UNIVERSIDADE DA BEIRA INTERIOR
Ciências

Biosynthesis of a G-Quadruplex—forming sequence and its stabilization by ligands

Josué Leandro de Oliveira e Carvalho

Dissertação para obtenção do Grau de Mestre em
Biotecnologia
(2º ciclo de estudos)

Orientador: Prof.ª Doutora Carla Cruz
Co-orientador: Prof. Doutor Eurico Cabrita

Covilhã, Junho de 2015

“If I have seen further it is by standing on the shoulders of giants”

Sir Issac Newton

To my parents,

for their never-ending love and support in all my efforts.

Acknowledgments

First of all, I would like to express my profound gratitude to my supervisor, Professor Carla Cruz, for all the support, guidance and patience throughout this past year. I value the freedom that was given to me in order to explore my research and to be involved in a variety of projects, which I believe will be very rewarding in the future.

I would like to thank Professor Eurico Cabrita for being constantly involved in this work, supporting the project with useful ideas and expertise.

To Health Sciences Research Centre for providing the conditions and equipment for the development of this research project.

I am also thankful and grateful to Patrícia Pereira, for all the precious support and assistance which allowed me to overcome the challenges of this research. I thank you for all the knowledge that you passed on to me.

To Eduarda Coutinho who did the sequencing experiments, thank you for all the valuable advices and availability during this last year, you were a pleasure to work with.

To my lab colleagues, thanks for all the moments of distraction and joy that you gave me, as well as the knowledge that we shared. To João for providing the synthetic molecules, thank you.

I express the most sincere gratitude and appreciation to my parents, for all the support during my whole life and for giving me the foundations which made me the person I am today. I love you both. Also, my brother Carlitos and my sister Filó, you're the best.

Last but not least, to Cátia, for your love, patience and fortitude. Thank you for everything.

Resumo

Além da forma B Watson e Crick do ADN duplex, os G-quadruplexes são estruturas de ADN de quatro cadeias, formadas *in vivo* pela auto-associação de sequências ricas em guaninas. Estas podem ser formadas por uma, duas ou quatro cadeias distintas de ADN e apresentar uma diversidade de topologias, definidas pela orientação da cadeia, tamanho dos *loops* e a sequência. G-quadruplexes podem ser encontrados nos telómeros, regiões de troca das imunoglobulinas e nas regiões dos promotores génicos. A localização biologicamente relevante no genoma faz com que estas estruturas altamente ordenadas sejam um alvo atrativo do desenho de fármacos e o desenvolvimento de ligandos altamente específicos que ligam e estabilizam o G-quadruplex com ação terapêutica. Neste trabalho, descreve-se a biossíntese da nova sequência de ADN rica em guaninas e formadora de G-quadruplex 58Sγ3, utilizando amplificação por plasmídeo. A recuperação e purificação do oligonucleótido 58Sγ3 é efetuada por cromatografia de exclusão molecular. A formação de G-quadruplex é promovida e a sua topologia é determinada por dicroísmo circular. A estabilização da estrutura do G-quadruplex com ligandos derivados de quinolina e naftaleno é estudada utilizando ensaios de estabilização térmica no dicroísmo circular, G4-FID e PCR-stop. Os resultados sugerem que 58Sγ3 adota uma estrutura G-quadruplex paralela em tampão 500 mM KCl e que os ligandos de naftaleno ligam e estabilizam a estrutura do G-quadruplex. Os ligandos demonstraram também ser específicos do G-quadruplex em relação ao ADN duplex além de inibir a *Taq* ADN polimerase. Este trabalho fornece evidência da formação de G-quadruplex nas regiões de troca das imunoglobulinas. Além disso, sugere que os derivados de naftaleno atuam como ligandos do G-quadruplex e que podem ser potencialmente utilizados para inibir a transcrição de genes em células tumorais.

Palavras-chave

G-quadruplex; ADN plasmídico; biossíntese; cromatografia de exclusão molecular; ligandos de quadruplex; dicroísmo circular; PCR-stop; G4-FID

Resumo Alargado

Durante muitos anos, o ADN foi caracterizado como uma “molécula passiva” cujas funções seriam unicamente o armazenamento de material genético no núcleo das células. A descoberta de Watson e Crick em 1953 abriu novas perspetivas para que fosse descoberto o segredo da vida por biólogos e geneticistas. As suas publicações revelaram a estrutura química do ADN como uma hélice dupla e explicaram como a informação genética era passada de geração em geração, como havia sido postulado um século antes por Mendel. Desde então, diversas estruturas alternativas de ADN derivadas do modelo de dupla hélice foram propostas e o seu papel biológico tem sido alvo de intensa investigação.

G-quadruplex é uma estrutura altamente ordenada que resulta da capacidade de sequências de ADN ricas em guaninas se organizarem espontaneamente em estruturas tetraméricas de quatro cadeias. A estrutura do G-quadruplex é constituída por diversas tétradas de guaninas empilhadas verticalmente, denominadas de G-quarteto, o núcleo da formação do G-quadruplex. Cada G-quarteto é formado pelo arranjo planar de quatro guaninas ligadas por pontes de hidrogénio Hoogsteen. A formação de G-quadruplexes é observada em solução em condições fisiológicas, sendo a sua formação e estabilização dependente de catiões monovalentes, especialmente potássio e sódio, sendo o potássio considerado mais relevante biologicamente dada a sua elevada concentração intracelular, quando comparado com os outros iões. Os G-quadruplexes podem ser formados por uma (intramolecular), duas ou quatro (intermolecular) cadeias distintas de ADN e apresentar diversas topologias, definidas pela orientação das cadeias (paralelo ou antiparalelo), tamanho e sequência dos *loops*. Estudos estruturais demonstraram esse mesmo polimorfismo estrutural, dependente das condições experimentais, a presença e coexistência de iões metálicos, a concentração do ião e condições de aglomerado molecular. Sequências capazes de formar G-quadruplex podem ser encontradas nos telómeros, nas regiões de promotores génicos (notavelmente nos oncogenes), nas regiões de troca das imunoglobulinas, assim como noutros locais menos comuns. Neste contexto, foi já proposto que o G-quadruplex esteja envolvido em diversos processos biológicos importantes como a transcrição, tradução, replicação e recombinação de ADN, além da manutenção da estabilidade do genoma. Deste modo, o G-quadruplex tem sido extensivamente estudado como um potencial alvo terapêutico e o desenvolvimento de pequenas moléculas, altamente específicas, que ligam e estabilizam a sua estrutura foi intensificado.

Diversos ligandos do G-quadruplex foram já desenvolvidos como agentes anticancerígenos, dada a sua habilidade para modular a atividade transcricional dos oncogenes e a atividade de enzimas relacionadas com o cancro tais como a telomerase. O ligando mais estudado é o TMPyP4, uma porfirina catiónica capaz de inibir a atividade da telomerase e a proliferação de diversas células tumorais, além de reprimir a transcrição de oncogenes tais como *c-myc, c-kit, VEGF* e *KRAS*.

Contudo as cargas positivas do TMPyP4 promoveram interações não-especificas como o ADN de dupla-cadeia e a inibição da proliferação de células saudáveis foi verificada. Por este motivo, o desenvolvimento de novos agentes de estabilização do G-quadruplex é da maior importância. Entre estes, ligandos aromáticos acíclicos com largos anéis capazes de emparelhar com as tétradas, funcionalizados com braços protonáveis capazes de interagir com as cadeias de ADN. Também o desenvolvimento de novas matrizes para o desenho destas moléculas terapêuticas é da maior importância, pelo que a identificação e produção de novas sequências do genoma, capazes de formar G-quadruplex, deve ser estimulada.

Neste trabalho, foi realizada a biossíntese de uma nova sequência capaz de formar G-quadruplex, denominada 58Sγ3. Esta é uma sequência de 58 pares de bases, encontrada num fragmento da região de troca Sγ3 de uma imunoglobulina murina, região esta que consiste em repetições degeneradas de guaninas, com um tamanho entre 2 e 10 mil bases e é necessária para o processo de recombinação dos anticorpos. A sequência 58Sγ3 foi produzida por uma estratégia que envolveu a clonagem da sequência no plasmídeo pVAX1-*lacZ*, a sua amplificação por replicação autónoma em *Escherichia coli DH5α* e consequente recuperação por lise alcalina. O isolamento e purificação da 58Sγ3 foram efetuados por digestão enzimática do plasmídeo, seguido de separação e purificação do fragmento de restrição por cromatografia de exclusão molecular. Esta estratégia provou ser mais rentável quando comparada ao processo caro e laborioso de síntese química de ADN, contudo mais morosa e com rendimentos baixos.

Após a purificação e isolamento da sequência 58Sγ3, a formação de G-quadruplex e a sua topologia foram estudados por dicroísmo circular. Esta técnica é utilizada para detetar a formação de G-quadruplexes e identificar as topologias paralela e antiparalela adotadas. Os resultados do dicroísmo circular sugerem que a sequência 58Sγ3 adota uma conformação paralela em solução de 500 mM KCl. Estudos com a porfirina TMPyP4, conhecida por induzir a formação de G-quadruplex, indicaram o mesmo resultado. Seguidamente, quatro ligandos derivados de naftaleno e quinolina foram testados para o reconhecimento e estabilização da estrutura do G-quadruplex. Foram efetuados ensaios de estabilização térmica no dicroísmo circular, nos quais se determinaram as temperaturas de fusão do G-quadruplex com e sem ligando. Recorrendo ao ensaio de PCR-stop também foi testada a indução e estabilização do G-quadruplex em sistemas biológicos, além da capacidade dos ligandos de inibir a enzima *Taq* ADN polimerase. O mesmo ensaio foi efetuado com uma sequência mutada por forma a descriminar se os ligandos reconheciam efetivamente a estrutura do G-quadruplex, ou se a inibição da enzima *Taq* era devida a interações não específicas. Por último, a afinidade e especificidade dos ligandos para o G-quadruplex em relação ao ADN dupla-cadeia foi avaliada pelo ensaio G4-FID, que usa um fluoróforo cuja emissão de fluorescência é multiplicada quando ligado ao ADN, mas quando desligado por competição com um ligando, perde-a por completo. Os três ligandos derivados de naftaleno demonstraram resultados promissores na estabilização do G-quadruplex da sequência 58Sγ3, sendo por isso moléculas promissoras para o desenvolvimento de agentes terapêuticos.

Abstract

In addition to the Watson and Crick B-form duplex DNA, G-quadruplexes are four-stranded DNA structures formed *in vivo* by the self-assembly of guanine-rich sequences. These can be formed by one, two or four separate strands of DNA and present a diversity of topologies, defined by the strand orientation, loop size and sequence. G-quadruplexes can be found in telomeres, immunoglobulin switch regions and gene promoter regions. The biological relevant location on the genome makes these high-order structures an attractive target for drug design and the development of highly specific ligands that bind and stabilize G-quadruplex with therapeutic activity. Herein, the biosynthesis of a novel G-rich quadruplex-forming DNA sequence 58Sγ3 is described by plasmid amplification. The recovery and purification of 58Sγ3 oligonucleotide using size-exclusion chromatography is presented. The G-quadruplex formation is promoted and its topology is determined by circular dichroism. The stabilization of the G-quadruplex structure with quinoline and naphthalene-based derivatives is studied using melting analysis, G4-FID and PCR-stop assays. The results suggest that 58Sγ3 folds into a parallel-stranded G-quadruplex structure in 500 mM KCl buffer and that naphthalene-based ligands bind and stabilize the G-quadruplex structure. The ligands are also found to be quadruplex-specific over duplex DNA and inhibit *Taq* DNA polymerase. This work provides evidence for G-quadruplex formation within the immunoglobulin switch regions. Furthermore, it is suggested that the novel ligands here reported act as potent specific G-quadruplex binders and may also potentially be used to inhibit genes transcription in tumor cells.

Keywords

G-quadruplex; plasmid DNA; biosynthesis; size-exclusion chromatography; quadruplex-ligands; circular dichroism; PCR-stop assay; G4-FID assay.

Table of Contents

CHAPTER 1 - INTRODUCTION 1

1.1 G-quadruplex 2

1.1.1 G-quadruplex sequence 4

1.1.2 G-quadruplex structure 5

1.1.3 G-quadruplex structural polymorphism 6

1.1.4 Metal ion coordination 11

1.2 Biological relevance of G-quadruplexes 12

1.2.1 Telomeres 13

1.2.2 Gene promoters 15

1.2.3 Immunoglobulin switch regions 18

1.3 G-quadruplex as a therapeutic target 19

1.4 Methods for studying G-quadruplex structure and ligand-binding 22

1.4.1 Methods for G-quadruplex structure determination 22

1.4.2 Methods for studying G-quadruplex—interactions 24

1.5 Plasmid technology 28

CHAPTER 2 – AIMS OF WORK 31

CHAPTER 3 – MATERIALS AND METHODS 33

3.1 Materials 33

3.2 Methods 33

3.2.1 Molecular cloning 33

3.2.1.1 Preparation of competent *E. coli* DH5α cells 33

3.2.1.2 pVAX1-*lacZ* production in *E. coli* DH5α 33

3.2.1.3 Plasmid purification using NZYTech NZYMiniprep kit 34

3.2.1.4 pVAX1-*lacZ* digestion 34

3.2.1.5 Oligonucleotides preparation 34

3.2.1.6 Ligation 35

3.2.1.7 Transformation 35

3.2.1.8 Confirmation of clones by automated sequencing 36

3.2.1.9 Preparation of *E. coli DH5a* cell banks for pVAX-G4 36

3.2.2 pVAX-G4 plasmid production and insert purification................................ 37

3.2.2.1 pVAX-G4 production 37

3.2.2.2 Plasmid recovery and purification... 37

3.2.2.3 Agarose gel electrophoresis 38

3.2.2.4 pVAX-G4 digestion.. 38

3.2.2.5 Size-exclusion Chromatography (SEC) 38

3.2.3 G-quadruplex characterization and interaction studies 38

3.2.3.1 Circular dichroism (CD) studies 38

3.2.3.2 CD melting 39

3.2.3.3 PCR stop assay 39

3.2.3.4 Fluorescence intercalator displacement assay (G4-FID)...................... 40

CHAPTER 4 – RESULTS AND DISCUSSION 41

4.1 Initial studies and G-quadruplex sequence selection 41

4.2 Construction of plasmid pVAX-G4.. 43

4.3 Biosynthesis of 58Sγ3 and purification.. 45

4.4 58Sγ3 G-quadruplex formation and structure assessment 49

4.5 Binding and stabilization of 58Sγ3 quadruplex with ligands 51

4.6 Ligands block *Taq* DNA polymerase in a concentration-dependent manner 55

4.7 Ligands affinity towards G-quadruplex over duplex DNA................................ 58

CHAPTER 5 – CONCLUSIONS AND FUTURE PERSPECTIVES 61

REFERENCES.. 63

List of Figures

Figure 1 - Molecular modelling representations of the three major nucleic acid duplex conformations 1
Figure 2 - Non-duplex DNA structures formed by repeating sequences 2
Figure 3 - The structure of the G-quartet, showing the hydrogen bonding arrangement between the four coplanar guanine bases 3
Figure 4 – Schematic representation of the G-quartet and the G-quadruplex structure.........5
Figure 5 - Schematic illustration of the four major groups of loops 6
Figure 6 - Guanines in anti and syn glycosidic conformations 7
Figure 7 - Four types of G-quartet cores 7
Figure 8 - Schematic representation of a G-quartet core grooves 8
Figure 9 - Schematic structure of unimolecular G-quadruplexes 9
Figure 10 - Schematic structure of bimolecular G-quadruplexes 9
Figure 11 - Schematic structure of a trimolecular G-quadruplex 10
Figure 12 - Schematic structure of tetramolecular G-quadruplexes........................ 10
Figure 13 – Structural polymorphism of G-quadruplex DNA 11
Figure 14 - Schematic view of the shelterin complex and the various associated proteins 13
Figure 15 – Schematic illustration of the telomeric G-quadruplex therapeutic strategy hypothesis........................ 14
Figure 16 - Schematic illustration of the promoter G-quadruplex therapeutic strategy hypothesis 16
Figure 17 - Recombination events in class switch recombination........................ 18
Figure 18 – Structures of the compounds TMPyP4, telomestatin and BRACO-19 20
Figure 19 – Schematic representation of three G-quartet—ligand binding modes 24
Figure 20 – Schematic representation of the G4-FID basis 26
Figure 21 - Schematic representation of the PCR-stop assay 28
Figure 22 – Schematic diagram of pVAX1/lacZ 35
Figure 23 – Schematic representation of the molecular cloning process for the construction of pVAX-G4. 35
Figure 24 - Representation of a 96-well plate and disposition of wells in a G4-FID assay. 40
Figure 25 - Schematic diagram of pPH600........................ 42
Figure 26 – 2% Agarose gel electrophoresis for the confirmation of annealing procedure 44
Figure 27 - Schematic diagram of the constructed plasmid pVAX-G4 45
Figure 28 – Transformed E. coli DH5α growth curve........................ 46
Figure 29 – Analysis of the purity and quality of pVAX-G4 by 1% agarose gel electrophoresis . 46
Figure 30 – Assessment of the effectiveness of the enzymatic double digestion................ 47

Figure 31 - Example of the chromatographic profiles obtained in the purification of 58Sγ3 at 1 mL/min and 0.75 mL/min ... 48
Figure 32 - Example of the chromatographic profiles obtained in the purification of 58Sγ3 at 0.5 mL/min ... 49
Figure 33 - CD titration spectra of 58Sγ3 G-quadruplex at 25 °C with increasing concentrations of KCl ... 50
Figure 34 - CD titration spectra of 58Sγ3 G-quadruplex at 25 °C with increasing concentrations of TMPyP4 ... 51
Figure 35 - Chemical structures of the ligands used in this work L1, L2, L3, L4 and TMPyP4. 52
Figure 36 - CD titration spectrum of 58Sγ3 G-quadruplex at 25 °C in 30 mM phosphate buffer (500 mM KCl) with increasing concentrations of (a) L_1, (b) L_2, (c) L_3 and (d) L_4 ... 52
Figure 37 - CD melting curves of 58Sγ3 G-quadruplex in the absence and in the presence of 8 molar equivalents of ligands ... 53
Figure 38 – CD spectra of the comparison between the folded (25 °C) and unfolded (95 °C) 58Sγ3 G-quadruplex. ... 54
Figure 39 - PCR-stop assay and the effect of the ligands on the 58 bp double-stranded PCR product ... 55
Figure 40 – Fluorescence quantification of the PCR-stop assay gel bands ... 56
Figure 41 – Effect of the ligands on the 58 bp double-stranded PCR product with a control mutated 58Sy3mu oligonucleotide ... 57
Figure 42 - TO displacement plot with the percentage of displacement of TO for each ligand with G-quadruplex and duplex DNA ... 58

