The impact of *zwf* overexpression on plasmid biopharmaceutical production by *Escherichia coli*

Rita Azevedo Fernandes

Thesis to obtain the Master of Science Degree in

Biotechnology

Supervisor: Professor Doutor Gabriel António Amaro Monteiro

Examination Committee

Chairperson: Professor Doutor Arsénio do Carmo Sales Mendes Fialho

Supervisor: Professor Doutor Gabriel António Amaro Monteiro

Member of the Committee: Professor Doutor Jorge Humberto Gomes Leitão

November, 2014

Abstract

Plasmid DNA vectors may find application as preventive or therapeutic DNA vaccines for viral, bacterial, or parasitic diseases or for other indications such as cancer, or gene therapy products. Therefore, the demand of pDNA is expected to rise significantly in the near future and the establishment of a high-yield and cost effective manufacturing process is one of the key challenges for the success of plasmid DNA market.

With the goal of creating an *E. coli* strain specifically adapted to meet the upstream and downstream processing challenges associated with large scale production of plasmid vectors, some mutations have been made in its genome. One main area of focus is modification of central carbon metabolism genes to increase flux toward nucleotide and amino acid precursor synthesis and reduction of byproducts, such as acetate. The *zwf* and *rpiA* are two important genes of the Pentose Phosphate Pathway and its overexpression was already tested and appeared to increase plasmid amplification rate. Therefore, the overexpression of the *zwf* and *rpiA* genes in GALG20 (MG1655*ΔendAΔrecAΔpgi*) is intended in this project, by replacing its original promoters with stronger ones.

The results obtained showed that *zwf* gene overexpression did not improve plasmid productivity in GALG20 strains, behaving similarly to GALG20rpiA.

Keywords: Plasmid DNA, *E. coli*, DNA vaccines, Metabolic Engineering, *zwf* gene

Resumo

A utilização de DNA plasmídico tem vindo a aumentar ao longo dos anos, podendo ter como aplicação vacinas de DNA para doenças virais, bacteriológicas ou parasíticas ou outras indicações como cancro. Assim, é expectável que a necessidade de DNA plasmídico aumente significativamente num futuro próximo, sendo necessária a utilização de um processo de produção de alto rendimento a baixo custo.

Com o objectivo de criar uma estirpe de *E. coli* adaptada aos desafios de *upstream* e *downstream processing* de DNA plasmídico em grande escala, determinadas mutações têm sido feitas ao seu genoma. Uma das principais estratégias utilizadas é a modificação de genes envolvidos no metabolismo central de carbono por forma a aumentar o fluxo na via de síntese de percursores de nucleótidos e aminoácidos e reduzir a formação de produtos secundários como o acetato. Assim, neste projecto foi feita a sobreexpressão de dois genes da via das pentoses fosfato, *zwf* e *rpiA*, na estirpe GALG20 (MG1655*ΔendAΔrecAΔpgi*) através da substituição dos promotores originais por outros com forças superiores.

Os resultados obtidos mostram que a sobreexpressão do gene *zwf* não aumentou a produtividade de DNA plasmídico na estirpe GALG20, comportando-se de forma semelhante a estirpes GALG20rpiA.

Palavras-chave: DNA plasmídico, *E. coli*, Vacinas de DNA, Engenharia Metabólica, Gene *zwf*

Acknowledgements

First of all I would like to thank to my supervisor Professor Gabriel Monteiro who gave me this great opportunity and was always helpful, available and supportive.

I would like to thank to Luís Carreira who helped me in the first months of my work in the lab and gave me the knowledge needed to start my own thesis work.

I would like to thank to Diana Marques, for all the support, patience and companionship throughout this year and for never losing the smile, even in the toughest times.

I would also like to thank to my lab colleagues and friends who made this year very special and unforgettable: João Trabuco, Ana Faria, Liliana Brito, Isabel Pinto, Cátia Jorge, Maria Vieira, Carlos Rodrigues and Fábio Gonçalves.

João Trabuco and Pedro Pereira deserve a special thank you for all the support and patience throughout this year and for never denying help. I could not forget Salomé Magalhães, Sofia Duarte, Jorge Paulo, Cláudia Alves, Ana Rosa and Luís Raiado for all the support and advice. I would also like to thank Ricardo Pereira for the help and support given and for making our lab work easier.

Finally, I would like to thank to my friends Catarina Freire, Beatriz Batista, Bruna Sousa and Rita Pitta for all the constant support and friendship and for being present through all these years. I also have to thank to my family who always supported my choices and never stopped believing in me. Last but not the least, I would like to thank to Pedro Rosa for the constant support and motivation and for always believing in me.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

pDNA - Bacterial Plasmid DNA
DNA - Deoxyribonucleic acid
RNA - Ribonucleic acid
tRNA - Transfer RNA
rRNA - Ribosomal RNA
gDNA - Genomic deoxyribonucleic acid
FDA - Food and Drug Administration
SFDA - State Food and Drug Administration
EMA - European Medicines Agency
PP - Pentose Phosphate
PPP - Pentose Phosphate Pathway
HIV - Human immunodeficiency virus
NADPH - Nicotinamide Adenine Dinucleotide Phosphate
ATP - Adenosine Tri-phosphate
R5P - Ribose 5-phosphate
E4P - Erythrose-4-phosphate
FAD - Flavin adenine dinucleotide
TCA - Tricarboxylic acid cycle
OAA - Oxaloacetate
AKG - α-ketoglutarate
LPS - Lipopolysaccharides
PEP - Phosphoenolpyruvate
PTS - Phosphotransferase System
G6P - Glucose 6-phosphate
Kan - Kanamycin
F6P - Fructose 6-phosphate
G3P - Glyceraldehyde-3-phosphate
PGP - Phosphoglycolate Phosphatase
3PG - 3-Phosphoglycerate
Pyr – Pyruvate
2PG - 2-Phosphoglycerate
AVV - Adeno-associated Viral Vector
LPLD - Lipoprotein Lipase Deficiency
6PGlac - 6-Phospho-Gluconolactone
6PG - 6-Phospho-Gluconate
Ru5P - Ribulose-5-phosphate
R5P - Ribose-5-phosphate
Rpi – Ribose-5-phosphate Isomerase
X5P - Xylulose 5-phosphate
S7P - Sedoheptulose 7-phosphate
AcA – Acetic acid
CIT - Citrate

1. Introduction

1.1 Plasmid DNA and Therapeutic Applications

During the last 40 years, plasmids have played a crucial role in the development of biotechnology, being probably the molecular tool most widely used for DNA manipulation, transfer, and gene expression in a variety of microorganisms and animal cells [1]. In the field of pharmaceutical biotechnology, plasmids are essential to produce heterologous proteins. Such proteins should substitute defective ones present in patients, or provide a lost function due to lack of the natural active protein.

An alternative to treat diseases related to the defective production of a protein, or missing genetic information, is the administration of the gene of interest to the patient, a strategy that is known as gene therapy or genetic medicine [1]. For a plasmid DNA vaccine, the target DNA sequence is cloned into a plasmid under the control of a eukaryotic promoter. After purification of the plasmid DNA, the final vaccine is injected into the recipient, where the target DNA sequence is expressed by the eukaryotic promoter [1].

The underlying concepts of gene therapy date back to the 1960s. In the early 1990s the transfer of genes to humans was reported. The first federally approved gene therapy clinical trials took place in 1990 when an adenosine deaminase (ADA)-deficient patient was given her own T cells engineered with a retroviral vector carrying a normal ADA gene [2]. This experiment paved the way for further clinical trials and since then, gene therapy and genetic vaccination have attracted much attention. As of 2013, over 2030 clinical trials have been completed, are ongoing or have been approved worldwide [3]. More than 60% of the trials have been performed in America and almost 30% in Europe (Figure 1). [4]

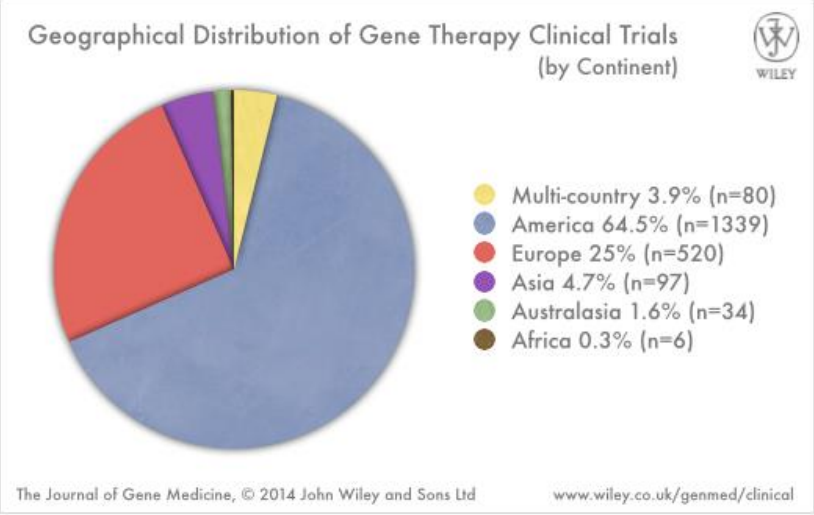


Figure 1 - Geographical distribution of Gene Therapy clinical trials (by continent) [4].

1.1.1 Advantages of Plasmid DNA in Therapeutic Applications

There exist several possible vectors to introduce the genetic information into human cells. The most relevant are virus (adenovirus or retrovirus) and plasmid DNA (pDNA), which can be used in aqueous solution (naked) or included in lipids or other formulations. Almost 18% of the trials for human gene therapy have been based on naked pDNA, whereas lipofection (which also requires pDNA production) counts for 5.3% of the trials [4]. Together, both approaches represent nearly 25% of the techniques used in clinical trials. On the other hand, the highly efficient transfection machinery of viruses makes them a very important tool for gene therapy, representing about 50% of the vectors used in current trials (Figure 2) [4].

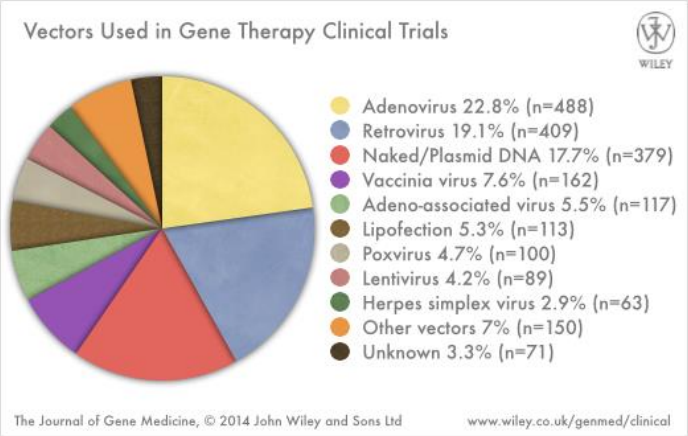


Figure 2 - Vectors used in Gene Therapy clinical trials [4].

Isolated cases of safety issues were reported for early clinical trials based on retroviral vectors, such as severe immune and inflammatory response that led to the death of one patient [5]. Therefore, although expression of the therapeutic gene following transfection with viral vectors is significantly higher than that using pDNA, the direct use of plasmid has important advantages over viruses. They are relatively simple to generate and safe to administer [6]. In contrast to vaccines that employ recombinant bacteria or viruses, genetic vaccines consist only of DNA or RNA, which is taken up and translated into protein by host cells [6]. Also, compared to retrovirus vectors, pDNA presents decreased risk of integration of the foreign DNA into the human chromosomes [6]. Another advantage of pDNA is that the gene size is not limited to the size of the viral capsid. Because they are not associated with a viral coat, naked nucleic acids are not generally subject to neutralizing antibody reactions that can hamper the clinical efficacy of vaccines based on recombinant viruses [6]. Further, the lack of a viral coat eliminates the possibility that immune responses to the coat will be immunodominant over immune responses to the desired transgene product [6].

Unfortunately, the immune response of pDNA when used for vaccination is usually lower than that of viral vectors [1]. Hence, large amounts of pDNA are necessary for a successful immunization. Since pDNA is produced in bacterial cultures, all the accumulated knowledge with such systems can be readily applied to manufacturing processes.

1.1.2 Clinical trials for pDNA-based gene therapy

In current clinical trials based on pDNA, the immunization against or the treatment of some severe illnesses, especially of cancer and cardiovascular diseases, is investigated, including malignant melanoma, pancreatic carcinoma, renal cell carcinoma, critical limb ischemia, and coronary heart disease [1]. Currently, 64.1% of the clinical trials are related to cancer, followed by 9.1% for monogenic diseases, 8.2% for infectious diseases and 7.8% for cardiovascular diseases (Figure 3) [4].

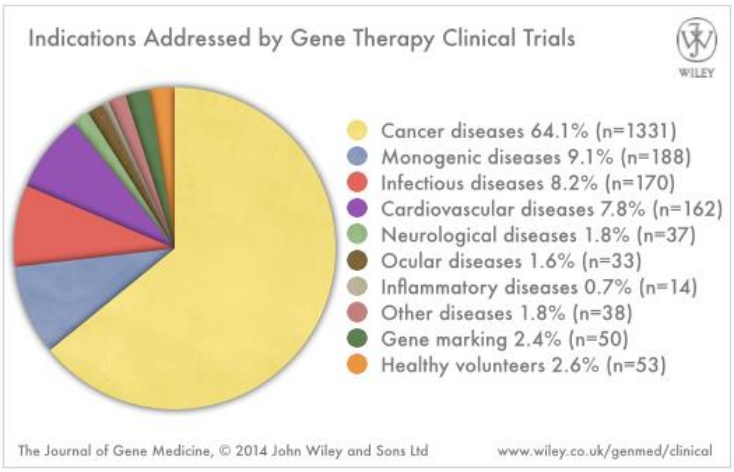


Figure 3 - Indications addressed by gene therapy clinical trials [4].

While some animal DNA vaccines are already being used, many human DNA vaccines are still being tested in clinical trials (Table 1) [7]. Successfully tested vaccines in animal models opened the door to the development of human vaccines, using procedures approved by ethic committees. Testing DNA vaccines in clinical trials is important to obtain information about adverse effects, safety and efficacy of the vaccine and vaccination procedure [8].

Table 1 - Human DNA vaccines currently being tested in clinical trials [7].

Condition	Phase
Seasonal Influenza	**I**
Metastatic Breast Cancer	**I**
Hemorrhagic Fever With Renal Syndrome	**II**
Chronic Hepatitis B	**II**
Condylomata Acuminata	**III**
Human Papillomavirus Infection	**IV**
Hepatitis B Infection; Chronic Infection; Viremia	**IV**

1.1.3 Plasmid DNA market

The global market for DNA vaccines, which was valued at just $78 million in 2013, is expected to grow to $2.7 billion by 2019 [9]. The technological and structural changes occurring in the vaccine industry are strong driving forces for revitalizing this particular market and providing new growth

opportunities. It is important that biotechnology and pharmaceutical firms understand the technology and market dynamics within this industry in order to best develop strategic plans and actions [10].

Four DNA vaccines have already been approved for veterinary use, being two of them prophylactic vaccines against infectious diseases, one for cancer immunotherapy and one for gene therapy (Table 2) [8].

Table 2 - DNA vaccines approved for veterinary use [8].

Type	Species	Target	Product/Company	License date/ country	Route of administration	Benefits
Prophylactic vaccine	Horses	West Nile Virus	West Nile-Innovator®/ Fort Dodge Animal Health	2005 USA	IM[1]	Production of protective antibodies
Prophylactic vaccine	Salmon	Infectious haematopoietic necrosis virus (IHNV)	Apex-IHN®/Novartis Animal Health	2005 Canada	IM	Stimulation of innate and adaptive immune responses improving the welfare and product yield
Immunotherapy of cancer	Dogs	Melanoma	Oncept™/Merial	2010 USA	ID[2] needle-free.	Production of antibodies capable of preventing the progress of the disease and prolonging the animal's life
Gene therapy	Swine	Growth hormone releasing hormone (GHRH)	LifeTide® SW5/VGX™ Animal Health	2008 Australia	IM followed by electroporation	Decrease in perinatal mortality and morbidity, increase of the number of pigs weaned per sow and productivity

Sadly, until now no licensed human DNA vaccines exist, but the results obtained with the commercialized veterinary vaccines and the various human clinical trials that are under study are increasing the reliability of this vaccine platform, greatly raising hopes for the successful development of human vaccines and therapies.

1.2 Plasmid biopharmaceutical production

The demand for plasmid DNA is anticipated to increase significantly as DNA vaccines and non-viral gene therapies enter phase 3 clinical trials and are approved for use. This increasing demand, along with renewed interest in pDNA as a therapeutic vector, has motivated research targeting the design of high-yield, cost-effective manufacturing processes [11]. Much effort has been directed towards downstream processing and fermentation strategies including factors that can enhance process performance such as the effects of growth medium, feeding mode and employment of high cell-density cultures. But it is also important to try to improve the earlier stages of production namely the cell host/plasmid selection where the choice of the *E. coli* host strain used to propagate the product containing-plasmid may have a significant influence on the quality and quantity of the purified DNA [12].

1.2.1 Concerns relative to plasmid DNA design

Plasmid DNA used for vaccination is based on constructed bacterial-derived plasmids that contain a eukaryotic gene expression cassette, responsible to code the immunogenic or immunomodulatory molecule of interest in mammalian cells, and thus, allowing their *in vivo* generation. DNA vaccine plasmids can be divided into two main structures: 1) the plasmid backbone, necessary for prokaryotic propagation and which contains a bacterial origin of replication and a bacterial selectable marker like an antibiotic resistance gene, and 2) a transcriptional unit, required for eukaryotic expression, which includes the promoter, used to drive optimal and high expression of the ORF (open reading frame) of interest in mammalian cells, and the Polyadenylation (poly A) sequence (Figure 4). Besides these two main structures it is also necessary to insert in the ORF of interest a specific consensus sequence present in the eukaryotic mRNA, called Kozak sequence (ACCATGG), which signals the start codon and is necessary to initiate protein synthesis in eukaryotic cells. To ensure correct translation termination and prevent read through, it is also important to add stop-codons in the ORF [8].

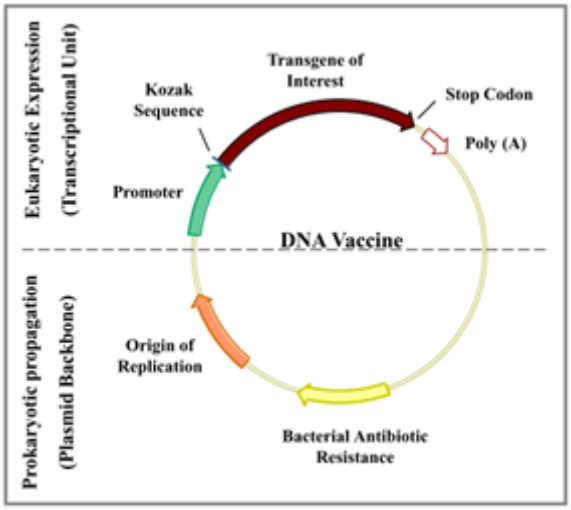


Figure 4 - Structure of DNA vaccine plasmids. The top of the figure shows the transcriptional unit (eukaryotic expression region), responsible for high-level expression of the antigen of interest in eukaryotic cells. The lower part of the figure (prokaryotic region) is responsible for propagation and vaccine plasmid maintenance in bacterial cells [8].

Many vector engineering efforts focus on modifications to the basic therapeutic backbone such as origin of replication, antibiotic resistance and promoters. These modifications aim to improve the production process by increasing yield, improving product homogeneity and quality, and/or ensuring the sequence fidelity of the final plasmid product [11].

One of the main obstacles during the developmental stages of a new therapeutic DNA molecule involves tackling a wide array of structural instability events occurring in/with pDNA and therefore assuring its structural integrity [13]. It is known that the type and overall organization of the genetic elements present in pDNA vaccines directly impact not only its bulk production, but also its shelf stability, efficacy and ultimately its clinical approval [13]. Therefore, elements prone to introduction of any type of structural instability (Table 3) should be avoided if possible [13].

Table 3 - Overview of major factors affecting plasmid structural stability [13].

Structural feature	Consequences
Plasmid size	• Large plasmids have a higher probability of containing potentially unstable regions or integrating into the host genome • Transfection efficiency decreases with plasmid size
Purine–pyrimidine/oligopurine–oligopyrimidine tracts; G-rich sequences	• Formation of Z-DNA, triplexes or quadruplexes that are involved in deletions and reduced plasmid yields
Chi sequences (5′-GCTGGTGG-3′)	• Loss of plasmid sequences and dimerization via the RecBCD pathway
Direct repeats	• Formation of DNA hairpin/loops • Loss of plasmid sequences or dimerization with concomitant loss or gain of genetic material
Inverted repeats	• Generation of cruciform structures • Formation of circular inverted head-to-head plasmid dimers
polyA sequences and other nuclease-sensitive regions	• Increased degradation by nucleases
Regions similar to gDNA	• Plasmid integration into host genome with disruption of potentially critical regions
Insertion sequences	• Gene activation • Gene down-regulation or inactivation • Deletions • Deleterious effects on plasmid replication and amplification

An important aspect concerning plasmid DNA design is the sequence integrity of the plasmid vector. It is essential to ensure that the sequence and structure of the plasmid is maintained throughout the duration of the process, due to its enormous impact on product safety, yield, and quality [11]. Given that, contamination of plasmid DNA by mobile elements is a serious regulatory concern, as these elements can alter the biological properties and safety profile of the vector DNA [11]. Strains with a high copy number of transposons, which are known to insert into plasmid DNA, are not recommended due to the fact that these elements can mediate recombination events such as transposition and horizontal gene transfer. For example, DH10B was found to have a mutation rate 13.5-fold higher than wild-type *E. coli* (MG1655 strain), mostly due to a significantly higher rate of insertion sequence (IS) transposition [11].

Another important characteristic that might affect plasmid stability is its size. At present, the size of the plasmid DNA being used in clinical trials is at the lower end of the possible range, typically lower than 10 kb [14]. Large plasmids have several drawbacks, such as a higher probability of containing potentially unstable regions or integrating into the host [15], lower transfection efficiency [16] and higher metabolic burden [17]. Given that, minicircles are emerging as new tool for *in vitro* and *in vivo* gene transfer, containing only the transcription unit required for expression of the therapeutic gene(s) [11].

The plasmid DNA structure is also an important aspect to consider. It is known that the supercoiled plasmid form is an important parameter to measure in the quality and control of the final product. The plasmid DNA that constitutes DNA vaccines is produced in bacteria as supercoiled or covalently closed circular (CCC) plasmid DNA. Single strand nicking results in the relaxation of supercoiled DNA into open circular DNA (OC) [18]. It is known that supercoiled plasmid has superior biological activity as compared to other plasmid forms [18]. Also, there is a greater risk of the linear form integrating into the host genome which could lead to deleterious effects. Therefore, the US Food and Drug Administration (FDA) recommends that vaccines contain at least 80% supercoiled plasmid [18].

In order to achieve an efficient gene expression, the pDNA must reach the nucleus intact and preferentially in the supercoiled isoform. Rational plasmid design can be of great help in order to avoid

nuclease degradation [13]. Regions that are particularly prone to nuclease degradation can be removed or modified to improve its resistance [19].

FDA requires a rigorous analysis for biodistribution, persistence and integration for approval of clinical trials involving pDNA [20]. One of the major concerns regarding the use of plasmid vectors as DNA vaccines is their possible integration into the host genome. This integration can have serious and even deleterious effects on host genome integrity, including the risk of tumorigenesis if the insertion reduces the activity of a tumor suppressor or increases the activity of an oncogene [20]. It may also result in chromosomal instability through the induction of chromosomal breaks or rearrangements. Given that, pDNA regions that share homology with host genome sequences and might increase the potential insertion frequency (even at a low frequency), should be eliminated, as well as all non-essential regions [13]. Another important aspect concerning safety is the presence of contaminants. International regulatory agencies are likely to set stringent specifications with regard to the level of endotoxins, RNA, protein, bacterial host DNA and quantities of linear and supercoiled DNA in the preparation and presence of toxic chemicals [14]. To avoid the presence of contaminants, there is the need to develop sensitive and quantitative analytical methods [21, 22]. Finally, FDA has long recommended that β-lactam antibiotics be avoided, as residual contamination in the final product could potentially cause an allergic reaction in sensitive individuals [11]. There is also the possibility of spread of antibiotic resistance genes to intestinal bacteria by horizontal gene transfer. To avoid these problems, there has been recently developed plasmid DNA that lacks the antibiotic resistance gene to further improve their safety profile [23].

1.2.2 Host strains

Currently, plasmid DNA is often produced using selected strains of *Escherichia coli* based on commercial availability, advantageous properties for mutagenesis or in many cases simply because standard laboratory-scale protocols had previously quoted using a particular host strain [12].

