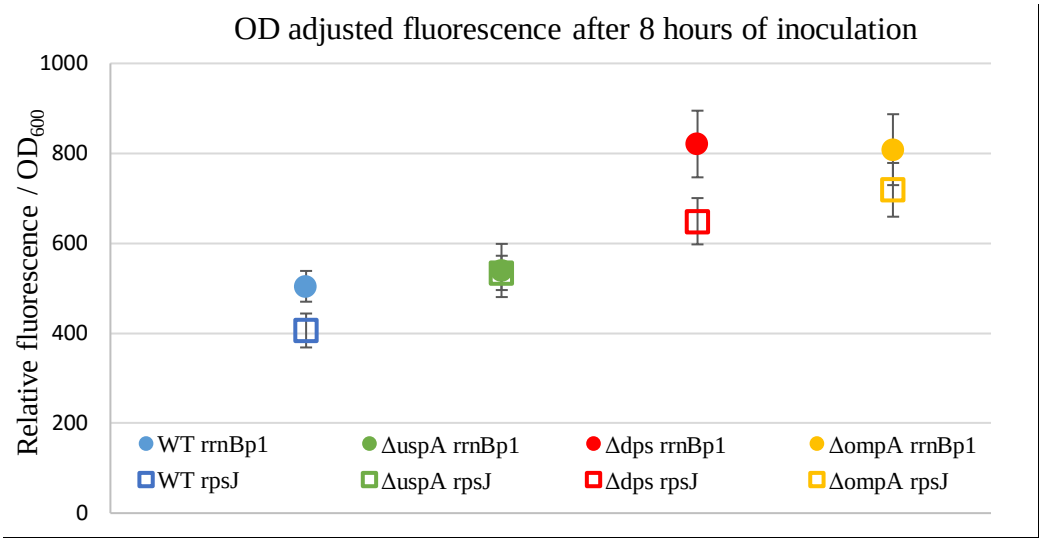


Figure 2.12: *levels of OD adjusted fluorescence from KEIO strains with removed kanamycin resistance cassette. Fluorescence from strains carrying the pMJ001 (rpsJ) plasmid are marked with squares, fluorescence from strains with the pMJ002 (rrnBp1) are marked with circles. Error bars are depicted as SDOM.*

Strains with removed kanamycin resistance cassettes showed similar results as strains from the KEIO collection. The *ΔompA* and *Δdps* strains achieved significantly higher fluorescence from both ribosomal promoters, compared to wild type cells. An ANOVA analysis and multiple t-tests revealed some other significant differences in the data:

1. OD adjusted fluorescence was significantly lower from rpsJ than rrnBp1, in *Δdps* and wild type strains.

2. OD adjusted fluorescence from rpsJ in *ΔuspA* strains was significantly higher, compared to OD adjusted fluorescence from rpsJ in wild type cells.

The *ΔompA* strain showed higher growth rates and increased level of ribosome expression, compared to wild type cells. This indicates that this strain had a higher translational capacity, and might be a more favourable host for recombinant protein production, compared to wild type *E. coli BW25113*.

Although an increase in growth rates was observed, the effect size was rather small, and just barely statistically significant. To engineer strains with significantly reduced proteome sizes it is most likely necessary to knock out multiple genes. CRISPR optimized MAGE recombineering (CRMAGE) might provide a convenient tool for multiplexed genome engineering in *E. coli* [Ronda et. al, 2016]. This novel method for generating targeted knockout mutations was investigated. All results regarding to the implementation of the CRMAGE system is described in section 2.5.

2.5 Generating knockouts with CRISPR optimized MAGE recombineering

CRISPR optimized MAGE recombineering (CRMAGE) was evaluated as a method to generate targeted knockout mutations in *E. coli BW25113*. During the CRMAGE protocol, mutagenic oligonucleotides (MAGE oligos) are used to generate mutations, while CRISPR/Cas9 are used for negative selection against non-mutated strains. The mechanism of CRMAGE is described in section 1.4.4. The CRMAGE system used in this work was previously tested and described by Ronda and colleagues [Ronda et. al, 2016]. The system featured three plasmids:

1. **pMA7CR_2.0**: Express λ-RED β-protein and Cas9. β-proteins are co-expressed with a dam methyltransferase gene (*dam*), which yield a *mutS* mutator phenotype. Cas9 is co-expressed with *recX*, which block the repair of double stranded breaks, caused by CRISPR/Cas9.

2. **pMAZ-SK**: Express gRNAs used to guide Cas9 to kill cells that was not successfully mutated during the MAGE cycle. Specific gRNA encoding regions were inserted, to design a specialized pMAZ-SK plasmid for negative selection for each of the different knockout mutations. This plasmid also expresses self-destruction gRNAs that guide Cas9 to cleave the plasmid's own backbone upon induction with L-rhamnose.

3. **pZS4Int-TetR**: Express a tetR repressor, necessary to closely regulate expression of Cas9 and sgRNA.

MAGE oligos were electroporated into cells as linear DNA.

CRMAGE was used to alter genes' initiation codons, from AUG to AGG. This was expected to cause functional knockouts by decreasing levels of gene expression by a factor of roughly 10,000 [Hecht et. al, 2017]. In addition, photo spacer adjacent motifs (PAMs) nearby the initiation codons were altered from NGG to NGN/NNG. PAMs are recognised by Cas9, and mutants with altered PAM sequences should not be killed by the CRISPR/Cas9 negative control mechanism [Ronda et. al, 2016].

In addition to *uspA*, *dps* and *ompA*, three other genes were targeted for knockout using CRMAGE:

- ***galK*** was knocked out to provide evidence that altering initiation codons in fact functionally stops gene expression. *galK* codes for a protein (galactokinase) which is vital for metabolism of D-galactose. It is easy to screen for loss of function in *galK*, by plating cells on MacConkey plates. On these plates, colonies capable of utilizing D-galactose will turn purple [Ronda et. al, 2016].

- The general stress response regulator, ***rpoS***, and the flagellar protein, ***flhD***, were knocked out. It has been reported that loss of expression of these genes yield increased growth rates of *E. coli BW25113* in a range of conditions [Price et. al, 2016]. *rpoS* regulates transcription of many stress related proteins. *flhD* is expected to constitute a large part of the *E. coli* proteome during exponential growth.

To target these six genes (*uspA*, *dps*, *ompA*, *galK rpoS* and *flhD*) for knockouts, specific MAGE oligos and gRNAs were designed.

2.5.1 MAGE oligos and gRNAs used to generate knockout mutations were designed *in silico*

MAGE oligos used to induce single nucleotide mutations, and gRNAs used for negative selection of non-mutated strains, were designed *in silico*. The general design procedure is described in section 5.8. The sequences of the MAGE oligos used to knock out different genes are presented in Table 2.3. gRNAs used for negative selection is also presented in Table 2.3. MAGE oligos and gRNAs to generate knockout mutation of the *galK* gene were not designed in this work, but were received from the authors of the original CRMAGE article [Ronda et. al, 2016]. In that work, this oligo and gRNA were successfully used to knock out the *galK* gene in *E. coli*, by altering the gene's initiation codon from AUG to AGG.

Table 2.3: *MAGE oligos and gRNAs used to induce knockout mutations using CRMAGE. Mutagenic nucleotides in MAGE oligos are highlighted. Off target scores describes the inverse probability of cas9 off-target binding and can range from 0-100.*

Gene	***uspA***	***dps***	***ompA***	***rpoS***	***flhD***	***galK***
Oligo strand	Forward	Reverse	Reverse	Reverse	Reverse	-
Mage oligo	CAGTCAT CGACAAC TTTATGT AAGGAGT AACACTA GGGCTTA TAAACAC ATTCTCA TCGCGGT CGACCTC	GGGACAT AACATCA AGAGGAT ATGAAAT TAGGAGT ACCGCTA AATTAGT TAAATCA AAAGCGA CAAATCT GC	ATGGCGT ATTTTGG ATGATAA CGAGGCG CAAAAAA GGAAAAA GACAGCT ATCGCGA TTGCAGT AGCACTG	GTTCCGT CAAGGGA TCACGGG TAGGAGC CACTTAG GAGTCAG AATACGC TAAAGTT CATGATT TAAAT	ATAAAAA TAAAGTT GGTTATT CTGGGTG GGAATAA GGCATAC ATCCGAG TTGCTGA AACACAT TTATGAC	GGCCGCG TGAATTT GATTGGT GAACACA CAGACTA GAACGAC GGTTTCG TTCTGCC CTGCGCG ATTGATT
gRNA strand	Forward	Forward	Reverse	Forward	Forward	-
gRNA	GAGCACG CCAGTCA TCGACAA CTTTAGT TTTAGAG CTAGAAA T	GAGCACG TTGCGGG TATAAAG CAGATGT TTTAGAG CTAGAAA T	GAGCACG ACAGCTA TCGCGAT TGCAGGT TTTAGAG CTAGAAA T	GAGCACG CGTATTC TGACTCA TAAGGGT TTTAGAG CTAGAAA T	GAGCACA ATGTGTT TCAGCAA CTCGGGT TTTAGAG CTAGAAA T	GAGCACA ACGAAAC CGTCGTT GTAGTGT TTTAGAG CTAGAAA T
Off-target score	84.3	70.7	94.1	83.3	82.8	-

All of the MAGE oligos in table 2.3 were expected to induce single point mutations (AUG →AGG) in the relevant gene's initiation codon. In addition, these oligos were designed to induce a NGG → NGN/NNG point mutation in a PAM motif close to the initiation codon. gRNAs were designed to guide Cas9 to cleave the DNA of cells that did not carry a mutation in the relevant PAM motif.

2.5.2 Specific gRNA encoding sequences, used for negative selection, were cloned into pMAZ-SK plasmid backbones

The CRMAGE system used in this work featured a negative-selection plasmid (pMAZ-SK). These plasmids carried sequences coding for guide RNAs (gRNAs) used to guide Cas9 to kill cells that were not mutated during the MAGE cycle. To construct systems able to knock out the different genes, the gRNA sequences listed in Table 2.3 were cloned into separate pMAZ-SK backbones, yielding one unique pMAZ-SK plasmid for each gene that was to be knocked out.

The pMAZ-SK plasmid, with gRNA for *galK* knockouts already inserted, was received from Ronda and colleagues [Ronda et. al, 2016]. This "pMAZ-SK::galK" plasmid was attempted to be reconstructed by inserting *galK* targeting gRNA sequences into 'empty' pMAZ-SK backbones. The verified correctly constructed pMAZ-SK::galK plasmid, received from Ronda and colleagues, were used as a positive control sample during colony PCR (Figure 2.13). The pMAZ-SK::galK plasmid constructed in this work was used for all further purposes.

The different pMAZ-SK plasmids were transformed into *E. coli BW25113*, and colony PCR was used to screen for strains carrying plasmids with correctly inserted gRNA sequences. The primers that were used to amplify the gRNA encoding regions are listed in table 5.3, Section 5.9.1. Primers were designed to yield equally sized PCR products for each of the different gRNA encoding regions. The PCR products were separated with gel electrophoresis. The resulting gel image is presented in Figure 2.13.

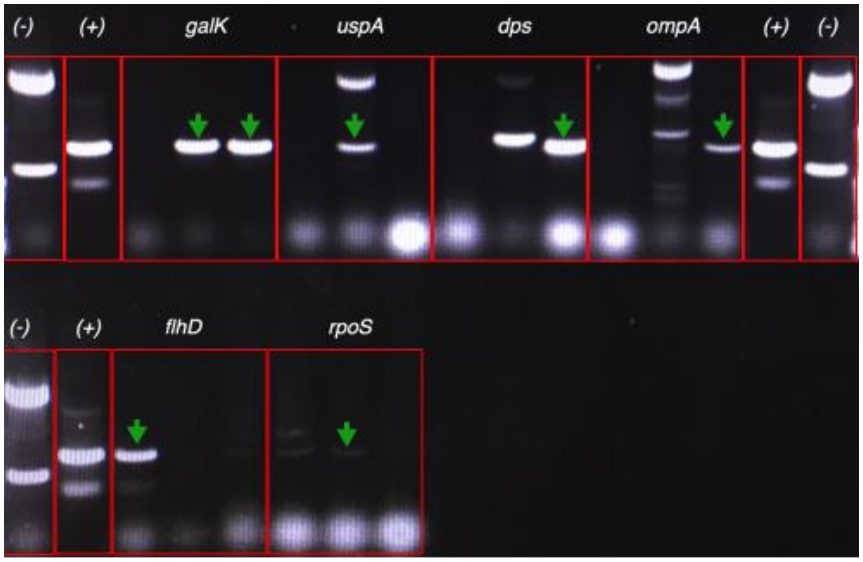


Figure 2.13: *Gel Image of PCR products from colony PCR. The text over each bracket annotates which pMAZ-SK plasmid the cells carried. e.g., the 'galK' bracket was loaded with PCR products of strains carrying pMAZ-SK plasmids designed to knock out the galK gene. pMAZ-SK plasmids without gRNA inserts were used as negative control (-). Purified pMAZ-SK plasmids with verified correctly inserted galK gRNA were used as positive control (+). Three separate colonies were tested for each strain. Green arrows highlights bands that indicated that gRNA coding sequences were successfully inserted in the plasmid backbones.*

The colony PCR results (Figure 2.13), indicated that the pMAZ-SK plasmids had been successfully constructed in at least one colony of each strain. These plasmids were cloned and purified.

