Construction of *Escherichia coli* strains for higher plasmid production: Evaluation of *lon* gene knockout and implementation of a marker-free system

Diana Sofia Cardoso Marques

Thesis to obtain the Master of Science Degree in

Biotechnology

Supervisors: Professor Doutor Gabriel António Amaro Monteiro
Professor Doutor Duarte Miguel de França Teixeira dos Prazeres

Examination Committee

Chairperson: Professor Doutor Arsénio do Carmo Sales Mendes Fialho, IST
Supervisor: Professor Doutor Gabriel António Amaro Monteiro, IST
Member of the Committee: Professor Doutor Jorge Humberto Gomes Leitão, IST

November 2014

“Science is not only a disciple of reason but, also, one of romance and passion.”

Stephen Hawking

Acknowledgments

First, I would like to thank to my supervisors, Professor Doutor Gabriel Monteiro and Professor Doutor Miguel Prazeres for accepting me and giving me the opportunity to work in this master thesis where I have learned so much and whose subject fascinates me immensely. I would also like to thank for all their help, guidance, availability, enormous patience and for receiving me even in less convenient times.

I would also like to thank to Jürgen Mairhofer, from Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria, for guidance, comprehension and availability.

Then, I would like to thank to Luís Carreira for helping me in the early days in the laboratory, for his patience, and for having spent the time to teach me most molecular biology techniques used throughout this work.

Furthermore, I would like to thank to my lab partners for all the help, patience to teach me and enlighten me and for all the counseling. First, João Trabuco for always being there for me, for all the encouragement, for taking the time to help me solve problems, and for his kind words, which so often helped to overcome adversity. Pedro Pereira for always supporting me, giving a friendly word in the most difficult moments and for all the help he gave me during this year. Sofia Duarte, for all patience, for always being available to clarify my doubts and for the useful suggestions to overcome the laboratory problems. My colleagues, Salomé Magalhães, Jorge Paulo, Ana Rosa, Marina Monteiro and Luis Raiado for all the guidance, support and counseling. Also, to Ricardo Pereira, for the constant support, availability and technical help; and to D. Rosa Gonçalves for the sympathy, generosity and for maintaining the lab material clean and organized.

I would also like to make a special acknowledgment to Rita Fernandes, for being the best lab partner I could ask for. Also, for all the unconditional support, companionship and for helping me to overcome the difficulties and worst moments.

To my friends and colleagues who were with me during this year in science: Isabel Pinto, Liliana Brito, Ana Faria, Cátia Jorge, Fábio Gonçalves, Carlos Rodrigues, Maria Vieira, Silvia Andrade, Nicole Rodrigues and António Grilo. To all of them, a special thank you for the friendship, the great times, for all the laughs, and for making this year unforgettable.

I could not fail to thank my friends, who were always available to me, always comforted me, and for being with me all these years. I would like to give special thanks to Tiago Mateus for believing so much in me, for all his patience, and for never letting me give up.

Finally, I would like to thank to my family. To my grandparents for keeping faith in me, for all the care, love and concern. To my sister, for all the patience, all the support, for the smile on days I needed most and for always being by my side. To my parents for supporting my dreams, believing in me so much, always being by my side, for all the love, concern, unconditional support and advices.

Thanks to all of you.

Abstract

Plasmid DNA (pDNA) vectors offer considerable benefits over viral systems in gene-based therapy and vaccination applications, including higher shelf stability, low immunogenicity and toxicity, and simple manufacture in large scale. However, also shows low transfection efficacy, and requires milligram scale of pharmaceutical-grade pDNA per patient, which implies extensive production efforts. In order to fulfil high pDNA production requirements, genetically engineered strains specially designed to achieve high pDNA yields are required. With this purpose, a Lon-protease deficient GALG20 strain was created, in this study. However, contrary to expected, GALG20Δ*lon* and GALG20 strains achieved similar volumetric and specific plasmid production yields. In order to fully evaluate *lon* gene mutation potential in pDNA production, large-scale cultivations using a fed-batch strategy, with reduced amounts of amino acids, and testing different temperature conditions should be performed.

Despite all the advantages of using pDNA, some safety concerns were raised due to possible integration of sequences from plasmid into cellular DNA or spread to intestinal bacteria. Moreover, the presence of such sequences in plasmid backbone was associated with diminished transfection efficiency and additional metabolic burden to the host during manufacture. In an attempt to address this problem, the other goal of this project was the transference of a marker free system from JM108*murselect* strain into GALG20, in order to achieve higher pDNA yields than the ones verified in JM108*murselect*. It was not possible to transfer the system to GALG20 probably because JM108*murselect* is *recA* deficient.

Keywords

Plasmid DNA

Antibiotic free system

Strain engineering

Lon protease

DNA vaccine

Plasmid selection

Resumo

O uso de DNA plasmídico (pDNA) como vetor oferece benefícios consideráveis em relação aos sistemas virais em aplicações como terapia génica e vacinação, como maior estabilidade, baixa imunogenicidade e toxicidade e simplicidade de produção em larga escala. No entanto, também apresenta baixa eficiência de transfecção e requere quantidades elevadas de pDNA por paciente, o que requer grandes esforços de produção. Com o objectivo de responder às elevadas necessidades de produção, é essencial o desenvolvimento de estirpes geneticamente modificadas especialmente para a produção de pDNA com elevados rendimentos. Com este intuito, foi desenvolvida, neste estudo, uma estirpe GALG20 deficiente na produção da protease Lon. Contrariamente ao esperado, as duas estirpes atingiram rendimentos específicos e volumétricos semelhantes. Para avaliar plenamente o impacto da mutação no gene *lon* na produção de DNA plasmídico, deveriam ter sido realizadas culturas em larga escala usando uma estratégia fed-batch, com quantidades reduzidas de aminoácidos e testando diferentes temperaturas.

Apesar de todas as vantagens associadas ao uso de DNA plasmídico como vector, foi demonstrada alguma preocupação relativas á sua segurança devido à possibilidade de integração de sequencias do plasmidio no DNA cromossomal do paciente, bem como em relação à possível propagação destas sequencias para as bactérias intestinais. Para além disso, a presença destas sequências nos plasmídeos está associada a uma menor eficiência de transfecção e a maior carga metabólica para o hospedeiro. Numa tentativa de solucionar este problema, o objectivo deste estudo passava também pela transferência de um sistema sem selecção associada a resistência a antibióticos, da estirpe JM108*murselect* para a estirpe GALG20, com o intuito de atingir maiores rendimentos na produção de DNA plasmídico. A transferência do sistema não foi possível, provavelmente pelo facto de a estirpe JM108*murselect* não expressar a recombinase RecA.

Palavras-chave

DNA plasmídico

Sistema livre de resistência a antibióticos

Engenharia de estirpes

Protease Lon

Vacinas de DNA

Seleção de plasmídeos

Contents

List of Figures

List of Tables

List of abbreviations

BGH - Bovine Growth Hormone

bp - base-pairs

CMV- Cytomegalovirus

DCW- Dry cell weight

DO- dissolved oxygen

EF-1α - human elongation factor 1α

EEU- Eukaryotic Expression Unit

E4P - erythrose-4-phosphate

Gnd - 6-phosphogluconate dehydrogenase

GT- Gene therapy

HCDC - high cell-density cultivations

HIV- Human Immunodeficiency virus

HPV- Human Papillomavirus

ISS - immune stimulatory DNA sequences

Kan^R - kanamycin resistance

LB - Luria Broth

MDS - multiple-deletion series

MHC- major histocompatibility complex

MurA - UDP-N-actetylglucosamine enolpyruvyl-transferase

Ori - Origin of replication

ORT- Operator–Repressor titration

pCOR -plasmid with conditional origin of replication

pDNA - plasmid DNA

PolyA - Polyadenylation

PCN - Plasmid Copy Number

pFAR - plasmids free of antibiotic resistance

PP - pentose phosphate pathway

PPU- Plasmid Propagation Unit

SC – Supercoiled

SEC - size-exclusion chromatography

SV40 - simian virus 40

TetA - tetracycline efflux protein

trxA -thioredoxin

UbC - ubiquitin C

UTR - untranslated leader region

1 Introduction

1.1 Plasmid DNA as a biotechnology product

Plasmid DNA (pDNA) is a new generation biotechnology's product that is just beginning to enter the marketplace [1]. The demand for this product has increased vastly in response to rapid advances in its use in gene therapy and vaccination [2], [3], and is being considered as a promising alternative to traditional protein vaccines or viral delivery methods for gene therapeutic applications [4], [5].

Plasmids are extra-chromosomal self-replicating cytoplasmic DNA elements present in archaea, eubacteria, and some yeast [6], [7], [8]. Microorganisms receive advantages from the acquired plasmids, once they can code for additional genetic information such as resistances against antibiotics or heavy metals, enzymes participating in the catabolism of unusual substrates, or may even confer the possibility to form surface antigens and toxins [6], [7].

During the last 40 years, plasmids were probably the molecular tool most widely used for DNA manipulation, transfer, and gene expression in a variety of microorganisms and animal cells, playing a crucial role in the development of biotechnology [6]. In the field of pharmaceutical biotechnology, plasmids have attracted much attention, for DNA vaccines and gene therapy applications [2], [3]. Gene therapy and DNA vaccines are based on the same principle, i.e. the introduction of nucleic acids in human/non-human receptor cells to restore, cancel, enhance or introduce a biochemical function [2]. In gene therapy, the introduction of a gene to specific cells in a patient aims to produce therapeutic biomolecules, usually proteins, to correct or to modulate disease, whereas in DNA-based vaccination the introduction of a gene to specific cells in the patient aims to produce the intended vaccine antigen to induce an immune response [3], [9].

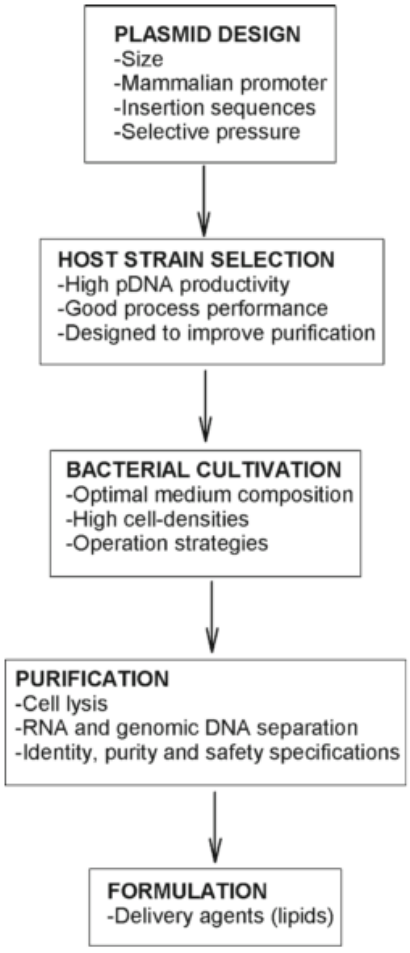

Figure 1 - Overview of the main steps for pDNA vaccine design and production. [6]

Gene therapy and DNA vaccination require the identification of the gene or genes related to a particular disease (inherited or acquired), the cloning of the therapeutic gene into a properly designed molecular vector (and its formulation), and the introduction of the gene into the patient [6]. The production of the plasmid encoding a gene of interest, for vaccination, begins with the design of the plasmid containing the gene of interest and the subsequent transformation into a bacterial cell, typically *Escherichia coli* [1], [2]. Cells are then

propagated in a bioreactor, in order to attain high yields of the plasmid. After cell recovery and lysis, plasmid DNA purification takes place, utilizing one of a plurality of purification methods and is finally formulated for delivery. After delivery to the eukaryotic cell, the gene of interest should be expressed [1], [6] (Figure 1).

Naked plasmid DNA as a vector has attracted a lot of interest since it offers several advantages over a viral vector. Both viral and non-viral vectors can carry out the transport of the transgenes towards the nuclei of the target cells [2], [10]. The use of non-viral vectors is expected to circumvent some of the problems occurring with viral vectors such as endogenous virus recombination, oncogenic effects and unexpected immune response [11]. Also naked DNA shows weak immunogenicity, simplicity of use, better safety [2], [12], [13] extended stability, which can be an advantage in vaccinating populations where sophisticated storage technologies are not available [14], [15]; and ease to manufacture [2], [12], [16], [17] however it also presents low transfection efficacy [2], [13], [15]. Non-viral gene therapy may require considerable amounts (milligram scale) of pharmaceutical-grade pDNA per patient since the efficacy and duration of gene expression is presently relatively low [2], [6]. For instances, and as an indication, gene therapy trials have used doses as high as 1250 μg (cystic fibrosis,) or 4000 μg (limb ischemia). Furthermore, repeated administration is required. This implies extensive production efforts [2].

The milligram range doses, the large patient populations and the productivity levels are inadequate for economically viable plasmid production. Although a few publications mention large scale plasmid production processes achieving productivity levels of several hundreds of milligrams per liter, which represents a large improvement over standard laboratory processes, significant process improvements continue to be necessary [2]. High yield pDNA processes are required to exploit the advantages associated with DNA vaccines and gene therapy, as well as to address new treatment options [2], [11]. In order to overcome the problems associated with the large amount of pDNA required, it is possible to use different adjuvants, and/or gene delivery methods such as needle-free jet injection, electroporation or a gene gun [6]. Currently, most of the methods to increase pDNA production are related with the purification steps, and much less is known about how to optimize the fermentation step. Through optimization of the fermentation, growth environment and biological system several improvements can be achieved in biomass productivity, plasmid yield, plasmid quality and production costs [2], [11].

1.2 Plasmid DNA market

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 [1]. DNA vaccines first sparked the interested of the scientific community in the early 1990s [18], especially when in 1993, two different research groups demonstrated that immunization with a DNA plasmid could protect mice against a lethal influenza challenge [13], [19]. The ability to elicit both humoral and cellular immune responses [18], [20], [21] combined with the versatility to a variety of animal models for a wide range of disease targets, and the simplicity of this vaccine approach generated a great deal of excitement and inspired additional

preclinical studies targeting a plethora of viral and non-viral antigens [18], [22], [19]. The first of several phase I trials [23], evaluated the efficacy of a DNA vaccine targeting human immunodeficiency virus type 1 (HIV-1) for therapeutic and prophylactic applications. Other studies shortly followed that targeted cancer or other HIV-1 antigens, influenza, human papillomavirus (HPV), hepatitis, and malaria [6], [24]. These studies demonstrated that DNA vaccines are well tolerated and safe, and there were no evidence of integration of DNA into host chromosomes, autoimmunity, or immunologic tolerance [25], [26].

This exponentially expanding interest has been demonstrated by the increase in number of patents in the area of DNA vaccine (from 1 in 1990 to 1949 in 2012) [27], [28]. Some of the most relevant are: “DNA vaccines against Rotavirus infections”, patented in 2000 by Hermann *et al*.; “Nucleotide vector vaccine for immunization against hepatitis”, in 2003, by Davis *et al*.; “DNA vaccination for treatment of multiple sclerosis and insulin dependent diabetes mellitus” patented in 2006, by Steinman *et al*. and also, “Composition and methods for vaccination against HSV-2”, in 2011, by Vilalta *et al*. [29]. Despite the good results obtained with the commercialized veterinary vaccines, no licensed human DNA vaccines exist [19], [18], [30]. In 2011 there were 43 clinical trials evaluating DNA vaccines for viral and nonviral diseases. The majority (62%) of these trials are investigating vaccines for HIV (33%) or cancers (29%).The remaining 38% of enrolling or active clinical trials are investigating vaccines for influenza, hepatitis B and C, HPV, and malaria [18], [31]. In 2005, the first DNA vaccine was released for immunization of horses against the West Nile Virus (Ft Dodge Animal Health). In 2012, in the veterinary field, five medical applications (four of them for vaccination purposes) based on pDNA had already been approved for clinical use [6]: two infectious disease vaccines for West Nile virus in horses, infectious haematopoietic necrosis virus in salmon (Novartis), a melanoma cancer vaccine for dogs (Merial) and a growth hormone releasing factor therapy for pigs (VGX Animal Health) [1], [26], [32] (**Table 1**). These licensures were important validations of the DNA vaccine platform because they illustrate its commercial potential [26]. It is generally believed that the advances in DNA vaccine technology will offer customized and tailored vaccines to any disease in the future, representing an important component of next generation prophylactic and therapeutic vaccines, efficient and economically accessible to peoples worldwide [19]. Despite all the advantages and perspectives associated with DNA vaccines based on plasmid vectors, they showed low immunogenicity when tested in human clinical trials, and compared with traditional vaccines, they induce weak immune responses [20], [26], [21]. Therefore, the improvement of vaccine efficacy has become a critical goal in the development of effective DNA vaccination protocols [20], [26], [21].

Table 1 – Current licensed DNA vaccine therapies [26] .

Vaccine target	Product name	Company involved	Date licensed and country	Target organisms	Benefits
West Nile virus	West Nile Innovator	Centers for Disease Control and Prevention and Fort Dodge Laboratories	2005 USA	Horses	Protects against West Nile virus infection
Infectious haematopoietic necrosis virus	Apex-IHN	Novartis	2005 Canada	Salmon	Improves animal welfare, increase food quality and quantity
Growth hormone releasing hormone	LifeTide-SW5	VGX Animal Health	2007 Australia	Swine and food animals*	Increases the number of piglets weaned in breeding sows; significantly decreases perinatal mortality and morbidity
Melanoma	Canine Melanoma Vaccine	Merial, Memorial Sloan–Kettering Cancer Center and The Animal Medical Center of New York	2007 USA, conditional license	Dogs	Treats aggressive forms of cancer of the mouth, nail bed, foot pad or other areas as an alternative to radiation and surgery

Regarding Gene therapy (GT), the first human gene therapy trial was in 1989 [1]. In September 2014, 2076 GT clinical trials have been completed, are ongoing or have been approved worldwide, with 2 already in the phase IV of the drug approval process, and with 1230 in phase I [33] (**Figure 2** A). Regarding its application, GT clinical trials are addressing a large variety of diseases such as cancer (64.1 %), infectious diseases (8.2 %) (like hepatitis and HIV) and cardiovascular diseases (7.8 %) [33] (**Figure 2** B), and the most common gene types transferred were genes coding for antigens (21.1 %), cytokines (17.2 %) and receptors (8.6 %) [33]. DNA naked vectors/ plasmid DNA already represent 17.7 % of the vectors used in gene therapy clinical trials, having just adenovirus and retrovirus ahead of them [33]. The highly efficient transfection machinery of viruses makes them a very important tool for GT, however the advantages of pDNA direct use are allowing the expansion of these vector in GT [6]. For example, they show decreased risk of integration of the foreign DNA into the human chromosomes, when compared with retroviruses. Also, the gene size is not limited to the size of the viral capsid [6].

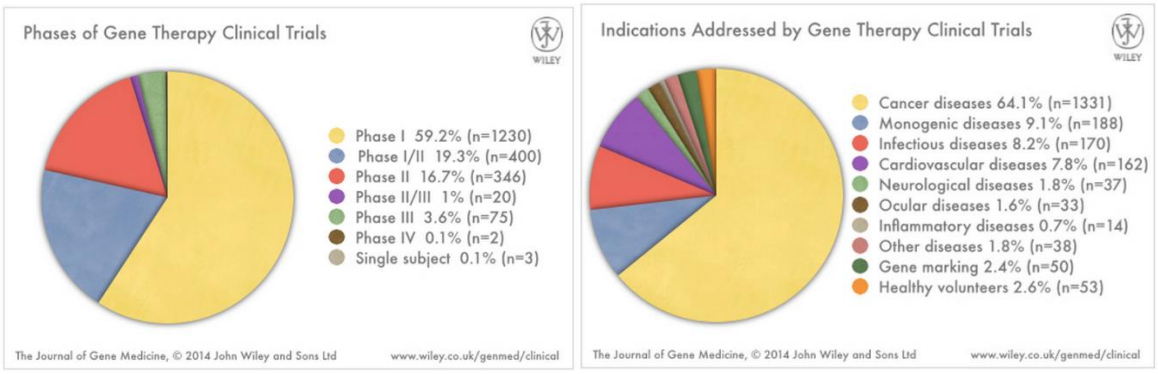

Figure 2- A) Phases of Gene Therapy Clinical Trials in 2014. **B)** Diseases targeted by gene therapy clinical trials in 2014 [33].

At the present 64.5 % of the GT clinical trials are being performed in America (64.5%), and in Europe (25%). Currently no study is being carried in Portugal [33].

1.3 Considerations for the Design of Plasmids for Gene Therapy and Vaccination

Manufacturing of DNA plasmids begins with the design and construction of the plasmid itself [3]. After a target gene has been defined, it is inserted into a plasmid, that will be used as a vector to introduce the genetic information into the patient's cells [6]. The main objectives in plasmid design include high copy number replication in *E. coli* (most frequently), ability to select for plasmid-containing cells, stable maintenance during bacterial fermentation, and genetic stability [3], [6]. Furthermore, the processes of manufacture must be scalable, reproducible and relatively inexpensive [3]. Plasmid size is a critical criterion in vector design, molecules should be as compact as possible [16]. All nonessential sequences should be removed so that the plasmids are as small as possible. In addition to the potential regulatory and therapeutic challenges, larger plasmids also create manufacturing hurdles by placing a metabolic burden on the host strain by increasing required resources for plasmid replication. This, in turn, results in reduced yields [2].

The organization of the genetic elements in therapeutic pDNA reflects its functionality, bulk manufacture and clinical use in the patient [2]. A therapeutic pDNA molecule is typically built around a modular structure [34], [19]. Thus, typical plasmid backbones used for gene therapeutics have several common features that fall into two categories: the plasmid propagation unit (replication region and selection market), responsible for propagation in the microbial host and the Eukaryotic expression unit (promoter, intron, signal sequence, transgene, transcription terminator poly A signal and immune stimulatory sequence) that drives the expression of the transgene in mammalian (eukaryotic) cells [2], [16], [35] (**Figure 3**).

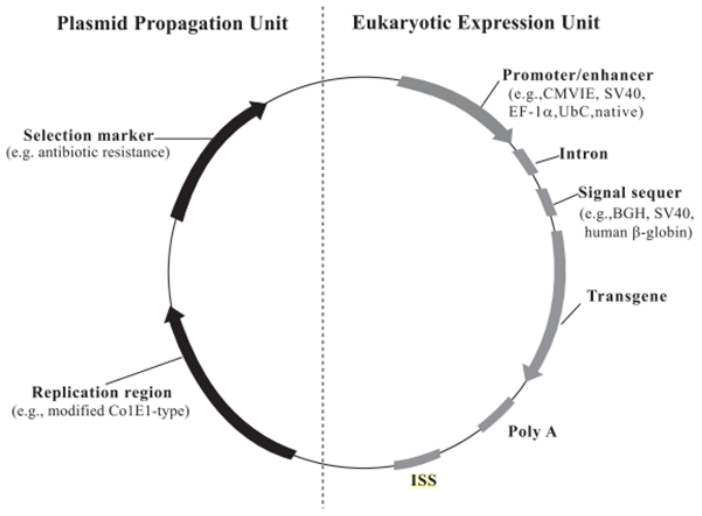

Figure 3 - Typical genetic elements of a pDNA vector. The plasmid consists of a Plasmid Propagation Unit (PPU) that operate in the microbial host and a Eukaryotic Expression Unit (EEU) that drives the protein synthesis in the eukaryotic cells. In the PPU, a replication region and a selection marker are employed, while the EEU comprises an enhancer/promoter region, intron, signal sequence, transgene and a transcriptional terminator (poly A). Immune stimulatory sequences (ISS) add adjuvanticity and may be localized in both units [2].

1.3.1 Plasmid propagation unit

The unit responsible for plasmid propagation in the microbial host contains a bacterial origin of replication and a selectable marker [2], [19].

Bacterial origin of replication

A critical element of all plasmids is the replication origin that ensures replication of the plasmid DNA during growth and division of the bacterial host [3], [8] and also determine the plasmid's host-range [19]. The replication origin determines the frequency of plasmid replication and the average number of plasmid copies per bacterial cell [3], [14]. High copy number is desirable, since it results in higher yields of plasmid per unit mass of bacteria [3], [36]. This region also allows a stable inheritance of the plasmid during bacterial growth, and determines the plasmid's host-range. It is important that the replication regions is narrow host-ranged, to reduce the probability for spread of the plasmid to the patient's own flora [2].

Some of the most commonly used gene therapy plasmids contain replication origins derived from pBR322, like the high copy number pUC derivatives [3], [37], [17]. Origin of replication of ColE1-type plasmid or pBR322 confers a relatively low copy number (50-100 copies per cell) [6]. pUC derivatives are a preferred origin of replication, once the regulatory constraints on the plasmids replication were removed by two single point mutations, enabling over 500 copies per cell [6], [2], [16]. This origin of replication is relatively small (~600 bp), well understood, have high copy number, and require no encoded proteins for replication [3].

However, it is not uncommon for a high copy number status to exacerbate a detrimental effect of certain plasmid elements on the bacterial growth, resulting in selective advantage of spontaneous plasmid deletion mutants, which manifests itself as structural plasmid instability. In addition, random, uncontrolled distribution of plasmids during bacterial cell division produces plasmid less cells, which, given some selective advantage, can outgrow cells with plasmids [36].

A number of alternative plasmid replication origins have also been described, including those derived from naturally occurring plasmids such as R1 [3], [17], R6K, RK2, and F. However, they are not widely used as cloning vectors in *E. coli* and their copy numbers are often relatively low [3], [36]. Thus, these origins have not played a significant role in plasmid-based gene therapy [3].

The addition of another origin of replication (ori) region that is active in mammalian cells allows prolonged persistence and expression of the vaccine gene in the transfected tissue. However, uncontrolled expression of the gene may induce immunological tolerance. Therefore, non-replicating plasmids are considered a factor that reduces risk of germline transmission. Accordingly, only prokaryotic and narrow host range replication regions should be present on therapeutic plasmids [2].

Selectable marker

A second critical element in most plasmids is a selectable marker such as an antibiotic-resistance gene. The marker is used to prevent growth of plasmid-free bacteria during fermentation [38]. This is important for

two reasons. First, many genetically engineered plasmids can be spontaneously lost from the host cell, despite being present at high copy number. Second, high copy number plasmids frequently reduce the growth rate of plasmid containing cells. As a result, plasmid-free cells can quickly outgrow plasmid-containing cells, even if generated at a very low rate [3], [36]. This is a major industrial problem once it leads to reduction of product yields and increased production costs during large-scale cultivation of vector-carrying bacteria [38].

This problem is usually overcome by including an antibiotic resistance gene on the plasmid [3]. The corresponding antibiotic is added to the growth so that only plasmid-containing cells that express the antibiotic resistance gene can grow [12]. This approach is widely used because it is simple and effective [39], but it does present some safety concerns. Regulatory agencies such as the Food and Drug Administration (FDA) have expressed concern that antibiotics added during fermentation may contaminate the final plasmid preparation [3], or lead to the spread of antibiotic resistance genes throughout the host's body [19], [26], also antibiotic resistance markers are associated with structural plasmid instabilities and decreased gene delivery efficiency. Currently, the neomycin/kanamycin resistance gene (neomycin phosphotransferase II, *nptII*), initially isolated from the transposon Tn5, is the most widely used selection marker in bacteria, plants, and mammalian cells because it is the only tolerated by regulatory authorities [34], [6], once kanamycin is not commonly used to treat human infections [26]. These drawbacks have initiated the development of various antibiotic marker-free selection approaches [34], [40], [21]. Some of them are discussed in section 1.5.

The selectable marker should be the only gene that is expressed in *E. coli*, because bacterial growth and plasmid production can be adversely affected by the expression of multiple genes, especially if the gene products are toxic [2].

1.3.2 Eukaryotic expression unit

Eukaryotic expression unit consists of the **gene of interest** and the elements necessary for high-level expression and targeting of the therapeutic component in mammalian cells [2], [41], [6]. The use of appropriate promoters, other regulatory elements, and mammalian maintenance devices ensures that the therapeutic gene or genes are adequately expressed in target human cells [36], [25]. The possibility of encoding multiple proteins in a single construct is an important advantage that DNA vaccines have over other platforms [42], [43].

Promoter

The amount of plasmid that is internalized *in vivo* has been estimated to be in the picogram range after injection into mouse muscle and in the picogram to femtogram range in tissues from 1 to 7 days after intravenous delivery of DNA. Because the plasmid will not replicate in the cells, the amount of plasmid

available for expression is very low. For this reason, a strong mammalian promoter/ terminator should be chosen to drive expression of the antigen gene [2], [20].

Most pDNA harbour promoters and enhancer regions from pathogenic viruses such as cytomegalovirus (CMV), simian virus 40 (SV40), or murine leukaemia virus [2], [44], [20]. For instance, plasmid vaccines with the CMV promoter have been in clinical trials and are versatile due to the promoter's activity in a variety of tissues and animal models [2], [19], [45]. Also, CMV promoters have shown to drive higher constitutive expression levels than alternative viral promoters (SV40) or cellular promoters such as the human ubiquitin C (UbC) or human elongation factor 1α (EF-1α) promoters [1], [44], [41]. Nevertheless, improvement in expression beyond that of current CMV based vectors may improve immunogenicity in humans. This may be gained by the use of modified CMV promoters. For example, incorporating the HTLV-1R-U5 region downstream of the CMV promoter, lead to increased expression and improved cellular immune responses to HIV DNA vaccines in mice and nonhuman primates. Improved expression and/or immunogenicity have also been observed with chimeric SV40-CMV promoters [1], [45]. However, as more than 50 % of the population in USA is infected with CMV and as the virus remains in the body throughout life, the use of its expression signals on vaccine plasmids may induce recombination events and form new chimeras of CMV [2].

Promoters derived from the antigen source may also be used for plasmid-based expression, though this approach would limit the ability to construct a single plasmid backbone for the expression of multiple antigens [44]. Promoters and enhancer regions have also been suggested from housekeeping genes encoding the mouse phosphoenolpyruvate carboxykinase and phosphoglycerate kinase. However, due to the risk of insertional mutagenesis and oncogenesis, highly inter-species-conserved sequences should be avoided [2], [40] . This risk can be reduced by the use of novel synthetic promoters selected by bioinformatic tools to have low homology to sequences potentially present in the recipient [2].

Modification of the Kozak consensus sequence, that is recognized in the cytoplasm by ribosomes and directs efficient transgene translation, can be made, to improve the expression of the transgene product. The optimization of the initiation start site for protein synthesis, can be useful because endogenous sites of viruses and bacteria might not be optimal for expression in mammalian cells [26], [46].

Intron

To augment the promoter activity, introns are introduced, which have a beneficial effect on the *in vivo* expression of the transgene, once optimal expression of some genes depends on the splicing of their transcripts. Most often the intron A from CMV is used [2], [45]. Typically, introns are included in the vector backbone downstream of the promoter, as part of an mRNA leader separating exon 1 and exon 2 [1], [45]. Also, the presence of an intron in the vector backbone downstream of the promoter can enhance the stability of mRNA and increase gene expression [45].