Since the beginning of the modern biotechnology era in the late 70s, *E. coli* has been widely used for molecular cloning methodologies and as a host to produce primary and secondary metabolites [24]. *E. coli* possesses a number of excellent properties, such as a rapid doubling time and growth rate, ease of high-cell-density fermentation, low production cost and most importantly a detailed knowledge of the metabolism and availability of excellent genetic tools for strain improvement [24]. Its genome is fully sequenced and can be easily manipulated by techniques that are maturing at a rapid pace [25]. On the other hand, *E. coli* has some disadvantages like endotoxin production, or lipopolysaccharides (LPS), and genetic instability, resulting in safety concerns surrounding its use [25]. For this reason, there has been some work investigating other microorganisms, such as the Gram-positive *Lactococcus lactis*, as hosts for biopharmaceutical production of pDNA [26]. Although not as efficient for plasmid production this bacteria produces neither endotoxin nor biogenic amines [26]. However, taking into consideration the benefits and drawbacks, *E. coli* is currently the most suitable organism for pDNA production on the industrial scale.

There are thousands of potential *E. coli* strains that can be used for pDNA production [27]. MG1655, J53 and W3110 represent some of the early *E. coli* K-12 strains that have been exposed to the fewest cycles of mutations. MG1655 and J53 are two mutational steps from the original *E. coli* K-12 whilst W3110 has undergone three rounds of mutations (Figure 5) [27]. The main purpose of these mutations was the development of new strains which could facilitate heterologous gene cloning and maintenance, and ultimately improve recombinant protein production [25]. Most of the chosen strains are derivatives of *E. coli* K-12 since the safety of these strains have been the most thoroughly investigated and hence is preferred by biosafety regulatory bodies, compared to others such as derivatives of *E. coli* B [24].

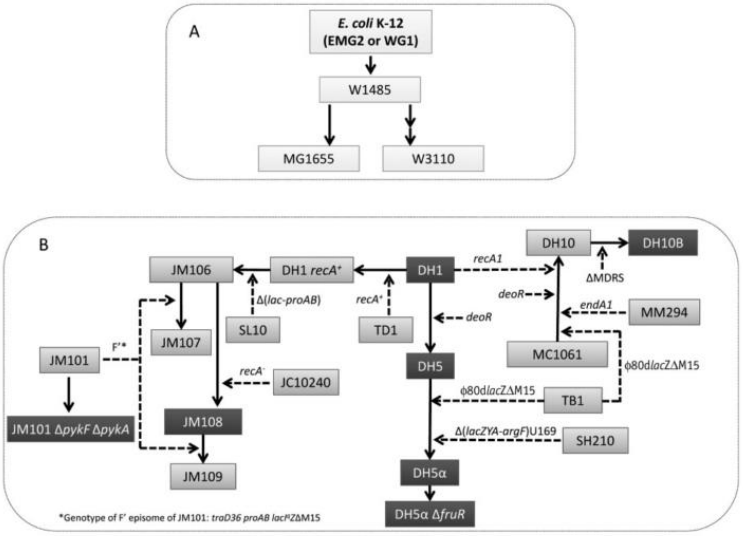


Figure 5 - *E. coli* K-12 and derivatives – creation of new strains and relationship between different strains. (A) Lineage of MG1655 and W3110, close relatives of wild-type *E. coli* K-12. (B) Generation of strains containing multiple mutations from MC1061, DH1, and JM101 [25].

Research on the use of wild-type K-12 strains such as MG1655 and W3110 is becoming more popular since their genomes have been completely sequenced and these strains are relatively free from cryptic mutations that are present in newer strain. Yau *et al.* [12] performed a study in which 17 strains were compared and it was observed that some of the existing commercial strains perform poorly compared to the older, less mutated ones. Also, Singer *et al.* [28] performed a study in which plasmid DNA production by 13 strains of *Escherichia coli* was compared. Shake flask studies indicated that BL21, SCS1-S, and SCS1-L strains generated the greatest DNA from glucose, while MC4100, JM105, and MG1655 generated the least [28].

However it is not possible to look at the genotype of a strain and draw accurate conclusions about the productivity and quality of plasmid DNA produced in that strain. Industrially relevant production of plasmid DNA depends on many factors including choice of plasmid, media components, fermentation type/strategy and downstream processing [28].

1.2.2.1 Host strain engineering

A particular strain used at laboratory scale may not necessarily be the best strain to support high yields of quality plasmid DNA at full scale production. Currently, plasmid DNA is often produced using strains of *Escherichia coli* and plasmid backbones that are known to be effective producers of recombinant proteins. However, these strains may not be the most favorable choice when plasmid DNA is the final product. In addition, it is known that many common laboratory strains like DH5α and DH10B have undergone a high degree of mutagenesis to improve their performance in cloning, library construction, and/or recombinant protein production applications [11]. The complex genotypes of many *E. coli* strains in industrial use today also greatly obstruct the ability to reliably predict the plasmid yield based on the genotype alone and makes it difficult to determine the impact of specific genetic variations [11].

A good demonstration of the relevance of strain genetic background is the study performed by Gonçalves *et al.* [29] in which single and double knockouts of genes *pykF* and *pykA* were performed in two different *E. coli* strains, MG1655 (wild-type genetic background) and DH5α (highly mutagenized genetic background). It was observed that these knockouts were only effective in the wild-type strain which demonstrates the relevance of strain genetic background and the importance to rationally design bacterial strains and plasmids specifically suited to the production of plasmid biopharmaceuticals [29].

Recently, many researchers have tried to develop well-characterized strains for pDNA production, using several strategies such as the ones outlined in Figure 6. These strain engineering efforts aim to improve plasmid stability, enhance product safety, increase yield, and/or facilitate downstream purification [11]. Using an alternative *E. coli* strain is important since it is known that the strain genetic background can have a big influence on plasmid production and in the outcome of strain engineering approaches [29]. But there are other strategies that can be used such as genome reduction. Pósfai *et al.* [30] performed a study in which researchers have constructed multiple-deletion series (MDS) strains of *E. coli*, containing a significantly reduced genome that is about 15% smaller

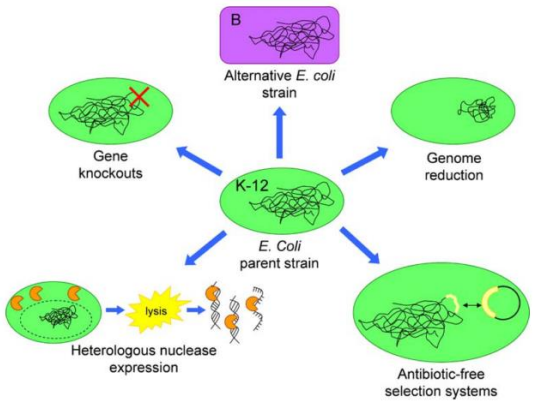


Figure 6 - Strain engineering strategies. The parent strain at the center of the diagram represents any *E. coli* K-12 strain, and the potential modifications are illustrated. Note that the engineering strategies are shown as independent of one another, but applying several strategies in parallel to a single strain is possible as well [11].

than that of the parental strain, MG1655. These strains have all of the mobile elements removed and showed no detectable transposon activity when compared to MG1655 and DH10B, therefore eliminating the main source of instability [30]. Compared to the parental strains, the reduced-genome strains showed identical growth rates and recombinant protein expression yields, but could be more efficiently transformed by electroporation and were able to propagate plasmids that were unstable in other strains [30].

Several researchers have addressed the concerns relative to antibiotic selection mechanisms by modifying the vector, host, or both to develop alternative plasmid selection systems. In particular, several researchers have chosen to manipulate essential *E. coli* genes to ensure efficient killing of plasmid-free cells. Cranenburgh *et al.* [31] decided to target *dapD*, an essential gene for diaminopimelate and lysine biosynthesis. The endogenous *dapD* locus was disrupted, and an ectopic copy of *dapD* under the control of a lac promoter was integrated into the chromosome. Transforming this strain with a high copy plasmid containing the lac operator resulted in sufficient titration of lac repressor away from *dapD* to give expression of the essential gene. As a result, only cells containing pDNA with the lac operator sequence survived in culture. Alternative selection systems based on antisense RNA have also recently been investigated. Mairhofer *et al.* [23] developed a system that uses the RNA I antisense transcript that is produced during ColE1 plasmid replication to regulate expression of an essential host gene. The results of the removal of the antibiotic resistance gene indicated an increase of the overall plasmid yield by 2-fold, compared to the conventional host/vector combination [23]. This result is possibly explained by a decrease in metabolic load and consequently an increased availability of metabolic precursors [23]. It is known that the major cause for metabolic burden exerted on the host cell, is the constitutive expression of the antibiotic resistance gene since the marker protein usually represents up to 20% of total cellular protein [23]. Also, the produced pDNA is 29% reduced in size, which is an additional advantage concerning transformation efficiency and therefore potency of the end product [23]. Finally, there is an increase in pDNA replication rate. Therefore, this study demonstrated that deleting the antibiotic resistance gene from the vector backbone is not only beneficial with regards to safety and potency of the end-product but also regarding the overall process performance [23].

To address removal of host genomic DNA as well as RNA, Nature Technology Corporation has developed *E. coli* hosts expressing periplasmic chimeric proteins that degrade both nucleic acids [32]. These chimeras include the plasmid-safe phage T5 D15 exonuclease linked to RNase A or S. This exonuclease specifically degrades linear and denatured DNA not affecting the fidelity of supercoiled plasmid DNA [11]. The chimeric proteins can be reintroduced into the cytoplasm before lysis by inner membrane permeabilization techniques or can begin to degrade host nucleic acids immediately after being released by cell lysis [11]. Both of these schemes significantly reduce the viscosity of the lysate, easing the later stages of purification. This approach to strain engineering also incorporated autolytic host strains, which have been used previously for recombinant protein production applications [11].

1.3 Effect of plasmid DNA synthesis on *E. coli* central carbon metabolism

Plasmid DNA synthesis can disturb *E. coli* gene regulation by altering levels of gene expression and carbon flux [25]. For the last three decades, many researchers observed that plasmid maintenance retards host growth, giving raise to lower biomass yields [25]. This metabolic burden can be caused by competition for cellular resources (energy, amino acids, etc.) between cell division and plasmid propagation activities, since plasmid replication and expression of the antibiotic resistance marker requires additional nutrients and energy [25], or by the perturbation of the *E. coli* host regulatory system affecting central metabolic pathways [33]. Because of the detrimental effect of this plasmid-imposed metabolic burden, cells that lose plasmid are able to grow faster during fermentation [33]. Over time, these non-productive cells become a significant fraction of the population contributing to problems such as plasmid loss and reduced productivity, which are of relevance to the industry [33].

However, the relation between plasmid DNA content and growth rate has not been fully explained since a direct comparison between plasmid-bearing and plasmid-free *E. coli* may be inappropriate [34]. Rozkov *et al.* [35] demonstrated that significant differences between a plasmid-free *E. coli* and the same host bearing a plasmid were due to expression of the kanamycin phosphotransferase gene, used as a plasmid marker. As mentioned before, the major cause for metabolic burden exerted on the host cell, is the constitutive expression of the antibiotic resistance gene since the marker protein usually represents up to 20% of total cellular protein [23]. Finally, it is also known that cultivation conditions affect plasmid DNA content significantly. Therefore, it appears that a direct comparison of plasmid-containing with plasmid-free cells may be inaccurate to analyse the metabolic limiting factors of plasmid copy number [34].

In spite of the problems related to the analysis of the effects of plasmid DNA production, recent reports have shown its effects on the central carbon metabolism of *E. coli*, namely in Glycolysis, the Tricarboxylic Acid Cycle and the Pentose Phosphate Pathway.

1.3.1 Effect of plasmid DNA synthesis on Glycolysis

The glycolytic pathway (Figure 7) is the main catabolic route of carbohydrates for the provision of energy and building precursors for biosynthesis [33]. Its functioning is essential under all conditions of growth because it produces seven (β-D-glucose 6-phosphate, β-D-fructofuranose 6-phosphate, dihydroxyacetone phosphate, 3-phospho-D-glycerate, glyceraldehyde-3-phosphate, phosphoenolpyruvate, and pyruvate) of the 13 precursor metabolites that are the starting materials for the biosynthesis of building blocks for macromolecules and other needed small molecules [36]. It oxidizes hexoses to generate ATP, reductants and pyruvate and it is also an amphibolic pathway (pathway that involves both catabolism and anabolism) [36]. Glucose and other sugars are fed into glycolysis to produce pyruvate from the common intermediate fructose diphosphate [33].

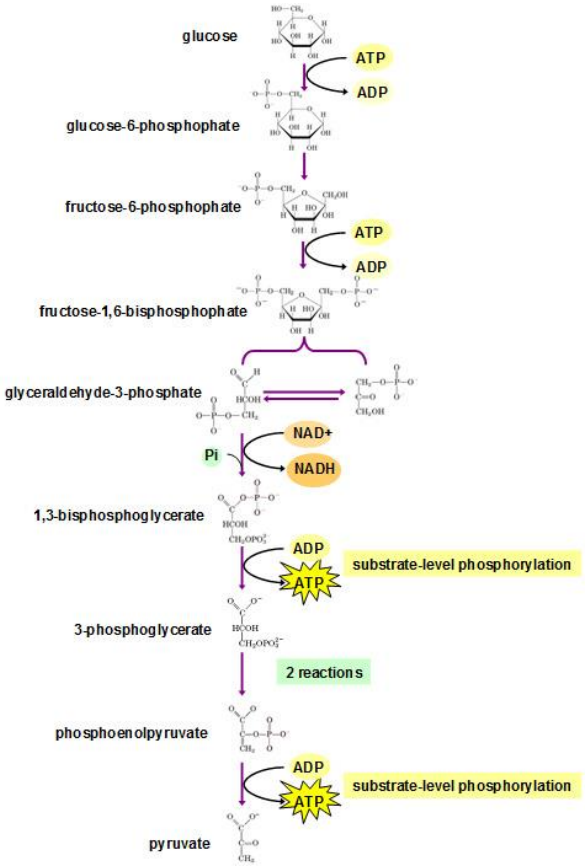


Figure 7 - The reactions of glycolysis convert glucose to pyruvate. ATP synthesis occurs in the reactions catalyzed by phosphoglycerate kinase and pyruvate kinase. These reactions are examples of substrate-level phosphorylation and are highlighted in yellow [37].

Previous studies reported different levels of glycolytic gene expression in plasmid-bearing cells versus plasmid-free cells [25]. Some results showed most of the glycolytic genes as down regulated [33], while other results have demonstrated up-regulation of the same genes, such as pyruvate kinase I (*pykF*) for cells carrying plasmids [34].

Ow *et al.* [33] demonstrated that in plasmid-bearing cells, except for *gapA* (encoding glyceraldehyde-phosphate dehydrogenase), most glycolytic genes were generally observed to be downregulated. Control of glycolysis resides in two key steps catalyzed by phosphofructokinase (Pfk) and pyruvate kinase (Pyk), which are the only two irreversible steps in glycolysis [33]. Two isoenzyme genes of phosphofructokinase (*pfkA* and *pfkB*), which essentially control the entry of substrates into glycolysis, were significantly downregulated by more than 1.5-fold [33]. The *pykF* gene encoding the enzyme catalyzing the final unidirectional step out of glycolysis was also downregulated [33]. However, no differential regulation for *pykA* was found. This is probably due to the fact that they are controlled by different transcription regulators; for example, FruR represses expression of *pykF*, but not *pykA* [33].

The down regulation of glycolytic enzymes suggests a reduction in glycolytic flux in the plasmid-bearing cells. Therefore, we can predict that any extra carbon flux arising from the reduced glycolytic flux might be redirected into the PP pathway [33].

On the contrary, Wang *et al*. [34] discovered that some genes involved in glucose transport (*ptsG*) as well as genes involved in glycolysis (*fba*, *gapA*, *eno, pykF*) were significantly up-regulated in plasmid-bearing cells relative to plasmid-free cells. It was also observed that the levels of glycolytic intermediates, such as glucose 6-phosphate, fructose 6-phosphate, glyceraldehyde-3-phosphate, phosphoglycolate phosphatase, 3-phosphoglycerate, 2-Phosphoglycerate, phosphoenolpyruvate and pyruvate were significantly increased in plasmid-bearing cells, during the early stationary phase [34].

1.3.2 Effect of plasmid DNA synthesis on the Pentose Phosphate (PP) pathway

Pentose Phosphate (PP) pathway is considered the second main destination for glucose (Figure 8), after Glycolysis. Its functions are: (a) the catabolism of carbon sources including sugars that cannot be utilized by other pathways like xylose and ribose, (b) provision of reducing power (NADPH) for biosynthesis and (c) biosynthesis of nucleic acids, amino acids, vitamins and ingredients of the cell lipopolysaccharide layer [33]. It can metabolize different sugars like xylose and ribose, and it is one of the pathways responsible for biosynthesis of the nucleotide-precursors ribose-5-phosphate (R5P) and erythrose-4-phosphate (E4P) [25]. This pathway comprises two phases: the oxidative generation of NADPH and the nonoxidative interconversion of sugars [38]. In the oxidative phase, NADPH is generated when Glucose 6-phosphate is oxidized to Ribose 5-phosphate (R5P) [38]. This sugar and its derivatives are components of RNA and DNA, as well as ATP, NADH, FAD, and coenzyme A [38]. In the nonoxidative phase, the pathway catalyzes the interconversion of different sugars in a series of nonoxidative reactions that can result in the synthesis of five-carbon sugars for nucleotide biosynthesis or the conversion of excess five-carbon sugars into intermediates of the glycolytic pathway [38].

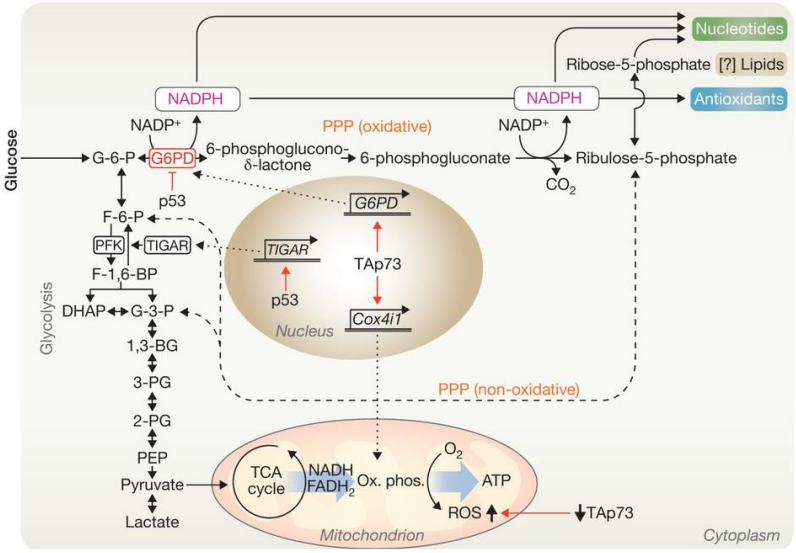


Figure 8 - The Pentose Phosphate (PP) Pathway [39].

It is known that the expression of many PP pathway enzymes is constitutive and not transcriptionally regulated, with only a few exceptions. Ow *et al*. [33] discovered that the expression of PP pathway genes remains constant despite changes in glycolytic flux. Therefore it is hypothesized that PP genes are steadily expressed at levels that can support multifold changes in metabolic flux. Only a few genes (*zwf*, *talA* and *tktB*), were differentially regulated in plasmid-bearing cells when comparing with plasmid-free cells [33]. Regarding *zwf* gene it was found to be 1.2 fold up-regulated while *talA* and *tktB* genes were found to be 1.5 fold down-regulated [33]. The *zwf* gene encodes for the first enzyme of the oxidative PP pathway (glucose-6-phosphate dehydrogenase) and it is subjected to transcriptional regulation by specific proteins [33]. For *talA* and *tktB*, these two genes are adjacent to each other and are transcribed from the same promoter [33].

On the contrary, Wang *et al.* [34] observed that the *zwf* and *gnd* genes were significantly down-regulated in early stationary in plasmid-bearing cells. It was also observed that the 6PGlac flux of plasmid-bearing cells decreased relative to that estimated in plasmid-free cells [34]. During the early stationary phase, levels of PP pathway intermediates, such as 6PGlac, 6PG, Ru5P, X5P, R5P, E4P, S7P were also significantly decreased in plasmid-bearing cells [34]. Furthermore, expression of the *rpiA* gene, encoding RPI enzyme, was significantly decreased in cells carrying a high-copy number plasmid relative to cells carrying a low-copy number plasmid [34].

1.3.3 Effect of plasmid DNA synthesis on the TriCarboxylic Acid (TCA) cycle

The tricarboxylic acid (TCA) cycle (Figure 9) is composed of eight reactions that oxidize the acetyl group from acetyl-CoA or from other sources [25]. This cycle is important in energy metabolism and biosynthesis and performs the complete oxidization of acetyl coenzyme A derived from glycolysis [33]. Also, several TCA intermediates are needed during amino acid, pyrimidine and heme biosynthesis [33]. Oxaloacetate (OAA) and α-ketoglutarate (AKG) are two TCA intermediates that play an important role in amino acids synthesis [25].

Most of the tricarboxylic acid cycle genes are up-regulated (including: *gltA*, *acnA*, *acnB, aceB, icdA*, *sucA*, *sucB*, *sucC*, *sucD*, *sdhC*, *fumA*, *fumB*, *fumC*) in plasmid-bearing cells relative to the plasmid-free strain [34].

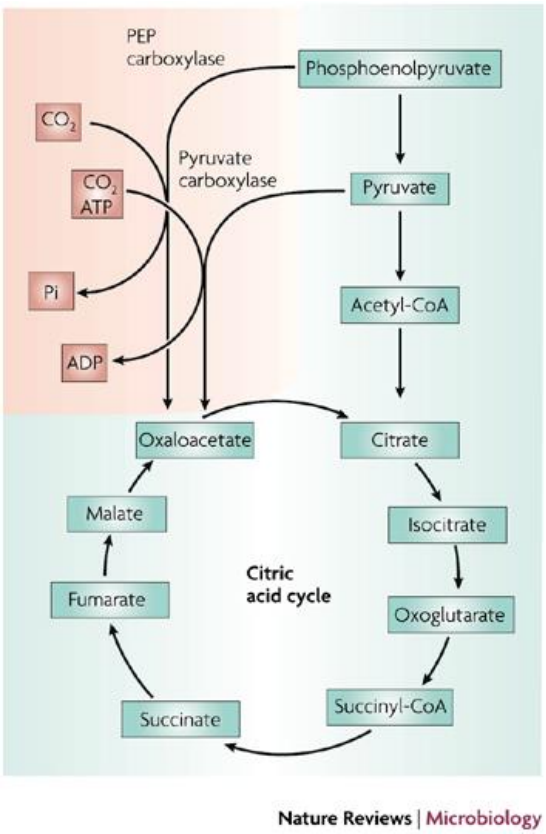


Figure 9 - The TCA cycle [40].

1.4 Metabolic engineering of *Escherichia coli* strains for plasmid DNA biopharmaceutical production

Metabolic engineering is the science that combines systematic analysis of metabolic and other pathways with molecular biological techniques to improve cellular properties by designing and implementing rational genetic modifications [41]. As such, metabolic engineering deals with the measurement of metabolic fluxes and elucidation of their control as determinants of metabolic function and cell physiology [41]. It involves manipulation of enzymatic, transport and regulatory functions of the cell by using recombinant DNA technology. Through the use of analytical techniques, metabolic fluxes of critical metabolic pathways in the cell or tissue of interest are determined. After that, molecular biological techniques are applied in order to enhance metabolic flux through that pathway of interest and/or minimize metabolic flow to undesired biosynthetically related products [41].

The development of recombinant DNA techniques has introduced a new dimension to pathway manipulation by offering, for the first time, the capability to construct specific metabolic configurations with novel, beneficial characteristics. Genetic engineering allows precise modification of specific enzymatic reactions in metabolic pathways, leading to the construction of well-defined genetic backgrounds. The redirection of cellular metabolism to create or enhance desirable attributes has been accomplished with a variety of novel techniques and applied towards a variety of goals [41].

Metabolic engineering is well suited for several biotechnological applications, especially in microbial fermentation. It can be used to design rational strategies for target selection for screening candidate drugs or designing gene therapies. It has also been applied to increase the production of chemicals that are already produced by the host organism, to produce desired chemical substances from less expensive feedstocks, and to generate products that are new to the host organism [41]. Other challenges associated with metabolic engineering are the biosynthesis of secondary metabolites, the generation of organisms with desirable growth characteristics, and the manipulation of pathways for the production of chiral compounds as intermediates in the synthesis of pharmaceutical products [41].

Currently, both in academia and industry there is the need for a technological and economical improvement in pDNA manufacturing [25]. Significant progress has been achieved in plasmid design and downstream processing, but there is still a demand for improved production strains [25]. This chapter will focus on engineering of *E. coli* strains for pDNA production, mainly on the gene mutations that can have an impact on pDNA biosynthesis.

1.4.1 Relevant genes for *E. coli* strain engineering

Many cell line engineering efforts have been made in order to improve plasmid DNA production by knockout or overexpression of rationally-selected genes (Table 4) [25]. One main area of focus is the modification of central carbon metabolism genes to increase flux toward nucleotide and amino acid precursor synthesis and reduction of byproducts, such as acetate [25]. Genes related to improving pDNA quality have also been common targets, as well as genes that are involved in various other cellular processes relevant to pDNA production, for instance genes involved in the stringent response and DNA replication [25]. In the next sections the main gene mutations that have recently been made will be discussed.

Table 4 - *E. coli* genes targeted for mutation to improve plasmid DNA production. Δ indicates gene knockout and ↑ indicates gene overexpression [25].