2.5.3 One *flhD* knockout strain was isolated after the CRMAGE procedure

To generate targeted knockouts with CRMAGE, *E. coli BW25113* carrying two CRMAGE plasmids (pMA7CR_2.0 and pZS4Int_TetR) were inoculated in LB medium with appropriate antibiotics. The MAGE protocol was started by inducing expression of Dam methyltransferase and λ-RED β-proteins, to increase recombineering efficiencies. Specific MAGE oligos and appropriate pMAZ-SK plasmids were transformed into cells, to generate single gene knockouts. Negative selection with CRISPR/Cas9 was initiated by inducing expression of RecX, Cas9 and gRNAs. After the CRMAGE protocol, cells were plated on LB plates with appropriate antibiotics.

Cells in which the *galK* gene was targeted for knockouts were plated on MacConkey plates. On these plates, strains capable of metabolizing galactose will turn purple [Ronda et. al, 2016]. As *galK* encodes for a protein (Galactokinase) that is necessary for galactose catabolism, it was expected that *galK* knockout mutants would *not* form purple colonies when grown on MacConkey plates. A MacConkey plate with strains treated with CRMAGE to knock out the *galK* gene is depicted in figure 2.14.

Figure 2.14: *Picture of a MacConkey plate inoculated with E. coli BW25113 strains treated with CRMAGE to knock out the galK gene. Isolated colonies are highlighted with green circles.*

All colonies on the MacConkey plate (Figure 2.14) were clearly purple. *galK* knockout mutants were expected to form white colonies when inoculated on these plates. Hence, this result indicated that *galK* was *not* knocked out during the CRMAGE protocol.

Relevant regions of the genomes of strains treated with CRMAGE to knock out each of the different genes (*galK*, *uspA*, *dps*, *ompA*, *rpoS* and *flhD)* where sequenced, to screen for successful knockout mutants. 4 colonies were picked for each strain, yielding a total 24 samples (4 colonies * 6 strains). Out of these 24 samples 1 single gene knockout mutant was identified. This correspond to an average mutation efficiency of ~4%. The one single gene knockout mutant identified was a *ΔflhD* strain.

The CRMAGE protocol is expected to generate single gene knockouts with a ~98% efficiency [Ronda et. al, 2016]. Traditional MAGE recombineering generally induce mutations with a ~5% efficiency [Ronda et. al, 2016]. The ~4% efficiency (1 out of 24) observed in this work may suggest that the MAGE protocol of CRMAGE worked, but that negative selection with CRISPR/Cas9 was not achieved. One possible reason for the lack of negative selection is that gRNAs may not have been correctly inserted in pMAZ-SK plasmids. Without gRNAs guiding Cas9, the CRISPR/Cas9 negative selection system will be non-functional. This was investigated by sequencing all the different pMAZ-SK plasmids.

2.5.4 Two pMAZ-SK plasmids were correctly constructed

Of the 6 different pMAZ-SK plasmids that were constructed (one plasmid for each targeted knockout) 2 had correctly inserted gRNA sequences. The two correctly constructed plasmids were pMAZ-SK::flhD and pMAZ-SK::uspA.

These results might provide some explanation to why no strains with knockouts in *dps*, *ompA*, *rpos* or *galK* were identified. However, mutation efficiencies were still low (~13%) when only considering strains treated with correctly constructed pMAZ-SK plasmids. This suggest that incorrectly constructed pMAZ-SK was not the sole reason for the low observed mutation efficiencies. The design of MAGE oligos and gRNAs might have been sub-optimal for some genes. Moreover the CRMAGE protocol might need further optimization to be universally effective in knocking out genes in *E. coli*.

3 Discussion

3.1 The fluorescence based reporter systems was verified as valid reporters for ribosome *expression*

The fluorescence based reporter systems were characterized in wild type *E. coli BW25113*. Both in strains carrying the pMJ001 plasmid (rpsJ promoter), and in strains carrying the pMJ002 plasmid (rrnBp1 promoter), fluorescence was heavily correlated with growth rate. It was expected that the activity from ribosomal promoters should be growth rate dependent. The observation that levels of fluorescence indeed was correlated with growth rates indicated that this system had merit as a reporter for levels of ribosome expression. This notion was further tested by growing cells in sub-lethal levels of chloramphenicol. Chloramphenicol inhibits translation, increasing the need for ribosomes, to maintain levels of protein synthesis [Scott et. al, 2010]. It was expected to observe a negative correlation between growth rates and levels of ribosome expression, when growing cells in increasing concentrations of chloramphenicol. Such a negative correlation was observed, further confirming the fluorescence based reporter system as a valid reporter for ribosome expression in *E. coli*.

It is important to notice that outputs from the fluorescence reporter systems does *not* directly equate to intracellular ribosome concentrations. Mainly two factors contribute to this:

1. **The system does not give information about ribosome degradation.**
   Red fluorescence proteins might be degraded at different rates than ribosomes.

2. **Expression of red fluorescent protein (Rfp) might give overestimated information about ribosome synthesis.**
   Although Rfp expression is under control of the same promoters as ribosomal constituents they might be synthesized at different rates. This is most likely the case for the pMJ002 plasmid, where the rRNA promoter rrnBp1 controls Rfp transcription. One transcription event from this promoter should roughly equate to the synthesis of one ribosome. However, one transcription of the *mCherry* gene will most likely yield multiple rfp proteins as the transcripts can be translated several times before they are degraded. As in the case of the pMJ002 plasmid, fluorescence from the strains carrying the pMJ001 plasmids will likely tend to overestimate levels of ribosome expression. Translation of rProteins are negatively regulated by translational feedback inhibition, that will not affect rfp expression [Lemke et. al, 2011].

Considering these limitations, the fluorescence reporter systems were used to estimate ribosome *expression*, not *concentrations*. Fluorescence measurements were analysed with focus on identifying strains that showed significantly altered levels of fluorescence, rather than quantitatively comparing fluorescence from each strain.

Although having some limitations, the results of the characterization experiments (Figures 2.2, 2.3 and 2.5) verified the fluorescence based systems as valid tools for evaluating ribosome expression in *E. coli*. In addition to these systems, growth rates were used to assess translational capacities. The rationale behind this approach is described in section 1.7.2. It is largely accepted that there is a positive linear correlation between ribosome concentrations and growth rates in exponentially growing cells [Forchhammer & Lindahl 1971; Bremer & Dennis 1996; Scott et al 2010; Klumpp et. al, 2013].

3.2 How should fluorescence measurements and growth rates be interpreted to evaluate translational capacities?

The fluorescence based reporter systems give *qualitative* estimations of transcription of ribosomal constituents. This information was used to compare levels of ribosome *expression* in the different knockout strains. When interpreting the data from these reporter systems it is worth noticing that the source of fluorescence is a recombinantly expressed protein (Rfp). Hence, elevated fluorescence measurements correlates to an increased level of recombinant expression. This may indicate that strains showing higher fluorescence have preferable traits as a host for recombinant protein production. However, it is hard to evaluate how much of this effect should be credited to increase in ribosomal promoter activities, and how much that can be credited to an actual increase in translational capacities. In this thesis, observed increases in levels of fluorescence were first and foremost interpreted as indications towards increased activity from ribosomal promoters.

Exponential phase growth rates were used to evaluate translational capacities. Proteome duplication is the growth rate limiting factor in exponential phases [Molenaar et. al, 2009; Scott et. al, 2014]. Hence, translational capacities are most likely positively correlated with growth rates during exponential growth. Elevated growth rates were interpreted as an indication towards increased translational capacities in this thesis. With this approach, it is important to notice that growth rates are measured in a specific set of conditions. It is possible that some knock out mutations might increase translational capacities, without increasing growth rates.

Specific knockout mutations can make strains less optimized for growth *in some given conditions*, but it may happen that they would outgrow other strains in a different environment. Moreover, a fairly large increase in growth rates must be achieved before it can be deemed statistically significant. Based on the calculated standard deviations in the growth rate data obtained in this work, an increase in growth rates of at least 4% was necessary for effects to be deemed significant. Even though a knockout mutation does not yield a significant increase in growth rates, it may have resulted in a subtle increase in translational capacities.

Even though growth rates are calculated *quantitatively* it is hard to convert this to meaningful quantitative data, directly describing translational capacities. A modelling approach to quantitatively link growth rates and cellular ribosome fractions are presented in section 3.6. However, growth rate data was first and foremost interpreted *qualitatively* in this thesis.

Fluorescence from strains carrying reporter plasmids and exponential phase growth rates of strains without plasmid inserts were used in combination to evaluate translational capacities. Concurrent increases in growth rates and levels of fluorescence were interpreted as strong indications towards elevated translational capacities. However, an increase *in one or the other* should not be easily dismisses as an indication towards *no* increase in translational capacities. It is a chance that translational capacities may be increased without increasing growth rates or activities from ribosomal promoters. The different combinations of *qualitative* experimental results, and possible explanations are listed in Table 3.1. No negative effects (decrease in growth rates or levels of fluorescence) were observed in this work. Hence, possible combinations of results with negative effects is not accounted for.

An observed increase in growth rates *or* levels of fluorescence might indicate elevated translational capacities. However, only strains in which both growth and fluorescence were increased were accepted as mutants with increased translational capacity in this work. This is not to say that other strains did *not* have increased translational capacities, but that additional analysis is necessary to investigate these mutants further.

Table 3.1: *Interpretation of different combinations of qualitative experimental results. 'Effects' refer to the observed effects of knocking out a gene, by comparing single gene knockout mutants and wild type cells.*

Effect on fluorescence	Effect on growth rate	Interpreted as	explanation
None	None	No indication towards increased translational capacity.	No phenotypical effect observed.
Increase	None	Weak indication towards increased translational capacity	The increase in translational capacity might have been too subtle to significantly influence growth rates. On the other hand, increased ribosome expression might have gone at the expense of metabolic proteins. More ribosomes will not equate to increase translational capacity if each ribosome is less efficiently supplied with energy and amino acids.
None	Increase	Weak indication towards increased translational capacity	Although ribosome expression was unaltered, the cellular ribosome fraction might have increased if the induced knockout mutation reduced the overall size of the proteome.
Increase	Increase	Strong indication towards increased translational capacity	Increased growth rates and levels of ribosome expression strongly suggest elevated translational capacities.

3.3 Substituting genes with kanamycin resistance markers did not yield significant increases in translational capacities

Strains in which genes had been replaced with kanamycin resistance markers were received from the KEIO collection. Three different single-gene *substitution* mutants were tested for improved translational capacities. The key results from experiments with the "KEIO strains" are summarized in Table 3.2.

Table 3.2: *Key results from the experiments with single gene knockout strains from KEIO collection. Cells were grown in LB medium for all experiments. Effects of knocking out genes were evaluated by comparing knockout mutants and wild type E. coli BW25113.*

Gene knocked out	Effect exponential phase growth rate	Effect on fluorescence from rpsJ	Effect on fluorescence from rrnBp1
uspA	No effect	No effect	No effect
dps	No effect	Increased by 60±30 %	Increased by 80±30 %
ompA	No effect	Increased by 60±30 %	Increased by 80±30 %

There was not observed any differences in growth characteristics between the different KEIO strains and wild type *E. coli BW25113.* This is in accordance with the results described in the original report, describing the KEIO collection [Baba et. al, 2006]. In that work, none of the ~4000 single gene knockout mutants showed significantly increased growth rates compared to wild type *E. coli BW25113.* In strains from the KEIO collection, the removed genes were replaced with kanamycin resistance cassettes, yielding strains with approximately unaltered proteome sizes. This might explain why none of the strains from the KEIO collection showed increased growth rates compared to wild type cells.