Lists of Tables

Table 1 - Primers used for pVAX-G4 sequencing 36

Table 2 - Primers used for pPH600 sequencing 42

Table 3 - QGRS sequences found and the respective length and G-Scores. 43

Table 4 - Sequences of forward and reverse oligonucleotides used for the construction of the 58Sγ3 insert. 44

Table 5 - Thermal Stability of 58Sγ3 with Ligands Measured by CD Melting Experiments. 54

Table 6 - Effect of the ligands on 58Sy3 measured by the PCR-stop assay. 57

List of Acronyms

°C	Degrees Celsius
ATP	Adenosine triphosphate
bcl2	B cell lymphoma 2
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BRAF	v-Raf murine sarcoma viral oncogene homolog B
C	Constant region
CD	Circular dichroism
DMS	Dimethyl sulfate
DNA	Deoxyribonucleic acid
ds	Double-stranded
DSC	Differential scanning calorimetry
DTT	Dithiothreitol
E	Efficiency of energy transfer
EDTA	Ethylenediaminetetraacetic acid
FAM	6-carboxyfluorescein
FRET	Fluorescence resonance energy transfer
G	Gibbs free energy
g/L	Gram per liter
G4-FID	G-quadruplex fluorescent intercalator displacement
gDNA	Genomic DNA
GMP	Guanosine monophosphate
H	Enthalpy
Hif-1α	Hypoxia-inducible factor 1α
hPOT1	Human protection of telomeres 1
hTERC	Human telomerase RNA component
hTERT	Human telomerase reverse transcriptase
IC_{50}	Half maximal inhibitory concentration
ICD	Induced Circular Dichroism
Ig	Immunoglobulins
ITC	Isothermal titration calorimetry
K_B	Binding constant
kb	Kilobases
KRAS	Kirsten rat sarcoma viral oncogene homolog
ln	Linear
M	Molar
m/v	Mass/Volume
mL	Milliliter
mM	Millimolar
n	Stoichiometry
NCBI	National Center for Biotechnology Information
NHE	Nuclease hypersensitive element
nm	Nanometer
NMR	Nuclear magnetic resonance

oc	Open-circular
OD	Optical density
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PDGF-A	Platelet-derived growth factor α polypeptide
QGRS	Quadruplex forming G-rich sequences
RAP1	Repressor activator protein 1
RAP1	Ras-related protein 1
RNA	Ribonucleic acid
RT-PCR	Real-time PCR
S	Heavy chain switch regions
S	Entropy
sc	Super-coiled
SDS	Sodium dodecyl sulfate
SEC	Size-exclusion chromatography
ss	Single-stranded
TAE	Tris-acetate-EDTA
TAMRA	6-carboxytetramethylrhodamine
TEM	Transmission electron microscopy
TIN2	TRF1-interacting nuclear factor 2
T_m	Melting temperature
TO	Thiazole Orange
TPP1	Tripeptidyl peptidase 1
TRAP	Telomeric repeat amplification protocol
TRF1	Telomere repeat binding factor 1
TRF2	Telomere repeat binding factor 2
Tris	Tris(hydroxymethyl)aminomethane
VDJ	Variable region
VEGF	Vascular endothelial growth factor
βTBP	Telomere end-binding protein β
θ	Ellipticity
µg	Microgram
µL	Microliter
µM	Micromolar

Chapter 1

Introduction

DNA is the primary genetic material of all cellular organisms and DNA viruses. Its sequence carries all the genetic information needed to build and maintain an organism. In April 1953, James Watson and Francis Crick published the world-renowned Nature paper proposing a structure model for the right-handed double helix DNA [1]. Watson and Crick suggested that DNA had two helical chains coiled around the same axis running through the complementary base pairs between each chain – adenine (A) with thymine (T) and guanine (G) with cytosine (C) using two or three hydrogen bonds, respectively [1,2]. This structure we know as B-form of DNA is the basic dominant form of DNA *in vivo*. However, DNA molecule is highly polymorphic and can adopt many forms depending on its sequence and environmental conditions [3]. The three major double helical forms are A-, B- and Z-DNA, all made of antiparallel strands (Figure 1).

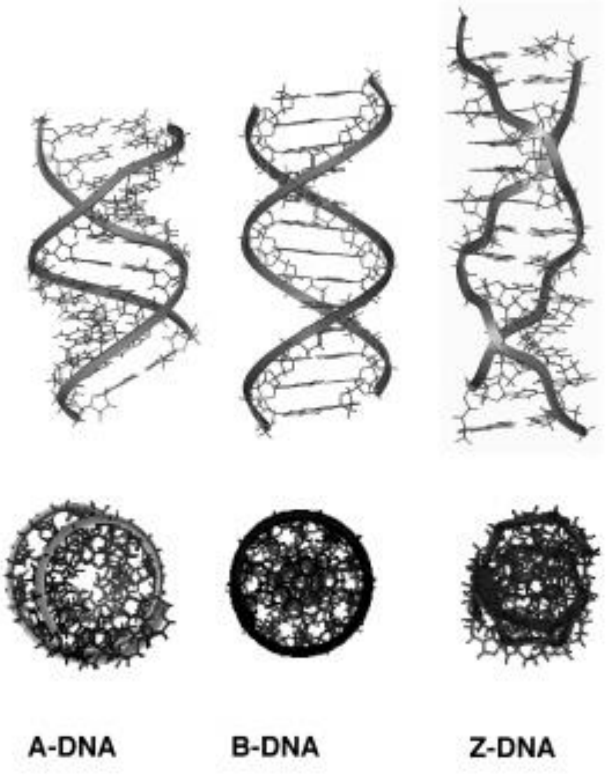


Figure 1 - Molecular modelling representations of the three major nucleic acid duplex conformations. Below: the orthogonal representations. Adapted from [4].

Other non-helical secondary structures exist such as triplexes, bent DNA, cruciforms, nodule DNA, slipped structures (hairpin), sticky DNA and G-quadruplexes (also referred to as tetraplexes, G4 DNA or simply quadruplexes) (Figure 2) [5]. While cruciform and triplex DNA retain the AT and GC Watson-Crick base pairing, G-quadruplex DNA does not involve any GC base pairs and requires the complementary DNA strands to unwind in order to allow the G-rich strands to fold into four-stranded structures [3]. The formation of this non-canonical DNA structures depends on sequence, topology (supercoiling), binding of proteins, DNA modifications, temperature, dehydration and ionic strength [5]. Evidence points to the important biological roles and implications of such diverse and dynamic structures [5].

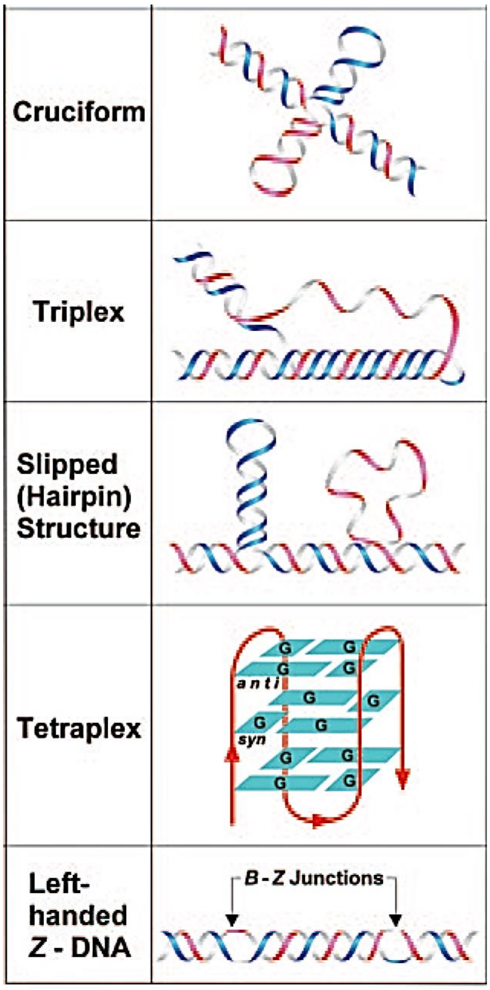


Figure 2 - Non-duplex DNA structures formed by repeating sequences. Adapted from [5].

1.1 G-quadruplex

The guanine quadruplex structure is the most studied non-canonical DNA conformation since the earliest physical studies of nucleic acids dated from the 1960’s [5]. The first observation of the self-assembly of guanylic acid, also known as guanosine monophosphate (GMP), was made in the twentieth century by Ivar Bang (1910). Bang observed the formation of a clear gel at high GMP concentrations (25 mg/mL) and pH 5 [6]. Later on in 1962, Gellert, Lipsett and Davies proposed the core of the guanine quadruplex (Figure 3).

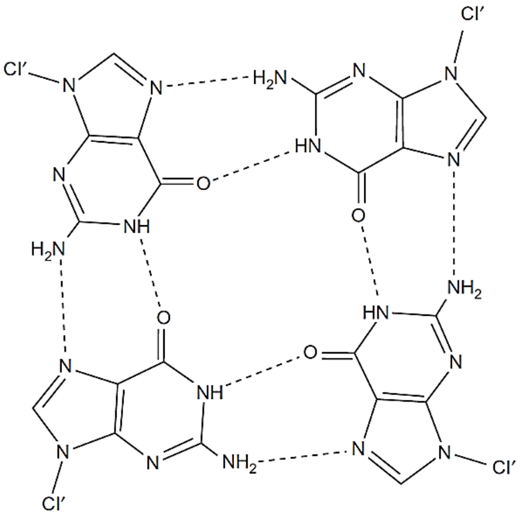


Figure 3 - The structure of the G-quartet, showing the hydrogen bonding arrangement between the four coplanar guanine bases. Taken from [8].

Through X-ray diffraction studies of fibers formed from 3’- and 5’-GMP gels, they found that tetrameric guanine residues arranged in vertically stacked hydrogen-bonded guanine tetrads (later termed as G-quartet or G-tetrad) [7]. Gellert *et al.* also reported the high stability of the structural arrangement provided by the hydrogen-bonding of the four guanine bases, with two hydrogen bonds between each pair involving the N1, N7, O6 and N2 atoms [8]. In 1978, Miles and Frazier showed that guanine tetrads stability depended on a central positive ion that interacts with the oxygen atoms of the guanines [9]. A decade later, new findings pointed to the formation of these four-stranded structures in guanine-rich sequences of immunoglobulin switch regions and in telomeric regions at the end of eukaryotic chromosomes [10,11]. These sequences present discrete runs of guanine tracts (G-tracts), which form compact structural arrangements, termed G-quadruplexes, instead of the continuous helices proposed years before by Watson and Crick [8]. Sen and Gilbert also postulated that the four-stranded structures formed in the immunoglobulin switch regions may have a role in meiosis [11]. In 2001, Schaffitzel *et al.* found strong evidence for the existence of quadruplex structures *in vivo* by staining the ciliate *Stylonychia lemnae* macronuclei with high affinity quadruplex-specific antibodies Sty49 [12]. More recently, G-quadruplexes were found in the mammalian cell nuclei by visualization using immunofluorescent antibodies evidencing its existence *in* vivo [13]. Guanine-rich sequences were also identified in oncogene promoter regions [14–16]. The technological advance in terms of nucleotide sequencing and conclusion of the human genome mapping in 2001 and the use of bioinformatics approaches led to the identification of new sequences with putative quadruplex formation in the DNA and even at RNA level (over 300,000 potential quadruplex-forming sequences) [8,17]. These guanine-rich sequences seem to be implicated in the function and regulation of cellular pathways at the transcriptional and translational levels, which generated much interest and investigation [2]. To date, several studies have been conducted to explore the native state of the quadruplex structures under physiological conditions.

1.1.1 G-quadruplex sequence

The availability of the complete human genome, as well as several other organisms genomes, provided the needed data in which the existence and identity of guanine-rich sequences could be systematically searched [8,18]. Potential quadruplex-forming sequences can be described by the following sequence motif:

$$G_m\ X_n\ G_m\ X_o\ G_m\ X_p\ G_m,$$

where m is the number of guanine residues in each G-tract, usually directly involved in G-tetrad formation [19]. X_n, X_o and X_p can be any combination of residues, including guanines, and are responsible for the loops formation. The G-tracts form the core of the quadruplex, while the loops, positioned on the exterior, help to maintain the overall structure intact. This sequence motif does not assume that all G-tracts are of equal length, and if one of the short G-tracts is longer than the others, the adjacent guanines will be located in the loops [19]. Both G-tracts and loops are usually limited to $3 \leq m \leq 5$ and $1 \leq X \leq 7$, respectively, but deviations to the rule are arguable [8,20]. Using this motif as a quadruplex probe, two distinct research groups conducted bioinformatics analysis using the ENSEMBL genome browser and both found approximatively 375,000 potential quadruplex sequences in the human genome, despite of the statistical and analytical approaches used being different [20,21]. These results indicate that such sequences are highly significant and non-random. Of these ≈375,000 potential quadruplex sequences, ≈223,000 are found in intergenic regions, ≈151,000 within genes and just ≈14,000 within exons [8]. However, the question if these sequences are capable of forming quadruplexes cannot be answered since the experimental data available on structure and stability is insufficient. The search of sequence similarities and distribution of those sequences by clusters with similar tertiary folds is one way to address this problem [21].

A number of informatics resources available online to search databases of quadruplex motifs and to find quadruplex-forming regions within a determined DNA sequence were developed over the years. Generally these search programs use the motif described above as a quadruplex probe with the same limits described. Of the several tools developed, only a few are still available and online. One of those programs is QGRS Mapper, developed by Paramjeet Bagga and his collaborators in 2006 [18]. The main goal of QGRS Mapper is to identify the presence of putative quadruplex forming G-rich sequences (QGRS) in any NCBI nucleotide sequence identified or user-provided sequence [18]. Using a scoring method called G-score, the program then returns the likelihood of the evaluated sequence to form a stable G-quadruplex. Higher scores denote better candidates for quadruplex-formation. The maximum possible G-score, using the default highest QGRS length of 30, is 105 [18]. The scoring method uses three principles based on previous studies: i) shorter loops are more common than longer ones; ii) G-quadruplex loops tend to be approximately equal in size; iii) sequences with greater number of G-quartets form more stable quadruplexes [18].

1.1.2 G-quadruplex structure

As mentioned before, Gellert, Lipsett and Davies introduced the core motif of all G-quadruplex structures, the G-quartet. Each quartet is composed of four guanines, arranged in a rotationally symmetric manner, connected by Hoogsteen hydrogen-bonds in which each guanine base forms two hydrogen bonds with its neighbors, from N1-O6 and N2-N7 (Figure 4) [2,8].

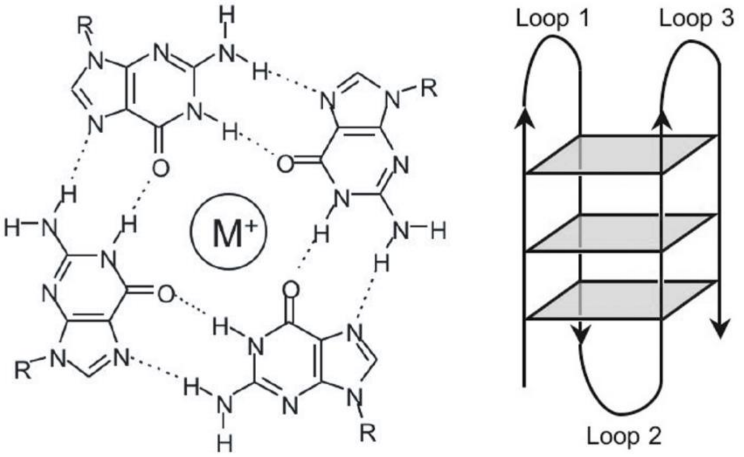


Figure 4 – Schematic representation of the G-quartet and the G-quadruplex structure. Left: hydrogen bond pattern in a G-quartet. A monovalent cation occupies the central position. Right: Schematic diagram of the vertically stacked G-quartets in the G-quadruplex structure. Taken from [20].

The fact that the G-quartets have two hydrogen bonds per base, contrarily to 1 or 1.5 bonds per base in duplex DNA Watson-Crick base pairing (A-T and G-C, respectively), suggests that the G-quadruplex structure is more stable form and presents higher melting temperatures than duplex DNA [22]. The association of guanines N7 atom in the assembly of G-quartets protects them from dimethyl sulfate (DMS) methylation. This unique feature is used to chemically discriminate G-quadruplex formation from other DNA structures (DMS methylation protection assay) [23]. These square planar structures are stabilized by monovalent cations such as potassium (K^+) and sodium (Na^+) and to a lesser extent ammonium (NH_4^+), which interact with the central electronegative carbonyl O6 atoms of the G-quartet core [2]. The most stable G-quadruplex structure is formed by the intercalation of K^+ between two adjacent G-quartets [24]. In fact, the G-quadruplex structure formed in the presence of K^+ is considered to be biologically more relevant due its higher intracellular concentration (≈140 mM) [25]. X-ray crystallography and nuclear magnetic resonance (NMR) studies of short oligonucleotide quadruplex structures provided a reliable set of hydrogen bond distances and angles. The hydrogen-bonds distance between nitrogen-oxygen in G-quartets ranges from 2.7-3.0 Å [8]. These G-quartets have large π-surfaces, hence they tend to stack on each other by π-π stacking (2-4 G-quartets) and enable cations to intercalate between them [2,26]. The stacked G-quartets overlap at a distance of 3.3 Å and are joined together by the sugar-phosphate backbone [27]. The bases that do not participate in the assembly of the G-quartets form loops with different lengths and sequence so that a variety of topologies can be formed (Figure 5).