Gene(s)	Product(s)	Mutation	Expected phenotype
recA	DNA strand exchange and recombination protein; protease and nuclease activity	Δ	Minimized recombination of cloned DNA, pairing and exchange between repeated DNA sequences
endA	DNA-specific endonuclease I	Δ	Decreased non-specific digestion of plasmid
rpiA	Ribose-5-phosphate isomerase A	↑	Increased biosynthesis of nucleotide precursors in pentose phosphate pathway
pykA pykF	Pyruvate kinase I, II	Δ	Increased pentose phosphate pathway and TCA cycle flux and reduced acetate synthesis
topA	Topoisomerase I	Δ	Increased RNAII R-loop formation (ColE1 plasmids)
polA	DNA polymerase I	↑	Increased availability of enzyme that extends the RNA primer template and removes RNA primers postreplication
ligA	DNA ligase	↑	Increased availability of enzyme that seals nicks during pDNA synthesis
gyrAB	DNA gyrase	↑	Increased availability of enzyme that negatively supercoils the covalently closed circular plasmid
priA priB priC	Primosome components	↑	Increased availability of primosomes (one lagging strand primosome is required per plasmid per replication)
trxA grx1	Thioredoxin, glutaredoxin	↑	Increased level of hydrogen donors for ribo-nucleotide reductase
zwf	Glucose 6-phosphate-1-dehydrogenase	↑	Increased pentose phosphate pathway flux
fruR	Transcriptional regulator	Δ	Deletion of global regulator of carbon flow through the central metabolic pathways
relA spoT	ppGpp synthetase I, II	Δ	Abolished stringent response to amino acid starvation
dcm	DNA-cytosine methyltransferase	Δ	Improved transgene expression and reduced immunogenicity

1.4.1.1 Genes related to plasmid properties

1.4.1.1.1 *endA* gene

The *endA* gene encodes DNA-specific endonuclease 1. This protein is a periplasmic enzyme that cleaves within duplex DNA no shorter than 7 bp [42]. Its cellular role is unknown, but it has been widely used as an experimental tool for DNA footprinting [42]. It is inhibited by tRNA and rRNA [42]. The deletion of this gene can improve the quality and final yields of plasmid preparations by eliminating non-specific degradation of pDNA by the endonuclease [25]. Mutants lacking *endA* show no apparent difference from wild type cells [42].

1.4.1.1.2 *recA* gene

recA gene codes for a protein, Recombinase A, essential for the *recBCD* pathway of homologous recombination, being necessary for the repair and maintenance of DNA. Generally, constitutive levels of RecA are maintained at between 1,000 and 10,000 monomers per cell but when induced by DNA damage, e.g. UV irradiation, it increases to 50 times its normal level [43].

RecA performs homologous recombination by catalyzing DNA strand exchange reaction. It serves as a regulatory protein to induce the SOS response to DNA damage by its action on the repressor protein LexA [43]. Activated RecA becomes a regulator of the SOS response by inducing the autocatalytic cleavage of LexA, which generally binds to the operator region of most of the DNA damage-inducible genes and suppresses their expression. As the cellular concentration of LexA decreases, the SOS genes become more frequently transcribed [43]. It also plays a role in the repair of Double strand breaks (DSBs) in which a separate intact region of DNA is used [43]. RecA polymerizes to form filaments around both single-strand and double-strand DNA and in doing so extends the DNA within the filament to 1.5 times its unbound length [43].

recA mutants, have less undesirable homologous recombination than wild-type cells. Homologous recombination can lead to both changes in the plasmid DNA as well as formation of plasmid multimers, which causes an increase in plasmid-free cells [25]. Phue *et al.* [44] performed a study in which BL21*ΔrecA* and BL21*ΔrecAΔendA* demonstrated to be superior producers of pDNA when compared to DH5α. Using glucose as the carbon source, BL21-derived strains showed significantly less acetate production and improved glucose utilization, as expected for a B strain [44].

1.4.1.1.3 *dcm* gene

The *dcm* gene codes for a methyltransferase present in *E. coli*, the DNA-cytosine methyltransferase (Dcm). DNA methylation consists of the addition of a methyl group to a base by a DNA methyltransferase enzyme [25]. Binding proteins can have high affinity to methylated DNA sequences and methylation sites may affect promoter activity [25]. Methylation patterns can also allow a species to distinguish its own DNA from foreign DNA [25].

Studies have shown that *Δdcm* mutant cells do not have a strong impact on plasmid yield and quality [45]. Extensive quality control analysis also demonstrated that dcm^- and dcm^+ plasmids were of similar high quality for a variety of quality attributes such as RNA primer removal, percentage of basic residues, and percentage of supercoiling [45].

On the other hand, plasmids produced in these dcm^- cells had a high transgene expression level in a human cell line and would therefore be recommended for gene therapy applications [45]. However, these plasmids also demonstrated to be less immunogenic and would not be ideal for DNA vaccine applications [45]. One hypothesis for reduced immunogenicity is that Dcm methylation sites are a pattern recognized by the innate immune system [45]. In conclusion, *dcm* methylation status affects plasmid-directed transgene expression and immunogenicity but not plasmid production [45]. Therefore, dcm^+ plasmid is recommended for DNA vaccines and dcm^- plasmid for DNA therapeutics and cell transfection reagents [45].

1.4.1.2 Central carbon metabolism genes

Central carbon metabolism is a logical target for genetic engineering strategies to increase plasmid DNA yield because of the potential to increase carbon flux to nucleotide precursors (Figure 10) [25]. Altering central metabolism to produce more energy and reducing power could also potentially improve plasmid yields [25]. Therefore, several changes have been made to these pathways (Figure 10).

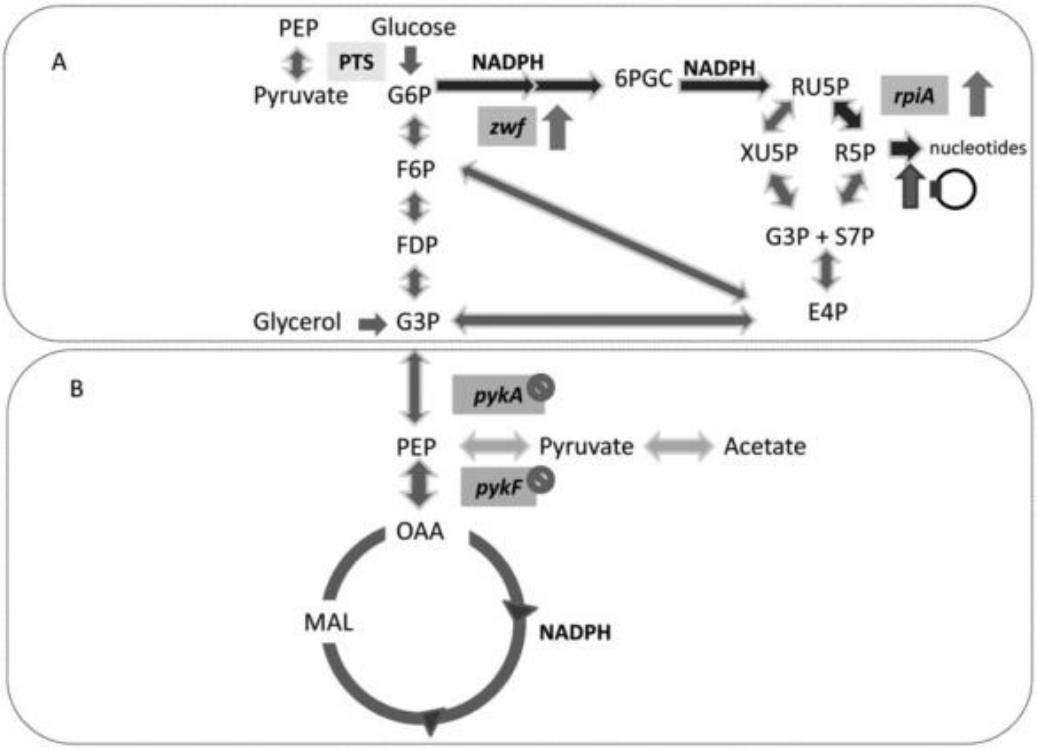


Figure 10 - Gene knockout and overexpression strategies to improve plasmid DNA production in *E. coli.* (A) Overexpression of the *zwf* and *rpiA* genes is proposed as a means to increase fluxes in the pentose phosphate pathway and nucleotide synthesis. (B) Knockouts of genes *pykF* and *pykA* are suggested to reduce acetate formation, increase fluxes in the pentose phosphate pathway and TCA cycle. Dark arrows represent high carbon flow in pentose phosphate pathway and light arrows represent less formation of pyruvate and acetate [25].

1.4.1.2.1 *pykF* and *pykA*

The *pykF* and *pykA* genes encode pyruvate kinase (Pyk) I and II, respectively, and the production of these two isoenzymes is independent [32]. However, both have a cooperative effect on transforming phosphoenolpyruvate (PEP) into pyruvate (PYR) at the final stage of glycolysis (Figure 10) [25].

Acetate secretion deviates carbon flux away from nucleotide synthesis and therefore it is disadvantageous for plasmid DNA production [25]. In addition, its accumulation in the medium reduces

cell growth and limits maximum cell concentration [46]. This reduction is due to the fact that the homeostatic mechanisms of *E. coli* require energy to adjust to the decrease in intracellular pH caused by acetate accumulation [46]. Thus for rapid growth, *E. coli* apparently requires maintenance of an optimum intracellular pH [46].

Luli *et al*. [46] developed a growth model based on the data obtained for acetate inhibition of growth rate at pH 7.0. The data obtained suggest that acetate is inhibitory very early in the growth of the culture, but that the inhibitory effect is not visualized until acetate has accumulated to a concentration of 5 g/L [46].

However, the exact effect of acetate on plasmid production is not yet clear. For instance, Wang *et al.* [34] showed that plasmid-bearing *E. coli* BL21 cells produce more acetate than plasmid-free cells. However, Carnes *et al.* [45] determined that, in contrast to recombinant protein production, secretion of this metabolic byproduct into the media had only a minor negative effect on plasmid replication.

Given its negative effect, it is important to diminish the accumulation of acetate by choosing host strains and growth conditions which minimize its production. Several strategies have been performed with the goal of increasing plasmid DNA production. For instance, Al Zaid Siddiquee *et al*. [47] performed a batch cultivation using both wild type *E. coli* K-12 and a *pykF*⁻ mutant. In batch culture, it was found that the *pykF* gene knockout has little effect on cell growth [47]. However, it was also found that flux through phosphoenolpyruvate carboxylase and malic enzyme was up-regulated in the *pykF*⁻ mutant as compared with the wild type, and acetate formation was significantly reduced [47]. The reduction of acetate production was expected since the synthesis of pyruvate would decrease, and phosphoenolpyruvate would be converted to oxaloacetate (OAA) without pyruvate kinase activity [25].

Cunningham *et al.* [48] discovered that an *E. coli pykF pykA* (PB25 strain) double knockout produces much less acetate than JM101, wild-type strain. It also exhibits a modest 15 to 25% reduction in growth rate, indicating that the isozymes are not essential for growth on glucose [48]. More importantly, when expressing GFPuv (pUC-based, high-copy-number plasmid), PB25 has threefold more plasmid than JM101 [48].

Gonçalves *et al*. [29] performed a study in which it was investigated these same mutations but in the nearly wild-type MG1655*ΔendAΔrecA* and in the commonly used DH5α strain using either glycerol or glucose as the primary carbon source. MG1655*ΔendAΔrecA* and DH5α grew faster than their respective *pykF* and *pykA* double knockout mutant strains in both glucose and glycerol [29]. However, GALG1011 (MG1655*ΔendAΔrecAΔpykFΔpykA*), produced less acetate than MG1655*ΔendAΔrecA* in both carbon sources. Also, strains with single mutations (*pykF* or *pykA)* demonstrated to be more efficient in producing pDNA than strains containing the double mutation [29]. The synthesis of pDNA was likely more advantageous when the pyruvate kinase activity was reduced, as opposed to eliminated, because glycolytic flux, although diminished, still enables additional ATP production and up-regulation of glucose flux through the Pentose Phosphate Pathway [29]. At the same time, carbon flux would be increased to the pentose phosphate pathway, improving nucleotide synthesis.

These results show that the outcome of rationally designed mutations is highly dependent on the strain's genetic background of the cells. Different strains with the same gene knockout can have diverse carbon fluxes and result in different ATP yields, favoring pDNA production or not.

1.4.1.2.2 *rpiA* gene

rpiA gene codes for ribose-5-phosphate isomerase A. There are two physically and genetically distinct ribose-5-phosphate isomerases present in *E. coli*. The constitutive ribose-5-phosphate isomerase A (RpiA) normally accounts for more than 99% of the ribose-5-phosphate isomerase activity in the cell [49]. This is a reversible enzyme that converts Ru5P into R5P, decreasing the flux into X5P (Figure 10). Therefore, a large part of the flux into Ru5P is used for synthesis of nucleotides.

Theoretically, overexpression of *rpiA* would enhance synthesis of the nucleotide precursor ribose-5-phosphate (R5P), which would lead to an increase in nucleotide formation, and consequently, plasmid DNA production. Wang *et al.* [34] discovered that the Rpi enzyme is deprived when *E. coli* carries a high copy number plasmid which in turn causes a decrease of the plasmid copy number. From the metabolic flux analysis, enzyme activity analysis and mRNA expression profiles, it was concluded that the *rpiA* gene is one of the important limiting factors for plasmid DNA replication [34]. Therefore, the authors performed overexpression of the *rpiA* gene in *E. coli* BL21. This event caused a 3-fold increase in plasmid copy number of a ColE1-derived plasmid during continuous culture, using defined medium and glucose as the carbon source [34].

In a more recent work, a cultivation was performed in order to obtain a deeper understanding of plasmid replication in GALG20 (MG1655*ΔpgiΔendAΔrecA)* strains overexpressing the *rpiA* gene [50]. The results indicated that plasmid amplification rate is increased in the new strains, more specifically until 14 hours of cultivation [50]. Increased production of Ribose-5-phosphate Isomerase (RpiA) allowed a faster plasmid replication until a certain point, where it became similar to GAL20 [50].

1.4.1.2.3 *zwf* gene

zwf gene codes for glucose-6-phosphate dehydrogenase (G6PDH) which is an enzyme responsible for the conversion of D-glucose 6-phosphate and $NADP^+$ into 6-phospho-D-glucono-1,5-lactone and NADPH (Figure 10). It is known that 22% of the transported glucose in the wild type *E. coli* strain is driven through glucose-6-phosphate dehydrogenase, Zwf, and part of this carbon flux is devoted to nucleotide, amino acid, and vitamin biosynthesis [51]. NADPH and nucleotides are required for biomass and pDNA production and they are intrinsically correlated in the PP pathway, composing the oxidative and non-oxidative phases, respectively [25]. Cunningham *et al.* [52] demonstrated through a mathematical model that increasing the availability of NADPH has a positive impact on pDNA production by increasing the reducing power available for pDNA.

It was discovered that the *zwf* gene knockout had a minor effect on the cell growth [53]. However, the profiles of the substrate and by-product concentration indicated that the corresponding

metabolic rates were changed due to the gene knockout [53]. The *zwf* mutant directed 98.9% and 87.0% of the total carbon flux through the first step of glycolysis (from G6P to F6P) and TCA cycle (from AcA to CIT),respectively, whereas the parent strain showed an obviously lower flux through the first step of the glycolysis (78.6%) and TCA cycle (73.1%) [53].

Overexpression of *zwf* has been investigated as a strategy to increase flux to the pentose phosphate pathway. Williams *et al.* [54] verified that simultaneous overexpression of *zwf* and *rpiA* in DH5α appeared to increase plasmid amplification rate (mg pDNA/L/OD_{600}/hr) but not final specific yield (mg/L/OD_{600}), in fed-batch fermentations using complex medium with glycerol as the carbon source. It appears that fermentation plasmid copy number (specific yield at plateau) within DH5α is largely set by process and vector-intrinsic factors [54]. While the rate of plasmid replication can be influenced by increasing levels of key replication proteins, the final copy number plateau is insensitive to such alterations [54]. In the same study, the overexpression of *zwf* alone as well as *zwf* and thioredoxin (*trxA)* did not show any effect on fermentation productivity [54].

In another study, transforming a wild-type strain with a high-copy number plasmid carrying the *zwf* gene, under the control of an inducible *trc* promoter, resulted in a recovery of the growth-rate from 0.46 h^{-1} (uninduced) to 0.64 h^{-1} (induced), decreasing plasmid-mediated metabolic burden [51]. These results indicate that engineering metabolic flux at the level of the G6P node can be used as a strategy to compensate for special metabolic demands of bacterial production strains [51].

1.4.1.2.4 *fruR* gene

fruR gene codes for a Fructose repressor (also known as Cra, Catabolite repressor activator) and is a global regulator that directs carbon flow through the central metabolic pathways via its influence on the transcription of carbon and energy metabolism genes [55]. Primarily, transcription of catabolic enzymes involved in the glycolytic pathway (*pfkA, pykF, gapA, pgk, eno*), glycolysis (*edd, eda*) and alternative sugar catabolism (*fruBKA, mtlADR*) are repressed, while enzymes of glyconeogenesis (*fbp, ppsA*), glyoxalate shunt (*aceBA*), TCA cycle (*acnA, icdA*) and electron transport chain (*cydAB*) are positively activated [55].

Ow *et al.* [55] evaluated the effect of a *fruR* knockout during batch fermentation. The results demonstrated a higher specific growth rate of 0.91 h^{-1} for the plasmid-bearing *fruR*$^-$ cells compared to its parental plasmid-bearing wildtype cells (0.75 h^{-1}) [55]. Also, in plasmid-bearing *fruR*$^-$ cells, expression of enzymes involved in sugar catabolism, glycolysis and transcription/translation processes were upregulated, while those related to gluconeogenesis, tricarboxylic acid cycle and stress response were downregulated [55]. It was also observed that the glycolytic genes *pykF*, *zwf*, and 6-phosphofructokinase I and II (*pfkA, pfkB*) were up-regulated when *fruR* was deleted [55]. However, this knockout did not increase plasmid final yield [55]. These results demonstrate that the inactivation of FruR global regulator in recombinant *E. coli* alters metabolic gene expression and significantly reduces growth retardation from the burden of maintaining a plasmid [55]. Later, Ow *et al.* [56] demonstrated that the same *fruR*-deficient DH5α strain improved pDNA yield under exponential

feeding in fed-batch conditions. Therefore it is possible to conclude that fermentation strategy directly affects plasmid amplification behavior.

1.4.1.2.5 *pgi* gene

pgi gene codes for a Phosphoglucose isomerase, which catalyzes the interconversion of glucose-6-phosphate and fructose-6-phosphate, an essential step of the glycolysis and gluconeogenesis pathways [57].

A recent study has demonstrated that the knockout of the *pgi* gene redirects glycolytic flux, increasing fluxes in the pentose phosphate pathway and enhancing nucleotide synthesis and NADPH production [29]. Glycolysis is down-regulated, but proceeds via the formation of fructose 6-phosphate and glyceraldehyde 3-phosphate [29]. GALG20 (MG1655*ΔendAΔrecAΔpgi*) produced three-fold more pVAX1GFP pDNA (11.6 mg/g DCW) than MG1655*ΔendAΔrecA* in 5+10 g/L glucose (3.6 mg/g DCW) [29]. In glycerol, GALG20 produced similar amounts of pDNA to MG1655*ΔendAΔrecA*. Surprisingly, this strain had the best performance in 20 g/L of glucose, producing 25-fold more pDNA (19.1 mg/g DCW) than the parental strain (0.8 mg/g DCW) [29]. Given these results, it is possible to conclude that the *pgi* gene is an important target to redirect carbon flux to the pentose phosphate pathway, in order to enhance nucleotide synthesis and consequently increase pDNA production, when using Glucose.

1.4.1.3 Glucose transport system (PTS)

The phosphotransferase transport system (PTS) is one of the mechanisms by which pyruvate is produced in *E. coli* growing aerobically in high glucose concentrations [58]. De Anda *et al.* [58], studied the effect of replacing the glucose phosphotransferase transport system (PTS) with an alternate glucose transport activity on growth kinetics, acetate accumulation and production of recombinant proteins. The resulting strain was grown in batch cultures and the results indicated that this engineered strain produced almost four-fold higher recombinant protein concentration relative to the wild-type strain [58]. Also, maximum acetate concentration reached only 0.39 g/L for the engineered strain, whereas a maximum of 2.8g/L accumulated in the wild-type strain [58]. The engineered strain also maintained similar production and growth rate capabilities, relative to the wild-type strain [58].

1.4.1.4 Nucleotide synthesis

Besides altering central carbon metabolism genes it is also important to improve the nucleotide availability in pDNA producing strains. Some genes have already been tested, such as *deoR* and *nupG*.

deoR codes for a transcriptional repressor, DeoR (Deoxyribose Regulator), which is involved in the negative expression of several genes of the deo operon that code for enzymes needed for

deoxynucleotide synthesis [59]. Therefore, it was expected that deoR mutants could produce more pDNA.

nupG codes for NupG protein which is a member of the major facilitator superfamily (MFS) of transporters and presumably functions as a nucleoside/proton symporter, participating in the regulation of genes involved in nucleotide synthesis [60]. Analysis of *nupG-lacZ* fusions has shown that expression of *nupG* is regulated by the CytR and DeoR transcriptional regulators [60]. Imported nucleosides serve as precursors of DNA and RNA, as well as of histidine and various co-factors [60]. It has been shown that *nupG* mutants can produce significantly more purine nucleosides than parental strains [61].

Given the relation between these two genes, single and double mutants were analyzed. Borja *et al.* [59] performed a study in which the genes *endA*, *recA*, *deoR* and *nupG* were inactivated in strain VH33 independently and in combination, in order to increase pDNA yields. It was expected that *deoR* mutants would produce more pDNA. When *nupG* was deleted in VH33, no change in pDNA production was observed as well as when both *deoR* and *nupG* mutations were combined [59]. On the other hand, the triple mutant VH33 (*ΔrecAΔdeoRΔnupG*) demonstrated to be the best. It accumulated low amounts of acetate (2.2 ± 0.1 g/L) and resulted in the best product yield (186 ± 15 mg/L), the highest ever reported for a batch cultivation of *E.coli*, maintaining a supercoiled fraction close to 80% [59]. This indicates that not only the deletion of genes involved in nucleotide catabolism are necessary to increase pDNA production, but also increasing the plasmid stability is needed to see a positive effect [59].

1.4.1.5 Stringent response

Stringent response has been studied in the last four decades, but many of the cellular mechanisms involved are still unclear. Studies have shown that the accumulation of unusual guanosine nucleotides, collectively called (p)ppGpp (guanosine-5´-diphosphate-3´-diphosphate), is the hallmark of the stringent response of *E. coli* [62]. This unusual nucleotide is produced not only in response to amino acid limitation but also in response to many different kinds of nutrient limitations and circumstances that cause growth arrest [63]. Such accumulation is known to be controlled by the activity of two enzymes, the ribosome-bound RelA enzyme (ppGpp synthetase I) that synthesises (p)ppGpp nucleotides upon the depletion of amino acids and the bifunctional SpoT enzyme (ppGpp synthetase II) that is responsible for maintaining the intracellular levels of (p)ppGpp nucleotides via enzymatic degradation [62]. The stringent response involves the control of the genetic expression by direct interaction of the (p)ppGpp nucleotides with the RNA polymerase (RNAP), activating the transcription of genes coding for stress-associated sigma factors and amino acid biosynthesis at the expense of those required for growth and proliferation, thus affecting bacterial chromosome and plasmid origin replication [63].

A strain lacking both RelA and SpoT is completely unable to produce ppGpp and the response to starvation of such a strain is called the relaxed response [63]. It was observed that plasmid yield is generally improved in *relA*$^-$ strains due to the relaxed response to amino acid starvation that allows plasmid replication to continue [64]. The isoleucine addition seems to enhance this effect in *relA*$^-$

strains [64]. The fermentation with isoleucine reached 633 mg/L and the one without isoleucine reached 398 mg/L [64]. Other mutations that eliminate the stringent response, such as *spoT*, could also be used in place of *relA*.

1.4.1.6 Other targets for genetic engineering

Faced with a plateau for plasmid replication, Williams *et al.* [54] hypothesized that it was probably determined by the efficiency of the replication origin and the percent of initiated replication cycles that are completed. The amplification slope may be set by a limiting factor, perhaps a protein or nucleotide necessary for plasmid DNA replication [54]. Given that, the genes responsible for these mechanisms have been analyzed, such as genes involved in DNA replication.

topA (Topoisomerase I)*, polA* (DNA Polymerase I)*, ligA (*DNA ligase) and *gyrAB* (DNA gyrase) are some of the genes that have been tested by overexpression or knockout. For instance, William *et al.* [54] demonstrated that overexpression of *polA* and *ligA* genes improved pDNA yields in shake-flask cultures, but not in the bioreactor. Moreover, it was also observed that *gyrAB* overexpression had a negative effect on pDNA production, decreasing the yield at both shake-flask and bioreactor scales [54]. However, the importance of these genes for pDNA production is still unclear.