Terminator/polyadenylation

On the other hand, poly A sequence signal (AAUAAA) inclusion next to the 3' extremity of the sequence of interest is essential for efficient transcriptional termination and exportation of the mRNA from the nucleus to the cytoplasm, playing also a very important role in mRNA stability and eukaryotic expression [19], [45], [26]. It has been found that the choice of the transcription terminator/polyA signal can have a dramatic effect on the strength of the promoter [41], [42]. The transcription terminators signals that are commonly used include those from bovine growth hormone (BGH), SV40, and human β-globin [1], [2], [44]. The combination of the cytomegalovirus (CMV) promoter and bovine growth hormone (BGH) terminator provides a high level of transcription [2], [45].

Immune stimulatory sequence

To enhance the potency of a DNA vaccine, immune stimulatory sequences (ISS's) are added to the plasmid. These sequences of genes encode biologically active molecules such as cytokines, chemokines and co-stimulatory molecules, which are able to strengthen the immune response to a vaccine, interacting with toll-like receptors and add adjuvanticity [2], [20], [21]. CpG motifs are 10– 20 fold underrepresented in mammalian genomic DNA and, when present, are often methylated. Immunostimulatory sequences such as unmethylated CpG can be utilized to enhance T lymphocyte recruitment or expansion (innate immune responses) [41], [21], [21] which might be beneficial in some forms of cancer gene therapy [36]. Increased immunostimulatory CpG content can be achieved by enhanced CpG content throughout the plasmid or by adding a block of CpG to a non-essential region [1], [21], [29]. However, unmethylated CpG from bacterial DNA or plasmids may contribute both to the potency of plasmid based DNA vaccines and to the limited duration of expression [1], [44], [41]. Alternatively, if the immune response needs to be minimized, CpG-depleted or CpG-ablated plasmid backbones can be used to decrease the CpG-motif load bacterial sequences can be removed from the plasmid DNA to produce minimized DNA vector [36], [11].

For some applications, such as gene replacement therapy, immune stimulation and short term expression is undesirable. In such situations, promoters and protein selectable marker genes can be modified to reduce or eliminate CpG motifs. For applications where any unmethylated CpG is undesirable, an alternative origin or modified host strain overproducing CpG methylase should be considered [1], [47].

Signal sequence

Finally, the signal sequence is also part of the eukaryotic expression unit. Adaptive immune responses can be enhanced by improving antigen processing for MHCI and or MHCII presentation [1], [20]. This can be accomplished by targeting heterologous proteins to various intracellular destinations [1], [20]. For secretion of the synthesized peptide to the extra-cellular milieu, a signal sequence is positioned in front of the vaccine gene. This codes for a signal peptide of about 20-40 amino acids, often derived from bovine proteins such as the plasminogen activator [2], [1]. As well, membrane-anchoring using human alkaline phosphatase (PLAP), endosomal targeting using human Lamp1, proteosomal targeting using murine Ubiquitin A76, or endoplasmic reticulum targeting may alter or enhance immune responses [1]. To avoid undesired immune

responses, the nature of the signal peptide should be considered. Statistical methods like the hidden Markov model have been used to predict and generate artificial signal peptide sequences for use in human cells. Such a strategy could be applied to DNA vaccine development in order to create more appropriate signal peptides [2].

1.4 Structural characteristics of plasmids

The structural characteristics of plasmids are important from a therapeutic point of view, since stability and efficacy depend in part of the topology of the plasmid [6].

pDNA may appear in three topological isoforms: linear, open circular, and supercoiled (SC), being the SC the main isoform. Their relative distribution fluctuate with the replication state of the molecule, the bacterial culture growth phase at the time of isolation, and the shear stress resulting from the DNA purification process [16]. Also, certain growth conditions can lead to significant amounts of relaxed, nicked, or even linear molecules. The SC form of pDNA is the physiologically active conformation and is the optimal form for the transfection of mammalian cells [25], [41]. In addition, SC-form has the advantage of displaying a reduced size [6]. Therefore, SC-form should be enriched during manufacturing and downstream processing [12], since a content of at least 80% of supercoiled pDNA in the final product is recommended [6], [12]. It is believed that the supercoiled form is better protected against enzymatic degradation and thus more plasmid can reach the nucleus of the target cell [6].

In adittion, pDNA is the negatively supercoiled B-form conformation. Within a generally B-like DNA molecule, specific regions may exist in a non-B form, due to special sequence compositions. Symmetry elements, like inverted repeats, mirror repeats, and direct repeats, also interfere with DNA conformation and can form cruciforms, intramolecular triplexes, and slipped mispaired structures [12], [6], [46]. Computer programs can be utilized to identify potential cruciform cleavage hot spots [1].

Plasmids can co-integrate, forming head-to-tail multimers that are two, three, or more times larger than the initial monomer [3]. There are also multimeric species of the whole monomeric plasmid molecule due to recombination and replication processes, termed concatamers, that are detrimental for currently established processes. Their formation is minimized in most bacterial strains used for large scale production of plasmid DNA, deficient in recombination genes (*recBCD*$^{-}$*, recA*$^{-}$) [12], [6].

1.4.1 pDNA copy number and stability

The term stability refers to the stable propagation of plasmids in prokaryotic cells and the integrity of the DNA sequence during cultivation. Considering that high copy number plasmids are unstable in *recA*$^{+}$ host strains due to multimerization, a *recA*$^{-}$ genomic background of the host cell should be preferred [12].

Plasmid copy number (PCN) is influenced by several factors like the efficiency of the replication origin and the percent of initiated replication cycles that are completed. Plasmid copy number will be reduced by sequences that cause inhibition of DNA polymerase III replication, stable RNA-DNA hybrids, unusual DNA structures and strong convergent promoters leading to head on transcription-replication collisions. Optimal vectors should have a higher copy number not just because of the promoter but also due to reduction of these factors [1].

1.4.2 Size

One of the first considerations in designing an efficient process should be the plasmid size [2], [41], once it was verified that pDNA size modulates gene transfer efficiency, and affect purification [6], [12]. The transfection activity of lipoplexes containing smaller plasmids was found to be greater than that of lipoplexes containing the same molarity of larger plasmids. DNA diffusion in cytoplasm of cells is strongly size dependent, with little or no diffusion for DNA molecules bigger than 2,000 bp. It is believed that binding effects to cytoskeletal elements, acting as molecular sieve, are not primarily responsible for the slowed diffusion of DNA molecules. The major problem seems to be molecular crowding and collisional interactions exerting a strong dependence on intracellular diffusion. Nuclear translocation is more frequently achieved during cell division, due to nuclear envelope disruption allowing DNA entry into the nucleus. When entry of pDNA in nuclei occurs via nuclear pores, it seems to be a size dependent process, as well. It was shown that the transfection of non-dividing smooth muscle cells by a minicircle, 2,900 bp in size, was 77 times more efficient than a plasmid 52,500 bp in length, supporting this hypothesis [12].

Thus, all nonessential sequences should be removed, so that the plasmids are as small as possible. In addition to the potential regulatory and therapeutic challenges, many larger plasmids also create manufacturing hurdles by placing a metabolic burden on the host cell line, which in turn, results in reduced yields [2].

1.4.3 Stability during Intracellular Routing

Efficient gene expression can only be achieved if the majority of plasmid molecules reach the nucleus intact and preferentially in the SC isoform [48], [40].

Once internalized, pDNA must overcome some major hurdles such as endolysosomal entrapment [48], [12], cytosolic sequestration, and nuclear exclusion of the DNA [12]. In addition to these physical barriers, the DNA is subjected to metabolic degradation [12], [48], contributing to poor efficiencies. Besides the use of different transfection methods and adjuvants shielding the DNA, the vector itself can be designed to maximize survival in hostile, nuclease-rich environments like the cytoplasm [48]. This implies that ‘hot- spot’ regions, which are particularly prone to nuclease digestion, should be removed or modified [48]. It was shown that pDNA molecules, possessing sequences that originate secondary structures with single-

stranded regions, are more susceptible to exo/endonuclease attacks. Two major hotspots were identified by nuclease S1 mapping of a commercial vector designed for the use in the development of DNA vaccines. Results showed that the plasmid was cleaved at spots located within the Bovine Growth Hormone (BGH) polyA signal and within the ColE1*ori* [12], [48], [41] due to six homopurine-rich sequences within the BGH polyA and the presence of an 11-base long inverted repeat within the prokaryotic ori [12], [41].

1.4.4 Safety and Potency

Regarding DNA vaccination, one major concern addressed by regulatory authorities is the integration of nucleic acid vectors in the host cell DNA [48] and consequent insertional mutagenesis [12], [25]. So far, very limited data demonstrating chromosomal integration have been reported [12]. In fact, only a single study involving the use of a PCR-based assay was able to detect rare integration events allowed the detection of independent genome integration events. The majority of recent studies that involved a number of different DNA vaccines and distinct organisms have found that integration occurs at a very low frequency, if at all [48]. Still, sequences homologous to human DNA should be avoided, or at least minimized, to reduce the possibility of recombination [12]. *De novo* DNA synthesis is also a promising approach to minimize any homology between pDNA and genomic DNA because the DNA sequence of coding regions can be altered without affecting the amino acid sequence of the translated protein [48]. Regarding the downstream purification process, it must remove impurities such as protein, RNA, chromosomal DNA, and endotoxins to acceptable levels [46].

In addition, to reduce the immunostimulatory response from unmethylated CpG-motifs recognition, different strategies have been developed to decrease the immunostimulatory properties of pDNA. The most straightforward approach consists in reducing the frequency of these inflammatory motifs by eliminating non-essential regions within the plasmid backbone. Bacterial minimal pDNA vectors will provide a reduced overall immunostimulatory profile and will, therefore, contribute to safer gene medicinal products [12]. Also, antibiotic resistance genes raise some safety concerns, discussed below [30].

Regarding DNA vaccines potency, an immune stimulus or adjuvant is usually needed to maximize the immunogenicity of the vaccine [25], [49]. Adjuvants generally interact with receptors (e.g., TLRs) present on immune cells [21], [25], [42]. This interaction results in rapid activation of cells leading to cytokine production and an increase in antigen presentation capacity, leading to enhancement of antigen-specific immune response [49], [25]. CpG motifs play an immunostimulatory role. It is believed that the correct number and placement of CpG within the plasmid are important [50], [25], [51]. Also the vaccine delivery system can protect the vaccine from degradation, facilitate cellular uptake, target specific cells or tissues, and ensure that the antigen and adjuvant are delivered together [25]. Potency can be increased, as well, by combining DNA vaccine with other vaccine modalities in prime/boost type approaches [21], [20]. First-generation DNA vaccines showed weak potency in humans when compared to mice, which is leading to several efforts in developing new ways to improve the potency of the technique by developing novel adjuvants, codon

optimization schemes and innovative methods of delivery that will make them more suitable and immunogenic for human vaccination [42].

1.5 Improving pDNA performance by avoiding antibiotic resistance genes

Successful non-viral gene delivery depends strongly on the design of the expression vector [12]. Regarding safety and potency of vectors the reduction of backbone sequences, e.g. minimizing or deleting selection markers, seems to be beneficial for overall performance of the delivery vehicle [12]. Antibiotics and antibiotic resistance genes have traditionally been used for the selection and maintenance of recombinant plasmids in hosts such as *E. coli*, but their continued use is undesirable in many areas of biotechnology, such as gene therapy protocols, or DNA vaccines that involve the direct introduction of plasmid DNA into patients [52], [53]. Regarding this concern, the European Pharmacopoeia 7.0 states, “Unless otherwise justified and authorised, antibiotic-resistance genes used as selectable genetic markers, particularly for clinically useful antibiotics, are not included in the vector construct. Other selection techniques for the recombinant plasmid are preferred”. As well, World Health Organization (WHO), the European Agency for the Evaluation of Medicinal Products (EMA) and the FDA included similar considerations in their guidelines [54], [55].

Antibiotic resistance genes and selectable markers in general are responsible for the main drawbacks in pDNA performance [12], as they cause safety problems due to possible integration of plasmid DNA into cellular DNA or spread of antibiotic resistance genes to intestinal bacteria by horizontal gene transfer [4], [46], [56]. Also, resistance genes may cause inflammatory reactions *in vivo*, hamper transfection efficiency by increased pDNA size and disturbing sequences [54], [12], and cause an additional metabolic load for the host during manufacturing [12], [52]. This metabolic burden caused by pDNA replication and consumption of precursor metabolites and energy [5] is manifested by inhibition of biomass growth, structural instabilities of cultivated pDNA and segregational instability, impairing the overall process performance [4], [54]. The major cause for metabolic burden exerted on the host cell, is the constitutive expression of the antibiotic resistance gene. Usually, the marker protein represents up to 20% of total cellular protein, thereby exceeding the levels that are needed for proper selection and maintenance [4], [54], [57]. Also, the amplification of plasmids harbouring an antibiotic selection marker under conditions mimicking those used at the industrial scale was found to increase the levels of the alarmone ppGpp and the production of inclusion bodies in an *E. coli* host, when compared with marker-free variants [34]. Due to the fact that these sequences represent a safety problem, are non-essential for the later product characteristics, and reduce the potency of the final pDNA-based product, it is desirable to exclude these sequences from the vector backbone [4], [5].

Another concern about the use of antibiotic resistance genes for selection is that bacterial cultures require a large amount of antibiotics, which are expensive compounds. It is difficult to avoid their degradation or inactivation, and antibiotics may not be completely effective, in particular in continuous culture conditions

limiting the scale-up. Moreover, residues of antibiotics could contaminate the final product even after purification [54].

The use of the very common Ampicillin resistance gene is not acceptable for therapeutic vectors due to hyper-reactivity of some patients to β-lactam antibiotics [56]. Allergic reactions may occur in up to 10% and anaphylaxis in approximately 0.01 % of patients respectively. The tetracycline efflux protein (TetA), another widespread resistance marker, is toxic to *E. coli* when incorporated in high copy vectors or when the culture is reaching the stationary phase [12]. Regarding DNA therapeutical applications, the kanamycin resistance is the only resistance gene tolerated by regulatory authorities, and encode for a aminoglycoside phosphotransferase (*nptII*) [4], [54].

Therefore, strategies have been developed to avoid or at least decrease the presence of bacterially active elements and to use alternative selection strategies [12], to further improve their safety profile [4], [12], [58].

1.5.1 Complementation of auxotrophic bacterial strains

To produce antibiotic-free plasmid by using this strategy, initially the bacterial strain is modified by introducing a deletion or a non- sense point mutation into an essential or conditionally essential chromosomal gene resulting in auxotrophy (**Figure 4**) [54], [59], [58]. Bacterial growth is restored upon introduction into those strains of a plasmid either carrying the deleted gene or coding a suppressor tRNA which allows a complete translation of the truncated protein. Several auxotrophic complementation approaches were performed: glycine [58], Nicotinamide adenine dinucleotide, Translation initiation factor 1, Arginine and Thymidine [54], [59].

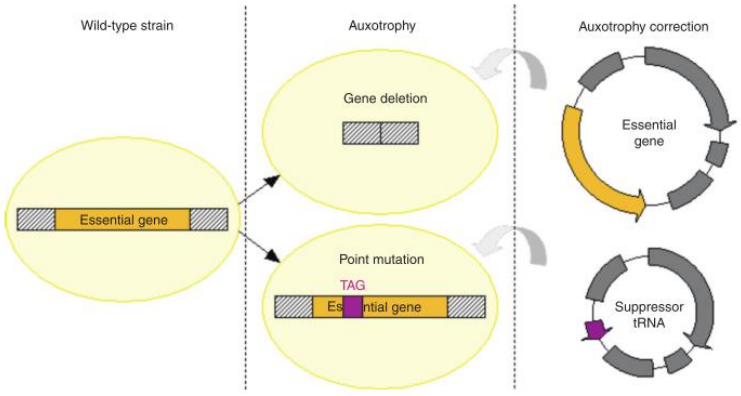

Figure 4- Auxotrophy of bacterial strain and prototrophic growth restoration by plasmid DNA, in case of gene deletion and non-sense point mutation [54].

1.5.2 Minicircles

Minicircles are double-stranded supercoiled expression cassettes generated by site-specific recombination [12], [34], providing a final product that is devoid of the bacterial pDNA backbone [34], [60], [15] (Figure 5

A). They are produced, in *E. coli,* from a parental plasmid, which contains a eukaryotic expression cassette flanked by either restriction enzyme- or recombinase- recognized specific sequences, respectively. After plasmid propagation, the expression cassette is released by either enzymatic digestion or intramolecular recombination and further purified [61]. The selection of these plasmids is still dependent on the use of antibiotics, but the *ori* and the antibiotic resistance gene are excised during manufacturing by a recombination step, yielding the so-called miniplasmid [12]. This miniplasmid has to be purified by additional chromatography steps, representing a major cost factor [12], [16], [54]. Due to the significant size reduction compared to conventional plasmid vectors, these DNA molecules seem to have higher transfection efficiencies and expression levels due to better bioavailability and permeability [9], [12], [15]. The major drawback of this system will be its costs compared to conventional fermentation processes due to their complexity [12].

1.5.3 Bacterial backbone-reduced plasmids

The pCOR (plasmids with conditional origin of replication) and the pFAR (plasmids free of antibiotic resistance) [34], [61] selection systems rely on the suppression of an amber nonsense mutation (UAG chain-terminating codons) introduced into an essential chromosomal bacterial gene (Figure 5 B) [62], [46]. For that purpose, the plasmids contain an amber suppressor tRNA (~100 bp in size) that inserts a certain amino acid instead of terminating translation [61]. The host strain for propagation of pCOR or pFAR has been modified to carry an amber stop codon in an essential gene that can only be complemented by the plasmid carrying the corresponding suppressor tRNA. pCOR was designed to overcome an arginine auxotrophy, caused by introduction of an amber mutation into the acetylornithine deacetylase gene (*argE*), by replacing the antibiotic resistance gene for a synthetic amber suppressor tRNA [34], [40]. pCOR plasmids have been produced with reasonable yields and several preclinical and clinical studies have been already performed. In contrast to pCOR, plasmids of the pFAR series harbour a high copy number pUC origin of replication. Here the amber mutation was introduced into the *thyA* gene, necessary for DNA synthesis and encoding for thymidylate synthase, consequently leading to a thymidine auxotrophic strain. Again, the complementation suppression is achieved by introduction of a suppressor tRNA on the plasmid backbone. *In vivo* and *in vitro* data for a pFAR plasmid encoding the luciferase reporter gene is available for mouse muscle, skin, and tumor cells [34].

1.5.4 Operator–Repressor Titration

Operator–Repressor titration (ORT) vector systems are designed in such a way that an operator sequence like *lacO* or *tetO*, placed on a multicopy plasmid, de-represses a chromosomal essential gene, controlled by a negatively regulated promoter (Figure 5 C) [12], [53], [59]. Expression of this gene is only possible upon plasmid attendance because the molar excess of the operator incorporated on the multicopy plasmid titrates the repressor from the chromosomal operator, allowing transcription of the essential gene [12], [62].

Cranenburgh *et al*. [52] developed an *E. coli* strain that allows the selection of plasmids by ORT in complex media [52], [34], [58]. This strain contains an ectopic copy of the *dapD* gene, encoding the tetrahydrodipicolinate *N*-succinyltransferase, involved in cell wall biosynthesis, whose expression is controlled by a *lac* operator [52], [34]. In the absence of *dapD*, mutants lyse [59], [54]. The pORT vectors were designed for use in repressor titration strains [52], and minimal setup for the pORT is 1,281 bp, containing a ColE1*ori*, a multiple cloning site (MCS) and two operator sequences. The advantage of such systems is small size and elimination of antibiotic resistance [40]. Due to this interference the reduction of the amount of prokaryotic sequence is of great importance [12]. pORT backbones have been used in some preclinical and clinical HIV- 1 DNA vaccine studies, as well as for plasmid stabilization in *Salmonella enterica* for the generation of live bacterial vaccines [34].

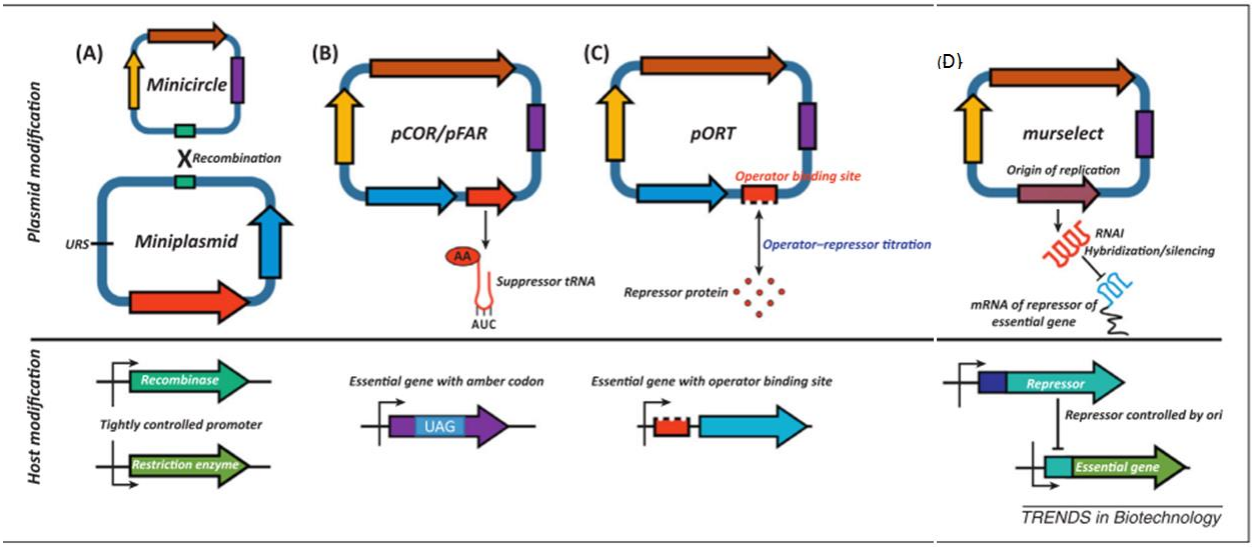

Figure 5 - Overview of available antibiotic-resistance-free plasmid selection systems showing necessary plasmid and *Escherichia coli* host modifications. **(A)** Minicircle technology: minicircle and miniplasmid are resolved from a parental plasmid by recombination. **(B)** Plasmid with conditional origin of replication (pCOR)/plasmid free of antibiotic resistance (pFAR): these plasmids encode a suppressor tRNA that can overcome an amber codon (TAG) that was introduced in an essential gene on the E. coli chromosome. **(C)** Plasmid with operator–repressor titration (pORT): multiple operator binding sites are integrated into the multicopy plasmid backbone, leading to a large excess of the plasmid-encoded operator compared to the chromosomal-encoded operator binding site that controls an essential gene. Titration of the repressor by the plasmid leads to expression of the essential gene. **(D)** murselect system: the plasmid only has to encode a ColE1-type origin of replication. Selection and maintenance is achieved by functionally linking the RNAI encoded on the origin of replication of the plasmid, to a repressor protein encoded on the host genome that further controls the expression of an essential gene. (Key: Red arrow, element conferring selection; blue arrow, origin of replication; yellow arrow, eukaryotic promoter; brown arrow, therapeutic gene; purple box, polyadenylation signal; green box, recombinase attachment site). (Adapted from [34])

1.5.5 Plasmid-derived Gene Silencing

This selection system is based on RNA/RNA interaction [34], [62]. Similarly to the operator–repressor titration strategy, an essential gene of *E. coli* was placed under the control of an operator–repressor system but, in that case, the repressor was fused to an RNAII sequence [54]. The replication of ColE1-type plasmids (e.g. pUC19) [34] is based on two RNAs that are encoded on the plasmids ori, namely RNA I and RNA II [6], [14], [63] (Figure 5 D). The key molecule, responsible for the regulation of PCN, is RNA I, which acts as a very efficient antisense molecule, inhibiting the replication when a certain threshold PCN is reached. It was shown that RNA I is capable of silencing chromosomal genes, fused to an RNA II antisense-target [12], [14]. Based on this finding, an artificial genetic circuit was engineered, wherein the plasmid replication machinery contributes to the key element for cell survival, and thereby enables antibiotic-free plasmid selection [34], [54]. A two-step regulatory system was designed, wherein an inducible repressor protein, namely TetR, subsequently represses the translation of an essential gene. The genome of *E. coli* was changed in a way that transcription of UDP-N-actetylglucosamine enolpyruvyl- transferase (*murA*), that encodes an enzyme involved in cell wall biosynthesis (catalyzing the first committed step in the assembly of peptidoglycan), is inhibited in the presence of the tetracycline repressor (TetR) [34]. The repressor gene is engineered in such a way that RNA I can bind to its mRNA and thus inhibit its translation. Consequently, cell survival requires plasmid replication within the cell, since the plasmid's replication machinery provides an essential advantage by suppressing the translation of the detrimental repressor protein TetR by RNA I. Due to the deduction of RNA I, the inhibitor of plasmid replication, the PCN gets increased and surpassing yields are achieved [12] (Figure 6). This approach allows the selection of plasmids without the addition of other extrinsic elements to the plasmid backbone [34], [54], the only sequence required for selection is the ColE1ori, included in common pUC-vectors; therefore, the size of replication-competent plasmids is reduced to a minimum [12]. This system is limited to existing ColE1 type origin containing plasmid vectors only [40].

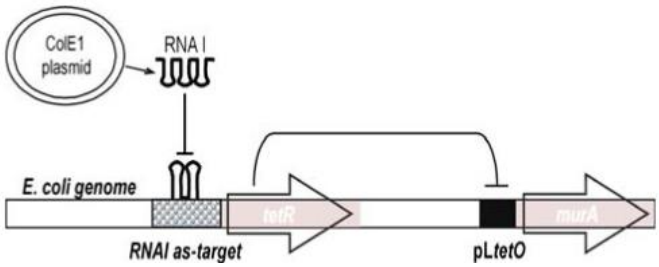

Figure 6- Plasmid-derived gene silencing. A repressor (TetR) controls translation of an essential gene (*murA*) by binding to an operator (pLtetO), thereby inhibiting cell growth. RNA I, derived from the ColE1ori, inhibits translation of the repressor by antisense hybridization to its mRNA. Since this results in expression of *murA*, plasmid-carrying cells are able to survive [12].

Some other alternatives to existing selection technologies are emerging. In 2012, El-Attar *et al.* developed a system applying the widely used household biocide triclosan as the selective agent and an endogenous

growth essential target gene, *fabI*, as the plasmid- borne marker in *Escherichia coli* [53], [54]. An antidote/poison selection technique was also developed where a F-plasmid ccd antidote-poison operon was modified for plasmid selection [54], [55], [59]. The antidote protein gene (ccdA) was incorporated onto the plasmid, and the ccdB poison gene was expressed from the chromosome; this scheme allowed plasmid stabilization [40], [54] . Similarly, to technique “Plasmid-derived Gene Silencing”, Luke *et al.* developed a system based on RNA/RNA interaction where RNAII represses expression of a chromosomally integrated constitutively expressed counter-selectable marker (*sacB*), allowing plasmid selection on sucrose [40], [54], [57]. The MIDGE vector is a linear covalently closed dumbbell-shaped molecule, which is generated by digesting the parent plasmid vector in vitro then, ligating a hairpin cap to its open ends, excluding the prokaryotic backbone and antibiotic-resistant genes [60], [64]. These vectors showed greatly enhanced immuno-compatibility, reduce the risk of spreading antibiotic-resistant genes into body microbial flora, and have a transcriptionally active structure [60]. There are also, a similar vector to MIDGE, which is a mini linear covalently closed DNA vector, as well, but is generated *in vivo* [64]. Some preclinical and clinical studies used to test systems without antibiotic are represented in Table 2.

Table 2- Preclinical and clinical studies using non-antibiotic plasmids [54].

Encoded gene	Preclinical results	Clinical results
pCOR vector		
Luciferase	Higher expression after intramuscular injection into muscle compared to pGL3 in mice	—
Interleukin-10	Beneficial effect on collagen-induced arthritis after muscle electrotransfer in mice	—
3 variants of human tumor necrosis factor-a soluble receptor I	Sustained therapeutic effect on collagen-induced arthritis after muscle electrotransfer in mice	—
Fibroblast growth factor 1	Pronounced therapeutic effect on collateral vessel formation to the ischemic hindlimb after intramuscular injection in rabbit Formation of collateral vessels and arterioles in ischemic muscles of hypercholesterolemic hamsters after intramuscular injection	Improved perfusion after intramuscular injection Improved amputation-free survival in patients with critical limb Ischemia after intramuscular injection Phase 1, 2 study
Murine neurotrophin 3	Partial prevention of cisplatin-induced neuropathy after muscle electrotransfer in mice	—
Murine erythropoietin	Efficient, long-lasting, and nontoxic treatment of β-thalassemic mouse anaemia	—
pFAR vector		
Luciferase	Equal or higher expression after electrotransfer into muscle, skin and tumor compared to pVAX2 in mice	—
pORT vector		
Part of HIV-1 gag protein fused to CTL epitopes	Highly immunogenic in mice after intramuscular or intradermal injection Cellular immune responses specific for multiple HIV-derived epitopes in rhesus macaques	Induction of HIV-1 specific response for most of the patients
Improved rearranged human papillomavirus type 16 E7	Strong cellular and humoral immune response including tumor protection and regression in mice	—
Leishmania infantum LACK antigen	Protective and immunostimulatory effect of DNA prime-poxvirus boost in a canine experimental model	—
Minicircles		
Human factor IX and α1-antitrypsin	Higher expression after mouse liver transfection compared to standard plasmid DNA	—
IFNγ	Higher expression than conventional plasmid Antiproliferative and antitumoral effect	—
Vascular endothelial growth factor	Higher VEGF expression after muscle injection in mice Similar expression in mouse heart	—
Manganese superoxide dismutase	Protection of mice from irradiation	—
Hypoxia-inducible factor-1 α	Improvement of transfection efficiency, duration of transgene expression and cardiac contractility	—

These systems represent progress in the area, but none have been widely adopted, possibly due to the need for mutant host strains [53], the fact that they are relatively recent and that the kanamycin selection marker remains allowed for gene therapy clinical trials [54].

In short, an ideal selection marker should have all the following properties:

- Reduced horizontal gene transfer, and if occur it t should not impact the recipient cell.
- The selection marker should have a limited impact on the plasmid size and should not induce immune activation.
- The marker cannot be toxic for eukaryotic cells.

- Plasmid maintenance should not require the presence of potentially harmful compounds.
- Plasmid yields should be high even in large culture, and the culture media required should be easily available and inexpensive to allow scaling-up [54].

Most of the presented strategies show the required characteristics, and data provided by the preclinical and clinical studies will allow the access to more information about efficiency and safety of the vectors [54].

