UNIVERSIDADE DA BEIRA INTERIOR
Ciências

Production and Purification of DNA G-quadruplex using pPH600 plasmid

Tiago André Afonso dos Santos

Dissertação para obtenção do Grau de Mestre em

Biotecnologia

(2º ciclo de estudos)

Orientadora: Prof.ª Dr.ª Carla Patrícia Alves Freire Madeira Cruz
Co-orientador: Prof. Dr. João António de Sampaio Rodrigues Queiroz

Covilhã, junho de 2015

“Attitude is a little thing that makes a big difference”

Winston Churchill

Acknowledgments

First of all, I would like to express my sincere gratitude to my supervisor Dr Carla Cruz and my co-supervisor Professor João Queiroz, for all the guidance, orientation, support and trust that giving me during this year. Your advices and scientific expertise are crucial to the development of this work.

I would also like to express my sincere gratitude to Professor Fani Sousa and Professor Ângela Sousa for the availability, criticism and scientific expertise in all the steps of my work.

My special thanks to Patricia Pereira for his teaching since my first day in the lab. Your friendship, advices, encouragement and scientific expertise are fundamental to the development of this work. Thank you.

I also thank to all the people involved in Health Sciences Research Centre of the University of Beira Interior, especially to the Biotechnology and Biomolecular Sciences group for all their help and to providing me all the resources. I would like to express a special acknowledge to Augusto and Ana Margarida for their friendship, complete availability and support.

To my friends, some of them far but the feeling that giving me was that are always present encourage me and supporting me.

To all of my family, especially to my parents and my brother for all affection, support and care, never doubting me and give me a strength to continue my thrill to success and happiness. I love you.

Finally, i want to thank to my girlfriend, Ana Sofia, for all the patiente, support and love that give me in all moments, always willing to help me in my critical moments and to have a friendly word to encourage me to follow my dreams. I love you.

Resumo

Quando pensamos na estrutura do DNA a imagem que vem imediatamente à nossa mente é a da icónica estrutura em dupla hélice que Watson e Crick descobriram em 1953. No entanto, para além desta estrutura em dupla hélice, o DNA tem capacidade para assumir outras conformações secundárias, não menos relevantes do ponto de vista biológico. Algumas importantes regiões do genoma humano têm um potencial incomum para formar estas estruturas durante o processo de transcrição. O plasmídeo pPH600 tem uma sequência de 604 pb da região de troca Sγ3 da imunoglobulina do ratinho que tem capacidade para formar G-quadruplex durante a transcrição. A predominância de G-loops é maior na topologia superenrolada (sc) do que nas topologias circular aberta (oc) ou linear (ln). O presente trabalho descreve a biosíntese do plasmídeo pPH600 em *E. coli* DH5α e as estratégias para purificação da isoforma sc do pPH600, diretamente a partir de uma amostra nativa (oc + sc) ou a partir do lisado de *E. coli* clarificado. As estratégias de purificação são baseadas na cromatografia de afinidade com aminoácidos, aproveitando o reconhecimento biológico com o plasmídeo. Para este propósito dois suportes foram preparados, L-triptofano e L-tirosina, através de imobilização covalente dos aminoácidos à Sepharose CL-6B através do braço espaçador 1,4-butanediol diglicidil éter. O suporte comercial L-arginina-Sepharose 4B foi também utilizado devido à eficiência na purificação da isoforma sc com diferentes plasmídeos utilizando baixas concentrações de sal para ligação e eluição da isoforma sc. Deste modo, foi realizado um *screening* inicial com a amostra nativa (oc + sc) de pPH600 para avaliar comportamento dos três suportes. O suporte que promove a separação das isoformas do pPH600 foi selecionado para purificação da isoforma sc com o lisado de *E. coli* clarificado. O suporte de L-tirosina foi o que demonstrou resultados proeminentes na separação das duas isoformas, permitindo a recuperação total da isoforma sc de pPH600, através de um gradiente por passos decrescente de 2,25 M para 0 M de $(NH_4)_2SO_4$ em HEPES 100 mM (pH 7,4) à temperatura de 10ºC. Subsequentemente, o lisado de *E. coli* clarificado foi injetado diretamente no suporte de L-tirosina para separar a isoform sc de pPH600 das impurezas internas do hospedeiro (gDNA, RNA, proteínas, endotoxinas e outras conformações de pPH600). A separação foi alcançada com um gradiente por passos decrescente de 2,25 para 1,95 M e posteriormente para 0 M $(NH_4)_2SO_4$ em HEPES 100 mM (pH 7,4) à temperatura de 10ºC. O mecanismo de reconhecimento molecular que permite a separação envolve não só as interações hidrofóbicas mas também outras interações não covalentes como, pontes de hidrogénio, empilhamento π-π e van der Waals. Os testes de avaliação do plasmídeo indicam que a isoforma sc de pPH600 resultante do passo de purificação tem um grau de pureza de 98,2%, com um reduzido nível de impurezas. Finalmente, a isoforma sc resultante da purificação foi transcrito e analisado por dicroísmo circular para comprovar a formação do G-loop.

Palavras-chave

G-quadruplex; sc pPH600; cromatografia de afinidade; *E. coli*; L-arginina; L-triptofano; L-tirosina

Resumo Alargado

Nos últimos anos o cancro e as doenças neurodegenerativas têm dominado o interesse dos investigadores devido ao aumento do número de casos diagnosticados e às previsões efetuadas por diversas entidades no sentido do aumento desse número nos próximos anos. Estas doenças têm causado um enorme impacto tanto na saúde como na economia dos países. Até à determinação da composição química e resolução da estrutura da dupla hélice em 1953, pensou-se que o DNA era apenas a molécula que armazenava a informação hereditária. No entanto, com a descoberta da dupla hélice foi possível determinar como a informação genética é transmitida de geração em geração. Muitos estudos foram posteriormente realizados para elucidar possíveis variações da estrutura e a função das mesmas. Várias estruturas secundárias receberam interesse, no entanto os G-quadruplex devido a inúmeros fatores têm recebido grande parte da atenção dos investigadores.

As guaninas têm a capacidade de se organizar e formar estruturas cíclicas através de ligações de hidrogénio, dando origem a G-quartetos. Os G-quartetos através de interações de empilhamento π-π e estabilização com catiões dão origem aos G-quadruplexes. Estas estruturas são polimórficas devido ao elevado número de possibilidades para as formar. Assim, deve ter-se em conta fatores como a quantidade de moléculas envolvidas na formação, sequência dessas moléculas e orientação das cadeias de DNA.

Os G-quadruplex são geralmente encontrados em regiões do genoma com enorme relevância, como são a região de troca da imunoglobulina, os telómeros e regiões regulatórias de oncogenes. Estas estruturas desempenham um papel fundamental em processos celulares essenciais como são a transcrição, a recombinação e a replicação e portanto, têm sido associados a inúmeras doenças.

O processo de transcrição da região de troca da imunoglobulina está diretamente envolvido no desenvolvimento normal das células B e consequentemente na resposta imune. Durante a transcrição formam-se os G-quadruplexes que são o alvo de ligação de diversas proteínas que podem estabilizar ou clivar as estruturas para manter a estabilidade genómica.

Recentemente, vários estudos foram realizados e descreveram a importância do grau de superenrolamento do DNA na formação destas estruturas. A importância deste fenómeno é relacionada não só com o armazenamento da informação genética no núcleo mas, também com a indução de estruturas secundárias devido à energia acumulada durante o superenrolamento do DNA.

Os plasmídeos são moléculas que possuem diferentes topologias: superenrolada, circular aberta e linear e por isso têm diferentes capacidades para formar G-loops. Os G-loops são caracterizados pela formação de um híbrido de RNA-DNA que estabiliza os G-quadruplex na cadeia rica em guaninas. Na literatura está descrito que a percentagem de G-loops formados nos plasmídeos em topologia superenrolada é superior à das outras topologias.

O plasmídeo utilizado foi o pPH600 que contém uma sequência do fragmento da região de troca da imunoglobulina Sγ3 do ratinho e que após transcrição forma G-quadruplexes na cadeia rica em guaninas.

Assim, a necessidade de produzir e purificar a isoforma superenrolada do pPH600 tem grande relevância quando o objetivo é a utilização destas estruturas para estudos *in vivo*.

A fermentação para amplificação do vetor foi efetuada após transformação das células de *E. coli* DH5 com o plasmídeo pPH600. Após fermentação as células forma recolhidas e foi efetuada a lise das células através de um kit comercial ou através do método de lise alcalina modificado.

A etapa cromatográfica foi realizada por cromatografia de afinidade. A cromatografia de afinidade foi a técnica selecionada para purificação de ácidos nucleicos uma vez que permite interações específicas entre o ligando e a biomolécula apresentando vantagens tanto processuais como económicas. Recentemente vários aminoácidos têm sido utilizados como ligandos na cromatografia de afinidade devido às múltiplas interações não covalentes que se podem estabelecer entre as duas unidades básicas, os aminoácidos e os nucleótidos. Neste estudo foi utilizado o suporte comercial, a L-arginina Sefarose e sintetizados dois outros suportes através da imobilização covalente do L-triptofano e da L-tirosina à Sefarose CL-6B.

O suporte L-arginina Sefarose tem sido utilizado para purificar diversos plasmídeos com eficiência e por isso foi incorporado neste trabalho. O L-triptofano e a L-tirosina foram selecionados devido às muitas similaridades que apresentam com o suporte de L-histidina, também utilizado recentemente para separar a isoforma superenrolada do plasmídeo de outros contaminantes.

Neste trabalho, pretende-se efetuar um *screening* inicial da purificação das isoformas do plasmídeo pPH600 com os três suportes e depois selecionar aquele em que melhores resultados forem obtidos para purificação da isoforma superenrolada através de lisado de *E. coli*. Neste caso, a L-tirosina apresentou os melhores resultados na purificação das isoformas permitindo uma completa separação através de um gradiente decrescente de 2,25 para 0 M de $(NH_4)_2SO_4$ em HEPES 100 Mm (pH 7.4). A separação completa da isoforma superenrolada de pPH600 das outras impurezas do lisado de *E. coli* foi obtida através de um gradiente decrescente de 2,25 para 1,95 e finalmente para 0 M de $(NH_4)_2SO_4$ em HEPES 100 mM (pH 7,4). Após purificação avaliou-se a qualidade e o nível de impurezas presentes na isoforma superenrolada. Verificou-se que a quantidade de pPH600 superenrolado recuperado foi de aproximadamente 56% e que a pureza da amostra era de aproximadamente 98%. No que diz respeito às impurezas, foram avaliadas através das técnicas definidas pelas agências reguladoras e verificou-se uma diminuição significativa sendo os valores abaixo dos limites aconselhados por estas agências.

Por fim, a amostra de pPH600 superenrolado purificada com o suporte de L-tirosina a partir do lisado de *E. coli* foi sujeita a transcrição e a observação dos G-loops no transcrito foi avaliada por CD.

Verificou-se que o G-quadruplex formado apresentava topologia paralela devido à presença de uma banda negativa a 240 nm e uma banda positiva a 265 nm.

Abstract

When we think in structure of DNA, the image that comes immediately to mind is the iconic structure in double helix discovered by Watson and Crick in 1953. However, in addition to this structure DNA can assume other secondary structures which are relevant in the biological context. Some important regions of human genome have unusual potential to form this structures upon transcription. The plasmid pPH600 have a sequence of Sγ3 immunoglobulin switch region of murine that are able to form G-quadruplex upon transcription. The G-loops predominance is more evidenced on supercoiled (sc) topology than on relaxed (oc) or linearized (ln) plasmid. The present work describe the biosynthesis of plasmid pPH600 in *E. coli* DH5α and the strategies for sc pPH600 purification, directly from native sample (oc + sc) and clarified *E. coli* lysate. The purification strategies are based on amino acid affinity chromatography taking advantage of biological recognition to pDNA. For this propose two supports were prepared, L-tryptophan Sepharose and L-tyrosine Sepharose, by covalent immobilisation using 1,4-butanediol diglycidyl ether spacer arm. The commercial support L-arginine Sepharose 4B was also used in the strategy for purifying sc pPH600 since it has already been efficiently applied to separate sc isoforms of different plasmids using mild binding and elution conditions. Therefore, an initial *screening* with pPH600 native sample (oc + sc) was performed to evaluate the behavior of three supports. The better support in isoform separation was selected to purify sc pPH600 directly from clarified lysate. L-tyrosine support shows the prominent result in separation of two isoforms, allowing the recovery of sc pPH600, through a decreasing stepwise gradient from 2.25 to 0 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4) with temperature at 10ºC. Thereafter, the clarified *E. coli* lysate sample was injected directly onto L-tyrosine support to separate sc pPH600 from internal impurities of *E. coli* (gDNA, RNA, proteins, endotoxins and other conformations of pPH600). The total separation of sc pPH600 was totally achieved using a stepwise gradient from 2.25, 1.95 and 0 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4) with temperature at 10ºC. The underlying mechanism is thought to involve not only hydrophobic but also other non-covalent interactions such as, hydrogen bonds, π-π stacking and van der Waals interactions. Plasmid assessment tests indicated that the sc pPH600 resultant from the purification step presented a purity degree of 98.2%, with an extremely reduced level of impurities.

Finally, the sc pPH600 resultant from purification was transcribed to induce the formation of G-quadruplex and it is confirmed by circular dichroism.

Keywords

G-quadruplex; sc pPH600; Affinity chromatography; *E. coli*; L-arginine; L-tryptophan; L-tyrosine.

Table of contents

List of figures

Chapter I – Introduction

Chapter II – Materials and methods

Chapter III – Results and discussion

List of tables

Chapter I – Introduction

Chapter II – Materials and methods

List of acronyms

A	Adenine
ADA	Adenosine deaminase
A-DNA	A-form of DNA
AID	Activation-induced cytidine deaminase
Aß	B-amyloid peptides
BCA	Bicinchoninic acid
Bcl-2	B cell lymphoma 2
B-DNA	B form of DNA
bp	Base pairs
C	Cytosine
C9ORF72	Chromosome 9 open reading frame 72
CD	Circular dichroism
c-Kit	V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
cm	centimeter
c-Myc	V-myc avian myelocytomatosis viral oncogene homolog
c-rich region	Cytosine rich region
CSR	Class switch recombination
DNA	Deoxyribonucleic acid
Duplex	DNA double strand
E. coli	*Escherichia coli*
EAEMP	European Agency for the evaluation of medical products
EDTA	Ethylene diamine tetra-acetic acid
EU	Endotoxin unit
FDA	Food and drug administration
FMR1	Fragile X mental retardation 1
FMRP	Fragile X mental retardation protein
G	Guanine
g	gram
G4	G-quadruplex
gDNA	Genomic DNA
g-rich region	Guanine rich region
h	hours
HEPES acid	HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
hTERT	Human telomerase reverse transcriptase
IgA	Immunoglobulin A
IgM	Immunoglobulin M
kbp	Kilo base pairs

K_D	Affinity constant
KRAS	Kirsten rat sarcoma viral oncogene homolog
L	liter
LAL	*Limulus* amebocyte lysate
LB	Luria Bertani
ln	linear
M	Molar
min	Minutes
miR1229	Micro RNA 1229
mL	milliliter
mM	millimolar
mRNA	Messenger RNA
MutSα	Mismatch repair protein
ng	nanogram
nm	nanometer
NMR	Nuclear magnetic resonance
NOP56	Nucleolar protein 56
oc	Open circular
°C	Celsius
OD_{600}	Optical density at 600 nm
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDGF-A	Platelet-derived growth factor subunit A
pDNA	Plasmid DNA
PEG	Polyethylene glycol
PNA	Peptide nucleic acid
pRb	Retinoblastoma protein
premiR-29	Pre micro RNA-29
PRNP	Prion protein
RCF	Relative centrifugal force
rDNA	Ribosomal DNA
RNA	Ribonucleic acid
rpm	Rotations per minute
S region	Switch region
sc	supercoiled
SCA36	Spinocerebral ataxia type 36
SCID	Severe combined immunodeficiency
SDS	Sodium dodecyl sulphate
SEM	Scanning electron microscopy
SPR	Surface plasmon resonance

T	Thymine
TB	Terrific broth
TF	Transcription factor
Tris	Tris(hydroximethyl)methylamine
TSS	Transcription start site
U	Uracyl
UV	Ultraviolet light
V	Volts
v/v	Volume/volume
VEGF-A	Vascular endothelial growth factor A
w/v	mass/volume
Z-DNA	Z-form of DNA, left handed-DNA
µg	microgram
µL	microliter

Chapter 1

Introduction

Since were discovered in 1869 by Friedrich Meischer, a Swiss Physician, the nucleic acids have been the subject of fields such biology, genetics and biochemistry. Meischer working at University of Tübingen, Germany in chemical composition of leukocytes. In his studies he isolate a hitherto unknown substance from the nuclei of cells. Meischer called the isolated substance *nuclein*. *Nuclein* is a crude extract, and contain many proteins. Subsequently, analysis and isolation of this substance showed that it had acidic nature and containing phosphorus, and its name was changed to nucleic acid [1, 2]. After that, nucleic acids continues to amaze and intrigue many researchers devoted to its study.

In 1919, Phoebus Levene was the first to suggest that nucleic acids was a linear chain composed by units of a sugar backbone that links to a phosphate groups and nitrogen bases. Levene identified D-ribose and 2-deoxy-D-ribose as the sugars occurring in RNA and DNA, respectively. He also identified the nitrogen bases, purines and pyrimidines. DNA is composed by cytosine (C) and thymine (T) that are pyrimidines and adenine (A) and guanine (G) that are purines. In RNA thymine is replaced by uracyl (U) that is a pyrimidine derivative. He called each unit of phosphate-sugar-base, nucleotide (Fig. 1). Based on this, Levene's work culminated with the suggestion of a tetranucleotide hypothesis for the structure of nucleic acids, in which nucleotides were always linked in the same order [3].

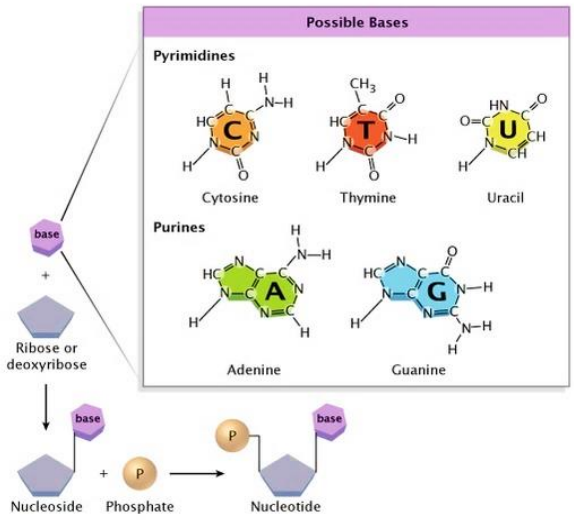


Figure 1 - Chemical Structure of DNA and RNA nucleotides (Adapted from [4]).

Later in 1940's decade, Erwin Chargaff and collaborators showed the lack of equality of four bases in most samples of DNA, unlike as required by the tetranucleotide hypothesis for the structure of nucleic acids. He was the first to develop micro-methods such as, paper chromatography and ultraviolet light absorption for accurate analysis of purines and pyrimidines and observed that the amount of adenine is equal to thymine and the amount of guanine is equal to cytosine, but the amount of guanine is not identical to thymine, nor adenine that of cytosine [5]. In this decade many other scientists such as, Avery, MacLeod, McCarthy, Hershey and Chase, contribute with their studies to prove that DNA was the carrier of hereditary information and not the proteins as previously thought.

In 1953, Francis Crick and James Watson discovered double helix of DNA known as B-DNA. This fact, marked a decisive milestone in the history of science in last century. Since then, the B-DNA has usually been regarded as the biologically relevant structure of DNA [6].

However, in addition to the normal Watson-Crick duplex, B-form, nucleic acids has potential to form a high number of alternative structural conformations under specific experimental or *in vivo* conditions, which are less familiar, although with biological relevance. The negative supercoiling of DNA can induce sequence-dependent conformational changes that give rise to alternative DNA conformations such as cruciforms, A-DNA, left-handed DNA (Z-DNA), triplexes, four-stranded DNA, known as G-quadruplexes (G4) and others [7].

First evidence that G4 behaved differently from all other nucleobases was first noted in 1910 by Ivar Bang, when he observed that it could spontaneously form a gel. However, only 52 years later, in 1962, the structure has been discovered by Davies and collaborators [8]. The structure has termed G4 because fold within extended guanine-rich regions into four stranded motifs, using an alternative base pairing arrangement [9]. These structures are generated by stacked G-quartets that are composed by four guanine bases in each of which interacts with its neighbours through Hoogsteen hydrogen bonding. These structures are further stabilized by cations (e.g., Na^+, K^+) that are in the center between the stacks, neutralizing the electrostatic repulsion of inwardly pointing guanine oxygen's [10].

Nowadays there is a growing evidence that G4 structures play important roles in various genetic processes [11]. They are often present in biologically regions, such as telomeres, regulatory regions of oncogenes and immunoglobulin switch regions. During last decades, studies indicated that formation of G4 in these regions may play important regulatory roles. Therefore, the polymorphism of G4 could make them valuable molecular targets to study biological processes and for possible therapeutic interventions in cancer and other diseases [12, 13].

The plasmid pPH600 contains a fragment of Sγ3 switch region of murine and upon transcription is able to form G4 structures. The supercoiling show more ability to form G4 as described by Duquette and collaborators [14].

Therefore there is a need to obtain purified sc pPH600 for G4 *in vitro* assessment. For this propose, affinity chromatography plays a powerful role in separation, identification and purification of sc pDNA. Recently, amino acids such as, L-arginine [15], L-lysine [16], L-histidine [17] and L-methionine [18] have been used as ligands in affinity chromatography to purify sc

pDNA from impurities. The principle of technique is based on the natural interactions that occur between amino acids and nucleic acids in the biological environment. Thus, the development of new methods to produce and purify vectors such as, pPH600 that contains a sequence that encodes to G4 became essential for the study.

Based on aforementioned, the main proposes of this work are the development of methods in a typically biotechnological process:

1. Biosynthesis of plasmid pPH600 that contains a sequence that encodes G-quadruplex.
2. *Screening* of binding/elution conditions of three affinity chromatography supports for isolation and purification of sc pPH600.
3. Isolation and purification of plasmid sc pPH600 isoform from *Escherichia coli* (*E. coli)* clarified lysate using the select support that allowed the desirable sc pPH600 with high yield and purity level.
4. *In vitro* transcription of sc pPH600 to induce G-quadruplex formation and it detection by circular dichroism (CD).

1.1 - Fundamentals of G4

1.1.1 – Formation and structural diversity of G4

Nucleic acids are flexible molecules that can assume different conformations. Predominantly, DNA in living systems is in double helix structure but in some regions of genome rich in guanines they can exist in an alternative form, known as G4, which can form from DNA or RNA (or other variants, such as PNA) [19, 20]. These structures, as the name suggests, are composed by a set of four guanine nucleotides associated with each other in a cyclic Hogsteen hydrogen-bond (Fig. 2A). These arrays forms G-quartets, which through π-π stacking interactions forms G4 (Fig. 2B). This interactions play an important role in stabilization of the structure [21].

Moreover, the structures are additionally stabilized by monovalent (Rb^+, NH_4^+, K^+, Na^+, Cs^+, Li^+), divalent (Sr^{2+}, Ca^{2+}, Ba^{2+}, Mg^{2+}), or trivalent (Tb^{3+}, Eu^{3+}) cations, through electrostatically coordination with the oxygen atoms of the adjacent stacked G-tetrads, neutralizing negative charged formed in central core of G4 (Fig. 2A and 2B) [22]. The order of cation ability to stabilize and/or induce G4 structures is as follow: $K^+ > NH_4^+ > Rb^+ > Na^+ > Cs^+ > Li^+$ for monovalent cations [23] and $Sr^{2+} > Ba^{2+} > Ca^{2+} > Mg^{2+}$ for divalent cations [24]. However, since K^+ and Na^+ are the main cations *in vivo*, G4 formation is favoured under physiological conditions. The K^+ form is the strongest known coordinating monocation for G4 due to its ionic radius that allow a better coordination with oxygen atoms and higher intracellular concentrations (~140 mM) than Na^+ (5-15 mM) [25]. The location of cations between tetrads depends on the nature of the ions: Na^+ ions are observed in a range of geometries, whereas K^+ ions are always equidistant between each tetrad plane [25].

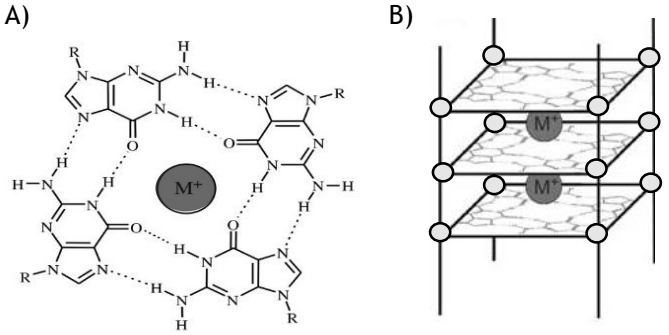


Figure 2 – A) The structure of a G-quartet arrangement with hydrogen Hogsteen bonds and monovalent cation in center of cavity. B) G4 structure. (Adapted from [21]).

Supercoiling affects the potential for DNA denaturation and structure formation, and supercoiling local can vary *in vivo* [14]. With the exception of extreme thermophiles in which, DNA is positively supercoiled to protect it from thermal denaturation, supercoiling has a negative sign, which means that the torsional tension reduce the helicity and facilitates strand separation [26]. In fact, to induce structural perturbations such as, formation of G4 structures the local unwinding or melting of DNA is necessary and both are generated by DNA supercoiling [27]. Since B-DNA is the minimum energy conformation the transition to G4 is unlikely to occur; however, the free energy generated by supercoiling may be enough to unwinding DNA and generate secondary structures [28]. In plasmids the torsional stress results in a build-up of free energy [29]. In this case, two duplexes wind helically up and down around each other and about a superhelix axis [29]. Since supercoiling affects G4 formation, plasmids can adopt different G4 conformations.