Empirical evidence [Baba et. al, 2006; Price et. al, 2016] and modelling approaches [Klumpp et. al, 2013; Scott et. al, 2014] suggest that cell's proteomes should be decreased, to increase translational capacities. Since the strains from the KEIO collection had genes *replaced* rather than *removed*, it is unlikely that the sizes, or partitionings, of the proteomes were significantly altered. This might explain why none of the ~4000 single gene knockout mutants from the KEIO collection showed increased growth characteristics compared to wild type *E. coli BW25113.*

Two strains (*Δdps::kan* and *ΔompA::kan*) showed significantly higher activities from both the ribosomal promoters (rpsJ and rrnBp1), compared to wild type cells. This indicate increased levels of ribosome expression. Even though intracellular levels of ribosomes might have been increased in some strains, it does not automatically indicate that these strains had higher translational capacities. Translational capacity is not only dictated by ribosome concentrations, but also by how efficiently each ribosome is supplied with amino acids. It is possible that the increased levels of ribosome synthesis went at the expense of production of metabolic proteins. This may have yielded strains with higher ribosome concentrations, but wherein each ribosome was used less efficiently. It is also possible that increased levels of ribosome synthesis were counteracted by increased levels of degradation. Sub-efficient use of ribosomes is a known trigger for ppGpp synthesis in *E. coli* [Klumpp et. al, 2014]. ppGpp mediates an increase of synthesis of enzymes that catalyse ribosome degradation.

As discussed in section 3.1, it is likely that the fluorescence based reporter system will tend to overestimate changes in levels of ribosome synthesis. Although the *Δdps::kan* and *ΔompA::kan* strains showed up to 2.1 fold higher fluorescence than wild type cells, it is likely that the difference in levels of ribosome synthesis was smaller. However, each data point was an average value of three parallels, and the results were treated with strict statistical analyses. Hence, the results (Table 3.2) still provide strong evidence that knockouts of *dps* and *ompA* increased ribosome expression in *E. coli BW25113*.

Although showing elevated levels of ribosome expression, it seems that none of the KEIO strains achieved increased translational capacities. As discussed in section 3.2, it is possible that these strains had somewhat higher translational capacities without showing phenotypical growth effects. Further analysis can be done to investigate this, but in the case of this thesis these results were not interpreted to indicate a significant increase in translational capacities. Maximization of growth rates is an important fitness strategy for bacteria [Bosdriesz et. al, 2015]. It is likely that *E. coli* has evolved to optimally tune its level of ribosome synthesis, to maximize growth rates in different conditions. The results described in this section (and earlier work [Baba et. al, 2006]) suggest that efforts to increase the translational capacities in *E. coli* should not focus solely on increasing ribosome synthesis. Rather, one could try to reduce the part of the proteome that contains genes not involved in protein synthesis. By reducing the non-essential part of the proteome, energy can be more efficiently focused towards growth or recombinant protein production.

3.4 Deletion of the *ompA* gene resulted in increased translational capacities in *E. coli BW25113*

The kanamycin resistance marker in the KEIO strains was flanked by FLP (*flp recombinase)* recognition sites. By expressing *flp recombinase* on a curable plasmid, the kanamycin resistance cassettes were removed from the KEIO strains. Resulting in a new set of knockout mutants, with gene *deletions*, rather than gene *substitutions.* Some differences were observed between KEIO strains with removed kanamycin resistance genes, and wild type *E. coli BW25113.* All knockout strains (*ΔompA*, *Δdps* and *ΔuspA)* showed increased activity from the rpsJ promoter, compared to wild type cells. The *ompA* knockout strain showed an increase in exponential phase growth rate of 6±4 %, compared to wild type cells. Key results are summarized in Table 3.3.

Table 3.3: *Key results from experiments with KEIO strains with removed kanamycin resistance cassettes. Effects of knocking out genes were evaluated by comparing knockout mutants and wild type E. coli BW25113.*

Gene knocked out	Effect on exponential phase growth rate	Effect on fluorescence from rpsJ	Effect on fluorescence from rrnBp1
uspA	No effect	Increased by 30±20 %	No effect
dps	No effect	Increased by 60±30 %	Increased by 60±30 %
ompA	Increased by 6±4 %	Increased by 80±30 %	Increased by 60±30 %

The *ΔompA* strain showed higher growth rates and increased level of ribosome expression, compared to wild type cells. This indicates that this strain had a higher translational capacity, and might be a more favourable host for recombinant protein production, compared to wild type *E. coli BW25113.* It is important to notice that the fractional insecurity of this increase was very high (67%). The calculated p-value (probability that the effect was random) from the T-test was 4.8%. This is barely within the 5% confidence level used in this work. However, the observed increase remained significant even when the data was screened for outliers using Chauvenet's criterion (Appendix B). This indicate that the observed increase in growth rates of *ΔompA* strains was *not* due to random errors.

ompA is an abundant protein in *E. coli* during exponential growth phases [Smith et. al, 2007]. The *ompA*-substitution (*ΔompA::kan)* introduced in the KEIO strains did not yield an increase in growth rates. Hence, it is reason to believe that the *ΔompA* strain showed increased growth rates because it had a leaner proteome, compared to wild type *E. coli BW25113.*

Neither of the *Δdps* or *ΔuspA* strains showed increased growth rates. dps and uspA are generally present in small concentrations during exponential growth [Kvint et. al, 2003; Calhoun & Kwon, 2011]. Although expression of dps and uspA might be upregulated during stress responses and entry to stationary phase, knockouts of these genes did not appear to create a significantly leaner proteome. Neither of these strains had longer exponential phases, compared to wild type cells. Nor did they achieve higher cell densities.

These results support the notion that genes with large translational burdens during exponential growth should be targeted for knockouts, to increase translational capacities [Valgepea et. al, 2015; Price et. al, 2016]. Knocking out the stress response related genes *dps* and *uspA* did not seem to increase translational capacities. However, only two out of several stress response related genes were knocked out in this work. It is possible that knocking out other stress response related genes can increase translational capacities. Price and colleagues observed that a knockout of the general stress response regulator (*rpoS)* yielded increased fitness of *E. coli BW25113* in a range of conditions [Price et. al, 2016].

3.5 CRISPR optimized MAGE recombineering failed to effectively generate knockout mutations.

A novel method developed by Ronda and colleagues [Ronda et. al, 2016] was investigated to generate targeted knockouts in *E. coli BW25113*. The method utilises traditional MAGE recombineering to generate mutations, while CRISPR/Cas9 is used for negative selection against non-mutated cells. This procedure was named CRISPR optimized MAGE recombineering (CRMAGE). By using CRMAGE Ronda and colleagues generated targeted knockouts in *E. coli* with a ~98% efficiency. Genes were knocked out by altering the genes' initiation codons from ATG to AGG. This is expected to cause functional knockouts by decreasing gene expression by a factor of ~10,000 [Heucht et. al, 2017].

A similar CRMAGE approach to the one described by Ronda and colleagues were used to target 6 different genes for knockout (*uspA*, *ompA*, *dps*, *galK*, *rpoS* and *flhD*). However, sequencing revealed that only the *flhD* gene was successfully knocked out. Moreover only 1 out of 4 sampled *ΔflhD* strains carried the desired knockout mutation. As only 1 out of 24 samples (4 samples per targeted gene * 6 targeted genes) showed the desired knockout mutation, the general efficiency mutation efficiency was ~4%.

The most probable explanation for the low mutation efficiency is that MAGE recombineering was successfully implemented, but that negative selection with CRISPR/Cas9 was not achieved. The pMAZ-SK plasmids are essential for negative selection during the CRMAGE procedure. These plasmids express gRNAs used to guide Cas9 to cleave the DNA of non-mutated cells. Unique pMAZ-SK plasmids with specific gRNAs must be made for each gene that is targeted for knockouts. This was done by cloning gRNA encoding oligonucleotides into pMAZ-SK backbones, received from Ronda and colleagues.

Sequencing revealed that only two pMAZ-SK plasmids were correctly (gRNA inserted in backbone). This were the pMAZ-SK::flhD and pMAZ-SK::uspA plasmids. This provided a possible explanation towards the low mutation efficiencies. However, when only accounting for strains treated with correctly constructed pMAZ-SK plasmids, mutation efficiencies were no still low (11% - 1 out of 8 samples). This suggest that incorrectly constructed pMAZ-SK plasmids were not the only factor decreasing mutation efficiencies.

It is possible that the MAGE oligos and/or gRNAs used to knock out the *flhD* and *uspA* genes were poorly designed. It is also possible that the CRMAGE protocol were conducted in a sub-optimal manner. CRMAGE is a novel method and some optimization might be necessary for it to be universally effective to knock out genes in *E. coli.*

3.6 A modelling approach to compare ribosome fractions in different strains

In this work, exponential phase growth rates were used as a reporter for translational capacities. This approach is based on the assumption that the rate limiting factor of exponential growth is protein synthesis [Molenaar et. al, 2009; Scott et. al, 2014]. In other words; the time between division events is dependent on the time it takes to duplicate the entire cellular proteome. Ribosome concentrations is the main determinant for rates of translation in bacteria, and it is generally accepted that there is a positive linear correlation between ribosome concentrations and growth rates in exponentially growing cells [Forchhammer & Lindahl 1971; Bremer & Dennis 1996; Scott et al 2010; Klumpp et. al, 2013]. Having some notion of the slope of this correlation might be beneficial for a range of applications. Estimating the slope of the correlation allows for *quantitative* estimations of ribosome concentrations, based on observed growth rates.

3.6.1 Justification of the need to *quantitatively* estimate ribosome concentrations

Even though a linear correlation between growth rates and ribosome concentrations exist, evaluating growth rates might be a sub-optimal way of comparing translational capacities. When comparing growth rates one does not take the steepness of the linear correlation between growth rates and ribosome concentrations into account. If the slope of the linear correlation is steep, an increase in growth rates might tend to overestimate the increase in ribosome concentrations. And vice versa for a moderate slope. This phenomenon is highlighted in Figure 3.1.

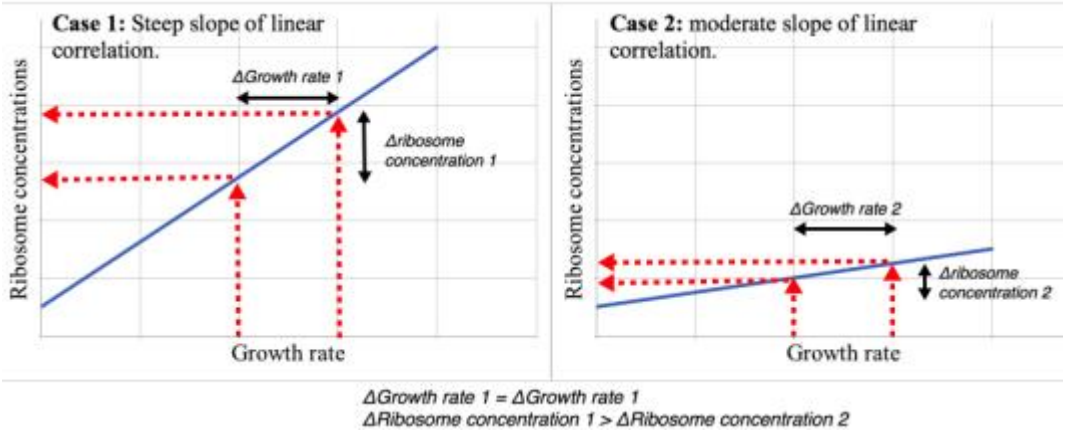


Figure 3.1: *Effects of slope steepness of the linear correlation between growth rate and ribosome concentrations.*

If not considering the slope of the linear correlation, one can only estimate changes in ribosome levels *qualitatively* based on *quantitative* growth rate inputs.

Growth rates might be a more direct phenotypical measure for translational capacities, compared to intracellular ribosome levels. Especially since rates of translation is dependent on both ribosome concentrations, and how effectively each ribosome is used. However, having some quantitative estimations of ribosome concentrations might be of interest. Ribosome levels might be a more intuitive way to describe to translational capacities for some applications. Moreover, having some quantitative estimations of ribosome concentrations makes for easier comparisons with earlier published data, in which intracellular RNA/protein ratio is a popular reporter for ribosome concentrations and translational capacities [Forchhammer & Lindahl 1971; Bremer & Dennis 1996; Scott et al 2010; Klumpp et. al, 2013].

3.6.2 Deduction of a model that estimates cellular ribosome fractions, based on exponential phase growth rates

To get some quantitative estimations of cellular ribosome levels a modelling approach was utilised. This approach was based on a model developed by Matthew Scott and colleagues [Scott et. al, 2014]. Based on this model, an equation to compare growth rates of two strains, and generate a comparison between the strains' intracellular ribosome concentrations was derived. The aim of the LEANPROT project (in which this work is a part of) is to create strains with reduced proteomes. Hence outputs were designed to be on the basis of ribosome fractions of entire proteome, rather than intracellular concentrations. Earlier works have identified ribosome *fractions* as the most relevant value to evaluate translational capacities in bacteria [Klumpp et. al, 2013; Scott et. al, 2014; Bosdriesz et. al, 2015]. Some key assumptions are important for the validity of the model:

- **Ribosome concentrations are optimally regulated to maximise growth rates.** The linear correlation between growth rates and ribosome concentrations is only valid if each ribosome translates at constant rates. Earlier work has revealed that ribosome levels in *E. coli* are optimally regulated to maximize growth rates [Bosdriesz et. al, 2015]. If ribosome levels are regulated sub-optimally, this model will tend to underestimate ribosome concentrations.