1.6 Host strain for plasmid production

After the design and construction of a DNA plasmid vector for vaccination or gene therapy, it must be established in a suitable production host strain [3], [65]. Desirable phenotypes include the ability to ferment to high cell density and high plasmid copy number, minimal generation of plasmid-free cells, minimal potential for genetic alterations to plasmid, and compatibility with subsequent purification procedures [3], [66] . The characteristics of the microbial host affect the quality of the purified DNA [2], [66]. Host encoded replication protein expression level variations between strains likely accounts for observable differences in plasmid properties such as percent open circular plasmid, steady state supercoiling density, catenation, multimerization and yield [46]. A number of safety concerns have been advanced concerning the microbial host. These include production of toxins and biogenic amines [2], [67] transferable antibiotic resistances and genetic instability [2], [66], including prophage induced promiscuity and rearrangement of pDNA [2].

Considering that vaccines require milligram dosages to elicit proper immune responses, the volumetric and specific yields of DNA must be maximized during production [68]. In order to achieve high DNA yields the gram-negative bacterium *E. coli* is the most commonly used host for the propagation of plasmid DNA [69], [67], [70] because it is very robust, capable of fast growth with minimal nutritional requirements and achieve high pDNA yields [37], [71], [69]. The most widely utilized high-density strains of *E. coli,* for DNA production [68], are DH5α, DH5, DH1, JM108 , SCS1-L and DH10B [72]. BL21 (an *E. coli* B strain) has recently been demonstrated to be a high yielding plasmid production host [1], [73], [37], exhibiting a five-fold increase in plasmid DNA yield when compared to strain DH5 α, in a study described by Yau *et al.* [74].

Some of the advantages of using *E. coli* cells for plasmid production, are associated with the fact that the genome is fully sequenced [71], [75] that they are easily modified by many genetic methods and display fast growth with high cell densities and minimum nutrient requirements [74], [76]. In addition, *E. coli* are classified as biologically safe vehicles for the propagation of many efficient gene cloning and expression vectors in all major national and international guidelines on biological safety for work with recombinant DNA technology [2]. On the other hand, *E. coli* has some disadvantages like highly immunogenic endotoxin production [71], lipopolysaccharides (LPS) , in its outer membrane [2], [67] and genetic instability, resulting in safety concerns surrounding its use [71]. For this reason, there has been some work investigating other microorganisms, such as the gram-positive, food-grade organism *Lactococcus lactis*, as hosts for

biopharmaceutical-grade pDNA [77]. However, taking into consideration the benefits and drawbacks, *E. coli* is currently the most suitable organism for pDNA production at industrial scale [71].

The choice of host strain may also affect downstream purification, e.g., carbohydrate producer should be avoided or strains that produce a large amount of capsule which can affect harvest and lysis characteristics [2].

The first *E. coli* K-12 strain was isolated in 1922, and since that time thousands of mutant strains have been produced, to facilitate cloning of heterologous genes and for the purpose of stably maintaining plasmid DNA for the production of recombinant proteins. However, it is still not known whether these mutations are beneficial for plasmid DNA production or not, since different strategies are used for protein expression [35], [66], [71]. In 2007, the top three patented high-yield pDNA fermentation processes used DH5, DH5α and JM108 as the main host strain [71].

Most of the strains used for pDNA production are selected by its previous use in laboratory-scale protocols and may be not suitable for process-like conditions [35]. For example, the typical challenges for high cell-density cultivations (HCDC) of *E. coli* remain as obstacles for the fast and efficient production of pDNA. Among them, aerobic acetate production is an important drawback, since it causes a loss of productivity and waste of carbon source [72], [78]. Also, to improve safety, *E. coli* strains that allow antibiotic-free plasmid selection have been developed [71].

In order to improve pDNA production, several new strains have been created by rationally mutating genes selected on the basis of *Escherichia coli* central metabolism and plasmid properties. However, the highly mutagenized genetic background of the strains used makes it difficult to ascertain the exact impact of those mutations [79].

The comparison of some of the most commonly used *E. coli* strains for plasmid production, with respect to higher plasmid yields and homogeneity of supercoiled plasmid (>90%), was performed in Patent WO2005098002 [80]. The seven strains compared were: DH1, DH5, DH5α, JM83, JM101, HB101, and JM108; in LB shake-flask media, from which JM108 was selected as the best, providing the highest amount of plasmid and 91% supercoiled isoform [1], [37]. Poorer performers included for example DH5α and DH1 [37].

1.7 Effects of plasmid DNA production on *E. coli*

Plasmid DNA maintenance and replication can induce a “metabolic burden” in *E. coli* [74], [81], by altering the cell regulatory status through a complex interaction between host and vector [74], [82]. The response of cells under energy limiting conditions includes the activation of alternative pathways for energy generation, resulting in cell growth alterations. Because of the detrimental effect of this plasmid-imposed metabolic burden, cells that lose plasmid are able to grow faster during fermentation. Over time, these non-productive

cells become a significant fraction of the population. This contributes to problems such as plasmid loss and reduced productivity, which are of relevance to the industry [83]. Overall, this plasmid-induced metabolic burden is reflected by alterations in host cell physiology and metabolism and can also elicit several stress responses which, in turn, can result in lower plasmid DNA yields due to the deleterious effects of these alterations on cell growth and plasmid synthesis [74].

Several research studies have shown that plasmid-bearing cells (P^+ cells) exhibited lower specific growth rates than plasmid-free cells (P^- cells), resulting in lower biomass yields at the end of fermentation. The degree of decrease in specific growth rate is dependent on plasmid type, size and copy number. This growth retardation in P^+ cells can be attributed to the competition between the two main biological activities that take place in recombinant cells, i.e., cell propagation and plasmid replication, for the use of intracellular resources such as amino acids, nucleotides and metabolic energy. Furthermore, these growth-arrested cells can experience protein deterioration and misfolding and difficult carbon and nitrogen assimilation due to the activation of stress responses [74].

One of the frequent observations during plasmid DNA production in *E. coli* is cell filamentation. Cell filamentation is seen when cells elongate and replicate their DNA, but are not able to septate and divide. This filamentation causes a decrease in growth rate or, eventually, no further cell division, leading to low biomass and pDNA productivity. Ow *et al*. [83] described that some genes coding for proteins involved in cell wall synthesis and septal division are downregulated in plasmid-bearing cells and there is evidence that the amplification of an essential cell division protein, FtsZ, could suppress cell filamentation in recombinant *E. coli* [83]. This uncoupling between chromosomal DNA replication and cell division can lead to cell cycle heterogeneity and give rise to cells containing more than one or two chromosomes. Metabolic burden can also reduce cellular viability as a result of the increased stress suffered by these cells [74].

1.7.1 Metabolic consequences of plasmid maintenance and replication

Metabolic burden, previously described, can occur due to the extra biosynthetic demands for plasmid synthesis, or the perturbation of the *E. coli* host regulatory system affecting central metabolic pathways [83]. Several studies correlate low growth imposed by plasmids with metabolic burden, since plasmid replication and expression of the antibiotic resistance marker requires additional nutrients and energy [74]**,** [84]. Many studies have been done in order to understand this phenomenon [74], however, the relationship between plasmid DNA content and growth rate has not been fully explained [71].

With the recent arrival of high-throughput analytical tools, the accumulation of vast amounts of transcriptome, proteome, metabolome and fluxome data has enable researchers to gain a deeper knowledge on plasmid-induced metabolic burden. Which allowed the simulation of *E. coli* metabolic behavior under specific conditions as well as to identify possible targets for a more systematic design and implementation of metabolic engineering towards the improvement of pDNA yields [74].

Recent reports have shown the effects of plasmid DNA on the central metabolism of *E. coli*, namely glycolysis, the tricarboxylic acid cycle (TCA), and the pentose phosphate pathway [71].

The **Glycolysis** pathway is the main catabolic route of carbohydrates for the provision of energy and building precursors for biosynthesis [71]**,** [83]. Glucose and other sugars are fed into glycolysis to produce pyruvate from the common intermediate fructose diphosphate for the TCA cycle [83]. The differences in the expression of these genes and proteins might be related with the different fermentation strategies and growth media used in these works [74].

The **pentose phosphate pathway** (PP) is the second major route for carbohydrate metabolism after glycolysis [83], and can metabolize different sugars like xylose and ribose, but is considered the second main destination for glucose. The PP pathway is also one of the pathways responsible for biosynthesis of the nucleotide-precursors ribose-5-phosphate (R5P) [68], [71]**,** [79] and erythrose-4-phosphate (E4P) [68], [71]. Another important product from the PP pathway is NADPH, synthesized by glucose 6-phosphate-1-dehydrogenase (Zwf) [68], [71]**,** [79] and 6-phosphogluconate dehydrogenase (Gnd). NADPH and nucleotides are required for biomass and plasmid DNA production and they are intrinsically correlated in the PP pathway, composing the oxidative and non-oxidative phases respectively [68], [71]. This metabolic pathways, as well as gluconeogenesis, are both downregulated in P^+ cells [74]. It was demonstrated through a mathematical model that increasing the availability of NADPH via transhydrogenase activity has a positive impact on plasmid DNA production by increasing the reducing power available for pDNA and antibiotic resistance marker synthesis. Cells carrying high copy plasmid DNA require extra synthesis of nucleotides and in this case the carbon flux directed to the PP pathway may be insufficient to cover the cell's metabolic needs [71] .

The **tricarboxylic acid** (TCA) cycle has important roles in both energy metabolism and biosynthesis [83], and is composed of eight reactions that oxidize the acetyl group from acetyl-CoA or from other sources. This cycle is important in energy metabolism and biosynthesis and is essential to complete the glycolysis pathway. Some TCA intermediates play an important role in amino acids synthesis like oxaloacetate (OAA) and α -ketoglutarate (AKG). In plasmid - carrying cells, most of the TCA genes were observed as up-regulated for different *E. coli* strains [71].

Interestingly, although the host cell is continuously producing plasmid DNA, the expression of some genes involved on purine and pirimidine synthesis (*dut, tdk, pyrG and gmk*) is lower in P^+ cells [74]. The levels of heat-shock proteins and heat-shock genes are higher in P^+ cells, corroborating the fact that plasmid-induced stress response in the host cell has similar characteristics to the heat-shock response [74].

It was also demonstrated that cells containing plasmid had higher expression of acetate synthesis genes, namely *ackA*, which may result in increased acetate production, a metabolic by-product of *E. coli* fermentations which is known to influence cell physiology [74].

Although not consensual, probably due to the different host strains plasmid vectors and culture conditions used; taken together, the proteomics and transcriptomics results obtained showed that P^+ cells might adopt different respiratory pathways for energy production and that most of the biosynthetic pathways are diminished in these cells, including the ones related with plasmid DNA replication [74].

1.8 Engineering strategies to overcome plasmid DNA-induced metabolic burden

Over the past years, efforts are being made towards the metabolic engineering of *E. coli* in order to achieve increased plasmid yields [74]**,** [85]. 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 by products, such as acetate [10], [12]. Genes related to improving pDNA quality have also been common targets, as have genes that are involved in various other cellular processes relevant to pDNA production such as the stringent response and DNA replication [87], [79].

Two genes that are often knocked out to improve plasmid yield and quality are *endA* and *recA*. The *endA* gene encodes DNA-specific endonuclease 1 [71], that protects cells from invading DNA, however, upon lysis, EndA will degrade plasmid DNA [1], [73]. Deletion of *endA* can improve the quality of plasmid preparations by eliminating non-specific degradation of DNA by the endonuclease [66], [71]. However, plasmid nicking and degradation can also be caused by other, non-EndA-mediated factors [37], [71]. *recA* encodes for a protein essential for the *recBCD* pathway of homologous recombination, and produces a DNA strand exchange and recombination protein with a variety of functions, such as ATP-dependent recombination by catalysing DNA strand exchange and insertions [6], [73]. *recA* mutation, minimizes recombination of cloned DNA [52], ensuring less undesirable homologous recombination than wild-type cells [68], [71], improving plasmid segregational stability [35]. Singer *et al.* [68] have observed a positive impact of *recA* mutation on plasmid DNA yield [71]. However, Yau *et al*. [66] observed that the effect of some mutations, such as *ΔendA* and *ΔrecA*, are very strain and/or plasmid dependent.

Methylation patterns allow a species to distinguish between its own DNA from foreign DNA [71]. One common methyltransferase present in *E. coli* is the DNA - cytosine methyltransferase (Dcm) [71]. The *dcm* gene encodes a DNA methylase that methylates the internal cytosine residues in the recognition sequences: 5′-CC*AGG-3′ or 5′-CC*TGG-3′ [1], [86]. Plasmids produced in *Δdcm* cells had a high transgene expression level in a human cell line. Despite no strong impact of *Δdcm* mutant cells on plasmid yield and quality, would be recommended for gene therapy applications, once they demonstrated to be less immunogenic, producing lower antibody responses for the influenza H5 hemagglutinin [71], [86]. There is also a dam methylase, that encodes a DNA methylase that methylates the internal adenosine in the recognition sequence 5′-GA*TC-3′ to N^6-methyladenosine. Dam methylation is involved in multiple cellular processes including chromosomal and ColE1. *dam*$^-$ strains are mutagenic [1], and poor growing [86], not being used for large scale plasmid production [1].

The most commonly used strains for plasmid production, were originally developed to facilitate the cloning of heterologous genes and for the production of recombinant proteins and, thus, may not be the most appropriate for pDNA manufacturing. A more rational approach has been pursued, whereby genes are selected and mutated on the basis of *E. coli* central carbon metabolism and plasmid properties [79]. Reducing acetate and enhancing nucleotide production are some of the strategies that are used to increase pDNA yields [79]. Knockouts of *pykF* and *pykA* have been explored to reduce acetic acid formation in pDNA production strains [79], [63], [78]. Previous studies investigated the effect of *pykF* knockout on plasmid-free *E. coli* cells. The growth rate of the mutant cells was slightly lower than wild type, and acetic acid formation was smaller in mutant cells [5], [71],[5]. Other studies demonstrated similar behavior in an *E. coli pykF pykA* double knockout bearing plasmid [71]**,** [74] (Table 3).

Other promising targets for pDNA production strain engineering are the pentose-phosphate-pathway gene *rpiA*, which codes for ribose-5-phosphate isomerase A, and gene *zwf,* which codes for Glucose 6-phosphate-1-dehydrogenase. Increasing expression of *rpiA* would enhance synthesis of the nucleotide precursor ribose-5-phosphate (R5P). As a result, an increase in nucleotide formation, and consequently, in plasmid DNA production is expected. In fact, overexpression of *rpiA* in *E. coli* BL21 showed 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 [71]. It was verified that simultaneous overexpression of *zwf* and *rpiA* in DH5α appeared to increase plasmid amplification rate but not final specific yield, in fed-batch fermentations using complex medium with glycerol as the carbon source. However, overexpression of *zwf* alone as well as *zwf* and thioredoxin (*trxA*) to enhance reducing power did not show any effect on plasmid DNA yield. In a separate study, *zwf* overexpression increased *E. coli* growth rate [71].

Gonçalves *et al.* also tested the effect of Phosphoglucose isomerase *(pgi)* gene knockout in plasmid production, in order to redirect the carbon flux into the PPP, enhance the synthesis of nucleotides, and also provide high amounts of reducing cofactors (i.e., NADPH) [79]. It was expected that the abolishment of the *pgi* gene would redirect glucose-6-phosphate preferentially to the PPP, but glycolysis would continue due to the generation of fructose-6-phosphate and glyceraldeyde-3-phosphate [79], [65]. Indeed a large excess of NADPH was detected in Pgi mutant strains. *pgi* gene knockout in MG1655 along with *endA* and *recA* knockout, originated a strain called GALG20 ((MG1655*ΔendAΔrecAΔpgi*), that was able to produce 25-fold more pDNA than its parental strain MG1655*ΔendAΔrecA*, in glucose, identifying *pgi* gene as a target for metabolic engineering in plasmid production strains [65], [79].

Table 3- *E. coli* genes targeted for mutation to improve plasmid DNA production. ∆ indicates gene knockout and ↑ indicates gene overexpression (Adapted from [71]).

Gene (s)	Product (s)	Mutation	Expected phenotype
recA	DNA strand exchange and recombination protein; protease and nuclease activity	∆	Minimized recombination of cloned DNA, pairing and exchange between repeated DNA sequences
endA	DNA-specific endonuclease I	∆	Decreased non-specific digestion of plasmid
rpiA	Ribose-5-phosphate isomerase A	↑	Increased biosynthesis of nucleotide precursors in pentose phosphate pathway
Pgi	Phosphoglucose isomerase	∆	Increasing fluxes in the pentose phosphate pathway enhancing nucleotide synthesis and NADPH production
pykA pykF	Pyruvate kinase I, II	∆	Increased pentose phosphate pathway and TCA cycle flux and reduced acetate synthesis
polA	DNA polymerase I	↑	Increased availability of enzyme that extends the RNA primer template and removes RNA primers post-replication
Zwf	Glucose 6-phosphate-1-dehydrogenase	↑	Increased pentose phosphate pathway flux
Dcm	DNA-cytosine methyltransferse	∆	Improved transgene expression and reduced immunogenicity

On the other hand, highly mutated strains tend to be unstable, i.e., they have a relatively high mutation rate leading to changes of the chromosomal as well as the pDNA. This can be a problem for safety, quality, and yield in regard to the production of pDNA as a drug [6].

Cunningham *et al* [5] developed a stoichiometric model of *E. coli* metabolism in order to determine its maximum theoretical plasmid-producing capacity, and to identify factors that significantly impact plasmid production. Such a model was developed for the production of a high copy plasmid under conditions of batch aerobic growth on glucose minimal medium. The highest theoretical yield (592 mg/g) resulted under conditions of no marker or acetate production, nil Pyk flux, and the maximum allowable transhydrogenase activity. These results suggest that specific plasmid yields can theoretically reach 12 times their current experimental maximum (51 mg/g) [5].

1.8.1 *lon* gene

In 2010, a study of Mairhofer *et al.*, pointed *lon*-deficiency as beneficial mutation regarding pDNA productivity [4]. During the development of a marker-free plasmid approach, it was verified that the strain used, JM108 was Lon-deficient due to a transposon insertion in the spacer region between the −10 and −35 box of the *lon* promoter [4].

lon gene [88] encodes a cytosolic [89] ATP-dependent protease, member of the AAA+ protein family (ATPases associated with various cellular activities) [90], [91], [92], and exists as a soluble, single-ringed homo-oligomer [90], [91], [93]. Lon protease plays important roles in maintaining cellular functions by selectively eliminating misfolded and damaged proteins [94], [95], [96], preventing aggregation [95]; and by participating in regulatory circuits, controlling the amount and availability of specific substrates [89], [97].

After recognition and binding of the substrate to the Lon protease, hydrolysis of ATP permits substrate unfolding. The unfolded substrate is then translocated into the proteolytic chamber of the protease where processive peptide bond cleavage takes place [89]. The mechanism by which Lon recognizes the substrates is not fully understood, but it is believed that it might result from exposure, at the protein surface, of hydrophobic patches or structural motifs that are normally hidden in the protein core [89].

lon gene was first genetically identified on the basis of its regulatory properties like UV sensitivity and mucoid phenotype of *E. coli lon* mutants [89], [98], but an increasing number of naturally unstable proteins have been identified as Lon substrates, pointing out the importance of Lon regulatory activity in essential biological processes [89], [93]. For instance, it is involved in the control of DNA methylation via the Dam methylase [89], in capsule synthesis [89], [97], in lysogeny of certain bacteriophages [99], in carbonylated protein degradation [89], acid tolerance [97], and in SOS response [89], [100], [97]. It is also involved in degradation of free ribosomal proteins, in amino acid starvation situations. In these cases, protein degradation by ATP- dependent proteases is stimulated in order to increase the amino acid pool that will be used for the synthesis of specific enzymes (e.g., amino acid biosynthetic enzymes) required for adaptation to this particular condition. In starvation situations, *E. coli* accumulates poly-P that binds to Lon and redirects its activity towards free ribosomal proteins, then stimulating Lon-mediated turnover of free ribosomal proteins. This not only regenerates the amino acid pool but could also help to reduce the translational rate during starvation [89], [97]. The concentration of protease Lon in *E. coli* increases when cells accumulate large amounts of abnormal proteins, such as at high temperatures, which is part of the heat-shock response [98], [99].

A recent study, by Wohlever *et al.* [101] demonstrated that Lon functions are even broader than previously thought, once it can function as a protease or a chaperone and reveal that some of its ATP-dependent biological activities do not require translocation [93], [101].

It was verified that *lon* gene is not essential for viability in *E. coli*, however, even viable cells display some cellular defects [97], [98]. *lon* mutants accumulate abnormal proteins, as Lon is the main protease involved in unfolded protein degradation, form mucoid colonies [89], [98] and long filaments due to failure to degrade rapidly proteins that regulate filamentation and the expression of genes for capsular polysaccharide synthesis [97], [98]. *E. coli lon* mutants also fail to adapt rapidly to a nutritional downshift because they are unable to degrade free ribosomal proteins which would normally generate an amino acids pool required for the synthesis of adaptive enzymes [89].

Considering *lon* mutants phenotype, Mairhofer *et al*. [4] associated the increase in PCN with the occurrence of uncharged-tRNAs, as a consequence of amino acid starvation, due to interference of uncharged- tRNAs with the regulatory RNAs of the ColE1-ori (See section 1.5.5).

1.9 Growth conditions

Plasmid DNA is produced from a fermentation process and the success of this fermentation process hinges on the interactions between the host organism harbouring the recombinant plasmid vector and the growth environment [70]. Plasmid fermentation processes ideally maximize both the volumetric yield (mg/L) and specific yield (mg/g DCW or mg/OD_{600}/L) of high quality SC plasmid, since other plasmid are difficult to remove during purification and are considered undesirable isoforms by regulatory agencies (FDA, 1996) [1], [80], [102]. The use of reduced growth rate coupled with high copy replication origins is the unifying principle in high quality, high yield plasmid fermentations [6], [80], [103]. In a study to determine the effects of fermentation strategy on plasmid quality, O'Kennedy *et al*. [104] found that higher growth rates in batch and fed-batch fermentations were associated with lower percentages of SC plasmid.

1.9.1 Media Components

The cultivation medium formulation dramatically affects the performance and nature of microbial processes [70]. Fermentation media commonly contain a carbon source, a nitrogen source, various salts, and trace metals [2], [105].

The **carbon source** provides energy and biomass and is usually the limiting nutrient in cultures. Glucose is the conventional carbon source [106]. It is inexpensive and metabolized very efficiently and therefore gives a higher cellular yield [69], [105]. However, high glucose levels are known to cause undesirable acetate production due to metabolic overflow [106]. This by-product inhibits cell growth and represses recombinant protein expression [105]. Glycerol is also used often as carbon source in batch cultures, once it does not cause as much acetate excretion [2], [68], [107].

Nitrogen may be provided in inorganic (e.g. ammonium salts) or organic sources [3]. Ammonia and ammonium salts, e.g., NH_4Cl or $(NH_4)_2SO_4$ are used in minimal media. On the contrary, semi-defined media

supply nitrogen either partly or entirely from complex components: yeast extracts, peptones, and/or casamino acids [2], [44].

Regarding **minerals and salts** they are necessary for bacterial growth, metabolism, and enzymatic reactions. Magnesium, phosphorus, potassium, and sulfur are typically added as distinct media components. Other essential minerals include calcium, copper, cobalt, iron, manganese, molybdenum, and zinc. They are required in smaller amounts and often supplied by adding a trace-minerals solution [2].

Use of animal-derived products (meat extracts or casein hydrolysates), in particular bovine products, in plasmid manufacture is unacceptable because of the risk of prion (transmissible spongiform encephalopathy) or virus contamination, and their use is therefore restricted by pharmaceutical authorities. All media components should be certified as animal product free [2], [3].

1.9.2 Media type

Media composition can drastically affect plasmid quality and yield [80], [105], and therefore the overall cost of production. The fermentation process requires a balanced medium that supplies adequate amounts of nutrients needed for energy, biomass, and cell maintenance [2], [80]. Also, media for plasmid production should support high nucleotide pools in cells and supply energy for replication while minimizing other cell activity [2]. In addition, it is probable that medium composition will directly bear on the physiology of the microorganisms by influencing their intricate regulatory systems [44]. The following factors should be considered when formulating media for therapeutic plasmid production: effect of components on plasmid yield and quality, biomass yield, lot-to-lot consistency, potential interference with downstream purification and regulatory concerns [1], [2].

The fermentation media can be minimal, semi defined and complex. **Semi-defined media** is often used for plasmid production. This media is composed of both a defined portion (defined carbon source, salts, trace elements, vitamins) and a complex portion (nitrogen source). The complex components commonly used are yeast extract, casamino acids, and peptones [2], because they are relatively simple to prepare and generally lead to high cell densities [70]. The use of semi defied media is associated with a decrease in reproducibility, which is not necessarily a problem for plasmid production; however, reliable sources should be used to prevent too much variability over time [2]. O'Kennedy *et al*. [108] evaluated plasmid stability when employing several media (complex LB versus semi-defined formulations) and found that a semi-defined medium supported the highest plasmid stability, while the complex LB medium offered the lowest plasmid stability [44].

Regarding **minimal media**, it supply only minimum nutritional requirements, and is prepared from purified components instead of complex, biological mixtures, which makes fermentation processes using minimal media highly reproducible [3], [70]. Despite the fact that this medium allows better process analysis, fermentation monitoring, better solubility, absence of inhibiting by-products upon sterilization, and less foam

formation during cultivation [2], lower cell densities are often obtained [70]. Finally, **complex media** that is mainly composed by complex components, have the disadvantage that these components are originated from biological materials; therefore, the composition of the medium underlies normal natural deviations that make the cultivation process less reproducible and problems with contaminant removal in downstream processing [2].

1.9.3 Fermentation Strategies and conditions

The fermentation process is greatly influenced by some cultivation parameters, that affect plasmid quality and yield, which are temperature, dissolved oxygen (DO) and pH [2]. The formation of nicked plasmids and multimers can be affected by temperature, pH, dissolved oxygen, nutrient concentration, and growth rate [2], [80]. Also the SC-plasmid percentage is directly affected by oxygen and temperature conditions [103].

The optimal temperature for *E. coli* growth is 37°C. However, lower temperatures (30-37°C) may be used in batch fermentation to cause a reduced maximum specific growth rate [37], [80], [103]. Higher temperature (42°C) can also be employed to induce selective plasmid amplification with some replication origins such as pUC, and pMMl. It has been reported a process for the production of R plasmids wherein plasmid production is maintained at a low level (by use of low temperature) to avoid retardation of growth due to pDNA synthesis; once the host cell population is high, plasmid production is induced by temperature shifting [2]. On the other hand, acidic media pH (6.2–6.8) and growth at 30°C have been associated with increased production of the ColE1-based plasmid, pBR322 [2], [44].

An important consideration in pDNA process development is the configuration and operating strategy to propagate the cells [2].

Shake flask cultivation

The majority of the process development studies are conducted firstly in shake flasks and only then in bench- scale bioreactors with advanced monitor control [109]. However, previous studies have shown that productivity data obtained from shake flask experiments often fail to predict the outcome of pDNA production in bench-scale bioreactor [37], [80], [109]. In fact, many plasmids optimized for high yield in shake flasks are poor fermentation performers [80]. Still, shake flask continues to be the most commonly used culture vessel in bioprocess development. The standard shake flask is an erlenmeyer flasks (100–2000 ml) filled with low volumes of media (10–25%), shaken to promote fluid mixing and gas–liquid mass transfer via surface aeration. Therefore, the major limitation of shake flasks is their dependence on surface aeration, leading to reduced oxygen transfer compared to that achieved in stirred tank bioreactors [110]. Some of the shake flasks also contain baffles, which can improve oxygen transfer rate of the cell culture and hopefully contribute to a production performance that is closer to the one observed in bioreactors [109]. The need for carrying out a vast number of cultures has resulted in the increasingly widespread development of shake-

flask cultivations, once they offer a solution to bioprocess development, reducing labour intensity and material costs of the vast number of fermentative experiments [110].

A study for pDNA production assessment, developed by Gonçalves *et al.* [109], shown accumulation of acetate and reduced pDNA and biomass production during shake flask cultivation, in comparison to the same conditions in controlled bioreactor. This result can be associated with the lack of control of pH and dissolved oxygenation in shake flask cultivations, which suggests that manual control of pH could improve pDNA production, especially in high acetate production systems [109]. Also the use of a carbon source or a strain, that minimizes acetate production could represent some improvements in shake flask cultivations [109].

Fermentation strategies

After shake flask cultivation, pDNA production is usually tested in bench-scale bioreactor. The most commonly used fermentation strategies are: Batch Fermentation, Fed-Batch Fermentation and Continuous Fermentation.

In **batch fermentation**, all nutrients are present at the time of inoculation and no nutrients are added during cultivation (Figure 7). During the exponential phase all nutrients are in excess; thus the specific growth rate will be essentially the maximum specific growth rate, μ_{max}, as predicted by Monod kinetics [2], [80], [5]. Batch fermentations are usually simpler than fed-batch or continuous fermentations, but they have fundamental disadvantages that result in limited pDNA yields. Some of them include substrate inhibition and salt precipitation at high nutrient concentrations in the batch medium, as well, extended growth in batch mode resulted in increased contamination by gDNA [104].

On the other hand, in **fed-batch** culture, nutrients are continuously or semi-continuously fed, while effluent is removed discontinuously (Figure 7). The fermentation starts with a batch phase [67]. Controlled feeding of the limiting nutrient begins once the cells have consumed the initial amount of substrate [2].This strategy is especially useful for plasmid production. Controlled addition of a limiting nutrient allows for control of growth rate smaller than μ_{max} [46], [67]. Also, fed-batch fermentation results in higher biomass yields, and reduced metabolic overflow from excess substrate, avoiding excessive formation of inhibitory acetate [80], [107].

Finally, **continuous** culture is most commonly performed as a chemostat, a perfectly mixed continuous-flow stirred-tank fermenter (Figure 7) [2]. Continuous processes improve profitability, in comparison to a batch process, and their efficacy and applicability is proven [44], however, maintaining sterility is a major challenge.