The importance of plasmid topology in G-loop formation was analysed by Duquette and collaborators by comparing formation of G-loops DNA on sc, oc and ln plasmids pPH600 and they observe that formation occurs more frequently in sc pPH600 [14]. More recently, another study performed by Gaynutdinov and collaborators shows that short G-rich peptide nucleic acids (PNA), stabilize G4 formation because have complementarity with C-rich strand, binding with more efficiency to supercoiled plasmids even after linearization of the plasmid [11]. In fact, this approach is very similar to formation of RNA-DNA hybrid; however, instead of using the RNA transcript to stabilize the structure, uses the PNA. Moreover, Lv and collaborators demonstrated that DNA gyrase that is commonly found in prokaryotes and are responsible to the formation of sc in their circular DNA and pDNA can readily drive the formation of G4 under physiological conditions [30]. This arise the importance of obtain pure sc pDNA which is more efficient in G4 formation.

G4 are diverse and polymorphic structures that can be correlated with several features, such as, number of molecules involved in formation of structure, chain orientation of DNA strands, guanine nucleoside glycosidic bonds and loop topology [31].

In the case of number of molecules involved in formation of structure, G4 can be intramolecular (uni-molecular) or intermolecular (bi-molecular or tetra-molecular) [31]. Intramolecular G4 formed by single-stranded are of intensive current research interest due to their potential formation in telomeres and oncogene promoter sequences. Inversely to tetramolecular G4, intramolecular structures are more complex and form quickly exhibiting high polymorphism [31].

The sequences with consecutive guanines in uni-molecular and bi-molecular G4 form loops that are positioned on the exterior of core, helping in stabilization of structure. Loops can be diagonal (Fig. 3E), lateral or chain-reversal (also termed propeller) (Fig. 3D), and the presence of a particular loop type in a structure is dependent on the number of G-quartets comprising a G4, on loop length and on sequence [32].

All G4 have four distinct phosphodiester chains and orientation of DNA strands may be parallel (Fig. 3A, 3B, 3C and 3D), anti-parallel (Fig. 3E) or hybrid (Fig. 3F). G4 conformation is influenced by both the DNA sequence and the conditions used to promote folding of G4 such, nature of stabilizing cation. In biological context, the tendency is to form parallel G4 structures in presence of K^+, due to its relevance in intracellular media [33].

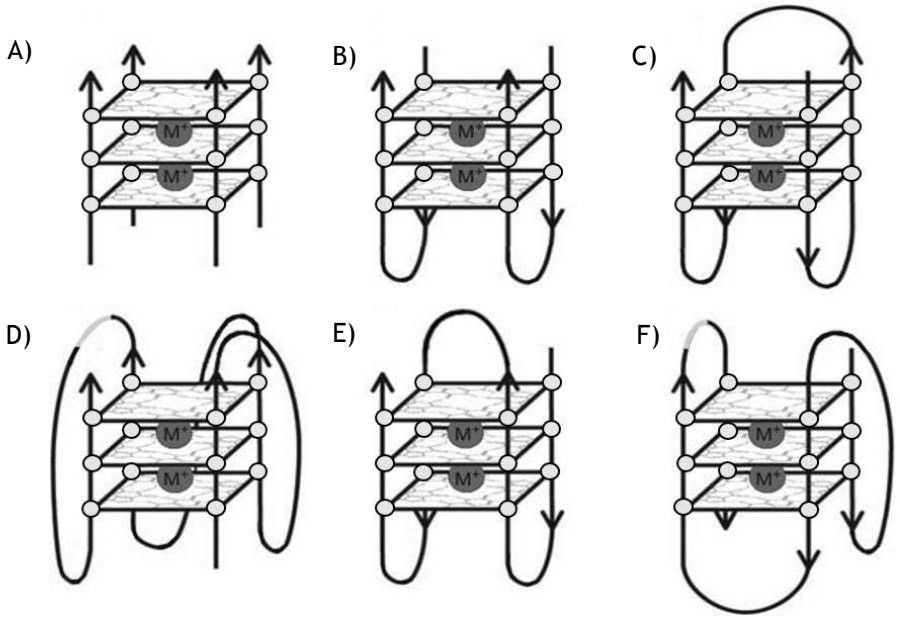


Figure 3 - G4 stoichiometries and folding patterns. A) Tetramolecular parallel structure; B) Bimolecular antiparallel structure; C) Unimolecular antiparallel structure with alternating parallel strands; D) Unimolecular parallel structure with three double chain reversal loops; E) Unimolecular antiparallel structure with adjacent parallel strands and a diagonal loop; F) Unimolecular mixed structure with three parallel and one antiparallel strands (Adapted from [21]).

All parallel G4 are consistent and have all guanine glycosidic angles in an *anti* conformation, not displaying significant polymorphisms. In contrast, anti-parallel G4 have both *syn* and *anti* guanines, arranged in a way that is particular for a given topology and set of strand orientations, since different topologies have the four strands in different positions relative to each other [34].

1.1.2 - Genome location and functions of G4

Actually, there are powerful tools that allows the prediction of G4-forming sequences in human genome. The sequences that are able to form G4 consists at least four runs of guanines which usually contains at least three guanine residues (e.g., $[G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}]_{\geq 4}$, where N is any nitrogen base (A; T; G or C)) [10]. If the runs of $G_{\geq 3}$ are shorter, much longer or mutated the structural stability decreases, significantly. This prediction was expressed as an algorithm, called *quadparser*. Huppert and Balasubramanian apply *quadparser* and reveal that human genome contains approximately 376.000 sequences that have potential to form G4, simultaneously, which are less than would be expectable by chance [35].

Predicted G4 sequences are not located randomly throughout the genome and tend to cluster together in particular important regions of human genome such as, telomeres, regulatory regions of oncogenes, rDNA and immunoglobulin switch region (S region). It is thought that the location of these sequences are a strong evidence of it crucial role in various cellular processes including telomere maintenance, replication, transcription, translation, genetic and epigenetic instability [36].

Telomeres are specialized nucleoproteins that protect extremities of linear chromosomes in order to distinguish them from unwanted double strand breaks. These nucleoproteins are constituted by repetitive sequences of variable lengths characterized by runs of guanines on one strand. Bearing runs of consecutive guanines, 5′ to 3′ telomeric strands from most eukaryotes may potentially fold into G4 [37]. At the G-rich telomere tails, the presence of G4 inhibits extension by telomerase, which is up-regulated in cancer cells, as well as negatively regulating oncogene's transcription [38].

Aside from telomeres, one of the first region considered for the presence of G4 was gene promoters. Notably, one evidence that G4 have an important function in regulatory processes is your high prevalence in oncogenes and low prevalence in tumor suppressor genes [39]. Based on hallmarks of cancer proposed by Hanahan and Weinberg [40] several oncogenes have been studied, among them are c-Myc, c-Kit and KRAS (self-sufficiency); pRb (insensitivity); Bcl-2 (evasion of apoptosis); VEGF-A (angiogenesis); hTERT (limitless replication); and PDGF-A (metastasis) [41]. Actually, various efforts have been done to design therapeutic targets with high specificity for these regions.

Eukaryotic rDNA is G-rich, rapidly transcribed and contain many sequences that can form G4 in non-template strand. Nucleolin is a very abundant protein in nucleolus and binds with high affinity (K_D=1 M) to G4 through two separated domains that recognize G4 as natural binding targets independent of sequence context. The strong interaction suggest that the presence of nucleolin stabilize G4, prevents renaturation of the duplex and is involved in rDNA transcription, replication, or recombination [42].

The class switch recombination (CSR) of the immunoglobulin loci is essential to normal B-cell development and immune response and the process occurs in response to antigen stimulation and costimulatory signals. The process allows mature B cells to alter effector portion modifying

the mode of antigen clearence without affecting antigen specificity. The mechanism of immunoglobulin isotype switching is regulated, irreversible and alter genomic structure by joins a rearranged and expressed variable region to a new downstream constant region, deleting the DNA between as an excised switch circle (Fig. 4). CSR is targeted to short and repetitive G-rich regions (20-80 bp), called switch regions. S regions are approximately 1–10 kb in length and are located upstream of each of the C_H regions that participate in switch recombination: Cμ, Cγ, Cε, and Cα. There is no S region upstream of Cδ, and RNA processing rather than DNA recombination regulates Cδ expression [43, 44].

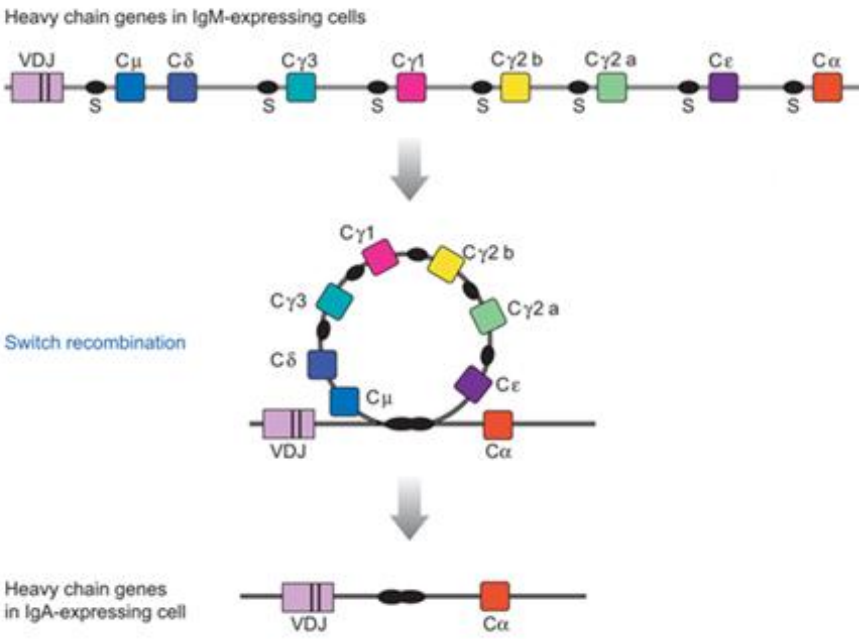


Figure 4 – Diagram of Class Switch Recombination from IgM to IgA (CSR). The murine heavy chain locus is shown before and after recombination from μ to *a*, which results in expression of IgA antibodies. Deleted sequences can be recovered in circular DNA molecules from B cells which have recently completed switch recombination. V, variable region; C, constant region; and S, switch region (Adapted from [43]).

1.1.3 – Transcription of S regions

The process of gene expression begins with transcription in cell nucleus. The transcription is initiated in transcription start site (TSS), which is located immediately downstream of promotor. The region is not transcribed but play a fundamental role in gene transcription.
The transcription factors (TFs) binds with specificity to nucleotide sequences in promotor and assist the binding of enzyme that catalyses RNA synthesis, called RNA polymerase [45].

The binding of the general TFs on their own produces low levels of transcriptional activity; however, the transcriptional activity is increased or decreased by other TFs that binds specifically to another regions of DNA called enhancers or silencers [45].

Potential G4 forming sequences were identified in the regions flanking the TSS suggesting and active role in transcription. Recent studies investigated the influence of transcription in G4 formation [46]. G4 can be induced far away from TSS, functioning as silencer or enhancer of transcription. Each S region has a dedicated promoter, and transcription through the S region is necessary to activate recombination and target it to specific S regions [47]. Transcription alters the structure of S region DNA, creating an active molecular partner for recombination.

Transcription of G-rich templates such as the S regions, either *in vitro* or *in vivo*, causes characteristic large loops to form, which are hundreds of base pairs in length and are called G-loops [48]. G-loops contains a stable, cotranscriptional RNA-DNA hybrid on the C-rich template strand and intramolecular G4 DNA interspersed within single-stranded regions on the G-rich strand (Fig. 5A) [48]. These structures are involved in a range of processes, such as regulation in *E. coli* plasmid replication [48]. The persistence of such hybrids in living cells can result in genomic instability, and conserved factors associated with RNA processing normally prevents the formation of such hybrids. The hybrid formation prevents further transcription, suggesting a regulatory role of these structures [48].

G4 DNA in G-loops is the target for TFs genetically linked to switch recombination. Two switch recombination proteins MutSα (Fig. 5B) and AID binds to G4 structures [46, 49]. The capacity to prevent or induce the formation of G4 with some transcription factors in S region seems to be a fundamental role in immunoglobulin class switch recombination.

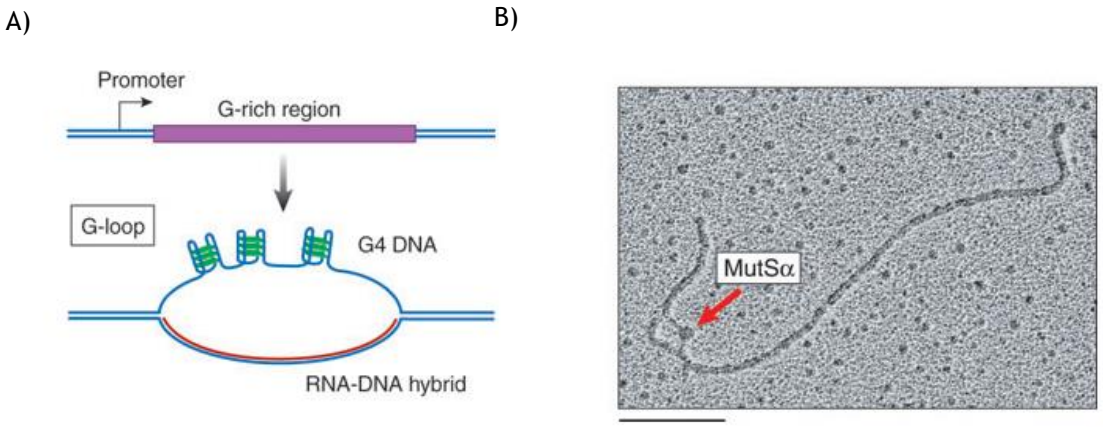


Figure 5 - G-loop formed in a transcribed G-rich region and bond of MutSα to G4 structure. (A) Transcription of G-rich region resulted in G-loop and cotranscriptional RNA-DNA hybrid on C-rich region; (B) Bond of MutSα to G4 structure in G-rich template (Adapted from [47]) .

1.1.4 – Circular dichroism of G4 structure

Circular dichroism (CD) is an excellent method for rapid determination of secondary structure and is widely used in structural studies of biomolecules. The technique takes advantage from the differential absorption of left-handed and right-handed circularly polarized light. The light is polarized by passing through suitable prisms or filters its electric field will oscillate sinusoidal in a single plane. This sinusoidal wave can be visualized as the resultant of two vectors of equal length, rotating in opposite directions, one in clockwise direction and the other in counter clockwise direction. In order to show signal the chromophore should be chiral (optically active) to absorb with different extent the circularly polarized light and to have different indices of refraction for the two waves. The result is the rotation of plane of light and the creation of a new vector, resulting from addition of the two vectors in opposite directions, which traces an ellipse. Generally the results are reported in terms of ellipticity (θ) in degrees [50]. For this reasons the technique is suitable for structural evaluations of biomolecules such as, proteins and nucleic acids.

In particular case of nucleic acids, CD arises from the asymmetric backbone sugars and the helical arrangement of its constituents. The task to determine secondary structures is not easy due to the length of nucleic acids and the empirically determination of structure [51]. Despite these drawbacks CD have advantages over the other methods because is handy, fast, highly sensitive and relative inexpensive, making technique ideal for tracing conformational transitions between nucleic acids arrangements [51]. Moreover, the technique is particularly powerful for monitoring structural changes resulting from changes in environmental conditions such as temperature, ionic strength, and pH [51].

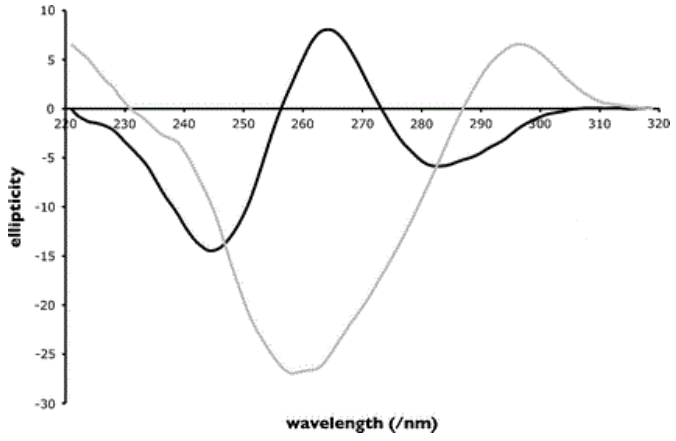


Figure 6 – Circular dichroism spectra with parallel (black line) and anti-parallel (grey line) G4 structure (Adapted from [20]).

In recent years the technique has been widely used to determine the topological conformation of G4. The CD spectra of a reference set of DNAs of known G4 structure, with examples of both anti-parallel and parallel types, have been obtained. In general, a peak in CD at 260 nm wavelength and at 240 nm is descriptive of an all-parallel structure, whereas a peak at 295 nm and a trough at 260 nm describes an antiparallel structure (Fig. 6) [20].

1.1.5 - Role of G4 in cancer and neurological diseases

Cancer, aging and neurological diseases represent an enormous source of health and financial burden in worldwide. Together affect hundreds of millions of people in world resulting in increased associated costs. G4 are intrinsically associated to aging process and disease development. The presence in telomeres, promoter regions of oncogenes and the formation of non-B DNA structure in triplex expansion-associated a neurological diseases and cancer are a strong evidence of its importance [52]. In this subsection it will be presented examples inform the literature about genes and proteins involved in these processes.

1.1.5.1 - Aging and cancer

The processes of aging and cancer are closely related and are directly associated to telomere shortening, risk of disease development, decreased of reparation ability and activity of regeneration processes [52]. Telomeres are DNA–protein complexes that cap the ends of linear chromosomes and provide protection against gene erosion at cell divisions, non-homologous end-joining's and nuclease attacks [53]. In this point of view, there seems to be a pathological paradox. On the one hand, telomere shortening accompanies the human aging, suggesting that erosion of chromosomes could be a natural mechanism to prevent the immortal replication of cells that characterizes cancer growth. On the other hand, telomere shortening is observed in many cancer cells that gained an unlimited proliferation potential [54]. Therefore, using telomeres to prevent aging may lead to an increase risk of cancer. This fact is supported by the concept that the incidence of cancer rises exponentially in the final decades of life [55], suggesting that human body try to fight aging and results in cancer. Since, telomeres are constituted by repetitive sequences of variable lengths characterized by runs of guanines on one strand may potentially fold into G4. The stabilization of G4 in telomeres results in telomerase inhibition, telomere shortening, cell growth suppression and induction of apoptosis through disruption of telomere maintenance [55].

In addition to their existence in telomere sequence, G4 are also found in promoter regions of oncogenes close to their TSS's [55]. As aforementioned there are many oncogenes that are able to form G4 structures and that can be the target for small molecules [55]. A study reported by

Hurley and collaborators, shows that the promoter region of c-Myc oncogene controls 80%-90% of the transcriptional activity of this gene [55].
Therefore, as in telomeres the stabilization of G4 results in telomerase inhibition, the stabilization in promotor regions of oncogenes inhibit transcription, supressing oncogenic expression [55].

1.1.5.2 - Neurological diseases

Neurological diseases in general are characterized, among other things, by mitochondrial dysfunction and oxidative stress [56]. The expansion of hexanucleotide GGGGCC repeat region in C9ORF72 gene in few hundred repeats causes frontotemporal dementia and amyotrophic lateral sclerosis [56]. G4 from the C9ORF72 gene may arise from their ability to bind and activate cellular Heme, towards to an oxidative stress [56]. In Alzheimer disease sequestration of Heme group by Aß peptide, the causal agent of Alzheimer disease, constitute a loss of function for mitochondrial activity but a gain of function in another way by bond of Aß peptide to Heme group [56]. Therefore, the sequestration of Heme group by G4 suggests that they have an active role in these diseases.
The expansions in FMR1 gene produce pre-mRNA that carry extended regions of G4 structures, abolishing transcription and initiating fragile X syndrome [57]. When the CGG expansion repeats exceed 200 they become hypermethylated, causing transcriptional silencing of the FMR1 gene, resulting in the loss of its encoded protein FMRP, which is a selective RNA-binding protein implicated in regulating dendritic mRNA transport and local protein synthesis at synapses. The expansion of the intronic GGCCTG hexanucleotide repeat in NOP56 causes a unique form of Spinocerebellar ataxia (SCA36), which shows not only ataxia but also motor neuron dysfunction [58]. This characteristic disease phenotype can be explained by the combination of RNA gain of function and miR1292 suppression. In all of cases aforementioned, produced pre-mRNAs appears to sequester essential RNA binding proteins and impar mRNA processing [36].
The PRNP gene encodes the prion protein, which has been implicated in various types of transmissible neurodegenerative spongiform encephalopathies, such as Creutzfeldt-Jakob disease [59]. The expansion of sequence CCCCATGGTGGTGGCTGGGGACAG from 5 to 10-14 repeats in the coding region of this gene cause a dominant form of familial Creutzfeldt-Jakob disease exhibiting early onset and slow progression, which is correlated with misfolding of corresponding prion protein and formation of insoluble protein aggregates in solution [60]. The mechanisms that drive expansions of this G4 repeats are not defined.

1.2 - Gene therapy and DNA vaccination

The significant progress of science during the last decades, namely in molecular biology and genetics, increase the understanding of genetic disorders, leading to revolutionary discoveries in techniques that are based on nucleic acids such as, gene therapy and DNA vaccination [61]. The idea of gene therapy has been around for some time, but only with the development of recombinant DNA technology and ability to transfer and express exogenous genes in mammalian cells received attention and progresses are made [62]. In fact, gene therapy is a relatively new paradigm in medicine with enormous therapeutic potential and is defined as the insertion of genetic material directly into cells of a human patient to correcting a genetic disorder or overexpressing proteins that are therapeutically useful [63]. The DNA vaccination is a type of gene therapy that consists in a pDNA build to express a protein of an aggressor agent inducing both humoral and cellular response [64]. In recent years, this techniques received the attention of researchers because the enormous potential to treat and prevent diseases.

1.2.1 - Gene therapy

Gene therapy can be divided into two categories: somatic gene therapy and germ line gene therapy [65]. The difference between the two categories becomes important and legislation allow the use of somatic gene therapy because the genetic change induced not pass to future generations, contrary to germ line therapy [65]. The major advantage of gene therapy is the treatment or elimination of the cause of disease, while most of conventional approaches treat the symptoms [63]. Initially, the idea of using recombinant DNA techniques was to treat or eliminate the cause of diseases that are based only in a single gene, better known as monogenic diseases [66]. In this case, the aim is the replace of expression of a dysfunctional gene by introducing one or more copies of the therapeutic gene in patient [66]. However, monogenic diseases are not the only target of gene therapy and in recent years many forms of cancer, cardiovascular diseases, neurological diseases and other non-treatable infectious diseases became the primary target [65]. The first attempt of gene therapy to treat a disease targeted a form of severe combined immune deficiency (SCID) due to defects in the gene encoding adenosine deaminase (ADA) [67].

Over the past years, the clinical applications of gene based therapies for treating or preventing various diseases, has been investigated and actually there are 2142 completed or active clinical trials approved (Fig. 7A), in especially to treat or preventing cancer (Fig. 7B).

The efficiency and safety of gene therapy is closely related with different vector systems that have been developed for gene transfer [68].

In general, there are two main categories of vectors, viral and non-viral, but all gene therapy applications depend on the fact that the genetic material needs to be delivered across the membrane and ultimately to the nucleus [68].

Viral vectors have better cell targeting properties due to its intrinsic characteristics and at present stage of development generally give the most efficient transfection [68]. The most used viral systems available for gene therapy are adenovirus, retrovirus, lentivirus and adeno-associated virus [68]. For many years it was thought that inactive virus, that have been modified by deleting some areas of their genomes to makes them more safe, would be the ideal vehicle to introduce genetic material in cells. However, a number of widely reported adverse events have focused the attention on associated risks such as, integration into host cell genome, which could possibly cause oncogene activation or tumor-supressor gene inactivation. These events raised safety concerns related to immunogenicity and toxicity [68].

Non-viral vectors based on pDNA give less efficient transfection; however, the safety concerns are clearer and production can be performed in a simply way when compared with the viral ones [69]. This characteristics made it vector very attractive for genetic vaccination and consequently its use in gene therapy arises in recent years (Fig. 7C).