- **Intracellular RNA/protein fractions is a valid reporter for ribosome concentrations.** This is a popular reporter for ribosome concentrations, used in a large range of earlier work [Forchhammer & Lindahl 1971 ; Bremer & Dennis 1996; Scott et al 2010; Klumpp et. al, 2013]. The quantitative empirical data implemented in this model is based on earlier work using this reporter system.

The deduction of the following model was done by Matthew Scott and colleagues [Scott et. al, 2014]. In this thesis, the equation proposed by Scott and colleagues (equation 5) was rewritten to compare ribosome fractions of two strains. Additionally, empirical parameters were estimated based on earlier published data.

For cells to double, levels of all cellular constituents (including proteins) must also double [Scott et. al, 2014]. During stable exponential growth, there is not net accumulation of proteins, and daughter cells are indistinguishable from mother cells. In this growth phase, the entire cellular protein content increase at the same rate (the growth rate). By neglecting protein turnover, the protein mass (M) accumulation leading to a division event can be written as:

$$\frac{dM}{dt} = \lambda M \qquad (1)$$

M = Total protein mass
λ = growth rate

Protein synthesis is conducted by a pool of active ribosomes (N_R^{act}) translating at an average rate "k" per ribosome. Hence, equation 1 can be formulated as:

$$\frac{dM}{dt} = \lambda M = kN_R^{act} \qquad (2)$$

k = rate of translation per actively translating ribosome (rate of elongation)
N_R^{act} = Number of actively translating ribosomes

Not all ribosomes are actively translating, and a subpopulation of ribosomes (N_R^{in}) will not be participating in protein synthesis. Ribosomes being recycled or paused during elongation constitutes the N_R^{in} subpopulation. As we can define ribosomes as either active or inactive, the number of active ribosomes can be written as:

$$N_R^{act} = N_R - N_R^{in} \qquad (3)$$

N_R = Total number of ribosomes
N_R^{in} = Number of inactive ribosomes

Combining equation 2 and 3 yields:

$$\lambda M = k(N_R - N_R^{in}) \qquad (4)$$

The total mass of the ribosome pool (M_R) can be expressed as the number of ribosomes (N_R) times the weight of each ribosome (m_R). Dividing each side of equation 4 by the total protein weight (M) yields a model connecting growth rate and ribosome fractions, as described by Matthew Scott et. al [Scott et. al, 2014]:

$$\lambda = \Upsilon\left(\phi_R - \phi_R^{in}\right) \qquad (5)$$

Υ = k/m_r

ϕ_R = Fraction of ribosomal proteins of entire proteome (g/g)

ϕ_R^{in} = Fraction of inactive ribosomes

The parameters ϕ_R^{in} and Υ are host specific, and are largely independent of growth medium [Scott et. al, 2014]. If protein synthesis is not inhibited (by for example antibiotics), empirical evidence suggests that both ϕ_R^{min} and Υ remain constant as growth rate varies [Bremer & Dennis, 1996; Scott et. al, 2010; Klumpp et. al, 2013]. Empirical evidence of this notion is presented in Figure 3.2, which features data from three separate studies.

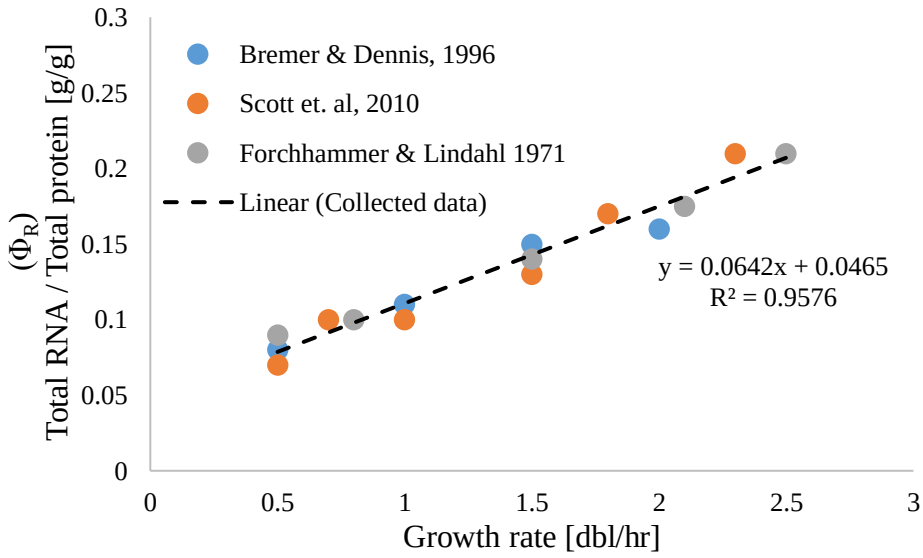


Figure 3.2: *Data from three separate studies, highlighting the linear relationship between ribosome fractions (measured as total RNA/total protein) and growth rates during exponential growth. Obtained from [Klumpp et. al, 2013]. All experiments were conducted with different strains of E. coli: Bremer & Dennis 1996 – E. coli b/r; Forchhammer & Lindahl 1971 – E. coli 15; Scott et al 2010 – E. coli EQ3. Growth rates were controlled by varying the growth media compositions.*

The empirical parameters ϕ_R^{in} and Υ have not been empirically estimated for the *E. coli* strain investigated in this work (*E. coli BW25113*). However, the three data sets, depicted in Figure 3.2, show remarkable linear correlation with each other ($R^2 \approx 0{,}96$ for linear regression through all data sets). Even though the experiments were conducted with different *E. coli* strains, in different labs. This suggests that the linear

regression through the three data sets in Figure 3.2 can provide an appreciably good estimation of ϕ_R^{in} and Υ, also for *E. coli BW25113.*

By interpreting the linear regression in Figure 3.2 with equation 5, one obtains that $\phi_R^{in} \approx 0{,}05$ and $\gamma \approx 0{,}06$ dbl/hr = 10.8hrs^{-1}. Although this might be a somewhat rough estimate, it was deemed good enough for the purposes of this work. The goal in this work was to compare strains, not to exactly pinpoint the intracellular ribosome levels in each strain. Having obtained the empirical values, the maximum ribosome fraction of two strains can be easily compared:

$$\frac{\Phi_R^1}{\Phi_R^2} = \frac{\frac{\lambda_1}{\gamma} + \Phi_R^{in}}{\frac{\lambda_2}{\gamma} + \Phi_R^{in}} = \frac{\frac{\lambda_1}{10.8hrs^{-1}} + 0.05}{\frac{\lambda_2}{10.8hrs^{-1}} + 0.05} \qquad (6)$$

Φ_R^i = Ribosome fraction of entire proteome for strain "i"

λ_i = Measured growth rate for strain "i"

Φ_R^{in} = Fraction inactive ribosomes (empirical constant)

γ = Elongation rate per active ribosome (empirical constant)

By using equation 6, ribosome fractions in different knockout strains can be compared, simply by inputting growth rates. Employing this method not only gives information about which strain produce more ribosomes, but also estimate *quantitative* differences in ribosome levels between strains.

3.6.3 Results of implementing the model on the growth data observed in this work.

The growth data from all experiments was treated with the model described in equation 6. The resulting outputs are listed in table 3.4. Effects of knocking out genes were estimated by comparing wild type *E. coli BW25113* and single gene knockout mutants.

Table 3.4: *Estimated ribosome fractions in different single gene knockout strains of E. coli BW25113. Ribosome fractions were estimated using equation 6. Cells were grown in LB medium for all experiments. Effects of knocking out genes were evaluated by comparing knockout mutants and wild type E. coli BW25113.*

Knockout method	**Gene knocked out**	**Estimated effect on ribosome fraction of entire proteome [g/g]**
Gene substitution (Strains from KEIO collection)	*uspA*	No effect
	dps	No effect
	ompA	No effect
Gene deletion (Keio strains with removed kanamycin resistance markers)	*uspA*	No effect
	dps	No effect
	ompA	Increased by 5±4 %

As for the growth rate data (Tables 3.2 and 3.3) the only significant effect was achieved by generating a gene *deletion* mutation of the *ompA* gene. The effects of single gene knockouts on cellular ribosome fractions were evaluated based on growth rates. Hence, it was expected that the ribosome fraction effects would have the same levels of significance as the growth rate effects (Table 3.2 and 3.3). Levels of significance was not altered (No insignificant effects were deemed significant, or vice versa) when treating growth rate data with equation 6. This indicate that the model was not prone to generate statistical errors. All error analysis related to the model is described in appendix B.

Based on a 6±4% increase in exponential phase growth rates, the model estimated that the *ΔompA* strain achieved a 5±4% higher cellular ribosome fraction, compared to wild type *E. coli BW25113*. Having this kind of a tool to estimate cellular ribosome fractions might yield more intuitive outputs to evaluate translational capacities. Moreover, it might make for easier data comparisons with other studies, using cellular RNA/protein fractions to estimate ribosome fractions.

4 Conclusions and future perspectives

How, and to what degree, the outlined goals of this project was fulfilled are summarized in this section. The main objectives of this work were to:

1. Implement and characterize reporter systems to evaluate translational capacities in *E. coli*.

2. Investigate what kinds of genes that should be knocked out, to increase translational capacities in *E. coli.*

3. Test genome engineering techniques that may be used to knock out genes from the *E. coli* chromosome. And to evaluate if knocking out the same genes, with different techniques, yield different phenotypes.

4.1 A fluorescence based reporter system was established, and used in combination with growth rates to evaluate translational capacities.

A fluorescence based reporter system for transcription of ribosomal constituents were established and characterized (Sections 1.7.1, 2.1 and 3.1). The system was verified as a valid reporter for levels of ribosome expression in *E. coli.* Due to the natural limitations of such systems it did not yield valid information about intracellular ribosome *concentrations.* Moreover, the system could only produce *qualitative* outputs because the correlation between levels of fluorescence and actual ribosome synthesis was not known.

To alleviate some of these limitations, the fluorescence based reporter system was used together with exponential phase growth rates to evaluate strains' translational capacities. It is generally accepted that there is a positive linear correlation between ribosome concentrations and growth rates in exponentially growing cells [Forchhammer & Lindahl 1971; Bremer & Dennis 1996; Scott et al 2010; Klumpp et. al, 2013]. Growth rates and outputs from the fluorescence based reporter system were used in combination to compare translational capacities in the different *E. coli* knockout mutants.

To make for easier comparisons with earlier published data, a model which quantitatively linked growth rates and intracellular ribosome fractions was suggested (Section 3.6).

4.2 Non-essential genes that are highly expressed during exponential growth phases should be targeted for knockouts, to increase translational capacities.

Three genes were knocked out from the *E. coli* chromosome: *dps, uspA* and *ompA.* Knockout of the stress response related genes, *dps* and *uspA*, did not yield strains with increased translational capacities. However, a deletion of the *ompA* gene generated strains with seemingly increased translational capacities (Sections 2.4 and 3.4). OmpA is a highly expressed protein in *E. coli* during exponential growth [Smith et. al, 2007]. It was postulated that *ΔompA* strains showed increased translational capacities because they had leaner proteomes, compared to wild type *E. coli BW25113.*

Only 3 genes were knocked out during this work, which may be too little to draw any final conclusions. However, the notion that genes with large translational burdens should be knocked out to increase translational capacities have been supported in earlier work: Studies have shown that the specific growth rate in *Bacillus subtilis* was increased by ~30% when the exponential phase proteome size was reduced by ~9% [Fischer & Sauer, 2005; Muntel et. al, 2014]. Another study suggested that deleting genes with higher protein expression cost led to greater growth advantage in *E. coli* [D´Souza et. al, 2014]. Price and colleagues estimated that non-essential protein production reduce growth rates in *E. coli* by more than 13% [Price et. al, 2016].

These results, and the ones obtained in this work, indicate that efforts to increase translational capacities in *E. coli* should focus on knocking out genes that are highly expressed during exponential growth. By knocking out non-essential genes with high translational burdens, energy can be more efficiently utilised towards growth and recombinant protein production.

4.3 Gene *deletions* should be generated to increase translational capacities.

Both gene *deletion* mutations and gene *substitution* mutations were generated in this work. In strains from the KEIO collection, genes were replaced with kanamycin resistance markers. None of the tested single gene knockout strains from the KEIO collection (*ΔuspA::kan*, *Δdps::kan* and *ΔompA::kan*) showed increased translational capacities (Sections 2.3 and 3.3). This was in accordance with the findings in the original report, describing the KEIO collection [Baba et. al, 2006]. In that work, none of the ~4000 single gene knockout mutants showed significantly increased growth rates compared to wild type *E. coli BW25113*.