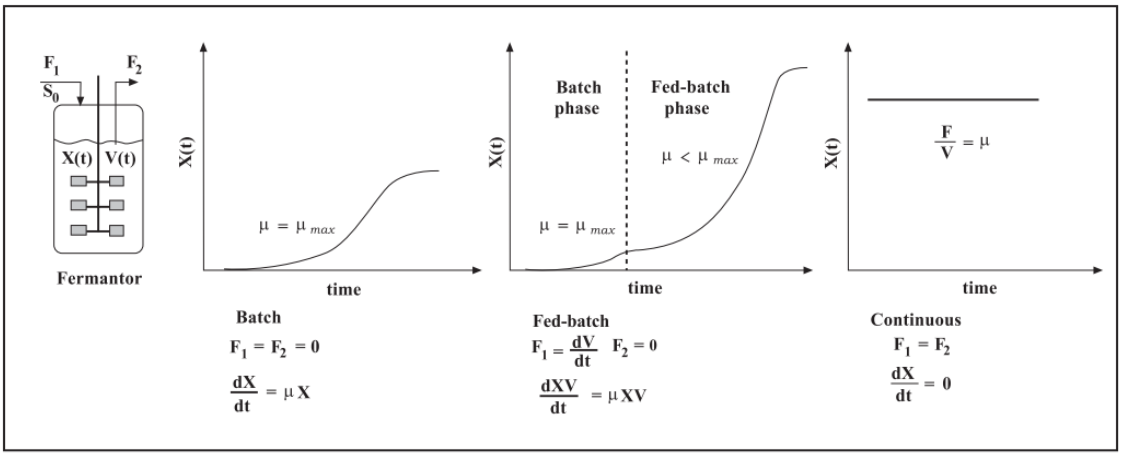

Figure 7 – Most frequent pDNA fermentation modes for plasmid production: Batch mode, Fed-batch and Continuous. The fermenter operates with pH and dissolved oxygen control. F1 is the input flow-rate, F2 is the output flow-rate, S_0 is the inlet limiting-substrate concentration, X is biomass concentration, V is culture volume, m is the specific growth rate and t is the time.

Alternatives to poorly predictable shake flask cultivations and expensive, time-consuming, labour intensive lab-scale bioreactor are necessary. Controlled micro-devices may represent a promising option [109], [110].

A schematic summary of the major factors controlling plasmid productivity in bioreactors is depicted in Figure 9.

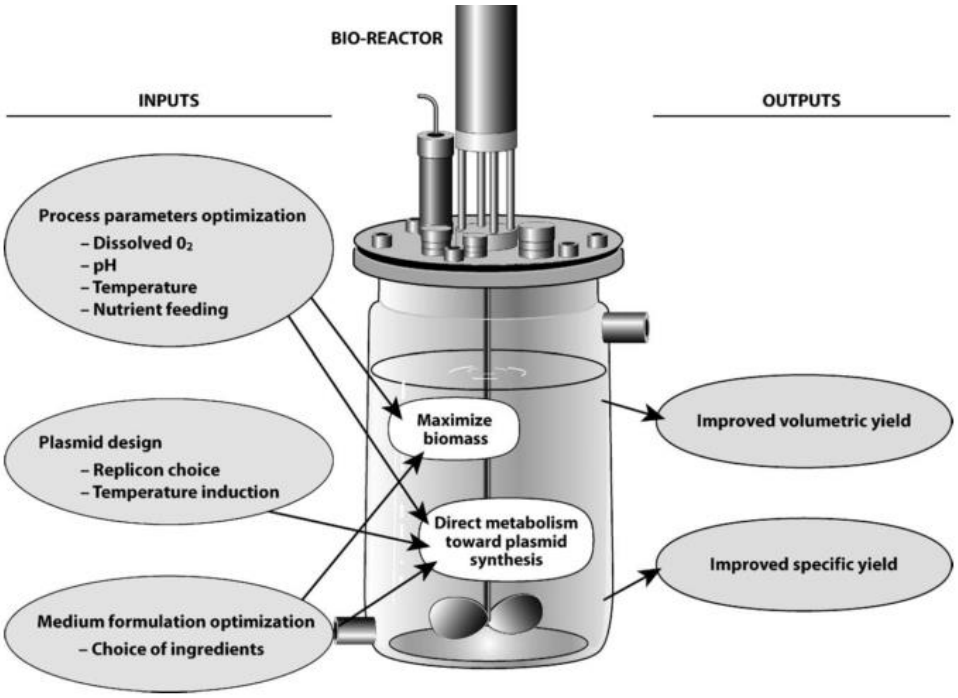

Figure 8 - Major factors controlling plasmid productivity in bioreactors [44].

In general, plasmid quality and yield is higher from fed-batch rather than batch fermentation. Furthermore, high plasmid homogeneity in the fermentation harvest is critical, since removal of nicked plasmid and dimers is extremely difficult due to similar properties to the desired supercoiled plasmid monomer product. Likewise, high yield is important since increased plasmid yield per gram of bacteria results in improved final product purity [46].

1.10 Downstream processing of plasmid DNA

After fermentation, a sequence of unit operations must be set up that are essentially aimed at eliminating impurities [111], [112], in order to obtain a final plasmid DNA preparation that complies with the approval specifications [111]. The order of steps is often as follows: cell harvest, cell lysis, clarification/concentration and finally purification [3], [44] (Figure 10). Cell recovery from the broth occurs by centrifugation or microfiltration [112], which provides a concentrated cell slurry, and removes the majority of the spent fermentation broth [44]. Cells are then resuspended and concentrated (10–20 times) in a buffer appropriate for the cell disruption and plasmid release [112], [113], [102], [114]. Cell lysis is a crucial step and the one at which most problems occur [112], which lead to the development of several early stage alternative mechanical, lysozyme or autolysis methods that may ultimately result in greatly improved industrial lysis technologies [80].

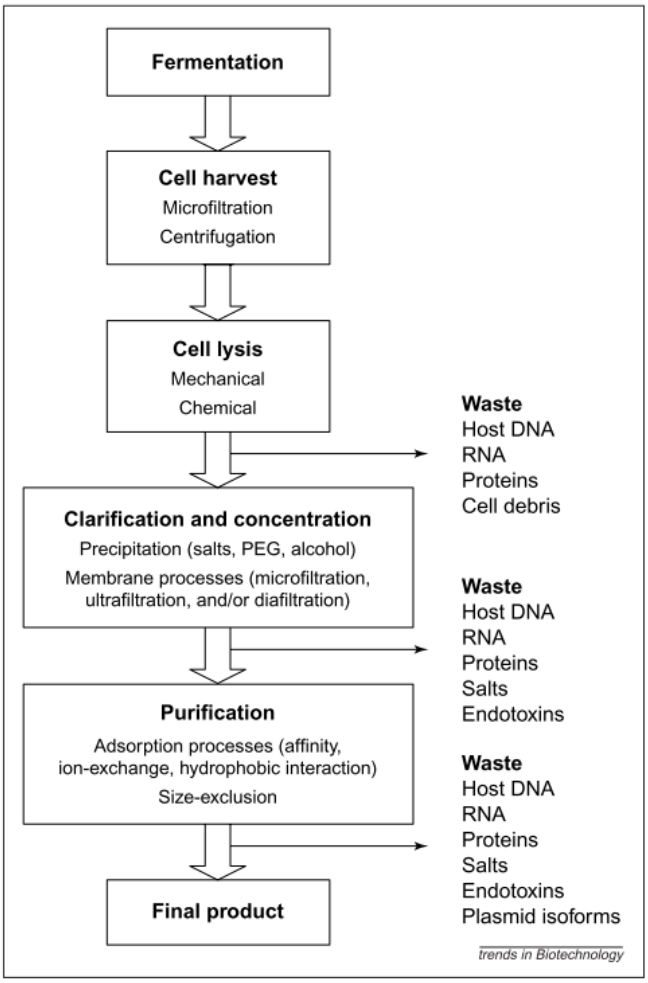

Figure 9 – Process flow sheet for the large-scale purification plasmid DNA. Representation of unit operations, and eliminated impurities during purification [112].

The strategy is to select and integrate appropriate purification operations in two distinct steps. In the first (clarification and concentration), which is performed after the lysis step, high-volumetric-capacity and low-resolution operations (e.g. salt precipitation) are performed in order to remove cell debris and structurally unrelated impurities, such as proteins and low molecular weight nucleic acids, while simultaneously concentrating and conditioning the plasmid-DNA preparations for the purification step [111]. In this second step, chromatography is used to

separate SC plasmid DNA from structurally related impurities such as relaxed or nicked circular and denatured plasmid DNA, genomic DNA (gDNA), high molecular weight RNA and endotoxins [3], [111]. It can be extremely challenging to separate alternate plasmid forms, so it is preferable to avoid generating them whenever possible, [3] which highlights the importance of plasmid design.

Plasmid isolation usually accounts for the majority of manufacturing costs, mainly, due to two factors 1) plasmid DNA makes up a small fraction of the cell mass produced by fermentation, and 2) separation of plasmid from host nucleic acids is difficult. Thus, the primary objective is to design a process that maximizes yield, purity, and safety, while minimizing cost [80].

1.11 Thesis motivation and aims

As pointed out previously, the use of plasmid DNA, for gene therapy and vaccination, is a recent technology with high potential in human and animal healthcare. One of the required improvements in this field is to avoid antibiotic resistance genes or other sequences used for selection and production of the therapeutic plasmid in *E. coli*, once these sequences are responsible for the main drawbacks in pDNA performance [12]. These sequences cause safety problems, are non-essential for the later product characteristics, hamper transfection efficiency by increased pDNA size and disturbing sequences [12], cause an additional metabolic load for the host [12], [52], and reduce the potency of the final pDNA-based product [115].

Mairhofer *et al*. [4], [115] shown that antibiotic resistance-free maintenance of plasmids is possible by using the ori-encoded RNAI instead of the antibiotic resistance markers. This system using a minimized plasmid that is devoid of any additional sequence elements on the plasmid backbone, and merely consists of the target expression cassette and the bacterial origin of replication [4]. pDNA selection and maintenance is achieved by functionally linking the RNAI, encoded on the ori of common ColE1-based plasmids, to a repressor protein, encoded on the host genome that further controls the expression of an essential gene (Figure 10). This approach allows gram-scale production of pDNA, devoid of antibiotic resistance genes or other additional sequences, in synthetic, chemically defined media, in contrary to other antibiotic-resistance-free pDNA approaches, dependent on additional sequences that need to be introduced in the plasmid backbone or have

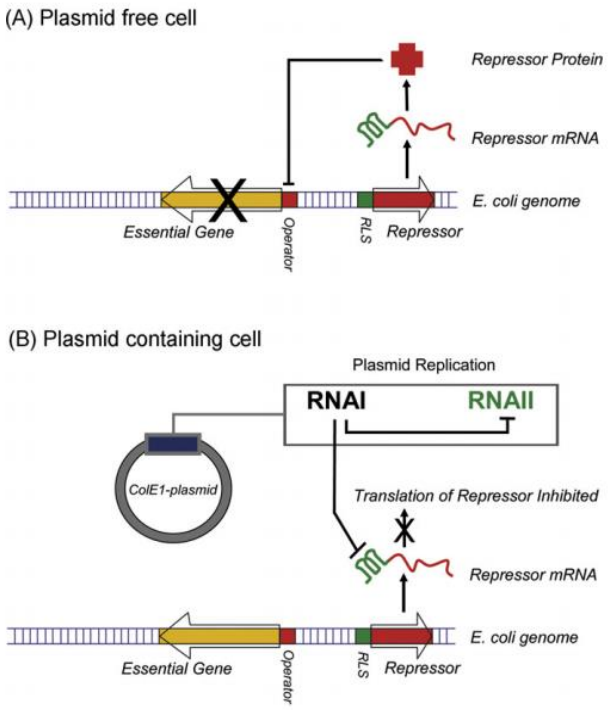

Figure 10- Schematic overview on the RNA-based plasmid maintenance system [4].

failed to be applicable in large-scale processes [4], [115]. The selection mechanism is based on RNA/RNA antisense interaction involving the naturally occurring RNA I derived from the origin of replication of the plasmid [115].

The system is based on ColE1-type origins of replication, one of the most commonly origins used for plasmid replication. This replicon is widely available, has been studied extensively and yields high copy numbers. ColE1-type plasmids require only host-encoded proteins for replication [14]. The ColE1 replication origin is a 600-bp cis-acting region from which DNA replication proceeds unidirectionally. An RNA preprimer, RNA II (555-nucleotides), is synthesized from a region close to the origin and must be cleaved by the host-encoded enzyme RNase H to release the 3' OH, which is elongated by the *E. coli* DNA polymerase I. This cleavage, however, is prevented if the complementary RNA I (a short 110-nucleotide antisense) binds to RNA II [14]. RNA I is constitutively transcribed and acts as a regulator of the frequency of initiation of replication, and thus of plasmid copy number, which acts as a very efficient anti-sense molecule. When they first interact, both RNA I and RNA II have folded structures [115], [14].

To develop this system, *E. coli* genome was modified by chromosomal engineering: an operator was introduced in front of an essential gene and the repressor gene corresponding to this operator is fused to an RNAII-like sequence (RLS) (Figure 10). In plasmid-free cells, the repressor gene is transcribed and mRNA thereof is translated to the repressor protein, which inhibits transcription of the essential gene by blocking its operator, which leads to cell growth inhibition in plasmid-free cells. Otherwise, in plasmid containing cells RNAI is deduced from plasmid replication control and translation of the repressor is inhibited with high efficiency by anti-sense hybridization of the RNAI to the RNAII-like sequence (RLS) fused to the mRNA of the repressor, allowing the expression of the essential gene, thus leading to cell survival [4] (**Figure 10**). The essential gene targeted for chromosomal engineering chosen was *murA*, which codes for the enzyme UDP-N-acetylglucosamine enolpyruvyl transferase and is involved in the first step of bacterial cell wall biosynthesis [116].

Using this system a 2-fold increase in specific pDNA content and productivity, in comparison with the same strain using antibiotic resistance as a selection mark, was attained. This huge increase could have been due to the decrease in metabolic load and consequently the increased availability of metabolic precursors, resulting from deletion of the constitutively expressed *nptII* gene; could be due to the decrease of plasmid size or the increase in pDNA replication rate; also, it could result from a combination of these factors [4].

During this study it was verified that JM108 is Lon-deficient due to a transposon insertion in the spacer region between the−10 and −35 box of the *lon* promoter [117]. In Lon-deficient strains, in consequence of amino acid starvation occurs the formation of uncharged t-RNAs. The uncharged tRNAs are able to interact with regulatory RNAs of the ColE1-ori, which suggests that the increase in specific pDNA content and productivity using this system could have influenced by *lon* gene mutation [4].

In conclusion, using the host strain JM108*murselct* for the production of marker-free pDNA it was possible to increase the overall plasmid yield 2-fold, compared to the conventional host/vector combination. Moreover, the produced pDNA is 29% reduced in size, which is an additional advantage concerning transformation efficiency and therefore potency of the product. It was demonstrated that the gram-scale production of antibiotic-resistance-free pDNA is feasible with the JM108*murselect* strain and that overall process performance benefits from deleting the constitutively expressed antibiotic marker gene[4], [115].

In a different study, developed by Gonçalves *et al.* [79] aiming to construct a high-yielding pDNA production strain, was created GALG20 (MG1655*ΔendAΔrecAΔpgi*) - a new pDNA production strain, constructed starting from MG1655 by knocking out the *pgi* gene in order to redirect carbon flux to the pentose phosphate pathway, enhance nucleotide synthesis, and, consequently, increase pDNA production [79]. This strain produced 25-fold more pDNA (19.1 mg/g DCW) than its parental strain, MG1655*ΔendAΔrecA* (0.8 mg/g DCW), in glucose. In addition to the high productivity achieved, the wild-type genetic background of this strain is also an advantage [79]**.**

Considering this, one of the goals of the present study is the transference of the marker-free system from JM108*murselect* into the high productivity strain GALG20 and quantifying the resulting plasmid DNA yields, in order to verify if in GALG20 the system achieve higher yields than JM108*murselect*. This study also aims to study the role of the *lon* gene in plasmid production and the possibility of using this mutation in the future as a target for cell engineering for plasmid production.

2 Materials and methods

2.1 Chemical and other reagents

In the polymerase chain reactions (PCR) performed, the polymerases used were the Nova Taq Hot Start Master Mix Kit (Novagen ®), Nova Taq DNA polymerase Kit (Novagen ®), and the Platinum PCR Supermix High Fidelity (Invitrogen®). For electrophoresis gel it was used Seakem LE agarose (Lonza ®)), and the DNA ladder used was the 10 kb DNA Ladder III from Nzytech. The L-arabinose (20%) solution was prepared from L-arabinose (Merck ®).

For phage P1 procedure were prepared the following solutions: 0.1 M calcium chloride ($CaCl_2$) (Merck ®); 0.1M Sodium Citrate (Merck ®); 10mM Magnesium sulphate ($MgSO_4$) (Merck ®). For cell cultivation studies, semi-defined medium was used, prepared from yeast extract (BD ®), Bacto Peptone (BD ®), $(NH_4)_2SO_4$ (Panreac ®), K_2HPO_4 (Panreac ®), KH2PO4 (Panreac ®), thiamine (Sigma Aldrich ®), $MgSO_4$ (Sigma Aldrich ®), Glucose (Panreac ®).

The antibiotics used: Ampicillin and Kanamycin are from Calbiotech®, while anhydro-tetracycline and Chloramphenicol are from Sigma Aldrich®.

2.2 Strains

The bacterial strains used in this study are indicated in Table 4 *E. coli* strain DH5α was used to replicate plasmids pKD13, pCP20, pkD46 and pVAX1-GFP. MG1655*Δpgi* was constructed by Geisa Gonçalves, former PhD student at IST. This strain was used as a starting point for the construction of the strains GALG20 and GALG20 *Δlon. endA* gene deletion in MG1655*Δpgi* was performed resorting to phage P1 protocol, and using JW2912-1 [*ΔendA*720::*kan*], from Keio collection, as donor strain [118]. *E. coli* K-12 strain JM108*murselect* [4] was donated by Jürgen Mairhofer from the Department of Biotechnology at the University of Natural Resources and Applied Life Sciences (Vienna, Austria). This strain was used as donor for the production of phage lysate used to infect cells, in the attempts to construct GALG20*murselect* and GALG20*murselectΔlon,* Gene deletions of *recA* gene*,* in the constructed strains, was carried out with P1 transduction, using as donor the strain JW2669-1 [*ΔrecA774::kan*], from the Keio collection [118].

Table 4 - Bacterial strains used in this study.

Strain	Genotype	Reference
DH5α	*F*-ϕ80*lacZΔM15 Δ(lacZYA- argF*)U169 *rec*A1 *endA*1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1	Invitrogen [119]
MG1655*ΔendA*	*F- λ -ilvG rfb-50 rph1ΔendA*	[79]
JW2912-1	F-, *Δ(araD-araB)567*, *ΔlacZ4787*(::rrnB-3), *λ-*, *ΔendA720::kan*,*rph-1*, *Δ(rhaD-rhaB)568*, *hsdR514*	[118]
JW2669-1	F-, *Δ(araD-araB)567*, *ΔlacZ4787*(::rrnB-3), *λ-*, *ΔrecA774::kan*,*rph-1*, *Δ(rhaD-rhaB)568*, *hsdR514*	[118]
GALG20	MG1655*ΔendAΔrecAΔpgi*	Developed in this study / [79]
JM108*murselect*	*F- thi Δ(lac-proAB) end A1 gyrA96 relA1 phx hsdR17 supE44 recA murselect*	[4]
GALG20 *Δlon*	MG1655*ΔendAΔrecAΔpgi Δlon*	Developed in this study

2.3 Plasmids

For gene knockouts three plasmids were used: pKD13, pCP20 and pkD46 (**Figure 11**). The pKD13 plasmid, carries kanamycin resistance gene flanked by FRT (FLP recognition target) sites, and was constructed by cloning of the Kanamycin gene into the pKD13 plasmid [120]. This plasmid is used to make an insertion cassette containing kanamycin resistance gene, surrounded by FRT sites. The pKD46 [121] plasmid comprises an Ampicillin resistance gene, and was made by insertion of the Red recombinase genes (β , γ , and exo) from phage l into pKD16 [122]. This plasmid is temperature-sensitive and can be cured by raising the temperature. The Red recombinase expression is induced with arabinose [79]. Plasmid pCP20 carries Chloramphenicol and ampicillin-resistance genes and carries the FLP recombinase. This plasmid shows temperature-sensitive replication and thermal induction of FLP synthesis [122]. It is used for the removal of the kanamycin cassette [122]. These constructions were performed at MIT and provided by Geisa Gonçalves, former PhD student at IST [79]. The main characteristics of the used plasmids are summarized in **Table 5**.

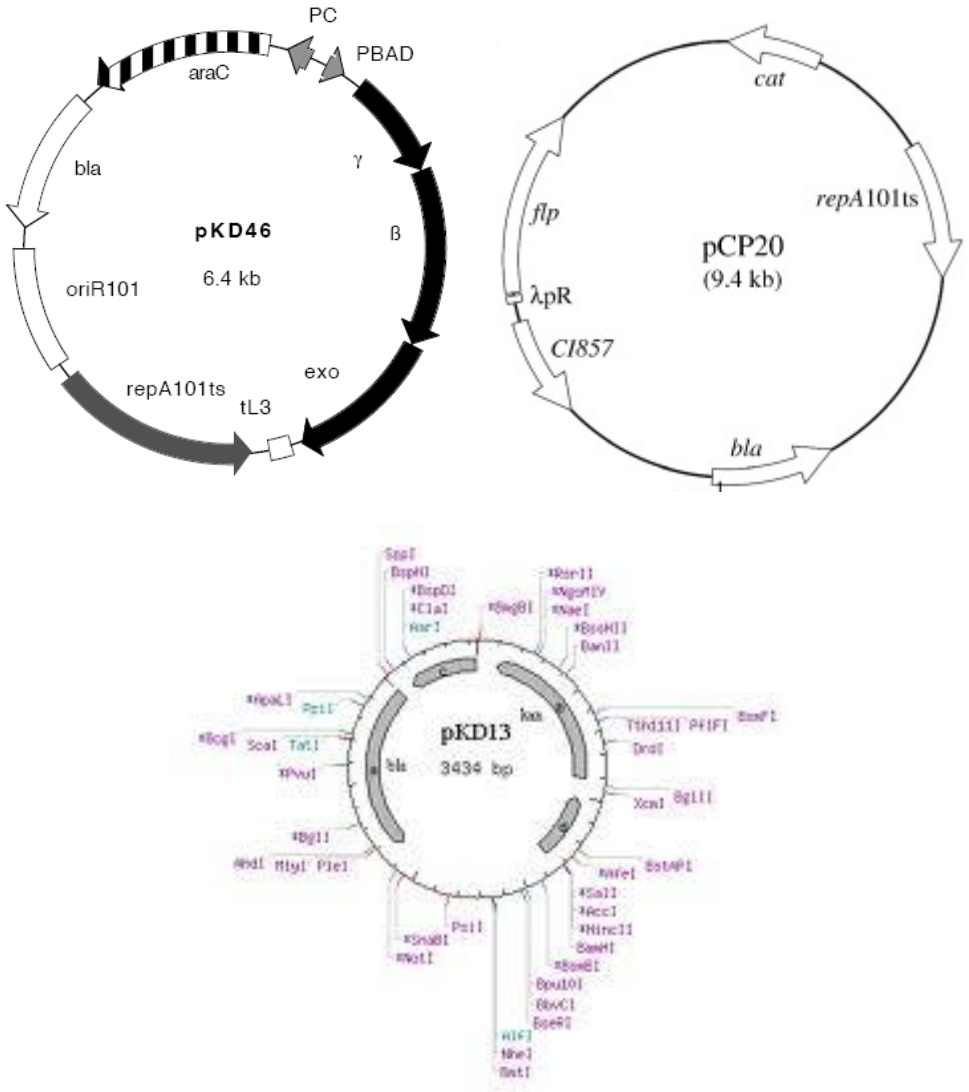

Figure 11 - Plasmids pKD46 [121], pCP20 [122] and pKD13 [123].

In order to explore the plasmid production potential of the strain GALG20 *Δlon,* and compare it with GALG20, these strains were transformed with the model plasmid pVAX1-GFP. This is a 3697 bp plasmid constructed from pVAX1LacZ (Invitrogen, Carlsbad, CA) [124] as described previously [125], and is one of the most commonly used backbones in DNA vaccine development [126].

The plasmid contains the green fluorescence protein (GFP) reporter gene under the control of the human cytomegalovirus (CMV) immediate-early promoter, the bovine growth hormone (BGH) polyadenylation sequence, a kanamycin resistance gene for selection in *E. coli*, and a pMB1 origin (pUC-derived) [127]. This origin contains two mutations, one single point mutation in the origin itself and *rom* gene deletion. This removes the regulatory constraints on the plasmids replication, which allows increased plasmid copy number [17].

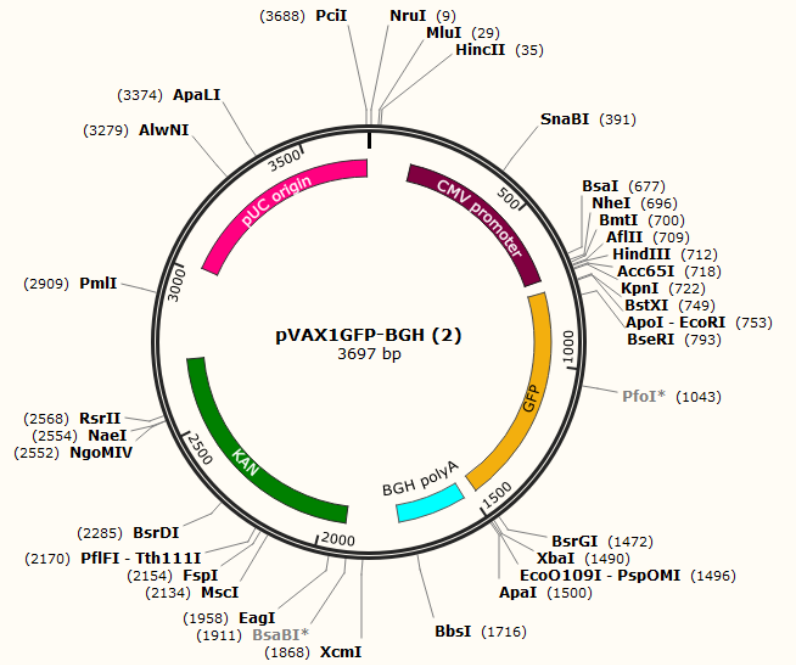

Figure 12 – Plasmid pVAX1GFP-BGH [125].

These plasmids were replicated in cells grown until middle exponential phase, according to the conditions described in section 2.5 and then harvested by centrifugation (3200g, 10 min, 4°C). The purification was made using the High Pure Plasmid Isolation Kit (Roche®). The plasmids were eluted in 100 µL of elution buffer (Tris-HCl Buffer, pH 8.5). Purified plasmids were used directly after purification or stored at -20°C. The plasmid concentration was determined using Nanodrop Spectrophotometer (Nanovue Plus, GE Healthcare®), and plasmid quality was assessed by gel electrophoresis (agarose 1% in TAE buffer 1x), in order to determine which were the predominant isoforms.

Table 5 – Main characteristics of used plasmids.

Plasmid	Size (bp)	Antibiotic resistance genes	Origin of replication	Purpose	Ref
pKD13	3434	Kanamycin and ampicillin	R6K gamma	Construction by PCR of insertion cassette containing kanamycin resistance	[123]
pKD46	6,329	Ampicillin	oriR101	Production of RecA recombinase	[121]
pCP20	~9400	Chloramphenicol an ampicillin	repA101-ts	Removal of the kanamycin cassette	[122]
pVAX1-GFP	3697	kanamycin	pMB1	Shake flask cultivations	[125]

2.4 Plasmid digestion

In order to verify if the restriction pattern of pKD46-purified plasmid corresponded to the expected pattern, plasmid was subjected to a restriction process. The restriction enzymes used for the restriction process were EcoRI and PstI (Figure 13). Two independent restriction reactions were performed with each enzyme, both with 8.7 µL of plasmid (1000 ng), 0.5 µL of enzyme, 5 µL of Buffer H and 35.8 µL of water to make a final volume of 50 µL per reaction. The reaction mixtures were incubated for 2 hours at 37°C. Then the reaction mixture was run in gel electrophoresis (agarose 1% in TAE buffer 1x), to verify the restriction patterns.

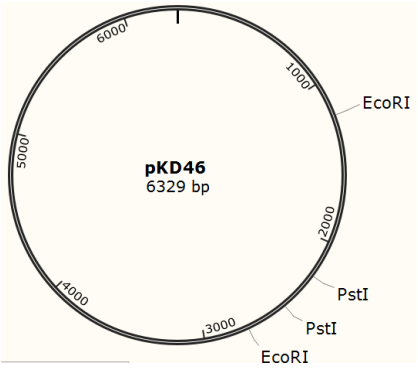

Figure 13 - Restriction recognition sites for restriction enzymes EcroRI and PstI, in plasmid pKD46.

2.5 Growth Media and conditions

E. coli DH5α cells were grown in liquid LB medium (Luria Broth, 20 g/L), from Sigma Aldrich, for plasmid replication. DH5α cells harboring pKD13 were grown in medium supplemented with 30 µg/mL of kanamycin at 37°C and 250 rpm, cells with pKD46 were grown in 100 µg/mL of ampicillin at 30°C and 250 rpm, whereas cells with pCP20 were grown in 50 µg/mL of Chloramphenicol at 30°C and 250 rpm. For gene knockout and phage P1 procedures *E. coli* strains were also grown in LB medium with the appropriate antibiotic, at 37°C and 250 rpm, except when specified otherwise.

When necessary, cells were grown in LB agar (2%) (Sigma Aldrich ®) plates, with the appropriate antibiotic, at 37°C, and then stored at 4°C.

Mutants were stored at -80°C in glycerol 15% (v/v) final concentration.

2.6 Transformation by electroporation

Cell transformation by electroporation is an established technique, consisting in the use of high-voltage electric shocks to introduce DNA into cells. Electroporation makes use of the fact that the cell membrane acts as an electrical capacitor that is unable to pass current (except through ion channels). Subjecting membranes to a high-voltage electric field results in their temporary breakdown and the formation of pores

that are large enough to allow macromolecules to enter or leave the cell [128] [129]. Considering that this technique yields a high frequency of both stable transformation and transient gene expression, that can be used with most cell types, and that requires few steps this was the chosen transformation technique used to insert plasmids and DNA cassettes in *E. coli* during this study.

Electrocompetent cells were prepared using cells grown until an OD_{600nm} of 0.5, in a 2 L shake flask with 50 mL LB, at 37°C and 250 rpm, inoculated with 500 µL of an overnight pre-inoculum.

Then 40 mL of the culture were centrifuged at 11000 g, during 15 min, at 4°C. After discarding the supernatant, the pellet was resuspended in 40 mL of ice cold sterile distilled water, and again centrifuged at the same conditions. Two more washing steps were performed under the same terms. After the last washing step, the supernatant was discarded, and pellet resuspended in 50 µL of ice cold sterile distilled water, for each electroporation cuvette prepared for electroporation. At this time, cells were electrocompetent. The electroporation cuvettes were previously placed on ice during 30 min.

