UNIVERSIDADE DA BEIRA INTERIOR
Ciências

Development of a factorial design for a therapeutic plasmid DNA biosynthesis

Luís Miguel Candeias Martins

Dissertação para obtenção do Grau de Mestre em

Biotecnologia

(2º ciclo de estudos)

Orientador: Prof. Doutor Luís António Paulino Passarinha

Co-orientador: Profª. Doutora Fani Pereira de Sousa

Covilhã, outubro de 2013

Acknowledgments

I would to thank to Professor Luís Passarinha and Professor Fani Sousa for the best supervision and guidance during this last year. It was because of them efforts that this work went through. They were always available to help and teach me whenever I needed to.

A very special thank to Professor Guillermo Gosset. It was because of his availability and help that part of this project went forward.

I would like express a special thank to Augusto Pedro, David Opolzer, Patrícia Pereira and Doctor Ângela Sousa for all help and availability in my work. They were part of this.

To the rest of the friends and colleagues in the Investigation Centre I thank the support and the nicest coffee-breaks ever.

To all my family, thanks for never give up on believing me, and for all your kindness even when you can't understand what I am doing so many hours in the lab. Love you all.

To my lovely sister, because I could not be around when you need me the most, thanks for understand.

Last but not least, to Patrícia Lindeza, for her help, love and beliefs.

Resumo

As aplicações terapêuticas de DNA plasmídico (pDNA) avançaram significativamente nos últimos anos. No mercado atual é possível verificar a existência de vários compostos terapêuticos que têm como base o pDNA, estando muitos outros a ser testados nas fases 2 e 3 de ensaios clínicos. A necessidade atual e futura de pDNA implica o desenvolvimento de bioprocessos cada vez mais eficientes para a sua produção. Geralmente o pDNA é produzido por culturas de *Escherichia coli*.

Foi previamente demonstrado que culturas específicas de *E. coli* que possuam um sistema de transporte do substrato alterado, que não sobrecarregue o metabolismo, são capazes de atingir altas densidades celulares. No entanto, as grandes quantidades de oxigénio exigido podem levar a condições microaeróbicas após algumas horas de cultivo, mesmo que em pequena escala. Normalmente, os problemas inerentes a estas culturas são a elevada necessidade de oxigénio e a acumulação de acetato, um subproduto metabólico que é sintetizado aerobicamente quando a taxa de captação de glicose ultrapassa um determinado limite. Nos últimos anos, várias pesquisas têm tido como foco o estudo da indução de DNA plasmídico bem como as estratégias de fermentação utilizando meios semi-definidos. Estes estudos conceberam resultados relevantes que nos permitam projetar uma plataforma de produção de DNA plasmídico otimizada. O objetivo principal deste estudo é melhorar o rendimento de produção de DNA plasmídico com potencial terapêutico através da cultura de uma estirpe de *Escherichia coli* recentemente desenvolvida.

A estratégia baseia-se na variação da composição dos meios de fermentação em termos de nutrientes e no desenvolvimento de um desenho experimental direcionado à via metabólica dos aminoácidos aromáticos.

A monitorização através de métodos analíticos válidos constitui uma vantagem, tanto no controlo como na otimização, dos intervenientes do processo fermentativo. A alteração da composição e concentração dos substratos utilizados afeta o crescimento de *Escherichia coli* VH33, e em simultâneo, a quantidade e qualidade do plasmídio em estudo. O modelo experimental desenvolvido confirma que a perturbação da via metabólica dos aminoácidos aromáticos influencia a produção DNA plasmídio.

Palavras-Chave

DNA plasmídico, *Escherichia coli* VH33, Box-Benhken design, Desenho Experimental.

Resumo Alargado

No início da década de 70, o aparecimento da Engenharia Genética foi uma consequência natural da grande quantidade de conhecimentos que se acumularam na área da Biologia Molecular, envolvendo principalmente conhecimentos sobre bactérias e vírus. No dias de hoje, o Homem pode intervir diretamente na programação genética dos microrganismos, não apenas para superprodução de um metabolito, mas ainda, de biomoléculas de interesse terapêutico como o DNA plasmídico.

O DNA plasmídico ou vetor define-se como uma molécula circular de DNA em dupla cadeia que ocorre naturalmente em bactérias. Os plasmídeos codificam genes que não são essenciais para o crescimento bacteriano, tornando-os vantajosos para a introdução de genes exógenos através da Tecnologia Recombinante. Estes genes exógenos contêm sequências corretas de DNA que quando transcritas e traduzidas nas células alvo culminam na ação terapêutica necessária. Atualmente, com o avanço da ciência e tecnologia os DNA plasmídicos são considerados uma biomolécula terapêutica fundamental à Vacinação e Terapia Génica. Os primeiros casos de sucesso no tratamento de doenças genéticas desencadearam um forte interesse da comunidade científica. Como alvo principal do uso desta tecnologia está o tratamento do cancro seguido das doenças monogénicas. No entanto, os dados estatísticos mostram que a grande maioria dos ensaios terapêuticos com DNA plasmídicos encontram-se na fase I. Este facto traduz a exigência e eficácia necessária para que este tipo tratamento seja aplicado na terapia de doenças humanas.

Para o sucesso deste tipo de terapia depende também a formulação que é feita para a entrega do DNA plasmídico nas células alvo. O vetor necessita de entrar na célula, ter capacidade de chegar ao núcleo de forma a transcrever a sua informação. Durante este "pequeno" percurso a biomolécula tem que ultrapassar todas as barreiras e superar o ambiente hostil sem qualquer dano no DNA plasmídico.

De acordo com a literatura, o alvo principal da Terapia Génica é o cancro, tendo como base o gene supressor de tumor p53 que conduz à apoptose de células cancerígenas. Quando por qualquer motivo ocorre uma mutação neste gene é necessário restaurar a sua correta informação genética.

A utilização de DNA plasmídico requer a sua produção e amplificação, para obtenção das quantidades necessárias para testes *in-vitro*, *in-vivo* e ensaios clínicos. A produção recombinante de DNA plasmídico tem em conta vários fatores, como o tipo de vetor, o sistema hospedeiro, as condições de fermentação e processo de recuperação. Apenas com a combinação destes fatores é possível obter um processo viável e efetivo para a produção deste tipo de biomoléculas terapêuticas.

Recentemente, a comunidade científica tem estado atenta ao processo de produção de DNA plasmídico por fermentação. Para que este processo seja possível, devem ter-se como base diversos estudos científicos que culminaram no aumento da qualidade e quantidade de DNA plasmídico produzido por esta metodologia.

As novas estratégias para a produção de vetores baseiam-se principalmente na otimização dos plasmídios, da célula hospedeira e dos parâmetros de fermentação. No caso da otimização dos plasmídeos os objetivos são direcionados para a diminuição do tamanho do vetor uma vez que facilita a transfeção da célula alvo, mas também manter um nível elevado de número de cópias no sistema hospedeiro. Relativamente à célula hospedeira a comunidade científica esforça-se por melhorar as vias metabólicas que canalizam o substrato para produção de produtos alvos. Focando esta abordagem é possível incrementar os rendimentos plasmídio-substrato consumido mas principalmente diminuir drasticamente a presença de produtos tóxicos tanto para a célula como para o plasmídeo. No caso dos parâmetros de fermentação, a seleção e otimização dos nutrientes do meio é um dos fatores mais relevantes, resultando no aumento das potencialidades de produção do sistema hospedeiro.

O trabalho realizado nesta dissertação vem ajudar a caminhar no sentido do aumento de produção de DNA plasmídico, tendo em conta os fatores acima descritos. A utilização de uma nova geração de estirpes de *E. coli* com a potencialidade de captar e metabolizar mais facilmente a glucose, permite direcionar os sub-produtos para a produção de DNA plasmídico. A variação da composição de nutrientes do meio possibilita compreender como é que esta bactéria se adapta e responde a estímulos. Esta estirpe revelou ser capaz de crescer e aumentar a produção do DNA plasmídico de 27 µg/mL para 54 µg/mL em concentrações de elevadas de glucose, como 30 g/L. Adicionalmente, os resultados constataram que o extrato de levedura deve ser preferecialmente usado em detrimento da triptona como fonte de azoto, obtendo-se concentrações superiores de biomassa e plasmídico, 2.3 g/L e 61 µg/mL respetivamente.

A avaliação da produção de plasmídeo tem um carácter importante, mas bastante relevante é a percentagem de pureza em termos da produção de ácidos nucleicos. Os dados revelaram que a substituição da fonte de carbono de glicerol por glucose no meio "Terrific Broth" consegue obter os mesmos efeitos na produtividade e pureza do plasmídio. Este trabalho ajudou a entender como a via dos aminoácidos aromáticos poderá influenciar a biossíntese do DNA plasmídico. Para isto, foi pertinente recorrer ao desenho experimental, especificamente o Box-Behnken design com uma resposta de superfície, permitindo avaliar a interação que ocorre entre os três aminoácidos aromáticos e determinar qual a concentração ótima na produção de DNA plasmídico. Esta abordagem permitiu criar um modelo com significância estatística, demonstrando quais são os parâmetros significativos e ao mesmo tempo fornecer as concentrações dos fatores que permitem alcançar os valores ótimos na produção do plasmídio em estudo. É importante de destacar o desenvolvimento de um método analítico de

acordo com os critérios internacionalmente aceites, da ***Food and Drug Administration*** (FDA) para a monitorização dos aminoácidos aromáticos, glucose e acetato (inibidor do crescimento da *E. coli*) foi totalmente validado com coeficientes de determinação (R^2) superiores a 0.99 para todos os compostos referidos.

Abstract

Therapeutic applications of plasmid DNA (pDNA) have significantly advanced during the last years. Currently, several pDNA-based drugs are already in the market, whereas several others have entered to phases 2 and 3 of clinical trials. The present and future demand for pDNA requires the development of efficient bioprocesses to produce it. Commonly, pDNA is produced by cultures of *Escherichia coli*. It has been previously demonstrated that specific strains of *E. coli* with a modified substrate transport system can be able to attain high cell densities in batch mode, due to the very low overflow metabolism displayed. However, the large amounts of oxygen demanded can lead to microaerobic conditions after some hours of cultivation, even at small-scale. Typically, the inherent problems for these cultures are the high oxygen demand and the accumulation of acetate, a metabolic by-product that is synthesized aerobically when the glucose rate exceeds limits.

In recent years, various researches have been focused on the study of induction of plasmid DNA as well as strategies for fermentation using semi-defined mediums. These studies conceived relevant results that allow us to design a production platform for enhanced plasmid DNA. The main goal of this study is to optimize the yield of therapeutic plasmid DNA by culture of recent developed strain of *Escherichia coli*. The strategy is based on the variation of composition of the fermentation media in terms of nutrients and by a development of an experimental design directed to aromatic amino acids pathway.

Monitoring through analytical methods is an advantage both in control and optimization of the intervenients in fermentation process. The change of composition and concentration of the substrates affect the growth of *Escherichia coli* VH33 and simultaneously, the quantity and quality of the plasmid in study. The experimental model confirms that the disturbing on the aromatic amino acid pathway influences the production plasmid DNA.

Keywords

Plasmid DNA, *Escherichia coli* VH33, Box-Benhken design, Experimental design.

Index

Index of figures

Index of tables

List of abbreviations

ATP	Adenosine Triphosphate
BB	Bacterial Backbone
BP	Pair bases
CCC	Covalently Closed Circular
CDM	Cell Dry Mass
CHO	Chorismate Pathway
CV	Coefficient of Variation
DNA	Deoxyribonucleic Acid
E. coli	*Escherichia coli*
E4P	Erythrose-4-phosphate
EC	Expression Cassette
EDTA	Ethylene-diamine tetraacetic acid
ELISA	Enzyme-Linked Immunosorbent Assay
FC	Flow Cytometry Systems
FDA	Food and Drug Administration
G6P	Glucose 6-phosphate
GalP	Galactose Permease
GC	Gas Chromatography
gDNA	Genomic Deoxyribonucleic Acid
HPLC	High Performance Liquid Chromatography
ICH	International Conference on Harmonization
IS	Insertion Sequence

Kb	Kilo bases
L	Linear forms
LLOQ	Lower Limit of Quantification
mRNA	Messenger ribonucleic acid
MS	Mass Spectrometry
OC	Open circular
OD_{600}	Optical Density at 600nm
OPA	O-phtalaldehyde Reagent
OPA-S	O phthalaldehyde-sodium Sulfite
ORI	Bacterial Origin of Replication
pDNA	Plasmid Deoxyribonucleic Acid
PEP	Phosphoenolpyruvate
PHE	Phenylalanine
Poly A	Polyadenylation Sequence
PPP	Pentose Phosphate Pathway
PTS	Phosphoenolpyruvate Carbohydrate Phosphotransferase System
PTS	Sugar Phosphotransferase System
PYR	Pyruvate
R5P	Precursors ribose-5-phosphate
RID	Refractive Index Detector
RNA	Ribonucleic acid
RSM	Response Surface Methodology
RT-PCR	Real Time Polymerase Chain Reaction
SC	Supercoiled

SDS	Sodium Dodecylsulfate Solution
SHIK	Shikimate Pathway
TAE	Tris–Acetate–EthyleneDiamine
TB	Terrific Broth
TCA	Tricarboxylic Acid Cycle
TRP	Tryptophan
TYR	Tyrosine
UV	Ultraviolet Light

Justifications and Objectives

The plasmid DNA production is becoming increasingly important as therapeutic approach make their way into clinical trials and eventually into the pharmaceutical product. The numerous clinical trials for plasmid DNA products have demonstrated the safety of the DNA vaccination method and indicate the potential of this relatively new field of therapeutics. This powerful bioproduct has become a viable option to treatment of cancer, as well as for the gene therapy and even for bacterial and viral diseases. Thus, research community have focused on the development integrate process between the upstream and downstream processing. However, the quality of final product is ultimately determined by fermentation strategy.

For these reasons, the biomass yield, plasmid yield and plasmid quality improvement can be reached through optimization of the growth environment of plasmid-producing organism.

Therefore, the objective of this work was to optimize the yield of therapeutic plasmid DNA by culture of *Escherichia coli* VH33, based on the variation of nutritional substrates, by a development of an experimental design directed to aromatic amino acids pathway and through establishment an analytical method for metabolites quantification.

Chapter 1-Revision of Literature

1.1. Definition of Plasmid

A plasmid is defined as a double stranded, circular DNA molecule capable of autonomous replication. Plasmids have systems which guarantee their autonomous replication also controlling the copy number and ensuring stable inheritance during cell division, however they do not carry genes essential for the growth of host cells [1]. Accordingly, plasmids potential assents in the correction of genetic defects and prevention or treatment of several diseases [2].

1.2. Vaccination and Gene therapy

For several years, bacterial plasmid DNA (pDNA) has been used as a biological tool for cloning and expression of recombinant proteins. Currently, has been considered as a potential biopharmaceutical, mostly for applications in gene therapy and DNA vaccination [3][4]. Gene-based therapeutics occurs with introduction of nucleic acids into cells with the intention of altering gene expression to prevent, halt or reverse a pathological process [5].

At the beginning of present century, the optimism of the gene therapy research community was encouraged by the first report of successful treatment of genetic diseases. As example, the majority of the clinical trials (81.5%) in gene therapy have addressed cancer, cardiovascular disease and inherited monogenic diseases, including lung, gynaecological, skin, urological, neurological and gastrointestinal tumours, according to figure 1 [6].

The DNA vaccination treatment based in gene therapy can lead to a strong and long-lasting immune response through the inoculation of a plasmid containing a interesting gene, which is subsequently expressed by the cellular machinery of the person receiving the DNA vaccine [7].

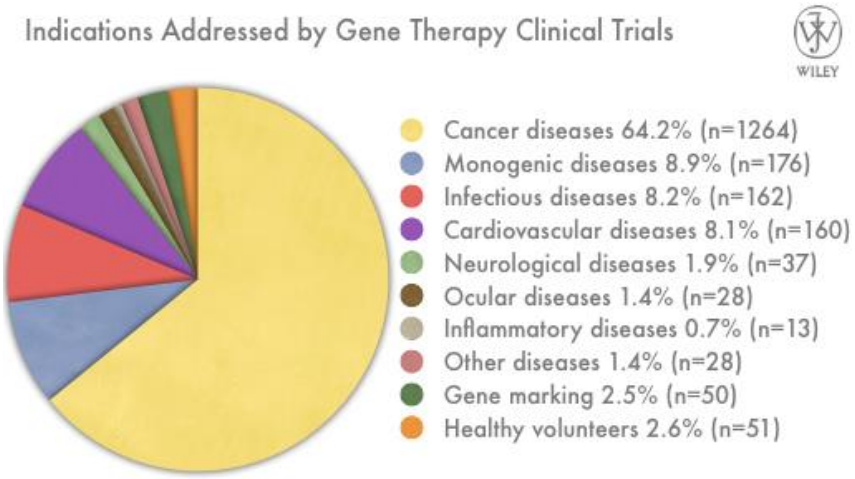

Figure 1- Indications addressed by gene therapy clinical trials (adapted from [6]).

DNA vaccines and gene therapy have been extensively evaluated in humans as recently addressed by Kutzler and co-workers. At these moment, there are, 72 products at Phase I, 20 at Phase II and 2 at Phase III human trials according table 1 [8].

Table 1- Number of open trials in Phase I, Phase II, Phase III (adapted from [8]).

Number of open trials	Category	Diseases and conditions treated
Phase I		
54	Cancer	Melanoma, glioblastoma, lymphoma, acute lymphoblastic leukemia,, and colorectal, prostate, pancreatic, lung and breast cancer
6	Cardiovascular	Re-endothelialization
5	Healthy volunteers	HIV vaccine safety
2	Infectious disease	Chonic hepatitis B and HIV-1
1	Neurological	Effect of human insulin-like growth factor for cubital tunnel syndrome
1	Ocular	Retinitis pigmentosa
3	Other	Erectile dysfunction, type I diabetes mellitus
Phase II		
10	Cancer	Melanoma, angioendothelioma, lung cancer and pancreatic adenocarcinoma
4	Cardiovascular	Peripheral ischaemic ulcers, limb ischaemia
3	Neurological	Relapsing remitting multiple sclerosis, diabetic neuropathy
3	ocular	Atrophic macular degeneration
Phase III		
1	Cancer	Lung cancer
1	Cardiovascular	Limb ischaemia

The improvement in gene transfer into cells and/or tissues *in vivo* has resulted largely from advances in vector technologies that include refined vector systems. Nevertheless, not all of the problems and limitations have been solved, unexpected barriers have emerged from clinical trials in course. During the test it was verified that the main barriers (figure 2) occurred on uptake, transport and uncoating; vector genome persistence; transcriptional activity and immune response [5].

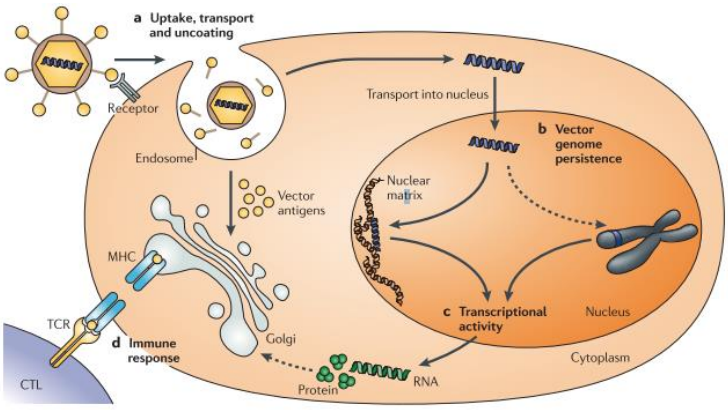

Figure 2- The four barriers to successful gene therapy. a) Vector bind to a cellular membrane and are internalized; b) The vector reaches the nucleus, where it can be further processed; c) Transcriptional activity; d) Immune response can limit the viability (adapted from [5]).

In brief, the distribution of the vector after administration is influenced by many parameters, which are the vascular supply and endothelial barriers to a particular organ, vector size and interactions between the vector ligand and the host cell receptor [5]. In order to improve gene delivery efficiency, it is desirable to understand how involucre-DNA complexes overcome the intracellular barriers. By understanding the preferred pathways and cellular mechanisms, it is possible to design more efficient vectors and treatment techniques. In particular, a complex must be brought to the surface of the target cells either by nonspecific electrostatic interactions or by specific binding to cellular membrane proteins. The complex is then internalized by the cell, usually by an endocytic mechanism such as receptor-mediated endocytosis or macropinocytosis. The involucre-DNA complexes must be subsequently transported from the cell periphery to the perinuclear region, it must escape its endocytic vesicle, and it must gain nuclear entry, all without having the therapeutic gene damaged by enzymes or chemically harsh environments. The plasmid must also dissociate from the involucre to allow for transcription [9].