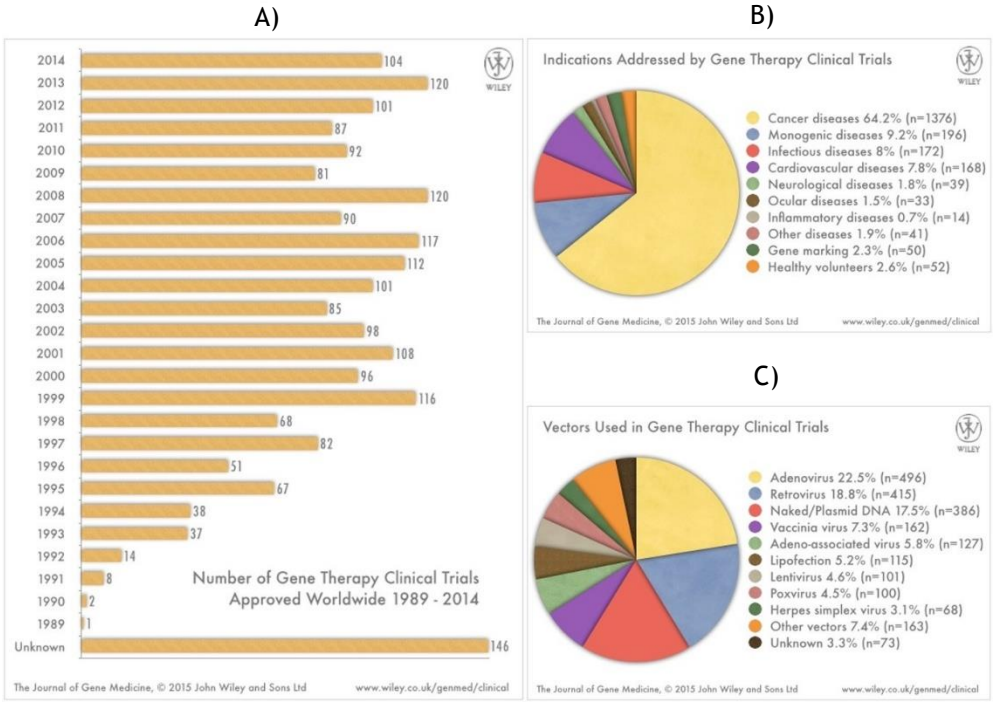


Figure 7 – Compiled data on gene therapy clinical trials. (A) - Clinical trials approved for Gene Therapy 1989 - 2015. The total corresponds to 2142 clinical trials approved. (B) – Diseases addressed and (C) – Vectors used in clinical trials.
(Available at *www.wiley.com/legacy/wileychi/genmed/clinical/, May 2015*)

1.2.2 DNA vaccines

Currently, vaccines can be classified into five types: live-attenuated, inactivated microorganisms, subunit vaccines (subunits, polysaccharides and protein conjugates), toxoids and genetic based [70]. However, the use of both live attenuated and inactivated vaccines raises several questions in terms of safety [71]. These limitations need to be surpass using new vaccine platforms that offer broader immunogenicity [71], as in case of DNA vaccines that consist in pDNA build to express an encoded protein following administration *in vivo* and subsequent transfection of target cells [64]. DNA vaccines have a wide range of economic and technical advantages when compared with conventional ones. DNA vaccines mimics the effects of live-attenuated vaccines because have the capacity to induce both immune and humoral response, through $CD4^+$ and $CD8^+$ T cell responses [72]. The mechanism consists in introduction of pDNA with a gene of interest in nucleus, where the cell machinery recognizes and induces the transcription of gene with subsequent protein production in cytoplasm [72]. The protein is recognized by specific cells and induce the humoral and immune response (Fig. 8) [72]. Moreover, DNA vaccines can also act as adjuvant of cellular response due to the presence of unmethylated CpG islands that are stimulatory DNA specific sequences [70]. The production cost in large scale is lower than the conventional vaccines, the quality control is much easier, since not need cold chain, because this vaccines are stable at room temperature and can be lyophilized, making distribution easier in remote locals with low economic resources [72]. Until moment, there are four DNA vaccines approved for veterinary use: against West Nile virus in horses; infectious hematopoietic necrosis virus in salmon; malignant melanoma in dog as well as delivery of growth hormone releasing hormone in pigs [72].

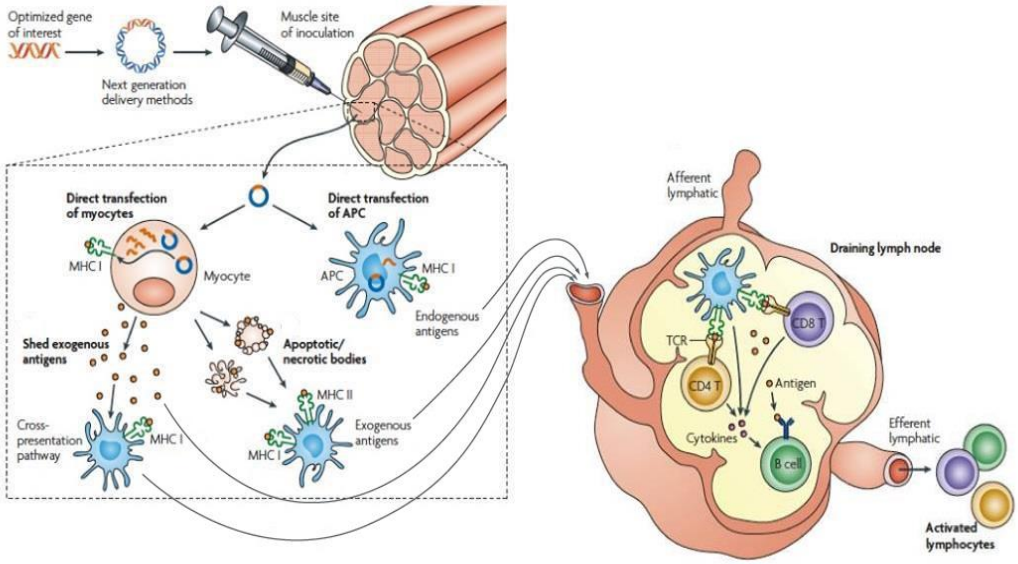


Figure 8 - Mechanism of DNA vaccines. Induction of immune and humoral responses. (Adapted from [72])

1.3 - pDNA

The importance of plasmids to molecular biology and their impact in biotechnology industry make them an indispensable molecular tool in life science research [73]. Plasmids are replicative extrachromosomal entities, normally the size of circular double stranded DNA is between 1 kbp and 1000 kbp, meaning that they are very large molecules [74]. Each strand of this molecule is composed by a linear polymer linked by phosphodiester bonds between the hydroxyl group of the 3' sugar carbon of a nucleotide and the phosphate group of the 5' sugar carbon of the adjacent nucleotide [75]. The phosphate groups are negatively charged when pH is maintained above 4 [75]. The winding of two anti-parallel DNA strands around each other and around a common axis forms the double helix structure with highly hydrophobic grooves accessible to molecule and solvent [76]. Plasmids contain an origin of replication that is recognized by DNA polymerase allowing growth in bacteria and amplification of large quantities of pDNA for purification [77]. They have also a gene coding for antibiotic resistance and a complete eukaryotic expression cassette composed of the transcribed region domain inserted between an enhancer and terminator. The region is flanking by restriction enzyme sites that allow insert the desired genes. Plasmids can exist in three different conformations (Fig. 9): supercoiled (sc); open circular (oc) and linear (ln) [75]. However, the most relevant isoform in biological context is supercoiled because transfection and gene expression is highly dependent on the degree of superhelicity [78] and can assume different secondary structures of DNA more efficiently [14]. The others conformations are derived from supercoiled by both single strand nick in phosphodiester backbone (oc) and double strand nick in phosphodiester backbone (ln) [76]. The nicking occur by specific nucleases, high temperatures and mechanical stresses [76].

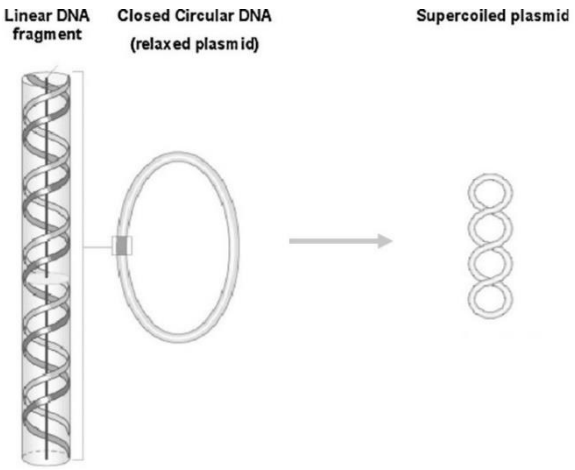


Figure 9 – Different conformations of pDNA: supercoiled (sc); open circular (oc) and linear (ln). (Adapted from [75])

1.3.1 – pDNA specifications

Plasmids are the crucial tool in many scientific areas, and large scale processes should be designated to produce a certain amount of plasmid within certain specifications. Thus, there is a need to obtain pure sc pDNA without impurities to allowing efficient analysis and applications. The quality of the plasmid should be assessed by safety, potency and purity, in order to validate the process [75]. For a product which intended to be used in humans, the removal of impurities during the manufacturing process is mandatory to avoid side effects once administered in patients [79]. The impurities for production of recombinant proteins in *E. coli* is also the same for pDNA production [79]. Therefore, in pDNA production the main impurities are host nucleic acids (gDNA and RNA), proteins and endotoxins).

The regulatory authorities such as, Food and Drug Administration (FDA) and European Agency for the Evaluation of Medical Products (EAEMP), have rigorous guidelines related to the presence of these impurities (Table 1). Recently, a new analytical method have been developed to specifically quantify the amount sc pDNA, oc pDNA and RNA present in pDNA sample [80].

Table 1 - Specifications and recommended assays for assessing pDNA quality. (Adapted from [75])

Requirement	Analytical assay	Specification
Proteins	BCA assay SDS-PAGE	Not detectable <0.01 µg/dose
RNA	Agarose gel electrophoresis	Not detectable
gDNA	Hybridisation blots, PCR, fluorescence	<0.05 µg/µg plasmid <0.01 µg/dose
Endotoxins	*Limulus* ameobocyte lysate (LAL) assay	<0.1 EU/µg plasmid <5 EU/kg body weight
Plasmid Homogeneity	Agarose gel electrophoresis	>90% supercoiled form
Potency	Cell transfection	According to application

1.3.2 - pDNA production

The development of processes to achieve the production of large quantities of highly purified supercoiled pDNA have the interest of industry and researchers due to the expansion of gene therapy and DNA vaccination [79].

The process development for pDNA production starts on a bench scale with the upstream process, in which the design of vector, choice of host, as well the selection and optimization of fermentation conditions, followed growth of cells and plasmid production and finally, the downstream process, in which the main goals are the cell lysis and purification of sc plasmid DNA (Fig. 10). These three stages of process development (upstream processing, fermentation and downstream processing) are integrated and must not be approached on an individual bases [79].

In general, plasmid production is performed in *E.coli*, a bacterium with a history of safe use in bio-industry, which growth to an optimal cell concentration, while pDNA is replicated autonomously in intracellular environment [81]. Although other host systems have been used [82]. A single bacterial cell may generate high plasmid number in appropriate conditions. There are two main goals associated with fermentation, the first is maximize the yield of pDNA obtained, while minimizing process that gives rise to impurities such as, gDNA, RNA, proteins and endotoxins, and improves the pDNA stability. In order to achieve this objective should be taking in consideration several parameters such as, the choice of host system and vector, as well the formulation of culture medium. Formulation of culture medium should consider the elemental composition of *E.coli*, pDNA replication and operative conditions. Finally, the second objective of the fermentation process is maximize the number of copies of sc pDNA in detriment of other isoforms that are usually treated as impurities [83, 84].

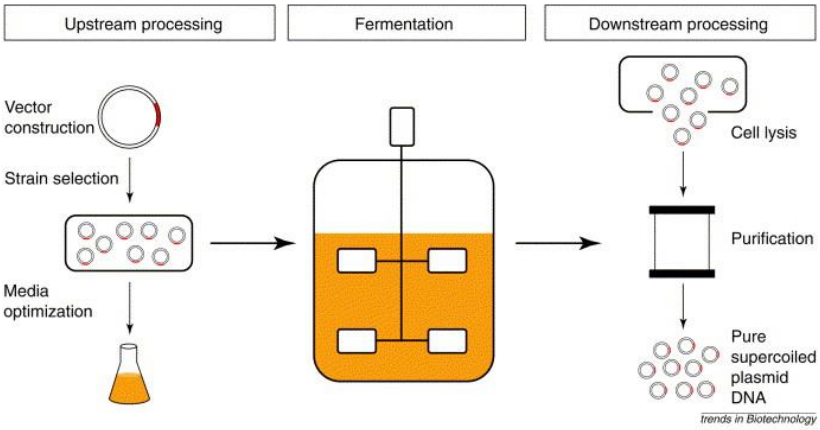


Figure 10 - The three stages of sc pDNA development process. The upstream, fermentation and downstream processing. (Adapted from [79])

1.3.3 – Downstream Processing

Since pDNA accounts for less than 3% of the total contents of an *E. coli* cell (depending on plasmid size and copy number) the major obstacles are encountered during downstream process stage (Fig. 11), in which is necessary eliminate other cellular components of the host strain that are usually treated as impurities [83].

After fermentation, the recovery of cells from the broth by microfiltration or centrifugation initiates the downstream processing [79]. Then, cells are ressuspended in appropriate buffer to initiate the cell lysis [79]. The buffer contains ethylene diamine tetra-acetic acid (EDTA) and often carbohydrates such as, glucose and sucrose to facilitate the lysis step and protect plasmid against mechanical stresses [79].

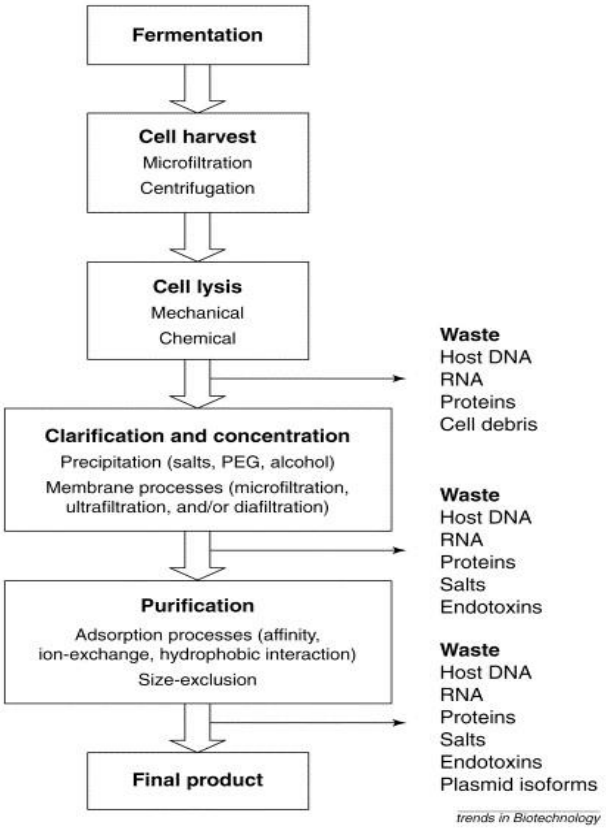


Figure 11 - Process flow sheet for the purification of sc pDNA. Each step of the process is a unit operation that can be improved. The eliminated impurities in each step are also indicated. (Adapted from [79]).

Although, some products are released by cell to extracellular environment, many others need disruption of host cell to release the desired product. The lysis is basically the disruption of cell and release of all intracellular components to medium by mechanical, enzymatic or chemical methods. Although it is also the most critical and troublesome of all unit operations in downstream processing [79] Many mechanical methods such as, sonication, bead milling microfluidisation and homogenisation have been studied; however drawbacks such as, the low plasmid recovery and shear stress that damage polynucleotides making these methods unsuitable to large-scale production [85]. Another alternative is the enzymatic strategy, but is not accepted by the regulatory agencies since have rigorous requirements for animal-derived materials. Thus, the most widely used method is a modification of alkaline lysis [86]. Alkaline lysis relies on the disruption of cells at high pH with NaOH and in presence of sodium dodecyl sulphate (SDS), followed by the release and denaturation of gDNA, cell wall material and most of the cellular proteins [86]. Thereafter, the release of gDNA increases significantly the viscosity of the solution. In addition, to maintain gDNA with the highest weight possible to subsequent step of precipitation the mixing of the solution should be very gentle.

The step after lysis procedure is the neutralization with salt, usually potassium acetate, which promotes the precipitation of denatured gDNA, protein-SDS complexes and cellular debris. This precipitate is removed by centrifugation or filtration, followed by precipitation of plasmid with 0.6-0.7 volumes of isopropanol or 2 volumes of ethanol. Thereafter, for the preparation of plasmids extracts to be applied in purification steps it is recommended the use of concentration and clarification techniques that are designed to remove proteins and host nucleic acids and to increase the plasmid mass fraction. The removal of high weight RNA is achieved with the presence of endogenous nucleases present in the final sample of alkaline lysis. The proteins are removed by salting out using a high concentration of chaotropic salts. After this the plasmid is normally concentrated with polyethylene glycol (PEG) that allows a buffer exchange to subsequent use in next purification step [86].

1.3.4 pDNA purification

In particular case of nucleic acids, the recently developments in gene therapy and DNA vaccination increases the interest in purifying pDNA; however many other applications arise. Recent reports show that sc pPH600 have more able to form G4 structures than oc or ln isoforms. For that reason the isolation and purification of pPH600 becomes essential to further studies of *in vivo* G4 stabilization [14]. Liquid chromatography is the operation of choice, either as a processing step or as an analytical tool for monitoring process development and quality control, since it is simple, robust, versatile and highly reproducible [76, 87]. The overall process has many concerns that should be taking in consideration. These concerns are related with characteristics of process namely, stationary phases and ligands, the structural nature of pDNA, like, size, shape and conformation, rheological properties of lysates, structural and chemical similarities between pDNA and impurities [76]. Most of the critical impurities share similar characteristics such as, size (gDNA, endotoxins), hydrophobicity (endotoxins) and negative charge (RNA, gDNA and endotoxins) with pDNA [88]. Moreover, the regulatory agencies consider also oc pDNA and ln pDNA impurities making process more difficult since this molecules have considerably similar characteristics with sc isoform [76]. In fact, the separation is only achieved because at atomic level exist some differences between these molecules [89].

Therefore, various types of chromatography have been used to separate sc pDNA from impurities taking advantage from physical, chemical and biological properties such as, negative charge (anion exchange chromatography), hydrophobicity (hydrophobic interaction chromatography and reversed-phase chromatography), size (size-exclusion chromatography) and biospecificity (affinity chromatography) [90].

1.3.4.1 Size exclusion chromatography

Size exclusion chromatography distinguishes and purify pDNA on the basis of size [91]. The reduction of plasmid hydrodynamic radius due to supercoiling is the basis for pDNA separation [79]. The chromatograms are characterized by two peaks, in which the first peak contains all forms of pDNA and gDNA and the second peak contains RNA and other small impurities such as, oligonucleotides, endotoxins and salts [92].

However, gDNA in the lysates is highly heterogeneous mixture that is generated by alkaline denaturation. It is made up of species with low (< 0.7 kbp) and high (>20 kbp) molecular mass DNA, either single or double stranded [93]. Thus, the separation of gDNA is extremely difficult, making process of separation poor and not reproducible. Smaller molecules such as, RNA, oligonucleotides, endotoxins and salts are easily separated from pDNA. Therefore, size exclusion chromatography is normally used together with other chromatographic techniques such as, anion exchange, hydrophobic interaction and affinity chromatography to obtain pure sc pDNA [76].

The technique is usually employed in polishing step enabling the removal of residual impurities and simultaneous buffer exchange [76]. The recovery of pure sc pDNA with about 70% could be achieved [76].

1.3.4.2 Anion-exchange chromatography

In general, anion-exchange chromatography of nucleic acids takes advantage of the interaction between the negatively charged phosphate groups in the backbone and the positively charged ligands like tertiary or quaternary amines on the matrix surface [90]. For binding and displacing of nucleic acids to ligands, the ionic strength of mobile phase is adjusted according to the aim. After binding, which usually occurs at low salt concentrations, a salt gradient is employed to displace the different nucleic acids that should elute in order of increasing charge density, which in turn is a function of chain length and conformation [81]. The size is an important parameter that should be taking in consideration, since how larger is molecule, the strongest is its binding to the charge matrix. This behavior is directly correlated with the number of phosphate groups in molecules. The spatial conformation is particularly important when the aim is the separation of pDNA isoforms (oc and sc). Since events like supercoiling or compaction can expose more the phosphate groups available to interact with charged matrix. Therefore, the separation of sc pDNA from oc pDNA is possible when a gradient is optimized in order to elute first oc pDNA followed by elution of sc pDNA [90]. However, it was suggested that base composition also affects the sc pDNA purification with anion-exchange chromatography [94]. Several inversions in retention time as a function of chain length were attributed to a higher AT content of the more retarded molecule [81]. Despite many reports about anion-exchange chromatography to purify pDNA, the process seems to be inefficient when the aim is the separation of sc pDNA from a complex mixture with several impurities. Thus, the resolution from impurities with a similar chemical composition and structure (gDNA, RNA) or charge (endotoxin) is insufficient.

1.3.4.3 Hydrophobic interaction chromatography

Hydrophobic interaction chromatography is a widely used technique that takes advantage of different hydrophobicity between pDNA conformations, single stranded nucleic acids, impurities and endotoxins to isolate and purify sc pDNA [95]. The high salt concentrations in mobile phase allow the elution of pDNA molecules, while impurities content of single strand such as, gDNA, RNA and endotoxins binding to hydrophobic ligands immobilized in a matrix [81]. The elution of these impurities is usually achieved with a decrease in salt concentration that allows the elution of species by increasing order of hydrophobicity or decreasing order of polarity [76]. The gDNA and RNA have more strongly interaction with matrix because the exposition degree of DNA bases is much greater than in pDNA [95]. Endotoxins are another

impurity, and have a lipid moiety allowing the strongly interaction with matrix [95]. These differences are also the principle to separation of sc pDNA from oc and ln.
In fact, sc pDNA have a more compact structure due to superhelicity and this allow a high exposition degree of DNA bases [81]. This difference allow the isolation and purification of sc pDNA based on hydrophobic interactions between hydrophobic grooves in pDNA and hydrophobic ligands.

1.3.4.4 Affinity chromatography

Unlike anion-exchange chromatography and hydrophobic chromatography, in affinity chromatography is not possible to pinpoint a single type of interaction in plasmid binding [96]. Affinity chromatography takes advantage of reversible biological interactions for the separation, analysis and purification of specific analytes within a sample [97]. In affinity chromatography, the key factor is the immobilized ligand, which can be bond by a spacer to an insoluble matrix, and determines the efficiency of chromatographic process [98]. The choice of matrix and conditions to be used should be taking in consideration, as well the properties of biomolecules, in other words, the physical, chemical and thermodynamic nature of the interactions [97]. Moreover, the matrix must be inert and should not interfere in separation processes. After binding of biomolecule by biological recognition with ligand immobilized in matrix and elution of unwanted materials, the elution steps of target can be performed specifically or non-specifically, by using a competitive ligand or by changing the pH, ionic strength or polarity, respectively [91]. In this context, affinity chromatography have the advantages of reduce additional steps, increasing yields and improving processes economics [91]. On the other hand the technique has some drawbacks associated with biological nature of ligands and process scale up. These ligands tend to be fragile and associated with low binding capacities. In the recent years many efforts have been done to design new ligands to purify a wide range of molecules.
The design of new ligands are based either in selectivity of natural systems as in capacity and durability of synthetic systems allow the application of several affinity approaches for nucleic acids purification (Table 2).

Table 2 - Different affinity chromatography approaches. (Adapted from [91])

Affinity type	Principle	Specific binding	Advantages	Limitations
Immobilized metal ion	Chelating ligands charged with divalent metal ions specifically interact with aromatic nitrogen atoms through π-d orbital overlap	Single-stranded nucleic acids (particularly purine bases)	Efficient resolution of RNA from gDNA and pDNA; High endotoxin removal; Separation of denatured pDNA	pDNA in the flowthrough; Incomplete RNA capture in complex mixtures; Co-elution of all DNA forms
Triple-helix	Specific sequences present on DNA are recognized by an immobilized oligonucleotide, forming a triple-helix	Double-stranded DNA	Discrimination of different plasmids based on their sequence; sc pDNA isolation in one chromatographic step; Reduction	Loss of pDNA during wash step; Low yields; Slow kinetics of triple-helix formation; Long chromatographic run times
Polymyxin B	Immobilized polymyxin B specifically recognizes the lipid structure of endotoxins	Endotoxins	Elimination of endotoxin contamination from pDNA preparations	Non-specific interaction of ligands with pDNA; Poor yields Toxicity of polymyxin B
Protein-DNA	A protein or protein complex immobilized on the matrix specifically recognizes a DNA motif	pDNA	Discrimination of different plasmids based on their sequence; pDNA isolation from clarified lysates; Elimination of proteins and RNA from preparation	Relatively low yields; Contamination with gDNA
Amino acid-DNA	Multiple interactions occur between immobilized amino acids	Sc pDNA	sc pDNA purification in a single chromatographic step; Efficient elimination of RNA, gDNA, proteins and endotoxins	Elution with high salt concentration

1.3.4.4.1 Aromatic amino acids as ligands in affinity chromatography

As aforementioned the key factor of any affinity separation is the type of ligand used within the column [98]. Amino acids are used as ligands because they act as electron acceptors, have high physical and chemical stability and are easy to obtain in large quantities at low cost. Several amino acids have been recently used as bio-specific ligands in affinity chromatography matrices to purify sc pDNA [15-18]. Namely, L-arginine have been recently used to fully separate sc pDNA from oc isoform through multiple non-covalent interactions including electrostatic and hydrogen bonds [99]. Other bio-specific support such as, L-histidine-agarose was described to separate pDNA isoforms from native sample (oc + sc)[17] and to purify sc pDNA from a clarified *E. coli* lysate [93]. The underlying interaction mechanism is mainly through hydrophobic interactions between the nucleic acid bases and L-histidine [93]. Also o-phospho-tyrosine was used to purify pre-miR29 [100]. Based on this purpose, aromatic amino acids can be explored as affinity ligands, since have an aromatic ring in side chain which allows amphipathic interactions (imino group hydrogen bonding, specific dipolar interactions, cation-π and hydrophobic interactions) [101]. The structure of nucleic acids is determined by two types of interaction: hydrogen bonding and π-π stacking [102], and are the main interactions that aromatic amino acids are able to do. Thus, L-tryptophan and L-tyrosine can be explored as affinity ligands since can promote interactions that occurs between nucleotides in double helix formation. Many techniques have been employed to characterize at atomic level the type and extension of interactions involved in affinity process between amino acids and pDNA. Surface Plasmon resonance (SPR) and nuclear magnetic resonance (NMR) were also used in our group to study both mechanism of binding and structural composition of the chromatographic supports [103-105].

In present work, the three stages of a biotechnological process have been explored in order to visualise the G4 structure. Firstly, the *E. coli* DH5α was used to produce the plasmid pPH600 that contains a fragment of Sγ3 switch region of murine and are able to form G4 structures. Since sc pPH600 have more efficiency in G4 structure formation there is a need to obtain sc pPH600 in detriment of another pPH600 conformations. Therefore, after plasmid production, the stage of downstream process is fundamental to obtain pPH600 without impurities with special focus on affinity chromatography. The biological interactions between amino acids (L-arginine, L-tryptophan and L-tyrosine) and pPH600 allows the fully separation of sc pPH600 from impurities. To prove the formation of G4 structures, the transcription of S region are performed *in vitro* and the resultant transcript are incubated with 50 mM KCl to stabilize the structure through K^+ for further evaluation in CD.
The objective of present work was the production, purification and induction of G4 structures for further *in vivo* detection of G4.