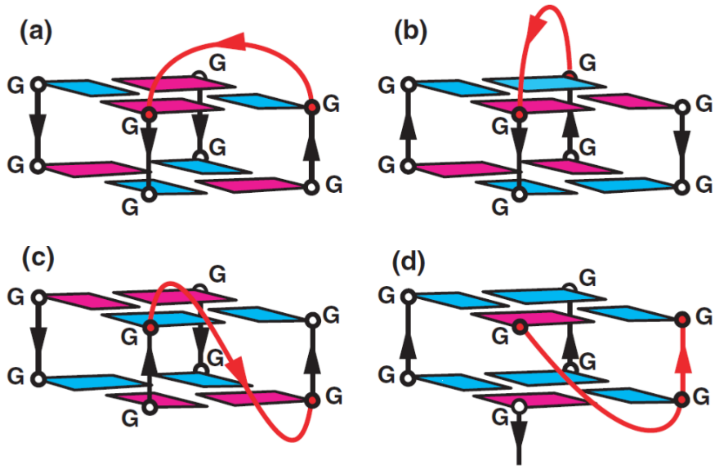


Figure 5 - Schematic illustration of the four major groups of loops. (a) Edge-wise, (b) diagonal, (c) double-chain reversal or propeller and (d) V-shaped loops. The loops connect individual strands bridging two G-tetrad planes. Color coding for illustration is as follows: *anti* guanines in blue and *syn* guanines in magenta. G-rich strands in black and connecting loops in red. Taken from [28].

The loop base composition and length determines the structural conformation and stability [28]. Depending on the size and sequence, the loops can be classified into four major groups: i) Edge-wise or lateral loops which connect two adjacent anti-parallel strands, generally composed of two or more residues (Figure 5a); ii) Diagonal loops which connect two opposing anti-parallel strands, generally composed of three or more residues (Figure 5b); iii) Double-chain reversal or propeller loops which connect adjacent parallel strands, and can be as small as one and as large as six or more residues (Figure 5c); iv) V-shaped loops which connect two corners of a G-quartet with a missing support column (Figure 5d) [28].

1.1.3 G-quadruplex structural polymorphism

G-quadruplexes display a wide variety of topologies, owing to the strand stoichiometry, polarity and orientation (parallel or antiparallel), additionally to the loop size and sequence as stated above [19]. The identity of the cation also contributes to the structural polymorphism of G-quadruplex structures. Regarding the strand stoichiometry, G-quadruplexes may be unimolecular, bimolecular and tetramolecular, whether it's formed by one, two or four separate strands of DNA or RNA, respectively. In principle, three strand arrangements are possible but have yet to be validated. Both the bimolecular and tetramolecular structures are intermolecular quadruplexes while the unimolecular form is a intramolecular quadruplex [19]. Guanine-rich sequences with potential to fold into intramolecular quadruplexes are comprised of four consecutive runs of guanines, separated by three loop regions of different lengths and sequences, while bimolecular quadruplexes are formed by sequences with two G-tracts [19]. On its turn, tetramolecular quadruplexes form from single-repeat G-tracts containing sequences [29]. Since there is directionality in the strands, generally described from 5' end to 3' end, the different strands that constitute the quadruplex may have different polarities. The adjacent strands can be parallel or antiparallel depending on the conformation of the guanine glycosidic torsion angles (Figure 6).

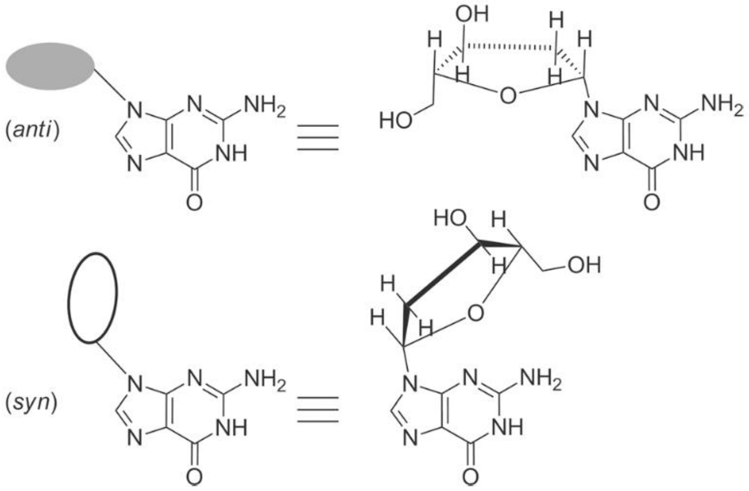


Figure 6 - Guanines in *anti* and *syn* glycosidic conformations. Taken from [26].

Parallel G-quadruplexes have all the guanines in *anti* conformations, being all the strands parallel to each other (Figure 7a); antiparallel G-quadruplexes have both *syn* and *anti* guanines, being at least one of the strands antiparallel to the others (Figure 7b-d) [26]. There are four types of G-quartet core configuration: i) All the strands in the same direction with all the guanines in *anti* or *syn* conformation (parallel G-quartet core, Figure 7a); ii) Three strands in one direction and the fourth in the opposite direction, where the stacked G-quartets adopt *anti-anti-anti-syn* and *syn-syn-syn-anti* alignments ((3+1) G-quartet core, also called hybrid/mixed core, Figure 7b); iii) Two adjacent strands in the same direction and the two other in opposite direction, where the stacked G-quartets adopt a *syn-syn-anti-anti* alignment (type 1 antiparallel G-quartet core, Figure 7c); iv) Alternating antiparallel strands with the diagonally opposite legs in the same direction, where the stacked G-quartets adopt an *anti-syn-anti-syn* alignment (type 2 antiparallel G-quartet core, Figure 7d) [26,30].

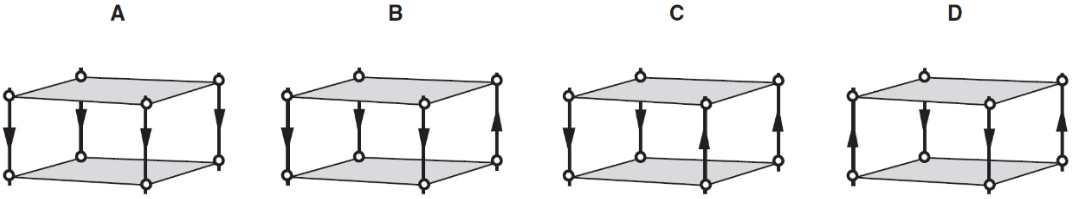


Figure 7 - Four types of G-quartet cores. (a) Parallel G-quartet core, (b) (3+1) G-quartet core, (c) antiparallel G-quartet core (up-up-down-down) and (d) antiparallel G-quartet core (up-down-up-down). Taken from [30].

One feature that arises from the guanine glycosidic torsion angles and differs significantly from duplex DNA is the number of grooves [16,26]. While duplex B-DNA has one minor and one major groove, in G-quartets there are four grooves with three different dimensions: wide, medium and narrow (Figure 8). A medium groove is formed by two adjacent guanines in an *anti* or a *syn* configuration; a narrow groove is formed by an *anti* and a *syn* guanine if the *syn* guanine is perpendicular to the *anti* guanine; finally, a wide groove is formed if the *anti* guanine is perpendicular to the *syn* guanine [16,26]. Some restrictions apply to adjacent guanines involved in the same G-quartet. If they are on parallel strands, they must have the same glycosidic torsion angles, and contrariwise if they are on antiparallel strands they must have opposite glycosidic torsion angles [16].

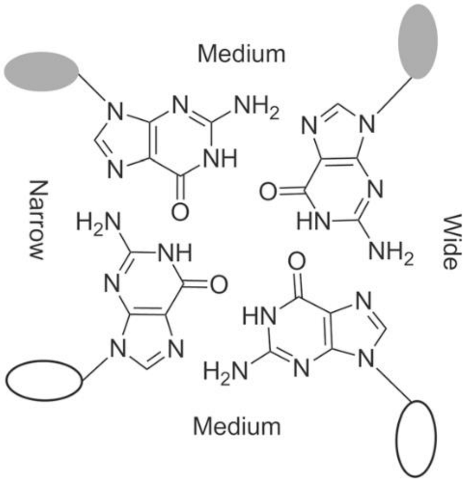


Figure 8 - Schematic representation of a G-quartet core grooves. Three types of grooves can be found: narrow, medium and wide. Taken from [26].

Putting together all the possible variations mentioned above along with the loop types, several different topologies were already determined and described in the literature.

A. Unimolecular G-quadruplex

G-quadruplexes formed by one strand display a variety of structures, as it can adopt three types of loops, namely the diagonal loop, the lateral loop, and the external loop [26]. Among them, the chair-type, with a folding pattern of antiparallel-stranded quadruplex and three lateral loop types (Figure 9a). The second model structure is the basket-type, which is also an antiparallel-stranded quadruplex and has one central diagonal loop and two edgewise loops (Figure 9b). Another model is the propeller-type, which is a parallel-stranded quadruplex with three double-chain reversal loops (Figure 9c). Furthermore, there's hybrid-types 1 and 2, a mixture of antiparallel/parallel-stranded structures formed by two lateral loops and one propeller loop with different arrangements (Figure 9d-e) [30]. The structural polymorphism was demonstrated by Webba da Silva who predicts that, with different loop length and sequence, unimolecular quadruplexes could give rise to 26 different topologies [31]. Recently, a NMR study was able to determine a new folding of the sequence d[$G_3ATG_3ACACAG_4ACG_3$] into a intramolecular antiparallel (3+1) G-quadruplex exhibiting three stacked G-quartets connected with the three types of loops possible: propeller, diagonal and edgewise loops of different lengths [32].

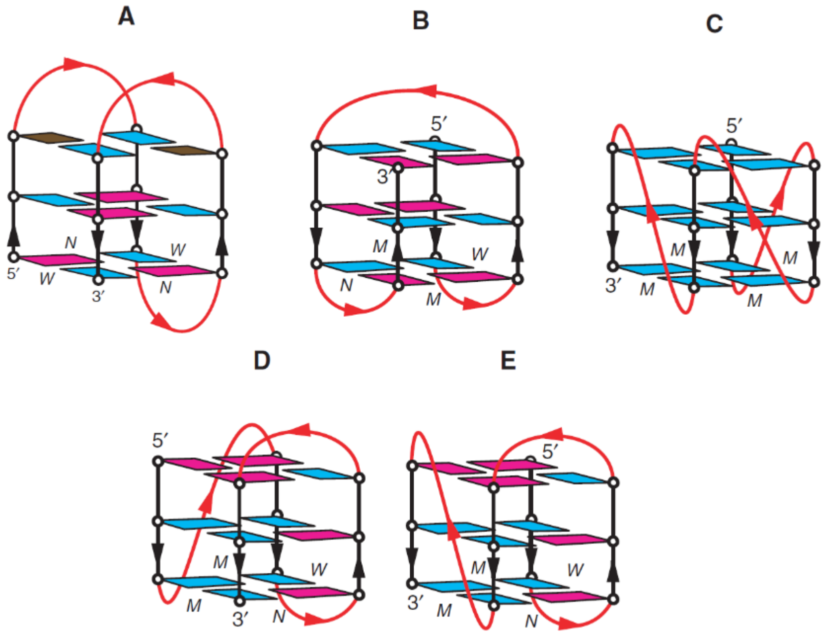


Figure 9 - Schematic structure of unimolecular G-quadruplexes. (a) Chair-type form G-quadruplex formed by variant human telomeric sequence d[A(GGGCTA)$_3$GGG] in K$^+$ solution; (b) Basket-type form observed for d[A(GGGTTA)$_3$GGG] in Na$^+$ solution; (c) Propeller-type form observed for d[A(GGGTTA)$_3$GGG] in a K$^+$-containing crystal; (d) (3 + 1) Form 1 observed for d[TA(GGGTTA)$_3$GGG] in K$^+$ solution; (e) (3 + 1) Form 2 observed for d[TA(GGGTTA)$_3$GGGTT] in K$^+$ solution. Adapted from [30].

B. Bimolecular G-quadruplex

The association of two G-rich DNA strands can also display a variety of topologies. The association leads to the formation of lateral-looped (Figure 10a-b), diagonal-looped (10c) and external-looped G-quadruplexes (Figure 10d), whether these contain two lateral (in one or opposite directions), two diagonal or two double-chain reversal loops, respectively [30]. The first three structures are antiparallel-stranded while the latter is parallel-stranded. There is also an unusual (3+1) mixed type bimolecular quadruplex observed for the three-repeat human telomeric sequence d[G$_3$(T$_2$AG$_3$)$_2$T] and the single-repeat human telomeric sequence d[TAGGGT] in Na$^+$ solution [26,30].

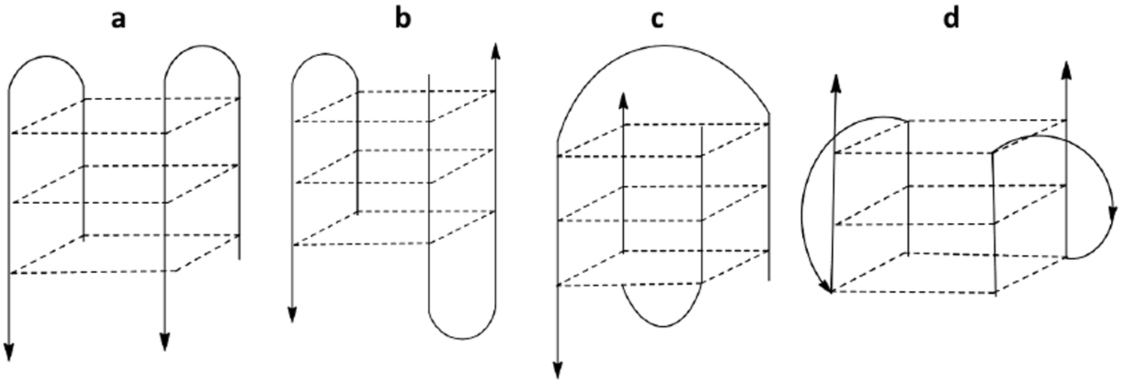


Figure 10 - Schematic structure of bimolecular G-quadruplexes. Taken from [33].

C. Trimolecular G-quadruplex

Despite being rare, in 2012 the Mergny group demonstrated the formation of an unprecedented trimolecular G-quadruplex structure [29]. Although it was a guided assembly of the G-quadruplex, this work proves the putative formation of three stranded quadruplexes under some special conditions. The structure contained some distinctive features as the fact that it had two duplex tails established in the upper and lower ends of the structure (Figure 11). Three single T1, T2 and T3 strands were used as guide strands to position the G-rich tracts in Li^+ solution, and then the structure formation was induced by adding Na^+ [29].

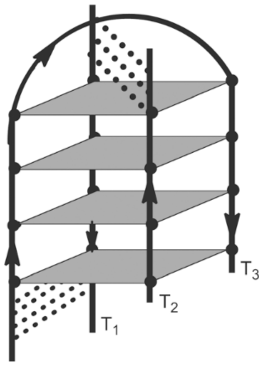


Figure 11 - Schematic structure of a trimolecular G-quadruplex. Two duplex tails are established at the upper and lower ends of the structure. T1: d[TGAGATGTACTATGAGGGGGTGTCATGGTAGAAGT]; T2: d[GGGGGTCATAGTACATCTCA]; T3: d[ACTTCTACCATGACAGGGGGTTTGGGGG]. Taken from [26].

D. Tetramolecular G-quadruplex

On its turn, the formation of tetramolecular quadruplexes seems much easier (Figure 12a). The crystal structure of hexanucleotide d[TG$_4$T] in Na^+ solution exhibit all four strands in a parallel-stranded structure, with all guanines in *anti* configuration, medium-size grooves and inexistent loops [26]. However, due to the glycosidic torsion angles of the guanines involved in the G-quartets formation and the different orientations of each independent strand, more complex and unusual structures are possible. For example, the tetranucleotide d[GGGT] was shown to adopt an interlocked dimeric G-quadruplex structure containing two parallel tetramolecular quadruplexes connected by an extra G-quartet (Figure 12b) [26].

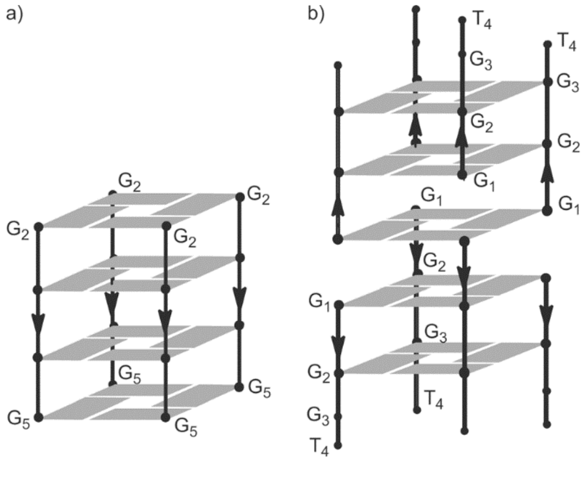
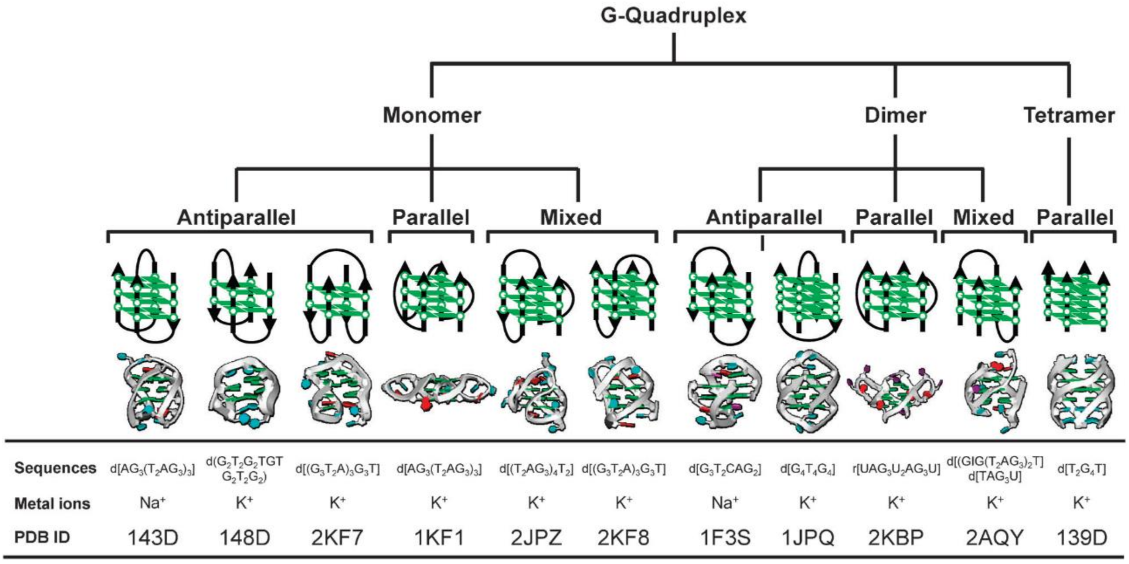


Figure 12 - Schematic structure of tetramolecular G-quadruplexes. (a) Tetrameric parallel-stranded G-quadruplex observed for the single-repeat human telomeric sequences d[TTAGGG] and d[TTAGGGT] in K^+ solution. (b) Interlocked dimeric G-quadruplex of d[GGGT] in K^+ solution. Adapted from [26].