1.3. Role of p53 pathway in cancer

The p53 protein is a tumor suppressor, allowing the selective induction of growth arrest and apoptosis due oncogenic or damage signalling. This protein has a crucial role against malignant cell transformation [10]. Majority of cancers have dysfunctional p53, either through mutation of the p53 gene itself or alterations in factors that modulate p53's stability and activity, like described in figure 3 [11]. The most important strategies to combat cancer are based in reactivation of p53 or related pathways in tumor cells [12]. Tumor suppressor p53 is a transcription factor that regulates a large number of genes and guards against genomic instability. Under multiple cellular stress conditions, p53 functions to block cell cycle progression transiently unless proper DNA repair occurs. Failure of DNA repair mechanisms leads to p53-mediated induction of cell death programs or to permanent cell cycle arrest known as cellular senescence. During neoplastic progression, p53 is often mutated and fails to efficiently perform these functions. Therefore, strategies to reactivate p53 provide an attractive approach for blocking tumor pathogenesis and its progression. However, several questions remain to be answered at this stage. For example, it is not certain if pharmacological activation of p53 will restore all of its multifaceted biological responses, assuming that the targeted cell is not killed following p53 activation. Thus, several research groups attempt assess and understand the effects of activated p53 on neoplastic progression [13].

Finally, the integration of therapeutic gene product into the human genome occurs through plasmid DNA that contains a therapeutic gene and a small control sequence. The therapeutic gene can replicate during cell division and thereby provide sustained expression of the therapeutic gene product [14].

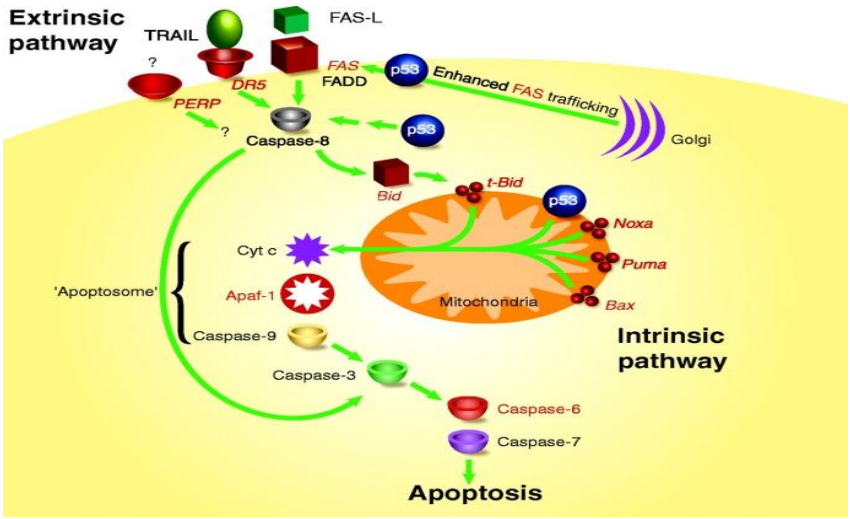

Figure 3- A model for p53-mediated apoptosis. This model depicts the involvement of p53 in the extrinsic and intrinsic apoptotic pathways. p53 target genes are shown in red (adapted from [11]).

1.4. Recombinant Technology and Plasmid DNA Production

1.4.1. Plasmid DNA vector selection

In this emerging technology the choice and design of plasmid DNA is crucial to success. This biomolecule has to combine sequences for replication and selection in *Escherichia coli* (bacterial region) with sequences necessary to express an encoded transgene in vertebrate cells (eukaryotic region) after delivery to an organism and transfection of target tissue cell [15]. Plasmid construction is relatively straightforward and permits the manipulation of a variety of regulatory elements that impact on gene transfer. For simplicity, an expression plasmid can be considered in two parts, which are: one is the mammalian expression cassette (EC) that contains the mammalian enhancer/promoter sequences for gene expression, reporter transgene or gene of interest and polyadenylation (poly A) sequence; the other is the bacterial backbone (BB) that contains a bacterial origin of replication (ori) and an antibiotic resistance gene or other selectable marker for plasmid amplification in bacteria [16, 17]. Additionally, plasmid DNA vectors (figure 4) intended for use as eukaryotic expression vehicles have to meet some pre-requisites such as [18]:

- a constitutive, inducible, or tissue-specific promoter with potent transcriptional activity and a transcription terminator;

- elements for optimized mRNA processing and translation, including Kozak sequence, translational termination, mRNA cleavage, polyadenylation as well as mRNA-splicing sites;

In general vectors are difficult to design due a numerous factors that affect them.

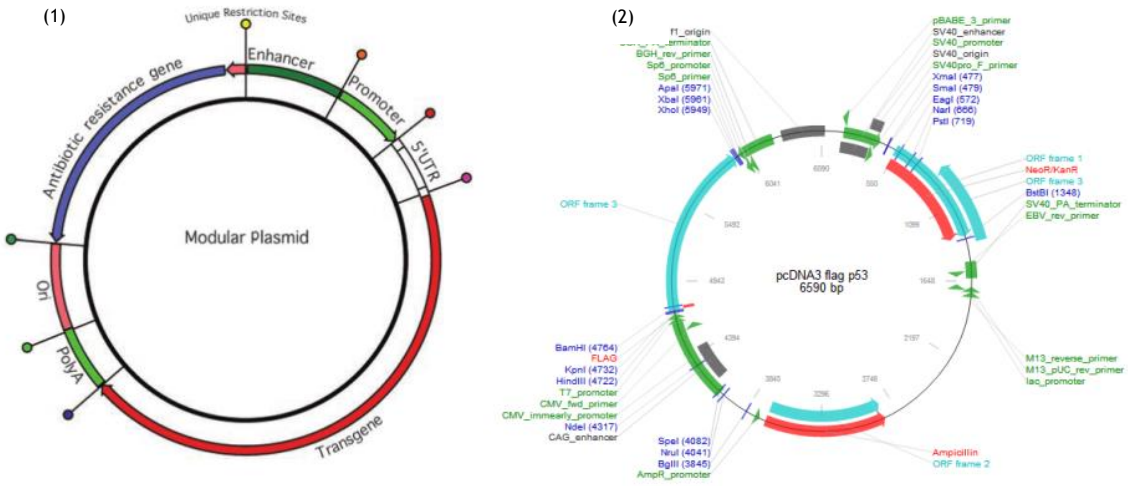

Figure 4- Diagram of modular plasmid design (1). Plasmid DNA of current study (2) (adapted from [15]).

As described in literature, these factors are associated to plasmid copy number, genetic fidelity and segregational stability of plasmid. Many vector engineering efforts are focused on modification to improve the production process by increasing yield, improving product homogeneity and quality, and/or ensuring the sequence fidelity of the final plasmid product [17, 19, 20].

1.4.2. Plasmid Structural Stability

The pDNA required for a therapeutic product should be homogeneous with respect to structural form and DNA sequence. This implies that the homogeneity and quality of the final pDNA product will be a function of the interactions which occur between the host, the plasmid and the growth environment [21]. Accordingly, plasmid vector must be controlled during the production process because of its enormous impact on product safety, yield, and quality. This structure stability is ensured through modifications in the strain and vector used for production of plasmid DNA [17].

The major factors affecting plasmid structural stability are plasmid size, purine-pyrimidie/oligopurine-oligopyrimidine tracts, G-rich sequences, direct repeats, inverted repeats, polyA sequence, nuclease sensitive regions, regions similar to genomic DNA, and insertion sequences [20].

1.4.2.1. Size

Several large molecular weight plasmids have been designed to accommodate large genes, tissue-specific expression regulatory elements, or multiple immunological protein genes [22]. On other hand, the increase of plasmid size leads to a general stress on the host cell and also conducts to a decrease in the maximum yield [23]. However, some experimental data in literature showed that the impact of increased plasmid size on the growth rate is strictly dependent on the host/vector system and not only on the plasmid size. To overcome a few problems with the increased use of large plasmids, the mobile elements (e.g. insertion sequences) were removed from bacterial genome [23]. Regarding on diffusion of DNA in the cytoplasm of cells is strongly size dependent, with little or no diffusion for DNA upper 2000 pair bases (bp). The transfection of cells by a minicircle 2.9 kb in size is 77 times more efficient than a plasmid 52.5 kb in length, according to figure 5. Moreover, the effect of plasmid size on promoter/enhancer activity in transfection assays revealed that expression declined as a function of vector size and more sharp decrease in reporter gene activity. The

major problem seems to be molecular crowding and collisional interactions exerting a strong effect on intracellular diffusion [2].

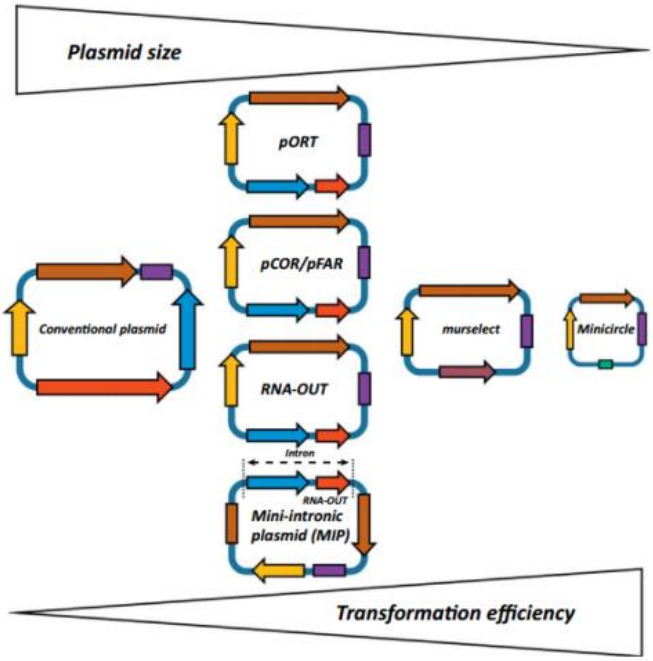

Figure 5- Ranking of alternative plasmid selection approaches according to plasmid size and transformation efficiency (adapted from [2]).

1.4.2.2. Purine–pyrimidine/oligopurine–oligopyrimidine tracts

Not all DNA has the familiar Watson-Crick double helical structure, but exhibit other forms such as Z-DNA, triplexes or quadruplexes that are involved in deletions and reduced plasmid yields. These different DNA structures occur more frequently than would be expected if the sequence of nucleotides were random. For example, homopurine–homopyrimidine sequences with the potential to form intramolecular triplexes appear in eukaryotes three to four times more frequently than they would be expected, while purine–pyrimidine tracts, which are able to form Z-DNA, are also over represented up to forty fold times [24].

Some roles have been suggested for non-Watson Crick structures:

- Z-DNA: transcriptional enhancer and terminal differentiation.
- Triplex: molecular switch.
- Quadruplex: gene regulation.

1.4.2.3. PolyA sequence and other nuclease sequence regions

The polyadenylation complex will be responsible for the 3-cut and addition of 200 to 250 adenine residues (polyA tail). Plasmid polyA sequences also have a strong effect on transgene expression, not only due to the impact at the post-transcriptional level, but also by modifying plasmid DNA resistance to nucleases.

Once intravenously administered supercoiled (SC) plasmid DNA is quickly converted to the open circular (OC) and linear (L) forms. As reported in the literature, molecules possessing base sequences that originate from secondary structures with single-stranded regions (hereafter named hot spots) are more susceptible to exo/endonucleases. Consequently, the elimination of these hot spots in poly A's, by replacing for a synthetic poly A sequence, or eliminate the hot spot by deletion of nucleotide sequences will result in plasmid DNA variants more resistant to nuclease action [25, 26].

1.4.2.4. Insertion sequence

Several research groups observed the *E. coli* insertion sequence IS1 in samples of plasmid DNA. The IS1 is the smallest (768bp) known transposable element among the bacterial insertion sequences and it is found naturally in *E. coli* strains. Typically, IS1 causes spontaneous insertion mutations with much higher frequency than other insertion sequence and was identified as the causative agent for mutations in both plasmid and chromosomal DNA. Moreover, the position of IS1 insertional mutation also affects plasmid replication and amplification [3, 27].

1.4.3. pDNA isoforms

In bacterial cultivation for plasmid productions, the different plasmid isoforms will be present. The most active and compact structure is the covalently closed circular (CCC) or supercoiled plasmid topology. If one strand is broken, the open circular forms results. This naturally occurs by bacterial plasmid production or while processing the biomass (enzymatic or mechanic degradation). Linear forms are generated if both strands are cleaved once at approximately the same position. As described below, there are other characteristics belonging to different isorforms and assist in its analysis [28].

1.4.3.1. Supercoiled pDNA

Currently, the perception is that the supercoiled form is the preferred form of DNA for use in gene therapy and vaccination as the supercoiled amount is deemed to be a measure of purity. The amount of the supercoiled isoform should be greater than 95 % for pharmaceutical grade DNA and to be supported by FDA requirements [23]. Supercoiled pDNA is considered as a nanoparticulate material, with the length of the SC plasmid ranging in the order of hundreds of nanometers whilst having a diameter between 9.9 and 13.4 nm and presents higher charge densities. Sometimes the supercoiling is removed to give the "relaxed circular DNA" which is fully intact with both strands uncut, but has been enzymatically relaxed. On the other hand, supercoiled denatured DNA is like the conventional supercoiled DNA, but has unpaired regions that make it slightly less compact [29, 30].

1.4.3.2. Open and Linear pDNA

A fraction of the pDNA molecules can exist with different topological conformations, open circular and linear, which arise through single-stranded and double-stranded nicks, respectively [30].

The linear form results from chemical/enzymatic cleavage of the phosphodiester bonds in opposite DNA strands, denatured forms exhibit a conformation where the hydrogen bonding between complementary strands at certain locations has been disrupted and oligomers are a consequence of homologous recombination [31]. The linear plasmid form has been deemed undesirable for clinical purposes, either by give lower levels of expression once transfected and lower efficacy but also due to an increased risk of recombination events and integration into genomic DNA [32].

Plasmid topology is important, especially high-quality pDNA preparations containing predominantly the SC isoform is highly desirable to further studies. However, realize the factors of the plasmid DNA stability during upstream and downstream process are essential too as described bellow [30].

1.4.4. Stability of plasmid DNA during cell culture, storage and alkaline lysis

Recently experimental data showed that during the most of exponential phase the percentage of the supercoiled isoform remains constant 94 %, however with the onset of the stationary phase occurs slight decrease to 90 %. These results indicate that the stress conditions found

during the stationary phase probably are a response of the bacterial DNA supercoiling control [17, 33, 34].

Concerning the cell pellet storage, this can be conveniently stored at 4 or -20º C for up to three or twelve weeks, respectively, without significant increase in the open circle or linear isoforms or detectable reduction in the yield of the SC isoform. More significantly it was observed reduction in high molecular weight RNA when the pellets were stored for more than two weeks at 4º C [28, 34].

Regarding the lysis step, the scientific community reported the better ranging concentration of NaOH in alkaline lysis was 80 to 100 mM, whereas concentrations greater than 150 mM leads to irreversible denaturation of the supercoiled pDNA. An interesting study indicates that *E. coli* cells pellet exposed to the alkaline lysis revealed no significant formation of open circular or linear isoforms, even after four hours of lysis [34]. However, the alkaline environment can be adverse to supercoiled plasmid DNA due to the prolonged exposure. On other hand, the increase in alkaline concentration did not affect RNA concentration significantly during short time lysis procedures. Prolonged alkaline lysis can significantly reduce RNA furthering the subsequent purification [34, 35].

1.4.4.1. Stability in Replication Process

Several natural plasmids are stable at their characteristic copy number, controlling their concentration and regulating the rate of replication on growing bacterial population. The minimal portion of a plasmid that replicates with the characteristic copy number of the parent plasmids called basic replicon. Replicons contain the ori but also the genes encoding specific replication initiator proteins that bind the ori and their regulating factors [36]. Control systems maintain the rate of replication in the steady state at an average of one replicative event per plasmid copy and cell cycle by correcting deviations from the average copy number in individual cells. Nevertheless, when a plasmid colonizes a new host, the concentration of existing negative regulators should be negligible. This seems desirable for successful establishment, since unhindered plasmid replication would permit the normal copy number to be reached in a short time. However, once the characteristic copy number is reached, keeping the average copy number in the population requires adjustments in this value for individual cells [37]. The control systems compromises increasing or decreasing the rate of replication per plasmid copy and cell cycle. Additionally, mechanisms controlling replication have been studied in various systems and several types of inhibitors have been detected [1, 37] such as;

- antisense RNA.

- both an antisense RNA and a protein.
- DNA sites for binding initiator proteins.

1.4.5. Host Selection

The ideal host strain should be able to grow to high cell density with high plasmid copy number, maintain genetic stability and be amenable to the downstream purification process. In 1922, the *E. coli* K-12 was the first strain isolated and since that time thousands of mutant strains have been produced. The status of *E. coli* K-12 as the most well understood model organism meant that it was the obvious candidate for the first bacterial genome sequencing project to be initiated. Many strains of *E. coli* have been created through a series of mutations to facilitate cloning of heterologous genes and maintain the stability of plasmid DNA. For example, *E. coli* K-12 strains like DH5α, DH10B, JM108, VH33, VH34 are typically used for plasmid DNA production [3, 38].

Strain	Genotype	Derivation
ABLE K	*lac (lacZ⁻) [Kan^r, mcrA⁻, mcrCB⁻, mcrF⁻, mrr⁻, hsdR($r_K^- m_K^-$)] [F′proAB, lacI^q,ZΔ M15, Tn10(Tet^r)]*	*E. coli* C
BL21 DE3	*F⁻ dcm ompT hsdS($r_B^- m_B^-$) galΔ (DE3)*	*E. coli* B
C600	*(F⁻ tonA21 thi-1 thr-1 leuB6 lacY1 glnV44 rfbC1 fhuA1 λ⁻)*	*E. coli* K-12
DH5α	*F⁻ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(r_K^- m_K^+), λ⁻*	*E. coli* K-12; derived from DH1
DH1	*endA1 recA1 gyrA96 thi-1 glnV44 relA1 hsdR17 (r_K^- m_K^+) λ⁻*	*E. coli* K-12; derived from strain 1100, parent of DH5α
DH10β	*F⁻ endA1 recA1 galU galK deoR nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ⁻*	*E. coli* K-12; MC1061 derivative
HB101	*F⁻ mcrB mrr hsdS20 (r_B^- m_B^-) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20 (Sm^R) glnV44 λ⁻*	*E. coli* K-12; hsd genes derived from *E. coli* B
J53	*proA met*	*E. coli* K-12; J5 derivative
JM107	*endA1 glnV44 thi-1 relA1 gyrA96 Δ(lac-proAB) [F′ traD36 proAB$^+$ lacI^q lacZΔM15] hsdR17(R_K^- m_K^+) λ⁻*	*E. coli* K-12: DH1 derivative
MG1655	*F⁻ λ⁻ ilvG- rfb-50 rph-1*	*E. coli* K-12 wild-type
SURE 2	*e14⁻(mcrA⁻), Δ (mcrCB-hsdSMR-mrr)171, endA1, supE44, thi-1, gyrA96, relA1, lac, recB, recJ, sbcC, umuC::Tn5(Kan^r), uvrC, [F′ proAB, lacI^{q}ΔM15, Tn10(Tet^r), Amy, Cam^r]*	*E. coli* ER1451 derivative
TG1	*supE thi-1 Δ(lac-proAB) Δ(mcrB-hsdSM)5 (r_K^- m_K^-) [F′ traD36, proAB, lacI^q ΔM15]*	*E. coli* K-12
TOP10	*F-mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 deoR nupG recA1 araD139 Δ(ara-leu)7697 galU galK rpsL(Str^R) endA1 λ⁻*	*E. coli* K-12; MC1061 derivative, genotype very similar to DH10β
TURBO	*[F′ proA$^+$B$^+$ lacI^q ΔlacZM15/fhuA2 Δ(lac-proAB) glnV zgb-210::Tn10 (Tet^R)] endA1 thi-1 Δ(hsdS-mcrB)5*	*E. coli* K-12
W3110	*F⁻ λ⁻ rph-1 INV (rrnD, rrnE)*	*E. coli* K-12 wild-type
XL1 Blue	*endA1 gyrA96 (nal^R) thi-1 recA1 relA1 lac glnV44 [F′::Tn10 proAB$^+$ lacIq Δ(lacZ)M15] hsdR17(r_K^- m_K^+)*	*E. coli* K-12
XL10 Gold	*endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tet^R [F′proAB lacI^{q}ZΔM15 Tn10(Tet^R Amy Cm^R)]*	*E. coli* K-12

Figure 6- List of *E. coli* strains commonly used in plasmid DNA production (adapted from [22]).