CHAPTER II

MATERIALS AND METHODS

2.1 - Materials

All solutions were freshly prepared with ultrapure double distilled water purified with MilliQ system from Millipore (Billerica, MA, USA). Solutions were filtered through 0.22 µm pore size membrane (Whatman, Dassel, Germany) and degassed before use.
Sepharose CL-6B was purchased from GE Healthcare (Uppsala, Sweden). $(NH_4)_2SO_4$, $NaHCO_3$ and K_2HPO_4 were obtained from Panreac (Barcelona, Spain). NaCl, HEPES acid and isopropanol were purchased from Fisher Scientific (Fair Lawn, NJ, USA). L-tryptophan, L-tyrosine, NaOH, $NaBH_4$ and KH_2PO_4 were obtained from Sigma-Aldrich (St. Louis, MO, USA). NZY Miniprep kit, NZY Maxiprep kit, Tris(hydroximethyl)methylamine (Tris) and GreenSafe Premium were acquired from NZYTech (Lisbon, Portugal). Hyper Ladder I (Bioline, London, UK) was used as DNA molecular weight marker. The Maxima® SYBR Green/Fluorescein qPCR Master Mix (Thermo Fisher Scientific Inc.) was used for gDNA quantification. All salts used were analytical grade.

2.1.1 - pDNA

The 3562 bp plasmid pPH600 used in this experiments, was kindly provided by Prof Nancy Maizels from Washington Medical School, USA. Plasmid pPH600 derivate from commercial plasmid pBluescript KS+. Plasmid encondes a sequence of a G-rich region of immunoglobulin switch region Sγ3 of murine that can form G4 (Fig. 12).

2.1.2 - Bacterial Strain

The bacterial strain *E.coli* DH5α was used in this experiments as system host to amplify the plasmid pPH600.

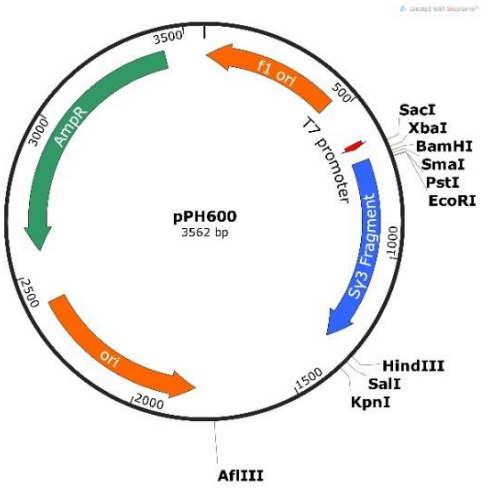


Figure 12 – Plasmid pPH600 map with more important unique cutters restriction enzymes. The restriction map was generated with SnapGene® Software.

2.2 – Methods

2.2.1 - Plasmid Production

This section provides information about experimental procedures that were performed since cell competence until fermentation. The aim is the production of high yields of pPH600 for further use in purification steps.

2.2.1.1 - Competent cells

A colony of *E. coli* DH5α cells was inoculated into LB media (50 mL) and grown overnight at 37°C and 250 rpm. From the culture containing the *E. coli* DH5α cells 2 mL was taken and added to sterile media (250 mL) and then grown at 37°C for approximately 2 h until an optical density 0.3 at 600 nm was reached. The *E. coli* DH5α cells were pelleted by centrifugation at 5000 rpm for 10 min at 4°C. The supernatant was removed and the pellet cells re-suspended in 12.5 mL of 100 mM $MgCl_2$ solution prechilled on ice. Thereafter, 2.5 mL of 100 mM $CaCl_2$ was added and then more 22.5 mL of the same solution was added. Next, cell suspension was then placed on ice for 30 min, and centrifuged for a further 10 min at 4000 rpm at 4°C. The supernatant was removed and the pellet ressuspended in 1 mL of 85 mM $CaCl_2$ with 15% glycerol. After that, aliquots of 100 µL were distributed for cryogenic tubes and stored at -80°C.

2.2.1.2 - Transformation

Previous *E.coli* DH5α competent cells (100 µL) were added to 0.5 and 1 µL of pPH600 (~ 100 ng). The solution was then placed on ice for 30 min. The samples were then heat shocked at 42°C for 1 min and placed on ice for 2 min. The cells were plated onto agar plates containing ampicillin (100 µg/mL) and grown overnight at 37°C overnight. The plates were then stored at 4°C.

2.2.1.3 - Plasmid extraction – NZYMiniprep

pDNA was isolated from cells according to the manufacturer's protocol. The protocol is based on a modified alkaline lysis procedure and is appropriate for small scale preparation of pDNA. Following lysis, the pDNA was pre-purified in NZYTech spin columns charged with a silica gel-based anion-exchange resin. First, 4 mL of harvest cells was ressuspended in A1 buffer (250 µL) and vortex was used to avoid any agglomerate of cells. Then, the homogeneous solutions was transferred to 1.5 mL eppendorfs. After that, A2 buffer (250 µL) was added, and the eppendorfs were carefully inverted 5-6 times. Then, the samples were incubated 4 min at room temperature. To neutralize lysis, A3 buffer (300 µL) was added, tubes were inverted 5-6 times. The lysate obtained was clarified by centrifugation at room temperature, 11000 RCF during 10 min. The supernatant was transferred to spin columns with a 2 mL tube collector and new centrifugation during 1 min in same conditions was performed to pass the solution through the column and bind pDNA to column. After that, AY buffer (500 µL) and A4 buffer (600 µL) were added to column to wash all impurities such, proteins, RNA, salts, nucleotides and small oligonucleotides. The elution of pDNA was obtained with A7 buffer which contains a high salt concentration. Finally samples were stored at -20°C.

2.2.1.4 - pPH600 digestion

Digests were done using 1 µg of pDNA. For digests, the restriction enzymes Hind III and PVU II were used. All digests were performed to a final volume of 20 µL in accordance to manufacturer's protocol. The digests were incubated at 37°C for 2 h and the digested pDNA was viewed on a 1% agarose gel electrophoresis.

2.2.1.5 - Master cell bank

A colony from plate with transformed *E.coli* DH5α was inoculated in flask with LB medium (10 mL) containing ampicillin (100 µg/mL). Growth was carried out at 37°C in 100 mL shake flask (250 rpm) until OD_{600} reach 0.7. Then, growth was suspended and glycerol in proportion (3:7 v/v) was added to cryopreserve cells. After that, aliquots of 1 mL were distributed for cryogenic tubes and stored at -80°C.

2.2.1.6 - Pre-inoculum and fermentation

The fermentation was initiated with inoculation of *E. coli* DH5α from master bank cell in LB-agar plates (35 g/L), which growth overnight at 37°C. Then, 125 mL of TB media (12 g/L tryptone, 24 g/L of yeast extract, 4 mL/L glycerol, 0.017 M KH_2PO_4 and 0.072 M K_2HPO_4) supplemented with 100 µg/mL of ampicillin was inoculated with one colony. The growth was carried out approximately during 5 h at 37°C. When OD_{600} reached 2.6, pre-fermentation was stopped and fermentation was initiated with an appropriate volume calculated through following equation:

$$OD_{pre-inoculum} \times V_{withdraw\ from\ pre-inoculum} = \left(V_{withdraw\ from\ pre-inoculum} \times V_{fermentation}\right) \times OD$$

Where:

OD = 0.2

Then, the 3562 bp pPH600 used in this work, was amplified by a cell culture of *E. coli* DH5α. Growth was carried out overnight at 37°C in 1 L shake flasks (250 rpm) using 250 mL of terrific broth medium (TB medium) (12 g/L of tryptone, 24 g/L of yeast extract, 4 ml/L glycerol, 0.017 M KH_2PO_4 and 0.072 M K_2HPO_4) supplemented with 100 µg/ml of ampicillin. Growth was suspended at late log phase (OD_{600nm} ≈ 8) After fermentation, cells were harvest by centrifugation, 5000 RCF (20 min, 4°C) and the pellet was stored at -20°C before further use.

2.2.2 - Cell lysis

2.2.2.1 - Plasmid extraction – Pre-purification with Maxiprep kit

First, 125 mL of harvest cells was ressuspended in 10 mL M1 Buffer and a homogeneous solution was obtained using vortex. After that, the extract was transferred to appropriate cell lysis tubes. Then, 10 mL M2 buffer was added to the solution, the tubes were carefully inverted 2-3 times and incubated 5 min at room temperature. After that, 10 mL M3 buffer was added, tubes were inverted 2-3 times and placed 20 min on ice. The obtained lysate was clarified by centrifugation at 4°C, 20000 RCF during 30 min. After that, one more centrifugation during 15 min was performed to remove completely cell debris, keeping the temperature and speed conditions. Following lysis, the pDNA was pre-purified in NZYTech columns charged with a silica-based anion-exchange resin. To proceed to pre-purification, the columns were equilibrated

with 10 mL QBT buffer. After centrifugation, the clarified lysate was added to column and washed with 2x30 mL QC buffer to remove impurities such as, proteins, RNA, salts, nucleotides and small oligonucleotides. Then, pure pDNA was eluted with 15 mL QF buffer, a high salt elution buffer, which neutralizes the positive charge of the resin by a pH shift to slightly alkaline conditions. Afterwards, the pDNA was concentrated by isopropanol precipitation. The final pellet was resuspended in 1 mL of 10 mM Tris–HCl (pH 8.0). The pDNA samples concentration was measured with a Nanophotometer™ P300 (Implen) and samples were stored at -20°C until use.

2.2.2.2 - Lysate extraction – Modified Alkaline Lysis

The *E.coli* lysate sample was prepared in accordance to modified alkaline lysis method described by Diogo and co-workers [95]. Initially, 125 mL of bacterial cells pellets recovered after fermentation were ressuspended in 10 mL of 50 mM glucose, 25 mM Tris-HCl, 10 mM EDTA, (pH 8.0) and vortex was used to obtain homogeneous solution. After that, lysis were performed adding 10 mL of 200 mM NaOH, 1% (w/v) SDS. After disruption of cells, cellular debris, proteins and gDNA were precipitated by adding 10 mL of 3 M potassium acetate. The precipitate was removed by centrifugation at 20000 RCF during 30 min at 4°C. The pDNA was precipitated with 0.7 volumes of isopropanol and incubated on ice for 30 min, followed by centrifugation at 16000 RCF at 4°C during 30 min. The pellets were then dissolved in 1 mL of 10 mM Tris-HCl (pH 8.0). Next, 2.5 M $(NH_4)_2SO_4$ was added, followed by 15 min of incubation on ice. Precipitated proteins and RNA were removed by centrifugation at 16000 RCF during 20 min at 4°C. Samples were stored at -20°C before further use.

2.2.3 - Bead morphology of activated Sepharose; L-tyrosine support and L-tryptophan support

Bead morphology was evaluated, before and after immobilization process with L-tryptophan and L-tyrosine, using scanning electron microscope (SEM) (Hitachi-S2700, Tokyo, Japan), operating at an accelarating voltage of kV, at variable magnifications. The samples were fixed on a brass stub using double-sided tape and then made electrically conductive by coating with gold using Emitech K550 sputter coater (London).

2.2.4 - Preparative chromatography

The initial *screening* was performed with pDNA isoforms of pPH600, pre-purified with Maxiprep kit, to evaluate behavior of L-arginine, L-tyrosine and L-tryptophan with plasmid. L-arginine support was commercially available and L-tyrosine column was previously prepared in our research group [106]. The L-tryptophan support was synthetized by coupling L-tryptophan to epoxy 1,4-butanediol diglycidyl ether activated Sepharose.

All chromatographic experiments were performed in Akta Pure 25 L with UNICORN™ 6.3 software (GE Healthcare, Uppsala, Sweden) at flow rate 1 mL/min. The absorbance was continuously measured at 260 nm. According to each experiment the temperature of column was correctly maintained at desirable temperature (4, 10, 15 and 25ºC), connecting a water-jacket column to circulate water bath. After purification, the fractions were pooled, according to the obtained chromatograms, concentrated and desalted with Vivaspin concentrators and analysed by 1% agarose gel electrophoresis.

2.2.4.1 - L-arginine-Sepharose 4B support

A standard 20 mL column was packed with L-arginine-Sepharose 4B gel. The gel matrix was a 4% agarose bead cross-linked with a 12 carbon atom spacer arm. The adsorption buffer used was 10 mM Tris-HCl (pH 8.0) and the elution solution was 1 M NaCl in 10 mM Tris-HCl (pH 8.0). The chromatographic system was prepared with 10 mM Tris-HCl (pH 8.0) in pump A (mobile phase A), and 1 M NaCl in 10 mM Tris-HCl (pH 8.0) in pump B (mobile phase B). Then, L-arginine-Sepharose 4B column was equilibrated with 110 mM NaCl in 10 mM Tris-HCl (pH 8.0) at flow rate 1 mL/min. After that, plasmid pPH600 were injected onto column by using a 100 µL loop at the same flow conditions. After injection the species that not have affinity to column elute and the species with strong interaction with column remain attached. The elution of remain species was achieved with increase of NaCl gradient from 110 mM to 500 mM in 10 mM Tris-HCl (pH 8.0).

2.2.4.2 - L-tryptophan-Sepharose CL-6B support

Sepharose CL-6B was activated according to the method previously described by Sundberg and Porath [107]. Briefly, 10 g of Sepharose CL-6B was washed with 250 mL of milli-Q water and suspended in 18 mL of 1 M NaOH solution. The solution was stirred at 25°C for 4 h. Then, 18 mL of 1,4-butanediol diglycidyl ether (bisoxirane) and aproximatelly, 40 mg of $NaBH_4$ was added slowly. The mixture was stirred for 16 h at 25°C. Sepharose CL-6B beads were washed with 250 mL of milli-Q water in a sintered glass funnel, followed by washing with different acetone-water solutions (20 mL) (1:9, 5:5, 9:1 v/v) and with 250 mL of milli-Q water. The gel was then suction-filtered to near dryness. The epoxy-activated sepharose CL-6B matrix (aproximatelly 3 g) was used to couple L-tryptophan. The L-tryptophan (14.7 mmol, 3 g) was dissolved in a

solution of 2 M sodium carbonate at pH 9. The mixture was stirred on the orbital at 55°C for 16 h. After that, the support was washed extensively with different solutions of acetone-milliQ water (1:9, 3:7, 5:5, 8:2 v/v) followed by washing with milli-Q water (3×200 mL).

For the chromatographic experiments, a standard 20 mL column was packed with a L-tryptophan-Sepharose CL-6B gel. The gel matrix was bead cross linked with 12 carbon atom spacer arm. The adsorption buffer used was $(NH_4)_2SO_4$ and the elution solution was 100 mM HEPES acid (pH 7.4). The chromatographic system was prepared with 100 mM HEPES acid (pH 7.4) in pump A (mobile phase A), and 3 M $(NH_4)_2SO_4$ in pump B (mobile phase B). Then, L-tryptophan-Sepharose CL-6B column was equilibrated with 2.65 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4) at flow rate 1 mL/min. Thereafter, pDNA (pPH600 and pVAX-*LacZ*) were injected onto column by using a 100 µL loop at the same flow conditions. After injection the species that not have affinity to column elute and the species with strong interaction with column remain attached. The elution of remain species was achieved with decrease of ionic strength. The gradient spetwise is decreased from 2.65 M to 0 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4).

2.2.4.3 - L-tyrosine-Sepharose CL-6B support

2.2.4.3.1 Pre-purification of sc pPH600 from a native sample (oc + sc)

A standard 20 mL column was packed with a L-tyrosine-Sepharose CL-6B gel. The gel matrix was bead cross linked with 12 carbon atom spacer arm. The adsorption buffer used was $(NH_4)_2SO_4$ and the elution solution was 100 mM HEPES acid (pH 7.4). The chromatographic system was prepared with 100 mM HEPES acid (pH 7.4) in pump A (mobile phase A), and 3 M $(NH_4)_2SO_4$ in pump B (mobile phase B). Then, L-tyrosine-Sepharose CL-6B column was equilibrated with 2.25 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4) at flow rate 1 mL/min. Thereafter, plasmid pPH600 were injected onto column by using a 100 µL loop at the same flow conditions. After injection the species that not have affinity to column elute and the species with strong interaction with column remain attached. The elution of remain species was achieved with decrease of ionic strength. The gradient stepwise is decreased from 2.25 M to 0 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4).

2.2.4.3.2 - Purification of clarified *E. coli* lysate

The column was previously packed with L-tyrosine-Sepharose CL-6B gel for initial *screening* of pPH600 isoforms. The adsorption buffer used was $(NH_4)_2SO_4$ and the elution solution was 100 mM HEPES acid (pH 7.4). The chromatographic system was prepared with 100 mM HEPES acid (pH 7.4) in pump A (mobile phase A), and 3 M $(NH_4)_2SO_4$ in pump B (mobile phase B). Then, L-tyrosine-Sepharose CL-6B column was equilibrated with 2.25 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4) at flow rate 1 mL/min. Thereafter, complex lysate was loaded onto the column using

a 200 µL loop at the same flow condition. The elution of bond species was carried out using a decreasing $(NH_4)_2SO_4$ stepwise gradient, first to 1.95 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4) and then to 0 M of $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4).

2.2.5 - Agarose gel electrophoresis

The pDNA isoforms was analysed by horizontal electrophoresis using 15 cm, 1% agarose gel (Hoefer, San Francisco, CA, USA) stained with GreenSafe Premium (0.01%; NZYTech) and visualized under UV light in a UVItec FireReader system (UVItec, Cambridge, UK). Electrophoresis was carried at 120 V, for 35 min, with TAE buffer (40 mM Tris base, 20 mM acetic acid and 1 mM EDTA, pH 8.0). Hyper Ladder I (Bioline, London, UK) was used as a DNA molecular weight marker.

2.2.6 - Analytical chromatography

The content of sc pPH600 present in clarified *E. coli* lysate and in purified fractions was determined with CIMac™ pDNA analytical column by a previously analytical method developed by Sousa and collaborators [80]. The pPH600 concentration in each sample was calculated by using a calibration curve made with pPH600 standards of 2.5 to 75 µg/mL (Fig. 13), purified with a commercial NZYTech kit. All samples were prepared by diluting the highest concentration with 200 mM Tris-HCl (pH 8.0). The analytic column was equilibrated with 600 mM NaCl in 200 mM Tris-HCl (pH 8.0) and after the injection of 20 µL of sample, a linear gradient of 10 min to 700 mM NaCl in 200 mM Tris-HCl (pH 8.0) was accomplished, at 1 mL/min.

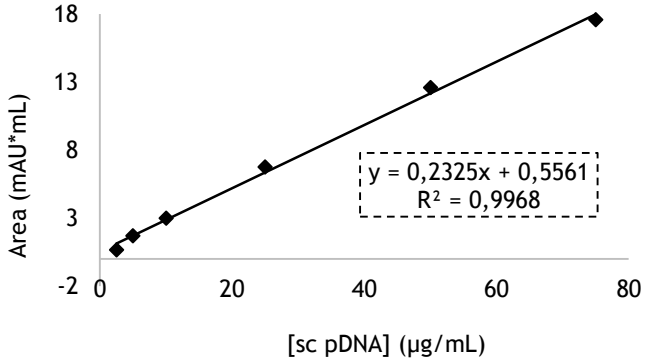


Figure 13 – Standard curve for sc pDNA quantification.

2.2.7 - Impurities assessment

The pPH600 quality control consists in the quantification of impurity levels, such as, proteins, endotoxins and gDNA, present in final sc pPH600 sample purified.

2.2.7.1 - gDNA quantification

Real-time polymerase chain reaction (PCR) was used to evaluate the existence of gDNA in the purified samples. The analyses were performed in an iQ5 Multicolor real-time PCR detection system (Bio-Rad), as previously described. Sense (5'-ACACGGTCCAGAACTCCTACG-3') and antisense (5'-CCGGTGCTTCTTCTGCGGGTAACGTCA-3') primers were used to amplify a 181-bp fragment of the 16S rRNA gene. PCR amplicons were quantified by following changes in fluorescence of the DNA binding dye SYBR Green I. The calibration curve to achieve the gDNA concentration was created by a serial dilutions of the *E. coli* DH5α gDNA sample (purified with the Wizard gDNA purification kit; Promega) in the range of 0.005 to 50 ng/µL (Fig. 14).

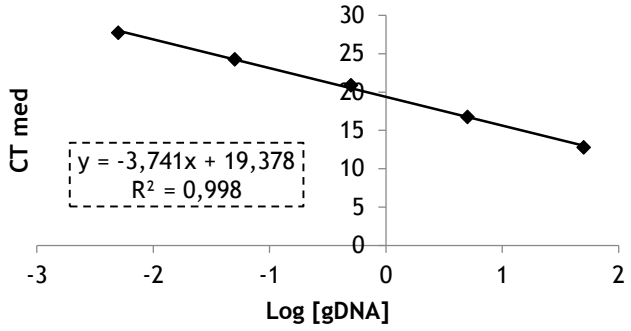


Figure 14 - Standard curve for the quantification of gDNA.

2.2.7.2 – Endotoxins evaluation

Endotoxin contamination was assessed by using the ToxiSensor™ Chromogenic *Limulus* Amoebocyte Lysate (LAL) Endotoxin Assay Kit (GenScript, USA, Inc.) according to the supplier's protocol. The calibration curve was constructed using a provided stock solution of 8 EU/mL (Fig. 15). To avoid external endotoxin interference, all the samples were diluted or dissolved with non-pyrogenic water, which was also used as blank. All the tubes and tips used to perform this quantification was endotoxin-free and the entire procedure was performed inside of a laminar flow cabine.

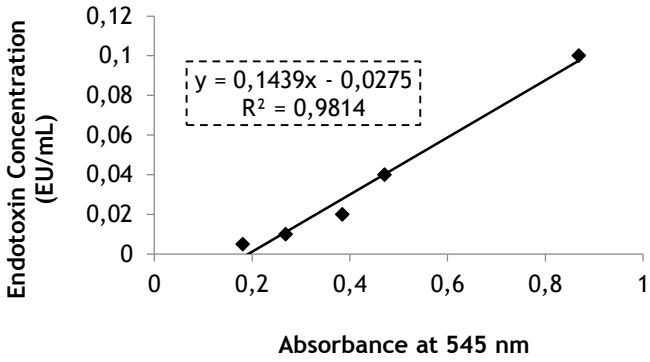


Figure 15 – Standard curve for the quantification of Endotoxin.

2.2.7.3 - Protein analysis

Protein quantification of pPH600 samples collected from the purification with L-tyrosine support and *E. coli* lysate, were measured using the micro-BCA (bicinchoninic acid) protein assay kit from Pierce in accordance with the specifications of the manufacturer's. A calibration curve was made with the standard protein bovine serum albumin (BSA) (0.01-0.150 mg/mL) diluted in 10 mM Tris-HCl (pH 8.0) (Fig. 16). A fraction of each sample (30 μL) was added to 200 μL of BCA reagent in a microplate and incubated for 30 min at 37ºC and then cooled to room temperature. The absorbance was measured at 570 nm in microplate reader.

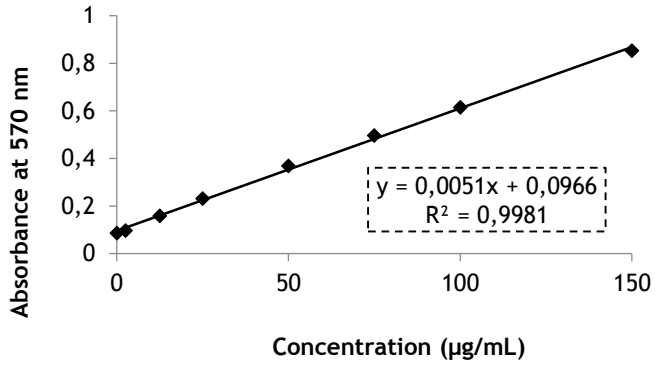


Figure 16 - Standard curve for the quantification of proteins.

2.2.8 – Plasmid transcription

Sc pPH600 transcription was enzymatically synthesized using MEGAscript® Kit, according to the manufacturer's recommendations. Briefly, the sc pPH600 containing a 604 bp PvuII-HindIII fragment of the murine Sγ3 switch region was linearized with HindIII enzyme and concentrated to a final concentration of 1 µg/µL. Then, the linearized plasmid was *in vitro* transcribed with the reagents provided in the kit for 1 h at 37 ºC. The reaction was terminated by adding 1/20th volume 0.5 M EDTA, 1/10th volume of 3 M sodium acetate and two volumes of ethanol, followed by incubation at –20 °C for at least 15 min. After centrifugation at 4 °C for 15 min at maximum speed, the supernatant was discarded. Then, the pellets were solubilized in water.

2.2.9 – Circular dichroism spectroscopy

CD spectra were acquired in Jasco J-1850 Spectrophotometer (Jasco, Easton, MD, USA) CD spectra using a quartz rectangular cell with an optical path length of 0.1 cm at temperature of 25 °C. The sc pPH600 and mRNA transcript samples with about 20 µg/mL were prepared in 10 mM Tris-HCl and 50 mM KCl, respectively. The spectral bandwidth was kept at 1 nm. The CD spectra were recorded from 200 to 320 nm at a scan speed of 10 nm/min. All measurements were conducted under a constant nitrogen gas flow, to purge the ozone generated by the light source of the instrument. The data were collected in triplicate and the average spectra are presented for each sample after subtracting the contribution of the buffer. The CD signal was converted to molar ellipticity. Noise in the data was smoothed using Jasco J-1850 software, including the Fast Fourier transform algorithm, which allows enhancement of most noisy spectra without distorting their peak shape.