The only observed increase in growth rates came from strains with *ompA deletions*. This is in accordance with earlier work, suggesting that expression of genes should be stopped all together to increase growth rates [Muntel et. al. 2014, D´Souza et. al, 2014, Price et. al, 2016]. These results suggest that gene *deletions*, not *substitutions*, should be generated to increase translational capacities in *E. coli*. This makes intuitive sense, as replacing a gene with another will not reduce proteome sizes or free up translational capacities.

Two possible ways of generating gene deletions were proposed in this work:

1. Use traditional lambda red recombineering to remove genes from the chromosome (Section 1.4.1)
2. Use CRISPR optimized MAGE recombineering (CRMAGE) to stop translation of genes by altering initiation codons from ATG to AGG (Section 1.4.4).

Although the CRMAGE approach failed to generate targeted knockouts in this work, it seems to be a promising technique going forward. The method has successfully been implemented to generate targeted knockouts in earlier work [Ronda et. al, 2016]. CRMAGE have the advantage over traditional lambda red recombineering in that the procedure allows for easier combination of knockout mutations [Ronda et. al, 2016]. Moreover, CRMAGE knockouts do not leave scars in the chromosome, like the lambda red recombineering approach does. To generate strains with significantly reduced proteomes, multiple knockout mutations should be generated. CRMAGE can provide a powerful method for generating multiple knockouts in *E. coli*, and further work should be invested in alleviating the problems that caused the CRMAGE approach to fail in this work.

5 Material and methods

5.1 Bacterial strain used

E. coli BW25113 was used for all purposes in this work. This particular strain was chosen because it was used to make the KEIO collection of single gene mutants [Baba et. al, 2006]. *E. coli BW25113* is a derivative of the *E. coli K-12* family. The genotype of *E. coli BW25113* is Δ(araD-araB)567, ΔlacZ4787::rrnB-3, LAM-, rph-1, Δ(rhaD-rhaB)568, hsdR514.

5.2 Working concentrations of antibiotics

The working concentrations of different antibiotics are listed in Table 5.1. When not otherwise stated, these concentrations were used for all medias and plates containing antibiotics.

Table 5.1: *Working concentrations of antibiotics*

Antibiotic	Working concentration [µg/ml]
Ampicillin	100
Kanamycin	50
Spectinomycin	100

5.3 Preparation of electrocompetent cells

The same protocol was used to make all strains electrocompetent. Cells were picked from frozen stock and inoculated over night in LB medium at 37°C, 225RPM. 1ml of each overnight culture was transferred to 100ml fresh LB medium. The cultures were grown to OD_{600} about 0.5, and chilled on ice for 20min. After 20 minutes on ice, the cultures were centrifuged for 15min, 4°C at 4000g. The supernatants were discarded and the pellets were resuspended in 100ml 10% glycerol. The centrifugation step was repeated 3 times, resuspending the pellets in progressively less volume of 10% glycerol (50ml, 10ml, 1ml). The cultures were kept on ice between all centrifugations. The final 1ml cell suspension in 10% glycerol was aliquoted into Eppendorf tubes (20µl in each tube) and stored at -80°C.

5.4 Electroporating plasmids into the different strains

The gene pulser Xcell™ electroporation system from biorad was used for all electroporations. Electro competent cells were thawed on ice. About 100ng plasmid DNA was added to a 20µl cell suspension with electro competent cells in 10% glycerol. The cell suspension was transferred to a 1mm electroporation cuvette, and electroporated at a voltage of 5kV, with a 1800ms response time. Immediately after electroporation, the cell suspension was transferred to 1ml SOC medium and incubated for 1hr at 37°C, 225RPM. After 1 hour of incubation, 50µl of cell suspension were plated out on LB plates with appropriate antibiotics and grown over night at 37°C (30°C for cells carrying the pCP20 plasmid). The following day, colonies were picked and inoculated over night in 10ml LB medium with the appropriate antibiotic. 500µl of overnight culture was added to 500µl of 50% glycerol and stored at -80°C.

5.5 Handling of strains from KEIO collection

The KEIO strains were received in LB medium suspensions, in Eppendorf tubes. 100µl of cell suspension, of each culture, was plated out on LB plates. Colonies were picked and inoculated in 50ml LB medium over night at 37°C, 225RPM. 500µl of the overnight cultures were transferred to tubes with 500µl 50% glycerol solution, and stored at -80°C.

5.6 Removing the kanamycin resistance gene from KEIO strains using FLP

The kanamycin resistance cassette was removed by following the protocol described by Datsenko and Wanner [Datsenko & Wanner, 2000]. A plasmid (pCP20) carrying the yeast Flp recombinase gene, an ampicillin resistance gene (*bla*), and a temperature sensitive replicon was electroporated into the KEIO strains (Electroporation protocol – section 5.4). As the pCP20 plasmid is heat curable, the electroporated strains were grown at 30°C over night. The following day, colonies were picked from plates and inoculated overnight in 5ml LB medium at 43°C, to induce FLP recombination. 50µl of 10^3 dilutions of the overnight cultures was plated on LB plates without antibiotics and inoculated over night at 30°C.

Screening for loss kanamycin resistance and pCP20 plasmid was done by patching colonies on LB plates with 50µg/ml kanamycin, LB plates with 100µg/ml ampicillin and LB plates without antibiotics. The plates were at 30°C for 24hours. Colonies that

showed growth only on LB plates without antibiotics was considered to be mutants with successfully removed kanamycin resistance cassette and cured pCP20 plasmid.

5.7 Generating knockouts with CRMAGE

The general mechanics of the CRMAGE system is described in section 1.4.4. The CRMAGE system used in this work consisted of three different plasmids. The system was previously tested and described by Ronda and colleagues [Ronda et. al, 2016].

4. **pMA7CR_2.0**: Express λ-RED β-protein and Cas9. β-proteins were co-expressed with a dam methyltransferase gene (*Dam*), which yield a *mutS* mutator phenotype. Cas9 was co-expressed with *recX*, which block the repair of double stranded breaks, caused by CRISPR/Cas9. The pMA7CR_2.0 plasmid carry an ampicillin resistance marker.

5. **pMAZ-SK**: Express gRNAs used to guide Cas9 to kill cells that was not successfully mutated during the MAGE cycle. The gRNA encoding region was changed, using USER cloning, to provide negative selection against different non-mutated genotypes. This plasmid also expresses self-destruction gRNAs that guide Cas9 to cleave the plasmids own backbone upon induction with L-rhamnose. The pMAZ-SK plasmid carry a kanamycin resistance marker.

6. **pZS4Int-TetR**: Express a tetR repressor, necessary to closely regulate expression of Cas9 and sgRNA. The pZS4Int-TetR plasmid carry a spectinomycin resistance marker.

Oligonucleotides used to induce mutations were electroporated into the cells as linear DNA. The genetic elements used during the CRMAGE protocol were under the control of different promoters. An overview of the promoters, and what genetic elements they controlled, is listed in table 5.2.

Table 5.2: *Promoters used to control expression of the different genetic elements used during the CRMAGE protocol.*

Promoter	Inducer	Elements controlled
pLtet	aTetracycline	RecX, Cas9 and gRNA
pBAD	L-arabinose	Dam methyltransferase and λ-RED β-proteins
pRham	L-rhamnose	Self-destruction gRNA

To induce knockout mutations using CRMAGE, mutagenic ssDNA oligonucleotides and sgRNA were first designed *in silico* (Section 5.8). pMAZ-SK plasmids were prepared by cloning in the appropriate gRNA (Section 5.9). The pMAZ-SK::uspA plasmid was created by cloning in the gRNA used for negative selection of *uspA* knockout mutants, and so on.

To start the CRMAGE protocol, 15 ml LB medium, with 100µg/ml ampicillin and 100µg/ml spectinomycin, was inoculated with 150µl overnight culture with *E. coli BW25113* carrying both the pZS4Int-TetR, - and the pMA7CR_2.0 plasmid (Electroporation protocol – section 5.8). These cultures were grown at 37°C until OD_{600} was in the range 0.5-06. L-arabinose was added to a concentration of 0.2%, and the cultures were grown for an additional 15 minutes, before being cooled in an ice-water bath for 20 minutes. Cultures were centrifuged at 6500g at 4°C for 5 minutes. Pellets were resuspended in 35ml ice-cold water and the centrifugation was repeated 3 times, resuspending in decreasing volumes of ice-cold water (15ml – 1ml - 500µl).

50µl of cells suspended in water was added to an oligo/plasmid mix, containing 0,5µl of ssDNA oligos at a concentration of 10pmol/µl and 250ng of the appropriate pMAZ-SK plasmid. The oligos/plasmid mix was electroporated into the cells at 1.8kV in a 1mm gap cuvette. 950µl of LB medium with 100µg/ml ampicillin and 100µg/ml spectinomycin was added immediately after electroporation, and cells were grown for 1 hour at 37°C, 225RPM. Kanamycin was added to a concentration of 50µg/ml, and the cells were grown at 37°C, 225RPM for subsequently 2 hours. After 3 hours of inoculation (2+1) aTetracycline was added to a concentration of 200ng/ml, and the cells were subsequently grown for 3 hours, before plating on selective media.

The pMAZ-SK plasmids were cured by first washing cells (centrifuging and resuspending) with LB. Cells were resuspended in 1ml LB medium with 100µg/ml ampicillin, 100µg/ml spectinomycin and 200ng/ml aTetracycline. L-rhamnose was

added to a concentration of 0.2% and cells were grown over night. To cure the other plasmids, cells were grown for about 50 generations in LB medium without antibiotics.

5.8 Designing oligonucleotides for CRMAGE

Oligonucleotides (oligos) used to generate knockout mutations, using CRMAGE, were designed *in silico*. The same general workflow was used to design oligos to generate all three knockout mutations (*ompA*, *dps* and *uspA*).

The whole genome of *E. coli BW25113* was downloaded to the open source software "Benchling" (www.benchling.com), from the NCBI database [NCBI database]. To determine what kind of knockout mutation should be generated, the genetic context of the gene was checked, using the EcoCyc database (www.ecocyc.org). The genetic context of the *uspA* gene is depicted in Figure 5.1.

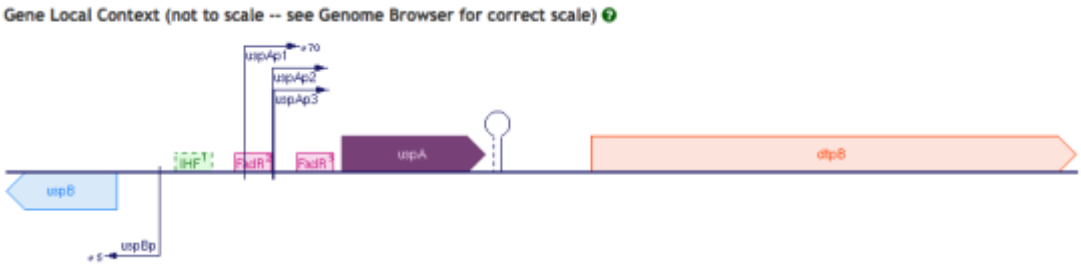


Figure 5.1: *Genetic context of the uspA gene. Obtained from (https://ecocyc.org/gene?orgid=ECOLI&id=EG11390-MONOMER#tab=TU, 07.03.2017).*

The *uspA* is not organized in operons with any other genes (this is also the case for *ompA* and *dps)*. To knock out these genes, a single point mutation changing the initiation codon from ATG to AGG was considered the best strategy. If a gene is the organized in a operon, the gene should be removed without altering the RBS. This is important in order to avoid generating unintended knockout mutations.

To design MAGE oligos able to cause these mutations a CRMAGE software developed by Ronda et al (http://staff.biosustain.dtu.dk/laeb/crmage/) was used. The gene sequence and the sequence 100bp upstream of the initiation codon (downstream if the gene was on the reverse strand) was copied into the CRMAGE software. The "T" in the initiation codon (ATG) was highlighted and changed to "G" (Figure 5.2).

Figure 5.2: *Input of uspA gene sequence, and the sequence 100bp upstream of the initiation codon, in the CRMAGE software (http://staff.biosustain.dtu.dk/laeb/crmage/). The nucleotide to be changed is marked in red.*

By inputting the sequence of a gene, and choosing a single point mutation, the CRMAGE software propose a suitable oligo to generate the mutation, using MAGE recombineering. The resulting output, from the input depicted in Figure 5.2, are shown in Figure 5.3.