Actually, the complex genotypes of many *E. coli* strains (figure 6) for industrial applications also greatly hinder the ability to reliably predict the plasmid yield based on the genotype alone. With these factors in mind, many researchers have used specific strategies to develop well-characterized strains for pDNA production. These strain engineering efforts often seek to preserve the sequence fidelity of both the plasmid and host genome and to address

downstream process issues [17]. Although much effort has been directed to downstream processing, further work is needed to improve the earlier stages of production namely the cell host/plasmid selection. The choice of the *E. coli* host strain used to propagate the product containing-plasmid may have sizeable influence on the quality and quantity of the purified DNA [23].

As expected, plasmid-bearing cells had a slower growth rate and an altered profile of central metabolic gene expression when compared to plasmid-free cells. The ability to recover a growth rate comparable to plasmid-free cells is particularly useful in plasmid DNA production as it reduces the selective pressure against plasmid-bearing cells and increasing plasmid stability. Nevertheless, *E. coli* strains with high plasmid copy numbers and high plasmid retention levels (segregational stability) are used as hosts for plasmid production, particularly if they are compatible with downstream processing and have minimum potential for genetic alterations [17, 23].

The great potentials of use *E. coli* strains are 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 [39]. Aerobic acetate production, as known as overflow metabolism, results from an imbalance between glycolysis and tricarboxylic acids cycle. Some of the strains commonly used for pDNA production present elevated overflow metabolism, including *E. coli* DH5α, DH1, JM105, JM109 [32, 40].

The Gram-negative bacterium *Escherichia coli* is the most commonly used host for the propagation of pDNA [2, 22, 40, 41], because it is very robust, capable of fast growth with minimal nutritional requirements, and can give high pDNA yields. The genome of *E. coli* is fully sequenced and can be easily manipulated by techniques that are well known and characterised. On the other hand, *E. coli* has some disadvantages like endotoxin production and genetic instability, resulting in safety concerns surrounding its use. However, taking into consideration the benefits and drawbacks, *E. coli* is currently the most suitable organism for pDNA production on the industrial scale [3, 28].

1.4.6. Central Carbon Metabolism

Plasmid DNA synthesis can disturb *E. coli* gene regulation, altering levels of gene expression and carbon flux. Glycolysis is the main pathway for glucose uptake and energy generation. It is composed of ten reactions catalyzed by specific enzymes that are coded by individual genes. The pentose phosphate pathway (PPP) can metabolize different sugars like, xylose and ribose, but is considered the second main destination for glucose [3]. When *E. coli* grows on glucose as carbon source, this sugar is imported by the phosphoenolpyruvate (PEP): sugar phosphotransferase system (PTS) that couples the transfer of the phosphate group from PEP

to this sugar, thus generating as products glucose 6-phosphate (G6P) and pyruvate (PYR). The pentose phosphate pathway is also one of the pathways responsible for biosynthesis of the nucleotide precursors ribose-5-phosphate (R5P) and erythrose-4-phosphate (E4P). Another important product from the PP pathway is NADPH, synthesized by glucose 6-phosphate-1-dehydrogenase and 6-phosphogluconate dehydrogenase [42]. 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 [43, 44].

Typically cells carrying high copy pDNA 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. The tricarboxylic acid cycle (TCA cycle) is extremely relevant 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 and α-ketoglutarate. In plasmid-carrying cells, most of the TCA genes were observed as up-regulated for different *E. coli* strains [3, 4, 45].

Thus, understanding the flux distribution and its regulation in central carbon metabolism (i.e. an "assessment" of metabolism) is fundamental for the progress of metabolic engineering [46]. With the knowledge of flux distribution (for instance precursors, reduction equivalents, etc), an effective improvement in the yield of the biosynthetic production of primary and secondary metabolites could be achieved [42]. Although the metabolic pathways have been known for a number of decades, the quantification and regulation of metabolic networks is still not fully understood due their numerous intervenient like shown in figure 7 [47].

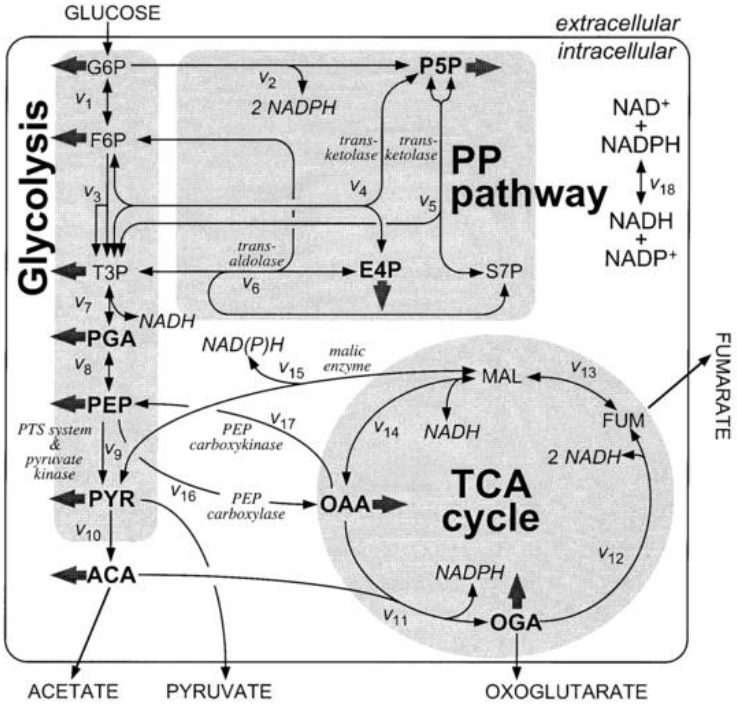

Figure 7- Bioreaction network of *E. coli* central carbon metabolism (adapted from [53]).

1.5. Importance of Environmental Conditions to Bacterial Growth

1.5.1. Medium Characterization

The formulation of the cultivation medium affects dramatically the performance and nature of microbial processes. This implies that, the type and source of nutrients available in the growth medium can have a profound effect on fermentation yields [48]. The growth of *E. coli* can be limited by other nutritional requirements including carbon, nitrogen, phosphorus, sulfur, magnesium and potassium iron, manganese, zinc, copper and specific growth factors [49]. All these compounds have precise functions during the bioactivity of cell growth and plasmid yield. Physiologically, hydrogen and oxygen form the basis of cellular water and with carbon as the main constituent of organic cell materials. Cellular cations and enzymes cofactors are provided by salts solutions, which are very important in *E. coli* growth. The type and concentration of these basal constituents applied in cultivation medium determines the amount of biomass produced, as well as plasmid volumetric yield and specific yield [48].

Additionally, the effects of medium composition for plasmid biosynthesis are closely intertwined with that affecting plasmid segregational stability. An important factor contributing to plasmid stability is the presence of an antibiotic, selecting cells containing the desired plasmid, in the cultivation medium [20]. A good selection of medium composition with association of specific cultivation conditions play an important role by controlling plasmid copy number, stability and the amount of biomass produced [50]. The design and development of an optimized cultivation medium can provide an extensive pay-off in a long-term strategy. Coupled with the powerful tool of statistical experimental design, the design of chemically defined cultivation medium will likely yield highly optimized media formulations [33, 51, 52].

1.5.1.1. Carbon Source

The growth of biomass was greatly influenced by the right proportion of carbon source. Much of the carbon source taken up by *E. coli* enters the pathways of energy yielding metabolism adenosine triphosphate (ATP) and is eventually secreted from the cell as CO_2 (the major product of energy yielding metabolism) [48].

The carbon sources most commonly used are glucose [21, 42, 53, 54] and glycerol [51, 55, 56]. Glycerol is preferable as a carbon source because it is metabolized more slowly as a

carbon source than glucose, in order to reduce the maximum specific growth rate during batch fermentation and to minimize acetate production during batch fermentation [57]. The addition of glucose also boosted both plasmid DNA volumetric and specific yields since glucose is needed for the synthesis of nucleotides in plasmid DNA [48, 58].

Several experiment focusing the effects of different carbon sources showed that the highest cell mass and plasmid production were obtained when glucose was used (figure 8), whereas both cell mass and plasmid productivity were reduced when lactose and sucrose were used [59]. On the other hand, glycerol exhibited highest specific plasmid productivity but poor cell growth. However, the application of glycerol avoids repression of intermediate metabolites and accumulation of inhibitive organic acids to some extent [59].

Strain	Carbon source (g/L)		pDNA yield (mg/g DCW)	pDNA yield (mg/L)
MG1655 *ΔendA ΔrecA*	Glucose	20	0.8±0.1	1.5±0.1
		10+10	0.8±0.1	1.9±1.2
		5+10	3.6±0.7	27.4±1.9
	Glycerol	20	11.5±0.8	79.3±1.4
		5+10	11.2±0.5	75.1±3.9
GALG1011	Glucose	20	1.6±0.3	3.2±0.4
		10+10	1.2±0.5	4.3±1.7
		5+10	6.6±0.4	42.1±3.5
	Glycerol	20	7.6±1.1	50.5±1.3
		5+10	2.7±0.2	20.9±0.2
GALG10	Glucose	20	0.99±0.1	3.45 ±0.3
		5+10	10.5±1.3	81.5±5.7
	Glycerol	5+10	8.6±0.6	69.2±3.9
GALG11	Glucose	20	0.65±0.1	2.4±0.3
		5+10	13.1±0.2	94.1±2.7
	Glycerol	5+10	10.3±0.6	79.4±7.0
GALG20	Glucose	20	19.1±1.5	140.8±0.8
		5+10	11.6±1.1	88.9±0.7
	Glycerol	5+10	10.1±0.1	65.5±1.4
DH5α	Glucose	20	0.8±0.1	1.3±0.1
		5+10	1.8±0.7	9.6±0.5
	Glycerol	20	4.4±0.3	34.7±0.6
DH5α *ΔpykF ΔpykA*	Glucose	20	0.4±0.1	2.6±0.2
		5+10	0.3±0.3	0.9±0.2
	Glycerol	20	1.5±0.3	5.9±1.1
DH5α *ΔpykF*	Glucose	5+10	5.9±0.1	36.9±0.4
DH5α *ΔpykA*	Glucose	5+10	0.9±0.1	1.5±0.1
JM101	Glucose	20	1.3±0.2	4.5±0.8
		5	2.5±0.4	35.5±6.8
	Glycerol	20	8.4±1.4	54.3±2.8
JM101 *ΔpykF ΔpykA*	Glucose	20	2.6±0.3	12.5±2.8
		5	5.3±1.7	28.5±4.3
	Glycerol	20	2.2±1.3	12.5±0.5

Figure 8- Effect of glucose and glycerol in different concentrations for pDNA biosynthesis in different *E. coli* strains (adapted from [4]).

1.5.1.2. Nitrogen Source

Two different types of nitrogen sources, yeast extract and, casamino acid were identified to be the key factors which affected plasmid productivity [59]. However, tryptone is associate as a nitrogen source exhibiting several advantages, such as carbohydrate-deficient unlike yeast extract that is an amino acid source [50]. Comparing yeast extract and casamino acid in fermentation media, the casamino acids showed less effectiveness nitrogen source in plasmid DNA production [59]. The addition of casamino acids seems affect the relative growth rates, decreasing plasmid production [60].

Media optimization needs to be carefully performed, because C/N ratio has a tremendous influence on plasmid specific yield. Even though no specific mechanism was known that could explain why plasmid levels varied with the C/N ratio, Bryers and co-workers demonstrated that kinetic constants for plasmid loss from recombinant *E. coli* were highly dependent on the C/N ratio [61]. According many researches the maximum plasmid specific yield obtained was C/N=3 [19, 48, 50, 60].

1.5.2. Culture conditions

A single plasmid can exist in a number of different forms in a cell. Changes in growth conditions, including temperature, nutrient concentrations, pH, oxygenation and growth phase can affect the presence of biological active supercoiled isoform [29].

1.5.2.1. Oxygen demand

A major factor for cells culture is the oxygen demand profile, which can be aerobic or anaerobic. The aerobic culture favors faster growth but anaerobic conditions are needed for the formation of certain desired products (ethanol, lactic acid). The pattern of metabolites produced by *E. coli* under conditions of high oxygen availability, microaerobic conditions and anaerobic regime fermentation has been widely analysed and studied during the last decades [57]. Basically, in the aerobic state the TCA cycle operates to oxidize pyruvate with the reductants formed coupling with the electron transport chain to generate the proton gradient, which is used for ATP production [62].

According to literature, the yield-constant for oxygen of an *E. coli* strain growing on glucose as a carbon source is close to 1, specifically, 1 g of oxygen is needed for the production of 1.06 g of *E. coli* biomass [49].

Recent investigations already use several strategies of oxygen supply because the oxygen availability in liquid phase it is closely associated with plasmid stability, growth cell rate and sub-products production. For example, oxygen limitation caused by a deficient oxygen supply can lead to a strong segregational instability in certain plasmid–host systems. Decreasing the growth rate by limiting dissolved oxygen increased plasmid specific yield and emerged as a viable strategy for maintaining productivity during scale up. Also, the production of toxic by-products decreases, mostly acetic acid, which severely limits growth and lead to cell death. [33, 52, 56, 63]. So, if pure oxygen is available instead of air supply, the oxygen transfer rate in liquid cultures would not become a limiting factor [49].

In anaerobic regime *E. coli* ferments glycerol in a pH-dependent bio system. The production of CO_2 from formate is required for glycerol fermentation to proceed. Therefore, high pH resulted in low CO_2 generation and availability (most CO_2 is converted to bicarbonate), thus resulting in very inefficient fermentation of glycerol. Hence, absence O_2 confirms a low potential for production of plasmid DNA [57].

1.5.2.2. Temperature and pH control

The control of external factors such as temperature or pH can modulate plasmid copy number or balance the metabolic burden effects, maintaining high copy number plasmids in high-yield fermentations [17]. Literature demonstrated that pH and temperature greatly influenced plasmid copy number showing these environmental factors (elevated pH and reduced temperature), positively influenced plasmid copy number by down-regulating the growth rate [33]. Temperature shifts allow the cultures reach higher cell densities and also no visible filamentation or loss of cell viability [64].

Relatively to pH, the overall external pH not changes drastically, because of the buffered medium applied, however the decrease of intracellular pH has a significant effect. Indeed, at lower pH affect the acetate producing pathways leading to less acetate levels [57, 62, 65].

1.5.2.3. Acetate, production and effect

The major metabolite of glucose metabolism is acetate. Because of its weak lipophilic nature, the protonated form of acetate is able to cross the cell membrane and act as an uncoupler of the proton motive force [65]. It can cause inhibition of cell growth and recombinant protein production and therefore has been recognized as a severe handicap during the upstream stage. The approaches to overcome this problem have to be carefully selected because acetate pathway plays an important physiological role in *E. coli* metabolism. In general, the

acetate production depends on many factors including the growth medium, glucose feeding strategy, growth conditions, and bacterial strain [66]. The acetate production pathway is used as a source of ATP formation under anaerobic and even aerobic condition [67]. The aerobic acetate production, known as overflow metabolism, results from an imbalance between glycolysis and tricarboxylic acids cycle [40]. In a rich medium, cells produce excess of acetyl-CoA that flows through acetyl-P (an intermediate of acetate pathway) to ATP with the associated excretion and accumulation of acetate in the extracellular environment. This phenomena leads to a sensitization intrinsically signals, changing the Acetyl-P levels that can significantly influence the growth rate of *E. coli* on glucose [67].

Additionally, acetate accumulation diminishes productivity as it represents a waste of carbon source that would otherwise be converted to biomass and product. It is probable that if cells were stressed under high glucose concentrations, the accumulated acetate originated from stress responses and not only from overflow metabolism. A relevant characteristic of the media formulated with high initial glucose concentration was the intrinsic high osmolality due to the elevated amount of media components leading to a strong growth inhibition when acetate accumulated to values above 5 g/L [68]. In addition, glycerol is preferable as a carbon source because it is utilized more slowly as a carbon source than glucose, in order to reduce the maximum specific growth rate during batch fermentation and to minimize acetate production [58].

Nevertheless, during the exponential phase of a typical batch fermentation the rate of metabolism and acetate production are high, which is detrimental to pDNA replication [48]. So the researchers developed *E. coli* mutants with alterations in specific enzymes that control acetate biosynthesis pathway which can improve the growth of *E. coli* and plasmid production [4, 42, 49].

1.6. Monitorization of fermentation cultures

The development of a fully monitoring and controlling system is very important during a fermentation process. The analysis of bioprocesses are more complex than pure chemical processes, that real-time access to physiologically relevant variables is not available, and bio products are large molecules that are extremely sensitive to the manufacturing process [69]. The commonly measurable parameters are pH, temperature, aeration, agitation and dissolved oxygen. However, monitorization of cells, substrate, metabolites and product concentrations during fermentation is crucial but more difficult. The solution to overcome this problem is developing analytical methods using High Performance Liquid Chromatography (HPLC)[70], Gas Chromatography (GC) [71], Mass Spectrometry (MS) [72] or Flow Cytometry (FC) [73] systems (table 2). These offline systems show several advantages, such as, fast and

simultaneously analysis of many components from fermentation broth, high reproducibility and acceptable benefit-cost [74].

Table 2- Examples of techniques used in a bioprocess monitoring for plasmid biosynthesis (adapted from [69]).

Monitoring	Techniques	Analytes
Biomass and Growth Biomass	Optical density measurements Gravimetric cell dry mass	Determination of biomass
	Dielectric spectroscopy Flow cytometry	Cell viability
Metabolites and Cell Compounds	HPLC methods	Sugars, amino acids and metabolites
	Proton transfer reaction mass spectrometry Gas Chromatographic	Volatile organic compounds
	2D multiwavelength fluorescence spectroscopy	Fluorescent components
Products from the Recombinant System	Western blotting	Target proteins and fragments
	Enzyme-linked immunosorbent assay (ELISA)	Protein quantity and quality
	Real-time PCR (RT-PCR)	Quantify messenger RNAs (mRNAs)
	Counting colony-forming units	Plasmid-containing cells and genetic stability

A HPLC system is a well-established technique for the analysis of complex mixtures, where the separation can be based on reverse phase, ion exchange, and affinity separation. Accomplish to a HPLC system, several detectors are used for concentration measurements, depending on the analyte nature. The most detectors used are ultraviolet/visible photometer, amperometric detector, potentiometric detector, refraction index detector [75].

According to literature and as an example of these systems, nucleotides and sugar phosphates were studied together with anion exchange chromatography by conductimetric detector, fluorometric and ultraviolet detection. Beside, carboxylic acids in TCA cycle were separated by cation exchange chromatography with ultraviolet detector or two detectors set up in series such as ultraviolet and refractive index [72].

Briefly, setting up an offline measurement platform is a nontrivial task as the set of selected techniques (i) must cover the wide diversity of all molecules of interest and (ii) should allow high sampling frequencies combined with low sample volumes, and (iii) cost–time demands must not exceed justifiable quantities. For more or less all molecule classes in a cellular

system, various analytical methods are available but only a limited number meet the above-mentioned criteria. Hence, careful selection and combination of methods is imperative [69].

1.7. Optimization Process

A main goal of process optimization and scale up consists in the development of appropriate process design (table 3) which improves the physiological conditions and the metabolic accuracy by minimizing microbial stress exposure [39].

The fundamental fact of bioprocessing is that the purity of the final product is proportional to the purity of the starting material. Thus, as the initial fermentation yield of plasmid DNA per gram of bacteria increases, the same occurs on purity of the final product. Although high cell densities are important to maximize volumetric yields from culture, an increase in the specific plasmid yield also leads to higher purity [76].

Table 3- Relationship between cause-effect on optimization process (adapted from [15]).