The overall structures, their sequences and Protein Data Bank (PDB) accession numbers of crystallography and NMR determined structures are listed in Figure 13.



Sequences	$\mathrm{d[AG_3(T_2AG_3)_3]}$	$\mathrm{d(G_2T_2G_2TGT\ G_2T_2G_2)}$	$\mathrm{d[(G_3T_2A)_3G_3T]}$	$\mathrm{d[AG_3(T_2AG_3)_3]}$	$\mathrm{d[(T_2AG_3)_4T_2]}$	$\mathrm{d[(G_3T_2A)_3G_3T]}$	$\mathrm{d[G_3T_2CAG_2]}$	$\mathrm{d[G_4T_4G_4]}$	$\mathrm{r[UAG_3U_2AG_3U]}$	$\mathrm{d[(GIG(T_2AG_3)_2T]}$ $\mathrm{d[TAG_3U]}$	$\mathrm{d[T_2G_4T]}$
Metal ions	Na^+	K^+	K^+	K^+	K^+	K^+	Na^+	K^+	K^+	K^+	K^+
PDB ID	143D	148D	2KF7	1KF1	2JPZ	2KF8	1F3S	1JPQ	2KBP	2AQY	139D

Figure 13 – Structural polymorphism of G-quadruplex DNA. Taken from [34].

1.1.4 Metal ion coordination

The stability and conformation of G-quadruplexes depend on cations. As stated before, the G-quartets central atoms O6 participate in the coordination of the cations. This feature is itself a difference from the duplex DNA, which is stabilized by magnesium (Mg^{2+}) ion while the G-quartet is not stabilized by such ion. In addition to the stabilization role of the cations, these ions help reducing the electronegative charges repulsion of each G-quartet [8,28]. The structural features of the G-quartets is optimal for the coordination of metal cations within its central cavities. G-quadruplex formation is induced by the presence of K^+, Na^+ and Rb^+ while cations such as Li^+ and Cs^+ have a limited ability to induce the formation. However, G-quadruplexes prefer K^+ over Na^+, reflecting the much greater energetic penalty for Na^+ dehydration [28]. Also, K^+ produces more stable quadruplexes than Na^+. This fact is due to the size of the cation, K^+ ion is larger than Na^+ and forms a much more compact unit upon stacking between two stacked G-quartets [19]. All together, these facts suggest that the formation of G-quadruplexes depends on the size and desolvation energy cost of the cation. From studies of the ions ability to induce gel formation in guanosine nucleotides it was possible to deduce the following order: $Sr^{2+}>K^+>>Na^+>Rb^+\sim Ba^{2+}>NH_4^+>Ca^{2+}> Mg^{2+}\sim Cs^+>>Li^+$ [16,35].

One interesting behavior of G-quadruplexes is the topology transition observed under different ionic conditions. The same sequence can adopt different G-quadruplex conformations in K^+ and Na^+. For example, based on NMR and crystallographic studies of the human telomeric DNA d[AGGG(TTAGGG)$_3$], Na^+ seems to favor an antiparallel conformation while K^+ induces a parallel or hybrid form [36,37]. Interestingly, RNA-quadruplexes formation is generally independent of the cation identity and bimolecular quadruplexes are not affected by the nature of the cation equally to the unimolecular quadruplexes [38,39].

1.2 Biological relevance of G-quadruplexes

Guanine-rich sequences with putative G-quadruplex formation are widely dispersed in eukaryotic and prokaryotic genomes. A number of critical regions in the eukaryotic genome have been reported to adopt G-quadruplex structures. These include promoter regions of oncogenes such as *c-myc* and *c-kit*, both short micro-satellite and longer mini-satellite repeats, ribosomal DNAs, as well as telomeres in eukaryotes and immunoglobulin heavy chain switch regions of higher vertebrates [16]. G-rich sequences were also found in the retinoblastoma susceptibility gene and upstream of the insulin gene [16]. These guanine-rich sequences have the potential to influence the gene metabolism processes, as well as key biological processes such as DNA replication and recombination, transcriptional regulation and genome stability [26]. The visualization of G-quadruplex structures *in vivo* was already reported. Besides the work with the ciliate *Stylonychia lemnae* described above, recent work reported by Biffi *et al.* described a monoclonal single chain antibody BG4 with high affinity and specificity for intramolecular G-quadruplex structures [13]. This work reported crucial evidence for the formation of G-quadruplex DNA within mammalian cells genomes, in a cellular context, the cell-cycle dependence of the structures (the occurrence of G-quadruplexes was maxed during the S phase) and provided an important basis to help understand the biological roles of G-quadruplex structures [13]. Additional works using monoclonal antibodies followed, strengthening the evidence of the G-quadruplex formation *in vivo* [40,41].

The existence of natural proteins that recognize G-quadruplexes *in vivo*, provided important insights in the location of G-quadruplex structures in the genome and reinforces the biological relevance of such sequences [28]. Such proteins bind, promote or disrupt the quadruplex formation, and over 30 proteins have been reported so far [2]. For instance, both telomere end-binding protein (βTBP) in *Oxytricha* and repressor activator protein 1 (RAP1) in *Saccharomyces cerevisiae* promote intermolecular G-quadruplex formation [28]. The MutSα protein, which is involved in mismatch repair, targets G-quadruplex DNA and promotes synapsis of transcriptionally activated immunoglobulin switch regions [28]. In addition, POT1 binding to DNA disrupts G-quadruplex formation at the telomeric G-rich overhangs, promoting telomerase activity [28]. Furthermore, the *Escherichia coli* RecQ protein can unwind G-quadruplex DNA and is essential for maintaining genomic stability of many organisms, being the RecQ DNA helicase family conserved from *E. coli* to humans [23,28]. Genome integrity is essential in order to maintain the normal cellular function, and malfunctioning DNA replication or repair processes can lead to genetic instability and disease [13]. In humans, the helicase-homologue protein ATRX was shown to bind G-rich sequences in both telomeres and euchromatin. It was proposed that such protein was able to modify the epigenetic state of G-quadruplex sites and to resolve its structure [13].

1.2.1 Telomeres

The telomere is a specialized nucleoprotein complex (termed *shelterin*) existent at the ends of all eukaryotic chromosomes [8,28]. The human telomere consists of tandem repeats of the motif d[TTAGGG] and is associated with a variety of telomeric proteins and other DNA-repair and damage response proteins [42]. The telomere function is to stabilize the termini of linear eukaryotic chromosomes, forming special T-loop-like structures, and protecting the chromosomes from unwanted recombination and degradation, while providing sites for recombination events and transcriptional silencing [28,42]. Telomeres are thought to play a critical role in cellular aging and cancer. Human telomeric DNA is typically 5-15 kilobases (kb) long duplex DNA with a single-stranded 3' overhang of 150-250 bases at the 3' extreme end. These G-rich single-stranded overhangs are attractive sites for potential G-quadruplex formation. The length of the duplex portion of the telomeres decreases progressively after each cell division cycle in somatic cells as a consequence of the end-replication effect. On the other hand, the single-stranded 3' overhangs can be elongated by telomerase, an enzyme with reverse transcriptase activity, which is expressed in the majority of cancer cells (80-85%) and primary tumors, thus maintaining the telomere-length homeostasis [8,28,42]. Telomerase is a ribonucleoprotein which contains an RNA template (subunit hTERC) from which its reverse transcriptase subunit hTERT copies and adds TTAGGG repeats to the single-stranded 3' overhangs of the telomeres [8,42]. In normal somatic cells, when telomeres reach critical short lengths (the Hayflick limit of about 40 cell divisions), cells enter irreversible p53 and Rb-dependent replicative senescence, and ultimately apoptosis. In cancer cells, where telomerase is expressed, it maintains the telomere length acting as a tumor promoter and helping the cells bypassing apoptosis and achieving cellular immortality [42]. The 3' single-stranded overhang is protected and maintained by several copies of hPOT1 (protection of telomeres 1), a single-stranded binding protein which interacts with other proteins of the *shelterin* complex to regulate telomerase-mediated telomere elongation. The *shelterin* complex proteins include the telomere repeat binding factors (TRF1 and TRF2), which are responsible for the T-loop formation; the associate-proteins (RAP1, TPP1, and TIN2) on its turn, are responsible for mediating the interaction between TRF1 and TRF2, as well as POT1 and TPP1 (Figure 14).

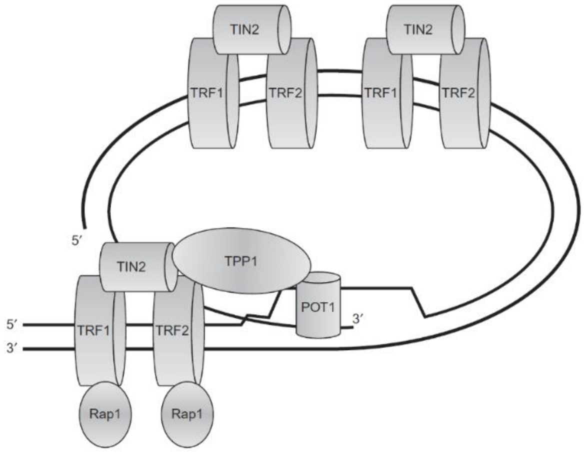


Figure 14 - Schematic view of the *shelterin* complex and the various associated proteins. Taken from [8].

T-loop formation involves the folding back of the single-stranded overhang and pairing with a complementary portion of the telomeric double-stranded region [8]. The disruption of telomeric DNA-hPOT1 association leads to quadruplex formation, deprotects telomeres and initiates DNA damage-response mediated cell death [28,42]. Thus, the development of small molecules that bind and stabilize the single strand G-quadruplex, competing with hPOT1 and initiating this response, is a viable and promising anti-cancer therapeutic strategy (Figure 15). Also, the formation of a stable G-quadruplex structure inhibits the activity of telomerase [19]. Therefore, much effort has been placed in the structural characterization of the telomeric G-quadruplex topologies in order to find scaffolds for anti-cancer drug development [28].

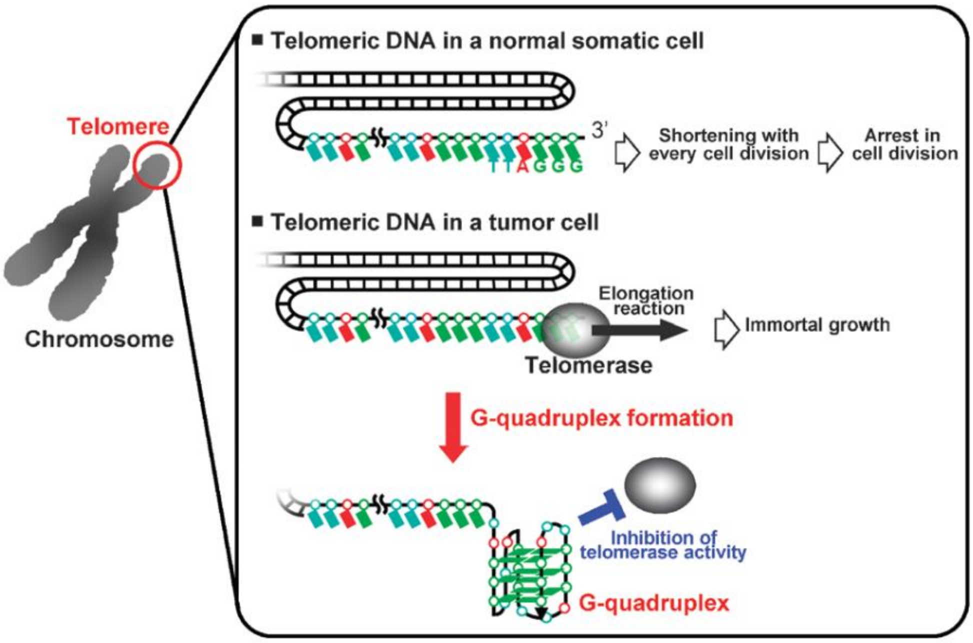

Telomeric DNA in a normal somatic cell
Telomere
3'
TTAGGG
Shortening with every cell division
Arrest in cell division
Chromosome
Telomeric DNA in a tumor cell
Elongation reaction
Immortal growth
Telomerase
G-quadruplex formation
Inhibition of telomerase activity
G-quadruplex

Figure 15 – Schematic illustration of the telomeric G-quadruplex therapeutic strategy hypothesis. The biological roles of telomeric DNA and telomerase are shown; G-quadruplex formation can inhibit telomerase activity. Taken from [34].

The human telomeric G-quadruplex structure is highly polymorphic and to date several structure models were proposed [8]. In 1992, the Patel group proposed a model for the single-repeat d[TTAGGGT) human telomere sequence which formed a parallel-stranded G-quadruplex in K^+ solution [43]. The structure was formed by three G-quartets with all guanines in *anti* conformation. Later on, the two-repeat human telomere sequence G-quadruplex model was proposed and the crystal structure was identified. The crystallography analysis demonstrated that the sequence d[TAGGGTTAGGGT] folds into a bimolecular quadruplex in K^+ solution, with all the strands in a parallel orientation and connected by propeller-type loops [36]. Following NMR studies of the same sequence confirmed the G-quadruplex topology [44]. Other NMR studies, involving the three-repeat human telomere sequence 5'-GGGTTAGGGTTAGGGT-3' demonstrated the formation of a peculiar G-quadruplex structure. This structure contained three strands in the same direction and one in the opposite direction (3+1 G-quadruplex) [45]. Such structure could occur *in vivo* in the T-loop region where the single-stranded overhang pairs with a complementary portion of the telomeric double-stranded region. Such results supported the biological and therapeutic implications of the G-quadruplex formation in the telomeres, and the use of small molecules for its stabilization [45]. The cation-dependent structural polymorphism was demonstrated by studies performed with the four-repeat human telomere sequence $d[AG_3(T_2AG_3)_3]$. In a solution containing Na^+, the sequence folds into a antiparallel G-quadruplex structure with a central diagonal loop and two edgewise loops, as determined by NMR [37]. However, in K^+ solution, a X-ray crystal structure of the same sequence demonstrated the folding of a intramolecular G-quadruplex composed of three G-quartets, with all four strands in parallel joined by three propeller-type loops [36]. Because the intracellular concentration of K^+ (≈140 mM) is higher than that of Na^+ (≈10 mM), the propeller-type G-quadruplex seems to be more physiological relevant. Further studies demonstrated the existence of other G-quadruplex conformations in the human telomeres such as hybrid-type mixed parallel/anti-parallel and basket-type quadruplexes [46–48].

1.2.2 Gene promoters

Apart from telomeric sequences, there are other biological relevant G-rich sequences that have the ability to form G-quadruplex structures, such as the gene promoter regions. The first evidence of the formation of unusual DNA conformations in gene promoters was found by Larsen and Weintraub who demonstrated that chicken ß-globin promoter adopted high order structures *in vivo* [49]. Since then, the use of bioinformatics approaches led to the discovery of putative quadruplex-forming sequences in the regulatory and promoter regions of oncogenes such as *c-myc, c-myb, c-Fos, c-kit, KRAS,* vascular endothelial growth factor (*VEGF*), platelet-derived growth factor α polypeptide (*PDGF-A*), *Rb, RET,* hypoxia-inducible factor 1α (*Hif-1a*), B cell lymphoma 2 (*bcl2*), *and hTERT* [50]. These proto-oncogenes are involved in growth and proliferation processes and their proximal promoter regions contain several G and C-rich regions. The formation of G-quadruplexes in these regions is believed to modulate the genes function and transcriptional activity. Moreover, these genes are important in cell signaling and are involved in a variety of cancers [50]. Therefore, the study of such G-quadruplex structures and the development of targeted-drugs, capable of interacting with the specific G-quadruplex and modulate the transcriptional activity of the associated gene, poses itself as a promising therapeutic strategy. This therapeutic approach is presented in Figure 16 in a broad perspective and represents the formation of G-quadruplex structures in gene promoters in a process that affects the transcriptional activity.

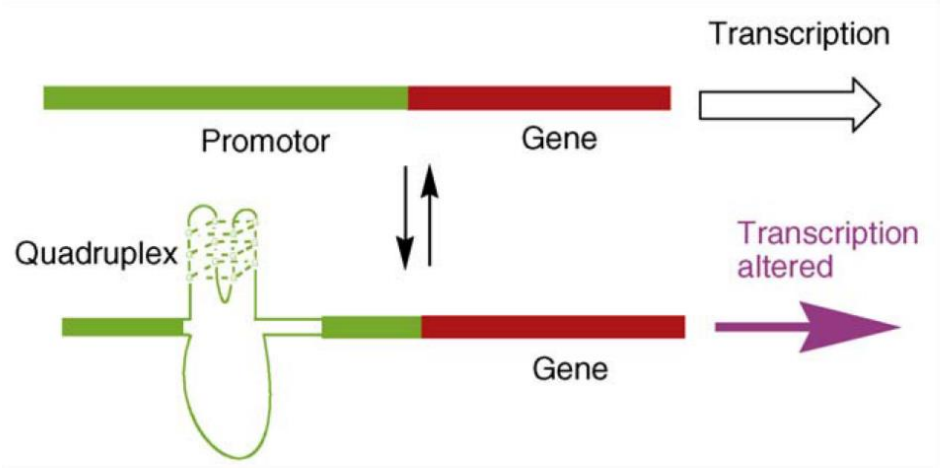


Figure 16 - Schematic illustration of the promoter G-quadruplex therapeutic strategy hypothesis. G-quadruplex forming region in the upstream (promoter) region (in green) of genes (in red). Taken from [42].

Some important oncogene promoter quadruplexes are described in more detail below.