To transform electrocompetent cells, 1 µL of DNA with a final concentration of approximately 10 pg/µl, was added to the cells. The mix was carefully transferred to a 2 mm electroporation cuvette, that was inserted on the electroporator (BTX ECM399) and one pulse of 2500 V was given. In each electroporation procedure was ensured that cells deposited across the bottom of the cuvette, and that no bubbles were introduced. Immediately after the pulse, 350 µL of LB medium were added to the cuvette, and cells incubated with agitation, for the cells to recover. The temperature and time used during recover depends on the plasmid or DNA fragment requirements. In addition, depending on the electroporated DNA fragment could be necessary to add antibiotic before recovery.

After incubation, cells were collected by centrifugation (11000 rpm, 5 min, 4°C), and most of the supernatant was discarded, leaving behind 100 µL used to resuspend the pellet and then plate in LB, and incubate overnight. Again, the supplementation with antibiotic and the appropriate temperature depends on the experiment.

2.7 Gene knockouts

2.7.1 *lon* gene Knockout

To evaluate the role of the *lon* gene in plasmid production, it was thought to knockout the *lon* gene form GALG20, to compare GALG20 (*lon*-), and GALG20 (*lon*+) growth curves and plasmid DNA productivities. In order to perform the knockout the method described by Datsenko and Wanner was adapted [122] (**Figure 14**). This method is named recombineering (homologous recombination-mediated genetic engineering), and provides a way to generate knockout mutations directly on the bacterial chromosome, in which the constructs are designed to the base pair and not dependent on suitable restriction sites [130].

This method can be used to disrupt chromosomal genes in *E. coli* [122], as well as to make deletions, point mutations, duplications, inversions, fusions and tags [130], and is based on the Red system [122], [130]. The recombination requires the phage λ Red recombinase, which is synthesized under the control of an inducible promoter on an easily curable, low copy number plasmid [122], [130], and consists of three proteins, Gam, Exo and Beta [130].

The basic strategy is to replace a chromosomal sequence with a cassette containing a selectable antibiotic resistance gene that is generated by PCR, by using hybrid primers with homology extensions (H1 and H2) to regions adjacent to the gene to be inactivated and to the template plasmid carrying antibiotic resistance gene that is flanked by FRT (FLP recognition target). The replacement of the chromosomal sequence is accomplished by Red-mediated recombination in the flanking homologies H1 and H2.

After selection, the resistance gene can also be eliminated by using a helper plasmid expressing the FLP recombinase, which acts on the directly repeated FRT sites flanking the resistance gene [122].

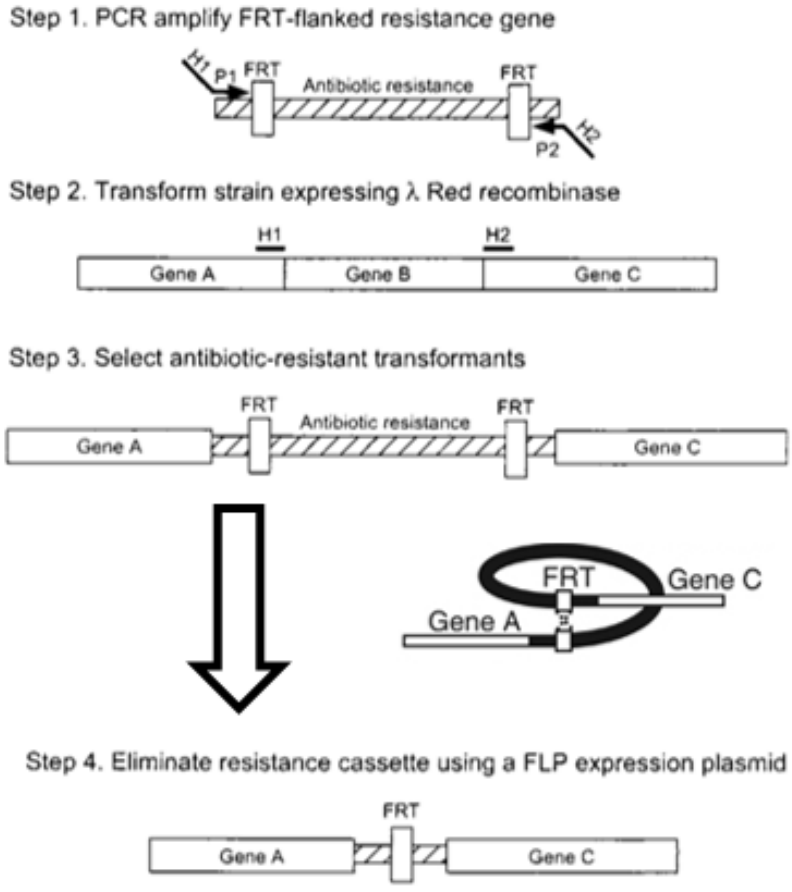

Figure 14 - Gene disruption strategy. H1 and H2 refer to the homology extensions. P1 and P2 refer to the priming sites. Adapted from [119] and [122].

The first step is to make the linear insertion cassette. The insertion cassette is composed by the selectable marker for Kanamycin resistance (Kan^R), surrounded by FRT sites, to facilitate later excision of the cassette, and regions adjacent to the gene to be inactivated (H1 and H2) (Figure 15).

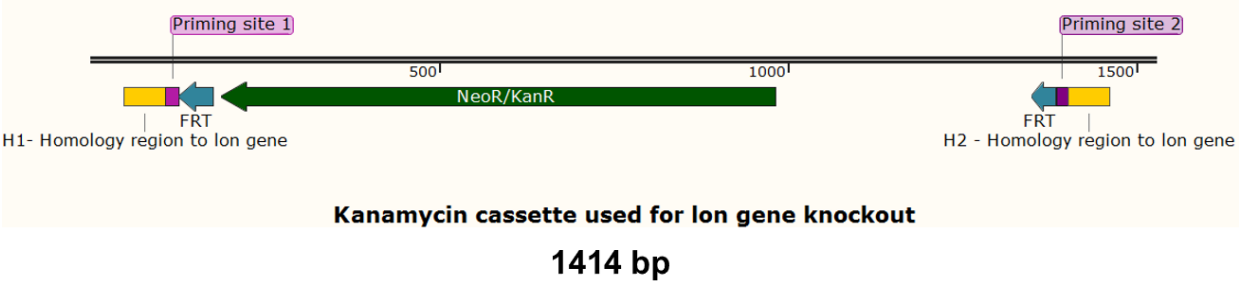

Figure 15 – Kanamycin cassette used for *lon* gene knockout, containing 1414 bp. In yellow is shown the homology regions (H1 and H2) to *lon* gene, in purple are represented the priming sites, in blue are highlighted the FRT regions where the FLP recombinase will operate to eliminate the kanamycin resistance gene, shown in green. (Scheme constructed using tool SnapGene.)

PCR amplification was used to construct this cassette. The DNA sequence of the *lon* gene was consulted in the EcoCyc database [117] and the drug resistance cassette used was the kanamycin gene from plasmid pKD13 (Figure 16) [120]. The Reverse primer used contained a homology sequence for *lon* gene, and a priming site coincident with the beginning of the Kanamycin resistance gene. The Forward primer also contains a homology sequence for *lon* gene, and a priming site coincident with the end of the kanamycin resistance gene. This primers were constructed with help of the tool Ape and for both primers was chosen a homology region with *lon* gene of 60 nucleotides, whereas the priming site 1 contains 20 nucleotides, and the priming site 2 contains 17 (Table 6). The evaluation of primers characteristic (melting temperature, CG content, hairpin and dimmers formation) was performed using IDTDNA [131] and SIGMA-ALDRICH [132]. These primers fulfil the requirements indicated by Datsenko and Wanner [20], of total primer size lesser than 100 nucleotides, and 50-70 nucleotides of homology to the area being targeted [122].

Table 6 - Primers used in Kan^R cassette generation for *lon* gene knockout. Homology sequences are represented by capital letters and correspond to the 5' and 3' flanking regions of the target sequence; priming site sequences are represented by small case letters. Tm-Temperature of melting; %GC- guanine cytosine content; F- Forward, R- Reverse.

Gene	Primer	Sequence	bp	Tm (°C)	%GC
lon	F	GACGTACATGTTAATAGATGGCGTGAAGCACAGTCGTGTCA TCTGATTACCTGGCGGAAAgtgtaggctggagctgcttc	80	71.3	50
	R	GCACTTGAATCCTTCAAGGTACGAACGCGCAGCAGTTATAT CAGGCCAGCCATCCCCTTAtccgtcgacctgcagtt	77	72.9	53.2

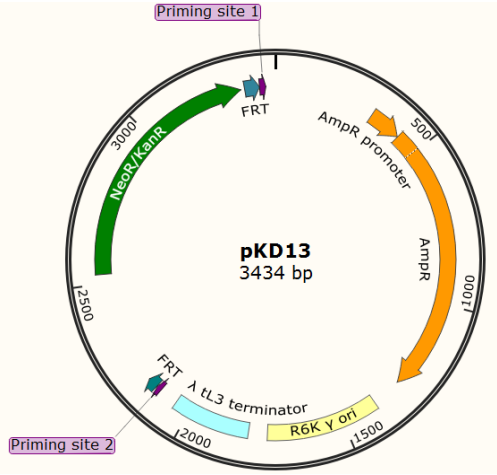

Figure 16 - Plasmid pKD13 used to construct the linear recombination cassette, with FRT regions (blue), priming sites (purple) and kanamycin resistance gene (green) represented.

After designing the primers, a PCR with the chimeric primers and the plasmid pKD13 (Figure 16) (containing the kanamycin resistance gene) was performed in order to make the linear recombination cassette, using Platinum PCR Supermix High Fidelity (Invitrogen), in a 50 µl reaction (Table 7). The PCR program is represented in Table 8. Latter, the PCR product was run in a 1% agarose gel with molecular weight markers (Ladder III from Nzytech) to confirm its size. If the size was in agreement with the expected, the PCR product was extracted, to make purified extract, from the gel (without exposure to UV or ethidium bromide) with the QIAquick Gel Extraction Kit (QIAGEN®), and the DNA concentration of the fragment was determined using Nanodrop Spectrophotometer (Nanovue Plus, GE).

Table 7- 50 µl PCR reaction mix for the construction of the cassette.

Component
45 µl Platinum PCR Supermix
1 µl template - pKD13
1 µl Primer Forward
1 µl Primer Reverse
2 µl Sterile MilliQ H_2O

Table 8- PCR program used to amplify the insertion cassette. The program was designed according to the specifications of the enzyme.

Insertion cassette			
		T (°C)	**Time**
	Incubation	94	2 min
35 cycles	Denaturation	94	25 s
	Annealing	55	25 s
	Extension	68	2 min

Then, the linear drug-resistant cassette made by PCR was transformed into recombination-competent cells (*E.coli* K-12 MG1655*ΔendAΔpgi*), by electroporation (Section 2.6). Previously, the host strain had been, transformed with the plasmid pKD46 (replicon is ampicillin resistant and temperature sensitive, which means that cells have to grow at 30°C), in order to express a λRED recombinase, that enables the integration of the cassette (removing wild-type sequence between the regions of homology). To induce recombinase, cells had to be grown in LB supplemented with L-arabinose, before electrotransformation with the cassette.

LB plates with kanamycin (30 µg/mL) were used to select successful recombinants and colony PCR was used to verify integration in the correct location. Primers check forward and reverse, were also designed with help of the tool Ape. The primer forward contains a 20 nucleotides length and the sequence can be found 107 nucleotides upstream of the initiation codon (ATG) of *lon* gene. The primer reverse also contains a 20 nucleotides length and the sequence can be found 225 nucleotides downstream of the termination codon (TAG). The primers were chosen according to size, melting temperature, CG content and process specifications. PCR conditions are represented in Table 8, using check primers represented in Table 10.

In case of integration in the right location, verified by agarose gel electrophoresis, positive clones were transformed with pCP20 (replicon is ampicillin and chloramphenicol resistant and temperature sensitive), and grown at 43°C to induce FLP recombinase that removes the kan^R cassette and cure all helper plasmids used. After this step, the selected colony were re-streaked on LB plates supplemented with ampicillin, kanamycin and chloramphenicol; and grown at 30°C overnight to confirm that all helper plasmids and kan^R cassette were fully cured. In the end, a second colony PCR to the region surrounding the modification should be performed to ensure the entire kan^R cassette removal.

2.8 Knockout of *endA* and *recA* genes by P1 transduction

Phage transduction is a useful genetic procedure for moving markers from one donor strain to a recipient strain [133], [134]. P1 is a bacteriophage of *E. coli* and other enteric bacteria [135], able to transfer fragments of approximately 100 kb [133]. P1 contains a rather sloppy packaging mechanism, and during growth and encapsidation of its DNA into the phage head, it occasionally packages random pieces of the chromosome, in place of phage genomes [136], [133]. The ability to package any piece of chromosomal DNA instead of phage DNA makes P1 a generalized transducing phage [134]. P1*vir* is a virulent phage mutant, traditionally used to prepare transducing lysates, that enters the lytic cycle upon infection, ensuring replication and lysis [134]. The phage is first grown on a strain containing the elements to be moved, and the resulting phage lysate is used to infect a second recipient strain [133]. During the replication and lysis of the phage in a culture of bacteria, a small percentage of the phage particles will contain a genome segment that contains the gene of interest [135]. Once a phage population has been generated from a donor host, the phage solution is used to infect a recipient host. Most of the bacteria are lysed by phage that packaged P1 genomes, but a fraction of the phage inject a genome segment derived from the donor host. This DNA pieces from the original host chromosome, can recombine and be permanently incorporated into the chromosome of the second strain, an event called transduction [133]. Successful P1 transduction of any one marker from a unique location in one bacterial strain to a second strain is a relatively rare event, so selection is required [133].The infected recipient bacteria are plated on a medium that selects for the genome segment of the donor bacteria. None of this would 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. Therefore, control of P1 infectivity is achieved by growing in the presence and absence of calcium [137]. Sodium Citrate works as a chelator, and lowers the concentration of free calcium, which prevents P1 infection [138].

The *endA* knockout in MG1655Δ*pgi* and the *recA* knockout in MG1655Δ*pgi*Δ*endA*Δ*lon* and in MG1655Δ*pgi*Δ*endA* were performed using P1 transduction protocol. In order to do that, JW2912-1 [*ΔendA*720::*kan*] strain from Keio collection [118], was used as donor strain for the first knockout. For the second, the strain used was JW2669-1 [Δ*recA*774::*kan*]. The donor strains contain a Kanamycin resistance cassette (kan^R) disrupting *end*A and *rec*A genes, providing selectable markers for the knockout of the genes.

*rec*A was the last gene to be deleted, once RecA Recombinase is required for the homologous recombination needed to integrate the donor DNA [134].

P1 transduction protocol can be divided in 3 important stages: preparation of donor phage stock, preparation of recipient cells and transduction.

For the **preparation of donor phage** stock, 30 µL of donor cells grown in LB supplemented with (30 µg/mL) kanamycin (37°C, overnight), were used to inoculate 5 mL LB medium with 0.1% glucose, 250 µL of $CaCl_2$ and kanamycin. After incubation at 37°C until OD_{600} = 0.15-0.25, 5 - 10 µL of P1 phage stock (ATCC #25404-

B1) were added. The donor cells continued incubating with phage P1, overnight. During that time, phage should infect and lysate donor cells, leading to clear up of the culture medium. Then 100 µL of chloroform were added, followed by vortex of the culture for 1 minute, to mix. Chloroform function is to kill the remaining live donor cells as well as help in the precipitation of cell debris. Chloroform is denser than water, so the organic phase, if present, will be at the bottom of the culture. Having in mind that the phage should be in the aqueous phase of the culture, 1 mL of this phase was recovered and centrifuged at maximum speed for 2 minutes. Again, the supernatant was recovered and 50 µL of chloroform were added, followed by vortex, to dissolve chloroform into the aqueous phase. Another centrifugation at maximum speed for 2 minutes was performed. At this stage, the supernatant is mainly composed by the donor phage stock that was later used to infect recipient cells.

For the **preparation of recipient cells,** 5mL of culture media, where cells were grown in LB at 37°C, overnight, were inoculated in 50 mL of LB, and incubated at 37°C until the OD_{600} was approximately 0.8-1.2, which corresponds to mid-exponential phase cells. At that time cells were centrifuged at 3200 *g,* 4°C during 15 minutes. After discarding the supernatant, cell pellet was resuspended in 300 µL of 10 mM $MgSO_4$ and 0.1 M $CaCl_2$, and then placed in ice (4°C) in 100 µL aliquots. This procedure was performed simultaneously with the preparation of donor phage stock.

The last step is the **transduction,** where the recipient cells were infected with donor phage stock. Recipient cells were incubated at 30°C for 30 min after the addition of 0, 30, 50, 70 and 120 µL of donor phage stock. During this time, P1 phage adsorbed and infected the recipient cells. The 0 µL sample was treated the same way the other samples, and was used as a control to make sure that the recipient cells did not grow on the antibiotic plate and give any false positives. To stop the infection process 100 µL of 0.1 M sodium citrate (pH 7.0) were added to each aliquot. Then cells were centrifuged at 5200 g for 1 min and the supernatant was discarded. Cell pellet was resuspended in 100 µL of 0.1 M sodium citrate, again centrifuged at 5200 g for 1 min, and the supernatant was discarded. One more sodium citrate wash was performed, before the last resuspension of the pellet in 100 µL of 0.1 M sodium citrate, followed by plating of the entire mixture on LB plates with kanamycin. 50-100 µL of the donor phage stock were directly plated on LB plate with Kanamycin, as a negative control to eliminate the possibility of live donor cells that survived to chloroform during the preparation of the donor phage stock. Plates were incubated at 37°C overnight, and in the next day colonies were re-streak from the non-control plates onto fresh plates because the original plates were covered with bacteriophage. Confirmation of Kanamycin resistance cassette insertion in place of the gene was performed by colony PCR, as described in Section 2.10

In case of kanamycin resistance gene integration in the right location, positive clones were transformed (Protocol described in section 2.5) with pCP20, and grown at 43°C to induce FLP recombinase that removes the kan^R cassette and cure all helper plasmids used. After this step, the selected colony were re-streaked

on LB plates supplemented with ampicillin, kanamycin and chloramphenicol; and grown at 30°C overnight to confirm that all helper plasmids and kan^R cassette were fully cured. In the end, a second colony PCR to the region surrounding the modification should be performed to ensure the entire kan^R cassette removal.

2.9 Transference of *murselect* trait by phage P1 transduction

In order to insert the *murselect* trait [115], [4] in MG1655*ΔpgiΔendA* and MG1655 *ΔpgiΔendAΔlon* the phage P1 transduction method was used. The protocol followed was similar to the one used for knockout of *end*A and *rec*A gene, described in section 2.8. However the donor phage stock preparation step, was not performed, once the P1 lysate carrying the *murselect* trait was sent from DBT, Austria. Therefore, the protocol was initiated in the preparation of recipient cells step, MG1655*ΔpgiΔendA* and MG1655*ΔpgiΔendAΔlon*. In this system after transduction, and sodium citrate washes the selection was made by incubation in plates containing Chloramphenicol (10 µg/mL) and anhydro-tetracycline (20 ng/µL).

The protocol was performed as previously described, and later repeated with some alterations. In addition to use volumes of 0, 30, 50, 70 and 120 µL of donor phage stock to infect the recipient cells, it were also used volumes of 5, 10, 20, 60, 80, 100 and 120 µL. Regarding the time of infection, a reduction from 30 minutes to only 20, was used. Furthermore, the time of incubation in plates was increased from 24 to 48 hours.

With the same purpose, of transference of *murselect* trait to MG1655*ΔpgiΔendA* and MG1655 *ΔpgiΔendAΔlon,* a different protocol for P1 transduction was also tested. As well, in this case the preparation of the donor phage stock was not performed. Initially the recipient cells were grown on 5 mL of LB supplemented with $CaCl_2$ (5 mM) at 37°C, 250 rpm, overnight. In the next day, 100 µL of donor phage stock were added to 300 µL of recipient cells, and incubated static at 37°C for 25 minutes. To stop the infection 5 ml of LB and 0.5 mL of sodium citrate were added. After centrifugation of the recipient cells at 4000*g*, 4°C for 15 minutes, the supernatant was discarded and the cells resuspended once more in 5 ml of LB and 0.5 mL of sodium citrate. Finally, cells were incubated 1 hour at 37°C to allow expression, and then plated in Chloramphenicol (10 µg/mL) and anhydro-tetracycline (20 ng/µL) LB plates, at 37°C, 250 rpm, overnight. Recipient cells not infected, and the phage lysate were also plated as controls.

2.10 Colony PCR

During gene knockouts experiments, it was necessary to verify, the replacement of the original genes by the DNA fragment containing the kanamycin resistance gene (or the *murselect* trait), and later the removal of the cassette. In order to do that for *pgi*, *endA* and *lon* gene knockout verification, colony PCR reactions were performed. Initially a colony of interest was touched, using a sterile pipet tip, and resuspended in 10.5 µL of sterile distilled water. The 25 µL PCR reactions contained Nova*Taq* Hot Start Master Mix, forward and reverse check primers, $MgCl_2$ and the sterile distilled water containing cells (Table 9). The Nova*Taq*™ Hot

Start Master Mix consist in a mixture of Nova*Taq* Hot Start DNA Polymerase, dNTPs and Nova*Taq* Buffer with $MgCl_2$.

Regarding *recA* gene knockout and *murselect* trait insertion verification, the protocol was slightly different. The DNA polymerase used was Nova*Taq*™ DNA polymerase, which does not have Hot Start activity. Therefore, an additional step is required for cell lysis and DNases denaturation. After touching the colony of interest, it was resuspended in 50 µL of sterile distilled water. Next, the tubes were placed in heat block (AccuBlock™ Digital Dry Baths, Labnet) at 99°C for 5 min, to lyse cells and denature DNases. Then, in order to remove the cell debris, solution was centrifuged at 12000 g for 1 min, and 10µl of the supernatant were transferred to new PCR tubes. The 25µl PCR reaction contained Nova*Taq* Master Mix, forward and reverse check primers, $MgCl_2$ and sterile distilled water containing DNA. Considering that the hot start DNA polymerase is inactive at room temperature, it requires an initial incubation step at 95°C for 10-15 min, using Nova*Taq*™ DNA polymerase, that step can be skipped (Table 9.The check primers and programs used were different for each gene and are represented bellow in Table 10 and Table 11.

Table 9 - Components for a 25 µl PCR reaction using NovaTaq™ Hot Start DNA Polymerase Kit and NovaTaq™ DNA Polymerase Kit.

NovaTaq™ Hot Start DNA Polymerase Kit		**NovaTaq™ DNA Polymerase Kit**	
Components	**Volume**	**Components**	**Volume**
Nova*Taq* Hot Start Master Mix	12.5 µL	Nova*Taq* Master Mix	12.5 µL
Check Primer Forward (~5 pmol/µL)	1 µL	Check Primer Forward (~5 pmol/µL)	1 µL
Check Primer Reverse (~5 pmol/µL)	1 µL	Check Primer Reverse (~5 pmol/µL)	1 µL
Sterile distilled water containing cells	10,5 µL	Sterile distilled water containing DNA	8 1 µL
		$MgCl_2$ (25 mM)	2.5 µl

Table 10 - Primer sequence and characteristics used to check for *endA, pgi, lon and recA* gene knockouts, and *murselect* trait insertion. These primers were designed using ApE software, IDT DNA [22] and Sigma Aldrich [23] and were synthesized by STABvida [24].

Gene	Primer	Sequence	N	Tm (°C)	%GC
endA	F	CGTCTATCGCTGTGTTCAC	19	53.2	52.6
	R	GGTTCAGGATGATAAATGCG	20	51.4	45
pgi	F	TCTGTGACTGGCGCTACAAT	20	56.5	50
	R	TAGGCCTGATAAGACGCGAC	20	56.5	55
lon	F	TTCCTCTATTCTCGGCGTTG	20	54.4	50
	R	GATCGACAAGATTGCTGCAG	20	54	50
recA	F	CTACAGTAACGAAGCCAAAG	20	50.9	45
	R	CTTTCTGATTCAGTTCCTGG	20	50.7	45
murA	F	GTACAACCGCCAGGTAGTG	19	55.5	57.9
	R	GTCTGATTTATCAGCGAGGC	20	53.3	50

Table 11- PCR programs used to check for *pgi, endA, lon and recA* gene knockout, and *muselect* trait.

		pgi		***endA***			***lon***		***recA***		***murA***	
		T (°C)	**Time**	**T(°C)**	**Time**		**T (°C)**	**Time**	**T (°C)**	**Time**	**T (°C)**	**Time**
	Incubation	95	15min	95	15min		95	10min	95	10min	-	
30 cycles	Denaturation	94	30 s	94	30 s	35 cycles	94	30s	94	30 s	94	1 min
	Annealing	64	30 s	57	30 s		54	1min	60	30 s	52	1:30
	Extension	72	1 min	72	1min		72	2:30	72	1 min	72	3 min
	Final elongation	72	10min	72	10min		72	10min	72	10min	72	10min

After PCR, amplified products were run in gel electrophoresis (agarose 1% in TAE buffer 1x). In order to confirm the size of the amplified fragments, molecular weight markers (Ladder III from Nzytech) were used.

2.11 Shake flask cultivation

For the cell, cultivation studies, single colonies of GALG20*Δlon* and GALG20 transformed with pVAX1-GFP were picked from the correspondent LB agar plate supplemented with 30 µg/mL of kanamycin, and then inoculated in LB medium containing kanamycin. Cultures were grown to mid-exponential phase, and then frozen at -80°C with glycerol as explained earlier (Section 2.5). The inoculum for the cultivation studies was prepared from these seed banks, in semi-defined medium (Bacto peptone, 10 g/L; yeast extract, 10 g/L; $(NH_4)_2SO_4$, 3 g/L; K_2HPO_4, 3 g/L; KH_2PO_4, 3.5 g/L; thiamine, 199 mg/L; $MgSO_4$, 1.99 g/L; Glucose, 20g/L [71]; trace element solution, 1 mL/L [76]) supplemented with 30 µg/mL of kanamycin. Cells were grown overnight at 37°C and 250 rpm, and then, used to inoculate batch cultures to an OD_{600} of 0.1. Cultures were grown in shake flasks with 50 mL of semi-defined medium supplemented with kanamycin, initial pH of 7.1, 37°C and agitation of 250 rpm. Cell cultures were grown during 24 hour, and monitored on an hourly basis for the first 12 hours and at the end of the 24 hours, for optical density measurements. At 12 hours were collected samples for plasmid DNA quantification. Cells were recovered and then purified using High Pure Plasmid Isolation Kit (Roche®), and plasmid concentration was determined using Nanodrop Spectrophotometer (Nanovue Plus, GE Healthcare ®), as described in more detail in Section 2.3. The plasmid concentration values obtained were used to determine the volumetric and specific plasmid production yield of each strain. Optical densities measured were used to construct a growth curve and calculate the growth rate of each cell.

3 Results and discussion

3.1 *pgi* gene knockout confirmation

In order to construct the desired strains: GALG20, GALG20*Δlon,* GALG20*murselec and,* GALG20*murselectΔlon,* the strain MG1655*Δpgi* was used as a starting point. First, this strain stored at -80°C was grown in solid medium and a colony PCR was performed to check for the *pgi* gene knockout, using primers represented in Table 10 and PCR program described in Table 11. The agarose gel of the PCR analysis is shown in Figure 17.

After performing the colony PCR analysis using the check primers for *pgi* gene there are three possible outcomes. The gene may be present, the kanamycin cassette may be inserted in *pgi locus*, or the PCR product may be the scar left by the removal of the cassette.

Figure 17 shows two different sized bands. In lanes 1, 2, 6, 7, 8, 9, 10 and 11 the product of amplification appears between the 200 and 400 bp DNA-ladder bands, corresponding to the scar left by the removal of the cassette (approximately 300 bp). Regarding lanes 3, 4 and 5, they exhibit a band of approximately 1650 bp, which corresponds to the size of the *pgi* gene.

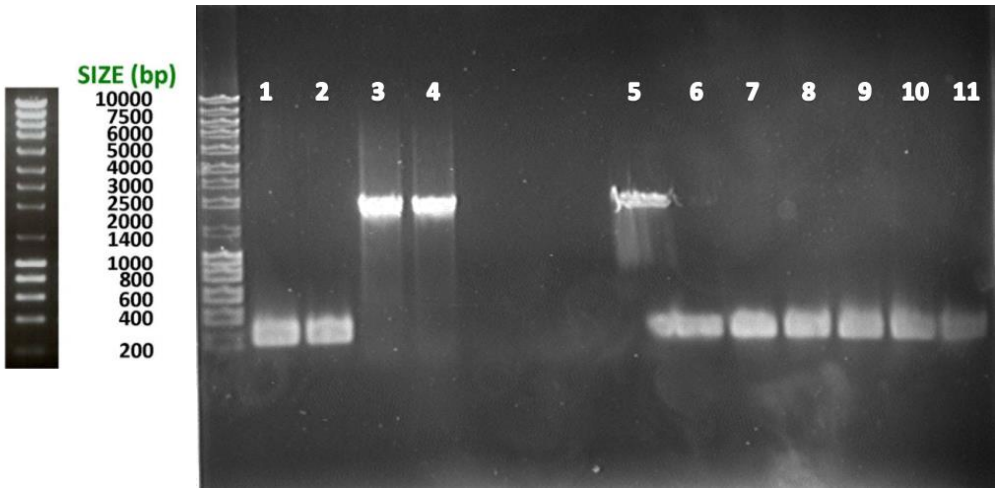

Figure 17- Agarose gel obtained from the colony PCR used to check for *pgi* gene knockout. In the first lane is Ladder III and the following lanes correspond to the analysis of different colonies. In lanes not identified by numbers no sample was added.

Considering that the goal is to construct strains lacking *pgi* gene and the kan^R cassette, the result of interest is the scar left by the removal of kanamycin cassette. That means that the gene was successfully replaced by the cassette containing the drug resistance gene, which was later removed by FLP recombinase, leaving a scar.

Therefore, colony corresponding to the product of amplification in lane 2, was chosen to proceed (randomly from colonies represented in lanes 1, 2, 6, 7, 8, 9, 10 and 11), and the following knockouts were performed in this strain.

3.2 *endA* gene knockout by phage P1 transduction

After confirming the *pgi* gene knockout, *end*A gene knockout was performed by phage P1 transduction, as described in section 2.8. Initially the phage donor stock was prepared from donor strain: JW2912-1

[*ΔendA*720::*kan*], from Keio collection; and then was used to infect the previously treated recipient cells MG1655*Δpgi*. After incubation, and selection in LB agar plates supplemented with Kanamycin, a colony PCR was used to confirm the insertion of the kanamycin resistance cassette in place of the gene, using the program and check primers shown in Table 9 and Table 10.