Optimization	Cause	Effect
Growth Rate	Reduce growth rate	High quality and high yield plasmid production Less acetate production Less plasmid instability Control Plasmid replication
Growth Conditions	Oxygen Temperature pH Dissolve oxygen	Increase supercoiled plasmid Plasmid stability
Media	Minimal media	Higher plasmid copy numbers and highly reproducible;
	Semi-defined and complex media	Higher cell densities and decrease in reproducibility

To optimize the fermentation process, several strategies can be addressed. According table 3 optimizations focus on rate and the physical parameters of cell growth and also substrates composition. In general, the purpose of these alterations is to increase cell density, stability and number of copies of plasmid [76].

1.7.1. Engineering *Escherichia coli* strains

Many cell line engineering efforts have sought to improve pDNA production by knockout or overexpression of rationally selected genes. One main area of focus is modification of central

carbon metabolism genes to increase flux toward nucleotide and amino acid precursor synthesis and reduction of byproducts, such as acetate [3].

A molecular solution to this problem is the replacement of the natural glucose uptake system in *E. coli* (phosphoenolpyruvate carbohydrate phosphotransferase system, PTS) by galactose permease (GalP) and glucokinase. Compared to the PTS system, the expression of GalP reduces the transport rate of glucose. Consequently, the acetate production was strongly reduced with little effect in growth rate [42, 43, 68].

Another approach followed to increase the availability of carbon for the synthesis of nucleic acids precursors was to reduce the activity of the pyruvate kinase by inactivation of the pykA and pykF genes [4, 40].

The aromatic amino acid biosynthetic pathway is coupled to central metabolism of *E. coli* and its regulation could affect the plasmid production. The first step in the aromatic or shikimate (SHIK) pathway consists in the condensation of PEP and erythrose 4 phosphate (E4P). This reaction is catalyzed by the DAHP (3-deoxy- D-arabinoheptulosonate 7 phosphate) synthase isozymes (AroF, AroG and AroH). The final product is chorismate (CHO), a precursor molecule for the synthesis of aromatic amino acids tryptophan (TRP), tyrosine (TYR) and phenylalanine (PHE), as showed in figure 9 [40, 42, 77].

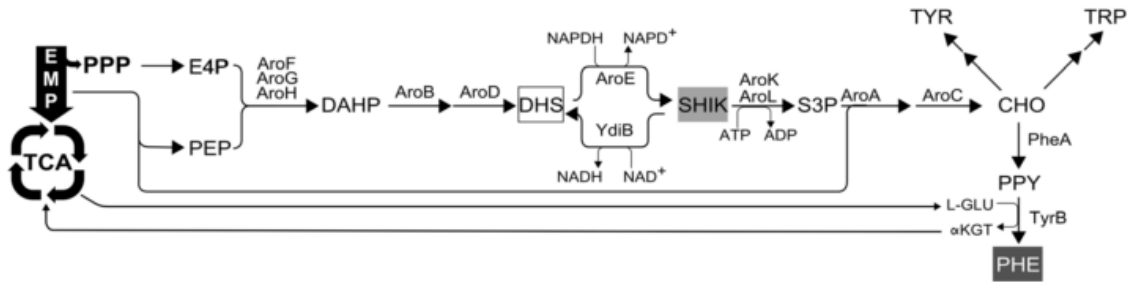

Figure 9- The common aromatic pathway leads to PHE, TYR and TRP production (adapted from [41]).

1.7.2. Experimental design

Due to the wide range of process variables and the metabolic complexity of cells, experimental design procedures are often used to determine the optimal operational conditions. According to the number of the factors to be investigated at a time, the experimental design can be classified into two categories: one-factor-at-a-time design (single-factor design) and factorial design (multiple-factor design) [78, 79].

The one-factor-at-a-time design is a traditional design, which investigates one-factor-at-a-time, while keeping the levels of other parameters constant. The level of the factor to be

investigated is then changed over a desired range to study its effects on a response. In contrast, factorial design is able to study the effects of more than one factor at two or more levels and enables depict the interactions among different factors. Typically, factorial design can be classified into two categories: full factorial design and fractional factorial design (that includes the Taguchi design, Central Composite design, Plackett–Burman design and Box–Behnken design) [79].

In a full factorial design, every combination of each factor level is tested. For example, the number of runs for a three-factor full factorial design is **axbxc**, which indicates that, the first factor is tested at **a** levels, the second factor is tested at **b** levels, while the third factor is tested at **c** levels. The most commonly used full factorial design is two-level design, which can be denoted by 2^n when there are n factors (Number of runs =2^n) [51, 79]**.**

The fractional factorial design is used when the number of runs is too large to be tested by experimental work. The Taguchi design using an orthogonal array, allows the effects of many factors with two or more levels on a response, to be studied in a relatively small number of runs. This design is a powerful and efficient method to find an optimal combination of factor levels. The Central composite design is a five-level fractional factorial design that consists in 2^n full factorial design, 2x*n* axial designs and *m* central designs. The axial design is identical to the central design except for one factor, which will take on levels either above the high level or below the low levels of the 2^n full factorial design [79].

Another approach to deal with a large number of factors, as example of the upstream stage of a bioprocess, is used Plackett-Burman design. This design identify the key factors of the respective production system, as called "Screening". Secondly, gradient method needs to be used to find the optimum variable ranges [78]. Plackett–Burman design allows testing of the largest number of factor effects with the least number of observations, and allows random error variability estimation and testing of the statistical significance of the parameters [78][51]. Then, the optimal values are obtained through mathematical methods such as response surface methodology (RSM) [80]. Besides analysing the effects of the independent variables, this experimental methodology generates a mathematical model, which can be used to describe the process for better understanding. As an example of RSM, the Box-Behnken design is a three level factorial design, which allows estimating and interpreting interactions between various variables at a time during the optimization process. It is suitable for exploration of quadratic responses and constructs a second order polynomial model with very few runs [79, 81, 82].

As an alternative approaches to optimization, it is can applied a mathematical representation of the neurological functioning of a brain, called as neural network, which is able to describe the interactive effects of various factors on a complicated process and has been applied successfully to construct models [82]. A neural network does not need any model or screening

before the development of a network and it may be applied on designed data or on the data that is not statistically designed. This powerful tool uses all the data making the model more accurate and can perform tasks that cannot be performed by linear programming [79, 82].

Briefly, there is a huge potential of all tools described above to develop models (fermentation conditions, physical parameters) and achieved optimal range concentrations of medium components (carbon and nitrogen sources, mineral salts, trace elements, vitamins and other growth factors) [78].

Chapter 2- Experimental Part

2.1. Reagents

Methanol (HPLC-grade) was acquired from VWR Internacional (Carnaxide, Portugal). The ethylenediaminetetraacetic acid (EDTA), o-phtalaldehyde (OPA), glucose, fructose, trehalose, *o*-phosphoric acid, sulfuric acid and acetic acid (HPLC grade) were acquired from Sigma-Aldrich (Sintra, Portugal). The L-tyrosine, L-phenylalanine and L-tryptophan (≥99 % purity) were acquired from Sigma-Aldrich (Sintra, Portugal). Sodium sulfite was acquired to José M. Vaz Pereira, S.A. (Sintra, Portugal). Sodium borate (≥98 % purity) was acquired to MERCK (Algés, Portugal).

2.2. Plasmid

The 6.07 kbp pcDNA3–FLAG–p53, ampicillin-resistant plasmid, with pcDNA3 based backbone. The pcDNA3–FLAG–p53 was purchased from Addgene (Cambridge, MA, USA).

2.3. Strain

In this study it was used *E. coli* VH33 that was gently provided by Professor Guillermo Gosset from the "Instituto de Biotecnología from Universidad Nacional Autónoma de México". Strain VH33 is derivative from *E. coli* W3110 in which the galP promoter was substituted by the trc promoter thus making constitutive the expression of galP. In *E. coli* VH33, glucose import is dependent on GalP and glucose phosphorylation on glucokinase activity. This bacterial strain was transformed with the 6.07 kb plasmid, which was described above.

2.4. Transformation of competent cells

The development of competent cells and process transformation was performed according to Current Protocols in Molecular Biology (1.8.4-1.8.5) (1997) with slight changes, as described below.

2.4.1. Development of *E. coli* competent cells

A single colony of *E. coli* was inoculated into 25 mL LB medium during overnight at 37°C. With this cells, a fresh LB medium was inoculated and the growth was performed at 37°C, 300 rpm and the inoculum grown to 0.35-0.40 of OD_{600}. The flask was chilled and the cells were centrifuged in falcon tubes. The supernatant was discarded and the pellet was resuspended with prechilled glycerol 10%. Again, the cells were centrifuged, glycerol was discarded and the pellet was ressuspended with GYT medium. The concentration of cells in this broth was approximately of $2\text{-}3.10^{10}$ cells per mL and total volume was distributed (100 µL) for eppendendorfs, which were rapidly frozen with liquid nitrogen and stored at -80°C.

2.4.2. Cell Transformation

An aliquot of competent cells was thawed and the plasmid was added to a suspension cells. This mixture was transferred to electroporation cuvette and incubated 5 minutes in ice. The cuvette was placed on an electroporation system where it was applied a pulse with 2500 V. After this, 250 µL of LB medium was added and the broth was transferred to a liquid sterile culture and incubated during two hours. Finally, the LB plates were supplemented with ampicillin for further inoculation.

2.5. Master and working cell bank

The Master cell bank was prepared from a single colony picked from agar plate and inoculated in Luria–Bertani medium supplemented with 50 µg/mL ampicillin. Cultures were grown to 0.6 OD_{600} at 250 rpm and 37°C. Then, 14 mL of cell culture was put into a falcon tube, followed by addiction of glycerol at final concentration glycerol 30 % (v/v). This mixture was distributed in 1 mL cryotubes and frozen -80°C. The working cell bank was prepared by growing an aliquot of master cell bank. The procedure and conditions used were the same as the master cell bank.

2.6. Bacterial cell growth and Plasmid DNA Biosynthesis

2.6.1. Culture Conditions

The *E. coli* strain VH33, was cultivated into different media conditions. All pre-culture and batch fermentations contained 50 µg/mL of ampicillin to maintain a selective growth and equal salt composition: 90 mM K_2HPO_4, 10 mM KH_2PO_4, 40 mM $(NH_4)_2SO_4$, 20 mM NaCl, 1.6 mM $MgSO_4.2H_2O$, 0.05 mM $CaCl_2$, 0.072mM $FeSO_4.7H_2O$. Salts solution were sterilized separately at 121°C.

To understand the effect of glucose on the reference TB medium in plasmid DNA production and purity, cells were grown in five different media. All media contained glucose (5 or 10 g/L) and 20 g/L of tryptone. Regarding yeast extract and glycerol, may be present or not at concentrations of 24 g/L, and 4 mL/L respectively. The nutrients composition of pre-culture was the same used in culture.

After the assays described above the next step was selection of best fermentation media. On selected medias it was varied the carbon source, glucose and/or glycerol. Besides this, the carbon source concentration was changed for cell growth. The glucose concentration was 5, 7.5, 20, 30 g/L and glycerol was 4 m/L, 20 mL/L to 30 mL/L. In these experiments, the pre-culture consisted of LB medium (absence of a specific carbon source)

All fermentations were carried out in 1L shake flasks containing 250 mL of medium at 250 rpm at 37°C and for further analysis the cells were harvested in the final of exponential growth phase after 14 hours of the fermentation stage. All cultures started with an OD600 of approximately 0.2 by inoculation with a pre-culture grown in 500 mL shake flasks containing 125 mL of medium at 250 rpm at 37°C. All fermentation (including pre-culture) was supplemented with 50 µg/mL of ampicillin.

2.7. Cultivation Conditions for the Box-Behnken design

2.7.1. Box-Behnken Design

A typical Box-Behnken design (BBD) was performed using the independent variables. In this study, the three aromatic amino acids (Tyrosine, Fenylalanine and Tryptophan) were the independent variable. Five replicates at the centre point level were also run to check if there was a non-linear relationship between the variables and the responses, leading to a total number of 17 trials. For predicting the optimal point, polynomial function was fitted to correlate the relationship between independent variables and response. The X_1, X_2 and X_3

were the factors used in design that correspond to tyrosine, phenylalanine and tryptophan medium concentration, respectively.

The *Design-Expert 7.0.0* was the software used for generation and evaluation of the statistical experimental design. Briefly, the Box Behnken design was applied in a semi-defined medium. Using glucose at 10 g/L as the main carbon source, yeast extract as potent nitrogen source, salt solution (90 mM K_2HPO_4, 10 mM KH_2PO_4, 40 mM $(NH_4)_2SO_4$, 20 mM NaCl, 1.6 mM $MgSO_4.2H_2O$, 0.05 mM $CaCl_2$, 0.072mM $FeSO_4.7H_2O$) and the three target amino acids (X1-tyrosine, X2-phenylalanine and X3-tryptophan). According the model levels pre-established the amino acids concentration varied according to: level -1, 0 and 1 which correspond respectively to 10, 50 and 100 µg/mL. In general, these experiments were carried out in 500 mL shake flasks containing 125 mL of medium at 250 rpm at 37°C. After fermentation, the cells were harvested in the final of exponential growth phase after 9 hours of growth. All fermentations' design started with an OD_{600} of approximately 0.2 by inoculation with pre-culture, which only contained 10 g/L of glucose, yeast extract, salts and ampicillin.

The aromatic amino acids used in experiments (described above) were prepared weighing 250 mg of each amino acid and dissolved in MiliQ water at a final concentration of 1 mg/mL. Since amino acids cannot be autoclaved (these compounds are degradable at higher temperatures), they were sterilized by 0.22 µm syringe filter into sterile Falcon tubes and stored at 4ºC in absence of light.

2.8. Alkaline cell lysis

Bacterial pellets were obtained from centrifugation of cell broth at 3900 × g for 30 min at 4°C (Beckman Coulter Allegra X22 centrifuge). The 100 mL of bacterial pellet was thawed and dissolved in 10 mL resuspension buffer (50 mM glucose, 25 mM Tris-HCl, 10 mM ethylene-diamine tetraacetic acid (EDTA), pH 8.0). Alkaline lysis was performed by adding 10 mL of a 200 mM NaOH, 1 % (w/v) sodium dodecylsulfate (SDS) solution. After 5 min of incubation at room temperature, cellular debris, gDNA and proteins were precipitated by adding and mixing 8 mL of prechilled 3 M potassium acetate, pH 5.0 (20 min on ice). The precipitate was removed by centrifuging twice at 20,000 × g (30 min, 4°C) with a Beckman Coulter Allegra 25R centrifuge. In this study, analysis and quantification of plasmid were accomplished from the extract of second centrifugation. In these clarified extracts there were no further purification steps.

2.9. Plasmid DNA quantification

The HPLC method based on hydrophobic interaction chromatography [83] was performed to measure concentration and purity of pDNA. In this method was applied a 4.6/100 mm HIC Source 15 PHE PE column (Amersham Biosciences) connected to an ÄKTA Purifier system (GE HealthCare Biosciences). Initially, the column was equilibrated with 1.5 M (NH4)2SO4 in 10 mM Tris-HCl buffer, pH 8.0. The loop of system allowed a 20 µL sample injection with 1 mL/min of flow rate. The absorbance of samples was recorded at 260 nm. The pDNA concentration in each sample was calculated using a calibration curve constructed with pDNA standards (1–100 µg/mL) purified with a commercial Qiagen kit (Hilden, Germany). The purity degree was defined as the percentage of the pDNA peak area related with the total area (area of all peaks on the chromatogram) [41].

2.10. Plasmid DNA quality analysis

The agarose gel electrophoresis experiments were performed using a 1 % agarose gel supplemented with green safe (0.5 µg/mL). Electrophoresis was carried out at 110 V for 30 min in Tris–Acetate–EthyleneDiamine (40mM Tris base, 20mM acetic acid and 1mM EDTA, pH 8.0) (TAE) buffer. The agarose gels were visualized under UV light in a Vilber Lourmat system (ILC Lda, Lisbon, Portugal).

2.11. Metabolites Assessment

The quantification of sugars, acetate and amino acids were performed by collecting a 1 mL of cell culture broth. This sample was centrifuged at 10000 rpm for 3 minutes. The supernatant was filtered through a 0.22 µm syringe filter for further HPLC analysis. The protocols for detection, validation and quantification of each compound are described below.

2.12. Detection, Validation and Assess of Sugars, Acetate and Amino Acids

2.12.1. Amino acids Validation

Amino acids stock solutions were prepared in MiliQ water at a final concentration of 1 mg/mL. For this, 10 mg of each amino acid were weighted and dissolved in MiliQ water, then added to a 10 mL volumetric flask, and filled up with MiliQ water. Tyrosine (only) was added

to 9.5 ml of MiliQ water, followed by 0.5 ml 30% NaOH. Working solutions of tyrosine, phenylalanine and tryptophan were obtained by direct dilution of the standards to final concentrations of 0.01, 0.1 and 1 mg/mL.

The derivatizing reagent (OPA-S) was carefully prepared by following order: 10 mg of OPA, 30 mg of sodium sulfite, 0.25 of ml MiliQ water, 0.25 ml of methanol. This is followed by the addition of 4.5 ml of sodium borate buffer (0.4 M boric acid adjusted to pH 10.4 with 6 M NaOH). The derivatizing reagent was prepared once weekly and stored at room temperature in an amber bottle [84]. All standards and working solutions were kept in glass tubes and stored at 4°C in the absence of light.

2.12.2. Glucose, Fructose, Trehalose and Acetate Validation

Individual stock solution of glucose, fructose and trehalose were prepared in MiliQ water at a final concentration of 200 mg/mL. For this, 2 g of each compound were weighted and dissolved in MiliQ water, then added to a 10 mL volumetric flask, and filled up with MiliQ water. Working solutions of glucose, fructose and trehalose were obtained by direct dilution of the standards to a final concentration of 25 and 100 mg/mL.

Regarding to acetate assessment, individual stock solution of acetate (acetic acid) was prepared in MiliQ water at a final concentration of 10 % (v/v). For this, 20 mL of acetic acid 50 % (v/v) (HPLC grade) was added to a 100 mL volumetric flask, and filled up with MiliQ water. Working solutions of acetate was obtained by direct dilution of the standard to final concentrations of 0.025; 0.1; 0.5 and 5 % (v/v). Standard and working solutions were kept in glass tubes and stored at 4°C.

2.12.3. Chromatographic and detection systems

2.12.3.1. Amino acids analysis

For the amino acids detection an electrochemical coulometric system were used. Therefore, a HPLC system (1260) from Agilent (Waldbronn, Germany), with an auto sampler and quaternary pump was used. The system was coupled to a Coulochem III from ESA (Dias de Sousa S.A., Lisbon, Portugal). Compounds separation was done using a Zorbax 300SB C_{18} reverse phase analytical column (5 µm, 250 x 4.6 mm i.d.). Electrochemical oxidation of the compounds was achieved using a high sensitivity dual electrode analytical cell (5011A) from ESA. Coulometric chromatography systems were controlled by Chemstation software supplied by Agilent Technologies.

Standard reactions were prepared as follows: to 2 ml glass vials were added 10 µl of sample (standard solution, biological sample, or blank) and 40 µl of OPA-S. These were allowed to react for a minimum of 5 min. Afterwards, samples were diluted with 100 µl of HPLC mobile phase and the vials were put on auto sampler. The auto sampler temperatures of coulometric system were kept at 4°C to avoid possible degradation of compounds. Routinely, a 20 µl volume was injected on column. All mobile phases were prepared with MilliQ water to a final volume of 1 or 2 L, filtered through a 0.2 µm pore nylon membrane and degassed for a minimum of 15 min before use. Storage of the mobile phases was done under 4°C, and before use, they were allowed to reach room temperature. For the coulometric system, separation of the compounds was achieved with a mobile phase consisting of 133 mM Na_2HPO4, 0.15 mM EDTA and 27 % Methanol (v/v) adjusted to pH 6.8 with *O*-phosphoric acid. The temperature of the column was set-up 30°C. An isocratic flow of 0.8 mL/min was used with a typical running time of 28 min. The potential of the analytical cell was set at +300 mV in the channel 1 and +650 mV in the channel 2, oxidation occurs in both channels. The range of calibration curve was set of 1 to 750 µg/mL.

2.12.3.2. Glucose, Fructose, Trehalose and Acetate Analysis

For the glucose, fructose, trehalose and acetate detection refractive index system were used. The system was coupled to a Refractive Index Detector (RID) (Agilent 1260 Infinity). Compounds separation was done using a Hi-Plex H ion-exchange analytical column with following dimension: 7.7x300 mm and 8 µm of pore size.