A. *c-myc*

The proto-oncogene *c-myc* encodes a multifunctional transcription factor thought to regulate 10–15% of all cellular genes and is involved in processes such as cell cycle regulation, apoptosis, metabolism, cellular differentiation and cell adhesion [23]. The overexpression of *c-myc* leads to cellular proliferation and inhibition of differentiation and is associated with a variety of human cancers, such as colon, breast, small-cell lung, osteosarcomas, glioblastomas, and myeloid leukemia [51]. The deregulation of the *c-myc* transcription factor is a consequence of gene amplification, translocations, altered ploidy or enhanced transcription owing to upstream signaling abnormalities [50]. Because of this, transcriptional modulation of the *c-myc* oncogene has been an attractive target for anti-cancer drug development. The regulation of *c-myc* transcription is complex and involves multiple promoters (P0, P1, P2, and P3) and seven nuclease hypersensitive elements (NHEs). The nuclear hypersensitivity element III_1 (NHE III_1), located -142 to -145 base pairs (bp) upstream of the P1 promoter of *c-myc*, is a G-rich strand that controls 80-90% of *c-myc* transcription and contains a 27 bp sequence (Pu27) with propensity to fold into a stable G-quadruplex structure [52]. Pu27 is capable of engaging in a slow equilibrium between B-form duplex DNA, single-stranded DNA and G-quadruplex DNA [51]. This 27 bp sequence contains six G-tracts with different lengths and can form a variety of intramolecular G-quadruplex structures depending on the G-tracts combinations [23]. Is worth to notice that for a quadruplex to form, the two complementary sequences must first separate to allow the G-rich strand to fold in G-quadruplex [53]. The first observation of a G-quadruplex formed by *c-myc* promoter in K^+ solution was made by the Simonsson group [14]. The G-quadruplex consisted of three G-quartets bound by two lateral loops and a central diagonal loop. Since Pu27 forms a dynamic mixture of four G-quadruplex loop isomers, smaller sequences derived from Pu27 started being used, in order to determine the physiologically more relevant quadruplex within this region [23]. Two different sequences, Myc-2345 and Myc-1245 (numbers correspond to the G-tract position), were analyzed by NMR and found to form intramolecular propeller-type G-quadruplexes in K^+ solution. The core of the structures was formed by four parallel strands, with all guanines in *anti* conformation and three propeller-type loops [23]. Another example is Myc22-G14T/G23T, in which two guanines have been mutated to thymines, and forms a parallel-stranded G-quadruplex with propeller-type loops [54].

B. *c-kit*

The *c-kit* proto-oncogene encodes a 145-160 kDa membrane bound receptor tyrosine kinase that constitutes a cell signaling system and can stimulate cell proliferation, differentiation, migration and survival [50]. The overexpression or activating mutations of *c-kit* may lead to aberrant function and oncogenic cellular transformations, being involved in a variety of human cancers such as mast cell tumors, germ cell tumors, ovarian carcinomas, malignant melanomas, gastrointestinal stromal tumors, small-cell lung cancer, neuroblastoma, and breast carcinoma [23]. Upstream of the *c-kit* transcription initiation site there's a G-rich strand which is essential for the promoter activity. *c-kit* quadruplexes are therefore attractive targets for anti-cancer drug development. Two G-quadruplex structures have been identified in the G-rich promoter region of *c-kit*, c-kit87up and c-kit21. c-kit87up is a 22 nucleotide sequence, located 87 base pairs upstream of the transcription initiation site, and was found to form a single G-quadruplex structure in K^+ [55]. The G-quadruplex structure formed was very peculiar as an isolated guanine was involved in the G-quartet core formation and it presented four distinct loops, two propeller-type, a single-residue loop and a five-residue loop [55]. This emphasizes the importance of the sequence as non-G-tract residues can participate in the G-quadruplex core formation. On its turn, c-kit21 was found to adopt a variety of conformations and mutations studies need to be employed in order to form a single G-quadruplex [23].

C. *bcl-2*

The human *bcl-2* gene encodes a 25 kDa mitochondrial membrane protein that blocks programmed cell death. The overexpression of *bcl-2* gene occurs in a variety of human cancers, such as B-cell and T-cell lymphomas, breast, cervical, non-small-cell lung, prostate, and colorectal cancers, which reduces the rate of cell death and also interferes with the therapeutic action by resisting apoptosis induced by chemotherapy [56]. Therefore, *bcl-2* gene has been considered as an important target for developing anti-cancer compounds. The human *bcl-2* gene is regulated by two promoters P1 and P2. A 39 bp G-rich sequence (bcl-2Pu39) located in the *bcl-2* P1 promoter plays a significant role in the regulation of *bcl-2* transcription [23]. The sequence contains six G-tracts (5'-AGGGGCGGGCGCGGGAGGAAGGGGGCGGGAGCGGGGCTG-3') with different sizes and was shown to adopt a mixed parallel/antiparallel G-quadruplex intramolecular G-quadruplex conformation [57]. Moreover, the structure contains two edgewise loops and one propeller-type loop and the middle four G-tracts are the ones that generate the predominant G-quadruplex structure (MidG4), since like *c-myc* this sequence can adopt multiple intramolecular G-quadruplexes (5'G4, MidG4, and 3'G4) [23,57].

1.2.3 Immunoglobulin switch regions

Other important genomic regions have the ability to form G-quadruplex structures. Of interest are the regions encoding immunoglobulin heavy chain switch (S) regions of higher vertebrates. These regions are critical for class switch recombination processes of B lymphocytes. B lymphocytes, or B cells, are responsible for the production of different isotypes of immunoglobulins (antibodies) for diverse pathogens [58]. The process that enables the production of immunoglobulins (Ig) to change from one isotype to another, termed class switch recombination, is a region-specific recombination process that bring an expressed variable (VDJ) region to a new constant (C) region during the differentiation of B lymphocytes to plasma cells [11,59]. The isotype switching from IgM to IgG3 and IgA is presented in Figure 17 as an example.

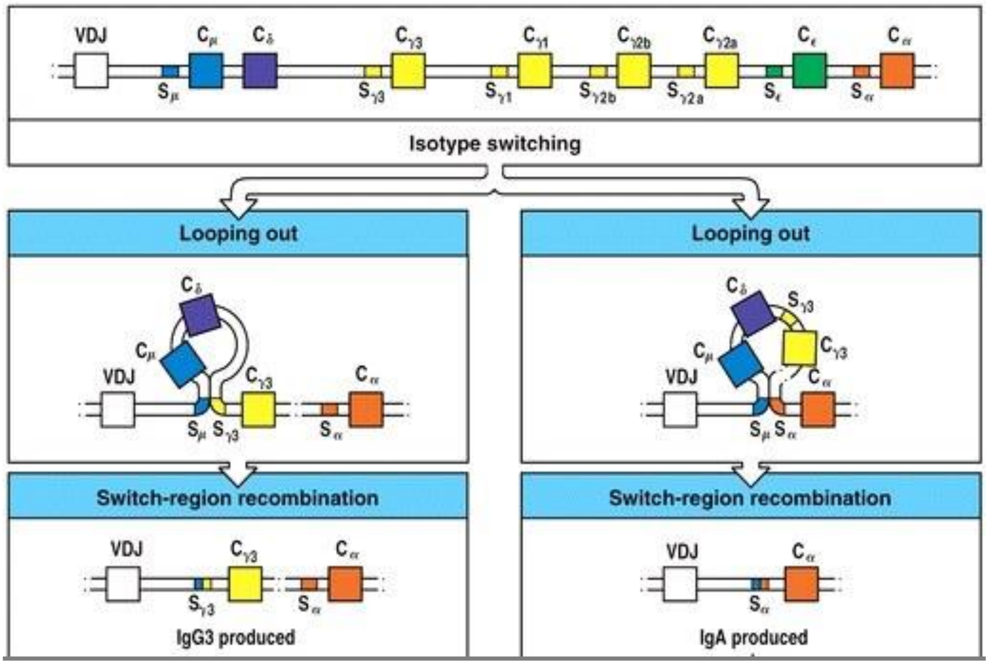

VDJ
C_μ
C_δ
$C_{\gamma3}$
$C_{\gamma1}$
$C_{\gamma2b}$
$C_{\gamma2a}$
C_ϵ
C_α
S_μ
$S_{\gamma3}$
$S_{\gamma1}$
$S_{\gamma2b}$
$S_{\gamma2a}$
S_ϵ
S_α
Isotype switching
Looping out
Looping out
Switch-region recombination
Switch-region recombination
IgG3 produced
IgA produced

Figure 17 - Recombination events in class switch recombination. During isotype switching, a portion of the DNA is looped out as switch regions recombine. A different constant regions is placed downstream of VDJ region. After recombination, the constant region directly downstream of the VDJ region will encode the immunoglobulin isotype produced on the surface of the B cell. Taken from [59].

The VDJ regions form a domain that recognize the antigen, while C region determines how the antigen is removed from the body [58]. During this recombination process numerous kb of DNA between the constant and variable regions are deleted. These processes are essential and if impaired, it can result in immunodeficiency [58]. The S regions of 2 to 10 kb in length, lie upstream of the constant-region genes and contain repetitions of highly degenerate G-rich consensus sequences which are 20-50 bp long [11,58]. Several repetitive motifs occur such as GGGGT, GAGCT and the conserved sequence (G)GGGGAGCTGGGG which is found in Sγ1, Sγ2b and Sγ3 S regions. Two different studies using the same Ig switch region sequences demonstrated the formation of stable G-quadruplex structures, suggesting a role of such structures in recombination events [11,15]. More recently, Maizels group reported the formation of a novel G-quadruplex structure called G-loops by using plasmid vectors containing inserts from murine switch regions [60]. For instance, pPH600, a plasmid containing a 604 bp fragment of the murine Sγ3 switch region, upon *in vitro* transcription of the G-rich regions formed G-loops observed by transmission electron microscopy (TEM) [60]. However, to my best knowledge, no studies regarding the determination of the G-quadruplex structure were performed so far.

1.3 G-quadruplex as a therapeutic target

Due to the biological relevant location of G-quadruplexes throughout the genome and the implication in key processes such as maintenance of chromosomal ends, transcription, translation, DNA replication and recombination, G-quadruplex is a promising and viable therapeutic target. This therapeutic strategy involves the development of selective drug-like small ligands that strongly bind and stabilize the G-quadruplex structures. Over the past years, there has been a remarkable effort on developing G-quadruplex ligands, especially to target human telomeres and oncogene promoters in order to block the action of telomerase and the oncogene transcription, respectively [8,42]. This concept was first validated through the demonstration that the compound 2,6-diamidodianthraquinone was capable of inhibiting the activity of telomerase by interacting with and stabilizing G-quadruplex structures [28]. G-quadruplex structure presents a large π-surface due to the guanine tetrads, approximately twice as large as that found in duplex DNA. Therefore, the majority of quadruplex-ligands have a large aromatic core, with a large π-surface, in order to maximize the π-π interactions [53]. Compounds containing polyaromatic heterocyclic ring systems, such as anthraquinones, acridines, naphthalenes, perylenes and porphyrins, capable of π-stacking interactions, are good candidates [28]. Another feature of these ligands, relates to the fact that G-quadruplexes, like all nucleic acids, have a high negative charge due to the backbone negative phosphate groups. Hence, cationic ligands will generally bind more strongly to the G-quadruplex structure [53]. On the other hand, the inclusion of sidechain specific functional groups, such as protonable side arms, once they participate in the recognition of G-quadruplex via actions in the grooves and enhance the interactions with the structure [53,61]. One important issue in the development of such small ligands, is their specificity for G-quadruplex over duplex DNA. This is a major problem since G-quadruplexes are far outnumbered by the duplex DNA in the cells [2]. The compounds synthesized must promote cell death in cancer cells, while ensuring that the toxicity to normal healthy cells is low.

Until now, some compounds with experimental evidence for G-quadruplex binding and therapeutic activity have been developed such as BRACO19 [62], telomestatin [63], and TMPyP4 [64] (Figure 18).

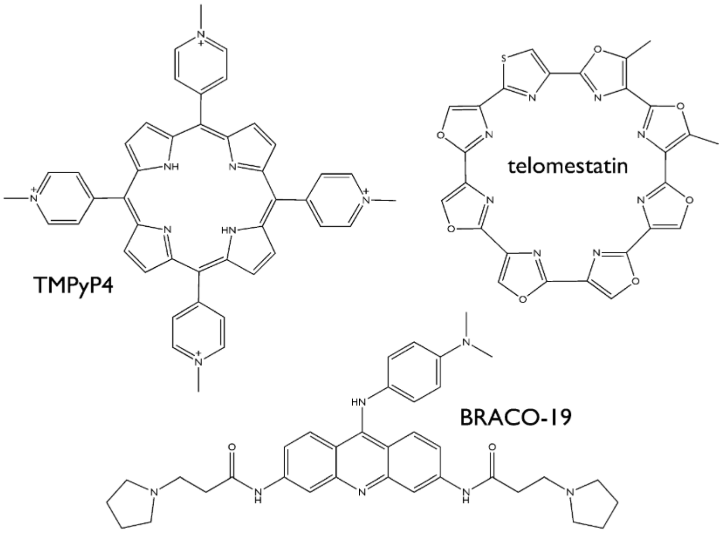


Figure 18 – Structures of the compounds TMPyP4, telomestatin and BRACO-19. Taken from [2].

BRACO19 is a 3,6,9-tri-substituted acridine that appears to directly target the telomeres and displayed the capacity to inhibit the catalytic function of telomerase in human cancer cells and destabilize the telomere complex [65]. BRACO19 induced tumor regression within 7-10 days of the initiation of treatment of prostate cancer cells. Its major limitations are lack of membrane permeability and small therapeutic window [65]. Telomestatin, a natural occurring macrocycle isolated from *Streptomyces anulatus,* is the tightest known G-quadruplex binder and telomerase inhibitor [65]. Telomestatin interacts preferentially with intramolecular G-quadruplex structures and has a 70-fold selectivity over duplex DNA. Telomestatin displayed promising anticancer activity in human cancer cells, within 3–5 weeks at the minimal effective concentration inhibited telomerase activity, reduced telomere length and caused apoptotic cell death in multiple myeloma cell lines [65]. TMPyP4, is a cationic porphyrin designed with attractive properties such as a fused planar ring system, positive charge and appropriate size to stack with the G-tetrads [65]. It has shown to inhibit human telomerase in HeLa cell extracts and to downregulate *c-myc*. However, TMPyP4 has only limited selectivity for G-quadruplex over duplex DNA [65]. Naphthalene-based ligands were also reported as having the necessary features to bind planar substrates such as nucleotides, mainly through stacking interactions [66]. Potent G-quadruplex binders containing naphthalene rings were reported by Neidle *et al.*, which were found to promote a dose dependent cell arrest of mutated *c-kit* cell lines [8,67]. However, some non-specific interactions were reported as well [67]. Despite the number of compounds already designed and synthesized, only one reached phase II clinical trials (Quarfloxin/CX-3543; (ClinicalTrials.gov identifier: NCT00780663), probably due to the specificity and selectivity requirements [68]. Quarfloxin is a fluoroquinolone which targets majorly *c-myc* promoter and ribosomal quadruplexes, with high specificity and selectivity, and demonstrated potent and tumor-selective activity *in vitro* and *in vivo* [65]. G-quadruplex sequences on its own can act as therapeutic agents as shown by Paula Bates group, who developed the G-quadruplex aptamer AS1411 which targets nucleolin, an overexpressed protein in cancer cells, and that shown good results in the treatment of renal cell carcinoma and myeloid leukemia [69]. Moreover, AS1411 has already reached phase II clinical trials (ClinicalTrials.gov identifier: NCT00740441) and seems a promising anticancer strategy [70]. AS1411 is a 26 nucleotide aptamer formed from the sequence d[$(GGT)_4TG(TGG)_4$] which folds into a bimolecular G-quadruplex [53]. The aptamer is nuclease resistant, similarly to unfolded sequences, and is active against a variety of cancer cell lines, whilst showing low toxicity against normal healthy cells [53].

The potential advantages and challenges of targeting gene promoter and telomeric G-quadruplexes/telomerase compared to proteins were summarized by Balasubramanian *et al.* and are presented below [50]:

A. Gene promoter G-quadruplex targets

Advantages

- Can target genes regardless of the druggability of the gene product;
- Lower likelihood of point mutations and resistance;
- Fewer copies of target, hence low concentration of inhibitor needed; this contrasts with the larger numbers (that is higher concentrations) of an overexpressing oncogenic protein or enzyme;
- Potential of unique sequence and three-dimensional structure for a given G-quadruplex — drug selectivity may be achievable by design;
- A number of relevant oncogenes and kinases are established as clinically validated targets in cancer: for example, *c-kit* and *BRAF*;
- Downregulation of expression of a target oncogene may be more important for tumor progression in a particular tumor type than telomerase (for example, *BRAF* in some melanomas), therefore, oncogene expression may be a more critical target for selective cell killing;
- High-throughput functional assays are available for many oncogenes and kinase targets, and are readily applicable to screening effects of promoter G-quadruplex targeting.

Challenges

- High affinity and selectivity needed, yet even the potency of state-of-the-art G-quadruplex ligands is considerably lower than that of typical enzyme inhibitors (often in the nanomolar range);
- Folding and structure can alter following ligand binding;
- As yet restricted diversity in available small-molecule inhibitors;
- Three-dimensional structures of very few promoter G-quadruplexes have been determined to date.

B. Telomeric G-quadruplex/telomerase targeting

Advantages

- Telomerase is expressed in most human cancers and not in normal somatic cells, so there is a possibility of broad clinical activity and limited cytotoxicity.

Challenges

- Telomerase is not yet a clinically validated target;
- High-throughput assays for telomerase inhibition by small molecules are not readily available and are under-developed;
- Telomere attrition resulting from telomerase inhibition is a slow process;
- Stem cells and germ cells express telomerase.

C. Protein targets

Advantages

- Straightforward if target has a well-defined active site;
- Large specialized compound libraries are available (for example, for kinases);
- Structural data are available on many existing targets.

Challenges

- Challenging if it involves protein-protein recognition;
- Active site changes following ligand binding;
- Undruggable if target is unstable or unfolded.