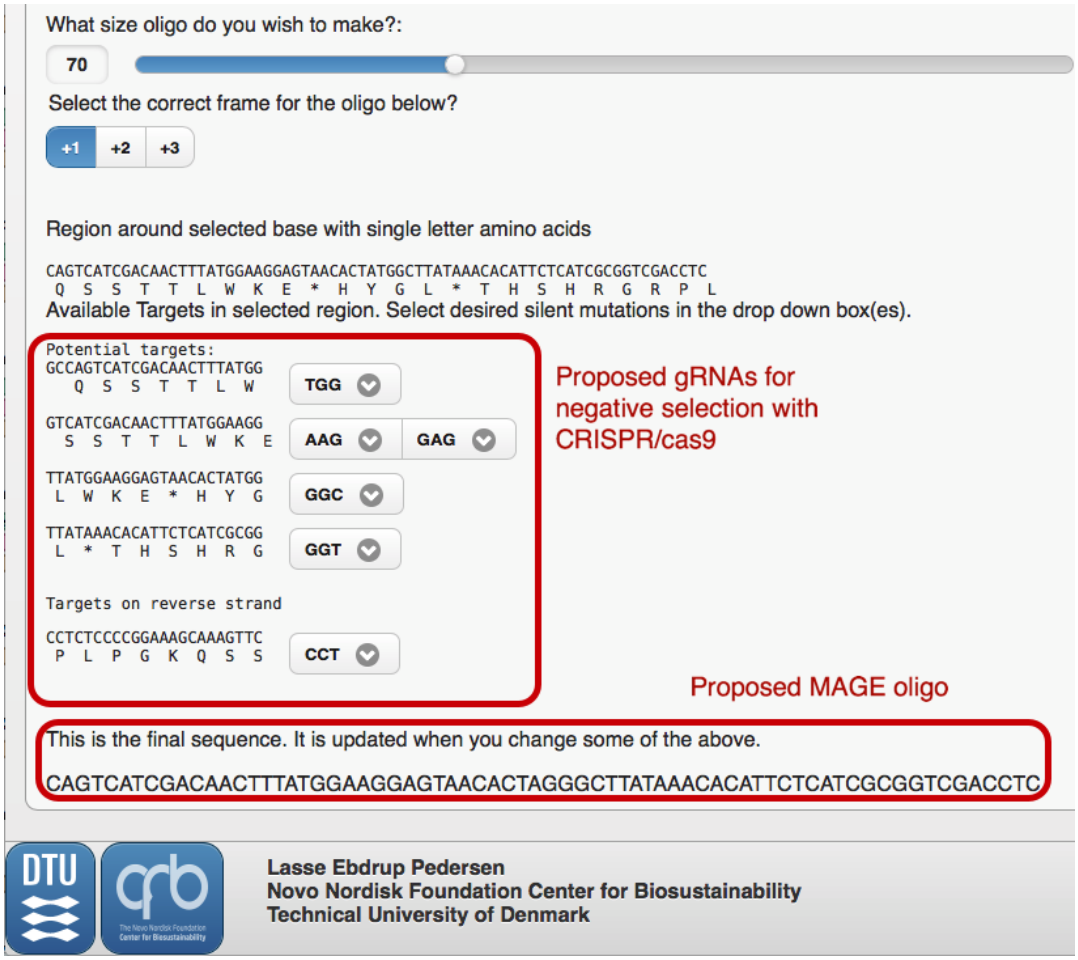


Figure 5.3: *Output from the CRMAGE software. Based on a desired size, the software proposes an oligo that will generate the desired mutation (figure 5.2) by MAGE recombineering. The software also proposes some suitable gRNA sequences, for use in negative selection with CRISPR/cas9.*

In addition to proposing a suitable MAGE oligo, the software propose several gRNA sequences for programming Cas9 to kill non-mutated cells. However, Benchling was considered a superior software to design gRNAs, compared to the CRMAGE software. To design gRNAs in Benchling, a region around the gene's initiation codon was highlighted and analysed with Benchling's built-in CRISPR suite. A suitable gRNA was selected, based on on-target scores (activity score 0-100), off-target scores (specificity score 0-100), and location relative to the MAGE oligo. Suitable gRNAs should have on-target and off-target scores larger than 50, and should have the PAM sequence (NGG) located within the MAGE oligo sequence. To ensure that the

designed gRNAs would program Cas9 to kill only non-mutated cells, the MAGE oligos from the CRMAGE software were altered to also induce a single point mutation in the PAM sequence next to the gRNA recognition site. A screenshot from the Benchling software, with the gRNA and the MAGE oligo used to silence *uspA*, is depicted in Figure 5.4.

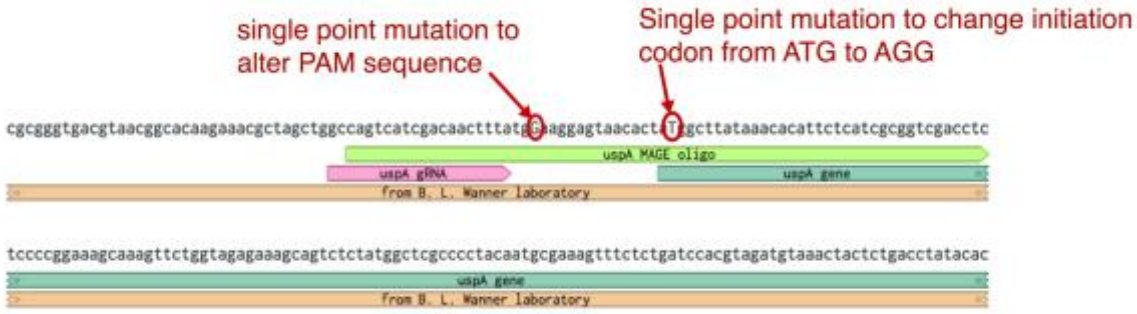


Figure 5.4: *Screenshot from the Benchling software. The MAGE oligo and gRNA sequence, used to silence the uspA gene with CRMAGE, are depicted. Generated mutations are highlighted with red circles. When possible, mutations in the PAM sequence were generated without altering the codons amino acid product (silent mutation).*

To ensure high frequencies of recombination between oligos and the bacterial genome, oligos were designed so that they would anneal to the lagging strand in the replication fork. The directions of replication, from the origin of replication in the *E. coli BW25113* chromosome, was identified by evaluating published literature on oligo design in *E. coli* [Gallagher et. al, 2014]. A map of the *BW25113* chromosome, with replication directions indicated, is depicted in Figure 5.5.

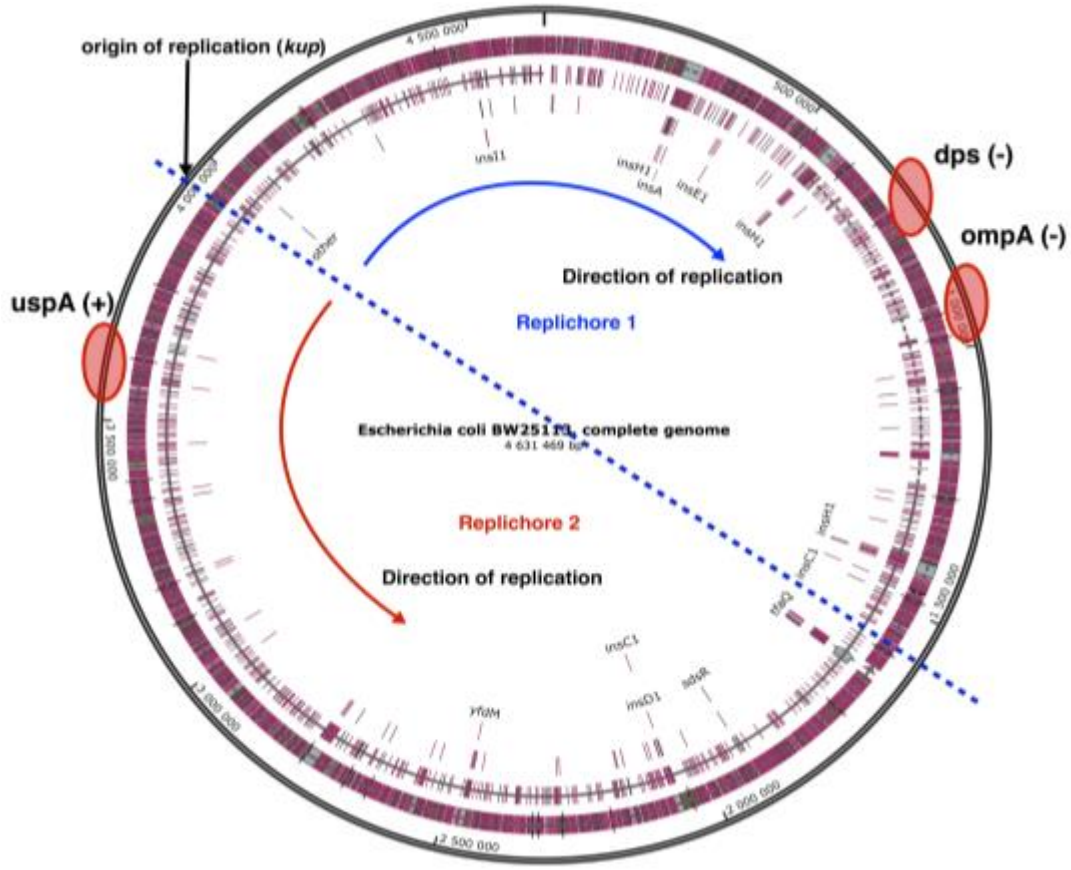


Figure 5.5: *Map of the E. coli BW25113 chromosome, with directions of replications indicated. In replichore 1, the lagging strand of the replication fork is on the forward strand. In replichore 2, the lagging strand is on the reverse strand [Gallagher et. al, 2014]. The circa placement of the three genes uspA, dps and ompA are marked with red circles. The annotation after the genes indicates which strand the gene is located on. (+) = forward strand, (-) = reverse strand.*

ompA and *dps* are both placed on the reverse strand in replichore 1 (Figure 5.5). Oligos with equal sequences as parts of these genes (with two mismatches, to generate mutations in PAM and initiation codon), will anneal to the lagging strand (forward strand) during replication. The same was the case for uspA, as this gene is placed on the forward strand in replichore 2. Oligos with equal sequences to this gene will anneal to the lagging strand (reverse strand) during replication.

5.9 Inserting specific gRNA encoding regions in pMAZ-SK plasmids

gRNAs used for negative selection during CRMAGE cycles where cloned into pMAZ-SK plasmid backbones, Yielding one specific pMAZ-SK construct for each gene that was targeted for knockout. USER-cloning was used to insert gRNA encoding regions in pMAZ-SK backbones. Colony PCR was used to screen for strains carrying pMAZ-SK constructs with gRNA encoding regions.

5.9.1 USER-cloning

20ng of linear pMAZ-SK backbone, amplified by using U-containing primers, was mixed with 10µl of gRNA encoding oligos and 1µl USER enzyme mix. For negative controls, water was added instead of gRNA encoding DNA. The DNA/enzyme mix was inoculated at 37°C for 15 min, followed by 15 min at 25°C. The entire volume of DNA/enzyme mix was used transform *E. coli BW25113*. The same general electroporation protocol as described in section 5.8 was used for all samples. The electroporated cells were plated on LB plates with 50µg/ml kanamycin.

5.9.2 Colony PCR

Colony PCR was used to screen for strains carrying pMAZ-SK plasmids with successfully inserted gRNA encoding regions. The primers used to screen for the different gRNA encoding regions are listed in table 5.3. The reverse primers were designed to recognise the pMAZ-SK backbone, while the forward primers would ligate specifically to the different gRNA encoding sequences. This allows for selection of quite large products, which should not be synthesized if the correct gRNA is not inserted in the pMAZ-SK backbone.

Table 5.3: *Primers used to identify correctly constructed pMAZ-SK plasmids during the colony PCR procedure*

Gene to be knocked out	gRNA encoding region	Forward primer	Reverse primer
Dps	GAGCACGTTGCGGGTATAAAGCAGATGTTTTAGAGCTAGAAAT	GTTGCGGGTATAAAGCAGAT	CGACCGCGTATTTCGTCTC
uspA	GAGCACGCCAGTCATCGACAACTTTAGTTTTAGAGCTAGAAAT	GCCAGTCATCGACAACTTTA	CGACCGCGTATTTCGTCTC
ompA	GAGCACGACAGCTATCGCGATTGCAGGTTTTAGAGCTAGAAAT	GACAGCTATCGCATTGCAG	CGACCGCGTATTTCGTCTC
flhD	GAGCACAATGTGTTTCAGCAACTCGGGTTTTAGAGCTAGAAAT	AATGTGTTTCAGCAACTCGG	CGACCGCGTATTTCGTCTC

Colonies from each plate (with strains carrying pMAZ-SK plasmids) were picked, replated, and added to a PCR mix containing:

- 1 µl 10X standard taq running buffer
- 0.2 µl 10mM dNTPs
- 0.2 µl 10µM forward primer
- 0.2 µl 10µM reverse primer
- 8.35 µl nuclease free water
- 0.05µl taq DNA polymerase

The conditions of the PCR reactions are listed:

1. 95.0°C for 10 minutes and 30 seconds (temperature incubation step to lyse cells)
2. 30 repeats of temperature cycling:
    1. 95.0°C for 20 seconds
    2. 51°C for 40 seconds
    3. 68°C for 1 minute
3. 68.0°C for 5 minutes

PCR products were analysed by gel electrophoresis on a 0.8% agarose gel. 1µl of PCR product was added to 9µl nuclease free water and 2µl loading buffer, and loaded to the agarose gel. Correct PCR products were identified by comparing the bands of each product with a positive, - and a negative control sample, and appropriate DNA ladders. The positive control was isolated pMAZ-SK plasmid with a verified correctly inserted gRNA coding region. Colonies with correct pMAZ-SK constructs were inoculated in

LB with 50µg/ml ampicillin, at 37°C, over night. Plasmids were isolated and purified with the plasmid miniprep kit from Zymo research.