In this system, a 2.5 mM H_2SO_4 solution was used as mobile phase at a flow rate of 0.6 mL/min with a typical running time of 18 min. The temperature of the column was set-up to 50°C. The auto sampler temperatures of refractive index system were kept at 4°C to minimize the degradation of any compound. All mobile phases were prepared with MilliQ water to a final volume of 1 or 2 L, filtered through a 0.2 µm pore nylon membrane and degassed for a minimum of 15 min before use. Storage of the mobile phases was done under 4°C, and before use, they were allowed to reach room temperature. The range of sugars calibration curve was set of 1 to 12 mg/mL and acetate calibration curve was set of 0.001 to 1 % (v/v).

Chapter 3-Results and Discussion

3.1. Monitorization of Fermentation Process

For the purpose of monitoring the fermentation process the two methods (Sugars and Acetate method and amino acids method) were validated according to parameters accepted by the Food and Drug Administration and the International Conference on Harmonization (ICH). For the sugars and acetate method validation, the compounds analysed were glucose, fructose, trehalose and acetate. On the other hand, tyrosine, phenylalanine and trytophan were analysed for the amino acids method validation. The validation was done with a five day validation protocol, evaluating the methods' selectivity, linearity, precision and accuracy.

3.1.1. Selectivity

Selectivity is the capability of a method to differentiate and quantify the analyte in the presence of other compounds in a complex sample. Since the fermentation broth is a complex mixture of compounds the conditions of HPLC and detectors systems were specifically selected to avoid the overlapping of chromatographic signals from different analytes. The selectivity of each method was evaluated by injecting blank (fermentation broth) mixtures of 10 different preparations (n=10) and no interference was observed at the respective retention times of compounds.

3.1.2. Linearity

The calibration curves for each compound (n=5) were obtained using increasing concentrations of glucose, fructose and acetate (sugars and acetate method) and tyrosine, phenylalanine and tryptophan (amino acids method). The data for the calibration curves was obtained by plotting the analyte area versus the analyte concentration.

Calibrators were prepared in the concentrations of 1, 2, 4, 6, 8, 10 and 12 mg/mL (Glucose, Fructose and Trehalose), 0.001, 0.005, 0.01, 0.025, 0.05, 0.1, 0.5 and 1% (v/v) of acetate and 1, 2.5, 5, 15, 50, 100, 200, 500, 750 µg/mL (TYR, PHE and TRYP). Using the linear regression equation of the curve, each calibrator was quantified in order to calculate the mean relative error or accuracy (bias) as shown in the following equation (1):

$$\% \text{ Bias} = \frac{\text{mesaured concentration} - \text{theoretical concentration}}{\text{theoretical concentration}} \times 100 \quad (1)$$

The bias values obtained from the calibration curve of sugars (Glucose, Fructose and Trehalose) were within the accepted parameters, therefore allowing correct quantification of the samples. However, the bias values obtained from the calibration curves of acetate and amino acids didn't allow correct quantification of the samples, since the error values were higher than the accepted values, especially at the lower concentrations. This fact is due to the wide calibration range adopted, and to compensate for heteroscedasticity weighted least squares regressions had to be adopted. Therefore, six weighting factors were evaluated for each analyte: $1/X$, $1/X^2$, $1/Y$, $1/Y^2$, $1/\sqrt{X}$ and $1/\sqrt{Y}$. The lowest relative error was obtained using the weighting factor $1/X^2$ for acetate and $1/Y^2$ for tyrosine, phenylalanine and tryptophan (table 4).

Table 4 - Linear data for Glucose, Fructose, Trehalose, Acetate, Tyrosine, Phenylalanine, Tryptophan.

	Weighting factor	**Calibration range**	**Regression equation**		**R^2**
			m	**b**	
Glucose	Not used	1-12 mg/mL	284767±10915	27984±53952	0.998±0.0009
Fructose			281266±11462	28745±51834	0.998±0.0010
Trehalose			274714±11075	47504±27041	0.998±0.0010
Acetate	$1/X^2$	0.001-1% (v/v)	1473478±76464	68.72±77.85	0.999±0.0015
Tyrosine	$1/Y^2$	1-750 µg/mL	1241.714±182.291	1063.501±419.436	0.993±0.0024
Phenyl-alanine			1218.805±209.809	682.786±522.025	0.993±0.0036
Tryptophan			1463.965±315.229	736.679±727.659	0.992±0.0035

The bias is considered valid if the value is situated between ± 15 % of the nominal concentration, except for the lowest calibrator (LLOQ) which is accepted at ± 20 %. With the application of regression equation analysis using weighting factors, the method showed linearity in the calibration range shown since the R^2 values were superior to 0.99 and the relative error was situated within the accepted intervals.

3.1.3. Within-Run, Precision and Accuracy

The precision and accuracy of a method should be measured using at least five determinations for each concentration [85], and at least three concentrations of the calibration range should be measured (including the highest and lowest concentration) [86].

The precision of an analytical method is defined as the repeatability of individual measurements of an analyte, when these are done repeatedly with multiple aliquots of the same sample, normally expressed by means of the coefficient of variation (% CV) and calculated by the following equation (2):

$$\%\ \mathrm{CV} = \frac{\text{standart desviation of measurements of analyte}}{\text{mean of measurements of analyte}} \times 100 \quad (2)$$

The accuracy describes the difference between the mean measured concentrations for a sample obtained by the linear equation, and the true concentration of the sample. Therefore, accuracy can be expressed as % bias.

Within-run precision was determined using three concentrations (1, 6 and 12 mg/mL) of sugars, three concentrations (0.001, 0.05 and 1% (v/v)) of acetate and four concentrations (1, 100, 500 and 750 µg/mL) of the amino acids. Each concentration was measured on a different day and in replicates of six measurements (table 5).

Table 5- Within-Run Precision an Accuracy (n=6) for all compounds.

	Concentration		Measured concentration		CV (%)	Bias (%)
Glucose	1	mg/mL	1.143±0.014	mg/mL	1,19	16,37
	6		5.785±0.283		4,89	-9,85
	12		12.077±0.084		0,70	0,53
Fructose	1	mg/mL	1.144±0.013	mg/mL	1,13	16,30
	6		5.783±0.283		4,89	-9,90
	12		12.069±0.084		0,69	0,45
Trehalose	1	mg/mL	1.147±0.019	mg/mL	1,66	16,19
	6		5.786±0.285		4,92	-9,88
	12		12.066±0.075		0,63	0,51
Acetate	0,001	%	0.001±1.54E-05	%	1,53	0,49
	0,05		0,053±3.08E-04		0,58	5,77
	1		1,004±7.66E-03		0,76	0,37
Tyrosine	1	µg/mL	0.96±0.04	µg/mL	3,73	-3,47
	100		91.62±4.77		5,20	-7,80
	500		475.09±38.86		8,18	5,33
	750		706.20±42.65		6,04	-5,89
Phenylalanine	1	µg/mL	1.01±0.06	µg/mL	5,57	7,94
	100		94.61±5.55		5,86	-5,07
	500		453.41±28.92		6,38	-0,86
	750		727.29±47.64		6,55	7,93
Tryptophan	1	µg/mL	0.92±0.06	µg/mL	6,08	-5,93
	100		93.63±8.89		9,50	-12,39
	500		474.30±39.65		8,36	5,30
	750		819.94±19.78		2,41	11,52

The CV obtained from the within-run precision was lower than 10 % for all analytes. All bias values were situated within ± 15 % of the nominal concentration, except at the LLOQ where they situated within ± 17 %.

3.1.4. Between-Run, Precision and Accuracy

The between-run precision was measured using seven concentrations for sugars, eight concentrations for acetate and nine concentrations for amino acids, during a five day period (table 6).

The precision is evaluated using two approaches, within-run and between-run repeatability. In the first approach, the precision during a single analytical run is assessed, while the latter

assesses precision with time and may involve different analysts, equipment, materials and laboratories [85].

Table 6- Between-Run Precision and Accuracy (n=5) for all compounds: Glucose, Fructose, Thehalose, Acetate, Tyrosine, Phenylalanine, Tryptophan.

	Concentration		Measured concentration		CV (%)	Bias (%)
Glucose	1	mg/mL	0.965±0.133	mg/mL	13,76	-3,50
	2		1.945±0.058		3,00	-2,71
	4		4.059±0.091		2,25	1,49
	6		6.028±0.153		2,54	0,47
	8		8.102±0.231		2,85	1,27
	10		9.938±0.136		1,37	-0,62
	12		11.961±0.167		1,39	-0,32
Fructose	1	mg/mL	0.963±0.129	mg/mL	13,45	-3,70
	2		1.946±0.058		2,99	-2,71
	4		4.057±0.093		2,29	1,42
	6		6.034±0.143		2,36	0,58
	8		8.100±0.229		2,83	1,25
	10		9.943±0.134		1,34	-0,57
	12		11.956±0.152		1,27	-0,36
Trehalose	1	mg/mL	0.960±0.127	mg/mL	13,21	-3,96
	2		1.941±0.048		2,49	-2,96
	4		4.069±0.088		2,17	1,74
	6		6.038±0.138		2,28	0,64
	8		8.086±0.241		2,98	1,07
	10		9.948±0.131		1,31	-0,52
	12		11.957±0.155		1,30	-0,36
Acetate	0,001	%	0.001±8.651E-06	%	0,86	0,72
	0,005		0.005±1.737E-04		3,57	-2,83
	0,01		0.009±3.699E-04		3,80	-2,74
	0,025		0.025±1.228E-03		4,86	0,81
	0,05		0.050±7.946E-04		1,59	0,05
	0,1		0.101±1.159E-03		1,15	1,16
	0,5		0.502±1.700E-02		3,38	0,41
	1		1.017±2.125E-02		2,09	1,63

Tyrosine	1	µg/mL	0.948±0.055	µg/mL	5,79	-5,82
	2,5		2.555±0.220		8,60	1,57
	5		5.392±0.297		5,51	7,04
	15		15.870±0.950		5,98	5,21
	50		53.811±0.944		5,47	6,86
	100		100.580±6.723		6,68	0,23
	200		200.806±8.666		4,32	0,25
	500		490.939±31.059		6,33	-2,16
	750		679.571±14.366		2,11	-10,40
Phenyl-Alanine	1	µg/mL	0.964±0.040	µg/mL	4,17	-3,91
	2,5		2.654±0.161		6,08	5,54
	5		5.289±0.405		7,66	5,01
	15		15.557±0.797		5,12	3,38
	50		51.502±2.957		5,74	2,66
	100		104.373±5.755		5,51	3,96
	200		196.542±14.526		7,39	-2,20
	500		493.215±34.202		6,93	-1,77
	750		693.170±43.302		6,25	-8,53
Tryptophan	1	µg/mL	1.012±0.073	µg/mL	7,24	0,73
	2,5		2.549±0.257		10,07	1,14
	5		4.981±0.443		8,88	-1,03
	15		15.154±1.554		10,25	0,19
	50		51.116±4.691		9,18	1,49
	100		101.637±8.028		7,90	1,13
	200		193.898±15.286		7,88	-3,66
	500		522.166±33.813		6,48	3,91
	750		765.568±69.404		9,07	1,42

The CV obtained from the between-run precision was lower than 14 % for all analytes. All bias values were situated within ± 10 % of the nominal concentration

3.1.5. Stability

The stability of target compounds was assessed through of three different concentrations which were subjected to three temperature conditions. The concentrations of sugars were 3, 5 and 9 mg/mL, of acetate were 0.0075, 0.075, 0.25 % and amino acids were 65, 350 and 650 µg/mL. In the case of temperature conditions, the vials (or "controls") were put overnight in auto-sampler of the HPLC system (assess the stability with auto-sampler temperature), in lab bench (the effect of room temperature) and in the freezer (followed by three cycles of

freeze-thaw). The stability was calculated dividing the average of the controls (submitted to conditions) signal with the average of the fresh daily control and was expressed in percentage. The stability values are shown in tables 7, 8 and 9.

Table 7- Stability at three different concentrations (n=3) for Trehalose (TREH), Glucose (GLUC) and Fructose (FRUC).
(T_{room} = 22ºC; $T_{auto\text{-}sampler}$= 4ºC).

	Stability (%)								
	T_{room}			**$T_{auto\text{-}sampler}$**			**Freeze-Thaw cycles**		
	Concentration (mg/mL)			**Concentration (mg/mL)**			**Concentration (mg/mL)**		
	3	5	9	3	5	9	3	5	9
TREH	100,19%	100,33%	100,25%	99,68%	98,08%	99,38%	110,18%	111,87%	108,63%
GLUC	100,05%	100,29%	100,20%	99,56%	97,84%	99,29%	111,20%	112,29%	109,66%
FRUC	100,03%	100,20%	100,05%	99,56%	97,80%	99,19%	111,17%	112,52%	109,68%

Table 8- Stability at three different concentrations (n=3) for Acetate (ACET).
(T_{room} = 22ºC; $T_{auto\text{-}sampler}$= 4ºC).

	Stability (%)								
	T_{room}			**$T_{auto\text{-}sampler}$**			**Freeze-Thaw cycles**		
	% (v/v)			% (v/v)			% (v/v)		
	0,0075	0,075	0,25	0,0075	0,075	0,25	0,0075	0,075	0,25
ACET	101,30%	100,38%	100,37%	99,41%	98,42%	105,30%	115,81%	106,58%	111,52%

Table 9- Stability at three different concentrations (n=3) for Tyrosine (TYR), Phenylalanine (PHE) and Tryptophan (TRP) (T_{room} = 22ºC; $T_{auto\text{-}sampler}$= 4ºC).

	Stability (%)								
	T_{room}			**$T_{auto\text{-}sampler}$**			**Freeze-Thaw cycles**		
	Concentration (µg/mL)			**Concentration (µg/mL)**			**Concentration (µg/mL)**		
	65	350	650	65	350	650	65	350	650
TYR	62,05%	61,25%	66,55%	75,02%	64,18%	82,24%	99,28%	94,96%	97,38%
PHE	71,48%	69,73%	75,65%	80,15%	67,33%	84,97%	93,13%	83,81%	84,26%
TRP	82,98%	75,49%	78,91%	74,55%	60,34%	76,98%	72,35%	78,34%	79,72%

The stability data shows that sugars and acetate have superior stability than amino acids. This was expected because the characteristics of the target amino acids (eg, hydrophobicity) and also the derivatizing hands-on becomes the molecule more unstable. However, it is found that the decrease of temperature leads to an increase of stability of all compounds.

Moreover, the freeze-thaw cycle not significantly affects the samples, which could be an advantage to further storage.

The results of both validation methods showed a good capacity to assess any target compound. Additionally, analyzing the within and between-run of each method is verified an acceptable range of CV and bias. Further, the stability experiments allowed understand how temperature would affect either on pre-analysis either in samples storage.

Moreover, this information confirms the robustness of methods and consequently a valid offline monitoring.

3.2. The effect of nutritional media composition in *E. coli* VH33 growth and plasmid DNA yield

3.2.1. Overall analysis of nutritional variation effect in bacterial growth and plasmid DNA yield

Over the last years, several plasmid DNA production processes have been proposed [4, 87, 89]. Additionally, the influence of medium composition has important role on overall process performance. Thus, is relevant assess the bacterial growth behaviour, plasmid DNA yield and purity in different broth fermentation formulations. The composition of each fermentation media are shown in table 10.

The selection of the composition of fermentation media was based on TB medium. In the literature, several results revealed the potential of this medium for plasmid DNA biosynthesis [48, 59, 60]. Since the bacterial strain applied in this work has a modified mechanism of glucose uptake, it was important to introduce the carbon source in the medium. However, for concentrations above 50 g/L glucose may cause toxicity to the cells [89]. To overcome this barrier was used 5 and 10 g/L of glucose. In addition to the effect of glucose, the presence or absence of yeast extract and glycerol allowed to evaluate if there is a reduction of contaminants (DNA and RNA) and consequently if a promoting in purity of plasmid levels can be achieved.

Table 10 - Nutrients composition of the six different media fermetations.

Media	Nutrient			
	Glucose (g/L)	Tryptone (g/L)	Yeast Extract (g/L)	Glicerol (mL/L)
Terrific Broth (TB)	-	20	24	4
A_1	5	20	24	4
B_1	10	20	24	-
C_1	10	20	24	4
D_1	10	20	-	4
E_1	10	20	-	-

Concerning to cell growth, the Terrific Broth and B_1 media showed the highest biomass concentration, 3.177 and 2.997 g/L respectively, than the remaining four fermentations. Conversely, the lower growth (0.756 g/L) was obtained on medium E_1, showing weak hypotheses to be a medium select for plasmid production. The results suggest that lack of yeast extract and conjugation among tryptone and glucose do not promote bacterial growth. Also the substitution of glycerol to glucose medium B_1, shows a small change in final cell growth (table 11). Indeed, the slight decrease in cell concentration occurs probably because glucose depletion leading the cells to enter stationary phase.

Table 11 - Cellular growth of the six different fermentations (1 OD_{600}= 0,45g/L of Cell Dry Mass).

	OD_{600} (14 h of fermentation)	**Cell Dry Mass (CDM) (g/L)**
Terrific Broth (TB)	7.06	3.177
A_1	5.38	2.421
B_1	6.66	2.997
C_1	5.07	2.282
D_1	1.89	0.8505
E_1	1.68	0.756

Also, the results revealed that the introduction of glucose in Terrific Broth, which corresponds to medium A_1, leads to a reduction in final cell growth. Additionally, the increased glucose concentration of 10 g/L (medium C_1), limited, even more, cell growth of this strain. In the literature it was been showed that the presence and internalization of two carbon sources causes cell stress once the cell machinery is overloaded [45]. In spite of metabolic pathways of glucose and glycerol are similar, the route of glycerol in the glycolysis pathway is typically smaller than glucose [43, 45]. Moreover, in the presence of both carbon sources may occur a diauxic response being firstly consumed all the glucose presents in the medium and then the glycerol. This fact may explain the decrease of CDM verified between the B_1 and C_1 experiments.

According to this experimental data, it was demonstrated that the presence of nitrogen source (yeast extract and tryptone) significantly influence *E. coli* VH33 growth. Comparing medium B_1 to E_1 it is apparent that the yeast extract provides more favorable conditions for growth. Moreover, the medium E_1 shows that the combination of tryptone and glucose does not promote growth (results in accordance with Losen and co-workers) [53], leading to the conclusion that in the case of using glucose as the main carbon source should be used yeast extract as a nitrogen source. This good synergy mostly happens due to a complete nutrient composition of yeast extract, which provides the amino acids, peptides, lipids and other nutrients to cell demands.

In sum, for a suitable cell growth is necessary select the composition fermentation medium, in order to maximize the production of the target product. In this case, *E. coli* VH33 requires for its growth only one source of carbon, especially glucose, but also with appropriate nitrogen source such as yeast extract.

3.2.1.1. Plasmid DNA yield and purity

In addition to the influence on cellular growth is even more important to understand how the changes affect the yield and purity of the target product. The results for the plasmid DNA volumetric and specific yields and purity are shown in table 12.

Table 12- Effect of different nutrients composition on plasmid DNA volumetric and specific yields, and purity.

	Peak Area			Concentration		Purity
Media	pDNA	gDNA	RNA	pDNA volumetric yield (µg/mL)	pDNA specific yield (µg pDNA/ mg CDM)	Peak Area
Terrific Broth (TB)	52.24±0.26	143.95±1.67	709.26±6.74	102.80±2.54	32.36	5.76
A_1	33.31±2.29	174.21±3.84	519.18±35.65	66.28±7.66	27.38	4.58
B_1	50.08±0.44	214.85±2.69	608.32±0.31	98.63±2.88	32.91	5.74
C_1	30.52±1.44	202.69±6.28	739.80±16.61	60.90±4.81	26.69	3.14
D_1	15.12±1.55	84.04±10.43	152.59±8.31	31.21±5.03	36.69	6.01
E_1	9.320±0.06	76.74±1.07	78.40±4.86	20.01±2.16	26.47	5.68

In concordance with the literature, the medium TB presents the best results in the production plasmid DNA [33, 53, 90]. In respect of other the five media, the medium B_1 showed excellent results, unlike medium E_1 which showed a low production of plasmid DNA. A more detailed analysis of the results described, it was found that the introduction of glucose in medium A_1 leads to a marked decrease in volumetric efficiency and specific yield of plasmid DNA, 66.26 µg/mL and 27.38 µg pDNA/mg cell dry mass (CDM) respectively. Less significant is the reduction in purity to 4.58, by decreasing RNA but especially by lower presence of plasmid DNA. Additionally, C_1 discloses a similar behavior with medium A_1, but in this case with slight increase in RNA biosynthesis. These data reinforce the idea that the presence of two carbon sources is harmful to cell growth and for production of plasmid DNA. Additionally, the increased glucose concentration enhances the production of RNA, raising the metabolic flux of *E. coli* to uptake this carbon source [45].