1.4 Methods for studying G-quadruplex structure and ligand-binding

1.4.1 Methods for G-quadruplex structure determination

The determination of high resolution models of G-quadruplex structures has relied on crystallographic and NMR spectroscopic methods. X-ray crystallography was the first technique used for a complete structure determination [2]. Using such technique for G-quadruplexes, the fundamental requirements apply: growing a crystal (ordering the material into 3D arrangements), collecting the scattered radiation and finally determining the phase angles for the amplitudes measured [53]. However the process is challenging and relatively slow. However, once crystals are obtained, detailed information about the structures can be acquired and to date over 50 crystal structures with good resolutions (below 1 Å in some cases) are available in the PDB server [2]. One disadvantage of this technique is that it can only determine the structure in solid state, which may be different from that adopted in solution. This is of the utmost importance, as the G-quadruplex structures are highly polymorphic and it can only describe the structure more easily crystallized [2]. In the case of NMR, new methodologies are in constant development to enable the accurate model determination and to unmistakably determine the variety of topologies possible for G-rich sequences. Nowadays, researchers can quickly and inexpensively synthesize isotope labelled G-quadruplex-forming oligonucleotides (with ^{15}N, ^{13}C and ^{31}P), that once folded can be used to determine the four guanine bases associated in each G-quartet formation [53]. Using this information, along with the glycosidic torsion angles determined, it is easy to link the phosphodiester backbone chains together and determine the G-quadruplex overall structure [53]. This technique requires much less sample preparation than crystallography, however it requires very pure and high-concentration samples (sometimes in the millimolar range) [2]. The simplest ^1H NMR spectrum in 90% H_2O/D_2O can provide valuable information, since there are relatively few protons in nucleic acids and the guanine NH_1 imino protons have a characteristic chemical shift when hydrogen bonded [2]. In general, imino proton chemical shifts >12.5 ppm are indicative of Watson–Crick base pairs (NH···H hydrogen bonds). Imino proton chemical shifts in the range of 10.5–12 ppm are indicative of guanine NH···O hydrogen bonds that appear in Hoogsteen alignments of the G-quartets [23]. More detailed information is provided by heteronuclear correlations, which allow complete assignment of the structure. Using NMR techniques is possible to study dynamics and kinetics, and therefore reveal the structures polymorphism. Some 30 G-quadruplex structures from NMR studies are currently in the PDB server [2].

Circular dichroism (CD) spectroscopy has also been extensively used to detect and identify G-quadruplex structures, most notably to distinguish parallel from antiparallel structures [2,71]. This is possible since the different conformations of DNA, B-form and G-quadruplex have CD characteristic spectra. Although CD results cannot be used to unambiguously determine the G-quadruplex structural type, they can typically limit the number of possibilities and represents one of simplest ways to predict the G-quadruplex topology [71]. Also, CD is a very powerful method for examining the cation dependence of quadruplex structure formation. The basis of CD spectroscopy are the use of circularly polarized light, which is emitted through a solution and interacts with the chiral species present. The chiral species will interact asymmetrically with the enantiomeric forms of light, i.e. the molecule absorbs the two components of light to a different extent, and the deviation of the direction and magnitude of the light is recorded [2]. Of this interaction results a measure of ellipticity (θ) in degrees, which is the difference of absorption between the two components of light, and the CD spectrum is obtained when the dichroism is measured as a function of wavelength in the UV range. The topology of a G-quadruplex structure (parallel, antiparallel or mixed) can be roughly assigned from the position and magnitude of the CD bands. In general, a maximum positive band in CD at ≈265-270 nm and a minimum negative band at ≈240 nm is indicative of a parallel-stranded G-quadruplex, whilst a maximum positive band at ≈290 and a minimum negative band at ≈265 nm describes an antiparallel-stranded G-quadruplex [23,72,73]. On its turn, duplex and single-stranded DNA characteristic bands are variable, depending on their sequence, but usually positive bands in at ≈275 and 220 nm, and a minimum band at ≈245 nm is indicative of duplex DNA. Topological arrangements of mixed parallel/antiparallel G-quadruplexes can also be detected, but cannot be guaranteed.

1.4.2 Methods for studying G-quadruplex—interactions

As discussed above, G-quadruplex structure has a very peculiar geometry, hence the small ligands can recognize and bind the structure through various binding modes (Figure 19), as determined by computational and chemical-biological approaches [23,26].

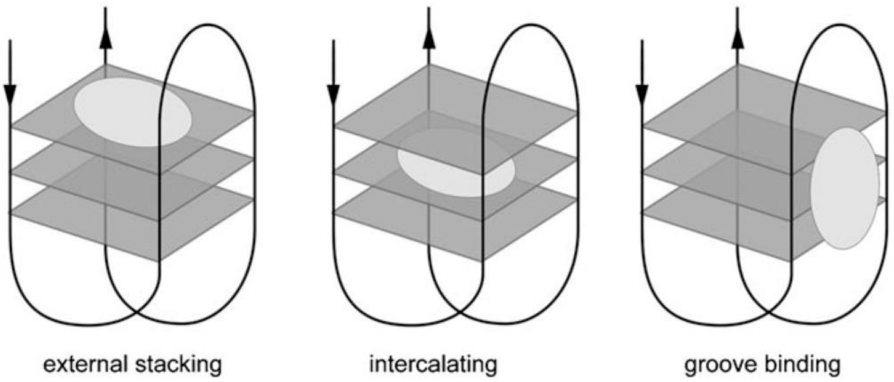


Figure 19 – Schematic representation of three G-quartet—ligand binding modes. Ligand–G4 complex with a) external stacking mode on the surface of the terminal quartet, b) intercalating mode between the stacks of G-tetrads, and c) groove binding mode. Taken from [23].

Three binding modes were already described for a variety of ligands: external stacking, intercalating and groove binding, whether the ligands interact through the terminal G-quartets, the middle G-quartets or the grooves/loops/backbones, respectively [23]. Ligands with large aromatic rings tend to interact with G-quadruplex structures mainly through π-stacking on the terminal G-quartets and less by intercalating between them [26]. This is due to the fact that the G-quadruplex is an extremely stable and rigid structure, but also because the structure distortion comprises a great energy cost [23]. Therefore, stacking in the outer G-quartets seems to be more energetically favorable and probable mode of binding. Ligands containing amino groups and (or) protonable side arms can interact trough the grooves, loops and the negatively charged phosphate backbone by electrostatic interactions or hydrogen bonds [26]. This remarkably improves the ligands water solubility. Also, long positively charged side arms can help small ligands to intercalate between the G-quartets and further stabilize the structure [26]. To gain a better insight of this G-quadruplex—ligand interactions, several techniques are routinely employed. Notably, spectroscopic methods, calorimetric techniques, melting experiments and biochemical assays. Some of these methods are described below and are also employed to study ligand binding to duplex DNA.

A. UV/Visible spectroscopy

A simple method for the identification of quadruplex-interacting ligands involves the use of UV/Vis spectroscopy [74]. It is based on the significant bathochromic shift and hypochromic effect on the bound chromophore Soret band (in the blue region of the visible spectrum), produced by the G-quadruplex, i.e. the peak shifts to longer wavelength and its absorbance decreases [23]. The hypochromic effect is thought to arise from the interaction between the electronic states of the chromophore and the DNA bases [23]. It has been postulated that the magnitude of the shifting can be related to the strength of interaction between the ligand and the G-quadruplex [75]. In addition to the qualitative information, by performing a titration of the DNA with the ligand, one can obtain quantitative information such as stoichiometries of binding constants [75]. For example, this method has been employed to study the interaction between methylene blue and *c-myc* quadruplexes [76]. This method is also commonly employed to study the stability of G-quadruplex structures by performing a UV-monitored melting experiment. In such experiment, melting temperature (T_m) and thermodynamic parameters can be determined from the change in the absorbance at 295 nm as a function of temperature [75]. Then, comparing the melting temperature values in the absence and presence of a ligand, it's possible to assess the interaction between the ligand and the G-quadruplex, as well as the stabilization (T_m increases) or destabilization (T_m decreases) of the structure. However, this method is slow and it is not straightforward to adequately deal with the complexities of multiple binding sites [8].

B. Circular dichroism (CD)

In addition the use of CD spectroscopy to distinguish between parallel and antiparallel folded quadruplexes, it is also used to characterize the complexes of G-quadruplex structures with ligands. Just like UV/Vis, CD spectroscopy has become an essential technique to study G-quadruplex—ligand interactions. In general, the quadruplex-ligands are nonchiral molecules and therefore are CD-inactive. Upon binding to quadruplexes, nonchiral molecules may give characteristic CD signals in the wavelength region corresponding to the absorbance of the ligand, due to induced CD (ICD) resulting from its bound state [23]. An ICD signal represents direct evidence of the interaction of the ligand with the G-quadruplex structure [23,71]. Intercalating compounds usually produce negative ICD signals or very small positive signals. Groove binding is generally indicated by the presence of a large positive induced CD signal upon titration of the compound into DNA [75]. It is also possible to monitor ligand-induced structural modifications by following the magnitude and position of the characteristic CD bands. In addition, CD can be used to monitor thermal unfolding of G-quadruplex structures and its stabilization by ligands (CD melting). This is usually performed by plotting the measured ellipticity at a specific wavelength, generally corresponding to the maximum of a positive band (around 260 nm for parallel structures or 290 nm for antiparallel structures), versus temperature [75]. Due to the low signal-to-noise ratio, CD is frequently limited to providing qualitative data although it can be used to determine equilibrium data.

C. Molecular fluorescence

Besides UV/Vis and CD spectroscopy, molecular fluorescence is one of the most commonly used techniques to study the interaction between ligands and G-quadruplex structures. Molecular fluorescence has the advantages of high sensitivity, selectivity and large linear concentration range [75]. Since the G-quadruplex ligands are usually compounds containing aromatic functional groups, these emit intense fluorescence due to the low-energy π->π* transition levels [75]. A popular molecular fluorescence technique to study ligand-G-quadruplex interactions is fluorescence resonance energy transfer (FRET) spectroscopy [77]. In this technique, the DNA sequence under study is labeled at the 5' and 3' ends with a donor and an acceptor fluorophore. Usually the donor used is 6-carboxyfluorescein (FAM) and the acceptor 6-carboxytetramethylrhodamine (TAMRA), respectively [75,77]. Upon excitation of the donor fluorophore, it transfers its energy to the acceptor, whereby the efficiency of energy transfer (E) depends on their distance and relative orientations. If the DNA structure suffers a modification (for example, unfolding), and consequently the distance between the two labeled sites changes, a measurable change in the efficiency of the energy transfer is observed [75]. Hence, FRET is used to determine the G-quadruplex T_m due to large differences between the fluorescence properties of folded and unfolded forms, being the ligand-induced shift in T_m a measure of the relative strength of interaction [75]. More recently in 2006, another assay was developed that allows the evaluation of the affinity and selectivity of ligands for G-quadruplex over duplex DNA [78]. G-quadruplex fluorescent intercalator displacement (G4-FID) is based on the competitive displacement of Thiazole Orange (TO) from quadruplex or duplex DNA when a ligand binds to the DNA (Figure 20). Since TO is non-fluorescent when free in solution but strongly fluorescent when bound to DNA, one can monitor the decrease of fluorescence induced by the ligand binding [79,80]. Some improvements to the protocol included the development if a high-throughput screening version of G4-FID, performed in 96-well microplates using a wide-spread equipment such as qPCR to perform the measurements. G4-FID is fast and easy isothermal assay, it doesn't requires modified oligonucleotides, and is able to analyze putative ligands that do not display fluorescence [80].

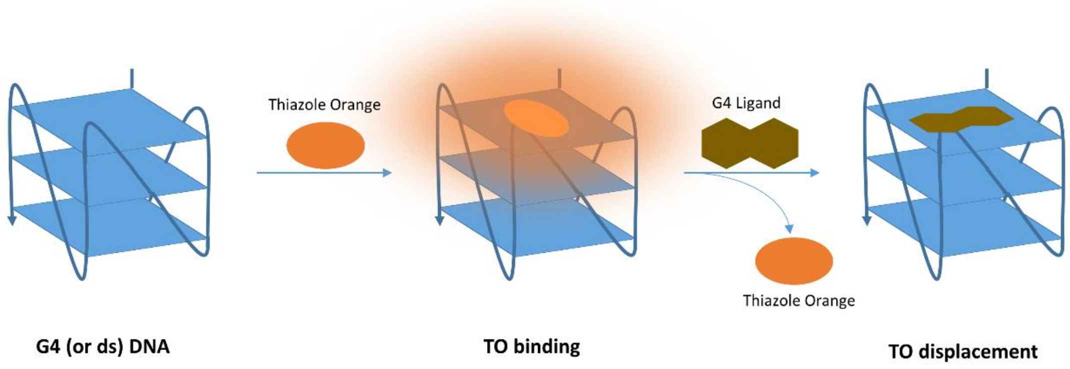


Figure 20 – Schematic representation of the G4-FID basis. G-quadruplex or duplex DNA is labelled by Thiazole Orange (TO); the fluorescence probe is then displaced by a small quadruplex-ligand. Adapted from [80].

D. Calorimetric techniques

Calorimetric techniques such as differential scanning calorimetry (DSC) and isothermal titration calorimetry (ITC) are used to determine reliable quantitative information on the thermodynamics of quadruplex-ligand interactions [23]. These methods require substantially more material, when compared to the spectroscopic methods, and may not be suitable for determining large binding constants [23]. ITC allows the monitoring of many bimolecular interactions without labelling the oligonucleotides. After performing a ligand titration and recording the heat response, one can plot a binding isotherm which is used to directly determine thermodynamic parameters such as binding enthalpy (ΔH°), the equilibrium binding constant (K_B), and binding stoichiometry (n) for an equilibrium binding system. From the equation $\Delta G = -RT \ln K = \Delta H - T\Delta S$, it is possible to calculate Gibbs free energy change (ΔG°) and entropy change (ΔS°) [23]. On its turn, DSC allows the continuous measurement of the apparent specific heat of a system as a function of temperature. DCS is a convenient and powerful manner of determining thermodynamic parameters such as calorimetric transition enthalpy (ΔH) and transition temperatures, T_m [23].

E. Nuclear Magnetic Resonance (NMR)

NMR experiments are also employed to study the interaction between ligands and G-quadruplexes, in particular, saturation transfer difference (STD) spectroscopy [81,82]. The technique is based on the principle that if a ligand shows two different signals because of a slow exchange between the bound state and the free state, a transfer of saturation is possible between both states [82]. When the DNA becomes saturated, ligands that are in exchange between a bound and the free state become saturated when bound to the DNA. By exchange, that saturation is further carried into solution where it is detected. By subtraction of this spectrum from a spectrum without DNA irradiation an NMR spectrum is obtained in which the resonance signals of the bound ligands can be assigned. Resonance signals from non-binders do not show up in the difference spectrum. The technique allows the user to obtain the ligand epitope maps, the parts of the ligand that bind to the G-quadruplex [82,83]. The ligand is normally used in an approximately 100-fold molar excess over the DNA [82].

F. Biochemical methods

Two methods with a polymerase chain reaction (PCR) step are frequently used to determine interaction between ligands and G-quadruplex structures. The first method is the telomeric repeat amplification protocol (TRAP) assay [8]. In this assay, the enzymatic activity of telomerase is measured by measuring the amount of TTAGGG repeats present, which is directly proportional. In the PCR step, two primers are used: the first one acts as a substrate for the addition of the TTAGGG repeats by telomerase in the presence of the quadruplex-ligand; the second primer is used as a reverse primer for PCR amplification. Then, the PCR products are resolved in a polyacrylamide gel electrophoresis and the data is used to construct a dose-response curve from which the ligand concentration required for 50% telomerase inhibition (IC_{50}) can be determined [8]. This assay, however, is prone to problems such as artifacts of amplification and unreliability of the PCR step, since the ligands can interfere in the second PCR step. Several variants of the TRAP assay are available commercially, for example, the TRAPEZE™ kit (www.millipore.com). Another method with a PCR step is the PCR stop assay (Figure 21) [84]. In this case, the quadruplex-forming sequence is annealed with the primer and incubated with the ligands being tested. Finally, primer extension is initiated by the addition of *Taq* DNA polymerase and incubation in a thermocycler with appropriate cycling conditions. The PCR products are submitted to gel electrophoresis analysis and the non-detection or decrease of concentration of PCR products confirms the interaction between the ligand and the DNA, as well as its relative strength [75]. There are two variations of the assay: on employs the use of ^{32}P-labelled DNA templates and a short complementary primer, where products can be discriminated between full-length or with major arrest sites; the other employs a non-labelled template DNA and a partially complementary primer, whereas products are full double-stranded DNA evaluated by gel electrophoresis and staining [23].

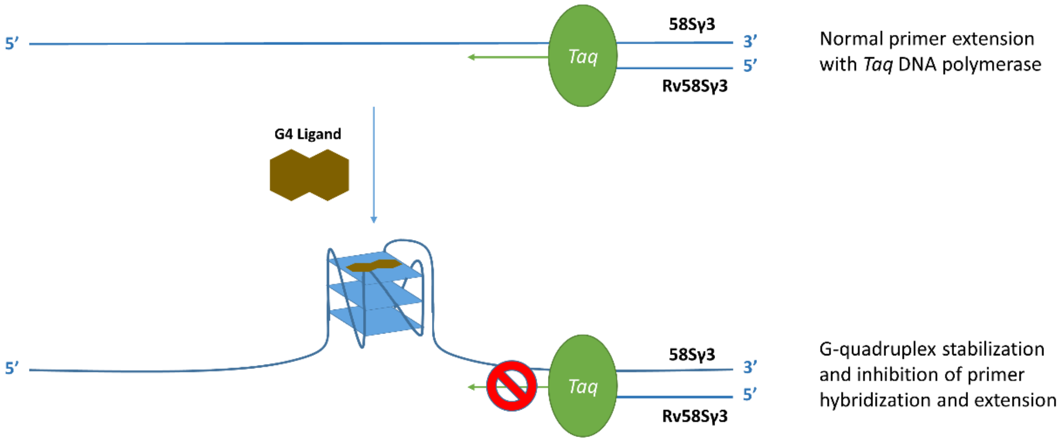


Figure 21 - Schematic representation of the PCR-stop assay. The induction and stabilization of G-quadruplex by a quadruplex-ligand inhibits *Taq* DNA polymerase and consequently primer extension.

1.5 Plasmid technology

Plasmids are circular, double-stranded DNA molecules that are independent from a cell chromosomal DNA and occur naturally in bacteria, yeast, and some higher eukaryotic cells [85]. Plasmid size ranges from few thousand base pairs to more than 100 kb, allowing the vectors to carry large DNA inserts. Plasmid DNA has been widely studied throughout the years for its use as a successful cloning tool due to the high replication rate, producing large numbers of copies of the desired DNA sequence [85]. The first constructed recombinant plasmid was reported by Stanley Cohen and his coworkers in 1973, making use of restriction enzymes, which are able to recognize and cut specific DNA sequences, and DNA ligase in order to re-attach such DNA sequences [86]. Generally, plasmids contain three functional regions: i) an origin of replication; ii) a drug-resistance gene for selection of transformed cells, and iii) a region where DNA can be inserted without interfering with plasmid replication or expression of the drug-resistance gene, the *polylinker* [85]. The plasmids most used in recombinant DNA technology are generally replicated in *Escherichia coli* (*E.coli*) hosts. *E. coli* has been extensively used as the cellular host for plasmid DNA amplification due to its easy transformation and fermentation, low nutritional requirements and rapid growth rate [87]. Therefore, this DNA amplification strategy has the advantage of being a relatively low cost process when compared to chemical synthesis, which is a still expensive, limited and labor intensive process.