The agarose gel where the samples of colony PCR were run is displayed in Figure 18. The check primers used originate a PCR product with 959 bp when the gene is present, and a PCR product of 1510 bp for the cassette containing the kanamycin resistance gene. In lane 2, 3, 4 and 5 was obtained a single band estimated to have 1000 bp, which corresponds to the approximate size of the gene. Respecting to lane 1, it shows two bands, one with the same size of the band in the other lanes, and the second band appears slightly above the 1400 bp DNA-ladder band, which corresponds to the expected band size in case of kanamycin resistance cassette insertion (1510 bp). The desired result would be just the band with 1510 bp that appears in lane 1. However, after testing several other colonies none of them showed only this band. The colony used to obtain the result shown in lane 1, and some of the other colonies with a similar result (results not shown), were also streaked in LB agar plates supplemented with kanamycin, and re-tested, in order to obtain isolated colonies without double results. Nevertheless, the outcome was the same two bands in all the attempts. Therefore, it was decided to use the colony whose result is in lane 1, and proceed to the next step of cassette removal. Depending on the result, could be necessary to return to this step.

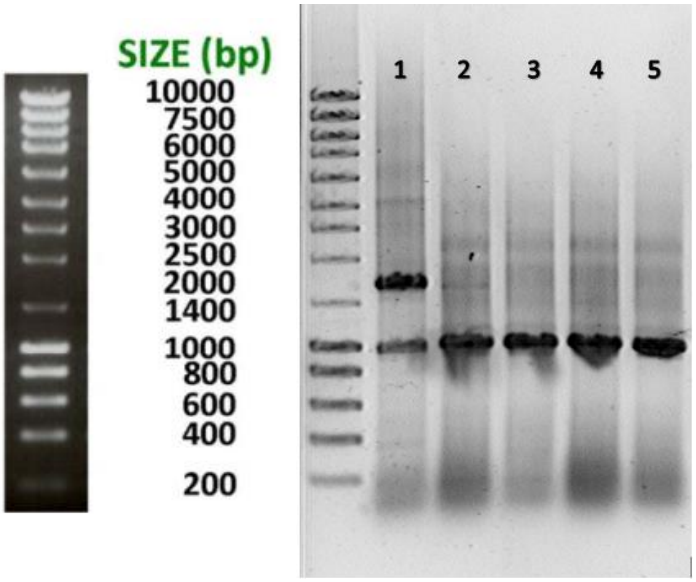

Figure 18 – Agarose gel analysis of a colony PCR result, used to verify the insertion of cassette containing kanamycin resistance gene in *endA* gene locus. First lane corresponds to Ladder III, and the following lanes correspond to different colonies analyzed.

The colony corresponding to the PCR product in lane

1, was transformed with pCP20 plasmid, and grown at 43°C to induce FLP recombinase that removes the Kan^R cassette and cure all helper plasmids used. Then, the selected colony was re-streaked on LB plates supplemented with ampicillin, kanamycin and chloramphenicol, to confirm that kan^R cassette was removed and all helper plasmids were fully cured. Then a Colony PCR was performed to ensure the removal of the kan^R cassette. The PCR was performed in the same conditions and with the same primers as in the PCR used to check for the insertion of the cassette, represented in Table 10 and Table 11.

In Figure 19 is shown the agarose gel where lanes B1, B2 and B3 correspond to the product of amplification from the colony PCR mentioned above. The amplicon exhibit approximately 350 bp, which corresponds to the scar left by the removal of the cassette with the drug resistance gene. The concern in the previous step related to the presence of the *end*A gene and the kan^R cassette at the same time, can be surpassed, once after the FLP recombinase-mediated recombination step, a single band was obtained, with the expected result.

At the same time that the colonies were tested for the cassette removal in *end*A locus, they were re-tested for the removal of kan^R cassette in *pgi* locus, to assure that the colonies used to construct the desired strains carry the expected knockouts. The confirmation of *end*A and *pgi* knockouts is shown in figure 19.

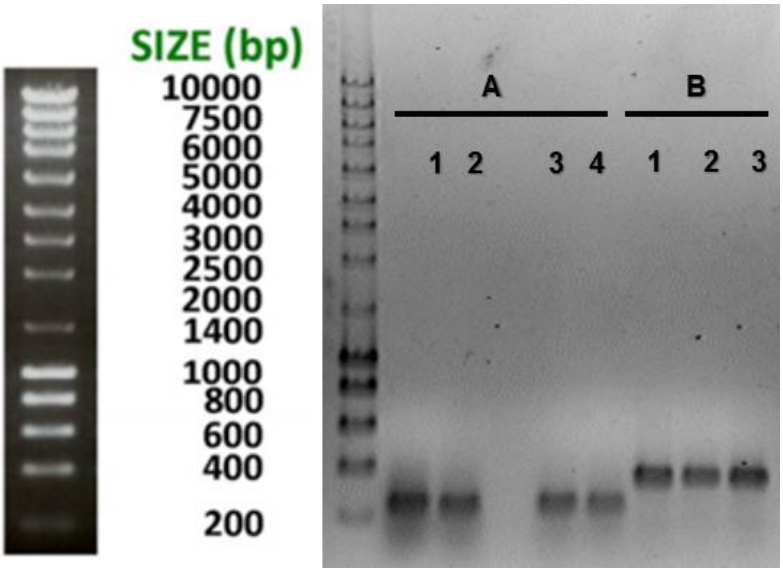

Figure 19 – Agarose gel showing the colony PCR result for the kan^R cassette removal for *pgi* gene **(A)**, and for the *endA* gene **(B)**. First lane corresponds to Ladder III, and the lanes A1, A2, A3 and A4 correspond to the test for *endA* gene, and lanes B5, B6 and B7 correspond to the test for *pgi* gene. Lanes with same number correspond to the same colony.

As it shown in figure 19, all of the tested colonies hold both knockouts, and the colony used for the next steps was colony B1, that was chosen arbitrarily among these colonies.

3.3 *lon* gene knockout by Datsenko and Wanner method

3.3.1 kanamycin resistance cassette construction

As explained in section 3.3, for the *lon* gene knockout the method [122], called recombineering, was adapted. The first step involves the PCR generation of a cassette with the kanamycin resistance gene, which replaces the *lon* gene by Red-mediated recombination. For the amplification of the cassette were used hybrid primers (Table 6), with a homology sequence for *lon* gene, and a priming site coincident with the beginning/end of the Kanamycin resistance gene. Using the chimeric primers and the plasmid pKD13, the PCR was performed to produce the linear recombination cassette (Figure 15). The homology regions to *lon* gene, the priming sites and the FRT regions in each extremity of the kanamycin resistance gene, compose the cassette, leading to a 1414 bp sequence.

After PCR, the amplification was verified in an agarose gel electrophoresis (Figure 20). A band with the approximate size of 1400 bp can be seen, corresponding to the intended size for the insertion cassette. It is also possible to observe a faint band that corresponds to the plasmid used for the PCR reaction. Considering that for fragment purification, only the desired PCR product is extracted from the gel, plasmid contamination does not represent problem. The purified band was not exposed to UV or ethidium bromide, once it was removed from a parallel lane.

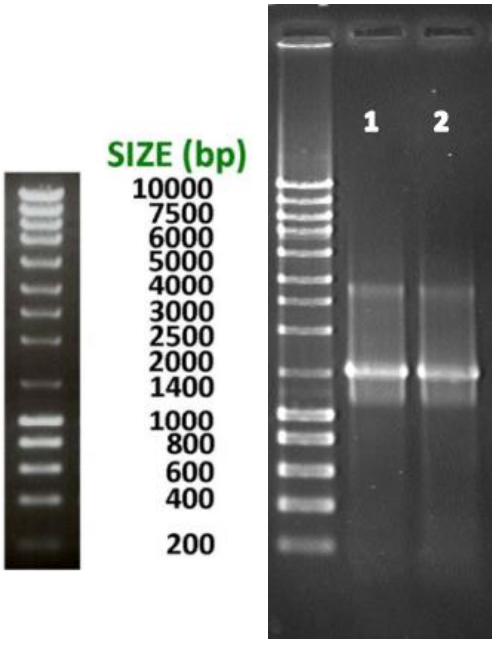

Figure 20 – Agarose gel obtained from the PCR used to produce the Kanamycin resistance cassette for *lon* gene replacement. In the first lane is Ladder III and the following lanes correspond to the PCR products.

3.3.2 Plasmid Digestion

After the kanamycin resistance cassette amplification and purification, several unsuccessful attempts to insert the cassette in MG1655Δ*end*Δ*pgi* strain were made. After eliminating several possible reasons for the failure in inserting the cassette, like problems in kan^R construction, transformation method, or colony PCR, it was thought that the lot of the pKD46 plasmid transformed into MG1655Δ*end*Δ*pgi* to promote recombination could have some problem and not being able to express Red-recombinase. In order to confirm the identity of pKD46 plasmid, the purified plasmid was digested using EcoRI and PstI restriction enzymes. It was verified that the restriction pattern did not corresponded to the expected when digesting pKD46 (data not shown). Thus pKD46 was produced and purified from another lot, and a new restriction reaction was performed to ensure that the plasmid obtained was pKD46 (Figure 21).

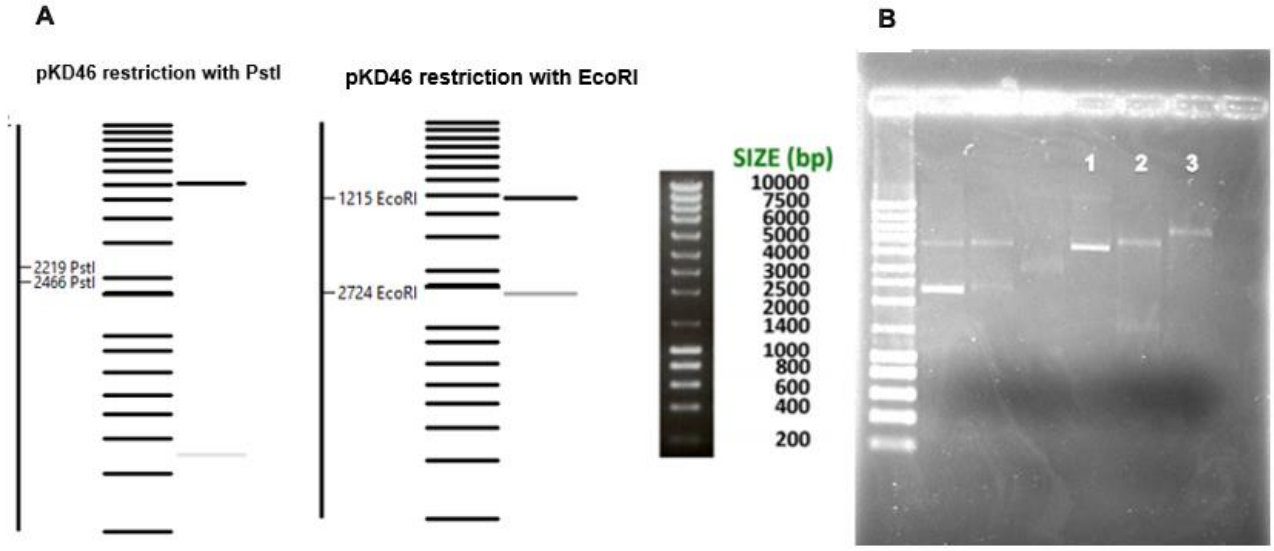

Figure 21- A) Prediction of restriction Reaction with restriction enzymes PstI, and EcoRI, obtained using Ape tool. **B)** Agarose gel analysis of restriction reactions in plasmid pKD46. First lane corresponds to Ladder III, lane 1- purified plasmid; lane 2- purified plasmid digested with EcoRI; lane 3 – purified plasmid digested with PstI. (The DNA ladder used in the prediction and in the agaroses gel is not the same.)

After restriction reaction with both enzymes, the mixture was loaded in an agarose gel to verify the restriction pattern. The expected results are indicated in Figure 21 A. The digestion using EcoR1 is expected to originate a 4820 bp band and a 1509 bp band, whereas PstI digestion generates a 6082 bp band and a 247 bp band. In Figure 21B in lane 2 is possible to see both expected bands resulting from EcoRI restriction, however in lane 3, the smaller band, product of PstI digestion, is not seen, probably due to its reduced size. Still, as the bigger band has the desired size it is possible to assume that the digestion was successful, and the purified plasmid corresponds to pKD46. This plasmid was transformed into MG1655Δ*end*Δ*pgi* strain to expressed Red-recombinase.

3.3.3 *lon* gene knockout

After purification of the cassette, it was transformed in strain MG1655*ΔpgiΔendA* harboring correct pKD46 plasmid, expressing Red-recombinase, to allow the insertion of kanamycin resistance cassette by exchange with the wild-type *lon* gene. To check for the insertion of the kan^R cassette, a colony PCR was performed, using the PCR program and the check primers designed for this purpose (Table 10 and Table 11), and the result can be seen in Figure 22.

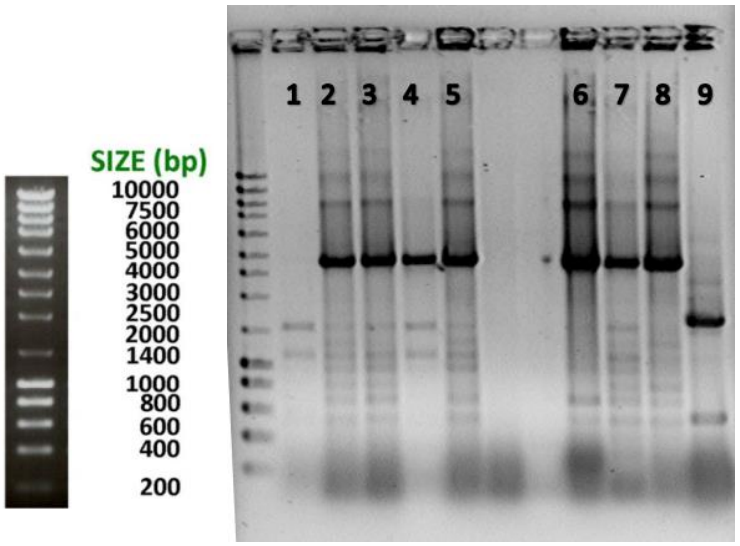

Figure 22 - Agarose gel analysis of a colony PCR result, used to check for cassette containing kanamycin resistance gene insertion in *lon* gene locus. First lane corresponds to Ladder III, and the following lanes correspond to different colonies analyzed.

Using the check primers described in Table 10, the product of amplification should be a 2727 bp fragment, in case of *lon* gene amplification. In case of kan^R cassette insertion the amplicon should have a size of 1526 bp. Finally, if the drug resistance cassette was removed, the size of the scar should be 303 bp. Figure 22 shows the Colony PCR result of 11 different colonies selected from the LB agar plates supplemented with kanamycin, where cells were incubated after transformation with insertion cassette. Two of the colonies (in lanes between lane 5 and 6) did not present any amplification. In turn, colonies whose amplification product can be found in lanes 2, 3, 4, 5, 6, 7 and 8, still possess the gene. This means that the Red-mediated recombination was not successful in the colonies that correspond to these amplicons. For the recombination to occur, Red recombinase must be synthesized, in this case from pKD46. Considering that the cells were transformed with the plasmid and then selected for ampicillin resistance (conferred by the plasmid), the explanation why recombination did not took place, is probably related with the efficiency of the method, or the extended size of the cassette.

On the contrary, in lane 1 and 9, the amplicon contains a band with a size slightly superior to the 1400 bp DNA-ladder band, that should corresponds to the 1526 bp fragment associated with the insertion of the

cassette containing the kanamycin resistance gene. Also, in both lanes the band corresponding to the *lon* gene amplification does not appear. Considering that despite the elevated number of cycles, and the reasonable time of extension used in the PCR program (Table 11) the band in lane 1 shows low intensity, the colony used to proceed to the next step is the one that corresponds to amplicon in lane 9. In all lanes, there are some unspecific bands, probably due to the PCR conditions, like low annealing temperature and high number of cycles.

The colony where the insertion of kan^R cassette was verified, corresponding to lane 9 in Figure 22, was then transformed with pCP20 plasmid that expresses FLP recombinase able to removes the Kan^R cassette. After incubation at 43°C to induce FLP recombinase and cure the helper plasmids, cells were tested in LB agar plates supplemented with ampicillin, kanamycin and chloramphenicol to confirm plasmid curing, and a colony PCR was performed to confirm cassette clearance. In case of cassette removal, the amplified product should be 303 bp long. From the colonies obtained after incubation at 43°C on LB agar plates, 6 colonies were chosen to be tested by colony PCR, as can be seen in Figure 23. All the colonies tested seem to have lost the kan^R cassette, as desired, with exception of colony represented in lane 3 that appears to have a low intensity band around the expected size for the kan^R cassette.

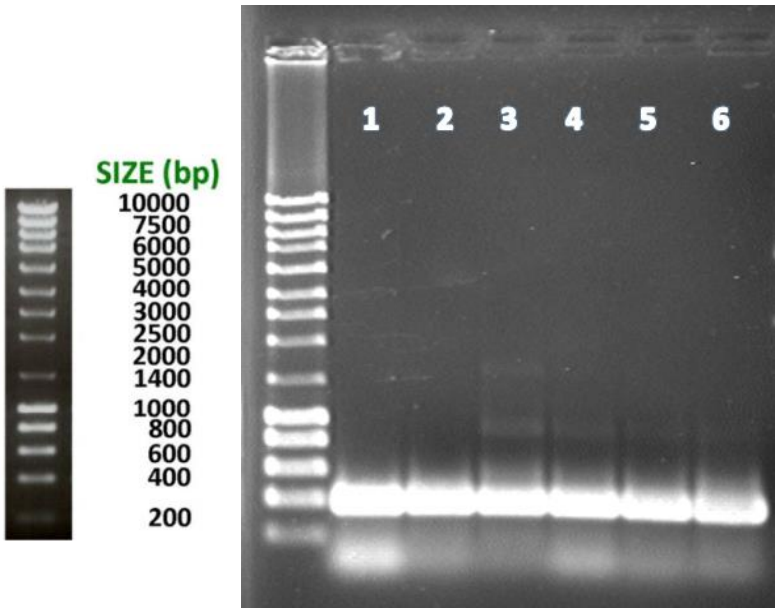

Figure 23 - Agarose gel obtained from the Colony PCR used to verify the removal of the kanamycin cassette from the *lon* gene locus. In the first lane is Ladder III and the following lanes correspond to different colonies analyzed.

Colony represented in lane 1 was chosen to proceed for the next knockouts, but before that, the sequence amplified was sent for sequencing to STABvida, using the same check primers used for the colony PCR.

The result was positive for the *lon* gene knockout and for the removal of the kanamycin resistance cassette, as the Colony PCR had shown (**Figure 24**).

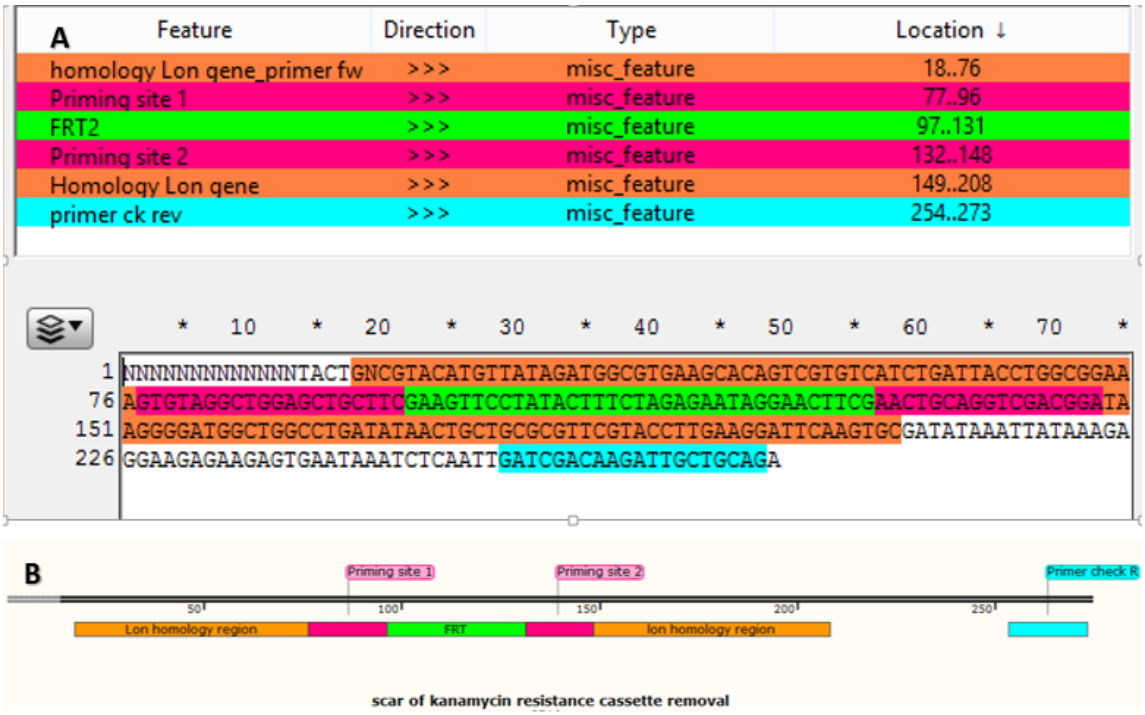

Figure 24- Sequencing result of the fragment obtained in lane 1 of the agaroses gel shown in Figure 23, displaying the result of the colony PCR used to test for *lon* gene knockout. A- Result from the sequencing reaction with each feature identified. B- Representation of the organization of each feature in genome after *lon* gene removal.

In the result of the amplification, the 5' extremity of the amplified fragment, particularly the region between the check primer and the homology region, was not possible to sequence (represented by N in place of the correct nucleotides), probably due to difficulties in the beginning of the sequencing, which is usual. In the 5' extremity of the sequenced product was found an insertion mutation, of one nucleotide. An insertion type of mutation can be particularly hazardous, once it can originate a frameshift mutation. In this case, the goal is the knockout of the gene, so the mutation is not relevant. The sequencing did not reveal any other mutation or irregularity, corroborating the previous positive result.

Cells from colony 1 (Figure 23) were stored at -80°C, and then submitted to the *rec*A gene knockout.

3.4 *recA* gene knockout by phage P1 transduction

This projects intends to create the strains GALG20 (MG1655Δpgi *ΔendA ΔrecA*) and GALG20*Δlon*. In order to construct both strains *recA* gene removal is required. The *recA* gene encodes a DNA strand exchange and recombination protein with protease and nuclease activity, which means that the gene removal will lead to minimized recombination of cloned DNA, pairing and exchange between repeated DNA sequences [52]. According to this information, the *recA* knockout should be the last to be performed, once the recombination efficiency is severally diminished. The gene knockout was performed simultaneously in both strains.

For the *recA* gene knockout JW2669-1 [*Δrec*A774::*kan*] strain, from the Keio collection [118], was used as donor cell. Donor cells were infected with phage P1, and produced the phage stock, used to infect treated recipient cells. After infection, cells were incubated and selected in LB agar plates supplemented with kanamycin, and some of the colonies tested, by colony PCR, for the insertion of the kanamycin resistance gene cassette in *recA* locus. The program and the check primers used in the colony PCR are displayed in Table 10 and Table 11.

Using the check primers previously indicated *recA* gene amplification is expected to originate a 2040 bp product. In case of kan^R cassette insertion in *recA* gene locus, the amplicon should be 2372 bp long. In the last step, when the kan^R cassette is removed, the scar should originate a 978 bp sized fragment.

3.4.1 *recA* gene knockout in MG1655Δ*endA*Δ*pgi*

For the construction of the GALG20 strain, the cells used as recipient cells were MG1655*ΔpgiΔendA*, specifically the cells stored at -80°C from colony correspondent to lane 1 in Figure 19. The result from the colony PCR performed after infection with phage stock is represented in Figure 25.

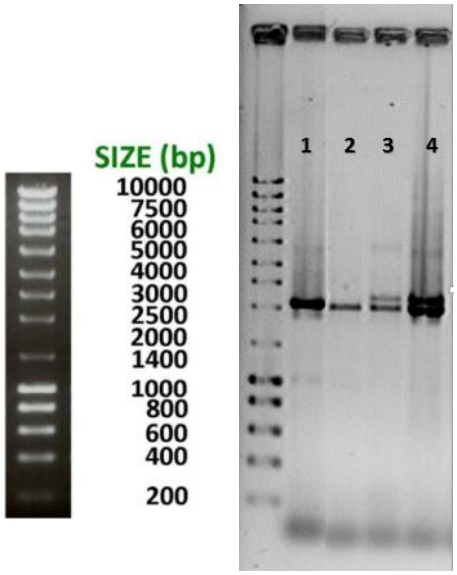

Figure 25 - Agarose gel analysis of a colony PCR result used to check for cassette containing kanamycin resistance gene insertion in *recA* gene locus. First lane corresponds to Ladder III, lane 1 is the positive control (colony with *recA* gene) the following lanes correspond to different colonies analyzed.

Lane 1, of Figure 25, shows a colony that was not exposed to phage stock and was used as a control; the visible band corresponds to *recA* gene amplification. Lane 2, 3 and 4 show the amplification product of three colonies exposed to phage stock. Transduction in colony represented in lane 2 was not successful, once the amplicon size is similar to the one in lane 1, corresponding to *recA* gene amplification. Colonies analyzed in lanes 3 and 4, show two bands, coincident with the size of *recA* gene amplification and kan^R cassette

insertion. The desired result for this stage would be the upper band, showing the insertion of the cassette. The general procedure was repeated, and a higher number of colonies were tested. After several attempts, it was not possible to isolate colonies displaying only the kan^R cassette insertion. Thus, it was decided to follow to next step of kanamycin resistance cassette removal, and according to the result decide if it was necessary to return to this step, or if the strain was complete. The colony chosen to continue to kan^R cassette removal was colony displayed in lane 4, once colony in lane 3 shows more intensity in the band corresponding to the gene than in the band corresponding to the cassette; on the contrary in lane 4 both bands show a similar intensity.

For kanamycin resistance gene removal, cells from colony in lane 4 were transformed by pCP20, incubated at 43°C for helper plasmids curing and for FLP recombinase induction and then re-streaked in LB agar plates supplemented with ampicillin, kanamycin and chloramphenicol to ensure plasmid curing and cassette removal. Colony PCR tested six colonies for kanamycin resistance cassette removal. Colonies tested did not growth in any antibiotic, after incubation in LB at 43 °C, and the agarose gel analysis is shown in Figure 26.

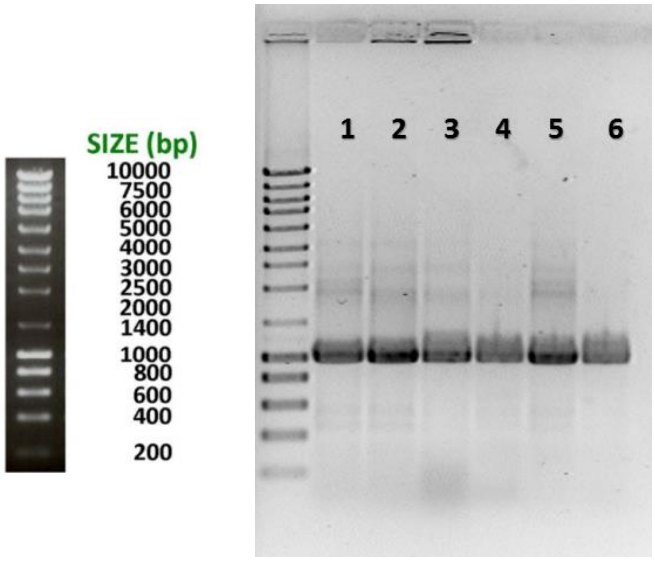

Figure 26 – Agarose gel obtained from the Colony PCR used to verify the removal of the kanamycin cassette from the *recA* gene locus. In the first lane is Ladder III and the following lanes correspond to different colonies analyzed.

All lanes exhibit a pronounced band around 1000 bp DNA ladder band, considering that the desired result would be a 978 bp product of amplification, the removal of the cassette was well achieved.

The kanamycin cassette removal from *recA* locus originates strain MG1655*ΔpgiΔendAΔrecA*, which corresponds to GALG20 genotype. This means that, at this time, one of the goal strains of this work was constructed. Thus, colony corresponding to lane 4 was stored at -80°C and later transformed with pVAX1-GFP and used for cultivation studies in shake flask.

3.4.2 *recA* gene knockout in MG1655Δ*endA*Δ*pgi*Δ*lon*

For the construction of the GALG20 *Δlon* strain, colony represented in lane 1, Figure 23 was submitted to phage P1 transduction for *recA* gene knockout. Colony PCR result can be seen in Figure 27.

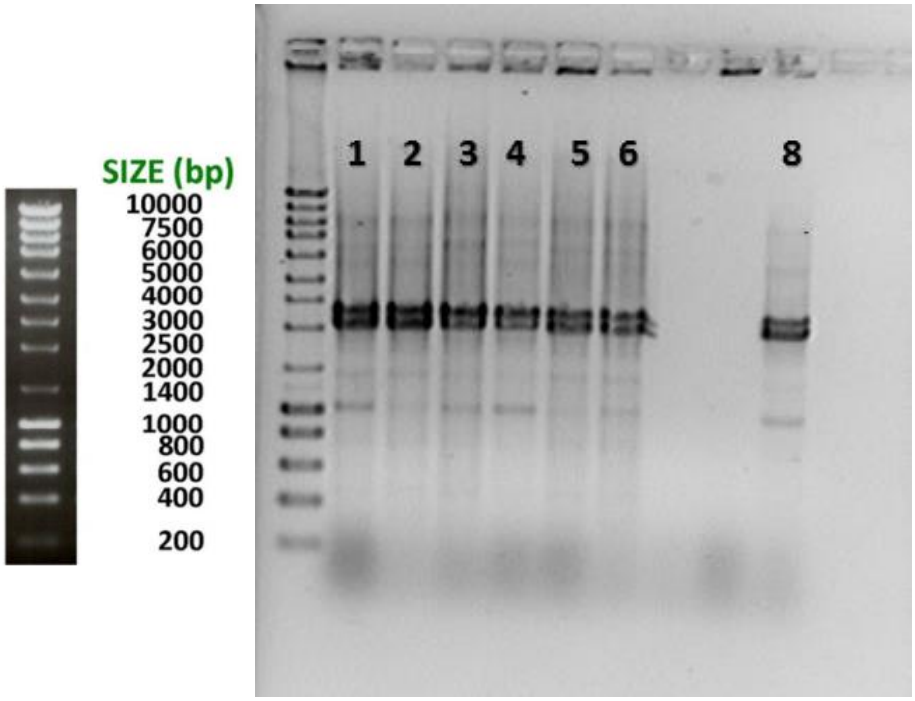

Figure 27– Agarose gel obtained from the Colony PCR used to verify the insertion of the kanamycin cassette in *recA* gene locus. First lane shows Ladder III and the following lanes correspond to different colonies analyzed.