The results for the tests D_1 and E_1 were considered relevant because they showed a good level of purity 6.01 and 5.68, respectively. The volumetric pDNA yield was quite low in both media and the plasmid specific yield seems contrary, the medium D_1 obtained 36.63 µg/mg (highest) and E_1 obtained 26.47 µg/ mg (the lowest). The information provided by these tests showed that the absence of yeast extract from the medium had a strong impact on production and consequent in plasmid yield. Therefore the presence of a potent source of nitrogen such as

yeast extract becomes crucial to develop an experimental design approach for the production of pcDNA3–FLAG–p53 in *Escherichia coli*.

Finally, the medium B_1 was found to be promising for process optimization in the production of the pDNA. The results of the fermentation medium B_1 with *E.coli* VH33 showed a volumetric and specific yield of 98.63 µg/mL and 32.91 µg/mg, respectively. The level of purity was acceptable, showing that exchange of glycerol by glucose leads to a slight increase in the load of genomic DNA. This is due to increased metabolic activity, especially for targeted cell growth [45].

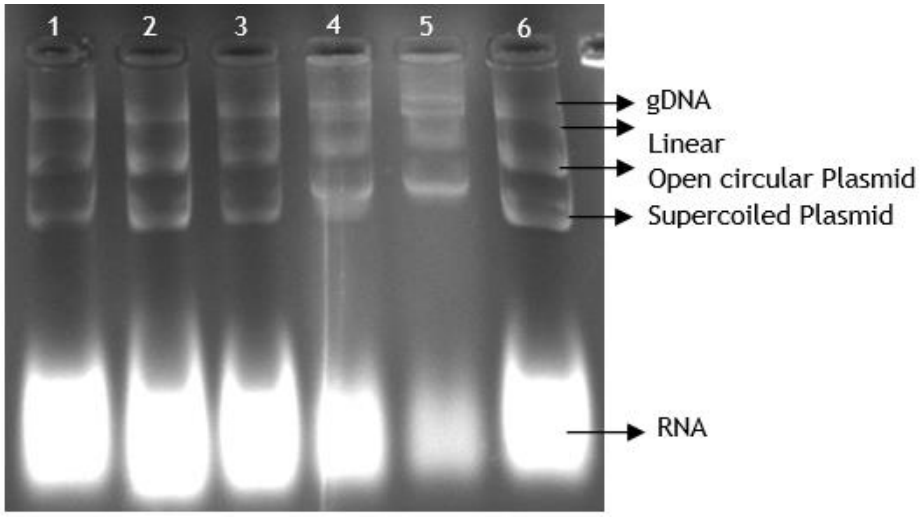

Figure 10- Agarose Gel Electrophoresis. The lane 1,2,3,4,5 and 6 correspond to lysates extracts from media A1,B1,C1,D1,E1 and TB, respectively.

After the upstream stage, the quality of the obtained plasmid is an important factor to be considered for the choice of a sustainable fermentation medium formulation. The agarose gel electrophoresis (figure 10) shows that the medium B_1 and TB produce more plasmid in supercoiled form (i.e. the most active form) than the remaining experiments. Further, the increase in cell density leads to an increase in low molecular weight RNA due to the needs of each cell to support its cellular integrity.

In contrast, the fermentation media D_1 and E_1 showed a poor quality and low yield of SC plasmid. This data supports the idea that definitely they were not a good option for future optimizations. Although, the conditions of these media led to a lower yield of contaminants, mainly RNA, but is a direct consequence of lower growth.

With the chromatographic system used was not possible verify and quantify which plasmid isoforms were major from distinct medium manipulation [31]. However, with the analysis of gel electrophoresis (figure 10) it can be observed the presence of three isoforms. The

topologies linear and circular-open occur as a result of degradation of lysis process but also because of degradation by extrinsic (fermentation) and intrinsic (nuclease) factors to the cell [25, 34].

3.2.2. The influence of different conditions of carbon source in the B_1 medium

The behavior of bacterial growth and production of bioproducts is closely tied with the formulations used during fermentation. In this case, the presence of a carbon source with alternating different concentrations could help to understand the best conditions for pcDNA3-FLAG-p53 biosynthesis in *E. coli* VH33.

Table 13- The effect of different carbon sources on DNA volumetric and specific yields, and purity. (All the media fermentations had 20 g/L tryptone, 24 g/L yeast extract, salts solution, ampicillin)

Medium	**Peak Area**			**Concentration**			**Purity**
Carbon source	pDNA	gDNA	RNA	Cell Dry Mass (g/L)	pDNA volumetric yield (µg/mL)	pDNA specific yield (µg pDNA/ mg CDM)	Peak Area Ratio (%)
Glucose 5 g/L	13.08±0.10	153.90±2.07	483.04±8.66	4.77	27.28±2.24	5.72	2.01
Glucose 7.5 g/L	19.10±0.33	173.37±1.28	472.17±2.32	4.18	38.89±2.69	9.30	2.87
Glucose 20 g/L	18.15±0.40	201.60±8.98	562.32±47.61	5.84	37.06±2.83	6.34	2.32
Glucose 30 g/L	27.20±0.20	208.46±1.84	640.70±1.05	5.23	54.52±2.43	10.43	3.10
Glycerol 10 mL/L	13.08±0.12	145.21±0.38	437.67±2.76	3.68	27.27±2.28	7.41	2.19
Glycerol 20 mL/L	14.97±0.09	153.35±0.43	501.22±2.43	3.71	30.92±2.23	8.33	2.24
Glycerol 30 mL/L	10.56±0.10	147.51±0.90	498.99±3.42	3.72	22.41±2.24	6,03	1.61

Analyzing the data from table 13, the relationship between the concentration of the carbon source and the production of plasmid it was not entirely straightforward. In the presence of glucose, for the concentration of 5 g/L to 30 g/L, there is an increased production of plasmid. However, the specific yield and purity of the plasmid are committed to growth rate and with

desired CDM. According to literature, the increase of glucose concentration leads to a higher final biomass concentration but decreases the specific growth rate (μ) [68]. This evidence supports the CDM data, since an increase of glucose from 7.5 to 20 g/L promote an increment in final cell concentration. The change of 20 g/L concentration of glucose to 30 g/L leads to a further increase of plasmid and "contaminants" gDNA and RNA levels. This effect is due to the amount of glucose available in the fermentation medium that is required for several host metabolic pathways (e.g. glycolysis and TCA cycle) [43].

As described in the literature [45, 64, 89], for the same period of fermentation when the glycerol is used, instead of glucose, the kinetics of cell growth is much reduced which result in a decrease of final cell concentration [45]. Additionally, an increase in glycerol levels does not result in highest CDM, therefore the concentrations used are perhaps on saturation limit levels. Furthermore, the increase of 20 mL/L to 30mL/L of glycerol translates into a loss of plasmid concentration and revealed that these conditions are unsuitable for its sustainable during cell generation.

Another important factor that can affect the previous results is the choice of pre-inoculums. In this case all fermentations started with a LB medium pre-culture and thus the bacteria had to adapt to new conditions during the fermentation trials. The new conditions were especially the presence of a carbon source, which induce a change in the cellular nutrient uptake system. It is clear that adaptation is faster and more efficient with glucose than glycerol. This happens due to gene alteration ($PTS^{-}GALP^{+}$) on the VH33, which conduct the glucose to inside the cell and for a central metabolism, according to Meza and co-workers [42]. So in this case, it was possible to achieve better cellular and plasmid yields.

The result obtained suggests that for a possible optimization and scale-up plasmid biosynthesis with *E. coli* VH33 would be preferential to use glucose as the carbon source. The advantage of this gene facilitates the incorporation and degradation of glucose in central carbon metabolism. So, it allows working with non-limiting concentration of substrate without development of a significant metabolic stress due to acetate inhibition [42].

3.3. Optimization of aromatic amino acids pathway to enhance plasmid DNA biosynthesis

3.3.1. Preliminary Assays of Box-Behnken Fermentation Design

A real-time monitoring of the fermentation process provides information that often would not be detected and can help improve the production of the bioproduct.

Initial assays allowed to identify how the cells responded in the presence of the three aromatic amino acids in presence only of glucose, as the main carbon source (a typical defined medium formulation). The table 14 shows the concentration of each analyte over fermentation time. It is possible to check the variation that occurs from the beginning to the end of the fermentation complementing the remaining experimental data. Although these data attended to outline and setup the strategy for construction of an experimental design, they have proven the potential of previously validated analytical methods depicted of results and discussion section.

Table 14- Quantification of analytes from defined medium (10 g/L glucose, aromatic amino acids at 50 µg/mL) during fermentation process (n=5).

Analytes	**Concentration of analytes in medium**		
	Time of fermentation (h)		
	0	7	16
Tyrosine (µg/mL)	41.25±6.43	24.66±3.25	8.35±1.35
Phenylalanine (µg/mL)	42.8523±4.85	24.19±2.30	9.10±1.56
Tryptophan (µg/mL)	39.22±5.12	31.35±2.30	24.04±2.08
Glucose (g/L)	10.69±0.23	9.06±0.17	5.46±0.18
Acetate (g/L)	0.003±0.001	0.038±0.002	1.02±0.056

In order to facilitate the interpretation of the previous table, the figure 11 clearly shows an accentuated depletion rate of the amino acids but at the same time the formation of by-products such as acetate. Additionally, to improve the analysis of the process should be collected the samples with lower interval time allowing calculate the rates of consumption and/or production more accurately (unfortunately, this approach was not done in study).

Additionally, the results show a greater internalization of tyrosine and phenylalanine than the tryptophan. These data can be explained by enzymes (transporter) which enable their uptake. In case of tryptophan, it is necessary the presence of three enzymes, while for phenylalanine and tyrosine are just two enzymes. On the other hand, of three aromatic amino acids, tryptophan is the only one that is degraded by *E. coli* (neither tyrosine nor phenylalanine serves as a carbon or nitrogen source) which may explain the difference of remaining aromatic amino acids levels in fermentation broth [91].

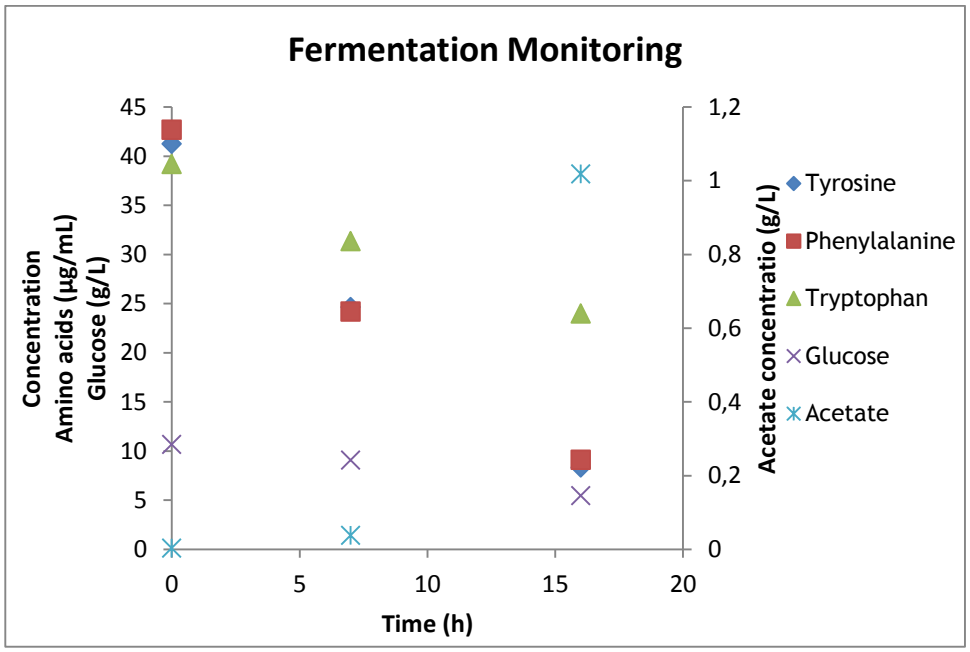

Figure 11- Aromatic amino acid glucose and acetate profiles from a fermentation using a defined medium formulation.

3.3.2. Box-Behnken Fermentation Design

In order to get the best performance during fermentation, all the factors involved need to be optimised. The classical method of optimization involves varying one factor at a time and keeping the others constant. To facilitate this typical process is commonly used the experimental design. This methodology allows reduce time and necessary experimental assays.

For the development of this work, the Box-Behnken factorial experimental design employed had three independent variables, aromatic amino acids (tyrosine, phenylalanine and tryptophan) (table 15). As described above, several enzymes of aromatic amino acid biosynthetic pathway are subject to feedback regulation. As reported in literature, the range concentration of aromatic amino acid enough to perturb their pathway, is approximately 15 to 500 µg/mL [91, 92]. In order to avoid the feedback inhibition caused by the introduction of three amino acids we applied lower aromatic amino acids concentrations, of approximately 10, 50 and 100 µg/ml.

Table 15 - Amino acids levels for Box-Benhken Design

Level	Amino acid concentration (µg/mL)
-1	10
0	50
1	100

The independent variables (TYR, PHE, TRP) was studied at three levels (1, 0, and +1), with 17 experimental runs (table 16). The condition of each run provides the development of Box-Behnken design.

Table 16- Runs of Box-Benhken Design.

Run	Factor 1 A: TYR (µg/mL)	Factor 2 B: PHE (µg/mL)	Factor 3 C: TRP (µg/mL)
1	1	-1	0
2	1	-1	0
3	**1**	**0**	**1**
4	0	-1	-1
5	0	0	0
6	**1**	**0**	**-1**
7	0	-1	1
8	-1	1	0
9	-1	0	-1
10	0	0	0
11	-1	-1	0
12	0	0	0
13	0	1	-1
14	**0**	**1**	**1**
15	0	0	0
16	0	0	0
17	-1	0	1

To complement the experimental design all fermentation (runs) was monitored. This monitorization involves the analysis of specific growth rate, final mass of cells, specific pDNA yield and purity (Table 17). The values of volumetric pDNA yield acted as an output (response) to the BBD. The parameters related to specific pDNA yield and purity could be considered as responses to the model development. However, none of these outputs were used due to the lack of statistical significance presented.

The criterion of purity in accordance with the specific yield of pDNA has special significance for the best interpretation of the data. From analysis of the table 16 it was showed that the best results for purity correspond to the runs where specific pDNA yield values were higher (Run 3, 6 and 14). These results are related to specific growth rate (µ) values of 0.3, 0.33 and 0.37 h^{-1} respectively. As mentioned in literature, the phenylalanine seems to be the factor that mostly affects the *E. coli* growth. However, with simultaneously addition of tyrosine and tryptophan the cell growth rate is not affected [91].

The analysis of the interaction between aromatic amino acids applied in the runs will aid a better understanding of these results, which will be described below.

Table 17- Fermentation runs results of Box-Benhken Design.

Run	Specific Growth Rate $\mu(h^{-1})$	Cell Dry Mass (g/L)	pDNA volumetric yield (µg/mL)	pDNA specific yield (µg pDNA/mg CDM)	Purity (%)
1	0,22	2,40	70,17	29,25	3,20
2	0,35	2,13	74,19	34,86	3,74
3	**0,30**	**1,46**	**78,72**	**53,83**	**4,36**
4	0,37	1,64	75,28	45,83	3,73
5	0,44	2,11	78,47	37,26	3,84
6	**0,33**	**1,63**	**85,88**	**52,72**	**4,18**
7	0,33	1,57	66,44	42,37	3,46
8	0,36	1,53	75,27	49,20	3,76
9	0,37	1,58	74,12	46,79	3,37
10	0,48	2,28	78,78	34,60	3,77
11	0,28	1,62	75,62	46,68	3,51
12	0,47	2,22	77,93	35,06	3,84
13	0,28	1,65	76,62	46,46	3,76
14	**0,37**	**1,53**	**82,59**	**54,06**	**4,15**
15	0,47	2,25	77,49	34,51	3,90
16	0,35	2,17	78,86	36,28	3,86
17	0,36	1,73	81,49	47,04	3,82

The analysis of run by agarose gel electrophoresis, as seen in figures 12 and 13, indicated a slight presence of two plasmid isoforms, especially supercoiled form. However, due to the low quality of electrophoresis is not possible to detect significant changes between runs of the experimental design performed. These gels were stained with *green safe* instead of ethidium bromide, which compromised the analysis of this test.

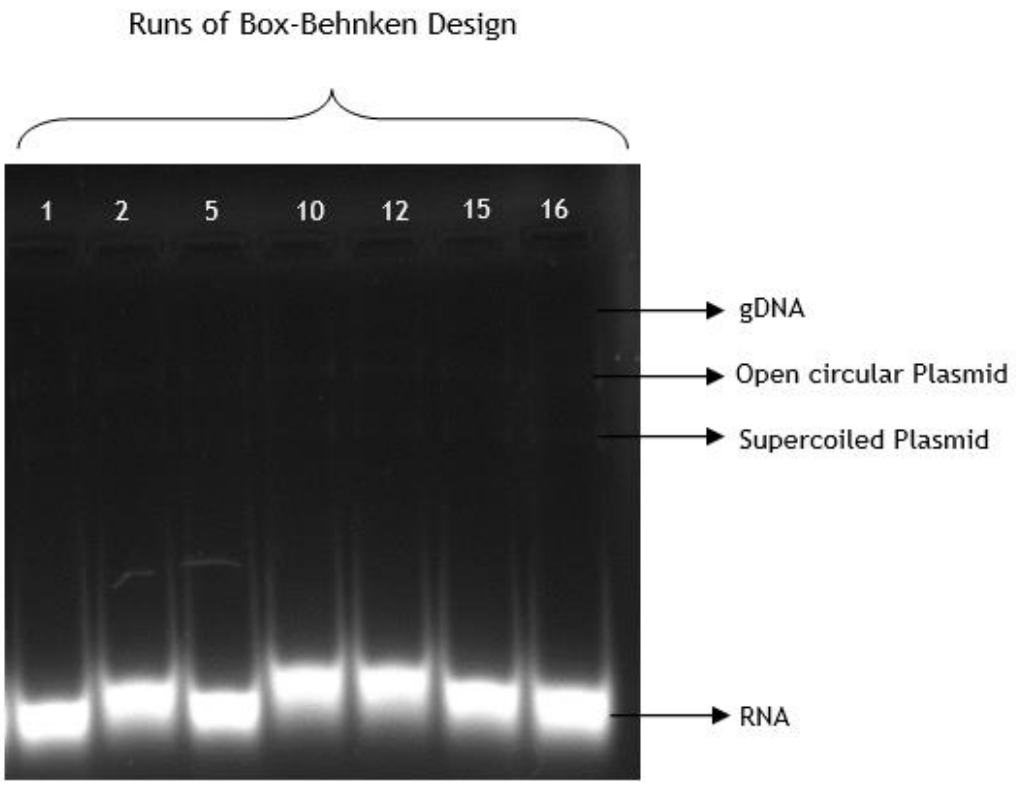

Figure 12- Agarose Gel Electrophoresis. The lanes 1, 2, 5, 10, 12, 15, 16 are runs of Box-Behken Design.