The use of constructed plasmids harboring G-rich sequences was already described in the literature [60,65,88,89]. Maizels group, upon G-rich plasmid transcription, reported the formation of characteristic G-loop structures, containing G-quadruplex DNA on the single-stranded regions of the G-rich (non-template) strand, and a stable RNA/DNA hybrid on the template strand [60]. Afterward, using electron microscopic imaging it was shown that some proteins targeted and bound specific structures within these G-loops [90]. All together these experiments provide evidence for the formation of G-quadruplex structures *in vivo*.

Plasmid DNA is often used for the amplification of a specific DNA fragment [88]. For the recovery of the DNA fragments from the plasmid backbone, the use of restriction enzymes and a suitable size-based purification method is required. The purification techniques commonly employed can be roughly grouped in two categories: i) chromatographic methods and ii) using agarose or polyacrylamide gel electrophoresis. Gel electrophoresis although commonly used, gives poor yields and the oligonucleotides are often contaminated with soluble agarose or acrylamide which requires further time-consuming purification methods [91]. In the case of chromatography, four modes were already employed and described to the purification of restriction fragments: ion-exchange, hydrophobic, reversed-phase and size-exclusion chromatography (SEC) [92,93]. Although generally offering lower resolution than the other chromatographic techniques, SEC offers other advantages: easy instrumentation, isocratic elution, great freedom of buffer choice and elution in strict order of size [94]. SEC is based on the principle that analytes of different sizes will elute through a porous stationary phase at different rates, being that larger DNA molecules elute first while smaller DNA molecules elute last [92]. Small molecules can penetrate each of the stationary phase pores while the larger molecules are excluded and limited to the interparticle volume. Diverse chromatographic porous media are available for SEC, such as silica and agarose-based media [92]. Separations carried in SEC do not depend entirely on the size of the analyte but also on its shape. DNA molecules with sizes ranging from 20 to 1000 bp behave mainly like negatively-charged rods with limited flexibility, as proteins in contrast, act more like globular molecules [95]. In an ideal SEC process the solute should not interact with the stationary-size, therefore the use of a buffer with appropriate ionic strength (often using NaCl) is advisable [94].

In this work, plasmid technology was used for the biosynthesis of the novel G-rich quadruplex-forming DNA sequence 58Sγ3. A method for the purification of 58Sγ3 oligonucleotide was also developed, using size-exclusion chromatography with an agarose-based matrix. Finally, the study of the interaction between four novel ligands and the 58Sγ3 G-quadruplex structure, as well as its stabilization, was performed using circular dichroism, PCR-stop and G4-FID assays.

Chapter 2

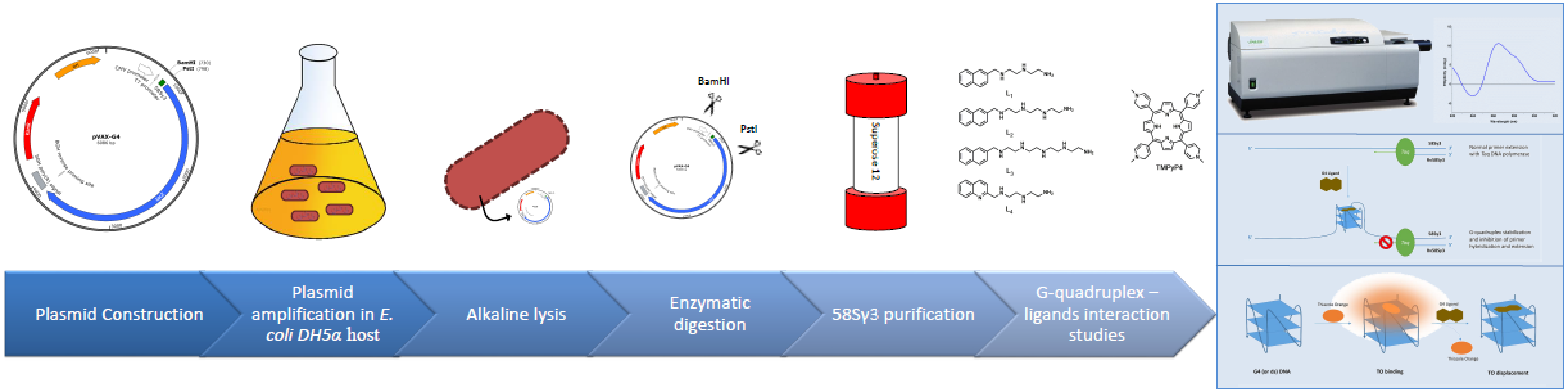
Aims of Work

The general scientific objectives of the present work are to contribute to the discovery and development of new G-quadruplex scaffolds for anti-cancer drug development and the study of novel compounds that stabilize G-quadruplex, with possible antitumor activity. The main objectives are the biosynthesis of a new G-quadruplex—forming DNA sequence and study of its interaction and stabilization by novel naphthalene and quinoline-based ligands.

For this purpose, the present dissertation is divided in three interrelated tasks:

1. Construction of a plasmid vector harboring the G-quadruplex sequence, in order to amplify the DNA using the replication machinery of a prokaryotic bacterial host.
2. Enzymatic digestion of the constructed plasmid and purification of the G-quadruplex—forming sequence by Size-exclusion Chromatography.
3. Biophysical and biochemical studies, namely Circular Dichroism, PCR-stop and G4-FID assays, of the interaction between the novel quadruplex-ligands and the G-quadruplex structure, to determine the ligands binding and stabilization of the structure.

The following scheme shows a graphical abstract of the overall experimental procedures used in this dissertation.


pVAX-G4
CMV promoter
T7 promoter
BamHI
PstI
BGH reverse priming site
BGH poly(A) signal
Superose 12
L1
L2
L3
L4
TMPyP4
Normal primer extension with Taq DNA polymerase
G4 ligand
G-quadruplex stabilization and inhibition of primer hybridization and extension
Thiazole Orange
G4 (or ds) DNA
TO binding
TO displacement
Plasmid Construction
Plasmid amplification in E. coli DH5α host
Alkaline lysis
Enzymatic digestion
58Sy3 purification
G-quadruplex – ligands interaction studies

Chapter 3

Materials and Methods

3.1 Materials

All the buffers and solutions were prepared with ultra-pure grade double distilled water purified with a Milli-Q system from Millipore (Billerica, MA, USA). The GreenSafe Premium, NZYMaxiprep and NZYMiniprep kits, restriction enzymes Speedy BamHI and Speedy PstI, dNTP and NZYDNA Ladder VI were purchased from NZYTech (Lisbon, Portugal). The 6.05-kbp plasmid pVAX1-*lacZ* was obtained from Invitrogen (Carlsband, CA, USA). Agarose was obtained from Hoefer (San Francisco, CA, USA). HyperLadder I was bought from Bioline (London, UK). The oligonucleotide sequences were obtained with HPLC purification from Stab Vida Inc. (Lisbon, Portugal). GeneJET™ gel extraction kit and Dream*Taq* DNA polymerase were obtained from Thermo Fisher Scientific (Waltham, MA, USA). T4 DNA Ligase was purchased from Promega (Madison, WI, USA). TMPyP4 was obtained from Tokyo Chemical Industry (Tokyo, Japan). Compounds L_1, L_2, L_3 and L_4 were synthesized by João Ferreira under the supervision of Carla Cruz.

3.2 Methods

3.2.1 Molecular cloning

3.2.1.1 Preparation of competent *E. coli* DH5α cells

A single colony of *Escherichia coli* (*E.* coli) *DH5a* was used to inoculate 50 mL of LB media (25 g/L, 10 g/L tryptone, 10 g/L yeast extract and 5 g/L NaCl) and incubated overnight at 37 °C and 250 rpm in a shake flask. The next day, 2 mL of the overnight culture was inoculated into 250 mL of LB media and incubated at 37 °C and 250 rpm until the optical density (OD) at 600 nm reached 0.3. The culture was transferred to sterile 50 mL polypropylene tubes and placed in ice for 5-10 minutes. Cells were then harvested by centrifugation at 5000 g for 10 minutes at 4 °C and re-suspended in 12.5 mL ice-cold $MgCl_2$ 100 mM solution. Cells were re-harvested at 4000 g for 10 minutes at 4 °C and the bacterial pellet was re-suspended in 2.5 mL ice-cold $CaCl_2$ 100 mM solution with further addition of 22.5 mL of the same solution and placed on ice for half an hour. Afterwards, cells were harvested by centrifugation again at 4000 g for 10 minutes at 4 °C and the pellet was re-suspended in 1 mL ice-cold $CaCl_2$ 85 mM solution with 15% glycerol. The competent cells were stored in cryogenic tubes at -80 °C until used.

3.2.1.2 pVAX1-*lacZ* production in *E. coli* DH5α

E. coli DH5a cell banks harboring plasmid vector pVAX1-*lacZ* were used to inoculate 50 mL of LB media (25 g/L) supplemented with 30 µg/mL kanamycin being cultivated overnight at 37 °C and 250 rpm in a shake flask until the OD_{600nm} reached 8. Afterwards, 2 mL of cultured volume was centrifuged at 11,000 g for 2 minutes and the pellets were further used to obtain plasmid DNA using NZYMiniprep kit.

3.2.1.3 Plasmid purification using NZYTech NZYMiniprep kit

The pellet previously obtained were re-suspended in 250 µL cold A1 buffer which contains RNase A followed by vortex until complete suspension. 250 µL A2 buffer (lysis buffer containing NaOH/SDS and supplemented with RNase) were added and the solution was gently mixed by inverting the tube 4 to 6 times. Then, 300 µL A3 buffer (neutralization buffer containing high salt concentration) were added and mixed gently by inverting the tube 4 to 6 times. The mixture was then centrifuged at 11,000 g for 10 minutes. The supernatant was transferred into NZYTech spin column, followed by centrifugation at 11,000 g for 1 minute. After discarding the flow-through, 500 µL AY buffer was added and centrifuged for 1 minute at 11,000 g. The column was then washed by adding 600 µL A4 buffer and centrifuged at 11,000 g for 1 minute. In order to fully remove the washing buffer, the spin column was spun for 2 extra minutes at 11,000 g. The plasmid DNA was eluted by adding 50 µL AE buffer followed by incubation at room temperature for 1 minute and centrifugation at 11,000 g for 1 minute and stored at -20 °C for further use. The concentration of the purified plasmid DNA was measured with NANOPhotometer (Implen).

3.2.1.4 pVAX1-*lacZ* digestion

A double digestion of pVAX1-*lacZ* with restriction enzymes Speedy BamHI and PstI was performed in order to create cohesive ends in the vector. 1 µg of plasmid was digested with 1 µL (10 units) BamHI and 1 µL (10 units) PstI in 2 µL NZYSpeedyBuffer and sterilized ultrapure water in a final volume of 20 µL. The mixture was incubated for 1 hour at 37 °C. In order to purify the digested and linearized vector, the reaction mixture was loaded in a 1% agarose gel and run for 40 minutes at 140 volts next to a sample of undigested plasmid as control. Next, the band corresponding to the linearized vector was excised from the gel and purified using GeneJET™ Gel Extraction Kit. Concentration of the purified vector was determined by NANOPhotometer in order to perform the necessary calculations for the ligation reaction.

3.2.1.5 Oligonucleotides preparation

The reverse and forward sequences of the desired insert used for cloning contained cohesive ends for BamHI and PstI restriction sites, in order to perform insertion of the sequence into pVAX1-*lacZ polylinker* (Figure 22). The oligonucleotide insert was prepared by performing annealing of the complementary oligonucleotides. Both complementary oligonucleotides were re-suspended at 100 µM in an annealing buffer (10 mM Tris. 50 mM NaCl, 1 mM EDTA, pH 7.5-8). The annealing procedure was performed by adding 10 µL of each oligonucleotide and 80 µL of annealing buffer to a 1.5 mL tube and placing it in a heat block at 95 °C. After 2-3 minutes, the block was removed from the heating unit and the tube was let to cool on the bench top for 45 minutes while in the heat block. The success of the annealing method was assessed by comparison of single-stranded forward and reverse oligonucleotides and the double-stranded obtained from the annealing procedure in a 2% agarose gel electrophoresis.

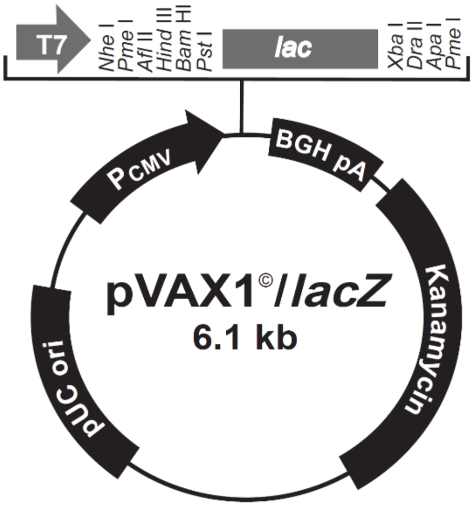

T7
Nhe I
Pme I
Afl II
Hind III
Bam HI
Pst I
lac
Xba I
Dra II
Apa I
Pme I
PCMV
BGH pA
Kanamycin
pUC ori
pVAX1©/lacZ
6.1 kb

Figure 22 – Schematic diagram of pVAX1/*lacZ*. Obtained from http://www.lifetechnologies.com /order/catalog/product/V26020

3.2.1.6 Ligation

For the ligation reaction, the masses of vector and insert were determined according to a 6:1 and 4:3 insert:vector molar ratios using equation (1). The required volumes of insert and vector were mixed with 1 µL T4 DNA ligase enzyme and 2 µL T4 ligase buffer (300 mM Tris-HCl pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP), to a final volume of 20 µL. The mixtures were then incubated for 3 hours at room temperature. After incubation, 10 µL were used for the transformation of competent *E. coli DH5a* cells, while the remaining 10 µL of the mixture were incubated overnight at 4 °C and then used for transformation. The cloning process is shown schematically in Figure 23.

$$Insert\ mass\ (ng) = \left(\frac{ng\ of vector \times kb\ size\ of\ insert}{kb\ size\ of\ vector}\right) \times molar\ ratio\ of\ \left(\frac{insert}{vector}\right) \quad (1)$$

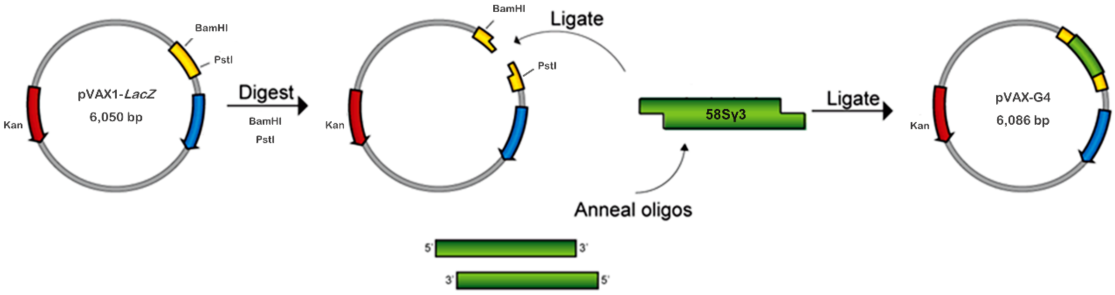

BamHI
PstI
pVAX1-LacZ
6,050 bp
Kan
Digest
BamHI
PstI
Kan
BamHI
Ligate
PstI
58Sy3
Ligate
Kan
pVAX-G4
6,086 bp
Anneal oligos
5'
3'
3'
5'

Figure 23 – Schematic representation of the molecular cloning process for the construction of pVAX-G4.

3.2.1.7 Transformation

Transformation of plasmid into *E. coli DH5a* cells was achieved by thermal shock. 10 µL of plasmid DNA were added to 100 µL of previously thawed on ice cells and left on ice for 30 minutes. The mixture was then heat-shocked at 42 °C for 1 minute to allow the plasmid DNA to enter and be retained inside the cells and once again placed on ice for 2 minutes. 200 µL of sterile LB medium were added to the cells and incubated at 37 °C and 250 rpm for 2 hours. Afterwards, the suspension was plated in LB agar plates (tryptone 10 g/L; yeast extract 5 g/L; NaCl 5 g/L and agar 15 g/L) supplemented with 30 µg/mL kanamycin, and incubated overnight at 37 °C. A negative control with cells missing the plasmid was used. Single colonies containing the plasmid of interest, selected to confirm the correct plasmid construction, were picked from the agar plate and used to inoculate 15 mL LB medium supplemented with 30 µg/mL kanamycin, and incubated overnight at 37 °C. 2 mL of the cultured volume were then centrifuged at 11,000 g for 2 minutes and the plasmid DNA was purified using NZYMiniprep kit (see section 3.2.2.2) followed by storage at -20 °C until further use.

3.2.1.8 Confirmation of clones by automated sequencing

In order to confirm the success of the cloning process, plasmid constructions with the expected insert were sequenced by Sanger sequencing method. About 150-200 ng of plasmid DNA was subjected to a pre-heat treatment at 98 °C for 5 minutes in 3.75 µL sterilized water. Then, 4 µL of Quick Start Kit sequencing mastermix (GenomeLab^{TM} GeXP, Beckman Coulter, Inc., Indianapolis, IN, USA) and 0.25 µM of primer (Table 3.1) were added to DNA sample in a reaction final volume of 10 µL. The PCR cycling conditions were the following: initial denaturation at 96 °C for 3 minutes, followed by 30 cycles of 20 seconds at 96 °C, 20 seconds at 55 °C and 4 minutes at 60 °C. The final extension was carried at 60 °C for 10 minutes. The DNA was then precipitated by adding 5 µL of Stop Solution/Glycogen mixture (2 µL of 3M Sodium Acetate (pH 5.2), 2 µL of 100 mM Na_2-EDTA (pH 8.0) and 1 µL of 20 mg/mL of glycogen) followed by 60 µL ice-cold 95% ethanol and mixed thoroughly. The mixture was centrifuged at 3,000 rpm and 4 °C for 30 minutes. The supernatant was carefully removed and the pellet was rinsed two times with 200 µL ice-cold 70% ethanol, being the sample centrifuged at 3,000 rpm and 4°C for a minimum of 5 minutes, for each rinse. After centrifugation, the supernatant was removed and the pellet carefully dried and re-suspended in 20 µL Sample Loading Solution. The re-suspended sample was transferred into the sample plate, covered with one drop of light mineral oil and loaded into the instrument. Sequence data was analyzed using GenomeLab system Beckman Coulter version 10.2 software.