Chapter 3

Results and discussion

In recent years, the interest for therapies that involve DNA arise with the developments of molecular biology techniques. The gene therapy and DNA vaccines are one of these and received the majority interest of researchers. DNA can assume different conformations and G4 are four stranded structures located in important areas of human genome. The DNA have different degrees of supercoiling that have influence in formation of secondary structures. Moreover, bacterial plasmids used as non-viral vectors for many therapies can assume different conformations and since supercoiling effect have influence in G4 formation, the purification of sc pPH600 is relevant for further analysis. Recently, a study performed by Duquette and collaborators show that plasmid pPH600 have more percentage of loops in sc isoform than in other isoforms (oc and ln) [14]. The pPH600 plasmid (3562 bp) is derivate from commercial plasmid pBluescript KS+ and encodes a sequence of immunoglobulin switch region Sγ3 of murine that can form G4 [14].

In this work, the first aim was the production of pPH600 which was essential to obtain quantities of sc pPH600 for further analysis in affinity chromatography. For this propose, competence was induced in *E. coli* DH5α cells for further transformation with pPH600 plasmid. Thereafter, the *E. coli* DH5α cells were produced and disrupted by modified alkaline lysis or pre purified with NZYTech silica columns to remove impurities such as, gDNA, RNA, proteins and endotoxins, resulting in a clarified *E.coli* lysate and a pPH600 native sample (oc + sc), respectively. Thereafter, a *screening* with pPH600 native sample (oc + sc) was performed to evaluate which of the three supports (L-arginine Sepharose; L-tryptophan Sepharose and L-tyrosine Sepharose) have better results in sc isolation. L-arginine Sepharose have been used in other works to purify different plasmids with interesting results because allow efficient isolation of sc pDNA and remove efficiently impurities [99]. However, there is a need to found another supports that are able to purify sc pDNA. L-tyrosine and L-tryptophan are hydrophobic and interact through hydrogen bonds, π-π interactions, vand der Waals and hydrophobic interactions, similar to L-histidine Sepharose previously used to purify sc pDNA from a pDNA native sample (oc + sc) [17] and from clarified *E. coli* lysate [93]. Then, the prominent support in sc pPH600 isolation was selected to perform chromatographic experiments with *E. coli* lysates. In this stage, the clarified *E. coli* lysate obtained from a modified alkaline lysis was injected onto selected support in order to elute single sc pPH600 from all other impurities (gDNA, RNA, proteins, endotoxins and oc pDNA).

Thereafter, the quality assessment of final product was evaluated in order to confirm several parameters such as, purity and recovery of sc pPH600.

Finally, the plasmid pPH600 was further subjected to *in vitro* transcription to induce the G4 formation in RNA transcript generated upon transcription of S region in pPH600. Then, circular dichroism was used to verify the presence and topology of G4.

3.1 - Characterization of plasmid pPH600

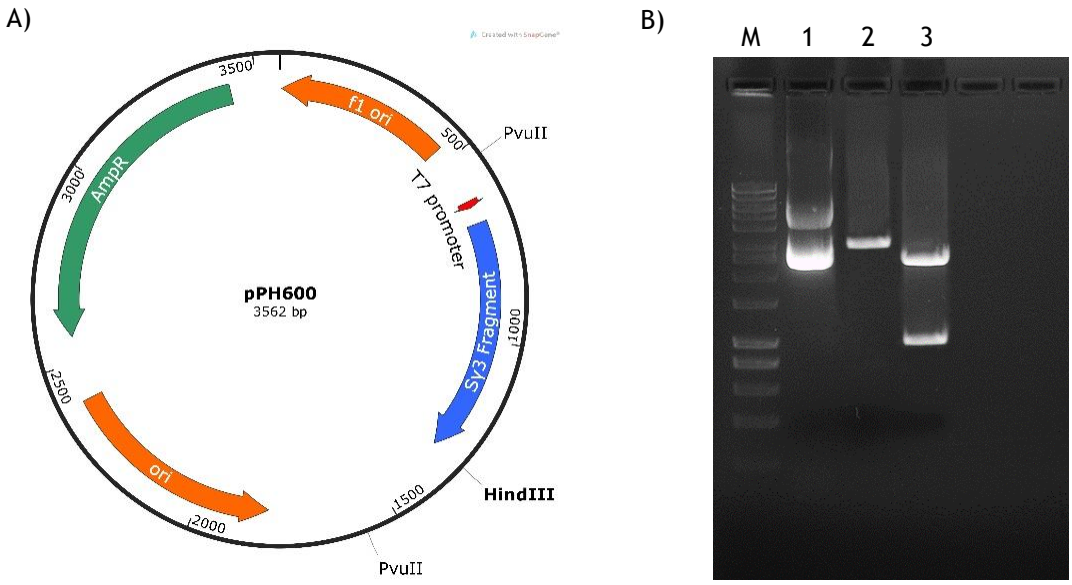


Figure 17 - A) Plasmid pPH600 map with ampicillin gene resistance, T7 promotor and Sγ3 fragment of murine. The restriction map was generated with SnapGene® Software and B) 1% agarose gel electrophoresis of plasmid pPH600 and restriction digests with HindIII and PvuII. M – Molecular weight marker; Lane 1 – Plasmid pPH600; Lane 2 – Plasmid digestion with HindIII; Lane 3 – Plasmid digestion with PvuII.

The plasmid pPH600 (3562 bp) used in this work is derived from pBluescript KS+ (2958 bp), which sequence is avalaible in GenBank and was consulted to perform the map of plasmid with Snapgene® Software. Plasmid backbone contains a gene that confers ampicillin resistance and a HindIII-PvuII 604 bp fragment of murine Sγ3 switch region that contains a *consensus* repeat d(CTGGGCAGCTCTGGGGGAGCTGGGGTAGGTTGGGAGTGTGGGGACCAGG) located just downstream of T7 promotor (Fig. 17A) [14]. In order to confirm if plasmid was in fact pPH600, resctrion digests were performed and analysed by 1% agarose gel electrophoresis. The restriction enzymes selected, HindIII and PvuII are the cutters of fragment, a single cutter and

a dual cutter, respectively. The enzymes were used single because the main goal is the confirmation of plasmid and not the extraction of fragment.

The number of fragments and the size of products obtained from restriction digests shown in the gel were the expected, confirming that plasmid is pPH600 (Fig. 17B). In lane 1 are observed two different bands, that corresponding to pPH600 isoforms oc and sc. The difference between the two bands is due to supercoiling state of pDNA. In sc isoform, pDNA is covalently closed circular and not rotate around phosphodiester backbone, decreseasing the hydrodynamic radius and causing less friction when runs through agarose gel. Unlike, oc isoform of pDNA increases the hydrodynamic radius because have a break in one strand of double helix, allow the movement around phosphodiester backbone and loss the sc state. Therefore, for the same pDNA, the isoforms runs diferently, sc isoform runs faster than oc isoform through agarose gel. In lane 2 the band observed correspond to ln isoform, that runs faster than oc isoform, because have a break in the two strands of double helix, but runs slower than sc isoform . In lane 3 the restriction enzymes cut the pDNA in two different locals, resulting in two fragments with different sizes. Therefore, the results of resctriction digests in 1% agarose gel eletrophoresis correspond to the restriction map in fig. 17A which shows the one restriction enzyme site for HindIII and two restriction enzyme sites for PvuII.

3.2 - pPH600 production

Plasmid pPH600 growth in *E.coli* DH5α, after competence of cells and transformation, because this strain have particular features that make it an ideal host for pDNA production. *E.coli* DH5α is probably the most used strain for plasmid production. This microorganism produces high yields and maintain the integrity of plasmid because have some mutations in the relA, endA, recA and GyrA to improve plasmid production [108]. The culture medium used was TB medium because is an enriched medium that helps *E. coli* strains maintain the growth rates. The medium have high quantities of tryptone and yeast extract which are required to enhance *E. coli* growth and to obtain high plasmid yields. Glycerol was added to provide an aditionall source of carbon. The potassium phosphates are present to prevent a drop in pH values. The culture strategy used in all of experiments was batch because the main advantage is its simplicity. In batch fermentation all of nutrients required to cell growth and plasmid production are present at the time of innoculation. The cell growth profile for bacterial cell density, obtained by the measuring of optical density at 600 nm is shown in fig. 18.

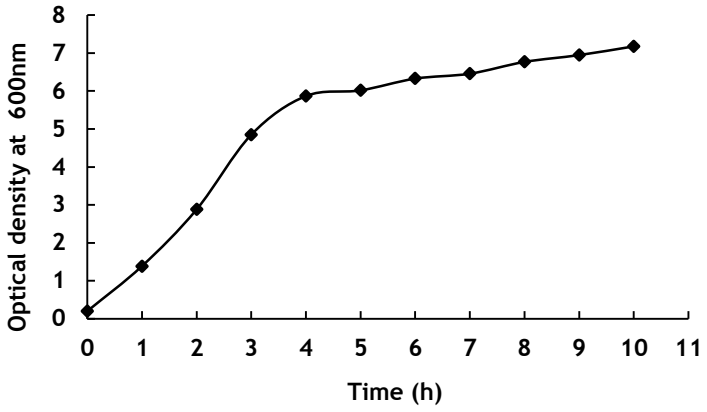


Figure 18 – Growth profile of *E. coli* DH5α harbouring plasmid pPH600 performed in 250 mL shake flasks at 37°C and 250 rpm in TB medium.

In fig. 18, OD_{600} reach approximately 7 and the fermentation was carried out for 10 h. Bacteria growth fast until 4 h, in which reach OD_{600} of approximately, 6. After that, the growth stabilize, and in further 6 h OD_{600} just only reach 7. This stabilization may be explained by metabolic burden that bacteria have been subjected during pDNA production [108]. Cooke and collaborators studied the impact of intrinsic DNA structures on processing of pDNA for gene therapy and DNA vaccination describe that cells containing pDNA with fragment of NHE, a sequence that are able to form G4, grows significantly faster than cells containing pDNA with other secondary structures, however they describe also that the reason for this increase in growth rate is unclear [109]. The results showed in fig. 18 represent growth curve of *E. coli* DH5α and is observed the same behavior that Cooke and collaborators describe.

3.3 - pPH600 extraction

The quality and quantity of pDNA sample obtained from lysis of *E.coli* DH5α is essential to perform reprodutible chromatographic experiments. For initial *screening*, the lysis of cells were performed with NZYMaxiprep kit that is designed for pre purification resulting in a pDNA native sample with oc and sc isoforms. First, cells were ressuspended in M1 buffer, the resultant solution should be homogeneous, whithout suspended particles.

Then, M2 buffer was added to disrupt cells through increased pH due to NaOH in presence of SDS, releasing pPH600, gDNA and celular proteins from cells to solution. In this stage occurs differential desnaturation of DNA, chromossomal DNA is desnatured, while pPH600 remains intact due to its small size. The next stage was the neutralization with potassium acetate, which leads to precipitation of SDS with gDNA, celular debris and proteins. All stages of

modified alkaline lysis were performed with extremely carefull to avoid break of chromossomal DNA. Thereafter, lysate was obtained and were pre purified in silica-based anion-exchange resin for NZYTech Maxiprep kit. After the extraction, the samples were evaluated by 1% agarose gel electrophoresis. The samples pre purified with NZYTech kit, not show impurities such as, gDNA and RNA. In fact, just show oc and sc pPH600 isoforms (Fig.19A).

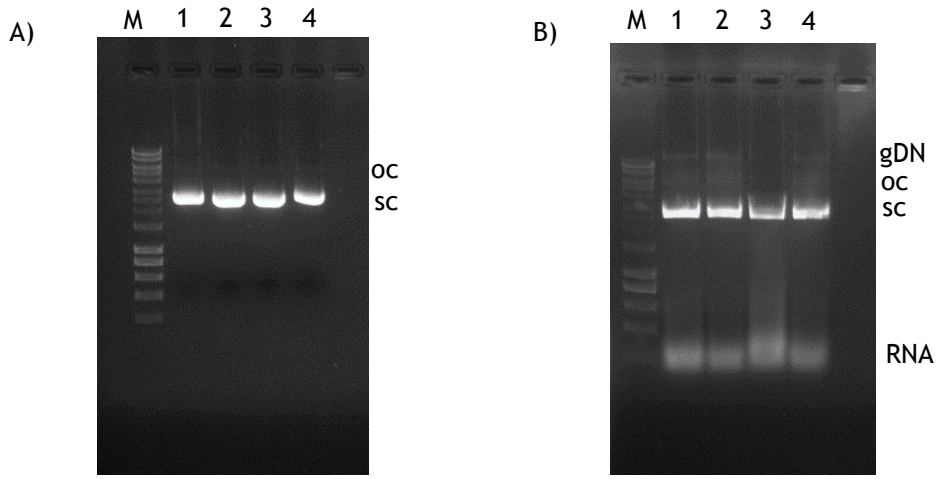


Figure 19 – pDNA Samples after extraction in 1% agarose gel electrophoresis. A) Pre purification with NZYTech Maxiprep kit; Lane M – Marker, Lane 1,2,3 e 4 – pPH600 native sample (oc + sc); B) Complex Lysate from *E. coli* DH5α; Lane M – Marker, Lane 1,2,3 e 4 – pPH600 isoforms, RNA and gDNA.

In laboratory scale there are many protocols and kits that are commercially avaliable to perform pre purification of plasmid samples, however, in large scale for therapeutic use, kits are not suitable due to presence of toxic products or difficulties in scale up process. To circuvent this drawback, clarified *E. coli* DH5α lysates were injected directly onto column. In lysate samples all of impurities (gDNA, RNA, proteins, endotoxins and oc pDNA) resultant from plasmid production in *E. coli* DH5α are present. The method to obtain lysate samples was also alkaline lysis; however, with some differences in composition of solution I and in pH of solution III. After alkaline lysis, the samples were analysed by 1% agarose gel electrophoresis and have present gDNA and RNA (Fig. 19B).

3.4. Characterization of L-tyrosine and L-tryptophan supports

Affinity chromatography is used to purify and separate biomolecules such as, proteins and nucleic acids and it is important characterize the support. In this work, the ligands L-tyrosine and L-tryptophan are immobilized in matrix Sepharose CL-6B. The efficiency of chromatographic processes is affected by bead morphology, and it is important the characterization of the matrix prior and after immobilization.

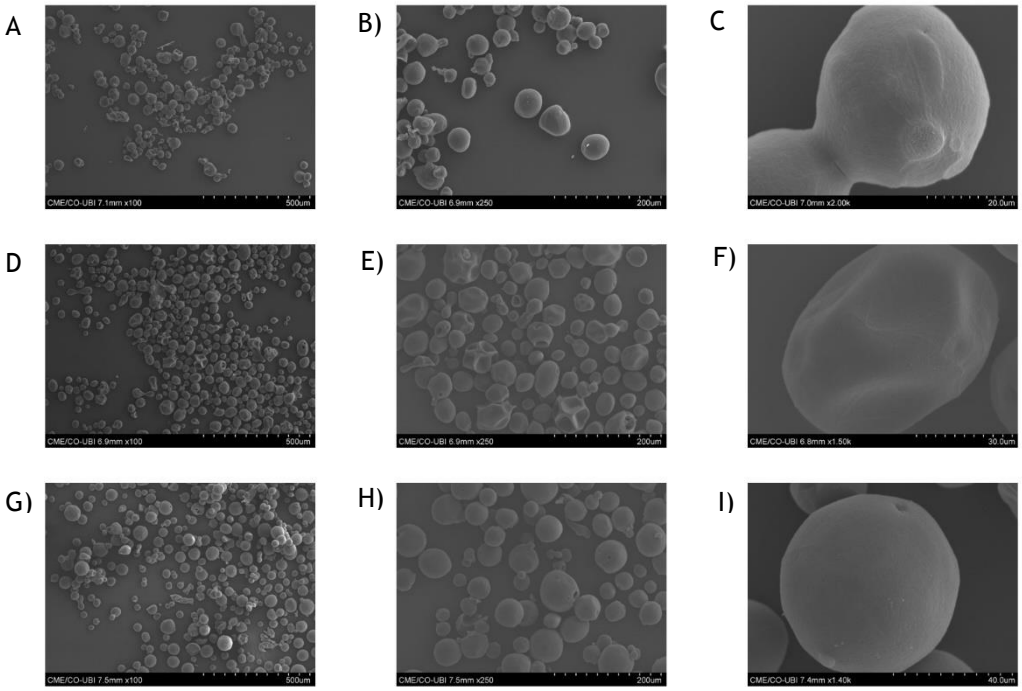


Figure 20 - SEM micrographs of Sepharose beads. (A-C) Epoxy activated Sepharose with no ligand at 100x, 250x and 2000x magnification. (D-E) L-tyrosine-Sepharose at 100x, 250x and 1500x magnification. (G-I) L-tryptophan-Sepharose at 100x, 250x and 1400x magnification.

This characterization is to verify the size, form and texture of Sepharose CL-6B beads and is done by SEM. The technique scans a focused electron beam over surface to create an image allowing to obtain information about surface topology of material. As shown in fig. 20, no significant differences could be observed before and after immobilization of Sepharose CL-6B with L-tyrosine and L-tryptophan suggesting that immobilization process did not cause visible alterations in bead morphology of support. This result guarantees the efficiency of chromatographic process concerning the flow properties is maintained, and the separation of pDNA is only influenced by the ligand.

3.5 - sc pPH600 purification with amino acid based chromatography

Interactions between proteins and nucleic acids play a central role in living systems by controlling the cellular expression. In many cases, proteins bind to highly specific nucleic acid base sequences [110]. This specific interaction takes advantage from biorecognition between two basic units, amino acids and nucleotides [89]. The strong interaction in biological context seems to be the result of multiple non-covalent interactions (hydrophobic, electrostatic, hydrogen bonds and π-π stacking interactions). In affinity chromatography using amino acids as ligands the aim is the separation of nucleic acids based on biological recognition by amino acids. L-arginine is the most used amino acid to purify nucleic acids. The most extend interactions that occur between arginine and nucleic acids are mostly electrostatic, however other multiple non-covalent interactions occur in less extension [99]. These characteristics, allowed the efficient separation of sc pDNA and pre-miR-29 [111, 112]. Also, L-histidine has been used to separate sc pDNA from a pDNA native sample (oc + sc) [17] and from clarified *E. coli* lysate [93]. L-tyrosine and L-tryptophan are both hydrophobic amino acids and interactions with DNA can be achieved, since strands are connected by hydrogen bonds between complementary nucleotides, results into the highly hydrophobic grooves [80]. Moreover, the charateristics of side chain of these amino acids, phenol in L-tyrosine and indole in L-tryptophan can be explored because aromatic rings have some unique and important properties such as, π-interactions and hydrogen bonds [113], which can be useful in separation of pPH600 isoforms.
In this work, was done a *screening* to evaluate the behavior of three amino acids affinity ligands after pre purification of pDNA isoforms with NZYTech Maxiprep kit. Thereafter the support in sc pDNA separation was selected and sc pPH600 was purified from clarified *E. coli* lysate.

3.5.1 – Pre-purification of sc pPH600 from native sample (oc+sc)

3.5.1.1 - sc pPH600 purification with L-arginine support

L-arginine was immobilized in gel matrix with 4% agarose bead cross-linked with a 12 carbon atom spacer arm (Fig. 21). L-arginine chromatography have been used to fully separate pDNA isoforms from a native sample (oc + sc) [99] and sc pDNA from all other impurities in clarified *E. coli* lysate samples [114]. The characteristics of L-arginine, namely its ability to interact in different conformations, the length of its side chain and its ability to produce hydrogen bonding geometries make it one of the most suitable amino acids for pDNA purification [111].

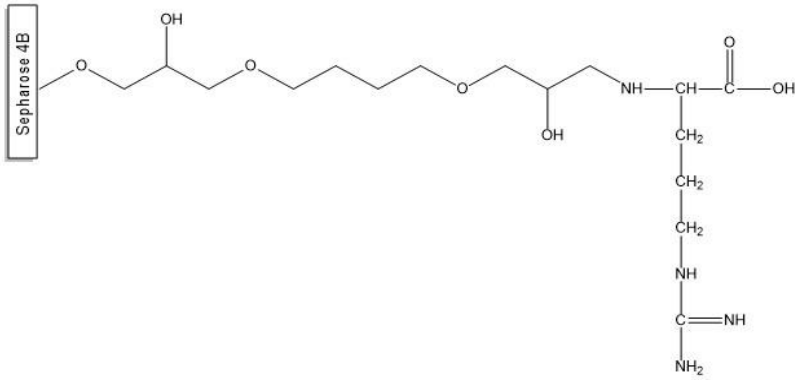


Figure 21 – Molecule structure of L-arginine coupled to Sepharose 4B through 12 carbon atom spacer arm. The molecular structure was generated by ChemBioDraw Ultra Software.

In order to obtain a selective and specific separation of sc pDNA, various parameters such as, type of salt, ionic strength and temperature should be considered. In previous work, the binding of plasmid pVAX-*LacZ* to L-arginine support was achieved using different salts such as, sodium citrate, sodium phosphate, $(NH_4)_2SO_4$ and NaCl; however, selective separation of sc pDNA was only obtained with $(NH_4)_2SO_4$ and NaCl [99]. NaCl is the most used salt in isolation of sc pDNA with L-arginine support because need low salt concentrations to elute pDNA and have lower impact in aquatic systems facilitating scale up of process. Temperature is another parameter that should be considered because have impact on sc pDNA isolation, since affects retention time of isoforms in L-arginine support and type of interactions involved between pDNA and L-arginine support [99].

In this work, L-arginine support was used to perform the isolation of sc pPH600. In order to achieve this separation, the type of salt choosen was NaCl and ionic strength was adjusted to obtain the better separation of pPH600 isoforms. The nature of interactions between pPH600 and L-arginine support was also evaluated.

Initially, a stepwise gradient was defined to evaluate the behavior of pPH600 with L-arginine support. The buffer used in this chromatographic experiments was 10 mM Tris-HCl (pH 8.0) because have a good capacity in maintaining pH at desired value and a low conductivity. This characteristic is particularly relevant, since L-arginine support is very effective in pDNA binding at low ionic strength [111]. Briefly, the chromatographic experiments was carried out by using a 100 µL loop to inject plasmid pPH600 native sample (oc + sc) onto column that was maintained at 4ºC. The stepwise gradient was adjusted to 10 mM Tris-HCl (pH 8.0) in first step and 500 mM NaCl in 10 mM Tris-HCl (pH 8.0), in second step. The chromatographic profile obtained when loading a pPH600 native sample (oc + sc) onto the L-arginine support is presented in fig. 22A.

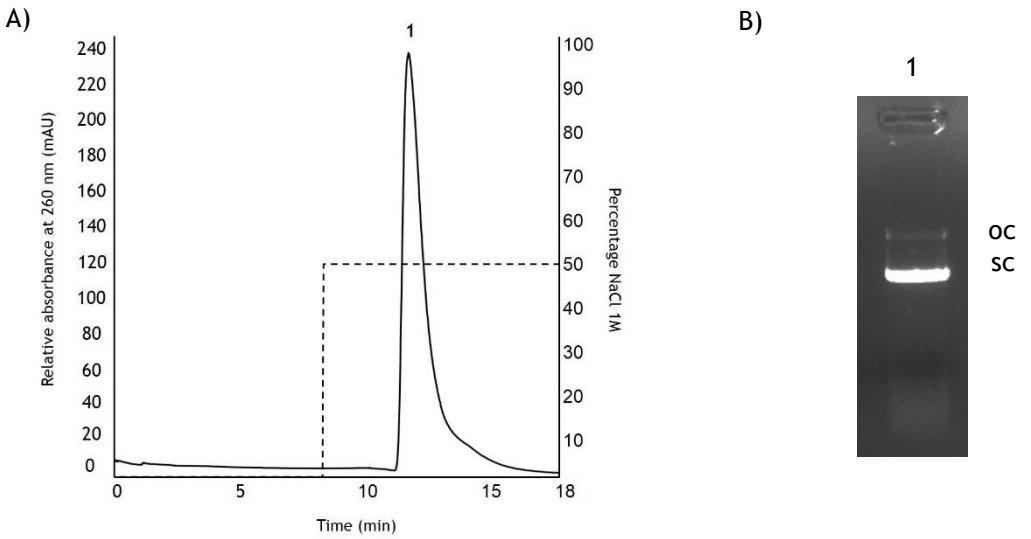


Figure 22 - A) Chromatographic profile of pPH600 isoforms from L-arginine support performing a stepwise gradient of 10 mM Tris-HCl (pH 8.0) and 500 mM NaCl in 10 mM Tris-HCl (pH 8.0). B) 1% agarose gel electrophoresis of pPH600 isoforms, Lane 1 – Peak 1 with oc and sc pPH600.

The respective chromatogram showed no peaks eluting with 10 mM Tris-HCl (pH 8.0) and a resolved peak eluting with 500 mM NaCl (peak 1) (Fig. 22A). To check the identity of the pDNA isoforms eluting in peak 1, the fraction recovered was analysed by agarose gel electrophoresis (Fig. 22B). The analysis revealed that the peak 1 corresponds to oc and sc pPH600 (Fig. 22B, lane 1).

The chromatographic profile and agarose gel electrophoresis suggests that pPH600 in presence of 10 mM Tris-HCl (pH 8.0) bond with high affinity to L-arginine support. The increment of ionic strength in mobile phase allows the elution of both pPH600 isoforms with 500 mM NaCl in 10 mM Tris-HCl (pH 8.0).

The results shows that the electrostatic interactions seems to be predominant between L-arginine support, that are positively charged due to guanidinium group and negative pPH600 [111].

An initial *screening* was performed to obtain the optimized binding/elution profiles for pPH600 with several concentrations of NaCl (100-500 mM) in 10 mM Tris-HCl (pH 8.0) (Table 3).

Table 3 – Summary of binding/elution profiles of plasmid pPH600 isoforms in different NaCl conditions.

	Binding/Elution profile in NaCl			
Plasmid isoform	**100 mM**	**120 mM**	**200 mM**	**500 mM**
oc	binding/elution	elution	elution	elution
sc	retention	binding/elution	binding/elution	elution

According to results in table 3, the behavior of pPH600 isoforms suggests that exists differences in biorecognition of pPH600 isoforms. This differences are the basis for sc pPH600 isolation and provides crucial information about interactions involved in biorecognition process.
Thus, despite electrostatic interactions promote the retention of two isoforms (Fig. 22), they are not specific, suggesting that specific interactions with sc pPH600 are related with different exposure degree of pPH600 bases in different isoforms due to supercoiling, as well other multiple non-covalent interactions (hydrophobic, hydrogen bonds and van der walls) (Fig. 23).