5.10 Characterizing the ribosome synthesis reporter systems by growing strains in chloramphenicol

Strains carrying reporter plasmids were picked from frozen stock and inoculated overnight in 10ml LB medium with 100µg/ml ampicillin. The overnight (ON) cultures were washed by centrifugation and resuspension with fresh LB without antibiotics. 30µl of washed ON cultures were added to 10ml LB with chloramphenicol concentrations of 0,- 2, - 4, - and 8µM. 200µl of each of these cell suspensions were added to a well in a 96 well plate. Cells were grown for 2 hours at 37°C, 225RPM. OD_{600} was measured every 30 minutes, and relative fluorescence was measured after 2 hours of inoculation.

5.11 Measuring growth and fluorescence

The translational capacities of the different knockout mutants were evaluated by measuring growth, and fluorescence from strains carrying reporter plasmids (pMJ002 or pMJ001). The same protocols were used for all strains.

5.11.1 Growth

Cells were picked from frozen stock solution and inoculated overnight in 20ml LB medium at 37°C, 225RPM. 250µL of overnight culture was added to 25ml fresh LB medium, in 100ml shaking flasks, and the new culture was inoculated at 37°C, 225RPM, for 8 hours. Optical density (OD_{600}) was measured at appropriate intervals with a photo spectrometer. Sampling was done by pipetting 300µl of culture, to a well in 96well plate. When OD_{600} values exceeded 0.4, samples were diluted with sterile LB medium. To minimize random uncertainty, each growth experiment was conducted in three parallels.

5.11.2 Fluorescence

Strains carrying reporter plasmids (pMJ001 or pMJ002) were inoculated following the same protocol described in the previous section (5.11.1). After 8 hours of inoculation, 300µl of culture was added to a 96 well plate (Thermo Fischer Scientific Nucleon 96 Flat Black). Fluorescence was read by a Tecan Infinite 200 pro series microplate reader. Excitation, - and emission wavelengths was set to respectively 580nm and 615nm. Z-position was calculated from the first well, and the gain was set to 100. The

optical density in the cultures was measured with the same method described in the previous section (5.11.1). Optical density and fluorescence were both measured after 8 hours of inoculation.

6 References

Bienick M. S., Young K. W., Klesmith J. R., Detwiler E. E., Tomek K. J., Whitehead T. A. (2014). The interrelationship between promoter strength, gene expression, and growth rate. PLoS ONE9:e109105.10.1371/journal.pone.0109105

Bremer, H., and P. P. Dennis. (1996). Modulation of chemical composition and other parameters of the cell by growth rate, p. 1527–1542. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washing- ton, D.C.

Bosdriesz, E., Molenaar, D., Teusink, B., & Bruggeman, F. J. (2015). How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *The Febs Journal*, *282*(10), 2029–2044. http://doi.org/10.1111/febs.13258

Cairrão, F., Chora, Â., Zilhão, R., Carpousis, A. J., & Arraiano, C. M. (2001). RNase II levels change according to the growth conditions: characterization of gmr, a new Escherichia coli gene involved in the modulation of RNase II. *Molecular Microbiology*, *39*(6), 1550-1561. doi:10.1046/j.1365-2958.2001.02342.x

Calhoun, L.N. and Kwon, Y.M. (2011), Structure, function and regulation of the DNA-binding protein Dps and its role in acid and oxidative stress resistance in *Escherichia coli*: a review. Journal of Applied Microbiology, 110: 375–386. doi:10.1111/j.1365-2672.2010.04890.x

Carneiro, S., Lourenço, A., Ferreira, E. C., & Rocha, I. (2011). Stringent response of *Escherichia coli*: revisiting the bibliome using literature mining. *Microbial Informatics and Experimentation*, *1*, 14. http://doi.org/10.1186/2042-5783-1-14

Chen, C., & Deutscher, M. P. (2010). RNase R is a highly unstable protein regulated by growth phase and stress. *RNA*, *16*(4), 667–672. http://doi.org/10.1261/rna.1981010

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., … Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (New York, N.Y.)*, *339*(6121), 819–823. http://doi.org/10.1126/science.1231143

Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(12), 6640–6645.

De Anda R, Lara AR, Hernandez V, Hernandez-Montalvo V, Gosset G, Bolivar F, Ramirez (2006): Replacement of the glucose phosphotransferase transport system by galactose permease reduces acetate accumulation and improves process performance of ***Escherichia coli*** for recombinant protein production without impairment of growth rate. ***Metab Eng*2006,** 8:**281–290.**

Dedhia N, Richins R, Mesina A, Chen W (1997) Improvement in recombinant protein production in ppGpp-deficien**t***Escherichia coli* **.** *Biotechnol Bioeng* 53**:**379–386.

Demain, A. L., & Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances*, *27*(3), 297-306. doi:http://dx.doi.org/10.1016/j.biotechadv.2009.01.008

Deutscher MP (2003) Degradation of stable RNA in bacteria. J Biol Chem 278: 45041–45044

Deutscher, M. P. (2006) Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Research, 34*(2), 659–666. http://doi.org/10.1093/nar/gkj472

Deutscher MP (2009) Maturation and degradation of ribosomal RNA in bacteria. Prog Mol Biol Transl Sci 85: 369–391

Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science, 346*(6213).

Dong H, Nilsson L, Kurland CG (1995) Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. J Bacteriol 177: 1497–1504

D'Souza G., Waschina S., Pande S., Bohl K., Kaleta C., Kost C. (2014). Less is more: selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. Evolution 68, 2559–2570.10.1111/evo.12468

El-Sharoud WM (2004) Ribosome inactivation for preservation: concepts and reservations. Sci Prog 87: 137–152

Fischer E., Sauer U. (2005). Large-scale in vivo flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. Nat. Genet. 37, 636–640.10.1038/ng1555

Gallagher, R. R., Li, Z., Lewis, A. O., & Isaacs, F. J. (2014). Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA. *Nat. Protocols, 9*(10), 2301-2316. doi:10.1038/nprot.2014.082
http://www.nature.com/nprot/journal/v9/n10/abs/nprot.2014.082.html#supplementary-information

Garneau, J. E., Dupuis, M.-E., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature, 468*(7320), 67-71. doi:http://www.nature.com/nature/journal/v468/n7320/abs/nature09523.html#supplementary-information

N. Gustavsson, A. Diez, T. Nyström (2002) The universal stress protein paralogues of *Escherichia coli* are co-ordinately regulated and co-operate in the defence against DNA damage Mol. Microbiol., 43 pp. 107–117

Hannig, G., & Makrides, S. C. (1998). Strategies for optimizing heterologous protein expression in Escherichia coli. *Trends in Biotechnology, 16*(2), 54-60. doi:http://dx.doi.org/10.1016/S0167-7799(97)01155-4

Hecht, A., Glasgow, J., Jaschke, P. R., Bawazer, L. A., Munson, M. S., Cochran, J. R., Salit, M. (2017). Measurements of translation initiation from all 64 codons in *E. coli*. *Nucleic Acids Research, 45*(7), 3615–3626. http://doi.org/10.1093/nar/gkx070

Hirvonen, Ross, Wozniak, Marasco, Anthony, Aiyar, Newburn and Gourse (2001) Contributions of UP elements and the transcription factor FIS to expression from seven *rrn* promoters in *Escherichia coli*. J. Bacteriol. November 2001 vol. 183 no. 21 6305-6314. doi: 10.1128/JB.183.21.6305-6314.2001

Huang, C., Jr., Lin, H., & Yang, X. (2012). Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. *Journal of Industrial Microbiology & Biotechnology, 39*(3), 383-399. doi:10.1007/s10295-011-1082-9

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology*, *169*(12), 5429–5433.

Ivanovski, G., Gubenšek, F., & Pungerčar, J. (2002). mRNA secondary structure can greatly affect production of recombinant phospholipase A2 toxins in bacteria. *Toxicon*, *40*(5), 543-549. doi:http://dx.doi.org/10.1016/S0041-0101(01)00250-1

Jain, C. (2002), Degradation of mRNA in Escherichia coli. IUBMB Life, 54: 315–321. doi:10.1080/15216540216036

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), 816.

Klumpp S, Scott M, Pedersen S, Hwa T (2013). Molecular crowding limits translation and cell growth. Proc Natl Acad Sci USA.;110:16754–16759.

Kvint, K., Nachin, L., Diez, A., & Nyström, T. (2003). The bacterial universal stress protein: function and regulation. *Current Opinion in Microbiology*, *6*(2), 140-145. doi:http://dx.doi.org/10.1016/S1369-5274(03)00025-0

LaRiviere FJ, Cole SE, Ferullo DJ, Moore MJ (2006) A late-acting quality control process for mature eukaryotic rRNAs. Mol Cell 24: 619–626

Lemke, J. J., Sanchez-Vazquez, P., Burgos, H. L., Hedberg, G., Ross, W., & Gourse, R. L. (2011). Direct regulation of *Escherichia coli* ribosomal protein promoters by the transcription factors ppGpp and DksA. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5712–5717. http://doi.org/10.1073/pnas.1019383108

Liang, S.T., Xu, Y.C., Dennis, P., and Bremer, H. (2000). mRNA composition and control of bacterial gene expression. J Bacteriol 182, 3037-3044.

Lower B.H. ,Yongsunthon R. ,Vellano F.P.III, Lower S.K. (2005) Simultaneous force and fluorescence measurements of a protein that forms a bond between a living bacterium and a solid surface. *J.Bacteriol*187: 2127–2137.

Mahalik, S., Sharma, A. K., & Mukherjee, K. J. (2014). Genome engineering for improved recombinant protein expression in Escherichia coli. *Microbial Cell Factories*, *13*(1), 177. doi:10.1186/s12934-014-0177-1

Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., & Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct*, *1*, 7. http://doi.org/10.1186/1745-6150-1-7

Molenaar D, Berlo RV, Ridder DD, Teusink B, van Berlo R & de Ridder D (2009) Shifts in growth strategies reflect tradeoffs in cellular economics. Mol Syst Biol 5, 323.

Mordor intelligence, Global biopharmaceuticals market growth, trends & forecasts (2016-2021). https://www.mordorintelligence.com/industry-reports/global-biopharmaceuticals-market-industry?gclid=CKLt6fWY_c8CFYrFcgodt7IAZw

Muntel J., Fromion V., Goelzer A., Maass S., Mader U., Buttner K., et al. (2014). Comprehensive absolute quantification of the cytosolic proteome of *Bacillus subtilis* by data independent, parallel fragmentation in liquid chromatography/mass spectrometry (LC/MSE). Mol. Cell. Proteomics 13, 1008–1019.10.1074/mcp.M113.032631

Nachin, L., U. Nannmark, and T. Nystrøm. 2005. Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. J. Bacteriol. **187:**6265–6272.

NCBI database,
https://www.ncbi.nlm.nih.gov/nuccore/CP009273.1?from=14168&to=15298&sat=4&sat_key=126354407

Nyerges, Á., Csörgő, B., Nagy, I., Bálint, B., Bihari, P., Lázár, V., . . . Pál, C. (2016). A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proceedings of the National Academy of Sciences*, *113*(9), 2502-2507. doi:10.1073/pnas.1520040113

Nyström, T., & Neidhardt, F. C. (1993) Isolation and properties of a mutant of Escherichia coli with an insertional inactivation of the uspA gene, which encodes a universal stress protein. *Journal of Bacteriology*, *175*(13), 3949–3956.

Nyström, T.C. Neidhardt (1994) Expression and role of the universal stress protein, UspA, of *Escherichia coli* during growth arrest Mol. Microbiol., 11 pp. 537–544

Nyström, T.C. Neidhardt (1996) Effects of overproducing the universal stress protein, UspA, in *Escherichia coli* K-1 J. Bacteriol., 178, pp. 927–930

Petersen, C. (1993) Translation and mRNA stability in bacteria: a complex relationship. In *Control of Messenger RNA Stability* (Belasco, J. G., and Brawerman, G. eds.). pp. 117–145, Academic Press, San Diego.