1.5 Fermentation strategy

Besides the strain engineering process, the fermentation strategy should be designed carefully to best suit the requirements of the new strain. The main challenge for cell culture is to maximize volumetric (mg pDNA/L culture broth) and specific (mg pDNA/g dry cell weight) yields while at the same time assuring the integrity of the primary sequence of the biosynthesized pDNA (i.e. absence of deletions and insertion sequences) and the predominance of the supercoiled topology [65]. In order to that, fermentation media and process needs to be carefully optimized considering some key factors, which will be discussed in this chapter.

1.5.1 Culture medium and Carbon source

Media composition can dramatically affect yields and consequently the overall cost of production. Options vary greatly, including minimal (defined) media to complex (semi-defined). Complex media formulations often contain ingredients like yeast extract, peptones, and other growth factors that may allow for a higher cell density, but may present challenges with reproducibility and with contaminant removal in downstream processing [66]. Minimal media contains known quantities of essential nutritional components including a carbon source, a nitrogen source, and salts, and excludes components known to be inhibitory to bacterial growth. Fermentation processes using minimal media are highly reproducible and plasmid copy number may even be higher [66].

A carbon source provides energy and biomass and is usually the limiting nutrient in cultures. This source is commonly included in the highest amounts [67]. Glucose is the conventional carbon

source as it is inexpensive and metabolized very efficiently and therefore gives a higher cellular yield [66]. However, it was previously demonstrated that high concentrations of glucose inhibited pDNA synthesis by MG1655*ΔendAΔrecA* and DH5α in shake flask experiments [29]. In aerobic conditions, acetate production occurs when high concentrations of glucose inhibit respiration (Crabtree effect) [29]. Under these conditions, around 15% of the glucose is converted in acetate [29]. Actually, when using high concentrations of glucose (20 g/L), high amounts of acetate were formed (5.74 and 4.72 g/L, MG1655*ΔendAΔrecA* and DH5α, respectively) and less pDNA was produced resulting in a volumetric and specific plasmid yield of approximately 1.5 mg/L and 0.8 mg/g DCW, respectively, for both MG1655*ΔendAΔrecA* and DH5α [29]. In contrast, when low amounts of glucose were used (5+10 g/L), higher amounts of pDNA were obtained for both strains. MG1655*ΔendAΔrecA* achieved a volumetric yield of 27.4mg/L and a specific yield of 3.6 mg/g DCW, while DH5α yielded 9.6 mg/L and 1.8 mg/g DCW. Additionally, null or minimum acetate were produced, for MG1655*ΔendAΔrecA* and DH5α, respectively, in the end of 24 hours [29]. Thus, acetate seems to play an important role in pDNA production.

Glycerol can also be used as a carbon source. It is well known that glycerol has the advantage of minimizing acetate formation in fermentation processes and it is becoming an inexpensive and attractive carbon source [29]. Zhi-nan *et al*. [68] performed a study in which the use of glycerol (4.0 g/L) resulted in highest specific plasmid productivity (37.9 mg DNA/g DCW), but poor cell growth, and delayed the peak time for plasmid yield.

Minimizing the formation of acetate during shake flask experiments can improve the outcomes regarding pDNA production [29]. Therefore, choosing a carbon source that minimizes acetate formation, such as glycerol, can be a solution when testing strains with high glycolysis activity [29]. However, this choice is highly dependent on the host strain used. A recent study was performed in which different strains were tested in a glucose and glycerol medium. pDNA synthesis was inhibited by a high initial concentration of glucose (20 g/L) and high amounts of acetate accumulated when using strains with high glycolysis activity (e.g. DH5α, MG1655*ΔendAΔrecA* and corresponding *pykA* and/or *pykF* mutants) [29]. Therefore, a reduction of the initial amount of glucose (5+10 g/L) or the use of glycerol as an alternative carbon source, positively affected these strains [29]. On the other hand, the novel strain GALG20 (MG1655*ΔendAΔrecAΔpgi*), which uses the Pentose Phosphate Pathway as the main route for glucose metabolism, was not inhibited by 20 g/L of glucose during shake flask cultivations, resulting in the production of high amounts of pDNA and a minimal production of acetate [29]. These results highlight the fact that the choice of the culture medium is highly dependent on the strain.

1.5.2 Feeding strategy

Currently, the majority of the process development for pDNA production strains and vectors are conducted in shake flasks, where screening process is done, and then only the best selected strains are tested in bench-scale bioreactors. While this approach is popular, previous studies have already demonstrated that productivity data obtained from shake flask experiments often fail to predict the outcome of pDNA production in bench-scale bioreactors [69]. Also, O'Kennedy *et al*. [70] observed

that plasmid degradation occurred at the end of the fermentation in shake flasks. This could be associated with a lack of control of important parameters such as pH and dissolved oxygen, leading to high concentrations of accumulated acetate which can be toxic and inhibit biomass formation [69]. Obviously, oxygen transfer rate should be controlled during pDNA production in order to maximize biomass formation and volumetric productivity of pDNA [69]. In addition, plasmid supercoiling percentage is known to be affected by oxygen and temperature conditions. A single drop in dissolved oxygen concentration to 5% of air saturation has led to rapid loss in plasmid stability [67]. Furthermore, the formation of nicked plasmids and multimers can be affected by many parameters, including temperature, pH, dissolved oxygen, nutrient concentration, and growth rate [67].

In a recent study it was confirmed that MG1655*ΔendAΔrecA* and DH5α grown in shake flasks at 20 g/L of glucose produced small amounts of pDNA (~2 mg/g DCW) and high amounts of acetate (~1.3-3.0 g/L) [69]. However, when both strains were grown in a controlled bioreactor with 20 g/L of initial glucose, the specific pDNA yield of MG1655*ΔendAΔrecA* and DH5α increased 5 and 4 fold, respectively [69]. This demonstrates the importance of controlling the fermentation parameters in pDNA production.

In a batch fermentation strategy all nutrients are provided in the beginning of the experiment and no nutrients are added during cultivation. The application of batch processes in controlled fermenters has led to an increase of pDNA yield per volume [67]. However, this approach has fundamental disadvantages that result in limited pDNA yields. This is due to substrate inhibition and salt precipitation at high nutrient concentrations in the batch medium [67]. Furthermore, the growth rate in batch fermentations cannot be controlled directly; it is therefore unlimited, while steadily changing during fermentation, and ceases only when one or more nutrients are depleted or if metabolic byproducts (such as acetate) inhibit growth of the cells [67].

In a fed-batch strategy, nutrients are continuously or semi-continuously fed, while effluent is removed discontinuously. The fermentation begins with a batch mode containing a non-inhibitory concentration of substrate. The cells grow at μ_{max} until the substrate is exhausted, at which point the nutrient feeding begins [67]. More importantly, feeding mode is associated with growth rate (μ) control, while in batch mode the growth rate is not controlled (μ = μmax) [70]. High growth rates have been associated with acetate production, plasmid instability, and lower percentages of super-coiled plasmid. On the other hand, a reduced growth rate alleviates growth rate dependent plasmid instability by providing time for plasmid replication to synchronize with cell division [67]. Using a fed-batch strategy it is possible to control the growth rate in a specific value (μ < μmax), resulting in high pDNA yield and high quality of the final product [70]. Also, fed-batch fermentation results in higher biomass yields because substrate is supplied at a rate that allows its almost total consumption [67]. As a result, conversion of substrate to biomass is very efficient and residual substrate concentration is approximately zero, never reaching inhibitory concentrations. Also, metabolic overflow from excess substrate is reduced, avoiding excessive formation of inhibitory acetate [67].

In spite of the benefits of a fed-batch strategy it is known that the success of a fermentation strategy is clearly strain and/or plasmid dependent. In the previously reported study, using a fed-batch fermentation strategy MG1655*ΔendAΔrecA* and DH5α accumulated substantial amounts of acetate

(~18 and 20 g/L, respectively) and produced low amounts of pDNA (~50 and 40 mg/L, respectively), while GALG20 accumulated low amounts of acetate (~7 g/L) and produced a high amounts of pDNA (~160 mg/L) [71]. These results highlight the importance of strain genetic background in the choice of the fermentation strategy.

1.5.3 Temperature

Generally, lower growth rate favors reduced selection against plasmid-containing cells because plasmid presence reduces the maximum growth rate. While the optimal temperature for *E. coli* growth is 37°C, lower temperatures (such as 30°C) may be used in the batch phase of fed-batch fermentation. Lower temperatures cause reduced maximum specific growth rate and reduced plasmid copy number [72]. This combination reduces metabolic burden and plasmid loss during the batch phase by minimizing growth difference between plasmid-containing and plasmid-free cells.

Consistent with this, a recent report compared 35°C versus 37°C growth during the batch phase, with subsequent fed-batch production at 37°C under restricted growth rate [73]. Production yields with 35°C batch phase growth were dramatically improved (250 mg/L versus 50 mg/L) [73].

Moreover, Carnes *et al.* [74] demonstrated that higher yields are obtained with pUC origin plasmids grown at 37°C or at 30°C with a shift to 37 or 42°C at mid-exponential phase to induce plasmid-copy-number amplification. These plasmids are of high copy number and are temperature-sensitive, with copy numbers increasing with temperature. The initial temperature set point of 30°C aims to keep the plasmid copy number at a minimum, thus reducing the metabolic load while accumulating biomass [74]. Temperature shifts to 37°C or 42°C were performed at OD_{600}=60 to increase the plasmid copy number prior to harvest [74]. Surprisingly, volumetric plasmid yields were 670 mg/L when shifted to 37°C, and 1070 mg/L when shifted to 42°C. This strategy allowed the cultures to reach higher cell densities, ultimately exceeding OD_{600nm}=100 with no loss of cell viability [74].

1.6 Downstream processing

A key challenge in the development of cost-effective manufacturing processes is improving the yield of downstream purification steps. Large-scale purification of pDNA is difficult due to the complex, dynamic structure of pDNA, viscous process streams, and the presence of impurities (e.g., RNA, genomic DNA) with similar properties to the desired product [11].

The choice of the host strain to solve problems associated with the purification process is essential in order to obtain a final high yield and quality plasmid DNA process. Several researchers have taken a different approach and sought to improve downstream purification by engineering the *E. coli* host strain to reduce the amount of contaminating genomic DNA and RNA in the cell lysate [11]. The separation of RNA from DNA is particularly challenging because of the similar physicochemical properties of both nucleic acids. One strategy for RNA removal is digestion of cell lysates with RNase A, which allows the resulting small RNA fragments to be more easily separated from DNA. Cooke *et al.* [75] integrated the gene coding for RNase A into the chromosome of *E. coli* JM107 under the

control of an IPTG-inducible promoter. This strategy was particularly successful because the RNase A is robust enough to withstand the conditions of high pH encountered during alkaline lysis [75].

As mentioned previously, to address removal of host genomic DNA as well as RNA, Hodgson and Williams [32] developed *E. coli* hosts expressing periplasmic chimeric proteins that degrade both nucleic acids. These chimeras include the plasmid-safe phage T5 D15 exonuclease linked to RNase A or S [32]. The exonuclease specifically degrades linear and denatured DNA while not affecting the fidelity of supercoiled plasmid DNA [32]. This approach significantly reduces the viscosity of the lysate, easing the later stages of purification [32].

The supercoiled plasmid form is also another important parameter to measure in the quality and control of the final product. Molecular biology and genetic studies have been performed in order to improve supercoiled plasmid yields. The DNA gyrase and sigma factor σS (*rpoS*) genes were observed to play an important role in the regulation of plasmid topology and could be potential gene mutation targets to increase the fraction of supercoiled plasmids produced [25].

1.7 The impact of *zwf* overexpression on plasmid DNA biopharmaceutical production by *Escherichia coli*

The goal of this project is to study the impact of *zwf* overexpression on plasmid DNA biopharmaceutical production. As mentioned before, the Pentose Phosphate Pathway is a logical target for genetic engineering strategies to increase plasmid DNA yield because of the potential to increase carbon flux to nucleotide precursors (Figure 10).

The *zwf* gene encodes for the first enzyme of the oxidative Pentose Phosphate Pathway (glucose-6-phosphate dehydrogenase, G6PDH) which is responsible for the conversion of D-glucose 6-phosphate and $NADP^+$ into 6-phospho-D-glucono-1,5-lactone and NADPH. Therefore, G6PDH (encoded by *zwf*) and RpiA will increase metabolic flux of G6P towards Ru5P.

Overexpression of *zwf* has already been investigated as a strategy to increase carbon flux to the PP pathway. Williams *et al*. [54] verified that simultaneous overexpression of *zwf* and *rpiA* in DH5α appeared to increase plasmid amplification rate (mg pDNA/L/OD_{600}/hr) but not final specific yield (mg/L/OD_{600}), in fed-batch fermentations using complex medium with glycerol as the carbon source. However, overexpression of *zwf* alone did not show any effect on fermentation productivity (both volumetric and specific. It was also reported by Wang *et al.* [34] that PP pathway genes (*zwf* and *gnd*) were significantly downregulated in early stationary phase growing *E. coli* cells carrying a plasmid.

Given the previous results, in this work it is intended to study the effect of *zwf* and *rpiA* overexpression. In order to overexpress the *zwf* gene three promoter replacements were performed, with three different strengths. The promoters used are constitutive meaning that their activity is dependent on the availability of RNA polymerase holoenzyme, but is not affected by any transcription factors [76]. These are recognized by *E. coli* σ^{70} RNAP. σ^{70} is the major *E. coli* sigma factor so there should be RNAP present to transcribe these promoters under most growth conditions (although maximally during exponential growth) [76].

The promoter sequences will be obtained from the Anderson promoter collection [77], a well-characterized collection of *E. coli* σ^{70} constitutive promoters recovered from a library screen by Chris Anderson. The *zwf* and *rpiA* promoter replacements will be performed in MG1655*ΔendAΔpgi*, an intermediate strain of the process of GALG20 (MG1655*ΔendAΔrecAΔpgi*) construction. It is not possible to use the latter due to the absence of *recA* gene, reducing the recombination efficiency. After the promoter replacements the resulting strains will be assessed through baffled shake-flask cultivations using glucose as the main carbon source and the well-known pVAX1GFP plasmid. Finally, plasmid DNA production by the different strains will be quantified and a comparative study with GALG20 will be performed.

2 Materials and Methods

2.1 Strains

The host strain used as the basis of this work is MG1655*ΔendAΔpgi*. This strain was constructed by Geisa Gonçalves, former PhD student at IST. This strain is the result of an intermediate step in the GALG20 (MG1655*ΔendAΔrecAΔpgi*) construction, lacking the *recA* gene knockout. The strains used in this project are presented in Table 5. In spite of originated from previous studies, GALG20rpiA5, GALG20rpiA10 and GALG20 were constructed again in this project starting from MG1655*ΔendAΔpgi.*

Table 5 – Strains used in this project and the corresponding genotype.

Strains	Genotype	Reference
GALG20	MG1655*ΔendAΔrecAΔpgi*	[78]
MG1655*ΔendAΔpgi*	*F- λ –ilvG rfb-50 rph1ΔendAΔpgi*	[78]
GALG20rpiA5	MG1655*ΔendAΔrecAΔpgi rpiA*$^{+}$	[50]
GALG20rpiA10	MG1655*ΔendAΔrecAΔpgi rpiA*$^{+}$	[50]
GALG20zwf3	MG1655*ΔendAΔrecAΔpgi zwf*$^{+}$	This work
GALG20zwf5	MG1655*ΔendAΔrecAΔpgi zwf*$^{+}$	This work
GALG20zwf10	MG1655*ΔendAΔrecAΔpgi zwf*$^{+}$	This work

2.2 Plasmids

To construct the different strains three plasmids were used (Figure 11 and 12). The first one, pKD13 plasmid (Figure 11), contains a Kanamycin resistance gene [79]. The second one, pKD46 plasmid (Figure 12) [80, 81], contains the Red recombinase gene from phage λ and an Ampicillin resistance gene. The third one, pCP20 plasmid (Figure 12) [81, 82], was built by cloning of the FLP recombinase and carries Chloramphenicol and Ampicillin resistance genes. These constructions were performed at MIT and provided by Geisa Gonçalves, former PhD student at IST.

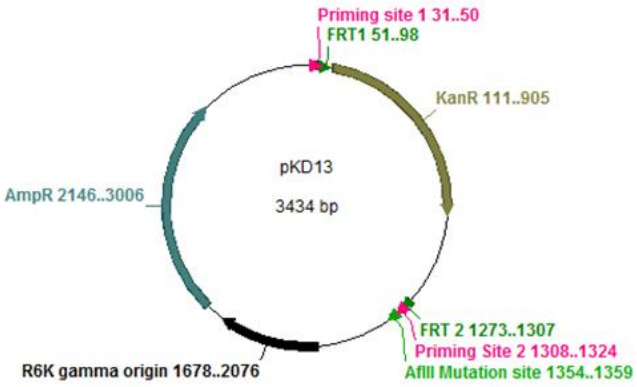


Figure 11 – Representation of pKD13 plasmid.

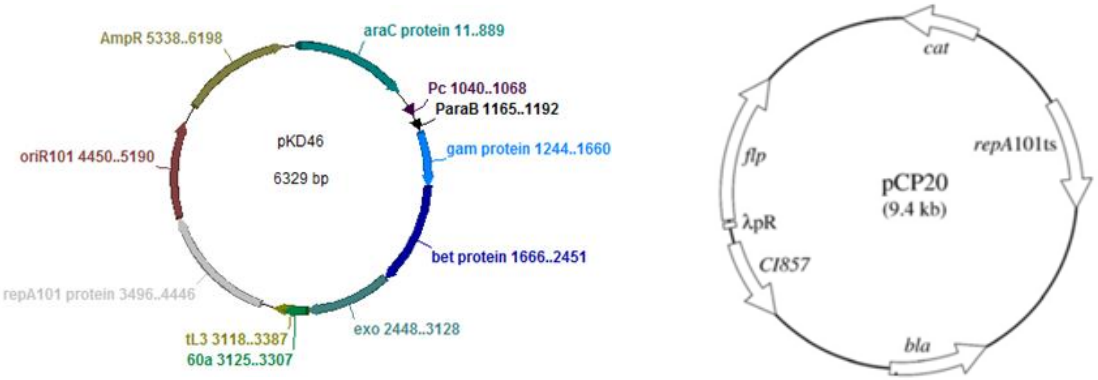


Figure 12 – Representation of pKD46 and pCP20 plasmids [81].

To quantify plasmid DNA production by the different strains, a fourth plasmid, pVAX1GFP (3697 bp), was used. This plasmid (Figure 13) was obtained by modification of the commercial plasmid pVAX1lacZ (6050 bp, Invitrogen), by replacement of the β-galactosidase reporter gene by the enhanced Green Fluorescent Protein (eGFP, referred to as GFP thereafter) gene [83]. This plasmid contains a Kanamycin resistance gene for bacterial selection and the human cytomegalovirus (CMV) immediate-early promoter, for expression in human cells. It also has a pUC origin of replication lacking the *rop* gene, which exerts control mechanisms to keep the copy number low. With such modification, pUC plasmids can reach copy numbers of more than 500 copies per cell [1].

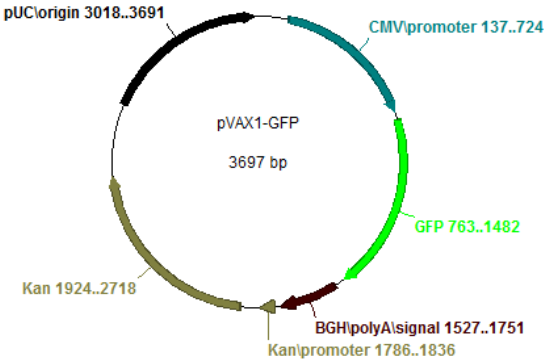


Figure 13 – Representation of pVAX1GFP plasmid.

2.3 Media, chemicals and other reagents

For cell growth, the medium used was LB (Luria-Bertani) from Sigma Aldrich (Molecular Biology tested). Some of the cell growths need the addition of antibiotics: Ampicillin (100 µg/mL), Kanamycin (30 µg/mL) and Chloramphenicol (50 µg/mL). These were obtained from Calbiotech.

The agar plates were made using LB-agar from NZYtech.

For PCR reactions, the Taq polymerase used was Nova Taq Hot Start Master Mix Kit (Novagen). The DNA ladder used was the 10 kb DNA Ladder III from NZYtech. Agarose used in electrophoresis gel was the SeaKem LE from Lonza.

For the electrocompetent cells protocol, L-arabinose 0.2% (Merck) was used.

In the P1 transduction protocol three solutions were used: 0.1M $CaCl_2$ (Merck ®), 10 mM $MgSO_4$(Merck ®), 0.1 M sodium citrate (Merck ®) pH 7.

To perform the shake-flask cultivations a complex medium was used. This medium contains a basal cultivation medium [Glucose (Panreac ®), 20g/L; Bacto peptone (BD ®), 10 g/L; yeast extract (BD ®), 10 g/L; $(NH_4)_2SO_4$ (Panreac ®), 3 g/L; K_2HPO_4 (Panreac ®), 3.5 g/L; KH_2PO_4 (Panreac ®), 3.5 g/L] with a final pH of 7.1. After autoclaving the basal cultivation medium the following components are added: Seed supplement solution, 8.3 mL/L (thiamine (Sigma Aldrich ®), 199 mg/L; $MgSO_4$ (Sigma Aldrich ®), 1.99 g/L); Trace elements solution, 1 mL/L ($FeCl_3$•6H2O, 27 g/L; $ZnCl_2$, 2 g/L; $COCl_2$•$6H_2O$, 2 g/L; Na_2MoO_4•$2H_2O$, 2 g/L; $CaCl_2$•$2H_2O$, 1 g/L; $CuCl_2$•$2H_2O$, 1.3 g/L; H_3Bo_3, 0.3 g/L; 1.2N HCl, 100 mL/L) [78]. This complex medium is then supplemented with 30 µg/mL of kanamycin.

2.4 PCR Primers

To perform this study several primers were used. Priming site sequences are represented by small case letters. Homology sequences are represented by capital letters and correspond to the 5' and 3' flanking regions of the target sequence. Promoter sequence is represented by underscored letters.

In order to evaluate if the previous mutational steps of MG1655 relative to *endA* and *pgi* genes were correct, four different primers were used.

Primer sequence for endA_ck_rev
GGTTCAGGATGATAAATGCG
Tm=49.7°C

Primer sequence for endA_ck_fwd
CGTCTATCGCTGTGTTCAC
Tm= 51.1°C

Primer sequence for pgi_ck_rev
TAGGCCTGATAAGACGCGAC
Tm= 64.1°C

Primer sequence for pgi_ck_fwd
TCTGTGACTGGCGCTACAAT
Tm=63.5°C

The primers related to *rpiA* promoter replacement were designed by Luís Carreira, former MSc student at IST and ordered from StabVida. These were designed with the goal of testing different promoter strengths, from a mild one (0.51) to a stronger one (1.00).

Primers sequences for rpiAFwd:
CGTTGTATGCTTATTTGCGGGGCGAAAGGGGATGCCTGCCATTGCGCGATATTTTATGGgtgtaggctggagctgcttc
Tm=94.5°C

Primer sequences for rpiA5Rev:
CGCCCATCCTACTGCTTTTTTCAATTCATCCTGCGTCATGATCGTTTCGCCTGTGGTATGAgctagcattatacctaggactgagctagctgtcagcccgtcgacctgcagtt
Tm=93.4°C, Promoter strength=0.51

Primer sequences for rpiA10Rev:
CGCCCATCCTACTGCTTTTTTCAATTCATCCTGCGTCATGATCGTTTCGCCTGTGGTATGAgctagcactgtacctaggactgagctagcgtcaacccgtcgacctgcagtt
Tm=94.8°C, Promoter strength=1

Primer sequences for rpiA_ck_Rev
GCGGTGGAACCTGTACCTACAC
Tm=66.5°C

Primer sequences for rpiA_ck_Fwd
GAACGGTGAACTGGTGCG
Tm=65.6°C

For the *zwf* promoter replacement, primers were designed using ApE software and obtained from STABvida.

Primer sequences for zwf_Fwd
TTACCCAATGGAAAACGATGATTTTTTTATCAGTTTTGCCGCACTTTGCGCGCTTTTCCC
gtgtaggctggagctgcttc
Tm=89°C

Primer sequence for zwf_Rev3
TCTCCTTAAGTTAACTAACCCGGTACTTAAGCCAGGGTATACTTGTAATTTTCgctagcattgtacctaggactgagctagccgtaaatccgtcgacctgcagtt
Tm=86°C, Promoter strength=0.33

Primer sequence for zwf_Rev5
TCTCCTTAAGTTAACTAACCCGGTACTTAAGCCAGGGTATACTTGTAATTTTCgctagcattatacctaggactgagctagctgtcagtccgtcgacctgcagtt
Tm=85°C, Promoter strength=0.51

Primer sequence for zwf_Rev10
TCTCCTTAAGTTAACTAACCCGGTACTTAAGCCAGGGTATACTTGTAATTTTCgctagcactgtacctaggactgagctagccgtcaatccgtcgacctgcagtt
Tm=86°C, Promoter strength=1

Primer sequence for zwf_ck_fwd
GCCTGTGTGCCGTGTTAATG
Tm=54°C

Primer sequence for zwf_ck_rev
GCAGCAATTTACGACGCGCA
Tm=59°C

To verify if the *recA* gene knockout was performed two primers were used. These were also designed by Luís Carreira.