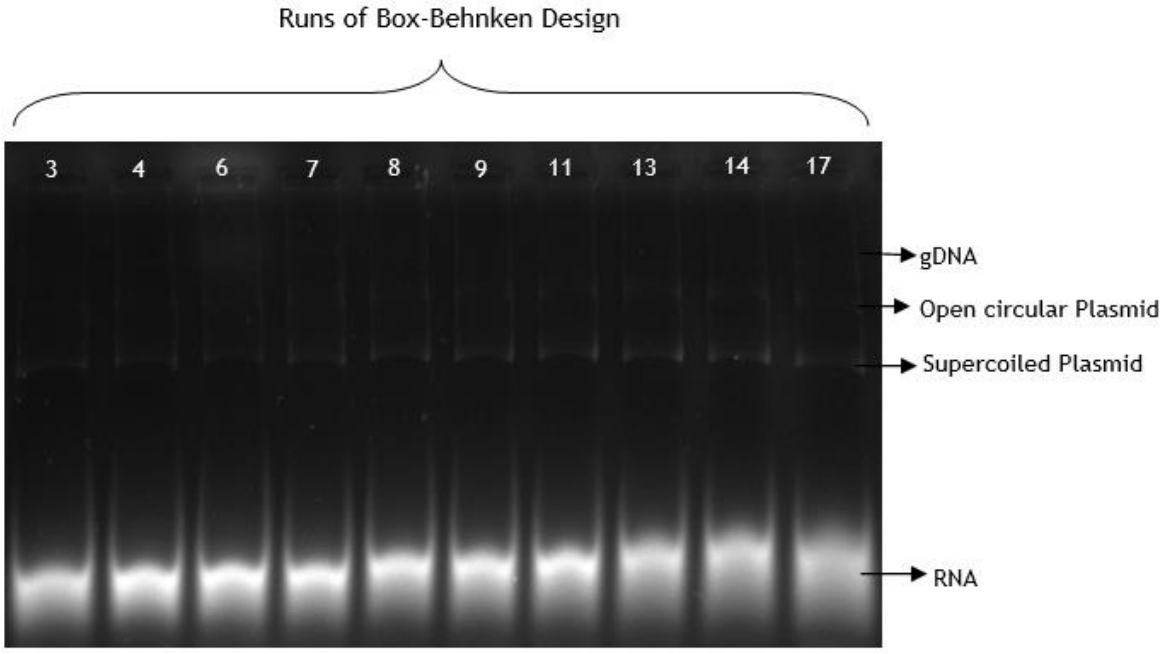

Figure 13- Agarose Gel Electrophoresis. The lanes 3, 4, 6, 7, 8, 9, 11, 13, 14, 17 are runs of Box-Behken Design.

3.3.2.1. ANOVA Statistical Analysis

For the Box-Behnken design, seventeen runs were performed and the results of the statistical experiment were analyzed with regard to the coded design matrix (table 15). To test the fit of the model, the regression equation (3) and determination coefficient (R^2) were evaluated the ANOVA data (figure 14).

$$\text{pDNA} = 78.304 + 0.308.\text{Tyr} + 2.647.\text{Phe} - 0.330.\text{Trp} + 1.093.\text{Tyr.Phe} - 3.631.\text{Tyr.Trp} + 3.704.\text{Phe.Trp} + 0.164.\text{Tyr}^2 - 4.656.\text{Phe}^2 + 1.584.\text{Trp}^2 \quad (3)$$

The model *F* value of 3.70 implies that the model is significant and there is only a 4.91% chance that a model *F* value could occur due to noise. The analysis of variance (ANOVA) (figure 13) quadratic regression model demonstrates that the model was significant for pDNA production with the low probability (P < 0.0491) of the *F* test. The fitted model is considered adequate if the *F* test is significant (P < 0.05) [93].

ANOVA for Response Surface Quadratic Model

Analysis of variance table [Partial sum of squares - Type III]

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F	
Model	268.95	9	29.88	3.70	0.0491	significant
A-TYR	*0.76*	*1*	*0.76*	*0.094*	*0.7678*	
B-PHE	*56.08*	*1*	*56.08*	*6.95*	*0.0336*	
C-TRP	*0.87*	*1*	*0.87*	*0.11*	*0.7518*	
AB	*4.78*	*1*	*4.78*	*0.59*	*0.4669*	
AC	*52.73*	*1*	*52.73*	*6.53*	*0.0378*	
BC	*54.87*	*1*	*54.87*	*6.80*	*0.0350*	
A^2	*0.11*	*1*	*0.11*	*0.014*	*0.9091*	
B^2	*91.29*	*1*	*91.29*	*11.31*	*0.0120*	
C^2	*10.56*	*1*	*10.56*	*1.31*	*0.2902*	
Residual	56.49	7	8.07			
Lack of Fit	*55.14*	*3*	*18.38*	*54.21*	*0.0011*	*significant*
Pure Error	*1.36*	*4*	*0.34*			
Cor Total	325.44	16				

Std. Dev.	2.84
Mean	76.94
C.V. %	3.69
R-Squared	0.8264
Adj R-Squared	0.6032

Figure 14- ANOVA for the quadratic response surface model fitting to the plasmid DNA biosynthesis.

Another important value is lack-of-fit test that was performed by comparing the variability of the current residual model to the variability between observations at replicate settings of the factors. The lack of fit is designed to determine whether the selected model is adequate to describe the observed data or whether a reformulation of the model should be applied. Despite the lack-of-fit value was significant, a contribution of other values such as coefficient variation, standard deviation and R^2 should support the validity of model.

Therefore the determination coefficient (R^2 =0.83) being a measure of goodness of fit to the model, indicates an acceptable degree of correlation between the observed value and predicted values. The coefficient of variation (CV), ratio of the standard error of estimate to the mean value of the observed response, is a measure of reproducibility of the model; generally, a model can be considered reasonably reproducible if its CV is not greater than 10% [93]. Hence, the low variation coefficient value (CV= 3.69%) obtained indicates a high precision and reliability of the experiments (figure 14).

The significance of regression coefficients was considered at an impact level of 95%. The *p*-values of the regression coefficients suggest that among the independent test variables; linear, quadratic, and interaction effects of tyrosine, phenylalanine and tryptophan are highly significant (according to ANOVA data). In this study, B, AC, BC, B^2 are significant model terms and had a potential effect on pDNA volumetric yield.

Moreover, it is observed that phenylalanine concentrations exerted more pronounced linear positively effect on pDNA production. The quadratic effect of tryptophan showed a good increment on plasmid production, unlike the phenylalanine.

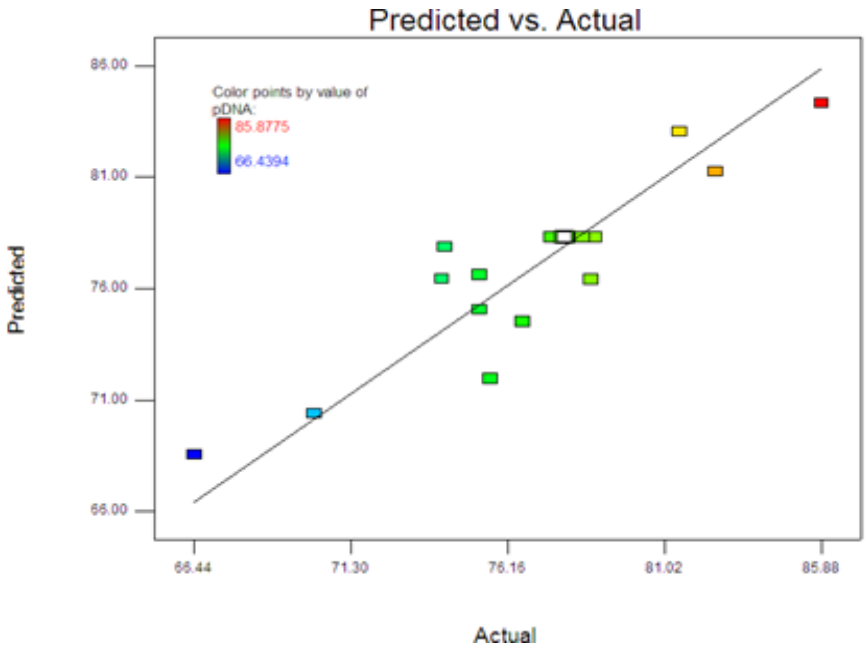

Figure 15- Predicated values versus observed values plot.

For a multiple regression model, rather than plotting the observed and predicted values versus each of the independent variables in turn, it is simpler to generate a single plot of observed values versus the predicted values, in order to verify linearity (Predicted vs Observed plot). Therefore, for continuous responses the actual versus predicted plot (figure 15) shows how well the model fits the data. The diagonal line represents of where predicted and observed values are the same. For a perfect fit, all the points would be on this diagonal. As expected, the best points belong to central point because it is from these that the whole model is developed. The furthermost points of the line have higher residual value than those are closer to the line. Additionally, the observation value with large residual is called outlier. In other words, it is an observation whose dependent-variable value is unusual given, its values on the predictor variables. An outlier may indicate a sample peculiarity or may indicate a data error or other problem. Analyzing the figure 15, seems no exist a significant observed value that could be considered an outlier.

3.3.2.2. Influence of variable interaction on Plasmid pcDNA3-FLAG-p53 DNA Production

In experimental designs test, if a variable influences another, this influence is called "effect". There are two different effects: the variable effects on another directly or via an interaction (or uses both mechanisms simultaneously) [81]. The following figures indicate the existence of interaction or not of interaction among the factors analyzed. Once, the graphic lines (red and black) are not parallel can be asserted that there is always interaction between factors. The interaction AB and AC is linear while BC interaction is non-linear.

Interaction AB

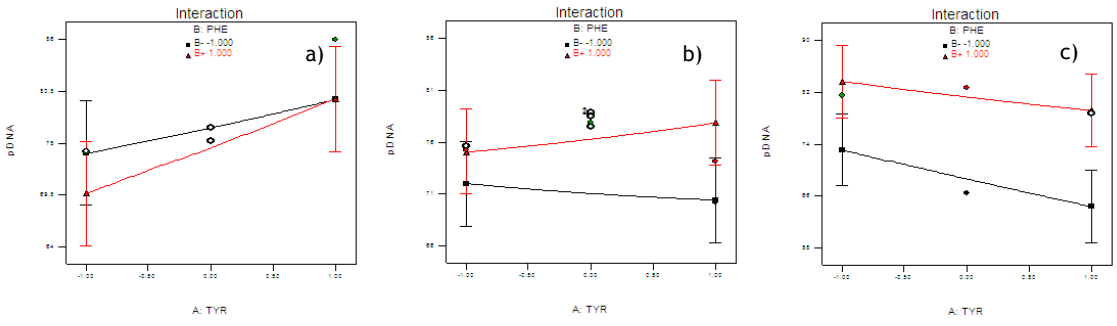

Figure 16- Interaction between Tyrosine and Phenylalanine plot.

The figure 16 (a) shows the interaction between AB factors with tryptophan at level -1. Basis on these data, an increase of tyrosine levels favors the production of plasmid DNA. This positive effect is more significant at +1 level for phenylalanine due to the slope of the line

(steeper). The tryptophan and phenylalanine at levels 0 and -1 (b), respectively, the increase tyrosine concentration reduces production of pDNA. However, when phenylalanine is at +1 level leads to an increase plasmid. In (c) tryptophan has +1 and with increment tyrosine is not favorable for production of pDNA for both levels (-1, +1) of phenylalanine (Figure 16, 17 and 18).

Interaction AC

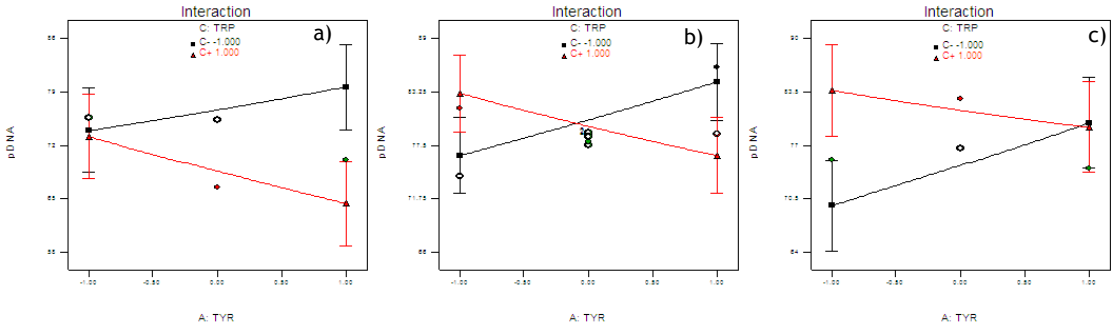

Figure 17- Interaction between Tyrosine and Tryptophan plot.

With the analysis of figure 17 is easier to visualize the interaction between A and C because at intersection of lines was obtained. This interaction is one of the most influence the model because it has a sum square of 52.73 and p-value of 0.0378. The graph (a) with the level of phenylalanine -1 revealed that with increasing concentration of tyrosine the effect is positive when tryptophan is at level -1 and negative at +1 level for plasmid production. The behavior of effects of phenylalanine at level 0 (b) are similar to (a) graph however the strength of the effect (positive and negative) is smaller. For +1 level of PHE (c) the situation is identical to the previous (b) but with a positive effect (TRP at -1 level) very pronounced.

Interaction BC

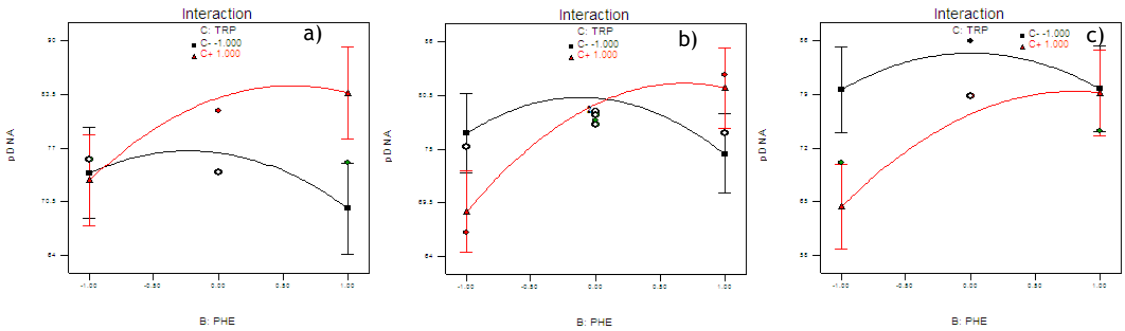

Figure 18- Interaction between Phenylalanine and Tryptophan plot.

The interaction between B and C (figure 18) happens non-linearly. This nonlinearity revealed that with increasing phenylalanine concentrations occurs a positive effect only until a certain concentration. From this point the interaction between B and C cannot exert any effect (tangent with slope = 0) or negative effect if the slope is negative. Analyzing the three graphs it is clear that the +1 level of TRP and with increase of phenylalanine the response is very similar. On the other hand, when TRP is at lowest level (-1) the transition from -1 to +1 level of PHE and TYR, translates into a low changes in the response. This happens due to the nonlinear behavior. The effect that seems to be more favorable to the production of pcDNA3.1-flag-p53 happens when tryptophan is at level 1 and accompanied by increased phenylalanine (a, b and c).

3.3.2.3. Optimization by Response Surface Methodology

Numerical optimization technique based on plasmid DNA production was carried out to determine the workable optimum conditions. The specific yield and purity of pDNA were not used as a response in the development of model, but response surface graph incorporation in this dissertation (figure 19) can improve the understanding of the experimental data. In case of pDNA specific yields, the best results occur with -1, +1, +1 levels of tyrosine, phenylalanine and tryptophan, respectively. In other words, these conditions suggest an increase the volumetric yield of plasmid DNA with the same biomass concentration. Moreover, analysis of response surface of purity indicate that the relationship between the three aromatic amino acids can be an advantage with their highest concentrations (+1,+1,+1 levels). However, it is necessary to examine this information very cautiously, since these data are not supported with adequate statistical significance.

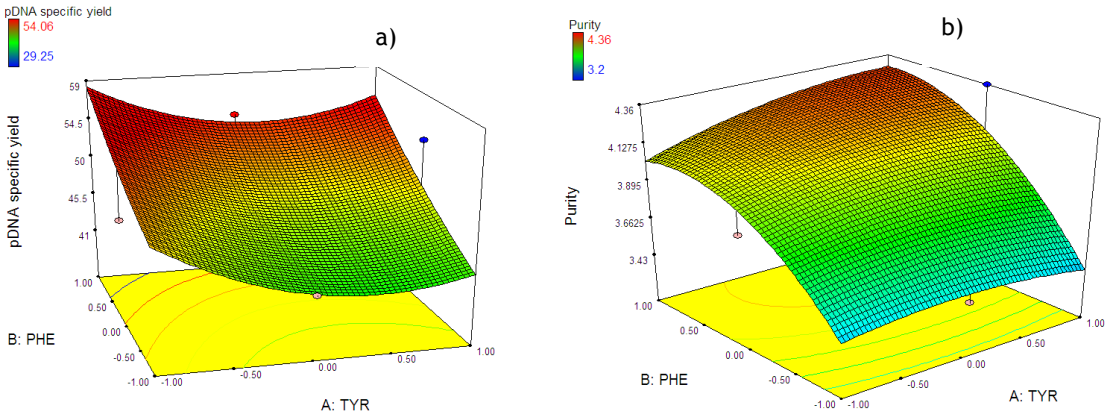

Figure 19- Response surface plot indicate the optimal pDNA specific yield (a), and purity (b).

In order to provide an ideal case for this specific pDNA production with statistical significance, the goal was set to maximized response (pDNA volumetric yield). The predicted optimum values of Tyrosine, Phenylalanine and Tryptophan corresponded to 10 µg/mL (-1 level), 78.5 µg/mL (0.57 level), and 100 µg/mL (+1 level), respectively, to achieve 84.53 µg/mL of plasmid volumetric yield (figure 20). Analyzing the graph of the response surface, the levels of tyrosine and phenylalanine that reflect a reddish area are the conditions to achieve better response (pDNA production) with triptophan at level +1.

The optimal variables levels were achieved based in whole information provided by BBD and RSM.

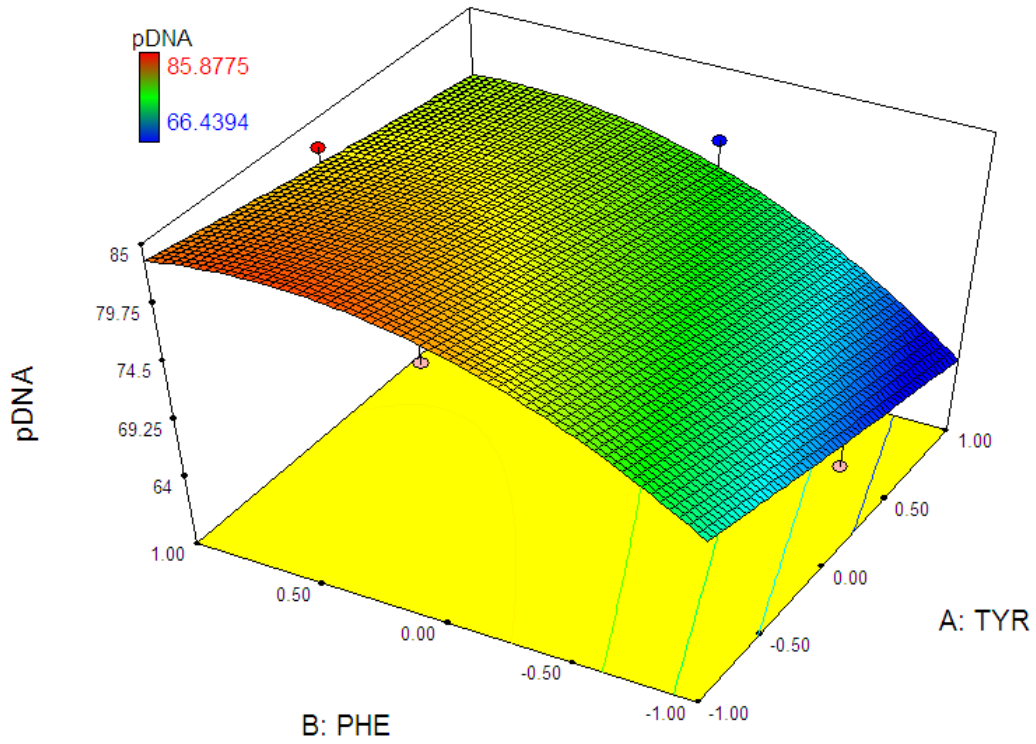

Figure 20- Response surface plot of optimal conditions of pDNA volumetric yield.

Chapter 4- Conclusions

In this study it was demonstrated the potential of a new strain of *E. coli* in therapeutic plasmid DNA biosynthesis. The strategies used have improved the knowledge of mechanisms of bacterial cell when submitted to different culture conditions. In this case, the use of *E. coli* VH33 requires the use of a carbon source, glucose. According to the literature, the best plasmid yield occurs through the use of glycerol. To overcome this barrier, research groups have developed a new generation of bacteria with better uptake system glucose. Thus, there is a higher efficiency in the production of plasmid as well as reducing the accumulation of inhibitory compounds such as acetate. The experimental data shown that an increase in glucose concentration leads a raise of plasmid volumetric yield. Moreover, the glucose uptake is much more efficient than glycerol and presence of these two carbon source inhibits the pcDNA3–FLAG–p53 production.