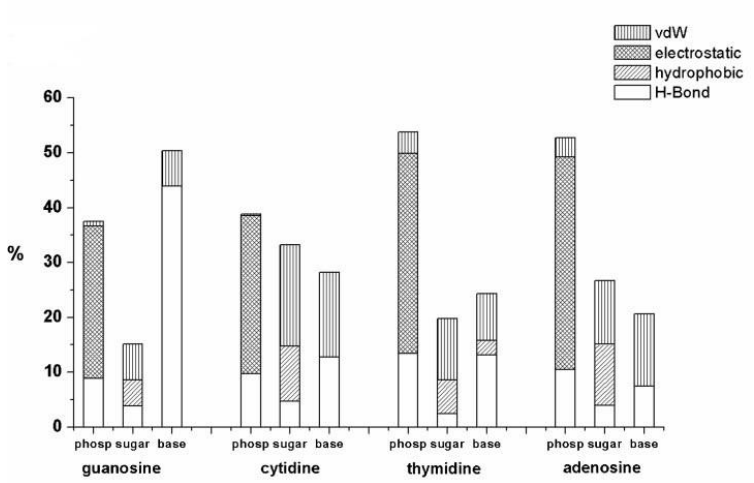


Figure 23 - Main interactions between L-arginine support and DNA bases (Adapted from [89])

In supercoiling, the winding of two anti-parallel strands, which are connected by hydrogen bonds between complementary nucleotides in each strand, result into highly hydrophobic grooves acessible to ligands and solvent [80]. This different exposure degree of DNA bases is probably the basis for specific recognition of pPH600 and L-arginine support. Moreover hydrogen bonds are not formed only between complementary nucleotides of double helix but also

between exposed DNA bases and L-arginine support contributing to affinity interactions between support and pPH600. DNA bases do not show same ability to form hydrogen bonds with L-arginine support and there are clear preferences for particular pairings of amino acids and DNA bases. L-arginine shows more affinity to guanine due to multiple hydrogen bonds that this DNA base can do (Fig. 23). Guanine is the only base in double strands that has two acceptor groups in the correct position to form hydrogen bonds with the guanidinium side chain of L-arginine [115].

The contacts between L-arginine and guanine are single or bidentade. In single contacts L-arginine interacts with either N^7 or O^6 atoms on guanine, but not both [89]. In bidentade contacts L-arginine interacts with both N^7 and O^6 atoms on guanine, these contacts occur more than single contacts [89].

The specific biorecognition of sc pPH600 by L-arginine support was probably due to hydrogen bonds that are formed between L-arginine and guanines in pPH600.

Since plasmids have different composition in terms of bases, the affinity for some bases is important. The full sequence of pPH600 it unknown but, have a fragment of 604 bp that encodes to G4, which have a high content of G-C (50,2% according with GenBank database), these characteristics make pPH600 suitable to be purified with L-arginine support.

The results obtained in table 3 also provide information about the range in which the isolation of sc pPH600 probably occurs. Analysing the table 3 it is possible conclude that the purification of sc pPH600 may be achieved in a small range of concentration between 100-120 mM NaCl.

Therefore, several experiments were performed in the same conditions of initial *screening* experiments; however with different salt conditions. The sc pPH600 retention was observed when salt concentrations were kept below 110 mM NaCl in 10 mM Tris-HCl (pH 8.0), being the sc pPH600 more strongly retained on the L-arginine support than oc that elutes immediately. The total elution of sc was observed at 500 mM NaCl in 10 mM Tris-HCl (pH 8.0). The chromatographic profile obtained when loading a pPH600 native sample (oc + sc) onto the L-arginine Sepharose is presented in Fig. 24A.

The respective chromatogram showed two resolved peaks eluting at 110 mM NaCl (peak 1) and 500 mM NaCl (peak 2) (Fig. 24A). To check the identity of the plasmid isoforms eluting within each peak, the two fractions recovered were analysed by agarose gel electrophoresis (Fig. 24B). The analysis revealed that the first peak corresponds to oc pPH600 (Fig. 24B, lane 1), whereas the second peak corresponds to the sc pPH600 (Fig. 24B, lane 2).

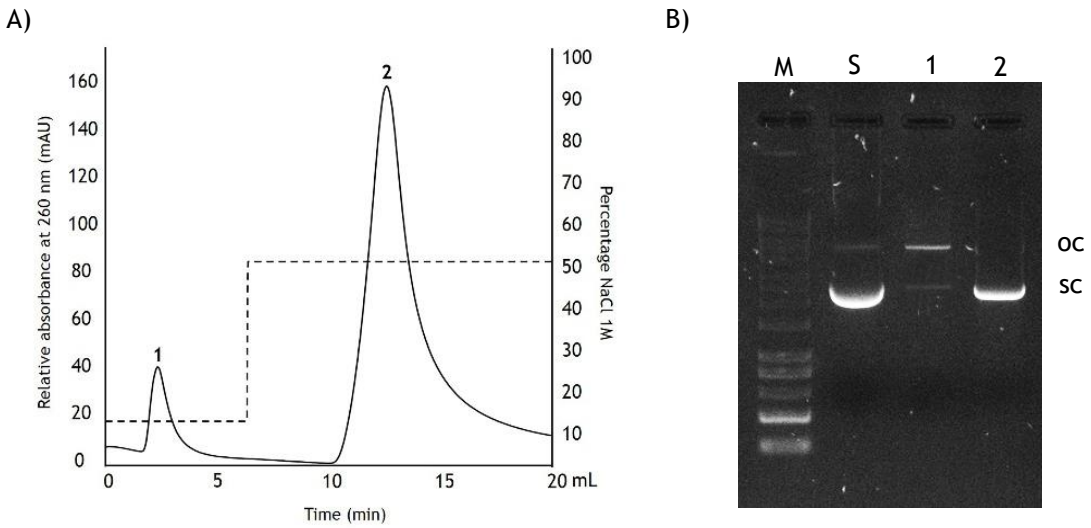


Figure 24 - A) Chromatographic profile of pPH600 isoforms from L-arginine support performing a stepwise gradient of 110 mM NaCl and 500 mM NaCl in 10 mM Tris-HCl (pH 8.0). B) 1% agarose gel electrophoresis of pPH600 isoforms, M – Marker; S – pPH600 native sample (oc + sc); Lane 1 – Peak 1 with oc and small quantity of sc isoform and Lane 2 – Peak 2 with sc isoform.

The temperature effect (10, 15 and 25ºC) in the retention of pPH600 on L-arginine support was also evaluated. The elution profiles obtained showed that when the temperature increases, the selectivity and specificity were affected, since retention increases for all the isoforms. In fact, this behavior is in accordance with predominant interactions occurring between L-arginine support and pPH600. As described by Sousa and collaborators the predominant interactions are electrostatic and occur mainly with phosphate groups of pDNA being affected by temperature [99].

It was also observed that the insert of 604 bp in pPH600, seems to be not influence in process of sc pPH600 purification. This conclusion is based on previous works in which, the separation of pVAX-*LacZ* isoforms was also obtained at low salt concentrations and the process shows selectivity and specificity [99].

In summary, the strategy of only two steps, the high resolution obtained and the low salt concentration required for the sc pDNA purification prove that this support is appropriate and economic to apply in the biopharmaceutical process.

3.5.1.2 - sc pPH600 purification with L-tryptophan support

The L-tryptophan Sepharose was prepared through covalent immobilisation of 1,4-butanediol diglycidyl ether directly onto Sepharose CL-6B (Fig. 25).
L-tryptophan is a hydrophobic and aromatic amino acid with indole ring in side chain. The indole ring makes L-tryptophan the bulkiest amino acid and allows amphipathic interactions (imino group hydrogen bonding, specific dipolar interactions, cation-π interactions, and hydrophobic interactions) [101]. Moreover, the special character of the indole ring provides the potential to stacking with other rigid ring structures that are found in nucleic acids.

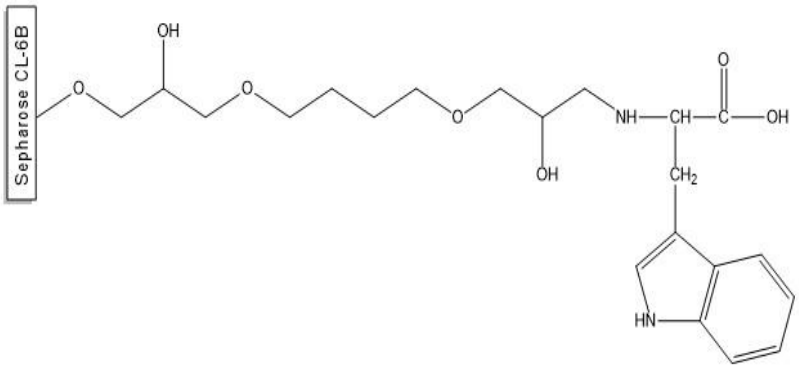


Figure 25 - Molecule structure of L-tryptophan coupled to Sepharose CL-6B through 12 carbon atom spacer arm. The molecular structure was generated by ChemBioDraw Ultra Software.

The purification of plasmids pPH600 and pVAX-*LacZ* with L-tryptophan Sepharose was novelty and many parameters needed to be taking into consideration. The buffer chosen was 100 mM HEPES acid (pH 7.4) because have strong buffer capacity with minimum contribution to conductivity. Moreover, a previous work developed by our group, in which SPR technique was used to determine affinity constants shows that the interactions between L-tryptophan support and pVAX-*LacZ* are more favourable with 100 mM HEPES acid (pH 7.4) than with 10 mM Tris-HCl (pH 8.0) [105]. The L-tryptophan was the most hydrophobic amino acid among 20 essential amino acids found in proteins, and hydrophobic interactions are explored with the use of $(NH_4)_2SO_4$ in mobile phase to bind sc and elute oc isoform [105].
Hence, the binding of the pPH600 native sample (oc + sc) was tested at different high salt concentrations in a range between 2.5 to 3 M $(NH_4)_2SO_4$ and the elution without salt in 100 mM of HEPES acid (pH 7.4), while temperature was maintained at 10ºC (Table 4).

Table 4 - Summary of binding/elution profiles of plasmid pPH600 isoforms in different $(NH_4)_2SO_4$ conditions.

	Binding/Elution Profile in $(NH_4)_2SO_4$			
Plasmid isoform	**0 M**	**2.5 M**	**2.7 M**	**3 M**
oc	elution	elution	binding/elution	binding
sc	elution	binding/elution	binding/elution	binding

The behavior of pPH600 isoforms with L-tryptophan support suggests differences in binding/elution patterns between isoforms.

In this case, the interactions of phosphate and sugar groups with L-tryptophan support can be ruled out because these groups are equaly exposed in both isoforms not allowing a selective distinction between isoforms [93]. Thus, the observed selectivity must be due to interactions between DNA bases and L-tryptophan support.

In table 4 are present the summary results of the range in which the isolation of sc pPH600 probably occurs. Analysing the results it is possible conclude that the purification of sc pPH600 may be achieved in a range between 2.5 M and 2.7 M $(NH_4)_2SO_4$.

Therefore, various experiments were performed to find the optimal concentration for separation of pPH600 isoforms. Thereafter, the column was equilibrated with 2.65 M $(NH_4)_2SO_4$ in 100 mM of HEPES acid (pH 7.4) to bind the sc pDNA isoform. After pPH600 native sample (oc + sc) injection, the oc isoform was immediately eluted and the sc remained bond to the support. Fig. 26A shows a chromatographic profile obtained from the injection of pPH600 native sample (oc + sc).

An agarose gel electrophoresis analysis of the fractions eluting from the column (Fig. 26B) proved that the first peak of unbond material corresponds to the oc isoform (Fig. 26B, lane 1), whereas the second peak was attributed to the sc isoform (Fig. 26B, lane 2). The elution of sc isoform was obtained at low salt concentrations, 0 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4), when supercoiling state returns to its original state, more compact than oc isoform but with less exposition of hydrophobic bases.

However, the analysis revealed that the separation of sc from oc was not complete, eluting a small quantity of sc isoform at 2.65 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4) and a small quantity of oc isoform at 0 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4) with temperature at 10ºC.

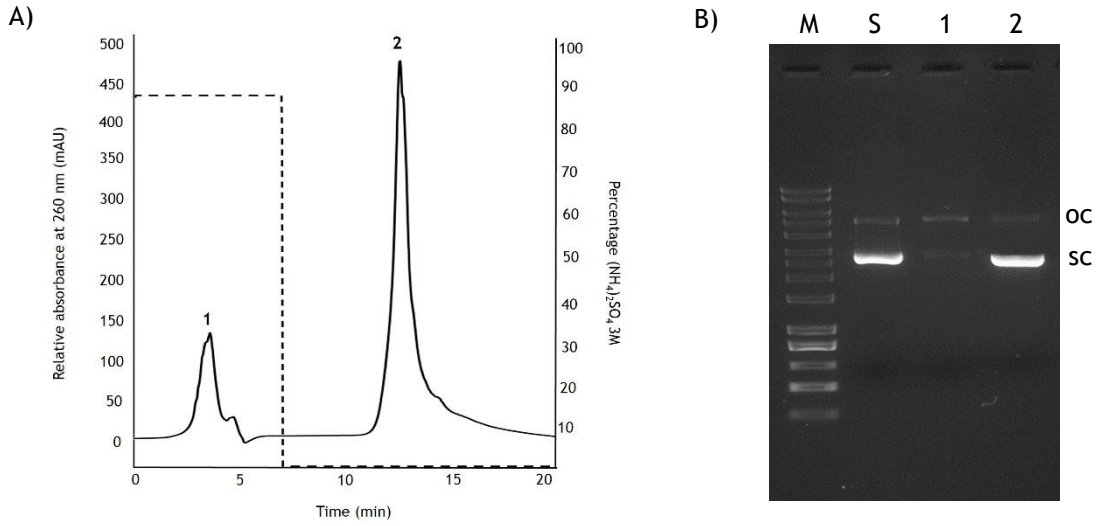


Figure 26 – A) Chromatographic profile of pPH600 isoforms from L-tryptophan support performing a stepwise gradient of 2.65 M $(NH_4)_2SO_4$ and 0 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4). B) 1% agarose gel electrophoresis of pPH600 isoforms, M – Marker; S – pPH600 native sample (oc + sc); Lane 1 – Peak 1 with oc and sc isoform and Lane 2 – Peak 2 with sc isoform.

This behavior is consistent with the fact that at high salt concentrations pDNA is tightly interwound. This is due to cations NH_4^+ that reduce electrostatic repulsion between phosphate groups in the DNA backbone. In fact, the supercoiling increases with salt concentration exposing the hydrophobic bases for the interaction with L-tryptophan support. Thus, the bases exposure occurs in the sc isoform in a higher extent than in oc isoform. Based on previous data, L-tryptophan, usually shows a preference for cytidine [89].

The aromatic group of L-tryptophan have some unique and additional capabilities that resulting not only in hydrophobic interactions but also in other multiple non-covalent interactions such as, hydrogen bonding and π-interactions that may play a fundamental role in molecular recognition of nucleic acids (Fig. 27) [113].

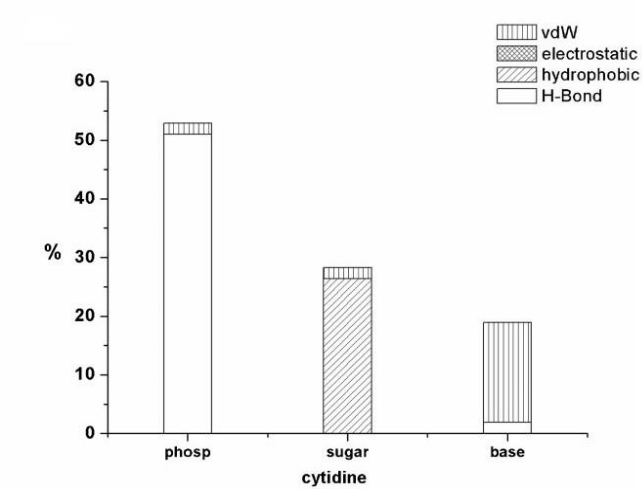


Figure 27 - Main interactions between L-tryptophan and cytidine (Adapted from [89]).

The L-tryptophan support, as well L-histidine, shows tendency to promote retention of sc pDNA in high salt concentrations. However, to elute sc pDNA it was necessary to decrease salt concentration, suggesting that ring stacking/hydrophobic interactions are the dominating effect, since are stronger at high salt concentrations [93].

In order to optimize the recovery of sc pDNA, the temperature was changed to 15ºC to increase the hydrophobic interactions. However, the separation was not possible since the retention time of sc isoform decreased, affecting selectivity and specificity. Moreover, this behavior of the L-tryptophan support is a strong evidence that the established interactions are not only hydrophobic, but also hydrogen bonds, van der Waals and π-π stacking as described by Sousa and collaborators [89]. Several experiments were also performed at 4ºC, however in this case the sc pPH600 retention time increased, not allowing the efficient separation of pPH600 isoforms due to lack of selectivity and specificity.

The plasmid pVAX-*LacZ* was also purified with L-tryptophan support because is a commercial plasmid used in gene therapy and DNA vaccination and also can be used as a model [105]. The behavior of pVAX-*LacZ* in L-tryptophan support was very similar to pPH600 purification. The ionic strength necessary to elute oc isoform and retain in column sc isoform was 2.7 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4), similar to salt concentration used to separate sc pPH600. As in sc pPH600 isolation the total elution of sc pVAX-*LacZ* was achieved when ionic strength was decreased to 0 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4). The temperature of experiments were also mantained at 10ºC. The chromatographic profile of sc pVAX-*LacZ* isolation is depicted in Fig. 28A.

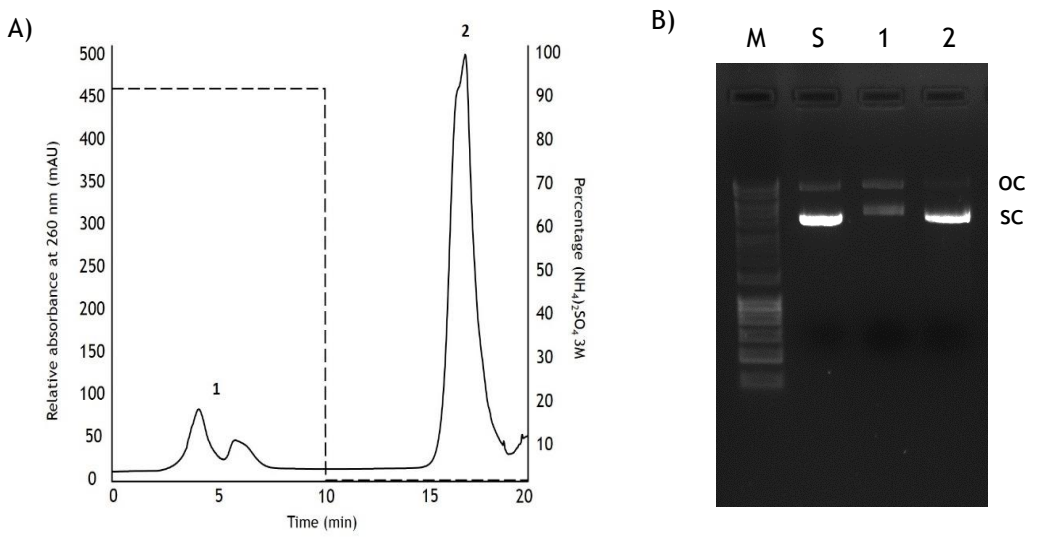


Figure 28 - A) Chromatographic profile of pVAX-*LacZ* isoforms from L-tryptophan support performing a stepwise gradient of 2.7 M $(NH_4)_2SO_4$ and 0 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4). B) 1% agarose gel electrophoresis of pVAX-*LacZ* isoforms, M – Marker; S – pVAX-*LacZ* native sample (oc + sc); Lane 1 – Peak 1 with oc and sc isoform and Lane 2 – Peak 2 with sc isoform.

An agarose gel electrophoresis analysis of the fractions eluting from the column (Fig. 28B) proved that the first peak of unbond material corresponds to the oc isoform (Fig. 28B, lane 1), whereas the second peak was attributed to the sc isoform (Fig. 28B, lane 2). However, as in sc purification of pPH600 the analysis revealed that the separation of sc from oc was not complete, eluting a small quantity of sc isoform at 2.7 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4) with temperature at 10ºC. According to the results obtained the behavior of pPH600 and pVAX-*LacZ* was very similar; however some differences were observed in ionic strength necessary to separate oc from sc isoform.

3.5.1.3 - sc pPH600 purification with L-tyrosine support

The support L-tyrosine Sepharose, was prepared by covalent immobilisation of 1,4-butanediol diglycidyl ether directly onto Sepharose CL-6B (Fig. 29). This amino acid is aromatic like L-tryptophan, however is less hydrophobic.

The aromatic ring in L-tyrosine is the phenol that is derived from benzene, which is present in another aromatic amino acid, the phenilalanine. The difference between the two amino acids is the hydroxyl group that can act in a hydrogen bond as donor.

The phenol ring as particular characteristics and like indole can interact with nucleic acids through hydrophobic interactions, hydrogen bonds, π-π interactions and other multiple non-covalent interactions .

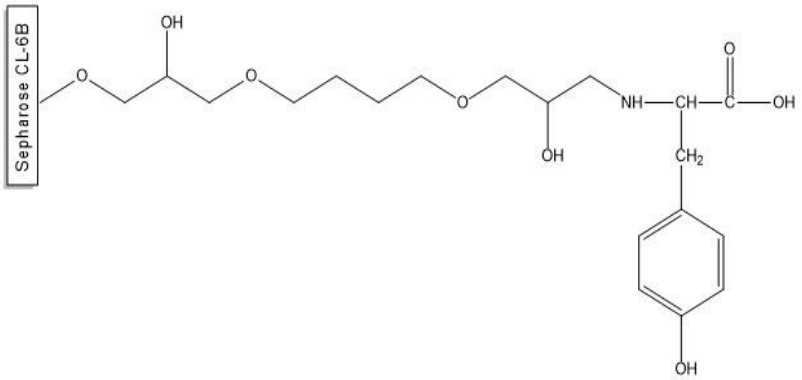


Figure 29 - Molecule structure of L-tyrosine coupled to Sepharose CL-6B through 12 carbon atom spacer arm. The molecular structure was generated by ChemBioDraw Ultra Software.

L-Tyrosine, unlike L-tryptophan has been already used in affinity chromatography to purify nucleic acids, although with some modifications in molecular structure. In work of Afonso and collaborators an affinity chromatography strategy using o-phospho-l-tyrosine was used to purify pre-miR29 with significant yields in purity and recovery [100]. However, for pDNA purification the use of L-tyrosine was quite novelty and various parameters should be optimized.

A recent work of our group was showed by SPR that 100 mM HEPES acid (pH 7.4) is the better buffer to isolate sc pDNA because have the minimum contribution to affinity interactions between L-tyrosine support and pDNA [106].

A first screen was also performed to select the better salt conditions for binding/elution, using concentrations of $(NH_4)_2SO_4$ between 2 and 2.5 M; however another experiment was done without salt in 100 mM HEPES acid (pH 7.4) in order to verify the elution of both isoforms (Table 5).

Table 5 - Summary of binding/elution profiles of plasmid pPH600 isoforms in different $(NH_4)_2SO_4$ conditions.

	Binding/Elution Profile in $(NH_4)_2SO_4$			
Plasmid isoform	**0 M**	**2 M**	**2.2 M**	**2.5 M**
oc	elution	elution	elution	binding
sc	elution	elution	binding/elution	binding

The behavior of sc pPH600 purification with L-tyrosine support was similar to observed with L-tryptophan; however the ionic strength used was lower. The table 5 shows that is possible purify sc pPH600 with L-tyrosine support because isoforms have different patterns of binding/elution, despite a decrease in salt concentration. However, the ionic strength to perform isolation of sc pPH600 remain high and this suggests that hydrophobic interactions as well, other multiple non-covalent interactions are presented (Fig. 30).

As previously mentioned, the supercoiling increases with salt concentration exposing the hydrophobic bases for the interaction with L-tyrosine support. Thus, the bases exposure occurs in the sc isoform in a higher extent than in oc isoform.

The L-tyrosine have particular affinity for guanine and the presence of hydroxyl group in aromatic ring allow the hydrogen bonds with pPH600 .

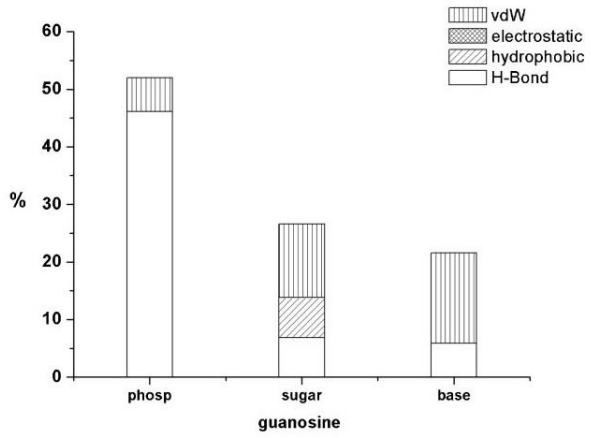


Figure 30 - Main interactions between L-tyrosine and guanine (Adapted from [89]).

The results obtained in table 5 also provide important information about the range in which the isolation of sc pPH600 probably occurs. Analysing the table 5 it is possible conclude that the purification of sc pPH600 may be achieved closer 2.2 M $(NH_4)_2SO_4$.

Thus, various experiments were performed in order to achieve the optimal concentration in which is possible isolate sc from oc pPH600. A total separation of sc was achieved only with salt concentration 2.25 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4) with temperature at 10ºC. The elution of sc isoform was obtained at low salt concentration, 0 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4), when the effect of salt is negligible as previously mentioned. Fig. 31A shows a chromatographic profile obtained after injection of pPH600 containing both isoforms (sc and oc) onto the L-tyrosine support.