Primer sequence for recA_ck_rev
GAAAAGACAGTTGGCAAAGGTTC
Tm=64.8°C

Primer sequence for recA_ck_fwd
CGAGTCTTGTACCGGTGGTT
Tm=63.9°C

2.5 Generation of Kanamycin cassette

For construction of the *rpiA* kan-cassette a PCR was performed using the different *rpiA* primers containing the different promoters (section 2.4). The PCR reactions were performed in 50 µl reactions containing Platinum PCR Supermix High Fidelity (45µL), pKD13 plasmid (2 µl), primer rpiAFwd (1.5 µl, 10µM stock), primer rpiA5Rev and rpiA10Rev separately (1.5 µl, 10µM stock). After the PCR reactions, the resulting products were analyzed by 1% agarose gel electrophoresis. The extraction of the Kan-cassette from the gel was performed using QIAquick Gel Extraction Kit from Qiagen®. After the gel extraction, the concentrations of the product were measured using NanoVue Plus Spectrophotometer (GE Healthcare®).

In order to obtain the Kanamycin cassette containing the different *zwf* promoters the procedure was the same.

2.6 Gel electrophoresis

Gel electrophoresis analyses were performed using horizontal electrophoresis tanks from VWR® and Electrophoresis Power supply – EPS 301 from Amersham Pharmacia Biotech using a voltage of 100-120V and duration of 1h-1h30, depending of the size of the gel. Gels are stained for 30 minutes with ethidium bromide (EtBr) and then observed in an Eagle Eye II ® Stratagene trans-illuminator, under UV light.

2.7 Electrocompetent cells

The purpose of this protocol is to make competent cells to introduce foreign DNA by electroporation. Electroporation uses short high-voltage pulses to overcome the barrier of the cell membrane. By applying an external electric field, which just surpasses the capacitance of the cell membrane, transient and reversible breakdown of the membrane can be induced. This transient, permeabilized state can be used to load cells with a variety of different molecules, either through simple diffusion in the case of small molecules, or through electrophoretically driven processes allowing passage through the destabilized membrane - as is the case for DNA transfer [84].

The first step of this protocol was to prepare a seed culture (5 mL) and incubate it overnight at appropriate conditions with LB medium and antibiotics/inducers if necessary (according to the step of the promoter replacement protocol). The next day, an appropriate volume of cells was transferred to a sterile 50 mL shake flask with LB medium, in order to start with an OD_{600nm}=0.1. This cell culture was supplemented with antibiotics/inducers if necessary, and incubated at appropriate conditions. When the culture reached an OD_{600nm}=0.5, 40 mL were transferred to a sterile Falcon tube (50mL) and centrifuged at 11,000 rpm for 15 min, at 4°C. After that, the supernatant was discarded and the pellet was resuspended in 40 mL of chilled ddH_2O and centrifuged again at the same conditions. This washing step was performed three times. After the last centrifugation, the supernatant was discarded and the pellet resuspended in an appropriate volume of chilled ddH_2O (according to the number of aliquots needed) and divided into several 50 µL aliquots. Finally the cells are ready to be electroporated. The electroporation cuvettes must be maintained for 1 hour in ethanol and then dried (if they are re-used). After that they must be put on ice, together with the ddH_2O.

2.8 Promoter replacements

The promoter replacement protocol was adapted from the method described by Datsenko and Wanner [85]. The method was the same for *rpiA* and *zwf* promoter replacements, differing only in the kanamycin cassette used. The sequences of the promoters used are represented in Table 6. Reported

activities of the promoters are given as the relative fluorescence of these plasmids in strain TG1 grown in LB media to saturation [77].

Table 6 – Sequences of the promoters used in the promoter replacement protocol [77].

Promoter sequence	Measured Strength
tttacggctagctcagtcctaggtacaatgctagc	0.33
ctgacagctagctcagtcctaggtataatgctagc	0.51
ttgacggctagctcagtcctaggtacagtgctagc	1

This procedure is based on the Red system which has showed to be very efficient. The basic strategy is to replace a chromosomal sequence with a selectable antibiotic resistance gene that is generated by PCR by using primers with 36-nt homology extensions (H1 and H2, Figure 14). This is accomplished by Red-mediated recombination in these flanking homologies achieved by the transformation of cells with a plasmid containing the λ-RED recombinase (pKD46 plasmid, Amp^R). After selection of the transformed colonies, the resistance gene can be eliminated by using a helper plasmid expressing the FLP recombinase (pCP20 plasmid, Amp^R and Cm^R), which acts on the directly repeated FRT (FLP recognition target) sites flanking the resistance gene. The Red and FLP helper plasmids can be simply cured by growth at 37°C because they are temperature-sensitive replicons [85].

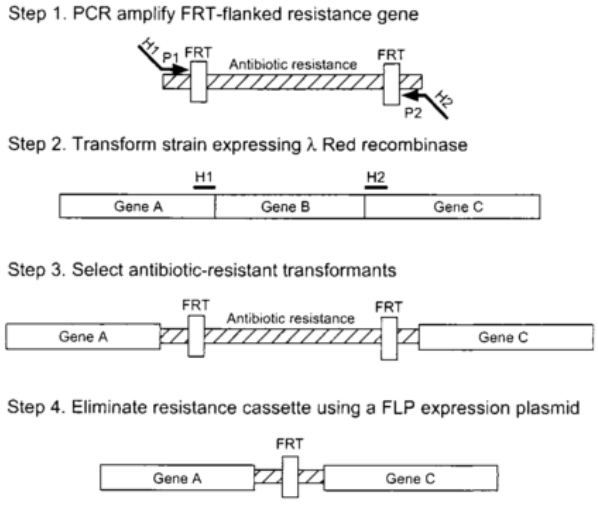


Figure 14 - Simple gene disruption strategy. H1 and H2 refer to the homology extensions or regions. P1 and P2 refer to priming sites [85].

The first step of this procedure was the generation of the Kanamycin cassette containing the desired sequence (*rpiA* or *zwf* new promoter), which protocol is already described in section 2.5. The resulting kanamycin cassette is shown in Figure 15.

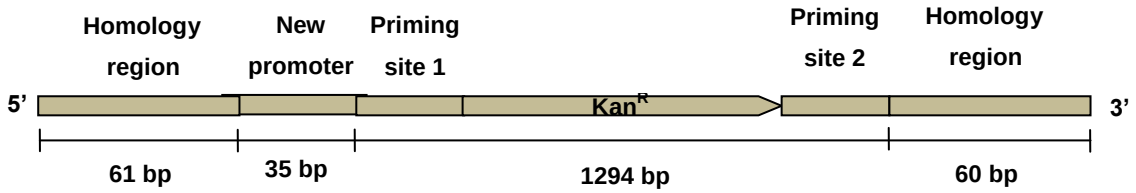


Figure 15 - Kanamycin cassette used in the promoter replacements (1450 bp).

This Kan-cassette contains a kanamycin resistance gene surrounded by FRT (FLP recognition target) sites to facilitate later excision of the Kanamycin resistance gene. The reverse primer used in the PCR reaction contains one homology sequence, the new promoter and a priming site coincident with the start of the Kan^R gene. The forward primer contains a priming site and a homology sequence coincident with the terminus of the Kan^R gene. Therefore, by performing the PCR reaction the product obtained consists on a kan-cassette flanked by large stretches of homology to the area in the genome that is targeted, as well as the promoter sequence on the upstream side (Figure 15).

After the generation of the kan-cassette, the first step of the protocol is to electroporate MG1655*ΔendAΔpgi* cells with pKD46 plasmid, which contains the λ-RED recombinase. In order to do that, after the electrocompetent cells protocol (section 2.7) the cells were electroporated with 1 µL of the pKD46 plasmid, using a 2,500 V pulse for 5 seconds. After the pulse, 350 µL of LB medium were immediately added and the cells were incubated at 30°C for 1 hour, in order to recover from the shock. After the recovery step, the cells were centrifuged at 12,000 g for 4 minutes and the majority of the supernatant was discarded. The remaining (*ca*. 100 µL) supernatant was used to resuspend the cell pellet which was plated onto LB/Amp100 plates and incubated overnight at 30°C.

After electroporation with pKD46 plasmid, the cells are capable of producing the λ-RED recombinase, which will allow kan-cassette insertion in the target genome by homologous recombination. Therefore, the following step is electroporation with the appropriate kan-cassette (*rpiA* or *zwf*). Therefore, a colony containing the pKD46 plasmid (Amp^R) was picked and grown at 30°C and 250 rpm overnight, in 5mL LB/Amp100 medium. The next day, an appropriate volume of cells was transferred to a sterile 50 mL shake flask with LB/Amp100 medium, in order to start with an OD_{600nm}=0.1. Also, 500 µL of L-arabinose (0.2%) were added to the culture in order to induce recombinase production. After the electrocompetent cells protocol (section 2.7) the cells were electroporated with 1 µL of the kan-cassette. After the pulse, 350 µL of LB medium were immediately added as well as 0.7 µL of ampicillin (200mg/mL) and 8.8 µL of L-arabinose (0.2 %). Subsequently, the cells were incubated at 30°C for 2 hours, in order to recover from the shock. After the recovery step, the cells were centrifuged at 12,000 g for 4 minutes and the majority of the supernatant was discarded. The remaining supernatant was used to resuspend the cell pellet and cells were plated onto LB/Kan30 plates and incubated at 37°C overnight.

The next day a colony PCR was performed in order to verify if the kan-cassette was inserted. The PCR conditions are described in section 2.10.

The final step of this protocol is the removal of the kanamycin resistance gene using pCP20 plasmid (Amp^R and Cm^R). For that, the resulting cells from the previous step were grown in 5ml LB/Kan30 at 37°C and 250 rpm overnight. The next day, the eletrocompetent cells protocol was the same as described previously (section 2.7) and then the cells were electroporated with 1 µL of the pCP20 plasmid. After the pulse, 350 µL of LB medium were immediately added and the cells were incubated at 30°C for 1 hour, in order to recover from the shock. After the recovery step, the cells were centrifuged at 12,000 g for 4 minutes and the majority of the supernatant was discarded. The remaining supernatant was used to resuspend the cell pellet and cells were plated onto LB/Cm50 plates and incubated at 30°C overnight.

The next day, a clone was isolated and plated onto a LB plate and incubated overnight at 43°C. The high temperature will induce FLP recombinase which will promote homologous recombination with the FRT sites in the target genome and subsequently remove the Kan-cassette. The high temperature will also cure all helper plasmids. The next day, several colonies were picked and plated onto LB/Amp100, LB/Kan30 and LB/Cm50 plates and incubated overnight at 30°C. This step is important in order to verify if the plasmids are fully cured. Finally, if no colonies the next day, the cells have lost all the plasmids and are ready to be used in further experiments. Therefore, the correspondent colonies were picked from the previous LB plate and a colony PCR was performed in order to verify if the mutational steps were successful. A sequencing step is also needed to ensure no unintended mutations.

2.9 P1 Transduction

P1 transduction (Figure 16) is used to transfer genetic material such as genes or mutations from one strain of bacteria (the donor) to another (the recipient). Fragments of ~100 kb can be transferred by the P1 bacteriophage [86]. The phage is first grown with a strain containing the elements to be moved, and the resulting phage lysate is used to infect a recipient strain. During growth and encapsidation of its DNA into the phage head, the phage occasionally packages the DNA of its bacterial host rather than its own phage chromosome into the protein capsid. Therefore, the lysate will contain bacterial DNA as well as phage DNA, and genetic recombination, catalyzed by enzymes of the recipient strain, will incorporate the bacterial fragments into the recipient chromosome [86]. The infected recipient bacteria are plated on a medium that selects for the genome segment of the donor bacteria. All of this would not work if the infectivity of the phage could not be controlled, as phage released from neighbor cells would infect and lyse the bacteria that had been infected with transducing particles [86]. Therefore, control of P1 infectivity is very important and is achieved by growing in the presence and absence of Calcium [86]. Sodium Citrate works as a chelator and it lowers the concentration of free calcium, essential for phage adsorption to the bacteria, thus minimizing secondary infection by residual phage [86].

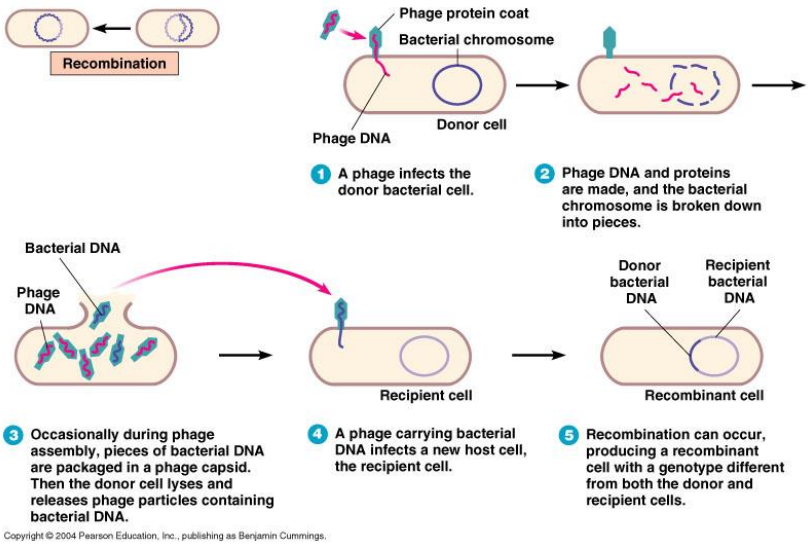


Figure 16 - Illustration of the P1 transduction mechanism [87].

This protocol was used to perform the *endA* and *recA* knockouts in MG1655. The strains used as donor cells are from the Keio collection [88]: JW2912-1 [Δ*endA*720::*kan*] and JW2669-1 [Δ*recA*774::*kan*]. The first part of the protocol is the preparation of the donor phage stock. A 5 mL LB culture tube of the donor cells with antibiotic was prepared, as well as a 5 mL LB culture tube with recipient cells without antibiotic, and incubated overnight at 37°C and 250 rpm. The next day, a 5 mL tube was prepared containing LB medium with 0.1% glucose (25 µL of 20% glucose), 5 mM $CaCl_2$ (250 µL of 0.1M $CaCl_2$), and the antibiotic. This solution was then inoculated with 30 µL of the donor cells and incubated at 37°C until the OD_{600nm} was between 0.15-0.25. When the cells reached that OD_{600nm} value, P1 phage (ATCC #25404-B1) was added to the culture (5-10 µL) and incubated for 5-6 hours in order to allow the phage to infect and lyse the donor cells. After that, if cell debris were visible (if not, it must be incubated overnight proceeding with the protocol in the following day), 100 µL of chloroform were added to the culture and mixed by vortex for 1 minute. The chloroform kills the remaining live donor cells as well as helps in the precipitation of cell debris. Since chloroform is denser than water, two separate phases are formed. Then, 1 mL of the aqueous phase was transferred to an eppendorf and centrifuged at maximum speed for 2 minutes. The supernatant was then transferred to a new eppendorf. Next, 50 µL of chloroform were added and the culture was mixed vigorously with vortex. The chloroform may or may not completely dissolve into the aqueous phase. Then, the culture was centrifuged again for 2 minutes at maximum speed. Finally, the supernatant was stored as donor phage stock at 4°C, being stable for at least 3 weeks.

The next step of the protocol is the preparation of the recipient cells. First, a 50 mL LB shake-flask was inoculated with 5 mL of stationary-phase recipient cell culture from the previous step and incubated at 37°C until the OD_{600nm} was approximately 0.8-1.2 (mid-exponential phase cells). After that, the cells were harvested by centrifugation at 4,000 g and 4°C for 15 minutes. The supernatant was discarded and the pellet was resuspended in 500 µL of 10 mM $MgSO_4$ + 10 mM $CaCl_2$. This solution was then divided into five 100 µL aliquots and 0, 30, 50, 70 and 120 µL of the donor phage

stock were added to each aliquot. The 0 µL sample is a control to make sure that the recipient cells will not grow on the antibiotic plate and give any false positives. The aliquots were incubated at 30°C for 30 minutes, to allow the donor phage to infect the recipient cells. Next, 100 µL of 0.1 M sodium citrate (pH 7.0) were added to each aliquot and mixed. This stops the infection process. Subsequently, the cells were centrifuged at 7000 rpm for 1 minute. The supernatant was discarded and the pellet was resuspended in 100 µL 0.1 M sodium citrate (pH 7.0). Then, the cells were centrifuged at 7,000 rpm for 1 minute, the supernatant was discarded and one more washing step was performed. After the final centrifugation, the cells were resuspended in 100 µL of 0.1 M sodium citrate (pH 7.0) and plated onto four LB plates with antibiotic (kanamycin, in this case). Also, 100 µL of the donor phage stock were plated onto a LB plate with antibiotic, as a negative control for the possibility that donor cells survived the chloroform killing during the preparation of the donor phage stock. Finally, all the plates were incubated at 37°C overnight.

The next day, the colonies obtained in the previous step were re-streaked onto fresh LB plates with antibiotic since the original plates are covered with bacteriophage. The plates are incubated at 37°C overnight, and the resulting colonies are ready to be used for further experiments. A colony PCR was performed in order to verify if the knockout was successful.

2.10 PCR conditions

Colony PCR reactions to verify the insertions and knockouts were performed in 25 µL reactions containing Nova Taq Hot Start Master Mix Kit (12.5 µl), primer forward (1 µl, 10µM stock), primer reverse (1 µl, 10µM stock) and ddH_2O (10.5 µl). Each colony was picked with a sterile tip and resuspended in 10.5 µl of ddH_2O, followed by the addition of the other components. In this case, the primers used were “check” according to the gene being tested (section 2.4). The PCR program was adapted to the specific primers used, varying only the annealing temperature: a 95°C incubation for 15 minutes; denature at 94°C for 30 seconds, annealing at proper annealing temperature for 30 seconds and extend at 72°C for 1 minute, being this cycle repeated 30 times; and a final extension at 72°C for 10 minutes.

The annealing temperatures used for the different primers were: pgi (64°C), endA (57°C), rpiA_ck (61°C), rpiA (66°C), zwf (61°C), recA (60°C).

The equipment used for PCR reactions was TGradient from Biometra ®.

2.11 Restriction analyses

The restriction analyses were performed using 1000 ng of DNA. A double digestion with EcoRI and NcoI was performed as well as two single digestions with PstI and EcoRI (Promega ®). Restriction reactions were done with 0.5 µL of enzyme (10u/µL), 1 µg of plasmid DNA and 5 µL of Buffer H within 25 µL reaction volume. The buffer was selected based on the maximum activity registered by each pair of enzymes. The mixture was then incubated for 2 hours at 37°C.

2.12 Cell Banks

The desired cells were grown overnight in a 5 mL LB tube (with antibiotic, if necessary). The next day 500 µL of the culture were transferred to a shake flask with 50 mL LB (with antibiotic, if necessary). When the cells reached an OD_{600nm}=1, 200 µL were transferred to a cryotube and 200 µL of Glycerol (30%) were added. The cells were then stored at -80°C.

2.13 Transformation of cells with pVAX1GFP plasmid

After the construction of the different strains these were electroporated with the pVAX1GFP plasmid in order to further evaluate plasmid DNA production. After the electrocompetent cells protocol (section 2.7) the cells were electroporated with 1 µL of the plasmid (2,500 V pulse) and incubated at 37°C for 1 hour in order to recover from the shock. After that, the cells were plated onto LB plates supplemented with kanamycin and incubated overnight at 37°C.

2.14 Shake-flask cultivations

To perform shake-flask cultivations a complex medium was used (section 2.3). The first step was the preparation of an overnight culture. The selected colonies were picked and incubated overnight in a LB tube with 5 mL of basal cultivation medium, 41.5 µL of seed supplement solution, 5 µL of trace elements solution (1 mL/L), 500 µL of glucose (20g/L) and 5 µL of kanamycin (30 µg/mL). The next day, the OD_{600nm} was measured and an appropriate volume of cells (in order to start cultivation with an OD_{600nm} = 0.1) was transferred into a shake-flask containing 45 mL of basal cultivation medium, 415 µL of seed supplement solution, 50 µL of trace elements solution (1 mL/L), 5 mL of Glucose (20 g/L) and 50 µL of kanamycin (30 µg/mL). The cells were incubated at 37°C and 250 rpm. This culture was grown for 24 hours and the OD_{600nm} was measured every hour during the first 12 hours and the final sample taken at 24 hours, in order to construct a growth curve. Sampling for DNA quantification was performed after 12 hours of growth. This experiment was performed four times.

2.15 Plasmid DNA quantification

Plasmid DNA quantification was performed using High Pure Plasmid Isolation Kit from (Roche®) and NanoVue Plus Spectrophotometer (GE Healthcare ®) equipment. During the cell growth, 1 mL of the culture was obtained after 12 hours of growth and used to purify PVAX1GFP plasmid. The sample was then centrifuged at 12000 rpm for 10 minutes and the supernatant discarded. The pellet was then stored at -20°C. To perform the isolation of the plasmid with Roche kit, an appropriate volume of water was added to the cell pellet in order to start with an OD_{600nm}=4 with

1mL of sample. After the isolation protocol, the plasmid concentration was quantified using NanoVue Plus Spectrophotometer (GE Healthcare ®).

3 Results and discussion

3.1 Kan-cassette amplification

The first step for promoter replacement was the amplification of *rpiA* and *zwf* kan-cassettes. This was performed as previously described (section 2.5) using pKD13 plasmid. This plasmid contains the kanamycin resistance gene as well as the FRT regions necessary to do the promoter replacements. Therefore, the product of this PCR will be the kan-cassette containing the specific promoter of each gene (*rpiA* or *zwf*) as illustrated in section 2.8 (Figure 15). After the PCR reaction, the resulting product was analyzed in an agarose gel electrophoresis, being the result presented in Figure 17.

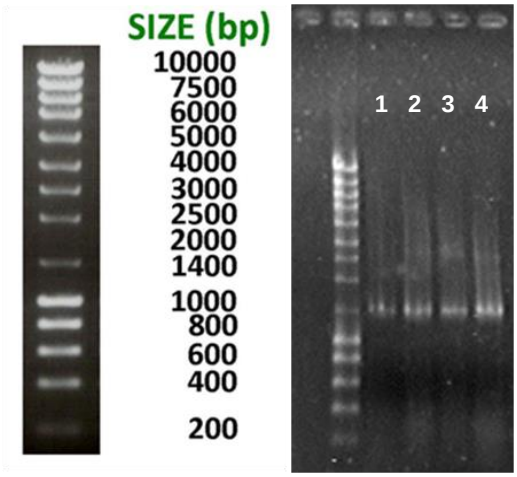


Figure 17 - Agarose gel analysis of PCR result to amplify *rpiA* kan-cassette. The different lanes represent 2 different pKD13 stocks and two different *rpiA* kan-cassette (rpiA5 (lane 1 and 2) and rpiA10 (lane 3 and 4)).

As seen in Figure 17, the kan-cassette sizes are correct (1450 bp) and so the PCR product was extracted using QIAquick Gel Extraction Kit from Qiagen®. After the gel extraction, the concentrations of each kan-cassette (strengths: 0.51 and 1) were measured using NanoVue Plus Spectrophotometer (GE Healthcare®).

To obtain the kan-cassette containing the different *zwf* gene promoters (strengths: 0.33, 0.51 and 1) the procedure was the same. After the PCR reaction, the resulting product was analyzed in an agarose gel electrophoresis, being the result presented in Figure 18.

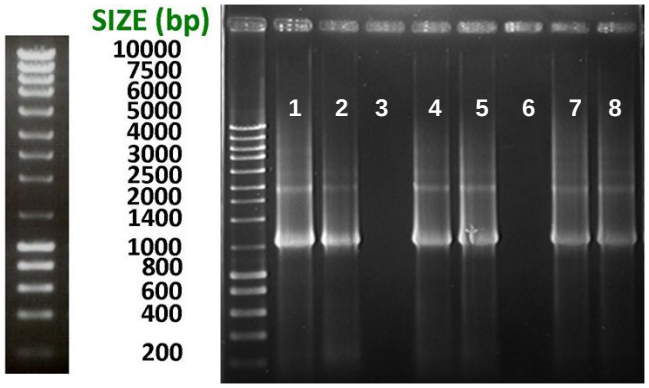


Figure 18 - Agarose gel analysis of PCR result to amplify *zwf* kan-cassette. The different lanes represent 3 different *zwf* kan-cassettes in duplicates (zwf3 (lane 1 and 2), zwf5 (lanes 4 and 5) and zwf10 (lane 7 and 8)).

Again, as seen in Figure 15, the kan-cassette size is correct (1450 bp) and so the PCR product was extracted from the gel. After that, the concentrations of each kan-cassette (zwf3, zwf5 and zwf10) were measured using NanoVue Plus Spectrophotometer (GE Healthcare®).

3.2 Assessment of MG1655*ΔendAΔpgi* strain's genotype and pKD46 plasmid's sequence

The first goal of this project was to perform the *rpiA* promoter replacements on MG1655*ΔendAΔpgi* strain. However, the first attempts were all unsuccessful. Different approaches were tried by changing some parts of the protocol but the insertion of the kan-cassette was not achieved. After several attempts in changing the experimental conditions, it was decided to test the materials used such as the starting strain's genotype and the pKD46 plasmid. This plasmid is essential for the insertion of the kan-cassette since it contains the gene that codes for the λ-RED recombinase, being essential for the homologous recombination between the kan-cassette and the genome to occur.