Once the host system was optimized it is also important develop a lab monitorization set up that can analyse and improve the upstream control processing. The method of sugars and acetate is important to monitor the glucose uptake of the fermentation medium and simultaneously assessing the production of the most inhibitory compound for *E. coli*, the acetate. On the other hand, the method of assessment of aromatic amino acid provides an additional information about what occurs in aromatic amino acids pathway. Both methods showed bias and CV within the parameters required by the FDA and ICH, which may be used for valid analysis.

The *E. coli* VH33 already has advantages with changes in their genome, which influence the aromatic amino acid pathway. So, it was important understand how this pathway affects the biosynthesis of plasmid. To assess this influence it was used the experimental design, specifically the Box-Behnken design. The development of the model with statistical significance reflects the relationship between the volumetric efficiency of pDNA and the concentration of the three amino acids. Moreover, it was possible to understand the interactions that occur between the different amino acids and plasmid DNA volumetric and massic yields. The significant interactions of model corresponded to phenylalanine, tyrosine-tryptophan, phenylalanine-tryptophan. Considering the interpretation of all statistical data of the developed model and response surface methodology was possible to obtain pseudo optimal conditions that increase the production of pDNA. So and basis on this scale, to achieve the best pDNA volumetric yield the concentrations of tyrosine, phenyalanine and tryptophan should be 10 µg/mL, 78.5 µg/mL and 100 µg/mL respectively.

Chapter 5-Future Trends

To extend and improve all this research study, the future work perspectives are:

- Conclude the investigation about the aromatic amino acids pathway understanding the effect of optimal conditions on production of by-products, such as, acetate.

- Scale-up to a mini-bioreactor plataform with a suitable model development using typical nutrients and raw materials from agro-food industries.

- Search new strategies for increasing pDNA yield at an early stage in the manufacturing process, such as increase plasmid copy number.

- Engineering the structure of the bacterial outer membrane to reduce endotoxin contamination.

Chapter 6- References

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

[2] P. H. Oliveira and J. Mairhofer, "Marker-free plasmids for biotechnological applications - implications and perspectives.," *Trends Biotechnol.*, pp. 1–9, Jul. 2013.

[3] 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.

[4] 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.

[5] 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.

[6] S. L. Ginn, I. E. Alexander, and M. L. Edelstein, "Gene therapy clinical trials worldwide to 2012 – an update," pp. 65–77, 2013.

[7] J. O. Josefsberg and B. Buckland, "Vaccine process technology.," *Biotechnol. Bioeng.*, vol. 109, no. 6, pp. 1443–60, Jun. 2012.

[8] 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.

[9] D. M. Drake and D. W. Pack, "Biochemical Investigation of Active Intracellular Transport of Polymeric Gene-Delivery Vectors," vol. 97, no. 4, pp. 1399–1413, 2008.

[10] V. M. Gaspar, I. J. Correia, A. Sousa, F. Silva, C. M. Paquete, J. a Queiroz, and F. Sousa, "Nanoparticle mediated delivery of pure P53 supercoiled plasmid DNA for gene therapy.," *J. Control. Release*, vol. 156, no. 2, pp. 212–22, Dec. 2011.

[11] S. Haupt, M. Berger, Z. Goldberg, and Y. Haupt, "Apoptosis - the p53 network.," *J. Cell Sci.*, vol. 116, no. Pt 20, pp. 4077–85, Oct. 2003.

[12] A. S. Azmi, "Pharmaceutical Reactivation of p53 Pathways in Cancer," *Curr. Pharm. Des.*, vol. 17, no. 6, pp. 534–535, 2011.

[13] M. Athar, C. A. Elmets, and L. Kopelovich2, "NIH Public Access," *Curr. Pharm. Des.*, vol. 17, no. 6, pp. 631–639, 2013.

[14] D. J. Glover, H. J. Lipps, and D. a Jans, "Towards safe, non-viral therapeutic gene expression in humans.," *Nat. Rev. Genet.*, vol. 6, no. 4, pp. 299–310, Apr. 2005.

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

[16] 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.

[17] 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.

[18] 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.

[19] J. a Williams, J. Luke, S. Langtry, S. Anderson, C. P. Hodgson, 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. 1129–43, Aug. 2009.

[20] 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.

[21] 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.

[22] J. Yang and Y. Yang, "Plasmid size can affect the ability of Escherichia coli to produce high-quality plasmids.," *Biotechnol. Lett.*, vol. 34, no. 11, pp. 2017–22, Nov. 2012.

[23] 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.

[24] J. R. Cooke, E. A. Mckie, J. M. Ward, and E. Keshavarz-moore, "Impact of intrinsic DNA structure on processing of plasmids for gene therapy and DNA vaccines," vol. 114, pp. 239–254, 2004.

[25] S. C. Ribeiro, G. a Monteiro, and D. M. F. Prazeres, "The role of polyadenylation signal secondary structures on the resistance of plasmid vectors to nucleases.," *J. Gene Med.*, vol. 6, no. 5, pp. 565–73, May 2004.

[26] 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.

[27] K. L. J. Prather, M. C. Edmonds, and J. W. Herod, "Identification and characterization of IS1 transposition in plasmid amplification mutants of *E. coli* clones producing DNA vaccines.," *Appl. Microbiol. Biotechnol.*, vol. 73, no. 4, pp. 815–26, Dec. 2006.

[28] M. Schleef and T. Schmidt, "Animal-free production of ccc-supercoiled plasmids for research and clinical applications.," *J. Gene Med.*, vol. 6 Suppl 1, pp. S45–53, Feb. 2004.

[29] Y. Han, S. Liu, J. Ho, M. K. Danquah, and G. M. Forde, "Using DNA as a drug—Bioprocessing and delivery strategies," *Chem. Eng. Res. Des.*, vol. 87, no. 3, pp. 343–348, Mar. 2009.

[30] A. Ghanem, R. Healey, and F. G. Adly, "Analytica Chimica Acta Current trends in separation of plasmid DNA vaccines : A review," *Anal. Chim. Acta*, vol. 760, pp. 1–15, 2013.

[31] M. M. Diogo, J. a. Queiroz, and D. M. F. Prazeres, "Chromatography of plasmid DNA," *J. Chromatogr. A*, vol. 1069, no. 1, pp. 3–22, Mar. 2005.

[32] A. Singer, M. A. Eiteman, and E. Altman, "DNA plasmid production in diVerent host strains of Escherichia coli," pp. 521–530, 2009.

[33] 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.

[34] S. S. Freitas, a. R. Azzoni, J. a. L. Santos, G. a. Monteiro, and D. M. F. Prazeres, "On the stability of plasmid DNA vectors during cell culture and purification," *Mol. Biotechnol.*, vol. 36, no. 2, pp. 151–158, May 2007.

[35] S. Haberl, M. Jarc, A. Strancar, M. Peterka, D. Hodžić, and D. Miklavčič, "Comparison of Alkaline Lysis with Electroextraction and Optimization of Electric Pulses to Extract Plasmid DNA from Escherichia coli.," *J. Membr. Biol.*, Jul. 2013.

[36] K. Nordström, "Plasmid R1--replication and its control.," *Plasmid*, vol. 55, no. 1, pp. 1–26, Jan. 2006.

[37] G. del Solar, R. Giraldo, M. J. Ruiz-Echevarría, M. Espinosa, and R. Díaz-Orejas, "Replication and control of circular bacterial plasmids.," *Microbiol. Mol. Biol. Rev.*, vol. 62, no. 2, pp. 434–64, Jun. 1998.

[38] R. R. Chaudhuri and I. R. Henderson, "The evolution of the Escherichia coli phylogeny.," *Infect. Genet. Evol.*, vol. 12, no. 2, pp. 214–26, Mar. 2012.

[39] F. R. Schmidt, "Optimization and scale up of industrial fermentation processes.," *Appl. Microbiol. Biotechnol.*, vol. 68, no. 4, pp. 425–35, Sep. 2005.

[40] 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, p. 132, Jan. 2012.

[41] a Sousa, F. Sousa, and J. a Queiroz, "Impact of lysine-affinity chromatography on supercoiled plasmid DNA purification.," *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.*, vol. 879, no. 30, pp. 3507–15, Nov. 2011.

[42] E. Meza, J. Becker, F. Bolivar, G. Gosset, and C. Wittmann, "Consequences of phosphoenolpyruvate:sugar phosphotranferase system and pyruvate kinase isozymes inactivation in central carbon metabolism flux distribution in Escherichia coli.," *Microb. Cell Fact.*, vol. 11, no. 1, p. 127, Jan. 2012.

[43] 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.

[44] S. Jahn, B. R. Haverkorn van Rijsewijk, U. Sauer, and K. Bettenbrock, "A role for EIIA(Ntr) in controlling fluxes in the central metabolism of *E. coli* K12.," *Biochim. Biophys. Acta*, vol. 1833, no. 12, pp. 2879–2889, Jul. 2013.

[45] K. Martínez-Gómez, N. Flores, H. M. Castañeda, G. Martínez-Batallar, G. Hernández-Chávez, O. T. Ramírez, G. Gosset, S. Encarnación, and F. Bolivar, "New insights into Escherichia coli metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol.," *Microb. Cell Fact.*, vol. 11, p. 46, Jan. 2012.

[46] V. F. Wendisch, M. Bott, and B. J. Eikmanns, "Metabolic engineering of Escherichia coli and Corynebacterium glutamicum for biotechnological production of organic acids and amino acids.," *Curr. Opin. Microbiol.*, vol. 9, no. 3, pp. 268–74, Jun. 2006.

[47] A. Buchholz, J. Hurlebaus, C. Wandrey, and R. Takors, "Metabolomics: quantification of intracellular metabolite dynamics.," *Biomol. Eng.*, vol. 19, no. 1, pp. 5–15, Jun. 2002.

[48] 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.

[49] J. Shiloach and R. Fass, "Growing *E. coli* to high cell density--a historical perspective on method development.," *Biotechnol. Adv.*, vol. 23, no. 5, pp. 345–57, Jul. 2005.

[50] F. Silva, J. a Queiroz, and F. C. Domingues, "Plasmid DNA fermentation strategies: influence on plasmid stability and cell physiology.," *Appl. Microbiol. Biotechnol.*, vol. 93, no. 6, pp. 2571–80, Mar. 2012.

[51] M. Ramı and J. M. Go, "Optimization of culture media for ethanol production from glycerol by Escherichia coli," vol. 7, pp. 3–9, 2011.

[52] D. M. F. Prazeres, G. N. M. Ferreira, G. A. Monteiro, C. L. Cooney, and J. M. S. Cabral, "pharmaceutical-grade plasmid DNA for gene therapy : problems and bottlenecks," vol. 17, no. April, pp. 169–174, 1999.

[53] M. Losen, B. Frölich, M. Pohl, and J. Büchs, "Effect of oxygen limitation and medium composition on Escherichia coli fermentation in shake-flask cultures.," *Biotechnol. Prog.*, vol. 20, no. 4, pp. 1062–8, 2004.

[54] M. Emmerling, M. Dauner, A. Ponti, M. Hochuli, T. Szyperski, K. Wüthrich, E. Bailey, U. Sauer, J. Fiaux, K. Wu, and J. E. Bailey, "Metabolic Flux Responses to Pyruvate Kinase Knockout in Escherichia coli Metabolic Flux Responses to Pyruvate Kinase Knockout in Escherichia coli," 2002.

[55] 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.

[56] D. M. Bower and P. S. Doyle, "Development of New Tools for the Production of Plasmid DNA Biopharmaceuticals by Development of New Tools for the Production of Plasmid DNA Biopharmaceuticals by," 2012.

[57] Y. Dharmadi, A. Murarka, and R. Gonzalez, "Anaerobic Fermentation of Glycerol by Escherichia coli : A New Platform for Metabolic Engineering," 2006.

[58] M. M. Salem-bekhit, M. Bayomi, F. Al-anazi, and S. Arabia, "pb YIELD PRODUCTION OF RECOMBINANT PLASMID DNA WITH ESCHERICHIA COLI IN FED-BATCH CULTURE BY PSEUDO-EXPONENTIAL," no. 22, pp. 3709–3715, 2012.

[59] 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.

[60] 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.

[61] S. Zheng, K. Friehs, N. He, X. Deng, and L. Qingbiao, "Optimization of medium components for plasmid production by recombinant *E. coli* DH5apha pUK21CMVb1.2," *Biotechnol. Bioprocess Eng.*, vol. 12, pp. 213–221, 2007.

[62] S. S. Levanon, K.-Y. San, and G. N. Bennett, "Effect of oxygen on the Escherichia coli ArcA and FNR regulation systems and metabolic responses.," *Biotechnol. Bioeng.*, vol. 89, no. 5, pp. 556–64, Mar. 2005.

[63] O. F. Restaino, D. Cimini, M. De Rosa, A. Catapano, and C. Schiraldi, "High cell density cultivation of Escherichia coli K4 in a microfiltration bioreactor: a step towards improvement of chondroitin precursor production.," *Microb. Cell Fact.*, vol. 10, no. 1, p. 10, Jan. 2011.

[64] 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.

[65] G. W. Luli and W. R. Strohl, "Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations.," *Appl. Environ. Microbiol.*, vol. 56, no. 4, pp. 1004–11, Apr. 1990.

[66] C. R. Dittrich, G. N. Bennett, and K. San, "Characterization of the Acetate-Producing Pathways in Escherichia coli," pp. 1062–1067, 2005.

[67] J. Y. H. Kim and H. J. Cha, "Down-regulation of acetate pathway through antisense strategy in Escherichia coli: improved foreign protein production.," *Biotechnol. Bioeng.*, vol. 83, no. 7, pp. 841–53, Sep. 2003.

[68] A. R. Lara, L. Caspeta, G. Gosset, and F. Bolı, "Utility of an Escherichia coli Strain Engineered in the Substrate Uptake System for Improved Culture Performance at High Glucose and Cell Concentrations : An Alternative to Fed-Batch Cultures," vol. 99, no. 4, pp. 893–901, 2008.

[69] G. Striedner and K. Bayer, "An Advanced Monitoring Platform for Rational Design of Recombinant Processes," no. December 2012, pp. 65–84, 2013.

[70] L. R. V. de Sá, M. A. L. de Oliveira, M. C. Cammarota, A. Matos, and V. S. Ferreira-Leitão, "Simultaneous analysis of carbohydrates and volatile fatty acids by HPLC for monitoring fermentative biohydrogen production," *Int. J. Hydrogen Energy*, vol. 36, no. 23, pp. 15177–15186, Nov. 2011.

[71] C. L. Winder, W. B. Dunn, S. Schuler, D. Broadhurst, R. Jarvis, G. M. Stephens, and R. Goodacre, "Global metabolic profiling of Escherichia coli cultures: an evaluation of methods for quenching and extraction of intracellular metabolites.," *Anal. Chem.*, vol. 80, no. 8, pp. 2939–48, Apr. 2008.

[72] B. Luo, K. Groenke, R. Takors, C. Wandrey, and M. Oldiges, "Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by liquid chromatography-mass spectrometry.," *J. Chromatogr. A*, vol. 1147, no. 2, pp. 153–64, Apr. 2007.

[73] 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.

[74] Y. Liu, F. Wang, and W. Lee, "On-line monitoring and controlling system for fermentation processes.," *Biochem. Eng. J.*, vol. 7, no. 1, pp. 17–25, Jan. 2001.

[75] K. Schügerl, "Progress in monitoring, modeling and control of bioprocesses during the last 20 years.," *J. Biotechnol.*, vol. 85, no. 2, pp. 149–73, Feb. 2001.

[76] A. E. Carnes, "Fermentation Design for the Manufacture of Therapeutic Plasmid DNA."

[77] J. Bongaerts, M. Krämer, U. Müller, L. Raeven, and M. Wubbolts, "Metabolic engineering for microbial production of aromatic amino acids and derived compounds.," *Metab. Eng.*, vol. 3, no. 4, pp. 289–300, Oct. 2001.

[78] P. Bakonyi, N. Nemestóthy, É. Lövitusz, and K. Bélafi-Bakó, "Application of Plackett-Burman experimental design to optimize biohydrogen fermentation by *E. coli* (XL1-BLUE)," *Int. J. Hydrogen Energy*, vol. 36, no. 21, pp. 13949–13954, Oct. 2011.

[79] J. Wang and W. Wan, "Experimental design methods for fermentative hydrogen production: A review," *Int. J. Hydrogen Energy*, vol. 34, no. 1, pp. 235–244, Jan. 2009.

[80] M. A. Bezerra, R. E. Santelli, E. P. Oliveira, L. S. Villar, and L. A. Escaleira, "Response surface methodology (RSM) as a tool for optimization in analytical chemistry.," *Talanta*, vol. 76, no. 5, pp. 965–77, Sep. 2008.

[81] S. L. C. Ferreira, R. E. Bruns, H. S. Ferreira, G. D. Matos, J. M. David, G. C. Brandão, E. G. P. da Silva, L. a Portugal, P. S. dos Reis, a S. Souza, and W. N. L. dos Santos, "Box-Behnken design: an alternative for the optimization of analytical methods.," *Anal. Chim. Acta*, vol. 597, no. 2, pp. 179–86, Aug. 2007.

[82] R. Nelofer, R. N. Ramanan, R. N. Z. R. A. Rahman, M. Basri, and A. B. Ariff, "Comparison of the estimation capabilities of response surface methodology and artificial neural network for the optimization of recombinant lipase production by *E. coli* BL21.," *J. Ind. Microbiol. Biotechnol.*, vol. 39, no. 2, pp. 243–54, Feb. 2012.

[83] M. M. Diogo, J. a Queiroz, and D. M. F. Prazeres, "Assessment of purity and quantification of plasmid DNA in process solutions using high-performance hydrophobic interaction chromatography.," *J. Chromatogr. A*, vol. 998, no. 1–2, pp. 109–17, May 2003.

[84] R. Bongiovanni, B. K. Yamamoto, and G. E. Jaskiw, "Improved method for the measurement of large neutral amino acids in biological matrices.," *J. Chromatogr. B. Biomed. Sci. Appl.*, vol. 754, no. 2, pp. 369–76, Apr. 2001.

[85] B. M. Validation, "Guidance for Industry Bioanalytical Method Validation Guidance for Industry Bioanalytical Method Validation," no. May, 2001.

[86] I. Conference, O. N. Harmonisation, O. F. Technical, R. For, R. Of, P. For, and A. Of, "ICH H ARMONISED T RIPARTITE G UIDELINE V ALIDATION OF A NALYTICAL P ROCEDURES :," vol. 1994, no. November 1996, 2005.

[87] 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.

[88] A. R. Lara, I. Knabben, L. Regestein, J. Sassi, L. Caspeta, O. T. Ramírez, and J. Büchs, "Comparison of oxygen enriched air vs. pressure cultivations to increase oxygen

transfer and to scale-up plasmid DNA production fermentations," *Eng. Life Sci.*, vol. 11, no. 4, pp. 382–386, Aug. 2011.

[89] R. Soto, L. Caspeta, B. Barrón, G. Gosset, O. T. Ramírez, and A. R. Lara, "High cell-density cultivation in batch mode for plasmid DNA production by a metabolically engineered *E. coli* strain with minimized overflow metabolism," *Biochem. Eng. J.*, vol. 56, no. 3, pp. 165–171, Oct. 2011.

[90] L. Regestein, a. Wolf, H.-J. Schneider, T. Maskow, F. Mertens, J. Büchs, and J. Lerchner, "A chip calorimeter for the monitoring of conventional bioreactors at elevated cell concentrations," *Thermochim. Acta*, vol. 544, pp. 10–16, Sep. 2012.

[91] T. Polen, M. Krämer, J. Bongaerts, M. Wubbolts, and V. F. Wendisch, "The global gene expression response of Escherichia coli to L-phenylalanine.," *J. Biotechnol.*, vol. 115, no. 3, pp. 221–37, Feb. 2005.

[92] G. a Sprenger, "From scratch to value: engineering Escherichia coli wild type cells to the production of L-phenylalanine and other fine chemicals derived from chorismate.," *Appl. Microbiol. Biotechnol.*, vol. 75, no. 4, pp. 739–49, Jun. 2007.

[93] S. E. Agarry and O. O. Ogunleye, "Box-Behnken design application to study enhanced bioremediation of soil artificially contaminated with spent engine oil using biostimulation strategy," *Int. J. Energy Environ. Eng.*, vol. 3, no. 1, p. 31, 2012.