All the colonies where amplification occurred (colonies represented in lanes 1, 2, 3, 4, 5, 6, 8), showed a triple result. It is possible to distinguish three individual bands with 2040 bp, 2372 bp and a low intensity band with 978 bp. The intended result would be a single band with 2372 bp corresponding to kanamycin resistance cassette insertion. Some previous procedures showed the same problem. As it was done before, the colonies were re-streaked, and a higher number of colonies were tested aiming to get colonies with a single amplification product, however no colonies were obtained.

Colony corresponding to the amplicon in lane 2 was chosen to be transformed with pCP20 plasmid, in order to remove the kanamycin cassette. The desired result is only the 978 bp band, corresponding to the scar of kanamycin resistance cassette removal.

After transformation and incubation, colonies without resistance to ampicillin, chloramphenicol and kanamycin were submitted to colony PCR to check for drug cassette removal (Figure 28).

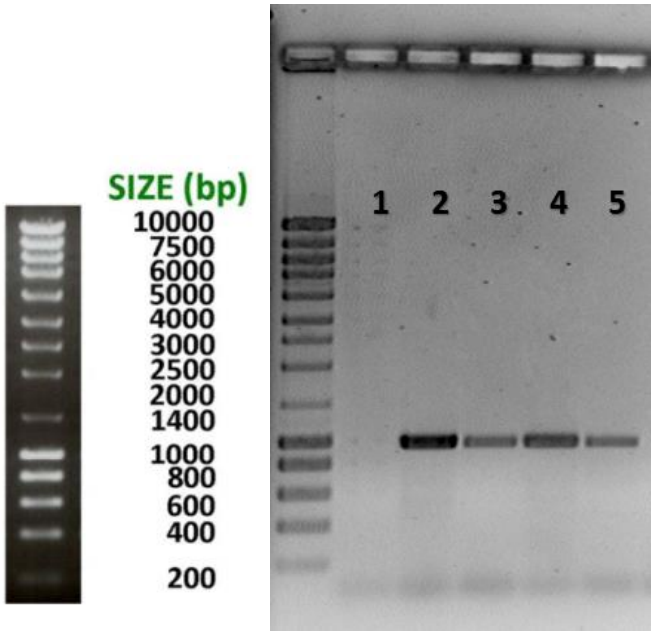

Figure 28 – Agarose gel analysis of a colony PCR result used to check for kanamycin resistance cassette removal of *recA* gene locus. First lane corresponds to Ladder III, lane 1 is a negative control (the PCR was performed without DNA template), the following lanes correspond to different colonies analyzed.

The colonies represented in lanes 2, 3, 4 and 5 exhibit a single band with 978 bp. The absence of bands with 2040 bp and 2372 bp indicate that the cassette was removed, and colonies with the gene and with the gene intersperse with kanamycin were eliminated. These colonies are GALG20 *Δlon* which is one of the strains aspired in the project. Colony corresponding to the amplicon in lane 2 was stored at -80°C and later transformed with pVAX1-GFP plasmid and used for cultivation studies in shake flask.

3.5 *murselect* trait insertion

Considering the previous highlighted disadvantages of antibiotic resistance genes, in plasmid backbone, and also the benefits of using the marker free system, provided by the *murselect* trait described in Mairhofer *et al.* study [4], the construction of GALG20*murselect* and GALG20*murselectΔlon* were also goals of this project. For the construction of these last two strains, *murselect* trait (cassette with marker free system) has to be inserted in MG1655*ΔpgiΔendA,* and in MG1655*ΔpgiΔendAΔlon*, once *recA* knockout should be the last knockout to be performed, as explained above.

Recipient cells MG1655*ΔpgiΔendA,* and MG1655*ΔpgiΔendAΔlon* were treated according to section 2.9, and then infected with the phage stock. Phage stock was provided by the Department of Biotechnology (DBT), University of Natural Resources and Life Sciences, Vienna, Austria.

After transduction, cells containing *murselect* cassette inserted on the chromosome were selected by incubating overnight on LB plates supplemented with chloramphenicol (10 µg/mL) and anhydro-tetracycline (20 ng/µL), at 37°C. Colonies obtained after incubation overnight were tested for the insertion of the *murselect* trait in *murA* locus, by colony PCR, using the PCR program and primers indicated in Table 10 and Table 11. Colony PCR result can be seen in Figure 29.

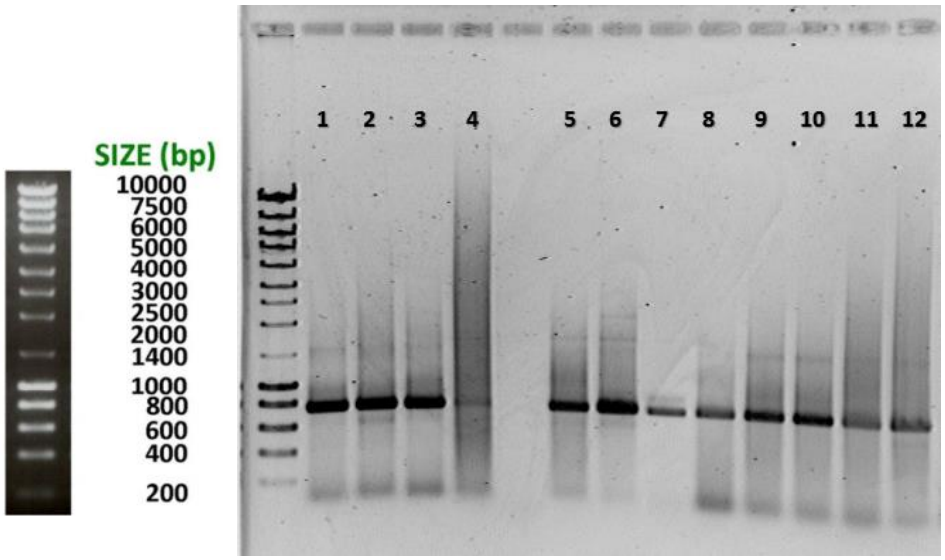

Figure 29 - Agarose gel analysis of a colony PCR result used to verify the insertion of *murselect* trait in *murA* gene locus. First lane corresponds to Ladder III, lanes from 1 to 4 correspond to results of colonies MG1655*ΔpgiΔendA,* and lanes from 5 to 12 show the results of colonies MG1655*ΔpgiΔendAΔlon.*

MG1655*ΔpgiΔendA* colonies are represented in lanes 1 to 4, while MG1655*ΔpgiΔendAΔlon* colonies are represented in lanes 5 to 12. In case of *murselect* trait insertion, the expected PCR product would be a 4273 bp fragment, otherwise, if the cassette is not inserted, the gene fragment amplified show a 751 bp fragment. As it is possible to notice in Figure 29, all colonies tested resulted negative for *murselect* trait insertion.

Several protocol modifications were performed, as extending the incubation period, vary the volume of phage stock used and increasing the $CaCl_2$ volume in donor cells preparation step to increase to adsorption of phage to cells. It was also tried a different transduction protocol. Despite all efforts, it was not possible to achieve any positive colony for the insertion of the *murselect* trait. The fact that JM108*murselect* is a *recA* minus strain was pointed as one of the possible reasons for the unsuccessful insertion of the cassette, once phage P1 replicate poorly in *recA* minus strains. In addition, the large size of the *murselect* trait can be an obstacle to recombination.

Considering that was not possible to confirm the *murselect* trait insertion, it was not possible to proceed for the GALG20*murselect* and GALG20*murselect Δlon* strains construction. As well, the impact of the marker free system implementation in GALG20 in plasmid production was not accessed.

3.6 Cultivation studies in shake flask

Gonçalves *et al*. have shown the benefits of GALG20 strain for plasmid production achieving the highest reported value for specific plasmid yield, for a constant temperature batch process. In a different study Mairhofer *et al.* suggested that *lon* gene mutation along with a marker free system using a minimized plasmid, were able to increase the plasmid DNA production in JM108*murselect* [4].

In this study, a *lon* gene mutation was performed in GALG20 in order to increase plasmid DNA production. After the construction of GALG20 and GAG20*Δlon* strains, cells were transformed with pVAX1-GFP plasmid and then grown in shake flask, to explore the potential of the *lon* gene knockout strain GALG20 for plasmid DNA production and cell growth rate, in comparison to GALG20.

Shake flask cultivations were performed in a batch fermentation strategy, which means that all nutrients are present at the time of inoculation and no nutrients were added during cultivation [2]. Thus during exponential phase all nutrients are in excess, which means that the specific growth rate will be the maximum specific growth rate, μ_{max}, as predicted by Monod kinetics [2], [80], [5]. As the nutrients will end up, the cells reach stationary phase.

The carbon source chosen to be used in this study was glucose, one of the most commonly used for cell cultivations [106]. It was already proved that in this carbon source, GALG20 achieves higher plasmid production yields in comparison to glycerol [79]. Furthermore, the *pgi* gene knockout redirects the carbon flux into PPP and reduces the acetate production, which was one of main disadvantages in using glucose.

Cells were cultivated in 50 mL of semi-defined media, supplemented with glucose and with kanamycin for selection of plasmid containing cells. The growth curve of GALG20 and GALG20Δ*lon* strains, in shake flask cultivations, during 24 hours are represented in **Figure 30**.

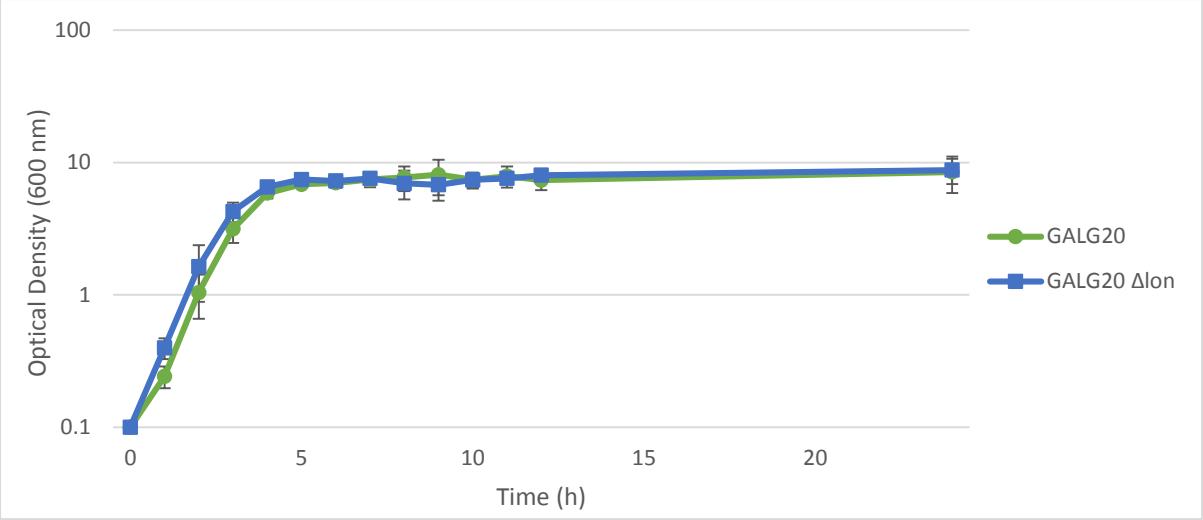

Figure 30 - Growth curve of GALG20 and GALG20Δ*lon*, in shake flask cultivations, during 24 hour, in semi-defined medium supplemented with 20g/L of glucose.

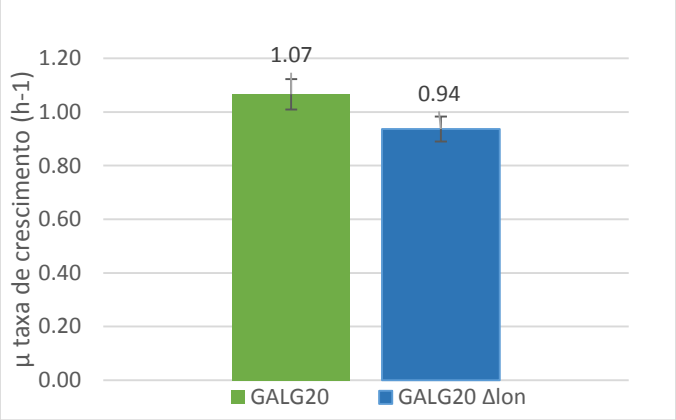

Figure 31 - Comparison between GALG20 and GALG20Δ*lon* growth rate, in shake flask cultivations in semi-defined medium supplemented with 20 g/L of glucose**.**

Figure 30 shows similar growth curves for GALG20 and GALG20Δ*lon*. In Figure 31 are represented the growth rates of both strains. The exponential phase occurs between 1 and 4 hours, and the growth rates were calculated based in the OD_{600nm} at that time points. GALG20 growth rate ($1.07 \pm 0.06\ h^{-1}$) is higher than GALG20Δ*lon* ($0.94 \pm 0.05\ h^{-1}$). The slight reduced growth rate in GALG20Δ*lon* in comparison with GALG20 can be explained by the adverse effects that a *lon* mutation can cause in cells. In situations of amino acid starvation, Lon protease of non-mutated cells, is stimulated to degrade free ribosomal proteins, in order to increase the amino acid pool that will be used for the synthesis of specific enzymes (e.g., amino acid biosynthetic enzymes) required for adaptation to this particular condition. This cellular response not only regenerates the amino acid pool but can also help to reduce the translational rate during starvation [89], [97]. In GALG20Δ*lon lon* mutation prevents the cellular response to amino acid starvation, making hard for the cells to adapt in the end of the exponential phase, when nutrients become scarce. In addition, Lon

protease is involved in several other cellular responses, so its absence may cause imbalance in some cellular pathways, delaying growth rate.

The growth rate achieved by GALG20 in this study (1.07 h^{-1}) is higher than previously reported. Gonçalves *et al.* reported a growth rate of 0.77 h^{-1}. Such difference was not expected, once the cultivation conditions were the same. The use of different lots or different brands of some products used in media composition, may have influenced growth rate. In addition, small differences in growth conditions may have a big impact in growth rate, like dissolved oxygen or pH that are difficult to control in shake flask cultivation [110].

3.6.1 Plasmid DNA production

In order to compare the pDNA production potencial of GALG20Δ*lon* with GALG20, at hour 12 of shake flask cultivation, samples were collected. As previously explained (Section 1.9.3) shake flask cultivation is usually the first cell culture strategy used in process development studies to predict pDNA production. However, this method normally achieves low yields, in comparison with other fermentation strategies, like fed batch fermentation [37]. Increasing the productivity of pDNA manufacturing processes requires the concomitant optimization of plasmid copy number (i.e., specific yield) and biomass concentration (i.e., volumetric yield) [2]. High specific plasmid yields are of great interest once it increases final purity and downstream purification efficiency, decreasing the manufacturing costs [37].

Collected samples were centrifuged for cell recovery and then purified using the High Pure Plasmid Isolation Kit (Roche ®) and their recommended protocol. After determination of the concentration using Nanodrop Spectrophotometer, volumetric and specific plasmid yields were estimated, and are represented in **Figure 32**.

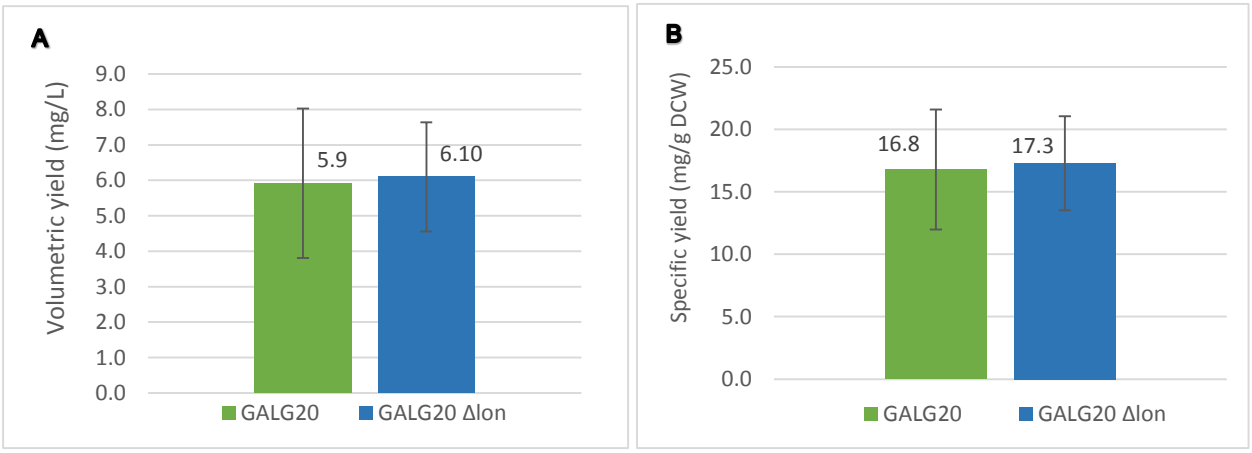

Figure 32 - Comparison of plasmid DNA yield between GALG20 and GALG20Δ*lon.* Samples were collected at 12 hours of shake flask cultivation at 37°C, in semi-defined media supplemented with 20 g/L of glucose. **A-** Plasmid DNA volumetric yield. **B-** Plasmid DNA specific yield.

The volumetric yield assessment shown that GALG20Δ*lon* exhibits a similar yield (17.3 ± 3.77 mg /L) to GALG20 (16.8 ± 4.79 mg/L). The same was verified in relation to the specific yield, once GALG20Δ*lon* displayed a yield of 6.1 ± 1.54 mg/g DCW, and GALG20 of 5.9 ± 2.11 mg/g DCW.

The expected result would be a higher plasmid DNA production in GALG20Δ*lon,* in comparison to GALG20 cells. Once it was expected that the uncharged-tRNAs in Lon deficient cells, that occur as consequence of amino acid starvation [89], [97], would interact with regulatory RNAs of the origin of replication of pDNA, influencing the mechanism of control of PCN. Thus leading to an increase in plasmid specific pDNA content and productivity [4].That was not verified, probably because the cellular response to amino acid starvation was not triggered, or at least not in a scale sufficient for the effects to be influencing the plasmid DNA yield.

Considering that shake flask experiments often fail to predict the outcome of pDNA production in bench-scale bioreactor [37], before concluding that the knockout of this gene does not increases plasmid DNA yields, more studies in large-scale using a fed-batch strategy should be performed. In addition, to fully evaluate the effect of *lon* gene knockout in plasmid production, could be useful grow GALG20Δ*lon* in feed-batch strategy adding just a reduced amount of amino acids, in order to trigger a cellular response to amino acid starvation, and increase uncharged tRNA in the cell. However, it would be necessary to evaluate the decrease in amino acids content in media, to ensure that the cells would still be viable and able to produce high quality, high yield pDNA.

Gonçalves *et al.* [79] reported a specific yield of 19.1 mg/g DCW, and a volumetric yield of 140.8 mg/L, in GALG20 strain, in media supplemented with 20 g/L of glucose. The imparity between the yields obtained by Gonçalves *et al.* and the ones obtained in this study are probably related to the method used for recovery and purification of the plasmid, and with the quantification method. In Gonçalves *et al.* study, pDNA was quantified from crude alkalyne lysates prepared from cell pellets (OD_{600}=10), and the resulting lysates were analyzed using a Gen-Pak FAX anion exchange column on an Agilent 1100 Series HPLC system. After purification pDNA eluted was detected at 260 nm with a diode array detector [79], while in the present study the purification and quantification was performed using Miniprep Kit from Roche ®, and NanoVue Plus Spectrophotometer (GE Healthcare ®), respectively. The first method uses the crude lysates which means that less plasmid is loss, leading to higher amounts of pDNA.

It was also hypothesized that the column used in Miniprep Kit become saturated, leading to plasmid losses. In order to test this possibility, samples with lower OD_{600} were purified. However, the values obtained were similar, so, probably, the reduced plasmid DNA yields were not related to column capacity.

In relation to the method used to read the plasmid concentration, the Nanodrop Spectrophotometer equipment is prone to measurements errors, what was verified by measuring some of the purified samples in a standard Spectrophotometer at $OD_{260\ nm}$. The values obtained were, between 1 or 2 times, superior to the ones obtained in Nanodrop equipment, which show that a different quantification strategy should be used to obtain more accurate results.

Some of the factors previously suggested to explain the different growth rate between the two studies, also apply to the plasmid production yields, like the lot-to-lot variations, and different brands of some of the compounds used in the cultivation medium. In addition, parameters like pH and dissolved oxygen are difficult to control in shake flask cultivations, which reduces the reproducibility of cultivations.

Despite the dissemblance between both studies, the results obtained in this study show consistency among them, which allows the comparison between GALG20 and GALG20Δ*lon.* More studies, in large scale, and testing different conditions should be performed with GALG20Δ*lon.* In addition, more accurate purification methods should be used, to evaluate the impact of *lon* gene mutation in plasmid production, and project a balance of the advantages and disadvantages in using a Lon protease deficient strain.

4 Conclusions and Future work

In this study a Lon-deficient GALG20 strain was created to assess the effect of *lon* gene knockout in plasmid production. The results obtained were compared with the GALG20 reference strain. It was verified that GALG20Δ*lon* growth rate (0.94 ± 0.05 h^{-1}) was slightly lower than the GALG20 (1.07 ± 0.06 h^{-1}). Considering the important roles played by Lon protease in several cellular responses [89], the decreased growth rate displayed by the *lon* gene mutated strain is in accordance with the expected results. Regarding plasmid DNA production by GALG20Δ*lon* strain, specific and volumetric plasmid DNA yield were determined. GALG20Δ*lon* shown similar specific (6.1 ± 1.54 mg/g DCW) and volumetric (17.3 ± 3.77 mg/L) yields in comparison to GALG20 (16.8 ± 4.79 mg/L), (5.9 ± 2.11 mg/g DCW). The expected result would be a higher plasmid DNA production in GALG20Δ*lon,* in comparison to GALG20 cells, resulting from the occurrence of uncharged-tRNAs in Lon deficient cells as consequence of amino acid starvation. That was not verified, probably because the cellular response to amino acid starvation was not triggered, or at least not in a scale sufficient for the effects to be influencing the plasmid DNA yield. These results were obtained in shake flask cultivations. Considering that shake flask experiments often fail to predict the outcome of pDNA production in bench-scale bioreactor [37], in the future would be pertinent to perform cultivations of GALG20Δ*lon* in large-scale using a fed-batch strategy. This way the effect of Lon deficiency in plasmid production would be more accurately measured, as well as the influences of this mutation in growth rate. Also could be of interest perform GALG20Δ*lon* cultivation experiments using a calculated amount of amino acids in a way that could trigger cellular response to amino acid starvation, without impairing cell viability. Lon protease is involved in heat shock response too [98], [99], so, experiments testing different temperatures could be of interest. Considering that the interaction between the uncharged tRNAs generated in GALG20Δ*lon* and the regulatory RNAs of the plasmid origin of replication, is not fully understood, more detailed studies in this field should be performed. Further experiments must be carried out to fully understand the potential of *lon* gene mutation in plasmid production.

The second goal of this study was the transference of the marker-free system developed by Mairhofer *et al*. [4] from JM108*murselect* strain into the high productivity strain GALG20, and the quantification of the resulting plasmid DNA yields, in order to verify if in GALG20 the system achieve higher yields than JM108*murselect*. However, this goal was not achieved despite all attempts. The step of insertion of *murselect* trait in GALG20 failed, and so it was not possible to assess the performance of the strain using the marker free system. Considering that the JM108*murselect* strain is *recA* minus, that could have been the reason for the unsuccessful insertion of the cassette procedure. In order to surpass this problem it could be of use repeat the procedure using a JM108*murselect* strain harboring a plasmid encoding *recA* gene, to ease the recombination step. Given that the marker free system developed by Mairhofer *et al.* [4], [12] shows so many advantages in comparison to the use of antibiotic resistance genes, and to the existing alternative systems, for plasmid containing cells selection, more efforts should be taken to implement this system in a high yield plasmid production strain combined with an optimized fermentation strategy. Despite the goal of

this project was not achieved, more work should be continued in order to surpass the difficulties in the insertion of the *murselect* trait in GALG20 in order to evaluate the full plasmid production potential.

Plasmid DNA vectors offer considerable benefits over viral systems in gene-based therapy and vaccination applications [127]. However, milligram scale of pDNA is required for clinical studies. In order to fulfill the demand, progress towards pDNA, rationally engineering *E. coli* host strains and vectors, as well as the development and optimization of fermentation strategies and medium design are required [71]. The development of wild-type genetic background strains designed specially to enhance pDNA production yield, like GALG20, holds great promising [71]. More studies pursuing less expensive and more effective media compositions are required, as well as new feeding strategies increasing plasmid DNA production. Improvements in vector design by reducing size and increasing genetic stability, safety and potency are critical. More studies destined to identify bottlenecks in plasmid DNA production should be carried out.

New methods combining high plasmid DNA yield with antibiotic-free maintenance systems are urgent. The interest in these systems has increased in last few years, as the disadvantages of antibiotic resistance genes were pointed out, and as the advantages of these systems were highlighted. However more studies, and more interest from companies and researchers to implement these emerging technologies are necessary.

The development of stoichiometric models of *E. coli* metabolism to determine its maximum theoretical plasmid-producing capacity, and to identify factors that significantly impact plasmid production, similar to the model created by Cunningham *et al.* [5], holds great promise.

Having in mind that engineered biological systems using only standardized biological components, to predict the outcome of biologically engineered unit operations, is far from developed, recent attempts to create and to implement modular, standardized biological components show the way for the future creation of highly predictable artificial biological systems. These synthetic biology frameworks promise to increase programmability and robustness in plasmid DNA production processes. The fast evolution in metabolic engineering and synthetic biology represents a high hope in strain engineering [85].

5 References

[1] C. P. Williams, James , Carnes, Aaron E , Hodgson, “Plasmid DNA Vaccine vector design: impact on efficacy, safety and upstream production,” *Biotechnol Adv.*, vol. 27, no. 4, pp. 353–370, 2010.

[2] A. Tejeda-Mansir and R. M. Montesinos, “Upstream processing of plasmid DNA for vaccine and gene therapy applications.,” *Recent Pat. Biotechnol.*, vol. 2, no. 3, pp. 156–72, Jan. 2008.

[3] E. Eastman and R. Durland, “Manufacturing and quality control of plasmid-based gene expression systems.,” *Adv. Drug Deliv. Rev.*, vol. 30, no. 1–3, pp. 33–48, Mar. 1998.

[4] J. Mairhofer, M. Cserjan-Puschmann, G. Striedner, K. Nöbauer, E. Razzazi-Fazeli, and R. Grabherr, “Marker-free plasmids for gene therapeutic applications--lack of antibiotic resistance gene substantially improves the manufacturing process.,” *J. Biotechnol.*, vol. 146, no. 3, pp. 130–7, Apr. 2010.

[5] D. S. Cunningham, R. R. Koepsel, M. M. Ataai, and M. M. Domach, “Factors affecting plasmid production in *Escherichia coli* from a resource allocation standpoint.,” *Microb. Cell Fact.*, vol. 8, p. 27, Jan. 2009.

[6] A. R. Lara and O. T. Ramírez, “Plasmid DNA Production for Therapeutic Applications,” in *Recombinant Gene Expression*, vol. 824, A. Lorence, Ed. Totowa, NJ: Humana Press, 2012, pp. 271–303.

[7] L. A. Palomares, S. Estrada-mondaca, and O. T. Ramírez, “Production of Recombinant Proteins Challenges and Solutions,” in *Methods in Molecular Biology*, vol. 267, Totowa: Humana Press Inc., 2004.

[8] A. Carattoli, “Plasmids in Gram negatives: molecular typing of resistance plasmids.,” *Int. J. Med. Microbiol.*, vol. 301, no. 8, pp. 654–8, Dec. 2011.

[9] M. a Kay, “State-of-the-art gene-based therapies: the road ahead.,” *Nat. Rev. Genet.*, vol. 12, no. 5, pp. 316–28, May 2011.

[10] M. a Liu, “DNA vaccines: an historical perspective and view to the future.,” *Immunol. Rev.*, vol. 239, no. 1, pp. 62–84, Jan. 2011.

[11] T. Niidome and L. Huang, “Gene therapy progress and prospects: nonviral vectors.,” *Gene Ther.*, vol. 9, no. 24, pp. 1647–52, Dec. 2002.

[12] J. Mairhofer and R. Grabherr, “Rational vector design for efficient non-viral gene delivery: challenges facing the use of plasmid DNA.,” *Mol. Biotechnol.*, vol. 39, no. 2, pp. 97–104, Jun. 2008.

[13] S. Ahmad, P. Sweeney, G. C. Sullivan, and M. Tangney, “DNA vaccination for prostate cancer , from preclinical to clinical trials - where we stand ?,” *Genet. Vaccines Ther.*, vol. 10, no. 9, pp. 1–9, 2012.

[14] I. Pfaffenzeller, J. Mairhofer, G. Striedner, K. Bayer, and R. Grabherr, “Using ColE1-derived RNA I for suppression of a bacterially encoded gene: implication for a novel plasmid addiction system.,” *Biotechnol. J.*, vol. 1, no. 6, pp. 675–81, Jun. 2006.

[15] S. Stenler, P. Blomberg, and E. Smith, “Safety and efficacy of DNA vaccines,” *Hum. Vaccin. Immunother.*, vol. 10, no. 5, pp. 1306–1308, 2014.

[16] E. G. Rodríguez, “Nonviral DNA vectors for immunization and therapy: design and methods for their obtention.,” *J. Mol. Med. (Berl).*, vol. 82, no. 8, pp. 500–9, Aug. 2004.

[17] D. M. Bower and K. L. J. Prather, “Development of new plasmid DNA vaccine vectors with R1-based replicons.,” *Microb. Cell Fact.*, vol. 11, no. 1, p. 107, Jan. 2012.

[18] B. Ferraro, M. P. Morrow, N. a Hutnick, T. H. Shin, C. E. Lucke, and D. B. Weiner, “Clinical applications of DNA vaccines: current progress.,” *Clin. Infect. Dis.*, vol. 53, no. 3, pp. 296–302, Aug. 2011.

[19] V. B. Pereira, M. Zurita-turk, T. Diniz, L. Saraiva, C. P. De Castro, B. M. Souza, P. M. Agresti, F. A. Lima, V. N. Pfeiffer, M. Santiago, P. Azevedo, C. S. Rocha, D. S. Pontes, V. Azevedo, and A. Miyoshi, “DNA Vaccines Approach : From Concepts to Applications,” *World J. Vaccines*, vol. 4, pp. 50–71, 2014.

[20] E. Signori, S. Iurescia, E. Massi, D. Fioretti, P. Chiarella, M. De Robertis, M. Rinaldi, G. Tonon, and V. M. Fazio, “DNA vaccination strategies for anti-tumour effective gene therapy protocols.,” *Cancer Immunol. Immunother.*, vol. 59, no. 10, pp. 1583–91, Oct. 2010.