The genotype of the starting strain was confirmed by verifying the *endA* and *pgi* genes knockouts. Therefore, colony PCR reactions were performed using *endA* and *pgi* primers (section 2.4) under the conditions presented in section 2.10.

Concerning the *pgi* gene, it was confirmed that the knockout was previously done successfully as the gene is absent. However, in the case of the *endA* gene assessment it was verified that all of the colonies tested contained this gene and so the *endA* knockout was not successfully achieved. Therefore, the next step was the *endA* gene knockout using P1 transduction (section 2.9). This procedure was successful and the final colonies were tested by colony PCR and agarose gel analysis in order to verify if the two genes were absent. The result is represented in Figure 19.

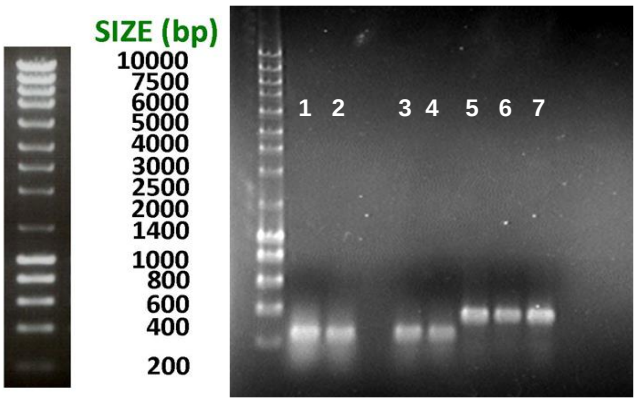


Figure 19 - Agarose gel analysis of Colony PCR result to assess *endA* and *pgi* genes knockout. The first 4 lanes represent 4 different colonies tested for *endA* gene knockout and the last 3 lanes represent the same colonies tested but for *pgi* gene knockout.

As seen in Figure 19 the two knockouts were successful since the different lanes represented contain bands correspondent to fragment sizes of approximately 300 bp. This indicates that both

genes are absent since it is known that the original *pgi* gene length is 1650 bp [57] and *endA* gene length is 708 bp [42]. Therefore, the strain's genotype is the expected and the cells were ready to be used in the next procedures.

After several failed experiments it was hypothesized that pKD46 plasmid's sequence might not be correct. Therefore, a restriction analyses was performed in order to confirm the sequence of the plasmid, by comparison with another pKD46 stock (donated by a lab colleague).

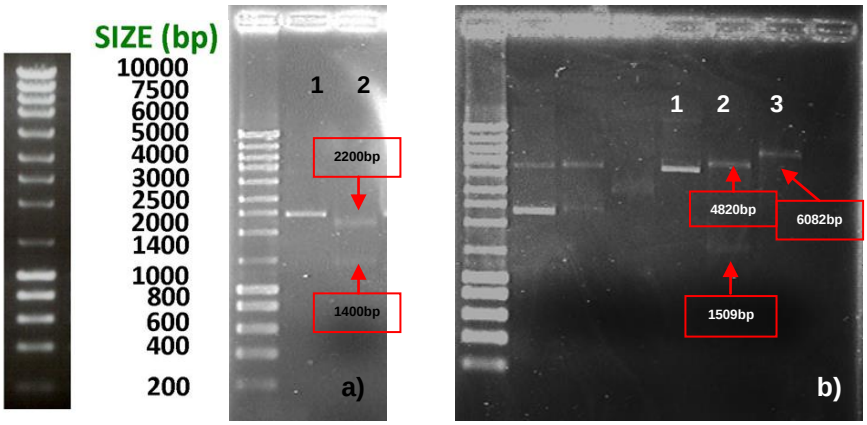


Figure 20 - Agarose gel analysis of a restriction analysis of pKD46 plasmids. a) Restriction analysis of pKD46 plasmid, old stock (lane 1, non-digested; lane 2, double digestion with EcoRI/NcoI); b) Restriction analysis of pKD46 plasmid, new stock (lane 1, non-digested; lane 2, digested with EcoRI; lane 3, digested with PstI).

In Figure 20a) the restriction analysis of pKD46 plasmid's initial stock is represented. The result of double digestion with EcoRI/NcoI enzymes should be the presence of three fragments with 4155 bp, 1509 bp and 665 bp. However, only two fragments are present, one with approximately 1400 bp and another with approximately 2200 bp. This demonstrates that the plasmid sequence is not correct. Therefore, another restriction analysis was performed with a new stock of pKD46 plasmid and the result is represented in Figure 20b). Digestion with EcoRI should originate two fragments of 4820 bp and 1509 bp, while digestion with PstI should result in two fragments of 6082 bp and 247 bp. In Figure 20b) lanes 2 and 3, is possible to verify the presence of the expected bands, although very faded, except the fragment of 247 bp resultant from the PstI digestion, which is not visible. This can be the result of low amounts of DNA which can give rise to very faded bands.

This experiment demonstrated that the pKD46 plasmid used in the first promoter replacements attempts was not correct, explaining the failures. Therefore, this stock was discarded and the new one was used in the further experiments.

3.3 *rpiA* promoter replacements

The *rpiA* promoter replacements were performed on the MG1655*ΔendAΔpgi* strain. The protocol (section 2.8) was executed and after the transformation step with the respective kan-cassette (rpiA5 and rpiA10) a colony PCR was performed in order to verify if it was inserted in the genome. For

that, rpiA_ck primers were used (section 2.4), which amplifies the region of interest, approximately 1500bp, represented in Figure 21 (40 bp upstream and downstream of the target region). After the PCR reaction, the resulting product was analyzed in agarose gel electrophoresis, yielding the gel in Figure 22.

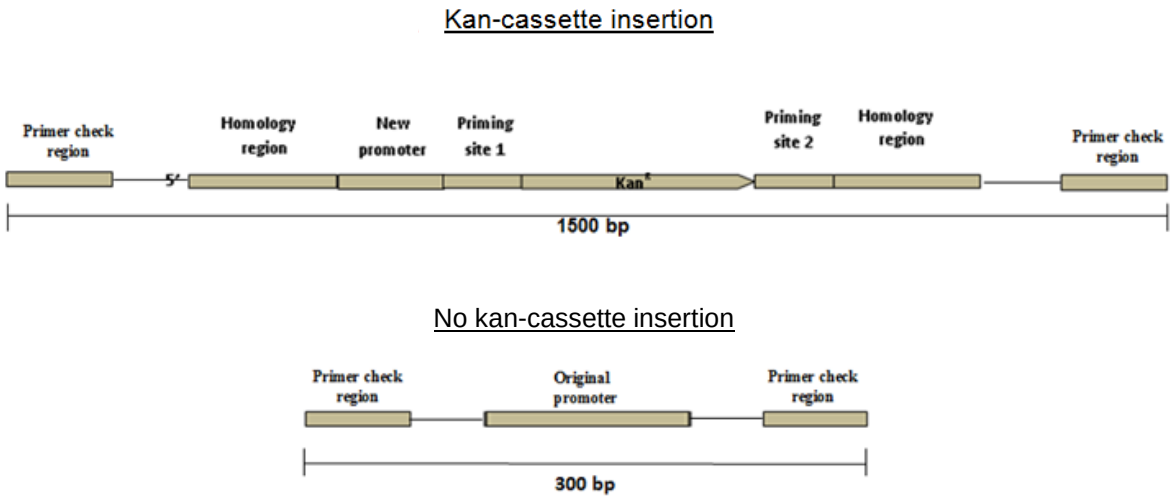


Figure 21 – Representation of the expected results for kan-cassette insertion and no insertion.

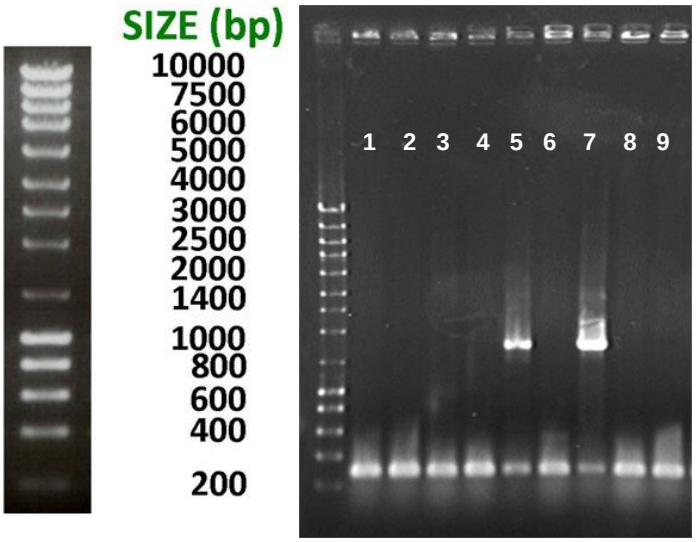


Figure 22 - Agarose gel analysis of a Colony PCR result to assess *rpiA* promoter replacement. The different lanes represent 9 different colonies tested (lanes 1 to 5, rpiA5; lanes 6 to 9, rpiA10).

As seen in Figure 22 only two colonies show a positive result (lane 5 and 7, rpiA5 and rpiA10 respectively) being represented a fragment of approximately 1500 bp correspondent to the insertion of the kan-cassette (Figure 21). However, a negative result is also present (a band correspondent to a fragment of approximately 300 bp) which can be due to the presence of cells with the insertion of the

kan-cassette and cells without it (mixed colonies). Therefore, these two colonies were re-streaked onto new LB/Kan30 in order to try to obtain isolated colonies. The next day, a PCR analysis was performed with a few of the new colonies and the resulting product was analyzed in agarose gel electrophoresis, yielding the gel in Figure 23.

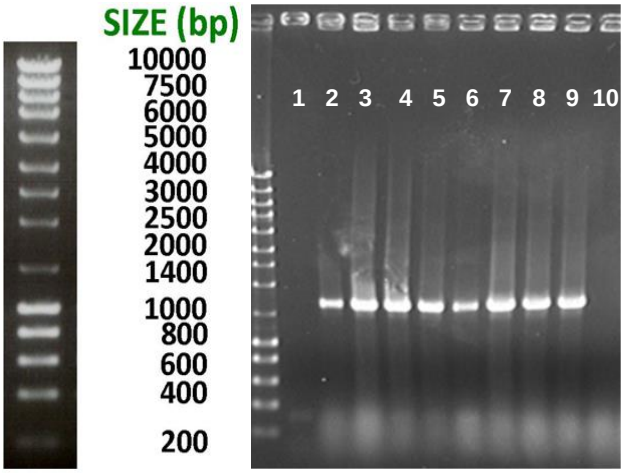


Figure 23 - Agarose gel analysis of a Colony PCR result to assess *rpiA* promoter replacement. The different lanes represent 10 different colonies tested (lanes 1 to 5, rpiA5; lanes 6 to 10, rpiA10)

As observed in Figure 23 most of the new isolated colonies show a positive result, since a band correspondent to a 1500 bp fragment is present. The lower band correspondent to a negative result is also absent. Therefore, the next step was the removal of the kanamycin resistance gene of two of the positive colonies (rpiA5 and rpiA10) using pCP20 plasmid (section 2.8). After the procedure the resultant colonies (chloramphenicol resistant) were plated onto a LB plate and incubated overnight at 43°C in order to cure all the plasmids and activate the recombinase that will perform the excision of the kan-cassette. The next day a colony PCR was performed to verify if the experiment was successful and the resulting product was analyzed in agarose gel electrophoresis. The expected result is represented in Figure 24 and the obtained result in Figure 25.

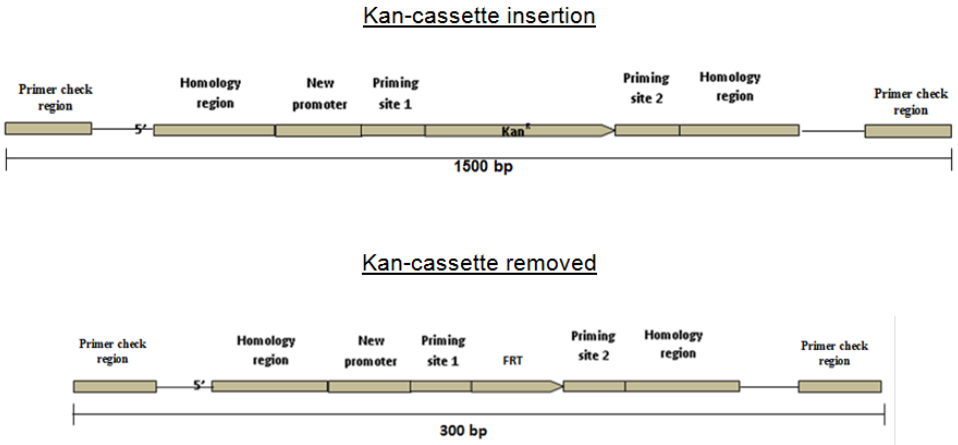


Figure 24 – Representation of the expected result of kan-cassette insertion and removal.

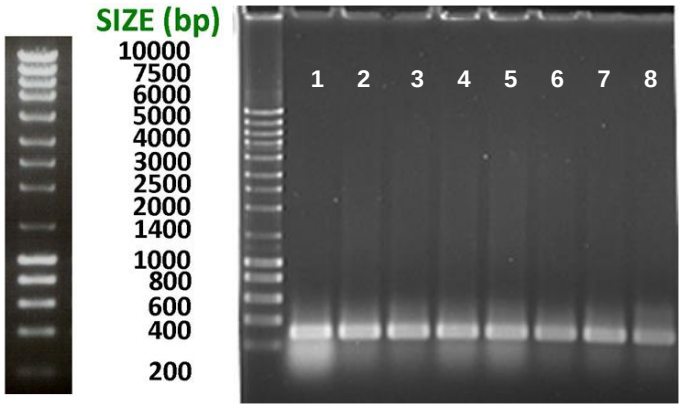


Figure 25 - Agarose gel analysis of a Colony PCR result to assess kanamycin gene removal for *rpiA* promoter replacement. The different lanes represent 8 different colonies tested (lanes 1 to 4, rpiA5; lanes 5 to 8, rpiA10).

As observed in Figure 25 the kanamycin resistance gene was successfully removed from all the colonies tested since the bands represent a fragment with approximately 300 bp which is correspondent to the size of the scar left by the removal of the kan-cassette. Some of these colonies were then plated onto LB/Kan30, LB/Amp100 and LB/Cm50 plates in order to verify if the cells do not have any antibiotic resistance. Finally, the colonies that did not grow in any of the antibiotics were picked again from the previous LB plate and submitted for sequencing and seed stocks were made and stored at -80°C. The results of the sequencing are presented in Figure 26 and Figure 27.

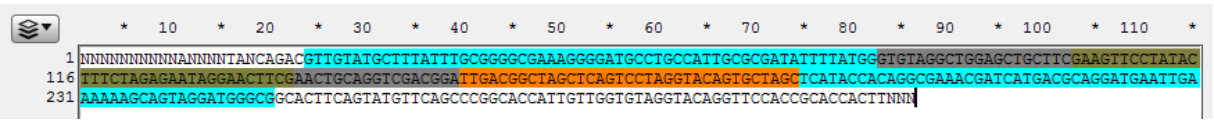


Figure 26 - Sequencing result for a rpiA5 colony. The orange color highlights the presence of the promoter with strength of 0.51. The blue, grey and green colors highlight the presence of homology, priming site and FRT regions, respectively.

Figure 27 - Sequencing result for a rpiA10 colony. The orange color highlights the presence of the promoter with strength of 1. The blue, grey and green colors highlight the presence of homology, priming site and FRT regions, respectively.

The sequencing results were positive and it is possible to verify that the new promoter was successfully inserted in both cases (rpiA5 and rpiA10) and there are no unintended mutations. Therefore, these cells are ready to be used in the next procedures. The following step was to perform

the *recA* gene knockout in order to obtain GALG20rpiA5 and GALG20rpiA10 strains (MG1655*ΔendAΔrecAΔpgi rpiA*$^{+}$).

3.4 *zwf* promoter replacements

The *zwf* promoter replacements were performed with the same procedure as for *rpiA* promoter replacement, on the MG1655*ΔendAΔpgi* strain. The protocol (section 2.8) was executed and after the transformation step with the respective kan-cassette (zwf3, zwf5 and zwf10) a colony PCR was performed in order to verify if it was inserted in the genome. For that, zwf_ck primers were used (section 2.4), which will amplify the region of interest with approximately 1500bp (40 bp upstream and downstream of the target region). The expected result is represented in Figure 28 and the obtained result in Figure 29.

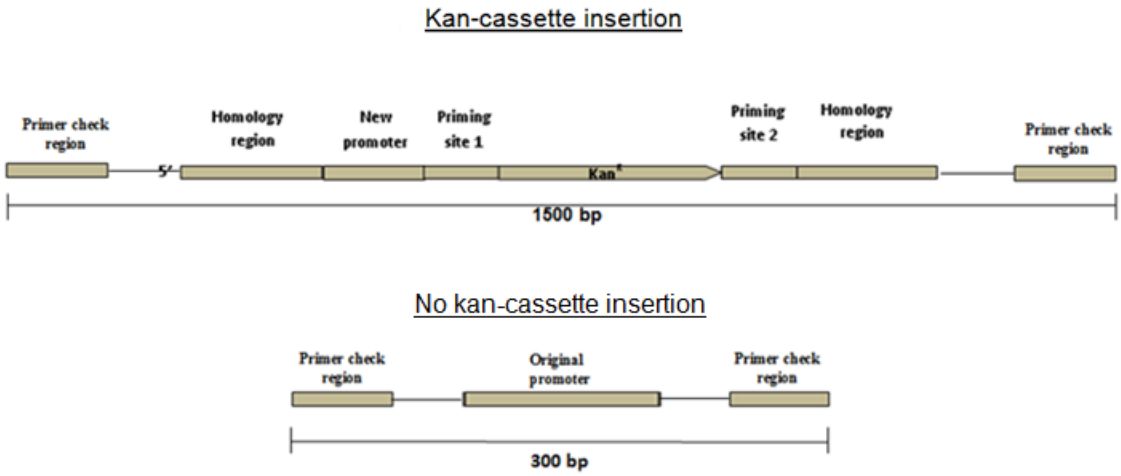


Figure 28 – Representation of the expected result for kan-cassette insertion and no insertion.

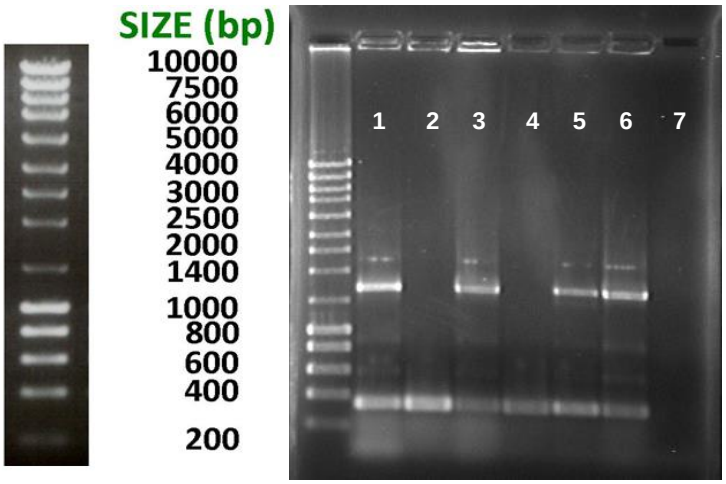


Figure 29 - Agarose gel analysis of a Colony PCR result to assess *zwf* promoter replacement. The different lanes represent 6 different colonies tested (lanes 1 and 2, zwf3; lanes 3 and 4, zwf5; lanes 5 and 6, zwf10) and a negative control (lane 7).

As seen in Figure 29 only four of the 6 colonies tested show a positive result (lanes 1, 3, 5 and 6) being represented a fragment of approximately 1560 bp correspondent to the insertion of the kan-cassette. However, a negative result is also present (a band correspondent to a fragment of approximately 345 bp) which can be due to the presence of cells with the insertion of the kan-cassette and cells without it (mixed colonies). Therefore, one positive colony of each strain was re-streaked onto a new LB/Kan30 plate in order to obtain isolated colonies. A PCR analysis was performed with a few of the new colonies and the resulting product was analyzed in agarose gel electrophoresis, yielding the gel in Figure 30.

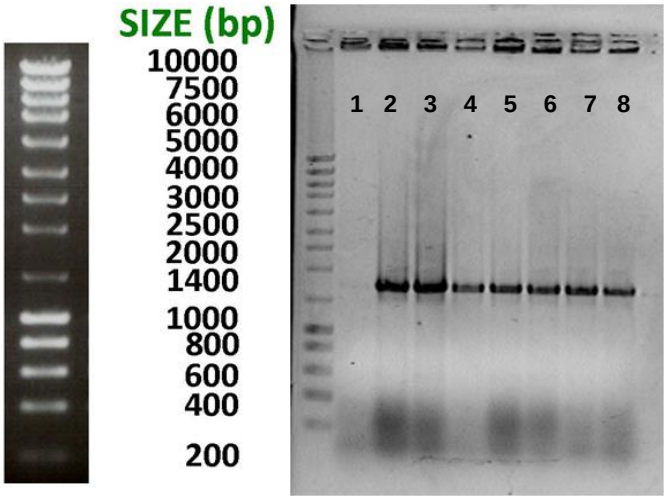


Figure 30 - Agarose gel analysis of a Colony PCR result to assess *zwf* promoter replacement. The different lanes represent 8 different colonies tested (lanes 1 to 3, zwf3; lanes 4 to 6, zwf5; lanes 7 and 8, zwf10).

With the previous procedure, colonies with only the positive result (1560 bp) were obtained. In Figure 30 it is possible to observe that the negative result (345 bp) is now absent and so the cells are ready to be used in the next experiments.

The next step was the removal of the kan-cassette from one positive colony of each strain (zwf3, zwf5 and zwf10) using pcp20 plasmid (section 2.8). After the procedure the resultant colonies (chloramphenicol resistant) were plated onto an LB plate and incubated overnight at 43°C in order to cure all the plasmids and activate the recombinase that will perform the excision of the kan-cassette. The next day a colony PCR was performed to verify if the procedure was successful and the resulting product was analyzed in agarose gel electrophoresis. The expected result is represented in Figure 31 and the obtained result in Figure 32.

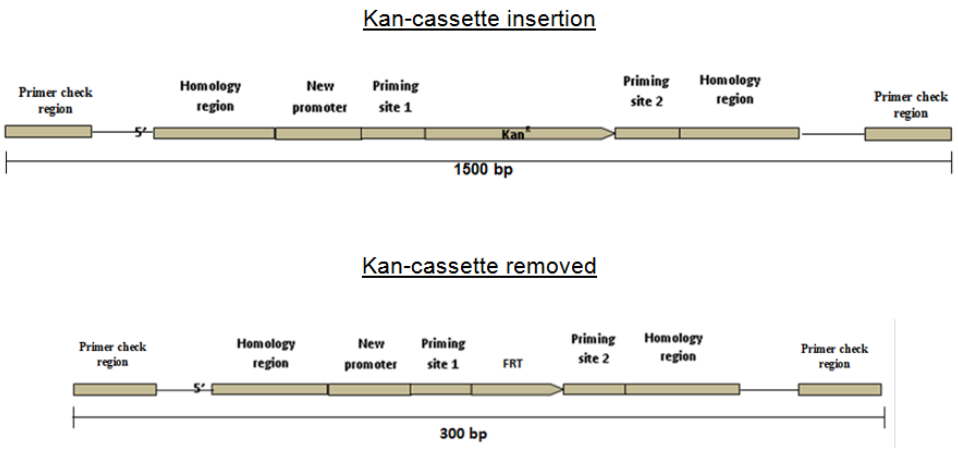


Figure 31 – Representation of the expected result for kan-cassette insertion and removal.

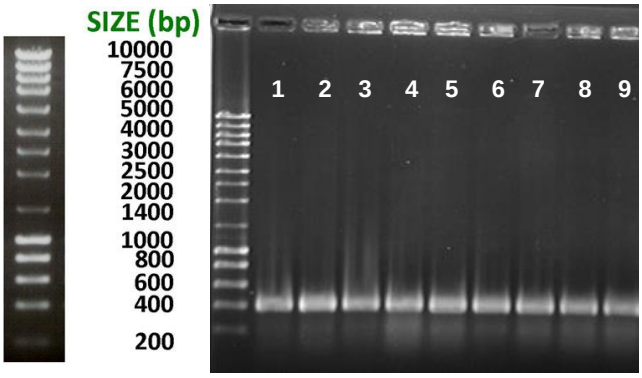


Figure 32 - Agarose gel analysis of a Colony PCR result to assess kanamycin gene removal for *zwf* promoter replacement. The different lanes represent 9 different colonies tested (lanes 1 to 3, zwf3; lanes 4 to 7, zwf5; lanes 8 and 9, zwf10).

It is possible to conclude that the kan-cassette was successfully removed from all the colonies tested since the bands represent a fragment with approximately 345 bp which is correspondent to the size of the scar left by the removal of the cassette (Figure 31 and Figure 32). Some of the these colonies were then plated onto LB plates supplemented with the different antibiotics used during the promoter replacement protocol (chloramphenicol, kanamycin and ampicilin) in order to verify if the cells do not have any antibiotic resistance. Finally, the colonies that did not grow in any of the antibiotics were picked again from the previous LB plate, submitted for sequencing and grown again in LB medium overnight. Next day seed stocks were made and stored at -80°C. The results of the sequencing are presented in Figure 33, 34 and 35.