Piir, K., Paier, A., Liiv, A., Tenson, T., & Maiväli, Ü. (2011). Ribosome degradation in growing bacteria. *EMBO Reports*, *12*(5), 458–462. http://doi.org/10.1038/embor.2011.47

Poteete, A. R. (2001), What makes the bacteriophage λ Red system useful for genetic engineering: molecular mechanism and biological function. FEMS Microbiology Letters, 201: 9–14. doi:10.1111/j.1574-6968.2001.tb10725.x

Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D. and Gold, L. (1992), Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. Molecular Microbiology, 6: 1219–1229. doi:10.1111/j.1365-2958.1992.tb01561.x

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, *5*, 172. http://doi.org/10.3389/fmicb.2014.00172

Ronda C, Pedersen LE, Sommer MO, Nielsen (2016) AT. CRMAGE: CRISPR optimized MAGE recombineering. Sci Rep. 6:19452.

Scott M., Gunderson C. W., Mateescu E. M., Zhang Z., Hwa T. (2010). Interdependence of cell growth and gene expression: origins and consequences. Science 330, 1099–1102.10.1126/science.1192588

Scott, M., Klumpp, S., Mateescu, E. M., & Hwa, T. (2014). Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Molecular Systems Biology*, *10*(8), 747. http://doi.org/10.15252/msb.20145379

Silvers JA, Champney WS (2005) Accumulation and turnover of 23S ribosomal RNA in azithromycin-inhibited ribonuclease mutant strains of Escherichia coli. Arch Microbiol 184: 66–77

Singh AB, Mukherjee KJ (2013) Supplementation of substrate uptake gene enhances the expression of rhIFN-beta in high cell density fed-batch cultures of *Escherichia coli* **.** *Mol Biotechnol* , 54:692–702.

Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila- Sanjurjo, A., Holton, J.M., and Cate, J.H. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. Science 310: 827–834.

Smith, S. G. J., Mahon, V., Lambert, M. A., & Fagan, R. P. (2007). A molecular Swiss army knife: OmpA structure, function and expression. *FEMS Microbiology Letters*, *273*(1), 1-11. doi:10.1111/j.1574-6968.2007.00778.x

Steitz, T.A. 2008. A structural understanding of the dynamic ribosome machine. Nat. Rev. Mol. Cell Biol. 9: 242–253.

Sørensen, H. P., & Mortensen, K. K. (2005). Advanced genetic strategies for recombinant protein expression in Escherichia coli. *Journal of Biotechnology*, *115*(2), 113-128. doi:http://dx.doi.org/10.1016/j.jbiotec.2004.08.004

Valgepea, K., Peebo, K., Adamberg, K., & Vilu, R. (2015). Lean-Proteome Strains – Next Step in Metabolic Engineering. *Frontiers in Bioengineering and Biotechnology*, *3*, 11. http://doi.org/10.3389/fbioe.2015.00011

Wang, Y. (2002). The Function of OmpA in Escherichia coli. *Biochemical and Biophysical Research Communications*, *292*(2), 396-401. doi:http://dx.doi.org/10.1006/bbrc.2002.6657

Widdel Friedrich (2010) Theory and measurement of bacterial growth. Grundpraktikum Mikrobiologie, 4. Sem. (B.Sc.) Universit t Bremen

Zhang, X., & Bremer, H. (1996). Effects of Fis on Ribosome Synthesis and Activity and on rRNA Promoter Activities inEscherichia coli. *Journal of Molecular Biology*, *259*(1), 27-40. doi:http://dx.doi.org/10.1006/jmbi.1996.0299

Zhao, D., Yuan, S., Xiong, B., Sun, H., Ye, L., Li, J., . . . Bi, C. (2016). Development of a fast and easy method for Escherichia coli genome editing with CRISPR/Cas9. *Microbial Cell Factories*, *15*(1), 205. doi:10.1186/s12934-016-0605-5

Zundel MA, Basturea GN, Deutscher MP (2009) Initiation of ribosome degradation during starvation in Escherichia coli. RNA 15: 977–983

Appendix A – Media recipes

LB medium

10g/l tryptone
5g/l yeast extract
5g/l NaCl

LA plates

LB medium with 15g/l agar
Antibiotics were added after autoclaving, when the medium had cooled to under 60°C

SOC medium

20g/l tryptone
5g/l yeast extract
4.8g/l $MgSO_4$
3,6g/l dextrose
0.5g/l NaCl
0.19g/l KCl

Minimal glucose medium

2g/l $(NH_4)_2SO_4$
13.6g/l KH_2PO_4
0.5g/l $MgSO_4*7H_2O$
0.5mg/l $FeSO_4$
2g/l CasAA
5g/l glucose
5ml/l glycerol
5g/l galactose

Appendix B - Error analysis and statistical methods

Error analysis

The calculations and methods used to estimate errors in the data, obtained in this work, are presented in this section.

Error in measured values

All measured data were presented as the mean value of at least three parallels. The error in each presented data point was estimated as the standard deviation of mean SDOM. SDOM was calculated with equation B.1 [Taylor, 1997].

$$SDOM = \sqrt{\frac{\frac{1}{N-1} * \sum_{i=1}^{N} (X_i - \bar{X})^2}{N}} \qquad \text{(B.1)}$$

N = number of individual measurements

X_i = Individual measurement value

$\bar{X}$ = Average value of individual measurements

Calculated SDOM have a 68% confidence interval [Taylor, 1997]. Hence, the chance that the means of two measurements series deviates from each other by one SDOM is 68%. In other words, if one measure a random set of samples and use only SDOM error bars to evaluate level of significance, every third sample will appear to be significantly different from the others, because of random errors. Using a 5% confidence interval, data points should deviate by at least 2 SDOMs to be considered significantly different.

Error in comparing numbers

When numbers were compared, by dividing one number with another, the numbers' fractional errors were added. Fractional errors were obtained by dividing the uncertainties by the corresponding measured values:

$$Frac.error\ in\ value\ A = \frac{Error\ in\ value\ A}{Size\ of\ value\ A}$$

$$Error\ in\ value\ \left(\frac{A}{B}\right) = (Frac.error\ A + Frac.error\ B) * \frac{A}{B}$$

Error in calculating growth rates

OD_{600} measurements were used to calculate growth rates by using equation B.2 [Widdel 2010].

$$\mu = \frac{\ln(OD_2) - \ln(OD_1)}{t_2 - t_1} \qquad \text{(B.2)}$$

μ = Growth rate

The error in ln(OD) data points were calculated with equation B.3

$$\delta\ln(OD) = \left|\frac{dLn(OD)}{dOD}\right| * \delta OD = \frac{\delta OD}{OD} \qquad \text{(B.3)}$$

δln(OD) = Error in ln(OD) data
δOD = Error in measured OD data (SDOM)

The error in the calculated growth rates was estimated by evaluating the steepest and most moderate linear regression line through the lnOD data:

$$Slope\ of\ steepest\ fit = \frac{[\ln(OD_1) - \delta\ln(OD_1)] - [\ln(OD_2) + \delta\ln(OD_2)]}{t_1 - t_2} \qquad \text{(B.4)}$$

$$Slope\ of\ most\ moderate\ fit = \frac{[\ln(OD_1) + \delta\ln(OD_1)] - [\ln(OD_2) - \delta\ln(OD_2)]}{t_1 - t_2} \qquad \text{(B.5)}$$

$$\delta\mu = \frac{|Slope\ of\ best\ fit - slope\ of\ steepest\ fit| + |Slope\ of\ best\ fit - slope\ of\ most\ moderate\ fit|}{2} \qquad \text{(B.6)}$$

δμ = estimated error in growth rate

Error in estimating relative ribosome content

The relative ribosome fraction between two samples was calculated with equation 6

$$\frac{\Phi^1_{R,max}}{\Phi^2_{R,max}} = \frac{\frac{\lambda^1_{max}}{\gamma} + \Phi^{min}_R}{\frac{\lambda^2_{max}}{\gamma} + \Phi^{min}_R} = \frac{\frac{\lambda^1_{max}}{10.8} + 0.05}{\frac{\lambda^2_{max}}{10.8} + 0.05} \qquad (6)$$

$\Phi^i_{R,max} = maximum\ ribosome\ fraction\ of\ entire\ proteome\ for\ strain\ "i"$

λ^i_{max} = maximum achieved growth rate for strain "i"

Φ^{min}_R = Fraction inactive ribosomes (empirical constant)

γ = Elongation rate per active ribosome (empirical constant)

Using error propagation rules [Taylor, 1997] the error in a relative ribosome content comparison can be expressed by equation B.7.

$$\delta\left(\frac{\Phi^1_R}{\Phi^2_R}\right) = \frac{\left|\frac{\delta\lambda^1}{10.8}\right|}{\frac{\lambda^1+0.05}{10.8}} + \frac{\left|\frac{\delta\lambda^2}{10.8}\right|}{\frac{\lambda^2+0.05}{10.8}} \qquad (B.7)$$

6.1 Statistical methods

ANOVA analysis

Analysis of variance (ANOVA) is a statistical method used to identify how different factors contribute to variance in a given set of data. The observed variance in a set of output values is partitioned into groups, where each group is attributable to a different source of variation. The ANOVA analysis uses a F-test to compare variation within, and between, groups, and uses a statistical F-distribution to evaluate if the means of groups are significantly different.

Four assumptions are used in the ANOVA analysis:

1. Expected values of errors in measurements are zero
2. The variances in all errors are equal
3. Errors are independent and random
4. Errors are normally distributed

Statistical significance calculated with ANOVA is independent of constant biases and scaling errors.
All ANOVA analysis was conducted in Matlab. A simple script, using Matlab's built-in ANOVA functions, was used:

```
function anovaX(data,response,labels)
% columns correspond to different categories
% rows correspond to different sample points
% response determines which column of the data matrix is the response variable
% labels contain the names of the different categories (columns) in  'data'
d=data;
d(:,response)=[];
n=size(data,2)-1;
vd=cell(1,n);
for i=1:n
   vd(i)={d(:,i)};
end
cat=labels;
cat(:,response)=[];
vec=[1:n];
[p,tbl,stats]=anovan(data(:,response),vd,'model','interaction','varnames',cat,'sstype',1,'continuous',vec);
end
```

This script returns an ANOVA table, like the one depicted in Figure B.1.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Strain	541806.6	1	541806.6	146.66	0
Reporter plasmid	36602.5	1	36602.5	9.91	0.0051
Strain*Reporter plasmid	114.8	1	114.8	0.03	0.8618
Error	73885.3	20	3694.3		
Total	652409.3	23			

Figure B.1: *ANOVA table. Sum sq. = sum of squares of data within group; d.f.= Degrees of freedom; Mean sq. = mean square of error within a group: F = output value from statistical F-test.*

Prob>F is the most important output. It describes the probability that the corresponding source of variance did not have a significant effect. Setting a level of significance to 5%, only sources with Prob>F values equal to, or lower than 0.05 was accepted as significant. The 4 first values are used to calculate the Prob>F values.

T-test

The T-test evaluates whether the means of two groups are significantly different from each other. The test assumes that errors are normally distributed, and compares the difference in group means and the variance in the groups. By accounting for the variance within groups, T-tests maintain reliability even when random uncertainties in measurements are high.

All T-test calculations were done by using the t-test function in excel:

=TTEST(array1;array2;2;3)

Array 1 and array 2 refers to the two data sets that are to be compared. The first number (2) indicates that a two-tailed T-test should be used. The second number (3) programs excel to assume that the variance of errors is not identical in the two data sets. The function returns a single value (p-value), describing the probability that the observed outcome would occur given that the null hypothesis similar means is true. As in the case of the ANOVA analysis, only data sets with p-values equal to or lower than 0.05 were accepted as significantly different.

Rejection of data – Chauvenet's criterion

In cases where it was suspected that outliers corrupted data, Chauvenet's criterion was used to evaluate if single data points should be rejected. This criterion is used to identify outliers, based on standard deviations in a set of data.
In a series of "n" measurements with an average value "avg" and a standard deviation "st.dev" the data point "x" should only be rejected if:

$$1 - prob\left[\left(\left|\frac{x - avg}{st.dev}\right|\right) * st.dev\right] * \frac{1}{n} > 0.5$$

Following standard distribution statistics, this implies that the following criterion should be fulfilled when considering to reject data:

$$\left|\frac{x - avg}{st.dev}\right| * \frac{1}{n} > 1.96$$

In other words: data points should be at least 1.96*n standard deviation away from the average, if they are to be rejected. Where "n" is the number of parallel measurements.