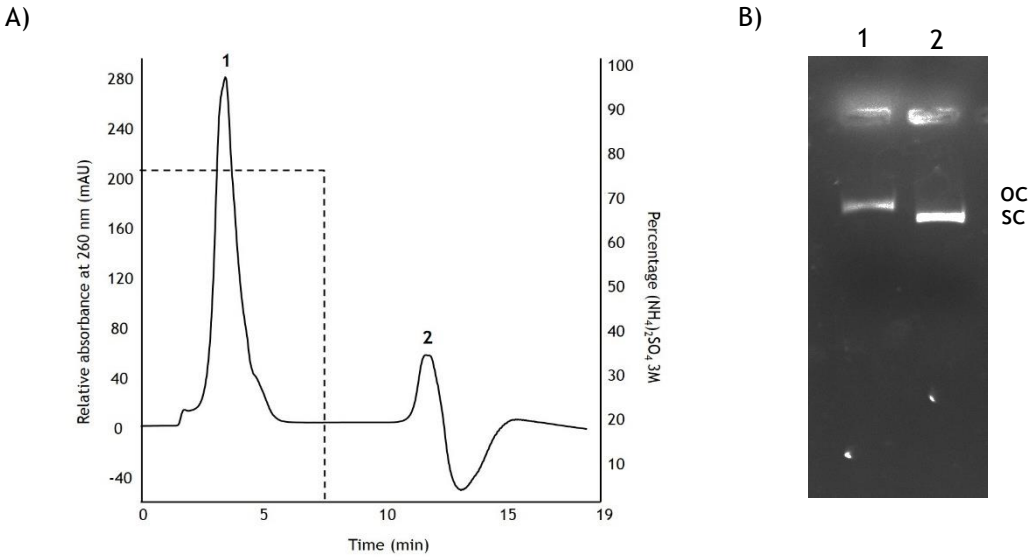


Figure 31 - A) Chromatographic profile of pPH600 isoforms from L-tyrosine support performing a stepwise gradient of 2.25 M $(NH_4)_2SO_4$ and 0 M $(NH_4)_2SO_4$ in 100 mM HEPES acid pH (7.4). B) 1% agarose gel electrophoresis of pPH600 isoforms, Lane 1 – Peak 1 with oc and sc isoform and Lane 2 – Peak 2 with sc isoform.

An agarose gel electrophoresis analysis of the fractions eluting from the column (Fig. 31B) proved that the first peak corresponds to the oc isoform (Fig. 31B, lane 1), whereas the second peak was attributed to the sc isoform (Fig. 31B, lane 2). In this case, as depicted by agarose gel electrophoresis (Fig. 31B) the total separation of pPH600 isoforms was achieved. The temperature effect in isoform separation of pPH600 on L-tyrosine was also evaluated. The elution profiles showed that when the temperature was 15 and 25ºC, the selectivity and specificity were affected, since the sc isoform retention decreases. Thus, the separation of two isoforms was only achieved at 10ºC. Several experiments were also performed at 4ºC; however in this case the sc pPH600 retention time increased, not allowing the efficient separation of pPH600 isoforms due to lack of selectivity and specificity.

Comparing the three chromatographic strategies used to purify the sc pPH600, it is evident that in the cases of L-tryptophan and L-tyrosine the retention of the sc pPH600 in the column increase with the salt, and can only be eluted by decreasing the salt concentration. The magnitude of the ionic strength necessary to promote these interactions depends on the hydrophobic character of the molecules. Those studies involving L-tryptophan and L-tyrosine supports have proved that by using an $(NH_4)_2SO_4$ gradient, oc pPH600 elutes in the flowthrough, whereas sc pPH600 interact with the columns.

However, pPH600 isoforms separation is more effective with L-tyrosine support than with L-tryptophan support, suggesting that the less hydrophobic nature of L-tyrosine compared with

L-tryptophan may interfere in the separation of isoforms. In fact, the interactions occurring between L-tyrosine and sc pDNA are multiple non-covalent, and are responsible for the selectivity achieved. This fact may be related with hydroxyl group of L-tyrosine that are able to form hydrogen bonds with guanine, π-π stacking between aromatic ring and DNA bases and van der Waals interactions [89].

For L-arginine support, the pPH600 isoforms separation was achieved with low ionic strength, indicating the presence of strong interactions with the charged phosphate groups in the pPH600 backbone. In this case, to separate sc from oc a slight increase of NaCl concentration was necessary, confirming the predominance of electrostatic interactions and a specific recognition for the sc isoform.

By comparing the results obtained with the three supports, it is noticeable that L-tyrosine support showed the appropriate selectivity for sc pPH600 isolation using a simple strategy with only two steps.

The behavior of three supports with temperature was also evaluated and in the case of L-arginine support, an increase in temperature affects significantly the efficiency of process and even with an optimized gradient the purification of sc isoform was not completely achieved because the recovery of sc isoform also induced the co-elution of oc. This result may be attributed to other interactions, namely hydrophobic interactions between aliphatic side chain or 12-carbon spacer arm and pDNA that arise when temperature increase, even at low salt conditions (< 300 mM NaCl) as described by Sousa and collaborators [99]. On the contrary, L-tryptophan and L-tyrosine supports are more hydrophobic and consequently the separation of the isoforms was not achieved at low temperature (4ºC). In this case when temperature was maintained at 4ºC the process was not efficient because sc pDNA fraction had always some contamination with oc isoform using L-tryptophan support.

The behavior of the L-tyrosine support is similar to the L-histidine support, since shows tendency to promote elution of oc pPH600 and retention of sc pPH600 with high salt concentrations [93]. However, to promote the pDNA elution it was necessary to decrease the salt concentration, suggesting that π-π stacking and is the dominant interaction, since hydrogen bonds do not weaken with lower salt concentrations [93].

Only when the temperature was increased to 10ºC, the total separation of sc pPH600 was achieved with L-tyrosine support. Also, with temperature at 10ºC and using L-tryptophan support, the fraction of sc is contaminated with oc, suggesting that using this strategy is not possible to isolate the pPH600 isoforms. Comparing with L-histidine support, the sc isoform retention decreased as the temperature increased [93]. This behavior suggests that the increasing of temperature may affect the sc isoform conformation inducing sc pDNA torsional strain changes, leading to relaxation and to a consequent decrease in the interaction of this molecule with the L-histidine support [93]. Moreover, pPH600 has a sequence that is able to form G4 and it was reported that an increase in temperature can disrupt the interactions that stabilize these structures affecting the retention [116].

3.5.2 – Purification of sc pPH600 from clarified *E. coli* lysate

Although the L-arginine support is able to separate sc from oc pPH600 under mild conditions, the total recovery of sc pPH600 is not possible, eluting sc pPH600 in first peak together with oc pPH600. In this context, L-tyrosine support showed the promising result in separation of two isoforms, allowing the total recovery of sc pPH600. Based on these considerations, L-tyrosine support was chosen to purify sc pPH600 directly from clarified *E. coli* lysate. The influence of impurities present in the clarified lysate (different nucleic acid species such as native oc pDNA, total *E. coli* RNA, gDNA, and proteins and endotoxins) was studied, on the loading of pPH600 to the column at different ionic strength and temperature conditions.

An initial *screening* was performed to choose the best stepwise gradient for separation of sc pPH600 from impurities. The conditions defined were 2.25 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4) in first elution step and 0 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4) in third elution step. The obtained results in initial *screening* with pPH600 native sample (oc + sc) were the basis for this decision. The second elution gradient was adjusted and in this case was defined at 1.6 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4). Thus, the column was equilibrated 2.25 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4), at 10ºC and the stepwise gradient was further defined to 1.6 M and 0 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4).

Fig. 32A shows the chromatographic profile obtained by injection of pPH600 clarified *E. coli* lysate. An agarose gel electrophoresis analysis of the fractions eluting from the column (Fig. 32B) proved that the first peak corresponds to oc and ln isoforms and a small content of gDNA (Fig. 32B, lane 1), the second peak was the sc pPH600 with a small contamination of RNA, suggesting that the ionic strength need to be incresead in order to avoid the elution of RNA (Fig. 32B, lane 2), and the third peak strongly bond species such as, RNA and gDNA (Fig. 32B, lane 3). In third peak a small quantity of sc pPH600 also elutes (Fig. 32B, lane 3).

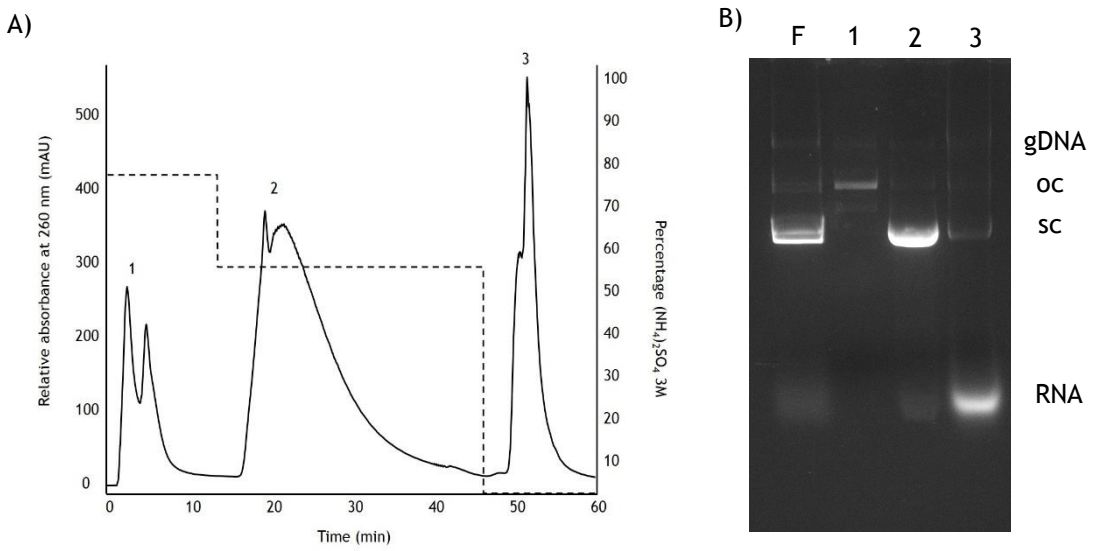


Figure 32 - A) Chromatographic profile showing the optimization of sc pPH600 purification from clarified *E. coli* lysate with L-tyrosine support. B) 1% Agarose gel electrophoresis for the analysis of collected fractions from L-tryptophan support. Lane F – Clarified *E.coli* lysate injected onto column; Lane 1 – gDNA, oc and ln pPH600 recovered from peak 1; Lane 2 – sc pPH600 and RNA recovered from peak 2 and Lane 3 – gDNA, RNA and small quantity of sc pPH600 recovered from peak 3.

Various experiments were been done to optimize the second elution step of gradient and the ionic strength in which is possible elute pPH600 without contaminats was 1.95 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4). Therefore, after the initial *screening* to optimize the ionic strength in second elution step, the column was equilibrated with 2.25 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4), at 10ºC. After injection of the clarified *E. coli* lysate sample, the elution of oc occurred immediately in the first elution step performed with 2.25 M of $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4), whereas the sc pPH600 remained bond, eluting only when the salt concentration was decreased to 1.95 M of $(NH_4)_2SO_4$. The RNA and gDNA were eluted only with 100 mM HEPES acid (pH 7.4) (without salt). As expected, these fractions of the total denatured gDNA and RNA were retarded in the column, since the single strands present in denatured gDNA enables an interaction with the support [93]. Thus, the remaining portion of the material adsorbed to the column was eluted only after decreasing the ionic strength to 100 mM HEPES acid (pH 7.4).

Fig. 33A shows the chromatographic profile obtained by injection of pPH600 clarified *E. coli* lysate. An agarose gel electrophoresis analysis of the fractions eluting from the column (Fig. 33B) proved that the first peak of unbond material corresponds to oc and small content of gDNA and sc pPH600 (Fig. 33B, lane 1), whereas the second peak was attributed to the sc pPH600 (Fig. 33B, lane 2). Finally, the third peak strongly bond species such as, RNA and gDNA (Fig.

33B, lane 3). In third peak a small quantity of sc pPH600 also elutes, although as depicted by agarose gel electrophoresis (Fig. 33B) the sc isoform was purified in the second peak. The gDNA also elutes in first peak because it is composed by species with different sizes (0.7-20 kbp) and hydrophobicity is related with size [93].

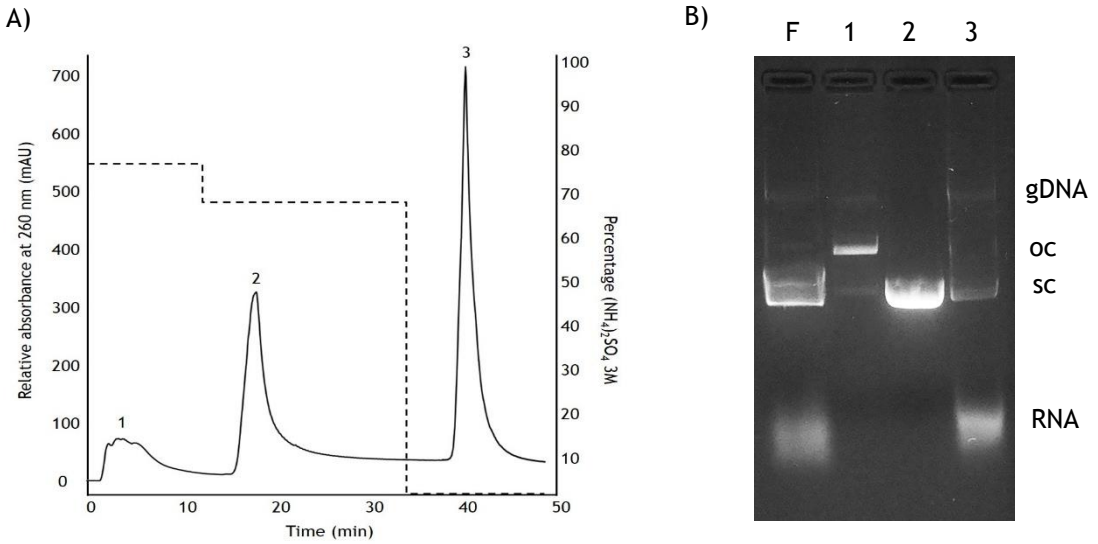


Figure 33 - A) Chromatographic profile showing the purification of sc pPH600 from clarified *E. coli* lysate with L-tyrosine support. B) 1% Agarose gel electrophoresis for the analysis of collected fractions from L-tryptophan support. Lane F – Clarified *E.coli* lysate injected onto column; Lane 1 – gDNA and oc pPH600 recovered from peak 1; Lane 2 – sc pPH600 recovered from peak 2 and Lane 3 – gDNA, RNA and small quantity of sc pPH600 recovered from peak 3.

The temperature effect on adsorption of pDNA isoforms and other *E. coli* nucleic acids on the L-tyrosine support was also evaluated. The increase of temperature (15 and 25ºC), shows that oc pPH600, RNA and gDNA did not alter the retention pattern. However the sc pPH600 was affected by temperature, decreasing the retention time and not allowing an effective separation. Various experiments were also performed at 4ºC and showed that RNA and gDNA were not affected by temperature, however, the separation of sc pPH600 was not possible since the low temperature affects the selectivity and specificty of support. The optimal temperature that was possible isolate sc pPH600 from impurities was 10ºC.

3.5.6 – pDNA quantification and purity assessment

The recovery yield and quality of the sc pDNA were assessed by a CIMac™ pDNA analytical column, according to the modified analytical method previously developed by Ângela Sousa and collaborators [80]. Briefly, the method allows the elution of RNA species in the flowtrough at 600 mM NaCl in 200 mM Tris-HCl (pH 8.0), oc and sc isoforms are separated during 10 min of linear gradient from 600 to 700 mM NaCl in 200 mM Tris-HCl (pH 8.0). After analysis of each peak with CIMac™ pDNA analytical column, the results of concentration, recovery yield and purity are depicted in table 6.

The recovery of sc pPH600 in each peak was related with the initial amount present in clarified *E. coli* lysate (Table 6). Finally, the purity degree of sc pPH600 present in peak 2 was determined from the ratio between the pDNA peak area and the sum of all the peak areas that appear in analytical chromatogram.

As shown in the table 6, this purification strategy revealed a sc pPH600 recovery yield of 56.28% and a purity of 98.23%. The total recuperation of sc pPH600 was not total and can be related with the use of $(NH_4)_2SO_4$ as adsorption buffer. Also in purification with L-histidine the recovery yield of sc pDNA was only 40%. Nevertheless, the sc pPH600 sample recovered can be used in therapeutic applications because it presents a homogeneity higher than 97% of sc isoform, according to the guidelines recommended by regulatory agencies.

Table 6 - Quantitative analysis of the sc pDNA, recovery yield and purity in each peak recovered from clarified lysate using L-tyrosine support.

Sample	Sc pDNA (µg/mL)	Volume (µL)	Sc pDNA (µg)	Recovery yield (%)	Purity (%)
Lysate sample	552.23	50	27.61	-	-
Peak 1	0.67	450	0.24	0.87	-
Peak 2	25.90	600	15.54	56.28	98.23
Peak 3	6.65	500	2.99	10.84	-

Also, through the analysis of agarose gel electrophoresis is possible to verify that RNA was totally eliminated in third peak as presented in fig. 33B, lane 3.

Protein, gDNA and endotoxins were determined by using the micro-BCA method, real-time PCR and kinetic-QCL Limulus amoebocyte lysate assays, respectively. The results presented in table 7 show a significant reduction of all impurities by comparing with the clarified lysate sample with the sc pPH600 purified with L-tyrosine support.

The analysis of protein content shows that after purification with L-tyrosine support, the sample of sc pPH600 in peak 2 did not present detectable levels of proteins. The presence of endotoxins was also evaluated and a significant decrease was observed, demonstrating that endotoxins, due to this hydrophobic character, bind to the L-tyrosine support more strongly than pPH600 (Table 7). The efficiency of L-tyrosine support to eliminate gDNA was also significant because the residual amount of gDNA in the sc pPH600 sample was 0.912 ng/µg. Overall, the content of impurities present in the sc pPH600 sample is in accordance with requirements of regulatory agencies.

Table 7 - Protein, endotoxin and gDNA assessment from clarified lysate sample and sc pDNA sample recovered using L-tyrosine support.

Sample	Protein (µg/mL)	Endotoxin (EU/µg sc pDNA)	gDNA (ng/ µg sc pDNA)
Regulatory agencies specifications [81, 117]	undetectable	< 0.1	< 2
Lysate sample	143.467	2.663	13.541
Peak 2 (sc pPH600)	undetectable	0.033	0.912

3.5.7 – *In vitro* transcription and G4 characterization

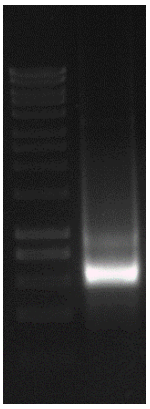
The transcription is the process of making RNA copy of a DNA template and are one of the key processes in gene expression. After purification of sc pPH600, the obtention of G4 structure was achieved with *in vitro* transcription of sc pPH600. The post-transcription digestion with DNase results in clean RNA bands (Fig.34A). In this case, the formation of G4 structures can occur in RNA because have the same sequence of strand in DNA that are able to form these structures with exception of change for "U" instead "T".

The G4 formation is detected by CD spectroscopy after transcription within sc pPH600 and incubation with a solution of 50 mM KCl to stabilize the structure. The transcript showed a CD spectrum having a positive band around 265 nm and a negative band around 240 nm suggesting, parallel G4 folding topology (Fig. 34B) [118].

Indeed, control sc pPH600 without transcription was also acquired and the CD spectrum showed a positive band at 275 nm and a negative band at 245 nm characteristic of typical B-DNA form as expected for pDNA (Fig. 34 B) [119].

In further experiments the aim is the transcription of pPH600 and G4 formation in DNA strand. These formation in double strand of DNA occur because the RNA produced remains bond to the template DNA strand, leaving the coding strand free and forming a loop with G4 structures [48]. These structures are called RNA-DNA hybrids [14].

A)



B)

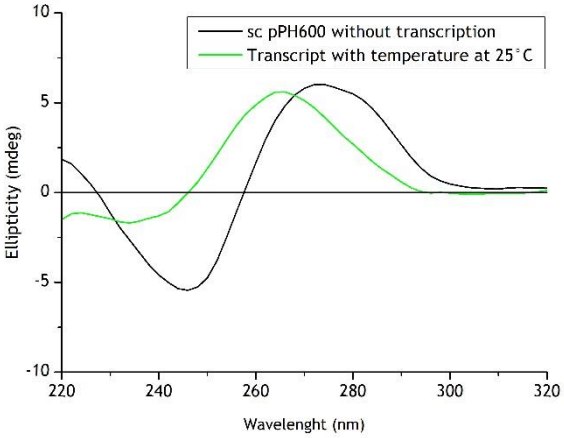


Figure 34 – RNA G4 formation upon transcription of plasmid pPH600. A) 1% agarose gel electrophoresis of transcript; Lane M – Marker and Lane 1 – Transcript with 604 bp. B) CD spectrum; sc pPH600 (black line) with characteristic negative band at 245 nm and positive band at 275 nm; transcript with temperature at 25°C (green line) with characteristic negative band at 240 nm and positive band at 265 nm.

Chapter IV

Conclusions

For many years it was thought that DNA was only the molecule which simply store the hereditary information in nucleus. The discover of double helix in 1953 by Watson and Crick opened new horizons to biologists and geneticists, leading to a new important discoveries about structure and function of DNA. In fact, besides the well-known double helix, DNA can adopt diverse structural conformations including duplexes, triplexes and quadruplexes. Four stranded structures known as G4 are located in important regions of genome and have been implicated in cellular processes such as, transcription, replication and recombination.

Nowadays, these structures are considered potential targets to develop new therapeutics for many diseases including many types of cancer and neurological diseases. The plasmids are non-viral vectors that have a complete eukaryotic expression cassette composed of the transcribed region domain that allow insert of the desired genes. Recent reports show that sc pPH600, which contains a fragment of the murine immunoglobulin Sγ3 switch region have more able to form G4 structures than oc or ln isoforms upon transcription. For that reason the production and purification of pPH600 becomes essential to further studies of *in vivo* G4 stabilization.

In present study the plasmid pPH600 was amplified in *E. coli* DH5α and the recombinant host grows significantly faster reaching an OD_{600} of 6 in just 4 h. This behavior suggest that the presence of a sequence that is able to form G4 have influence in growth of *E. coli* DH5α and consequently, in pPH600 amplification. Moreover, the analysis of pPH600 samples recovered by lysis through kit and modified alkaline lysis revealed a higher content of sc pPH600.

The plasmid pPH600 obtained by amplification in recombinant host *E. coli* DH5α was purified by affinity chromatography with three different supports. Recently, some amino acids have been used as affinity ligands, producing interesting results in pDNA purification. For that reason, the amino acids, L-arginine, L-tryptophan and L-tyrosine were chosen as affinity ligands.

The commercial L-arginine support was used to compare the binding/elution conditions.

The supports L-tyrosine and L-tryptophan were prepared by covalent immobilisation of 1,4-butanediol diglycidyl ether onto Sepharose CL-6B.

For characterization of supports the bead morphology of activated Sepharose CL-6B was compared with L-tryptophan and L-tyrosine supports by SEM. This experiment showed no significant alteration in characteristics before and after immobilization of the ligands in the structure of Sepharose CL-6B. Thus, the chromatographic efficiency was not affected by immobilisation.

In order to evaluate the behavior of pPH600 with supports an initial *screening* with pPH600 native sample (oc + sc) were performed. The *screening* allow the comparison of temperature and salt conditions necessary to purify sc pPH600. The L-tyrosine support showed the complete separation of two isoforms by decreasing a stepwise gradient from 2.25 to 0 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4).

The L-arginine and L-tryptophan supports showed interesting results however, the total recuperation of sc pPH600 was not achieved.

Although the separation of sc pPH600 from pre purified samples have been achieved, the use in therapeutic applications is not suitable because the kit used to perform the pre-purification have toxic products and enzymes.

Based on these considerations, the L-tyrosine support was used to separate sc pPH600 from clarified *E. coli* lysate. The initial *screening* allow determine the first elution step of stepwise gradient, 2.25 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4), in which occurs the elution of small quantity of gDNA and oc pPH600. The second elution step of stepwise gradient was optimized to perform the elution of sc pPH600 without impurities and was defined in 1.95 M $(NH_4)_2SO_4$ in 100 mM HEPES acid (pH 7.4). Finally the elution of gDNA and RNA was achieved with 100 mM HEPES acid (pH 7.4) without $(NH_4)_2SO_4$.

The quality and purity of sc pPH600 resultant from purification with L-tyrosine support was evaluated by using the guidelines of regulatory agencies. This purification strategy revealed a sc pPH600 recovery yield of 56.28% and a purity of 98.23%, higher than the 97% recommended by regulatory agencies. The evaluation of impurities (gDNA, RNA, endotoxins and proteins) reveal that this strategy reduce considerably the levels of impurities, making product suitable for further *in vivo* studies.

The formation of in G-loop was assessed by transcription of pPH600 and confirmed by CD. The results reveal that the formed transcripts had a G4 parallel topology because the CD spectrum showed a negative band at 240 nm and a positive band at 265 nm, characteristic of parallel G4. Moreover, was possible to conclude that after transcription there was G-loops formation in plasmid with the stabilization of complementary strand.

In conclusion, this affinity strategy based on new affinity ligands reveal promising results for pDNA purification. In particular, the L-tyrosine support showed higher efficiency in separate sc pDNA from other impurities present in clarified *E. coli* lysate. Moreover using a stepwise gradient with only two steps revealed technically and economically advantages of this strategy.

Chapter V

Future considerations

In order to continue this work, more optimizations on the chromatographic process has to be done. The purification with L-tryptophan support need to be optimized namely to separate sc pDNA from a complex mixture of clarified *E. coli* lysate in order to evaluate the behavior of support with other host nucleic acids and impurities. In purification of sc pDNA from clarified *E. coli* lysate with L-tyrosine support the yield and purity could be improved using design of experiments.