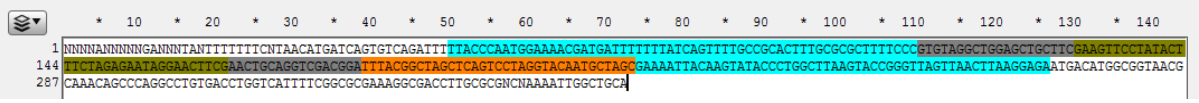


Figure 33 - Sequencing result for a zwf3 colony. The orange color highlights the presence of the promoter with strength of 0.3. The blue, grey and green colors highlight the presence of homology, priming site and FRT regions, respectively.

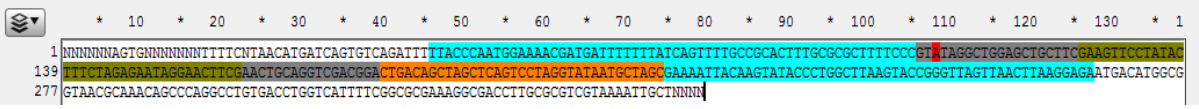


Figure 34- Sequencing result for a zwf5 colony. The orange color highlights the presence of the promoter with strength of 0.5. The blue, grey and green colors highlight the presence of homology, priming site and FRT regions, respectively.

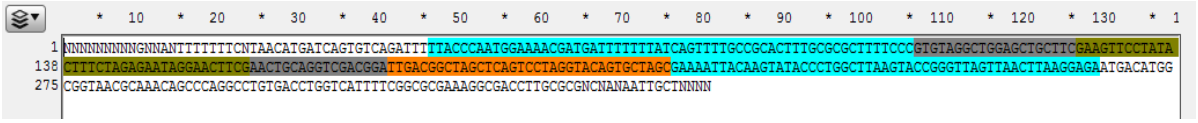


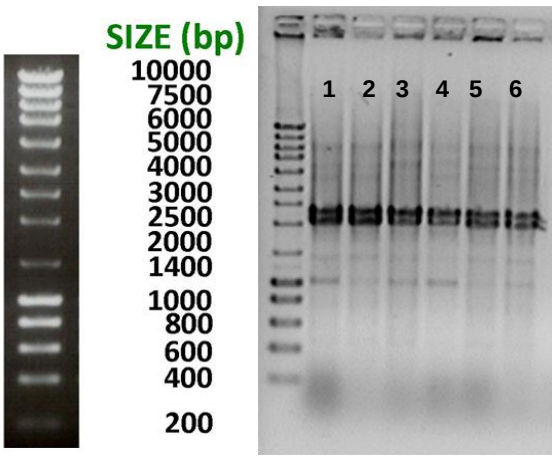
Figure 35 - Sequencing result for a zwf10 colony. The orange color highlights the presence of the promoter with strength of 1. The blue, grey and green colors highlight the presence of homology, priming site and FRT regions, respectively.

It is possible to observe that in the case of zwf5 colony, there is a mismatch in the priming site region. However, since this mutation does not alter the reading frame it is not necessary to eliminate it. In the case of zwf3 and zwf10 colonies there are no unintended mutations present. Therefore, these cells are ready to be used in the next procedures. The following step is to perform the *recA* gene knockout in order to obtain GALG20zwf3, GALG20zwf5 and GALG20zwf10 strains (MG1655*ΔendAΔrecAΔpgi zwf*$^{+}$).

3.5 Knockout of the *recA* gene

The final step in the construction of the different strains was the *recA* gene knockout. As mentioned before, this mutation is necessary in order to reduce undesirable homologous recombination events which could lead to both changes in the plasmid DNA as well as formation of plasmid multimers, causing an increase in plasmid-free cells [32].

Deletion of *recA* gene was performed by P1 transduction as explained in section 2.9. After infection with P1 phage the cells were plated onto LB/Kan30 plates and the day after analyzed by colony PCR in order to verify if the kan-cassette insertion was successful. The resulting product was then analyzed in agarose gel electrophoresis, yielding the gels in Figure 36 and 37.



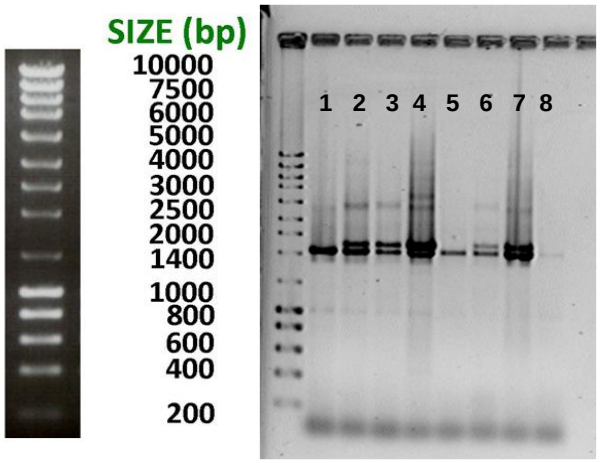
Expected results

Kan-cassette insertion – 2372 bp

recA gene – 2040 bp

Kan-cassette removal – 978 bp

Figure 36 - Agarose gel analysis of a Colony PCR result to assess kan-cassette insertion to perform *recA* gene knockout. The different lanes represent 6 different colonies tested (lanes 1 and 2, rpiA5; lanes 3 and 4, rpiA10; lanes 5 and 6, MG1655*ΔendAΔpgi*).



Expected results

Kan-cassette insertion – 2372 bp

recA gene – 2040 bp

Kan-cassette removal – 978 bp

Figure 37 - Agarose gel analysis of a Colony PCR result to assess kan-cassette insertion to perform *recA* gene knockout. The different lanes represent 8 different colonies tested (lanes 1 to 3, zwf3; lanes 4 to 6, zwf5; lanes 7 and 8, zwf10).

A positive result, correspondent to the kan-cassette insertion, is represented by a band of 2372 bp, whereas a negative result, correspondent to no kan-cassette insertion, is represented by a band of 2040 bp (original *recA* gene). As seen in Figure 36 and 37 both bands are present, as well as some unspecific bands and one at 978 bp which corresponds to the size of the scar that results from gene knockout. The presence of these unexpected bands can be due to the presence of cells with the kan-cassette insertion as well as cells without it (mixed colonies), therefore, one colony of each strain was re-streaked onto a new LB/Kan30 plate. After this step, a new colony PCR reaction and agarose gel electrophoresis analysis was performed with a few of the new colonies but the result was the same, with the three bands present. Given that, colonies containing the kan-cassette were selected

and then subjected to transformation with the pCP20 plasmid in order to remove the Kanamycin resistance gene. Next day, resultant colonies were plated onto LB plates and incubated overnight at 43°C in order to cure all the plasmids and activate the recombinase that will perform the kan-cassette excision. Finally, a colony PCR reaction was performed to verify if the gene knockout was successful and the resultant product was then analyzed in agarose gel electrophoresis, yielding the gel in Figure 38.

Expected results

Kan-cassette insertion – 2372 bp

recA gene – 2040 bp

Kan-cassette removal – 978 bp

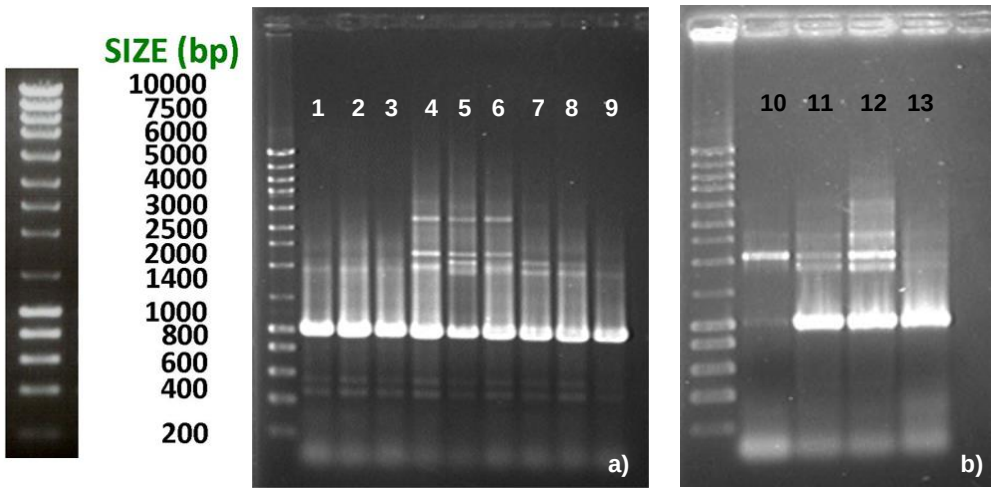
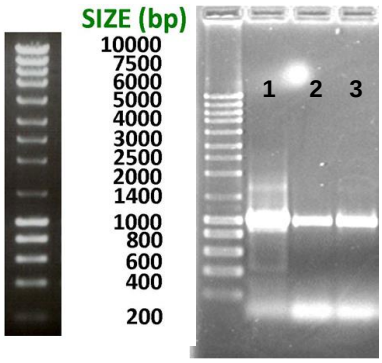


Figure 38 - Agarose gel analysis of a Colony PCR result to assess kan-cassette removal and *recA* gene knockout. a) The different lanes represent 9 different colonies tested (lanes 1 and 2, rpiA10; lanes 3 and 4, rpiA5; lanes 5 and 6, zwf10; lanes 7 and 8, zwf5; lane 9, zwf3); b) the different lanes represent 4 different colonies tested (lane 10, zwf3; lanes 11 to 13, MG1655*ΔendΔApgi*).

As observed in Figure 38, the different lanes present a band correspondent to the kan-cassette removal (978 bp) as well as other bands with approximately 2000 bp. Colonies 1 to 3, 7 to 9 and 13 were plated onto LB/Kan30 and LB/Cm50 plates and incubated overnight at 30°C in order to verify if the cells do not have any antibiotic resistance. Next day, the colonies that did not grow in any of the antibiotics are ready to be used in further experiments. These colonies were then grown again in LB medium overnight and seed stocks were made and stored at -80°C.

In the case of colonies 4 to 6 and 10 to 12, the results of the agarose gel analysis present some bands above 2000 bp. Therefore, colonies correspondent to lane 5 and 6 (zwf10) were re-streaked onto new LB plates and incubated overnight at 43°C to try to obtain an isolated positive result and eliminate the possible remaining cells containing the kan-cassette.



Expected results

Kan-cassette insertion – 2372 bp

recA gene – 2040 bp

Kan-cassette removal – 978 bp

Figure 39 - Agarose gel analysis of a Colony PCR result to assess kan-cassette removal and *recA* gene knockout, after a second 43°C incubation. The different lanes represent 3 different colonies tested (lane 1, rpiA5; lanes 2 and 3, zwf10).

After the second 43°C incubation step, the agarose gel analysis performed (Figure 39) presented an isolated positive result. Therefore, the correspondent colonies (zwf10) were plated onto LB/Kan30 and LB/Cm50 plates and incubated overnight at 30°C to verify if the cells were antibiotic resistant. The colonies that did not grow in any of the plates were picked again from the previous LB plates and grown overnight in LB medium and seed stocks were made and stored at -80°C. In addition, the cells were submitted for sequencing which results confirmed the knockout of *recA* gene.

After the sequencing steps and the assessment of the sizes of the modified regions by PCR analysis it is possible to conclude that the strains were successfully constructed containing the desirable modifications without unintended mutations. Therefore, these are ready to be evaluated through shake-flask cultivations.

3.6 Shake-flask cultivations

After the construction of the different strains it was necessary to evaluate cell growth behavior as well as pDNA production. In order to do that, all the strains were electroporated with pVAX1GFP plasmid (Kan^R) and cultivated in shake-flasks. Therefore, the strains used in shake-flask cultivations were: GALG20, GALG20rpiA5, GALG20rpiA10, GALG20zwf3, GALG20zwf5, GALG20zwf10.

The growth curves of the different strains are represented in Figure 40.

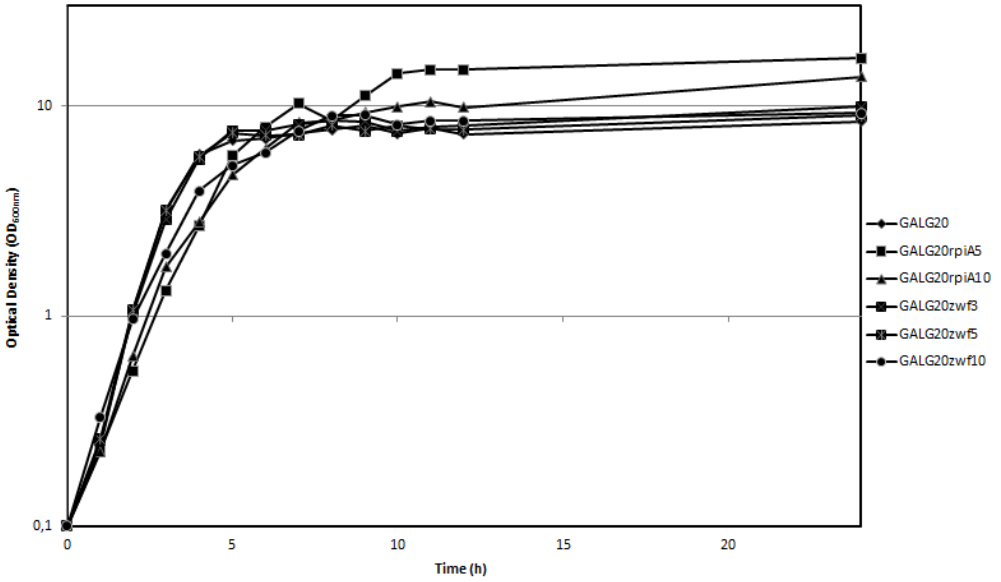


Figure 40 –Growth curves of the different *E.coli* strains used in the shake-flask cultivations.

As seen in Figure 40, all the strains showed a similar growth behavior, reaching stationary phase after approximately 6 hours. GALG20rpiA5 presented the highest growth rate (0.71±0.02 h^{-1}) while GALG20zwf10 presented the lowest (0.54±0.15 h^{-1}) (Table 7). Taking into consideration the standard deviation, the values obtained are reasonable and close to the ones described in previous studies [50,78].

Table 7- Growth rate of the strains used in the shake-flask cultivations in comparison with values obtained in previous studies.

Strain	Growth rate (h^{-1})	Growth rate from previous studies (h^{-1})
GALG20	0.68±0.05	0.77±0.05 [78]; 0.85 [50]
GALG20rpiA5	0.71±0.02	0.90 [50]
GALG20rpiA10	0.64±0.09	0.89 [50]
GALG20zwf3	0.68±0.05	-
GALG20zwf5	0.67±0.04	-
GALG20zwf10	0.54±0.15	-

The final OD_{600nm} value achieved was lower than expected, in all the strains. This might be due to problems in the cultivation conditions, such as temperature or agitation shifts, different lots of the components or different shake-flasks which can cause alterations in the aeration, among others. These parameters are hard to control and can be responsible for differences between studies. In addition, although the strains used have the same genotype as the ones from previous studies, in this

project they were constructed again starting from MG1655 strain. This might also add some variability to the study.

3.7 pDNA quantification

After the shake-flask cultivations pDNA production by the different strains was quantified using NanoVue Plus Spectrophotometer (GE Healthcare ®). pVAX1GFP plasmid was purified from the different strains using Roche kit and then its concentration was measured.

In order to obtain relevant information about pDNA productivity two parameters were determined. Specific plasmid yield (mg/g DCW) indicates plasmid productivity per gram of cells while volumetric plasmid yield (mg/L) is relative to the total productivity of the cell culture. The first parameter is very important since it demonstrates which cells produce higher amounts of DNA, while the second is dependent on the growth behavior of the cell (more biomass formation, more plasmid DNA produced). In Figure 41 and Figure 42 is represented the volumetric plasmid yield (mg/L) and the specific plasmid yield (mg/g DCW), respectively, of each strain after 12 hours of growth.

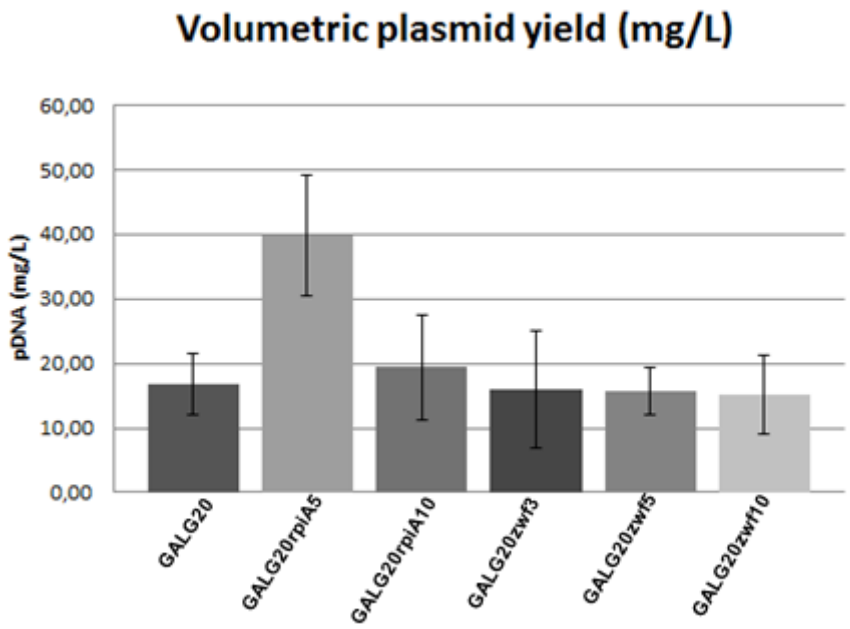


Figure 41 – Volumetric plasmid yield (mg/L) calculated for the different strains used in shake-flask cultivations.

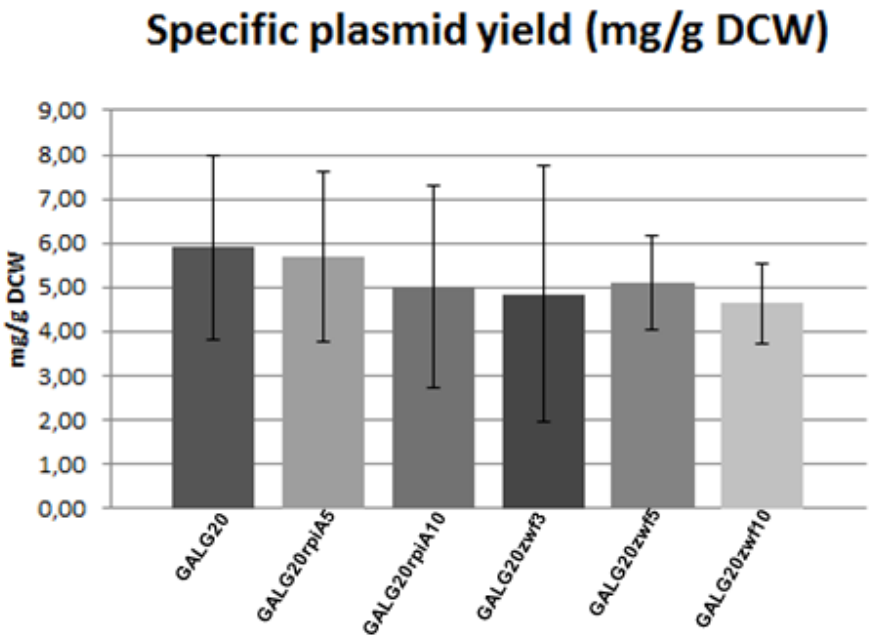


Figure 42 – Specific plasmid yield (mg/g DCW) calculated for the different strains used in shake-flask cultivations

As observed in Figures 41 and 42, all the strains produced similar amounts of DNA. In order to have a better understanding of the values obtained in this study, in Table 8 these values are compared with the ones obtained in previous studies.

Table 8 – Plasmid DNA productivity of the different strains used in shake-flask cultivations in comparison with values obtained in previous studies.

Strain	pDNA (mg/L)	pDNA (mg/L), previous studies	pDNA (mg/g DCW)	pDNA (mg/g DCW), previous studies
GALG20	16.8 ± 4.8	90.1 ± 1.9 [78] 125 [50]*	5.9 ± 2.1	10.3 ± 0.5 [78]; 13.3 [50]*
GALG20rpiA5	39.8 ± 9.3	30 [50]*	5.7 ± 1.9	3.8 [50]*
GALG20rpiA10	19.4 ± 8.2	90 [50]*	5.0 ± 2.3	9.6 [50]*
GALG20zwf3	16.0 ± 9.1	-	4.9 ± 2.9	-
GALG20zwf5	15.8 ± 3.7	-	5.1 ± 1.1	-
GALG20zwf10	15.2 ± 6.1	-	4.6 ± 0.9	-

*values obtained after 24 hours of growth

Relative to volumetric plasmid yield, GALG20rpiA5 appears to have higher productivity in comparison with the other strains. This value was expected since this strain also had higher biomass formation, achieving the highest OD_{600nm} values in comparison with the other strains (Figure 40). Therefore, with more cell biomass, more total pDNA is produced. GALG20rpiA10 had a lower volumetric yield, since it had a lower growth rate. Concerning specific plasmid yield, both GALG20rpiA

strains show a similar productivity. Wang *et al.* [34] performed a study in which overexpression of the *rpiA* gene in *E. coli* BL21 caused a 3-fold increase in plasmid copy number of a ColE1-derived plasmid during continuous culture, using defined medium and glucose as the carbon source [34]. However, since the strain used was different, the values obtained are not directly comparable. In a more recent study, it was demonstrated that increased production of Ribose-5-phosphate Isomerase permitted a faster plasmid replication until a certain point, where it becomes similar to GAL20 [50]. The calculated values were higher than the ones in this project (Table 8), however these were obtained after 24 hours of growth, while the ones relative to this study were obtained only after 12 hours of growth. In the previous study it was observed a 3.5 fold increase in the specific pDNA yield of GALG20 between 12 and 24 hours, while GALG20rpiA5 and GALG20rpiA10 displayed a 2.6 and 1,022 fold increase [50]. Regarding the productivity after 12 hours of growth, all the strains presented a value of approximately 4 mg/g DCW and 30 mg/L. Taking this into consideration, it is possible to conclude that the values obtained in the two studies are similar and the small differences can be explained by variability in the experimental conditions, as mentioned before.

GALG20zwf3, GALG20zwf5 and GALG20zwf10 strains had very similar volumetric and specific productivities. The values presented in the literature are not comparable since the strain used was DH5α, but it is documented that the overexpression of *zwf* gene alone had no effect on fermentation productivity [54]. In this case it is possible to verify that there is no significant increase in specific plasmid productive. Relative to volumetric yield, these strains were the ones with the lower values in comparison with the other strains.

GALG20 strain was expected to have high specific and volumetric plasmid DNA productivities. Since the main via of utilization of carbon in this strain is the PPP, glycolysis is down-regulated, and low amount of acetate is formed, which can be an advantage for pDNA production. Using a fed-batch strategy, this strain achieved productivities of ~160 mg/L [78]. However, the plasmid productivities of this strain in this study were lower than expected. As mentioned before, control of the cultivation parameters is difficult in the case of shake-flask cultivations and so differences between experiments are expected.

The results obtained in this study were, in general, between acceptable values. However, there were some changes in the protocol used that might explain the observed differences. Although the growth conditions were the same as in previous studies [50, 78] small unintended differences can have a big influence on the final results. Perhaps some of the components used in the growth media are from a different brand or a different lot which can influence its activity and consequently cell growth. Another important aspect is the plasmid DNA purification method. In this case, plasmid DNA purification and quantification was performed using Miniprep Kit from Roche® and NanoVue Plus Spectrophotometer (GE Healthcare ®), respectively. In the previous studies used as comparison [50, 78] plasmid DNA was quantified from crude alkaline lysates prepared from cell pellets (OD_{600nm} = 10) using the hydrophobic interaction HPLC method described before by Diogo *et al.* [21]. This method is much more precise and reduces plasmid DNA losses during the procedure, therefore resulting in higher amounts of DNA. Another possible reason related to plasmid purification method might be the capacity of the Miniprep column. It was hypothesized that the column might become clogged or

saturated and was losing pDNA. To test that hypothesis, instead of purifying a sample with OD_{600nm}=4, samples with OD_{600nm}=1 and OD_{600nm}=2 were used. But the difference in the results was not significate which indicates that this might not be the problem. Also, the use of NanoVue Plus Spectrophotometer is associated with the occurrence of errors in the measurements. To test that, some plasmid DNA samples were quantified using a standard spectrophotometer and OD_{260nm} was measured. The concentration of pDNA was calculated and the values obtained were superior relative to Nanodrop measurements with a ratio between 1 and 2. This demonstrates that the use of NanoVue Plus Spectrophotometer was not the best approach to quantify plasmid DNA, being important to compare results from the two methods. Finally, it is also known that shake-flask cultivations are not a very precise method to evaluate plasmid DNA productivity due to the lack of control over important parameters such as pH and dissolved oxygen.

4 Conclusions and future work

Plasmid DNA market is expected to increase in the next years, being necessary to create a system with high productivity and low costs. Several studies have been performed with the goal of improving both upstream and downstream processes of plasmid DNA production process. However there is still no consensus on what strategy is the best in all parameters.