[21] L. Li, F. Saade, and N. Petrovsky, “The future of human DNA vaccines.,” *J. Biotechnol.*, vol. 162, no. 2–3, pp. 171–82, Dec. 2012.

[22] H. Büning, “Gene therapy enters the pharma market: the short story of a long journey.,” *EMBO Mol. Med.*, vol. 5, no. 1, pp. 1–3, Jan. 2013.

[23] U.S Department of Health and Human Services, “FDA - U.S. Food and Drug Administraton.” [Online]. Available: http://www.fda.gov/default.htm. [Accessed: 21-Oct-2014].

[24] M. A. Liu, “DNA vaccines : a review,” *J. Intern. Med.*, vol. 253, pp. 402–410, 2003.

[25] M. a Liu and J. B. Ulmer, “Human clinical trials of plasmid DNA vaccines.,” *Adv. Genet.*, vol. 55, no. 05, pp. 25–40, Jan. 2005.

[26] M. a Kutzler and D. B. Weiner, “DNA vaccines: ready for prime time?,” *Nat. Rev. Genet.*, vol. 9, no. 10, pp. 776–88, Oct. 2008.

[27] “The United States Patent and Trademark Office.” [Online]. Available: http://www.uspto.gov. [Accessed: 15-Jan-2014].

[28] A. Ghanem, R. Healey, and F. G. Adly, “Current trends in separation of plasmid DNA vaccines: a review.,” *Anal. Chim. Acta*, vol. 760, pp. 1–15, Jan. 2013.

[29] R. Saha, S. Killian, and R. S. Donofrio, “DNA Vaccines : A Mini Review,” *Recent Pat. DNA Gene Seq.*, vol. 5, no. 2, pp. 1–5, 2011.

[30] R. P. Deering, S. Kommareddy, J. B. Ulmer, L. A. Brito, and A. J. Geall, “Nucleic acid vaccines : prospects for non-viral delivery of mRNA vaccines,” *Expert Opin. Drug Deliv.*, vol. 11, no. 6, pp. 885–899, 2014.

[31] “iaviReport - The publication on AIDS vaccine research.” [Online]. Available: http://www.iavireport.org/Pages/default.aspx. [Accessed: 20-Oct-2014].

[32] L. Frelin and M. Sällberg, “A small step closer to the Holy Grail of DNA vaccines: undisputed clinical benefit in humans.,” *Genome Med.*, vol. 1, no. 1, p. 15, Jan. 2009.

[33] “Gene Therapy Clinical Trials Worldwide.” [Online]. Available: www.wiley.com//legacy/wileychi/genmed/clinical/. [Accessed: 12-Oct-2014].

[34] P. H. Oliveira and J. Mairhofer, “Marker-free plasmids for biotechnological applications - implications and perspectives.,” *Trends Biotechnol.*, vol. 31, no. 9, pp. 539–47, Sep. 2013.

[35] D. M. Bower and K. L. J. Prather, “Engineering of bacterial strains and vectors for the production of plasmid DNA.,” *Appl. Microbiol. Biotechnol.*, vol. 82, no. 5, pp. 805–13, Apr. 2009.

[36] O. Tolmachov, “Designing Plasmid Vectors,” in *Methods in Molecular Biology*, vol. 542, W. Walther and U. S. Stein, Eds. Totowa, NJ: Humana Press, 2009, pp. 117–129.

[37] C. P. Williams, James A, Lukea, Jeremy, Langtry Sarah, Andersona, Sheryl , Hodgsona and A. E. Carnes, “Generic plasmid DNA production platform incorporating low metabolic burden seed-stock and fed-batch fermentation processes,” *Biotechnol Bioeng.*, vol. 103, no. 6, pp. 1129–1143, 2009.

[38] P. Hägg, J. W. de Pohl, F. Abdulkarim, and L. a Isaksson, “A host/plasmid system that is not dependent on antibiotics and antibiotic resistance genes for stable plasmid maintenance in Escherichia coli.,” *J. Biotechnol.*, vol. 111, no. 1, pp. 17–30, Jul. 2004.

[39] P. M. Bennett, “Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria.,” *Br. J. Pharmacol.*, vol. 153 Suppl, pp. S347–57, Mar. 2008.

[40] J. M. Luke, J. M. Vincent, S. X. Du, U. Gerdemann, a M. Leen, R. G. Whalen, C. P. Hodgson, and J. a Williams, “Improved antibiotic-free plasmid vector design by incorporation of transient expression enhancers.,” *Gene Ther.*, vol. 18, no. 4, pp. 334–43, May 2011.

[41] D. R. Gill, I. a Pringle, and S. C. Hyde, “Progress and prospects: the design and production of plasmid vectors.,” *Gene Ther.*, vol. 16, no. 2, pp. 165–71, Feb. 2009.

[42] M. Ingolotti, O. Kawalekar, D. J. Shedlock, K. Muthumani, and D. B. Weiner, “DNA vaccines for targeting bacterial infections,” *Expert Rev Vaccines.*, vol. 9, no. 7, pp. 747–763, 2010.

[43] K. Kumaragurubaran and K. Kaliaperumal, “DNA vaccine: the miniature miracle,” *Vet. World*, vol. 6, no. 3, p. 228, 2013.

[44] K. J. Prather, S. Sagar, J. Murphy, and M. Chartrain, “Industrial scale production of plasmid DNA for vaccine and gene therapy: plasmid design, production, and purification,” *Enzyme Microb. Technol.*, vol. 33, no. 7, pp. 865–883, Dec. 2003.

[45] F. Saade and N. Petrovsky, “Technologies for enhanced efficacy of DNA vaccines,” *Expert Rev. Vaccines*, vol. 11, no. 2, pp. 189–209, 2012.

[46] J. Williams, “Vector Design for Improved DNA Vaccine Efficacy, Safety and Production,” *Vaccines*, vol. 1, no. 3, pp. 225–249, Jun. 2013.

[47] S. a Plotkin, “Vaccines: past, present and future.,” *Nat. Med.*, vol. 11, no. 4 Suppl, pp. S5–11, Apr. 2005.

[48] P. H. Oliveira, K. J. Prather, D. M. F. Prazeres, and G. a Monteiro, “Structural instability of plasmid biopharmaceuticals: challenges and implications.,” *Trends Biotechnol.*, vol. 27, no. 9, pp. 503–11, Sep. 2009.

[49] C. Tang and G. A. Pietersz, “Intracellular detection and immune signaling pathways of DNA vaccines,” *Expert Rev. Vaccines*, vol. 8, no. 9, pp. 1161–1170, 2009.

[50] C. Gustafsson, J. Minshull, S. Govindarajan, J. Ness, A. Villalobos, M. Welch, O. B. Drive, A. Suite, and M. Park, “Engineering Genes for Predictable Protein Expression,” *Protein Expr Purif .*, vol. 83, no. 1, pp. 37–46, 2013.

[51] C. Elena, P. Ravasi, M. E. Castelli, S. Peirú, and H. G. Menzella, “Expression of codon optimized genes in microbial systems: current industrial applications and perspectives.,” *Front. Microbiol.*, vol. 5, no. February, p. 21, Jan. 2014.

[52] R. M. Cranenburgh, J. a Hanak, S. G. Williams, and D. J. Sherratt, “*Escherichia coli* strains that allow antibiotic-free plasmid selection and maintenance by repressor titration.,” *Nucleic Acids Res.*, vol. 29, no. 5, p. E26, Mar. 2001.

[53] L. M. R. El-Attar, S. Scott, S. Goh, and L. Good, “A pestivirus DNA vaccine based on a non-antibiotic resistance *Escherichia coli* essential gene marker.,” *Vaccine*, vol. 30, no. 9, pp. 1702–9, Feb. 2012.

[54] G. Vandermeulen, C. Marie, D. Scherman, and V. Préat, “New generation of plasmid backbones devoid of antibiotic resistance marker for gene therapy trials.,” *Mol. Ther.*, vol. 19, no. 11, pp. 1942–9, Nov. 2011.

[55] A. Reschner, S. Scohy, G. Vandermeulen, M. Daukandt, C. Jacques, B. Michel, H. Nauwynck, F. Xhonneux, V. Préat, A. Vanderplasschen, and C. Szpirer, “Use of Staby ® technology for development and production of DNA vaccines free of antibiotic resistance gene,” *Hum. Vaccin. Immunother.*, vol. 9, no. 10, pp. 2203–2210, 2013.

[56] J.-C. Galán, F. González-Candelas, J.-M. Rolain, and R. Cantón, “Antibiotics as selectors and accelerators of diversity in the mechanisms of resistance: from the resistome to genetic plasticity in the β-lactamases world.,” *Front. Microbiol.*, vol. 4, no. 9, pp. 1–17, Jan. 2013.

[57] R. N. Trivedi, P. Akhtar, J. Meade, P. Bartlow, M. M. Ataai, S. a Khan, and M. M. Domach, “High-Level Production of Plasmid DNA by *E. coli* DH5α/sacB by Introducing inc Mutations.,” *Appl. Environ. Microbiol.*, no. September, Sep. 2014.

[58] L. Vidal, J. Pinsach, G. Striedner, G. Caminal, and P. Ferrer, “Development of an antibiotic-free plasmid selection system based on glycine auxotrophy for recombinant protein overproduction in *Escherichia coli.*,” *J. Biotechnol.*, vol. 134, no. 1–2, pp. 127–36, Mar. 2008.

[59] H. S. Garmory, M. W. Leckenby, K. F. Griffin, S. J. Elvin, R. R. Taylor, M. G. Hartley, J. A. J. Hanak, E. D. Williamson, and R. M. Cranenburgh, “Antibiotic-Free Plasmid Stabilization by Operator-Repressor Titration for Vaccine Delivery by Using Live Salmonella enterica Serovar Typhimurium,” *Infect. Immun.*, vol. 73, no. 4, pp. 2005–2011, 2011.

[60] N. Nafissi, S. Alqawlaq, E. a Lee, M. Foldvari, P. a Spagnuolo, and R. a Slavcev, “DNA ministrings: highly safe and effective gene delivery vectors.,” *Mol. Ther. Nucleic Acids*, vol. 3, no. April, p. e165, Jan. 2014.

[61] C. Marie, G. Vandermeulen, M. Quiviger, M. Richard, V. Préat, and D. Scherman, “pFARs , Plasmids free of antibiotic resistance markers , display high-level transgene expression in muscle , skin and tumour cells,” *J. Gene Med.*, vol. 12, no. March, pp. 323–332, 2010.

[62] I. Peubez, N. Chaudet, C. Mignon, G. Hild, S. Husson, V. Courtois, K. De Luca, D. Speck, and R. Sodoyer, “Antibiotic-free selection in E. coli: new considerations for optimal design and improved production.,” *Microb. Cell Fact.*, vol. 9, p. 65, Jan. 2010.

[63] M. Wunderlich, H. Taymaz-Nikerel, G. Gosset, O. T. Ramírez, and A. R. Lara, “Effect of growth rate on plasmid DNA production and metabolic performance of engineered *Escherichia coli* strains.,” *J. Biosci. Bioeng.*, vol. 117, no. 3, pp. 336–42, Mar. 2014.

[64] N. Nafissi and R. Slavcev, “Construction and characterization of an in-vivo linear covalently closed DNA vector production system.,” *Microb. Cell Fact.*, vol. 11, no. 1, p. 154, Jan. 2012.

[65] G. a L. Gonçalves, K. L. J. Prather, G. a Monteiro, A. E. Carnes, and D. M. F. Prazeres, “Plasmid DNA production with *Escherichia coli* GALG20, a pgi-gene knockout strain: Fermentation strategies and impact on downstream processing.,” *J. Biotechnol.*, vol. 186, pp. 119–27, Sep. 2014.

[66] S. Y. Yau, E. Keshavarz-Moore, and J. Ward, “Host strain influences on supercoiled plasmid DNA production in *Escherichia coli*: Implications for efficient design of large-scale processes.,” *Biotechnol. Bioeng.*, vol. 101, no. 3, pp. 529–44, Oct. 2008.

[67] O. R. Hernández, M. L. Fernández, J. Valdes, M. P. Peña, and E. M. Díaz, “Scalable Technology to Produce Pharmaceutical Grade Plasmid DNA for Gene Therapy,” in *Gene Therapy - Developments and Future Perspectives*, InTech, 2005, pp. 53–70.

[68] A. Singer, M. a Eiteman, and E. Altman, “DNA plasmid production in different host strains of *Escherichia coli.*,” *J. Ind. Microbiol. Biotechnol.*, vol. 36, no. 4, pp. 521–30, Apr. 2009.

[69] K. Marisch, K. Bayer, T. Scharl, J. Mairhofer, P. M. Krempl, K. Hummel, E. Razzazi-Fazeli, and G. Striedner, “A comparative analysis of industrial *Escherichia coli* K-12 and B strains in high-glucose batch cultivations on process-, transcriptome- and proteome level.,” *PLoS One*, vol. 8, no. 8, p. e70516, Jan. 2013.

[70] M. K. Danquah and G. M. Forde, “Growth medium selection and its economic impact on plasmid DNA production.,” *J. Biosci. Bioeng.*, vol. 104, no. 6, pp. 490–7, Dec. 2007.

[71] G. a L. Gonçalves, D. M. Bower, D. M. F. Prazeres, G. a Monteiro, and K. L. J. Prather, “Rational engineering of *Escherichia coli* strains for plasmid biopharmaceutical manufacturing.,” *Biotechnol. J.*, vol. 7, no. 2, pp. 251–61, Feb. 2012.

[72] G. M. Borja, E. Meza Mora, B. Barrón, G. Gosset, O. T. Ramírez, and A. R. Lara, “Engineering *Escherichia coli* to increase plasmid DNA production in high cell-density cultivations in batch mode.,” *Microb. Cell Fact.*, vol. 11, no. 40, p. 132, Jan. 2012.

[73] J.-N. Phue, S. J. Lee, L. Trinh, and J. Shiloach, “Modified Escherichia coli B (BL21), a superior producer of plasmid DNA compared with *Escherichia coli* K (DH5alpha).,” *Biotechnol. Bioeng.*, vol. 101, no. 4, pp. 831–6, Nov. 2008.

[74] F. Silva, J. a Queiroz, and F. C. Domingues, “Evaluating metabolic stress and plasmid stability in plasmid DNA production by *Escherichia coli*.,” *Biotechnol. Adv.*, vol. 30, no. 3, pp. 691–708, 2012.

[75] J. Schorr, P. Moritz, and M. Schleef, “Production of Plasmid DNA in Industrial Quantities According to cGMP Guidelines,” in *Methods in Molecular Medicine*, vol. 29, Totowa: Humana Press Inc., 2000, pp. 11–21.

[76] K. Listner, L. Bentley, J. Okonkowski, C. Kistler, R. Wnek, a Caparoni, B. Junker, D. Robinson, P. Salmon, and M. Chartrain, “Development of a highly productive and scalable plasmid DNA production platform.,” *Biotechnol. Prog.*, vol. 22, no. 5, pp. 1335–45, 2006.

[77] N. Noreen, W. Y. Hooi, A. Baradaran, M. Rosfarizan, C. C. Sieo, M. I. Rosli, K. Yusoff, and A. R. Raha, “*Lactococcus lactis* M4, a potential host for the expression of heterologous proteins.,” *Microb. Cell Fact.*, vol. 10, no. 1, p. 28, Jan. 2011.

[78] T. E. Pablos, R. Soto, E. M. Mora, S. Le Borgne, O. T. Ramírez, G. Gosset, and A. R. Lara, “Enhanced production of plasmid DNA by engineered *Escherichia coli* strains.,” *J. Biotechnol.*, vol. 158, no. 4, pp. 211–4, Apr. 2012.

[79] G. a L. Gonçalves, D. M. F. Prazeres, G. a Monteiro, and K. L. J. Prather, “De novo creation of MG1655-derived *E. coli* strains specifically designed for plasmid DNA production.,” *Appl. Microbiol. Biotechnol.*, vol. 97, no. 2, pp. 611–20, Jan. 2013.

[80] A. E. Carnes and J. a Williams, “Plasmid DNA manufacturing technology.,” *Recent Pat. Biotechnol.*, vol. 1, no. 2, pp. 151–66, Jan. 2007.

[81] J. Mairhofer, T. Scharl, K. Marisch, M. Cserjan-Puschmann, and G. Striedner, “Comparative transcription profiling and in-depth characterization of plasmid-based and plasmid-free *Escherichia coli* expression systems under production conditions.,” *Appl. Environ. Microbiol.*, vol. 79, no. 12, pp. 3802–12, Jun. 2013.

[82] K. Bohle and a Ross, “Plasmid DNA production for pharmaceutical use: role of specific growth rate and impact on process design.,” *Biotechnol. Bioeng.*, vol. 108, no. 9, pp. 2099–106, Sep. 2011.

[83] D. S.-W. Ow, P. M. Nissom, R. Philp, S. K.-W. Oh, and M. G.-S. Yap, “Global transcriptional analysis of metabolic burden due to plasmid maintenance in *Escherichia coli* DH5α during batch fermentation,” *Enzyme Microb. Technol.*, vol. 39, no. 3, pp. 391–398, Jul. 2006.

[84] V. G. Yadav, M. De Mey, C. Giaw Lim, P. Kumaran Ajikumar, and G. Stephanopoulos, “The future of metabolic engineering and synthetic biology: Towards a systematic practice,” *Metab. Eng.*, vol. 14, no. 3, pp. 233–241, May 2012.

[85] E. Leonard, D. Nielsen, K. Solomon, and K. J. Prather, “Engineering microbes with synthetic biology frameworks.,” *Trends Biotechnol.*, vol. 26, no. 12, pp. 674–81, Dec. 2008.

[86] A. E. Carnes, J. M. Luke, J. M. Vincent, A. Schukar, S. Anderson, C. P. Hodgson, and J. a Williams, “Plasmid DNA fermentation strain and process-specific effects on vector yield, quality, and transgene expression.,” *Biotechnol. Bioeng.*, vol. 108, no. 2, pp. 354–63, Feb. 2011.

[87] K. E. J. Tyo, K. Kocharin, and J. Nielsen, “Toward design-based engineering of industrial microbes.,” *Curr. Opin. Microbiol.*, vol. 13, no. 3, pp. 255–62, Jun. 2010.

[88] “lon gene.” [Online]. Available: http://ecocyc.org/ECOLI/substring-search?type=NIL&object=lon&quickSearch=Quick+Search. [Accessed: 19-Sep-2013].

[89] V. Tsilibaris, G. Maenhaut-Michel, and L. Van Melderen, “Biological roles of the Lon ATP-dependent protease.,” *Res. Microbiol.*, vol. 157, no. 8, pp. 701–13, Oct. 2006.

[90] D. Vineyard, J. Patterson-Ward, and I. Lee, “Single-turnover kinetic experiments confirm the existence of high-and low-affinity ATPase sites in *Escherichia coli* Lon protease,” *Biochemistry*, vol. 45, no. 2, pp. 4602–4610, 2006.

[91] N. Chondrogianni, I. Petropoulos, S. Grimm, K. Georgila, B. Catalgol, B. Friguet, T. Grune, and E. S. Gonos, “Protein damage, repair and proteolysis.,” *Mol. Aspects Med.*, vol. 35, pp. 1–71, Feb. 2014.

[92] T. V Rotanova, I. Botos, E. E. Melnikov, F. Rasulova, A. Gustchina, and M. R. Maurizi, “Slicing a protease : Structural features of the ATP-dependent Lon proteases gleaned from investigations of isolated domains,” *Protein Sci.*, vol. 15, pp. 1815–1828, 2006.

[93] I. Lee and C. K. Suzuki, “Functional mechanics of the ATP-dependent Lon protease- lessons from endogenous protein and synthetic peptide substrates.,” *Biochim. Biophys. Acta*, vol. 1784, no. 5, pp. 727–35, May 2008.

[94] I. Lee, A. J. Berdis, and C. K. Suzuki, “Recent developments in the mechanistic enzymology of the ATP-dependent Lon protease from *Escherichia coli*: highlights from kinetic studies.,” *Mol. Biosyst.*, vol. 2, no. 10, pp. 477–83, Oct. 2006.

[95] J. Heuveling, A. Possling, and R. Hengge, “A role for Lon protease in the control of the acid resistance genes of *Escherichia coli,*” *Mol. Microbiol.*, vol. 69, no. 2, pp. 534–547, Jul. 2008.

[96] E. Gur, M. Vishkautzan, and R. T. Sauer, “Protein unfolding and degradation by the AAA+ Lon protease.,” *Protein Sci.*, vol. 21, no. 2, pp. 268–78, Feb. 2012.

[97] L. Van Melderen and A. Aertsen, “Regulation and quality control by Lon-dependent proteolysis,” *Res. Microbiol.*, vol. 160, no. 9, pp. 645–651, 2009.

[98] A. H. G. Which, D. T. Chin, S. A. Goff, T. Webstern, T. Smithn, and A. L. Goldberg, “Sequence of the *lon* Gene in *Escherichia coli*,” *J. Biol. Chem.*, vol. 263, no. 24, pp. 11718–11728, 1988.

[99] M. J. Smith, “Bacterial Protease Lon Is a Site-specific DNA-binding Protein,” *J. Biol. Chem.*, vol. 272, no. 1, pp. 534–538, Jan. 1997.

[100] H. Nicoloff and D. I. Andersson, “Lon protease inactivation, or translocation of the lon gene, potentiate bacterial evolution to antibiotic resistance.,” *Mol. Microbiol.*, vol. 90, no. 6, pp. 1233–48, Dec. 2013.

[101] M. L. Wohlever, T. a Baker, and R. T. Sauer, “Roles of the N domain of the AAA+ Lon protease in substrate recognition, allosteric regulation and chaperone activity.,” *Mol. Microbiol.*, vol. 91, no. 1, pp. 66–78, Jan. 2014.

[102] P. A. Shamlou, “Scaleable processes for the manufacture of therapeutic quantities of plasmid DNA.,” *Biotechnol. Appl. Biochem.*, vol. 37, no. Pt 3, pp. 207–18, Jun. 2003.

[103] F. C. Domingues, “Influence of Growth Conditions on Plasmid DNA Production,” *J. Microbiol. Biotechnol.*, vol. 19, no. December 2008, pp. 51–54, 2009.

[104] R. D. O’Kennedy, J. M. Ward, and E. Keshavarz-Moore, “Effects of fermentation strategy on the characteristics of plasmid DNA production.,” *Biotechnol. Appl. Biochem.*, vol. 37, no. Pt 1, pp. 83–90, Feb. 2003.

[105] Z.-N. Xu, W.-H. Shen, H. Chen, and P.-L. Cen, “Effects of medium composition on the production of plasmid DNA vector potentially for human gene therapy.,” *J. Zhejiang Univ. Sci. B*, vol. 6, no. 5, pp. 396–400, May 2005.

[106] L. G. Fuentes, A. R. Lara, L. M. Martínez, O. T. Ramírez, A. Martínez, F. Bolívar, and G. Gosset, “Modification of glucose import capacity in *Escherichia coli*: physiologic consequences and utility for improving DNA vaccine production.,” *Microb. Cell Fact.*, vol. 12, p. 42, Jan. 2013.

[107] A. E. Carnes, C. P. Hodgson, and J. a Williams, “Inducible *Escherichia coli* fermentation for increased plasmid DNA production.,” *Biotechnol. Appl. Biochem.*, vol. 45, no. Pt 3, pp. 155–66, Nov. 2006.

[108] R. D. O’Kennedy, C. Baldwin, and E. Keshavarz-Moore, “Effects of growth medium selection on plasmid DNA production and initial processing steps.,” *J. Biotechnol.*, vol. 76, no. 2–3, pp. 175–83, Jan. 2000.

[109] G. a L. Gonçalves, K. L. J. Prather, G. a Monteiro, and D. M. F. Prazeres, “Engineering of *Escherichia coli* strains for plasmid biopharmaceutical production: scale-up challenges.,” *Vaccine*, vol. 32, no. 24, pp. 2847–50, May 2014.

[110] F. Garcia-Ochoa and E. Gomez, “Bioreactor scale-up and oxygen transfer rate in microbial processes: an overview.,” *Biotechnol. Adv.*, vol. 27, no. 2, pp. 153–76, 2009.

[111] G. N. Ferreira, G. a Monteiro, D. M. Prazeres, and J. M. Cabral, “Downstream processing of plasmid DNA for gene therapy and DNA vaccine applications.,” *Trends Biotechnol.*, vol. 18, no. 9, pp. 380–8, Sep. 2000.

[112] D. M. F. Prazeres, G. N. M. Ferreira, G. A. Monteiro, C. L. Cooney, and J. M. S. Cabral, “Large-scale production of pharmaceutical-grade plasmid DNA for gene therapy : problems and bottlenecks,” *TIBTECH*, vol. 17, no. April, pp. 169–174, 1999.

[113] J. C. Nunes, A. M. Morão, C. Nunes, M. T. Pessoa de Amorim, I. C. Escobar, and J. a. Queiroz, “Plasmid DNA recovery from fermentation broths by a combined process of micro- and ultrafiltration: Modeling and application,” *J. Memb. Sci.*, vol. 415–416, pp. 24–35, Oct. 2012.

[114] H. C. B. Doly, J., “A rapid alkaline extraction procedure for screening recombinant plasmid DNA,” *Nucleic Acids Res.*, vol. 7, no. 6, 1979.

[115] J. Mairhofer, I. Pfaffenzeller, D. Merz, and R. Grabherr, “A novel antibiotic free plasmid selection system: advances in safe and efficient DNA therapy.,” *Biotechnol. J.*, vol. 3, no. 1, pp. 83–9, Jan. 2008.

[116] “murA gene - UDP-N-acetylglucosamine enolpyruvoyl transferase.” [Online]. Available: http://ecocyc.org/ECOLI/NEW-IMAGE?type=GENE&object=EG11358&redirect=T. [Accessed: 20-Jan-2014].

[117] “lon gene- DNA-binding, ATP-dependent protease La.” [Online]. Available: http://ecocyc.org/ECOLI/NEW-IMAGE?type=GENE&object=EG10542&redirect=T. [Accessed: 20-Jan-2014].

[118] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. a Datsenko, M. Tomita, B. L. Wanner, and H. Mori, “Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection.,” *Mol. Syst. Biol.*, vol. 2, p. 2006.0008, Jan. 2006.

[119] Invitrogen, “DH5 alpha.” [Online]. Available: http://www.lifetechnologies.com/pt/en/home/brands/invitrogen.html. [Accessed: 20-Oct-2014].

[120] “Template plasmid pKD13, complete sequence.” [Online]. Available: http://www.ncbi.nlm.nih.gov/nuccore/15554335. [Accessed: 16-Sep-2014].

[121] “Red recombinase plasmid pKD46, complete sequence.” [Online]. Available: http://www.ncbi.nlm.nih.gov/nuccore/AY048746.1. [Accessed: 30-Oct-2013].

[122] K. a Datsenko and B. L. Wanner, “One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 12, pp. 6640–5, Jun. 2000.

[123] “Template plasmid pKD13, complete sequence.” [Online]. Available: http://www.ncbi.nlm.nih.gov/nuccore/15554335. [Accessed: 30-Oct-2013].

[124] “pVAX1 Vector.” [Online]. Available: in order to minimize the possibility of chromosomal integration. [Accessed: 27-Sep-2014].

[125] A. R. Azzoni, S. C. Ribeiro, G. A. Monteiro, and D. M. F. Prazeres, “The impact of polyadenylation signals on plasmid nuclease-resistance and transgene expression,” no. April, pp. 392–402, 2007.

[126] G. a L. Gonçalves, P. H. Oliveira, A. G. Gomes, K. L. J. Prather, L. a Lewis, D. M. F. Prazeres, and G. a Monteiro, “Evidence that the insertion events of IS2 transposition are biased towards abrupt compositional shifts in target DNA and modulated by a diverse set of culture parameters.,” *Appl. Microbiol. Biotechnol.*, vol. 98, no. 15, pp. 6609–19, Aug. 2014.

[127] L. Raiado-Pereira, J. De La Vega, D. M. F. Prazeres, and M. Mateus, “Development of a phenyl membrane chromatography-based process yielding pharmaceutical grade plasmid deoxyribonucleic acid for mammalian cells transfection.,” *J. Chromatogr. A*, vol. 1337, pp. 67–74, Apr. 2014.

[128] H. Potter and R. Heller, “Transfection by electroporation,” *Curr Protoc Mol Biol.*, no. 813, 2003.

[129] W. J. Dower, J. F. Miller, C. W. Ragsdale, M. B. Group, B. Laboratories, and S. Medical, “High efficiency transformation of E.coli by high voltage electroporation,” vol. 16, no. 13, pp. 6127–6146, 1988.

[130] J. A. Sawitzke, L. C. Thomason, M. Bubunenko, X. Li, and D. L. Court, “Recombineering : Using Drug Cassettes to Knock out Genes in vivo,” *Methods Enzym.*, vol. 533, pp. 79–102, 2013.

[131] “IDT DNA.” [Online]. Available: http://eu.idtdna.com/site. [Accessed: 04-Nov-2014].

[132] “SIGMA-ALDRICH.” [Online]. Available: http://www.sigmaaldrich.com/portugal.html. [Accessed: 04-Nov-2013].

[133] L. C. Thomason, N. Costantino, and D. L. Court, “E. coli genome manipulation by P1 transduction.,” *Curr. Protoc. Mol. Biol.*, vol. Chapter 1, no. July, p. Unit 1.17, Jul. 2007.

[134] N. Sternberg and R. Maurer, “Bacteriophage-Mediated Generalized Transduction in *Escherichia coli* and Salmonella typhimurium,” *METHODS Enzymol.*, vol. 204, pp. 18–43, 1991.

[135] M. B. Lobocka, D. J. Rose, G. P. Iii, M. Rusin, A. Samojedny, H. Lehnherr, M. B. Yarmolinsky, and F. R. Blattner, “Genome of Bacteriophage P1,” *J. Bacteriol.*, vol. 186, no. 21, pp. 7032–7068, 2004.

[136] H. Huang and M. Masters, “Bacteriophage P1 pac sites inserted into the chromosome greatly increase packaging and transduction of *Escherichia coli* genomic DNA.,” *Virology*, vol. 468–470C, pp. 274–282, Sep. 2014.

[137] M. J. Donath, M. a Dominguez, and S. T. Withers, “Development of an automated platform for high-throughput P1-phage transduction of *Escherichia coli*.,” *J. Lab. Autom.*, vol. 16, no. 2, pp. 141–7, Apr. 2011.

[138] L. R. Zeph, M. a Onaga, and G. Stotzky, “Transduction of *Escherichia coli* by bacteriophage P1 in soil.,” *Appl. Environ. Microbiol.*, vol. 54, no. 7, pp. 1731–7, Jul. 1988.