The lack of suitable affinity ligands to purify with efficiency pDNA have been received the interest of researchers. The amino acids can interact with nucleotides through multiple non-covalent interactions, however neither all amino acids have the same characteristics. The use of dipeptides based on amino acids seems to be a suitable alternative for conjugation of different properties. For example the use of a dipeptide L-tyrosine-L-arginine can improve the efficiency of sc pDNA separation because probably have high predominance of electrostatic interactions due to L-arginine and at the same time high prevalence of hydrophobic and π-π interactions from L-tyrosine.

The transcription need to be optimized in order to confirm the formation of G-loop and consequently G4 structures in DNA. For this proposes the post transcription treatment should be done with RNase A to degrade mRNA free.

Moreover, the transcription can be induced *in vivo* to confirm the presence of G4 and to stabilize these structures with ligands not allowing the transcription of some genes that are the cause of pathologies such as, cancer and neurodegenerative diseases.

Chapter VI

Bibliography

[1] R. Dahm, Friedrich Miescher and the discovery of DNA, Developmental biology, 278 (2005) 274-288.

[2] J. Stenesh, Nucleic Acids, in: Biochemistry, Springer US, 1998, pp. 171-200.

[3] I. Hargittai, The tetranucleotide hypothesis: a centennial, Structural Chemistry, 20 (2009) 753-756.

[4] L. Pray, Discovery of DNA Structure and Function: Watson and Crick, Nature Education, 1 (2008) 100.

[5] K.L. Manchester, Historical Opinion: Erwin Chargaff and his 'rules' for the base composition of DNA: why did he fail to see the possibility of complementarity?, Trends in Biochemical Sciences, 33 (2008) 65-70.

[6] N.W. Luedtke, Targeting G-Quadruplex DNA with Small Molecules, CHIMIA International Journal for Chemistry, 63 (2009) 134-139.

[7] X. Du, D. Wojtowicz, A.A. Bowers, D. Levens, C.J. Benham, T.M. Przytycka, The genome-wide distribution of non-B DNA motifs is shaped by operon structure and suggests the transcriptional importance of non-B DNA structures in Escherichia coli, Nucleic Acids Research, 41 (2013) 5965-5977.

[8] M. Gellert, M.N. Lipsett, D.R. Davies, Helix formation by guanylic acid, Proceedings of the National Academy of Sciences of the United States of America, 48 (1962) 2013-2018.

[9] G.W. Collie, G.N. Parkinson, The application of DNA and RNA G-quadruplexes to therapeutic medicines, Chemical Society reviews, 40 (2011) 5867-5892.

[10] J. Bidzinska, G. Cimino-Reale, N. Zaffaroni, M. Folini, G-quadruplex structures in the human genome as novel therapeutic targets, Molecules, 18 (2013) 12368-12395.

[11] T.I. Gaynutdinov, E.A. Englund, D.H. Appella, M.I. Onyshchenko, R.D. Neumann, I.G. Panyutin, G-Quadruplex Formation Between G-Rich PNA and Homologous Sequences in Oligonucleotides and Supercoiled Plasmid DNA, Nucleic acid therapeutics, 25 (2015) 78-84.

[12] L. Yuan, T. Tian, Y. Chen, S. Yan, X. Xing, Z. Zhang, Q. Zhai, L. Xu, S. Wang, X. Weng, B. Yuan, Y. Feng, X. Zhou, Existence of G-quadruplex structures in promoter region of oncogenes confirmed by G-quadruplex DNA cross-linking strategy, Scientific Reports, 3 (2013) 1811.

[13] A. Henderson, Y. Wu, Y.C. Huang, E.A. Chavez, J. Platt, F.B. Johnson, R.M. Brosh, Jr., D. Sen, P.M. Lansdorp, Detection of G-quadruplex DNA in mammalian cells, Nucleic acids research, 42 (2014) 860-869.

[14] M.L. Duquette, P. Handa, J.A. Vincent, A.F. Taylor, N. Maizels, Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA, Genes & development, 18 (2004) 1618-1629.

[15] F. Sousa, J.A. Queiroz, Supercoiled plasmid quality assessment by analytical arginine-affinity chromatography, Journal of chromatography A, 1218 (2011) 124-129.

[16] A. Sousa, F. Sousa, J.A. Queiroz, Impact of lysine-affinity chromatography on supercoiled plasmid DNA purification, Journal of chromatography B, Analytical technologies in the biomedical and life sciences, 879 (2011) 3507-3515.

[17] F. Sousa, C.T. Tomaz, D.M. Prazeres, J.A. Queiroz, Separation of supercoiled and open circular plasmid DNA isoforms by chromatography with a histidine-agarose support, Analytical biochemistry, 343 (2005) 183-185.

[18] J.F. Valente, A. Sousa, J.A. Queiroz, F. Sousa, Selective purification of supercoiled p53-encoding pDNA with L-methionine-agarose matrix, Analytical biochemistry, 459 (2014) 61-69.

[19] A. Verdian Doghaei, M.R. Housaindokht, M.R. Bozorgmehr, Molecular crowding effects on conformation and stability of G-quadruplex DNA structure: insights from molecular dynamics simulation, Journal of theoretical biology, 364 (2015) 103-112.

[20] J.L. Huppert, Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes, Chemical Society reviews, 37 (2008) 1375-1384.

[21] L.B. König Sebastian, C. Evans Amanda, L. Huppert Julian, Seven essential questions on G-quadruplexes, BioMolecular Concepts, 1 (2010) 197-213.

[22] P.L. Tran, A. De Cian, J. Gros, R. Moriyama, J.L. Mergny, Tetramolecular quadruplex stability and assembly, Topics in current chemistry, 330 (2013) 243-273.

[23] A. Wong, G. Wu, Selective Binding of Monovalent Cations to the Stacking G-Quartet Structure Formed by Guanosine 5‘-Monophosphate: A Solid-State NMR Study, Journal of the American Chemical Society, 125 (2003) 13895-13905.

[24] E.A. Venczel, D. Sen, Parallel and antiparallel G-DNA structures from a complex telomeric sequence, Biochemistry, 32 (1993) 6220-6228.

[25] D. Yang, K. Okamoto, Structural insights into G-quadruplexes: towards new anticancer drugs, Future medicinal chemistry, 2 (2010) 619-646.

[26] G. Witz, A. Stasiak, DNA supercoiling and its role in DNA decatenation and unknotting, Nucleic Acids Research, 38 (2010) 2119-2133.

[27] D. Sun, L.H. Hurley, The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression, Journal of medicinal chemistry, 52 (2009) 2863-2874.

[28] R.P. Bowater, Supercoiled DNA: Structure, in: eLS, John Wiley & Sons, Ltd, 2001.

[29] D.M.F. Prazeres, Structure, in: Plasmid Biopharmaceuticals, John Wiley & Sons, Inc., 2011, pp. 85-128.

[30] J. YiqináLee, DNA gyrase-driven generation of a G-quadruplex from plasmid DNA, Chemical Communications, 49 (2013) 8317-8319.

[31] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, Quadruplex DNA: sequence, topology and structure, Nucleic acids research, 34 (2006) 5402-5415.

[32] S. Balasubramanian, S. Neidle, G-quadruplex nucleic acids as therapeutic targets, Current opinion in chemical biology, 13 (2009) 345-353.

[33] T.M. Bryan, P. Baumann, G-quadruplexes: from guanine gels to chemotherapeutics, Molecular biotechnology, 49 (2011) 198-208.

[34] X. Cang, J. Sponer, T.E. Cheatham, 3rd, Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes, Nucleic acids research, 39 (2011) 4499-4512.

[35] J.L. Huppert, S. Balasubramanian, Prevalence of quadruplexes in the human genome, Nucleic acids research, 33 (2005) 2908-2916.

[36] N. Maizels, L.T. Gray, The G4 genome, PLoS genetics, 9 (2013) e1003468.

[37] P.L. Tran, J.L. Mergny, P. Alberti, Stability of telomeric G-quadruplexes, Nucleic acids research, 39 (2011) 3282-3294.

[38] A.M. Zahler, J.R. Williamson, T.R. Cech, D.M. Prescott, Inhibition of telomerase by G-quartet DNA structures, Nature, 350 (1991) 718-720.

[39] J. Eddy, N. Maizels, Gene function correlates with potential for G4 DNA formation in the human genome, Nucleic acids research, 34 (2006) 3887-3896.

[40] D. Hanahan, R.A. Weinberg, The hallmarks of cancer, Cell, 100 (2000) 57-70.

[41] T.A. Brooks, S. Kendrick, L. Hurley, Making sense of G-quadruplex and i-motif functions in oncogene promoters, The FEBS journal, 277 (2010) 3459-3469.

[42] L.A. Hanakahi, H. Sun, N. Maizels, High affinity interactions of nucleolin with G-G-paired rDNA, The Journal of biological chemistry, 274 (1999) 15908-15912.

[43] J. Stavnezer, J.E. Guikema, C.E. Schrader, Mechanism and regulation of class switch recombination, Annual review of immunology, 26 (2008) 261-292.

[44] H. Leung, N. Maizels, Transcriptional regulatory elements stimulate recombination in extrachromosomal substrates carrying immunoglobulin switch-region sequences, Proceedings of the National Academy of Sciences of the United States of America, 89 (1992) 4154-4158.

[45] J.M. Vaquerizas, S.K. Kummerfeld, S.A. Teichmann, N.M. Luscombe, A census of human transcription factors: function, expression and evolution, Nature reviews. Genetics, 10 (2009) 252-263.

[46] E.D. Larson, M.L. Duquette, W.J. Cummings, R.J. Streiff, N. Maizels, MutSalpha binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions, Current biology : CB, 15 (2005) 470-474.

[47] N. Maizels, Dynamic roles for G4 DNA in the biology of eukaryotic cells, Nature structural & molecular biology, 13 (2006) 1055-1059.

[48] K.J. Neaves, J.L. Huppert, R.M. Henderson, J.M. Edwardson, Direct visualization of G-quadruplexes in DNA using atomic force microscopy, Nucleic acids research, 37 (2009) 6269-6275.

[49] M.L. Duquette, P. Pham, M.F. Goodman, N. Maizels, AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation, Oncogene, 24 (2005) 5791-5798.

[50] N.J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure, Nature protocols, 1 (2006) 2876-2890.

[51] G.R. Bishop, J.B. Chaires, Characterization of DNA structures by circular dichroism, Current protocols in nucleic acid chemistry, Chapter 7 (2003) 7.11.11-17.11.18.

[52] V. Brazda, L. Haronikova, J.C. Liao, M. Fojta, DNA and RNA quadruplex-binding proteins, International journal of molecular sciences, 15 (2014) 17493-17517.

[53] E.H. Blackburn, Switching and signaling at the telomere, Cell, 106 (2001) 661-673.

[54] O. Calcagnile, D. Gisselsson, Telomere dysfunction and telomerase activation in cancer--a pathological paradox?, Cytogenetic and genome research, 118 (2007) 270-276.

[55] T. Shalaby, G. Fiaschetti, K. Nagasawa, K. Shin-ya, M. Baumgartner, M. Grotzer, G-quadruplexes as potential therapeutic targets for embryonal tumors, Molecules, 18 (2013) 12500-12537.

[56] J.C. Grigg, N. Shumayrikh, D. Sen, G-Quadruplex Structures Formed by Expanded Hexanucleotide Repeat RNA and DNA from the Neurodegenerative Disease-Linked C9orf72 Gene Efficiently Sequester and Activate Heme, PLoS ONE, 9 (2014) e106449.

[57] M.R. Santoro, S.M. Bray, S.T. Warren, Molecular mechanisms of fragile X syndrome: a twenty-year perspective, Annual review of pathology, 7 (2012) 219-245.

[58] H. Kobayashi, K. Abe, T. Matsuura, Y. Ikeda, T. Hitomi, Y. Akechi, T. Habu, W. Liu, H. Okuda, A. Koizumi, Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement, American journal of human genetics, 89 (2011) 121-130.

[59] S. Mead, M. Poulter, J. Uphill, J. Beck, J. Whitfield, T.E. Webb, T. Campbell, G. Adamson, P. Deriziotis, S.J. Tabrizi, H. Hummerich, C. Verzilli, M.P. Alpers, J.C. Whittaker, J. Collinge, Genetic risk factors for variant Creutzfeldt-Jakob disease: a genome-wide association study, The Lancet. Neurology, 8 (2009) 57-66.

[60] S. Mead, T.E. Webb, T.A. Campbell, J. Beck, J.M. Linehan, S. Rutherfoord, S. Joiner, J.D. Wadsworth, J. Heckmann, S. Wroe, L. Doey, A. King, J. Collinge, Inherited prion disease with 5-OPRI: phenotype modification by repeat length and codon 129, Neurology, 69 (2007) 730-738.

[61] C. Voß, Downstream Processing of Plasmid DNA for Gene Therapy and Genetic Vaccination, Chemical Engineering & Technology, 31 (2008) 858-863.

[62] M.A. Kay, State-of-the-art gene-based therapies: the road ahead, Nature Reviews. Genetics, 12 (2011) 316-328.

[63] A. Mountain, Gene therapy: the first decade, Trends in biotechnology, 18 (2000) 119-128.

[64] M.A. Liu, DNA vaccines: an historical perspective and view to the future, Immunological Reviews, 239 (2011) 62-84.

[65] T. Wirth, N. Parker, S. Yla-Herttuala, History of gene therapy, Gene, 525 (2013) 162-169.

[66] M.H. Porteus, J.P. Connelly, S.M. Pruett, A look to future directions in gene therapy research for monogenic diseases, PLoS genetics, 2 (2006) e133.

[67] R.M. Blaese, K.W. Culver, A.D. Miller, C.S. Carter, T. Fleisher, M. Clerici, G. Shearer, L. Chang, Y. Chiang, P. Tolstoshev, J.J. Greenblatt, S.A. Rosenberg, H. Klein, M. Berger, C.A. Mullen, W.J. Ramsey, L. Muul, R.A. Morgan, W.F. Anderson, T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years, Science, 270 (1995) 475-480.

[68] N. Nayerossadat, T. Maedeh, P.A. Ali, Viral and nonviral delivery systems for gene delivery, Advanced biomedical research, 1 (2012) 27.

[69] H. Yin, R.L. Kanasty, A.A. Eltoukhy, A.J. Vegas, J.R. Dorkin, D.G. Anderson, Non-viral vectors for gene-based therapy, Nature reviews. Genetics, 15 (2014) 541-555.

[70] M. Ingolotti, O. Kawalekar, D.J. Shedlock, K. Muthumani, D.B. Weiner, DNA vaccines for targeting bacterial infections, Expert review of vaccines, 9 (2010) 747-763.

[71] B. Ferraro, M.P. Morrow, N.A. Hutnick, T.H. Shin, C.E. Lucke, D.B. Weiner, Clinical applications of DNA vaccines: current progress, Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 53 (2011) 296-302.

[72] M.A. Kutzler, D.B. Weiner, DNA vaccines: ready for prime time?, Nature reviews. Genetics, 9 (2008) 776-788.

[73] D.M.F. Prazeres, G.A. Monteiro, Plasmid Biopharmaceuticals, Microbiology Spectrum, 2 (2014) 1-18.

[74] W. Schumann, The Biology of Plasmids, in: Plasmids for Therapy and Vaccination, Wiley-VCH Verlag GmbH, 2007, pp. 1-28.

[75] G.N.M. Ferreira, Chromatographic Approaches in the Purification of Plasmid DNA for Therapy and Vaccination, Chemical Engineering & Technology, 28 (2005) 1285-1294.

[76] A. Ghanem, R. Healey, F.G. Adly, Current trends in separation of plasmid DNA vaccines: a review, Analytica chimica acta, 760 (2013) 1-15.

[77] G. del Solar, R. Giraldo, M.J. Ruiz-Echevarría, M. Espinosa, R. Díaz-Orejas, Replication and Control of Circular Bacterial Plasmids, Microbiology and Molecular Biology Reviews, 62 (1998) 434-464.

[78] J.C. Diaz Ricci, M.E. Hernandez, Plasmid effects on Escherichia coli metabolism, Critical reviews in biotechnology, 20 (2000) 79-108.

[79] G.N.M. Ferreira, G.A. Monteiro, D.M.F. Prazeres, J.M.S. Cabral, Downstream processing of plasmid DNA for gene therapy and DNA vaccine applications, Trends in Biotechnology, 18 (2000) 380-388.

[80] A. Sousa, A.M. Almeida, U. Cernigoj, F. Sousa, J.A. Queiroz, Histamine monolith versatility to purify supercoiled plasmid deoxyribonucleic acid from Escherichia coli lysate, Journal of chromatography. A, 1355 (2014) 125-133.

[81] M.M. Diogo, J.A. Queiroz, D.M.F. Prazeres, Chromatography of plasmid DNA, Journal of Chromatography A, 1069 (2005) 3-22.

[82] G.J. Gram, A. Fomsgaard, M. Thorn, S.M. Madsen, J. Glenting, Immunological analysis of a Lactococcus lactis-based DNA vaccine expressing HIV gp120, Genetic Vaccines and Therapy, 5 (2007) 3-3.

[83] F. Silva, L. Passarinha, F. Sousa, J.A. Queiroz, F.C. Domingues, Influence of growth conditions on plasmid DNA production, Journal of microbiology and biotechnology, 19 (2009) 1408-1414.

[84] P.A. Shamlou, Scaleable processes for the manufacture of therapeutic quantities of plasmid DNA, Biotechnology and applied biochemistry, 37 (2003) 207-218.

[85] D.M.F. Prazeres, G.N.M. Ferreira, Design of flowsheets for the recovery and purification of plasmids for gene therapy and DNA vaccination, Chemical Engineering and Processing: Process Intensification, 43 (2004) 609-624.

[86] D.M.F. Prazeres, G.N.M. Ferreira, G.A. Monteiro, C.L. Cooney, J.M.S. Cabral, Large-scale production of pharmaceutical-grade plasmid DNA for gene therapy: problems and bottlenecks, Trends in Biotechnology, 17 (1999) 169-174.

[87] C. Caramelo-Nunes, C.T. Tomaz, Specific recognition of supercoiled plasmid DNA by affinity chromatography using a synthetic aromatic ligand, Methods in molecular biology, 1286 (2015) 47-54.

[88] A. Sousa, F. Sousa, J.A. Queiroz, Differential interactions of plasmid DNA, RNA and genomic DNA with amino acid-based affinity matrices, Journal of separation science, 33 (2010) 2610-2618.

[89] F. Sousa, C. Cruz, J.A. Queiroz, Amino acids-nucleotides biomolecular recognition: from biological occurrence to affinity chromatography, Journal of molecular recognition : JMR, 23 (2010) 505-518.

[90] D.M. Prazeres, T. Schluep, C. Cooney, Preparative purification of supercoiled plasmid DNA using anion-exchange chromatography, Journal of chromatography. A, 806 (1998) 31-45.

[91] F. Sousa, D.M. Prazeres, J.A. Queiroz, Affinity chromatography approaches to overcome the challenges of purifying plasmid DNA, Trends in biotechnology, 26 (2008) 518-525.

[92] L.-Z. Li, Y. Liu, M.-S. Sun, Y.-M. Shao, Effect of salt on purification of plasmid DNA using size-exclusion chromatography, Journal of Chromatography A, 1139 (2007) 228-235.

[93] F. Sousa, S. Freitas, A.R. Azzoni, D.M. Prazeres, J. Queiroz, Selective purification of supercoiled plasmid DNA from clarified cell lysates with a single histidine-agarose chromatography step, Biotechnology and applied biochemistry, 45 (2006) 131-140.

[94] A. Eon-Duval, G. Burke, Purification of pharmaceutical-grade plasmid DNA by anion-exchange chromatography in an RNase-free process, Journal of chromatography. B, Analytical technologies in the biomedical and life sciences, 804 (2004) 327-335.

[95] M.M. Diogo, J.A. Queiroz, G.A. Monteiro, S.A.M. Martins, G.N.M. Ferreira, D.M.F. Prazeres, Purification of a cystic fibrosis plasmid vector for gene therapy using hydrophobic interaction chromatography, Biotechnology and Bioengineering, 68 (2000) 576-583.

[96] D.M.F. Prazeres, Final Purification, in: Plasmid Biopharmaceuticals, John Wiley & Sons, Inc., 2011, pp. 497-542.

[97] D.S. Hage, Affinity Chromatography: A Review of Clinical Applications, Clinical Chemistry, 45 (1999) 593-615.

[98] C. Caramelo-Nunes, D. Bicho, P. Almeida, J.C. Marcos, C.T. Tomaz, Dynamic binding capacity and specificity of 3,8-diamino-6-phenylphenanthridine-Sepharose support for purification of supercoiled plasmid deoxyribonucleic acid, Journal of chromatography. A, 1307 (2013) 91-98.

[99] F. Sousa, T. Matos, D.M. Prazeres, J.A. Queiroz, Specific recognition of supercoiled plasmid DNA in arginine affinity chromatography, Analytical biochemistry, 374 (2008) 432-434.

[100] A. Afonso, P. Pereira, J.A. Queiroz, A. Sousa, F. Sousa, Purification of pre-miR-29 by a new O-phospho-l-tyrosine affinity chromatographic strategy optimized using design of experiments, Journal of chromatography. A, 1343 (2014) 119-127.

[101] A. Holt, R.F. de Almeida, T.K. Nyholm, L.M. Loura, A.E. Daily, R.W. Staffhorst, D.T. Rijkers, R.E. Koeppe, 2nd, M. Prieto, J.A. Killian, Is there a preferential interaction between cholesterol and tryptophan residues in membrane proteins?, Biochemistry, 47 (2008) 2638-2649.

[102] C.M. Baker, G.H. Grant, Role of aromatic amino acids in protein-nucleic acid recognition, Biopolymers, 85 (2007) 456-470.

[103] C. Cruz, A. Sousa, F. Sousa, J.A. Queiroz, Affinity analysis between immobilized l-arginine and plasmid isoforms provided by surface plasmon resonance, Analytical Methods, 5 (2013) 1682-1686.

[104] C. Cruz, E.J. Cabrita, J.A. Queiroz, Screening nucleotide binding to amino acid-coated supports by surface plasmon resonance and nuclear magnetic resonance, Analytical and bioanalytical chemistry, 401 (2011) 983-993.

[105] Tiago Santos, Josué Carvalho, Cyrille Feijó, Marta C. Corvo, Eurico J. Cabrita, J.A. Queiroz, C. Cruz., Binding affinity between L-tryptophan containing dipeptide derivatives and nucleic acids, International Journal of Biological Macromolecules, (2015) submitted.

[106] Carla Cruz, Soraia Ferreira, Joana F.A. Valente, Marta C. Corvo, Eurico J. Cabrita, Fani Sousa, J.A. Queiroz, Synthesis, affinity analysis and application of dipeptides derived from L-tyrosine in plasmid DNA purification., Journal of Biochemistry, (2015) submitted.

[107] L. Sundberg, J. Porath, Preparation of adsorbents for biospecific affinity chromatography : I. Attachment of group-containing ligands to insoluble polymers by means of bifunctional oxiranes, Journal of Chromatography A, 90 (1974) 87-98.

[108] F. Silva, J.A. Queiroz, F.C. Domingues, Evaluating metabolic stress and plasmid stability in plasmid DNA production by Escherichia coli, Biotechnology advances, 30 (2012) 691-708.

[109] J.R. Cooke, E.A. McKie, J.M. Ward, E. Keshavarz-Moore, Impact of intrinsic DNA structure on processing of plasmids for gene therapy and DNA vaccines, Journal of biotechnology, 114 (2004) 239-254.

[110] R. Mayer, F. Toulme, T. Montenay-Garestier, C. Helene, The role of tyrosine in the association of proteins and nucleic acids. Specific recognition of single-stranded nucleic acids by tyrosine-containing peptides, The Journal of biological chemistry, 254 (1979) 75-82.

[111] F. Sousa, D.M. Prazeres, J.A. Queiroz, Binding and elution strategy for improved performance of arginine affinity chromatography in supercoiled plasmid DNA purification, Biomedical chromatography : BMC, 23 (2009) 160-165.

[112] P. Pereira, A. Sousa, J. Queiroz, I. Correia, A. Figueiras, F. Sousa, Purification of pre-miR-29 by arginine-affinity chromatography, Journal of chromatography. B, Analytical technologies in the biomedical and life sciences, 951-952 (2014) 16-23.

[113] D.A. Dougherty, Cation-π Interactions Involving Aromatic Amino Acids, The Journal of Nutrition, 137 (2007) 1504S-1508S.

[114] F. Sousa, D.M. Prazeres, J.A. Queiroz, Improvement of transfection efficiency by using supercoiled plasmid DNA purified with arginine affinity chromatography, The journal of gene medicine, 11 (2009) 79-88.

[115] G. Lancelot, C. Helene, Selective recognition of nucleic acids by proteins: the specificity of guanine interaction with carboxylate ions, Proceedings of the National Academy of Sciences of the United States of America, 74 (1977) 4872-4875.

[116] A. Sousa, F. Sousa, D.M. Prazeres, J.A. Queiroz, Histidine affinity chromatography of homo-oligonucleotides. Role of multiple interactions on retention, Biomedical chromatography : BMC, 23 (2009) 745-753.

[117] J. Stadler, R. Lemmens, T. Nyhammar, Plasmid DNA purification, The Journal of Gene Medicine, 6 (2004) S54-S66.

[118] A.R. Haeusler, C.J. Donnelly, G. Periz, E.A.J. Simko, P.G. Shaw, M.-S. Kim, N.J. Maragakis, J.C. Troncoso, A. Pandey, R. Sattler, J.D. Rothstein, J. Wang, C9orf72 nucleotide repeat structures initiate molecular cascades of disease, Nature, 507 (2014) 195-200.

[119] F. Sousa, D.M.F. Prazeres, J.A. Queiroz, Circular dichroism investigation of the effect of plasmid DNA structure on retention in histidine chromatography, Archives of Biochemistry and Biophysics, 467 (2007) 154-162.