The goal of this project was to verify the influence of *zwf* gene overexpression on plasmid DNA production on GALG20 strains. In order to that, three promoter replacements were performed using three different strengths (0.33, 0.51 and 1). There are a few previous studies that indicate that overexpression of this gene would have a positive impact on pDNA production by increasing the reducing power available for pDNA [52]. However, in this study a high productivity by these strains was not observed. On the contrary, they showed a similar or worse behavior when compared to the other strains evaluated. Williams *et al.* [54] verified that simultaneous overexpression of *zwf* and *rpiA* in DH5α appeared to increase plasmid amplification rate (mg pDNA/L/OD_{600}/hr). On the other hand, overexpression of *zwf* alone as well as *zwf* and thioredoxin (*trxA)* did not show any effect on fermentation productivity [54]. Therefore, as future work, it is necessary to investigate the influence of simultaneous overexpression of *rpiA* and *zwf* genes in GALG20 strains. It was previously observed that GALG20rpiA strains have an increased plasmid amplification rate [34,50] so it would be useful to test whether these strains overexpressing also *zwf* gene would produce higher amounts of plasmid DNA relative to GALG20.

The other strains evaluated in this study did not present any significant result. Highest volumetric plasmid yield was obtained by GALG20rpiA5 strain (39.8 mg/L) while the lowest was obtained by GALG20zwf10 (15.2 mg/L). Relative to specific plasmid yield, GALG20 presented a productivity of 5.9 mg/g DCW followed by GALG20rpiA5 with 5.7 mg/g DCW. Again, GALG20zwf10 gave the lowest values with only 4.6 mg/g DCW. In comparison with previous studies all the strains had lower productivities than expected, in particular GALG20. The differences observed might be explained by the difficulty in controlling certain growth parameters. Although the growth medium was the same, certain parameters are difficult to control and small changes to the protocol can have big influences on the final plasmid DNA productivity.

There are several feeding strategies that can be used in the process of pDNA production. However it is known that productivity data obtained from shake flask experiments often fail to predict the outcome of pDNA production in bench-scale bioreactors [54]. The lack of control of important parameters such as pH and dissolved oxygen can lead to high concentrations of accumulated acetate which can be toxic and inhibit biomass formation [69]. In addition, it was observed that plasmid degradation can occur at the end of the fermentation in shake flasks as well as the formation of nicked plasmids and multimers [67, 70]. In this study lower productivities were observed relative to previous studies, which might be due to the limitations of shake-flask cultivations and difficulties in controlling important parameters. Therefore, it would be important to test the strains constructed in this study using batch and fed-batch strategies. Growth conditions should also be optimized in order to try to obtain higher plasmid DNA productivities. Finally, precise plasmid DNA purification and quantification methods must be used in order to obtain reliable values.

In conclusion, there are still a lot of studies to perform in order to identify a strategy that increases plasmid productivity and quality, reducing the production costs. In addition to modifying genes related to plasmid productivity it is also necessary to increase stability and safety. It is known that plasmid productivity reaches a plateau in which the cells cannot produce higher quantities. Williams *et al.* [54] hypothesized that it was probably determined by the efficiency of the replication origin and the percent of initiated replication cycles that are completed. The amplification slope may be set by a limiting factor, perhaps a protein or nucleotide necessary for plasmid DNA replication [54]. Given that, the genes responsible for these mechanisms have been analyzed, such as genes involved in DNA replication. For instance, *topA* (Topoisomerase I)*, polA* (DNA Polymerase I)*, ligA (*DNA ligase) and *gyrAB* (DNA gyrase) are some of the genes that have been tested by overexpression or knockout. Therefore, a combination of several strategies must be performed in which different upstream and downstream processes are evaluated, namely fermentation strategy, ideal host strain and effective purification methods, with the goal of creating the best system for plasmid DNA production.

5 Bibliography

[1] Lara, A. R., & Ramírez, O. T. (2012). Plasmid DNA Production for Therapeutic Applications. (A. Lorence, Ed.)*Methods in Molecular Biology*, *824*, 271–303. doi:10.1007/978-1-61779-433-9

[2] Prazeres, D. F.; Santos, J. A. L., 2007. Production and Purification of Adenovirus Vectors for Gene Therapy. In: S. C. Gad., ed. *Handbook of Pharmaceutical Biotechnology.* New Jersey: John Wiley & Sons, Inc., pp. 1291-1289.

[3] Market Wired. [Online].; 2014 [cited 2014 September 28. Available from: http://www.marketwired.com/press-release/gene-therapy-technologies-markets-companies-report-includes-market-forecasts-2023-contains-1885236.htm

[4] The Journal of Gene Medicine. [Online].; 2014 [cited 2014 September 28. Available from: http://www.wiley.com/legacy/wileychi/genmed/clinical/

[5] Raper, S. E., Chirmule, N., Lee, F. S., et al., (2003) Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. *Molecular Genetics and Metabolism, 80* , 148–158.

[6] NP Restifo, H Ying, L Hwang, and W. L. (2000). The promise of nucleic acid vaccines. *Gene Therapy,* 7, 89–92.

[7] Clinicaltrials.gov. [Online].; 2014 [cited 2014 October 15. Available from: http://www.clinicaltrials.gov/ct2/results?term=DNA+vaccines&recr=&rslt=&type=&cond=&intr=&titles=&outc=&spons=&lead=&id=&state1=&cntry1=&state2=&cntry2=&state3=&cntry3=&locn=&gndr=&phase=2&phase=3&rcv_s=&rcv_e=&lup_s=&lup_e=

[8] Pereira, V. B., Zurita-Turk, M., Saraiva, T. D. L., De Castro, C. P., Souza, B. M., Mancha Agresti, P., … Miyoshi, A. (2014). DNA Vaccines Approach: From Concepts to Applications. *World Journal of Vaccines*, *04*(02), 50–71. doi:10.4236/wjv.2014.42008

[9] bccResearch. [Online].; 2014 [cited 2014 September 28. Available from: http://www.bccresearch.com/pressroom/bio/global-market-dna-vaccines-reach-$2.7-billion-2019

[10] bccResearch. [Online].; 2014 [cited 2014 October 14. Available from: http://www.bccresearch.com/market-research/biotechnology/dna-vaccines-markets-bio067a.html

[11] Bower, D. M., & Prather, K. L. J. (2009). Engineering of bacterial strains and vectors for the production of plasmid DNA. *Applied Microbiology and Biotechnology*, *82*(5), 805–13. doi:10.1007/s00253-009-1889-8

[12] Yau, S. Y., Keshavarz-Moore, E., & Ward, J. (2008). Host strain influences on supercoiled plasmid DNA production in *Escherichia coli*: Implications for efficient design of large-scale processes. *Biotechnology and Bioengineering*, *101*(3), 529–44. doi:10.1002/bit.21915

[13] Oliveira, P. H., Prather, K. J., Prazeres, D. M. F., & Monteiro, G. a. (2009). Structural instability of plasmid biopharmaceuticals: challenges and implications. *Trends in Biotechnology*, *27*(9), 503–11. doi:10.1016/j.tibtech.2009.06.004

[14] Levy, M. S., O'Kennedy, R. D., Ayazi-Shamlou, P., & Dunnill, P. (2000). Biochemical engineering approaches to the challenges of producing pure plasmid DNA. *Trends in Biotechnology, 18*(7), 296–305.

[15] Ertl, P.F. and Thomsen, L.L. (2003) Technical issues in construction of nucleic acid vaccines. *Methods 31*, 199–206

[16] Kreiss, P. et al. (1999) Plasmid DNA size does not affect the physicochemical properties of lipoplexes but modulates gene transfer efficiency. *Nucleic Acids Research, 27*, 3792–3798

[17] Rozkov, A. et al. (2004) Characterization of the metabolic burden on *Escherichia coli* DH1 cells imposed by the presence of a plasmid containing a gene therapy sequence. *Biotechnology and Bioengineering, 88*, 909–915

[18] Pillai, V. B., Hellerstein, M., Yu, T., Amara, R. R., Robinsona, H. L.. (2008). Comparative studies on *in vitro* expression and *in vivo* immunogenicity of supercoiled and open circular forms of plasmid DNA vaccines. *Vaccine 26*, 1136 - 1141.

[19] Azzoni, A.R. et al. (2007) The impact of polyadenylation signals on plasmid nuclease-resistance and transgene expression. *The Journal of Gene Medicine, 9*, 392–402

[20] FDA. Guidance for industry: considerations for plasmid DNA vaccines for infectious disease indications. US Food and Drug Administration, 2007

[21] Diogo, M. M., Queiroz, J. a, & Prazeres, D. M. F. (2003). Assessment of purity and quantification of plasmid DNA in process solutions using high-performance hydrophobic interaction chromatography. *Journal of chromatography A, 998*(1-2), 109–17. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12862377

[22] Martins S., Prazeres D., Cabral J., Monteiro G. (2003) Comparison of real-time PCR and hybridization assays for the detection of Escherichia coli genomic DNA in process samples and pharmaceutical-grade plasmid DNA products. *Analytical Biochemistry, 322*, 127-129.

[23] Mairhofer, J., Cserjan-Puschmann, M., Striedner, G., Nöbauer, K., Razzazi-Fazeli, E., & Grabherr, R. (2010). Marker-free plasmids for gene therapeutic applications--lack of antibiotic resistance gene substantially improves the manufacturing process. *Journal of Biotechnology, 146*(3), 130–7. doi:10.1016/j.jbiotec.2010.01.025

[24] Chen, X., Zhou, L., Tian, K., Kumar, A., Singh, S., Prior, B. a, & Wang, Z. (2013). Metabolic engineering of *Escherichia coli*: A sustainable industrial platform for bio-based chemical production. *Biotechnology Advances, 31*(8), 1200–23. doi:10.1016/j.biotechadv.2013.02.009

[25] Gonçalves, G. a L., Bower, D. M., Prazeres, D. M. F., Monteiro, G. A., & Prather, K. L. J. (2012). Rational engineering of *Escherichia coli* strains for plasmid biopharmaceutical manufacturing. *Biotechnology Journal*, *7*(2), 251–61. doi:10.1002/biot.201100062

[26] Glenting, J., & Wessels, S. (2005). Ensuring safety of DNA vaccines. *Microbial Cell Factories, 4*, 26. doi:10.1186/1475-2859-4-26

[27] Bachmann, B. J., (1996). Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12. In: Neidhardt FC, et al. (ed), *Escherichia coli and Salmonella: cellular and molecular biology.* Washington, DC: ASM Press, 2nd ed, , pp 2460–2488.

[28] Singer, A., Eiteman, M. a, & Altman, E. (2009). DNA plasmid production in different host strains of *Escherichia coli*. *Journal of Industrial Microbiology & Biotechnology*, *36*(4), 521–30. doi:10.1007/s10295-008-0522-7

[29] Gonçalves, G. a L., Prazeres, D. M. F., Monteiro, G. a, & Prather, K. L. J. (2013). De novo creation of MG1655-derived *E. coli* strains specifically designed for plasmid DNA production. *Applied Microbiology and Biotechnology, 97*(2), 611–20. doi:10.1007/s00253-012-4308-5

[30] Pósfai, G., Plunkett, G., Fehér, T., Frisch, D., Keil, G. M., Umenhoffer, K., … Blattner, F. R. (2006). Emergent properties of reduced-genome *Escherichia coli*. *Science (New York, N.Y.)*, *312*(5776), 1044–6. doi:10.1126/science.1126439

[31] Cranenburgh RM, Hanak JA, Williams SG, Sherratt DJ (2001). *Escherichia coli* strains that allow antibiotic-free plasmid selection and maintenance by repressor titration. *Nucleic Acids Research, 29*:E26

[32] Hodgson CP, Williams JA (2006). Improved strains of *E. coli* for plasmid DNA production. WO 2006/026125 A2

[33] Ow, D. S.-W., Nissom, P. M., Philp, R., Oh, S. K.-W., & Yap, M. G.-S. (2006). Global transcriptional analysis of metabolic burden due to plasmid maintenance in *Escherichia coli* DH5α during batch fermentation. *Enzyme and Microbial Technology, 39*(3), 391–398. doi:10.1016/j.enzmictec.2005.11.048

[34] Wang, Z., Xiang, L., Shao, J., Wegrzyn, A., & Wegrzyn, G. (2006). Effects of the presence of ColE1 plasmid DNA in *Escherichia coli* on the host cell metabolism. *Microbial Cell Factories, 5*, 34. doi:10.1186/1475-2859-5-34

[35] Rozkov A, Avignone-Rossa CA, Ertl PF, Jones P, O'Kennedy RD, Smith JJ, Dale JW, Bushell ME. (2004). Characterization of the metabolic burden on *Escherichia coli* DH1 cells imposed by the presence of a plasmid containing a gene therapy sequence. *Biotechnology and Bioengineering, 88*, 909-915.

[36] EcoCyc. [Online].; 2014 [cited 2014 October 22. Available from: http://ecocyc.org/ECOLI/NEW-IMAGE?type=PATHWAY&object=GLYCOLYSIS

[37] Nature Education. [Online].; 2014 [cited 2014 October 22. Available from: http://www.nature.com/scitable/content/glycolysis-14897204

[38] NCBI. [Online].; 2002 [cited 2014 October 22. Available from: http://www.ncbi.nlm.nih.gov/books/NBK22416/

[39] Nature Cell Biology. [Online].; 2013 [cited 2014 October 22. Available from: http://www.nature.com/ncb/journal/v15/n8/fig_tab/ncb2810_F1.html

[40] Nature Reviews Microbiology. [Online].; 2014 [cited October 22. Available from: http://www.nature.com/nrmicro/journal/v5/n10/fig_tab/nrmicro1746_F4.html

[41] Koffas, M., Roberge, C., Lee, K., & Stephanopoulos, G. (1999). Metabolic Engineering. *Annual Review of Biomedical Engineering, 01*(96), 535–557.

[42] EcoCyc. [Online].; 2014 [cited 2014 October 22. Available from: http://www.ecocyc.org/ECOLI/NEW-IMAGE?type=ENZYME&object=EG11336-MONOMER

[43] EcoCyc. [Online].; 2014 [cited 2014 October 22. Available from: http://www.ecocyc.org/ECOLI/NEW-IMAGE?type=GENE&object=EG10823

[44] Phue, J.-N., Lee, S. J., Trinh, L., & Shiloach, J. (2008). Modified *Escherichia coli* B (BL21), a superior producer of plasmid DNA compared with *Escherichia coli* K (DH5alpha). *Biotechnology and Bioengineering, 101*(4), 831–6. doi:10.1002/bit.21973

[45] Carnes, A. E., Luke, J. M., Vincent, J. M., Schukar, A., Anderson, S., Hodgson, C. P., & Williams, J. a. (2011). Plasmid DNA fermentation strain and process-specific effects on vector yield, quality, and transgene expression. *Biotechnology and Bioengineering, 108*(2), 354–63. doi:10.1002/bit.22936

[46] Luli, G. W., & Strohl, W. R. (1990). Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Applied and Environmental Microbiology, 56*(4), 1004–11.

[47] Al Zaid Siddiquee, K., Arauzo-Bravo, M. J., & Shimizu, K. (2004). Metabolic flux analysis of *pykF* gene knockout *Escherichia coli* based on ^{13}C-labeling experiments together with measurements of enzyme activities and intracellular metabolite concentrations. *Applied Microbiology and Biotechnology, 63*(4), 407–17. doi:10.1007/s00253-003-1357-9

[48] Cunningham, D. S., Liu, Z., Domagalski, N., Koepsel, R. R., Ataai, M. M., & Domach, M. M. (2009). Pyruvate kinase-deficient *Escherichia coli* exhibits increased plasmid copy number and cyclic AMP levels. *Journal of Bacteriology, 191*(9), 3041–9. doi:10.1128/JB.01422-08

[49] EcoCyc. [Online].; 2014 [cited 2014 October 23. Available from: http://www.ecocyc.org/ECOLI/NEW-IMAGE?type=GENE&object=EG11443

[50] Carreira, L. (2013). Evaluation of *rpiA* promoter strength in plasmid DNA production. Master thesis, Instituto Superior Técnico da Universidade de Lisboa, Lisboa

[51] Flores, S., de Anda-Herrera, R., Gosset, G., & Bolívar, F. G. (2004). Growth-rate recovery of *Escherichia coli* cultures carrying a multicopy plasmid, by engineering of the pentose-phosphate pathway. *Biotechnology and Bioengineering, 87*(4), 485–94. doi:10.1002/bit.20137

[52] Cunningham, D. S., Koepsel, R. R., Ataai, M. M., Domach, M. M., (2009). Factors affecting plasmid production in *Escherichia coli* from a resource allocation standpoint. *Microbial Cell Factories*, *8*, 1–27.

[53] Zhao, J., Baba, T., Mori, H., & Shimizu, K. (2004). Effect of *zwf* gene knockout on the metabolism of *Escherichia coli* grown on glucose or acetate. *Metabolic Engineering, 6*(2), 164–74. doi:10.1016/j.ymben.2004.02.004

[54] Williams, J. a, Luke, J., Langtry, S., Anderson, S., Hodgson, C. P., & Carnes, A. E. (2009). Generic plasmid DNA production platform incorporating low metabolic burden seed-stock and fed-batch fermentation processes. *Biotechnology and Bioengineering, 103*(6), 1129–43. doi:10.1002/bit.22347

[55] Ow, D. S.-W., Lee, R. M.-Y., Nissom, P. M., Philp, R., Oh, S. K.-W., & Yap, M. G.-S. (2007). Inactivating FruR global regulator in plasmid-bearing *Escherichia coli* alters metabolic gene expression and improves growth rate. *Journal of Biotechnology, 131*(3), 261–9. doi:10.1016/j.jbiotec.2007.07.508

[56] Ow, D. S.-W., Yap, M. G.-S., Oh, S. K.-W., (2009). Enhancement of plasmid DNA yields during fed-batch culture of a *fruR*- knockout *Escherichia coli* strain. *Biotechnology and Applied Biochemistry,* 52, 53–59

[57] EcoCyc. [Online].; 2014 [cited 2014 October 23. Available from: http://ecocyc.org/ECOLI/NEW-IMAGE?type=GENE&object=EG10702

[58] De Anda, R., Lara, A. R., Hernández, V., Hernández-Montalvo, V., Gosset, G., Bolívar, F., & Ramírez, O. T. (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. *Metabolic Engineering, 8*(3), 281–90. doi:10.1016/j.ymben.2006.01.002

[59] Borja, G. M., Meza Mora, E., Barrón, B., Gosset, G., Ramírez, O. T., & Lara, A. R. (2012). Engineering *Escherichia coli* to increase plasmid DNA production in high cell-density cultivations in batch mode. *Microbial Cell Factories, 11*(40), 132. doi:10.1186/1475-2859-11-132

[60] EcoCyc. [Online].; 2014 [cited 2014 October 24. Available from: http://www.ecocyc.org/ECOLI/NEW-IMAGE?type=GENE&object=EG10664

[61] Matsui H, Kawasaki H, Shimaoka M, Takenaka Y, Kurahashi O: Method for producing purine nucleoside by fermentation. US Patent: 7, 608,432 B2.; 2009

[62] 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*(1), 14. doi:10.1186/2042-5783-1-14

[63] Magnusson, L. U., Farewell, A., & Nyström, T. (2005). ppGpp: a global regulator in *Escherichia coli*. *Trends in Microbiology, 13*(5), 236–42. doi:10.1016/j.tim.2005.03.008

[64] Carnes, A., & Williams, J. (2007). Plasmid DNA Manufacturing Technology. *Recent Patents on Biotechnology, 1(*2), 151–166. doi:10.2174/187220807780809436

[65] Gonçalves, G. a L., Prather, K. L. J., Monteiro, G. a, Carnes, A. E., & Prazeres, D. M. F. (2014). Plasmid DNA production with *Escherichia coli* GALG20, a pgi-gene knockout strain: Fermentation strategies and impact on downstream processing. *Journal of Biotechnology, 186*, 119–27. doi:10.1016/j.jbiotec.2014.06.008

[66] Peterson M, Brune B. (2008). Maximizing yields of plasmid DNA processes - High cell density processes can produce high yields without compromising quality. *BioPharm International., 16*.

[67] Tejeda-Mansir, A., & Montesinos, R. M. (2008). Upstream processing of plasmid DNA for vaccine and gene therapy applications. *Recent Patents on Biotechnology*, *2*(3), 156–72.

[68] Zhi-nan, X. U., Wen-he, S., Hao, C., & Pei-lin, C. E. N. (2005). Effects of medium composition on the production of plasmid DNA vector potentially for human gene therapy. *Journal of Zhejiang University SCIENCE*, *6B*(5), 396–400. doi:10.1631/jzus.2005.B0396

[69] Gonçalves, G. a L., Prather, K. L. J., Monteiro, G. a, & Prazeres, D. M. F. (2014). Engineering of *Escherichia coli* strains for plasmid biopharmaceutical production: scale-up challenges. *Vaccine*, *32*(24), 2847–50. doi:10.1016/j.vaccine.2014.02.023

[70] Kennedy, R. D. O., Ward, J. M., Keshavarz-moore, E., Street, G., & Wce, L. (2003). Effects of fermentation strategy on the characteristics of plasmid DNA production. *Biotechnology and Applied Biochemistry*, *37*, 83–90.

[71] Gonçalves, G. a L., Prather, K. L. J., Monteiro, G. a, Carnes, A. E., & Prazeres, D. M. F. (2014). Plasmid DNA production with *Escherichia coli* GALG20, a pgi-gene knockout strain: Fermentation strategies and impact on downstream processing. *Journal of Biotechnology, 186*, 119–27. doi:10.1016/j.jbiotec.2014.06.008

[72] James A Williams, Aaron E Carnes, and C. P. H. (2009). Plasmid DNA Vaccine vector design: impact on efficacy, safety and upstream production. *Biotechnology Advances, 27*(4), 353–370. doi:10.1016/j.biotechadv.2009.02.003.Plasmid

[73] O'Mahony, K., Freitag, R., Hilbrig, F., Müller, P., & Schumacher, I. (2007). Strategies for high titre plasmid DNA production in *Escherichia coli* DH5α. *Process Biochemistry, 42*(7), 1039–1049. doi:10.1016/j.procbio.2007.03.014

[74] Carnes, A. E., Hodgson, C. P., & Williams, J. A. (2006). Inducible *Escherichia coli* fermentation for increased plasmid DNA production. *Biotechnology and Applied Biochemistry, 45*, 155–166. doi:10.1042/BA20050223

[75] Cooke GD, Cranenburgh RM, Hanak JAJ, Dunnill P, Thatcher DR, Ward JM (2001). Purification of essentially RNA free plasmid DNA using a modified *Escherichia coli* host strain expressing ribonuclease A. *Journal of Biotechnology, 85*, 297–304

[76] Registry of Standard Biological Parts. [Online]. [cited 2014 October 25. Available from: http://parts.igem.org/Promoters/Catalog/Constitutive

[77] Registry of Standard Biological Parts. [Online]. [cited 2014 October 25. Available from: http://parts.igem.org/Promoters/Catalog/Anderson

[78] Gonçalves, G. (2013). Rational engineering of *Escherichia coli* strains for plasmid biopharmaceutical manufacturing. PhD thesis, Instituto Superior Técnico da Universidade de Lisboa, Lisboa

[79] Template plasmid pKD13, complete sequence. [Online]. [cited 2014 October 25. Available from: http://www.ncbi.nlm.nih.gov/nuccore/15554335

[80] Red recombinase plasmid pKD46, complete sequence. [Online]. [cited 2014 October 25. Available from: http://www.ncbi.nlm.nih.gov/nuccore/AY048746.1

[81] Doublet, B., Douard, G., Targant, H., Meunier, D., Madec, J.-Y., & Cloeckaert, A. (2008). Antibiotic marker modifications of lambda Red and FLP helper plasmids, pKD46 and pCP20, for inactivation of chromosomal genes using PCR products in multidrug-resistant strains. *Journal of Microbiological Methods, 75*(2), 359–61. doi:10.1016/j.mimet.2008.06.010

[82] Cherepanov, P. P., & Wackernagel, W. (1995). Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene, 158*(1), 9–14.

[83] Ribeiro, S. C., Monteiro, G. a, & Prazeres, D. M. F. (2004). The role of polyadenylation signal secondary structures on the resistance of plasmid vectors to nucleases. *The Journal of Gene Medicine*, *6*(5), 565–73. doi:10.1002/jgm.536

[84] Gehl, J. (2003). Electroporation: theory and methods, perspectives for drug delivery, gene therapy and research. *Acta Physiologica Scandinavica, 177*(4), 437–47. doi:10.1046/j.1365-201X.2003.01093.x

[85] 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), 6640–5. doi:10.1073/pnas.120163297

[86] Thomason, L. C., Costantino, N., & Court, D. L. (2007). *E . coli* Genome Manipulation by P1 Transduction. *Current Protocols in Molecular Biology*, *1.17*(July), 1–8. doi:10.1002/0471142727.mb0117s79

[87] Microbial genetics. [Online]. [cited 2014 October 26. Available from: http://classes.midlandstech.edu/carterp/Courses/bio225/chap08/Microbial%20Genetics%206.htm

[88] Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., … Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, *2*. doi:10.1038/msb4100050