Table 1 - Primers used for pVAX-G4 sequencing

Primers	Sequence
T7 Forward	5'-TAATACGACTCACTATAGGG-3'
BGH Reverse	5'-TAGAAGGCACAGTCGAGG-3'

3.2.1.9 Preparation of *E. coli DH5a* cell banks for pVAX-G4

The colonies confirmed to be harboring the correct plasmid construction were incubated at 37 °C and 250 rpm in 10 mL LB medium (25 g/L) supplemented with 30 µg/mL kanamycin, until the OD_{600nm} reached 0.7 (beginning of exponential phase). Afterwards, 3 mL of sterile 99%

glycerol was added to the culture medium and cell banks with the constructed plasmids were prepared by pipetting 1 mL of suspension in cryogenic tubes and stored at -80 °C.

3.2.2 pVAX-G4 plasmid production and insert purification

3.2.2.1 pVAX-G4 production

pVAX-G4, the pVAX1-*lacZ*-based plasmid bearing the desired 58-bp sequence, was amplified by autonomous replication in *E. coli* DH5α host. *E. coli* DH5α cells harboring the plasmid of interest were plated in an LB agar plate supplemented with 30 µg/mL kanamycin, and incubated overnight at 37 °C. Multiple colonies were used to inoculate 62.5 mL TB media (12 g/L tryptone, 24 g/L yeast extract, 4 mL/L glycerol, 0.017 M KH_2PO_4 and 0.072 M K_2HPO_4) and incubated until OD_{600nm} reached 2.6. Then, a determined volume of cell suspension calculated with equation (2) for a starting OD_{600nm} of 0.2 was added to 250 mL TB media, in 1 L shake flasks, supplemented with 30 µg/mL kanamycin and grown at 37 °C and 250 rpm. Cell growth was suspended at late log phase (OD_{600nm}≈8), cells were harvested by centrifugation at 5,000 g for 15 minutes at 4 °C, and the cell pellets were stored at -20 °C. During the incubation period, 1 mL aliquots of cell suspension were taken hourly in order to construct a growth curve. Data was plotted using OriginPro 8.0 (Microcal Software Inc, USA).

$$OD_i \times V_i = OD_f \times V_f \qquad (2)$$

3.2.2.2 Plasmid recovery and purification

Plasmid DNA was recovered by alkaline lysis and purified using NZYMaxiprep kit. The following procedure was set to be used with 125 mL cell suspension pellets. The pellet was re-suspended in 10 mL P1 buffer (25 mM Tris-HCl, 10 mM EDTA, 100 µg/mL RNase A, pH 8.0) using vortex. Cell lysis was achieved by the addition of 10 mL P2 buffer (200 mM NaOH, 1% (m/v) SDS, pH 8), mixed by gently inverting the tube 4 to 5 times and incubated at room temperature for 5 minutes. Then, 10 mL P3 buffer (3 M potassium acetate, glacial acetic acid to adjust pH, pH 5.5) were added and mixed by gently inverting the tube 4 to 5 times, in order to stop lysis and precipitate genomic DNA, followed by 10 minutes of incubation on ice. The alkaline lysate was centrifuged at 20,000 g for 30 minutes at 4 °C, in a Allegra™ 25R centrifuge (Beckman Coulter, Miami, FL, USA), in order to eliminate cell debris, gDNA and proteins. The supernatant was placed in new centrifuge tubes and spun in the same conditions in order to ensure complete elimination of cell debris, gDNA and proteins. The supernatant was then transferred into NZYTech Maxiprep anion-exchange column, previously equilibrated with 10 mL QBT buffer (750 mM NaCl, 50 mM MOPS, 15% (m/v) Isopropanol, 0.15% (m/v) Triton X-100, pH 7). After binding of the pDNA, the column was washed with 30 mL QC buffer (500 mM NaCl, 50 mM MOPS, 15% (m/v) isopropanol, pH 7) twice. Finally, the plasmid is eluted with 15 mL QF buffer (1.25 M NaCl, 50 mM Tris, 15% (m/v) isopropanol, pH 8.5) followed by the addition of 10.5 mL ice-cold isopropanol and gentle mixing by inverting 3 times. The mixture was centrifuged at 15,000 g for 30 minutes at 4 °C, the supernatant was discarded and the pDNA pellet was re-suspended in buffer 1 mL Tris-HCl. Concentration of the purified pDNA was measured with NANOPhotometer and the samples were stored at -20 °C for further use.

3.2.2.3 Agarose gel electrophoresis

The plasmid isoforms content and purity of each sample were assessed by horizontal electrophoresis using 15 cm, 1% agarose gels stained with 0.01% GreenSafe Premium and visualized under UV light in UVItec FireReader system (UVItec, Cambridge, UK). Electrophoresis was carried at 120 V, for 35 minutes, with TAE buffer (40 mM Tris base, 20 mM acetic acid and 1 mM EDTA, pH 8.0). Hyper Ladder I was used as a DNA molecular weight marker.

3.2.2.4 pVAX-G4 digestion

In order to excise the insert, a double digestion with the restriction enzymes Speedy BamHI and Speedy PstI was performed. The enzymatic reaction was scaled-up to 100 µg of plasmid in a total reaction volume of 400 µL. 30 µL (300 units) BamHI and 30 µL (300 units) PstI in 30 µL NZYSpeedyBuffer were added to the plasmid DNA sample followed by sterilized ultrapure water in order to make up the final volume. The mixture was incubated overnight at 37 °C. The effectiveness of the restriction digestion was assessed by horizontal agarose gel electrophoresis. 10 µL of the reaction mixture was loaded in a 2% agarose gel and run for 40 minutes at 120 volts. NZYDNA Ladder VI was used as a DNA molecular weight marker. Digested samples were stored at -20 °C for further use.

3.2.2.5 Size-exclusion Chromatography (SEC)

In order to separate and purify the restriction fragment from the plasmid backbone, size-exclusion chromatography was performed using a Superose 12® column (125 mm x 20 mm), an agarose-based matrix suitable for restriction fragments purification [92,94]. Separations were performed on ÄKTA Pure 25 L system with UNICORN 6.3 software. The column was washed with milliQ H_2O and equilibrated with 10 mM Tris-HCl pH 8.0 with 150 mM NaCl. The elution buffer used was the same as the equilibration buffer. Purified DNA restriction fragment was typically recovered at room temperature (≈ 20 °C) with a flow-rate of 0.5 mL/min, after injection of 1 mL of sample volume. The elution buffer was 10 mM Tris-HCl, pH 8 with 150 mM NaCl. The fractions were pooled according to the chromatograms obtained, concentrated and desalted with Vivaspin concentrators. The purity of the recovered samples was assessed by horizontal 2% agarose gel electrophoresis, run for 40 minutes at 120 volts.

3.2.3 G-quadruplex characterization and interaction studies

3.2.3.1 Circular dichroism (CD) studies

All CD measurements were performed at 25 °C, unless stated otherwise, with 6-15 µM strand concentration of oligonucleotides in a 30 mM phosphate buffer (15 mM KH_2PO_4, 15 mM K_2HPO_4, pH 7.1) containing 100 or 500 mM KCl. When required, oligonucleotides were annealed by heating at 95 °C for 15 minutes and slowly cooling to 4 °C for 2 hours. For CD titration with ligands, in order to study the association between the ligands and quadruplex structure, a 480 μM stock solution of each ligand was prepared in the same buffer as the oligos and the required volume was added directly in the cell. In the case of the KCl titration, a 2.5 M stock solution was used. CD spectra were recorded using a Jasco J-815 spectrapolarimeter equipped with a Peltier-type temperature control system (model CDF-426S/15), in a 1-mm quartz cell using an instrument scanning speed of 50 nm/min with a response time of 1 s over wavelengths ranging from 200 to 340. The recording bandwidth was 1 nm with 1 nm step size. The spectra were signal-averaged over three scans and baseline by subtracting a buffer spectrum. Data analysis and plotting was performed using OriginPro version 8.

3.2.3.2 CD melting

CD melting studies were performed in the temperature range 25-105 °C, at the heating rate of 5°C/min from 25 to 65 °C and 1°C/min from 65 to 105 °C by measuring the ellipticity at the wavelength of maximum variation upon oligonucleotide folding, 265 nm. Sampling was performed at every 1 °C with oligonucleotides at 6 μM strand concentration in the adequate buffer. Quadruplex samples were annealed as described above. CD values were normalized to the equation (3)

$$\mathrm{CD_{norm}} = (\mathrm{CD} - \mathrm{CD_{min}})/(\mathrm{CD_{max}} - \mathrm{CD_{min}}) \quad (3)$$

where CD is the signal at a given temperature; $\mathrm{CD_{max}}$ is the maximum signal observed at 265 nm; and $\mathrm{CD_{min}}$ is the minimum value obtained at 265 nm. Data was converted into fraction folded plots and fit to a Boltzmann distribution (OriginPro 8) without boundary constraints and the melting temperatures were determined from the two-state transition model used.

3.2.3.3 PCR stop assay

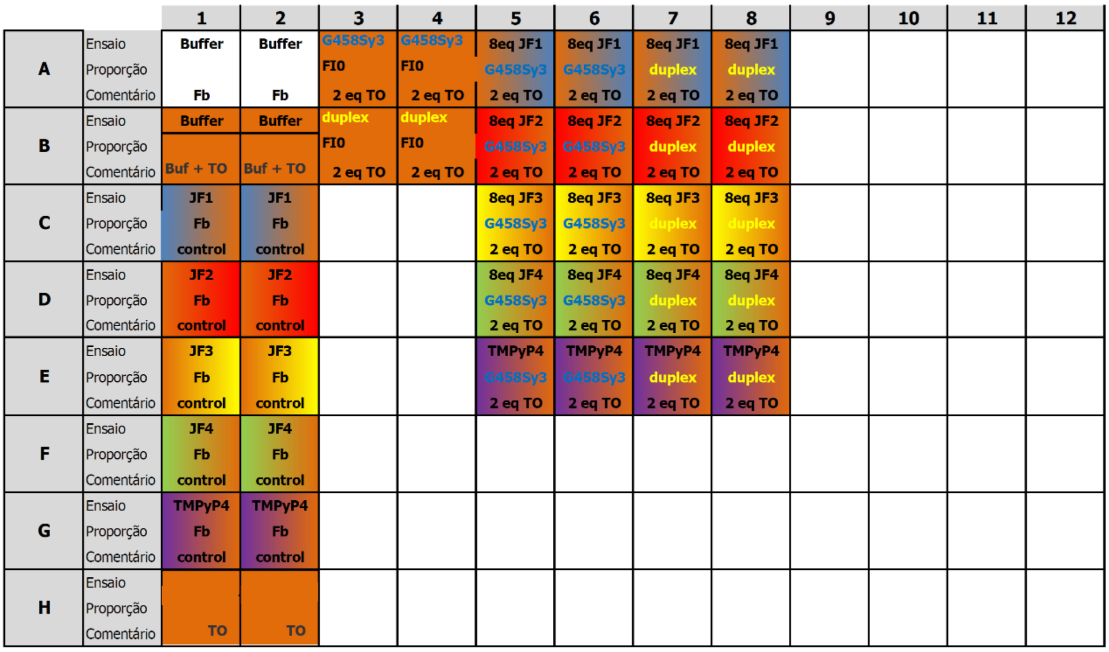
The stabilization of G-quadruplex structure by specific ligands was assessed by a PCR stop assay using a test oligonucleotide and a complementary oligonucleotide that hybridizes to the last G-repeats. 50 nM of primer and 200 nM of template DNA were mixed and annealed in the presence of 500 mM KCl by heating at 90 °C for 5 minutes followed by gradual cooling to room temperature over 1 h. The reactions were performed in a final volume of 20 μL. The annealed primer-template was mixed with 1x DreamTaq Green Buffer, 0.2 nM dNTP and increasing concentrations of the ligands (0-16 eq.). After the addition of the ligands, the reaction mixture was incubated for 30 minutes at room temperature. Primer extension was initiated by adding 1U Dream*Taq* DNA polymerase followed by incubation in a thermocycler with the following cycling conditions: 94 °C for 2 minutes, followed by 30 cycles of 94 °C for 30 s, 58 °C for 30 s, and 72 °C for 30 s. Amplified products were resolved on a 2% agarose gel stained with 0.01% GreenSafe Premium. Relative fluorescence was scanned and determined with ImageJ [96]. Relative fluorescence data was fit to a dose-response model and $\mathrm{IC_{50}}$ was determined for each ligand (OriginPro 8). Results represent the average of 2 independent experiments.

3.2.3.4 Fluorescence intercalator displacement assay (G4-FID)

The experiment was performed by using a modified protocol of the previously reported method [80]. Each run was performed in duplicate with 96-well microplates, in a final volume per well of 50 μL for each sample. Oligonucleotides were annealed as described above at 5 μM strand concentration and mixed with two molar equivalents (10 μM) of Thiazole Orange (TO). Eight equivalents of each ligand were added to both quadruplex and duplex DNA (40 μM). The fluorescence of each sample was measured at 25 °C in a qPCR instrument (Biorad CFX Connect™). Thiazole Orange was excited at 492 nm and the emission was collected at 516 nm. The percentage of TO displacement is calculated by the formula:

$$TO\ Displacement\ (\%) = 100 - ((FA/FA_0) \times 100)\ (4)$$

where the initial fluorescence with and without ligand (FI and FI_0, respectively) is subtracted by the background fluorescence F_b: $FA = FI - F_b$ or $FA_0 = FI_o - F_b$ [80]. A representation of a 96-well microplate and the disposition of the wells is presented in Figure 24.



		1	2	3	4	5	6	7	8	9	10	11	12
A	Ensaio Proporção Comentário	**Buffer** **Fb**	**Buffer** **Fb**	G458Sy3 **FI0** **2 eq TO**	G458Sy3 **FI0** **2 eq TO**	**8eq JF1** **G458Sy3** **2 eq TO**	**8eq JF1** **G458Sy3** **2 eq TO**	**8eq JF1** **duplex** **2 eq TO**	**8eq JF1** **duplex** **2 eq TO**				
B	Ensaio Proporção Comentário	**Buffer** **Buf + TO**	**Buffer** **Buf + TO**	**duplex** **FI0** **2 eq TO**	**duplex** **FI0** **2 eq TO**	**8eq JF2** **G458Sy3** **2 eq TO**	**8eq JF2** **G458Sy3** **2 eq TO**	**8eq JF2** **duplex** **2 eq TO**	**8eq JF2** **duplex** **2 eq TO**				
C	Ensaio Proporção Comentário	**JF1** **Fb** **control**	**JF1** **Fb** **control**			**8eq JF3** **G458Sy3** **2 eq TO**	**8eq JF3** **G458Sy3** **2 eq TO**	**8eq JF3** **duplex** **2 eq TO**	**8eq JF3** **duplex** **2 eq TO**				
D	Ensaio Proporção Comentário	**JF2** **Fb** **control**	**JF2** **Fb** **control**			**8eq JF4** **G458Sy3** **2 eq TO**	**8eq JF4** **G458Sy3** **2 eq TO**	**8eq JF4** **duplex** **2 eq TO**	**8eq JF4** **duplex** **2 eq TO**				
E	Ensaio Proporção Comentário	**JF3** **Fb** **control**	**JF3** **Fb** **control**			**TMPyP4** **G458Sy3** **2 eq TO**	**TMPyP4** **G458Sy3** **2 eq TO**	**TMPyP4** **duplex** **2 eq TO**	**TMPyP4** **duplex** **2 eq TO**				
F	Ensaio Proporção Comentário	**JF4** **Fb** **control**	**JF4** **Fb** **control**										
G	Ensaio Proporção Comentário	**TMPyP4** **Fb** **control**	**TMPyP4** **Fb** **control**										
H	Ensaio Proporção Comentário	**TO**	**TO**										

Figure 24 - Representation of a 96-well plate and disposition of wells in a G4-FID assay.

Chapter 4

Results and Discussion

Cancer is a major public health problem all over the world. According to the World Health Organization, approximately nine million new cases of cancer are detected per year. In 2014 in the United States only, around 1.6 million new cases were reported and 585000 deaths due to cancer occurred [97]. The common treatments for cancer are surgery, radiotherapy and chemotherapy. However, these treatments lack selectivity and can harm healthy cells as well, besides weakening the immune system. Because of these and other secondary effects, extensive research is being undertaken to find new alternatives for cancer treatment. Ever since G-quadruplexes were shown to exist *in vivo* and that the formation of such DNA structures were shown to be involved in cellular key processes, such as telomerase activity and oncogene transcription regulation, this has become a significant approach for anti-cancer drug development. The development of new small molecules with the ability to bind and stabilize the G-quadruplex structure, as well as the identification of new candidate sequences throughout the genome, and possible new G-quadruplex structures, is of the utmost importance in providing new scaffolds for this therapeutic approach. In this work, the production of a novel G-quadruplex—forming sequence is reported. The G-quadruplex structure is assessed and the interaction between the quadruplex and novel ligands is studied.

4.1 Initial studies and G-quadruplex sequence selection

The first approach to obtain a putative G-quadruplex—forming sequence was performed using plasmid pPH600, kindly provided by N. Maizels. pPH600 plasmid (3562 bp) is a derivative of pBluescript KS(+), which carries a 604 bp fragment of the murine Sγ3 switch region with repeats of the consensus sequence CTGGGCAGCTCTGGGGGAGCTGGGGTAGGTTGGGAGTGTGGG-GACCAGG [60]. Immunoglobulin switch regions are G-rich sequences, with high potential to fold into stable G-quadruplex structures. The objective was to replicate pPH600 plasmid in a suitable host, in this case *E. coli DH5a*, recover the plasmid and then excise a particular DNA sequence using the suitable restriction enzymes. Since pPH600 plasmid map was unknown, automated sequencing had to be performed in advance, in order to get information relative to the identity of the insert and its restriction map. All the necessary experimental procedures, namely cell transformation, plasmid production and recovery, and automated sequencing, were the same as described in the Materials and Methods section for pVAX-G4. Using primers T7 and T3 (Table 2), plasmid pPH600 was sequenced and the position and identity of the 604 bp insert was confirmed by comparison to the nonredundant nucleotide database at GenBank by using BLAST tool (NCBI) [98]. BLAST returned a significant alignment with Mouse Ig germline gamma-3 switch region (GenBank accession number M12182), with 100% identity, therefore confirming the insert's identity. Afterward, the restriction map of the plasmid's insert was constructed using the online software RestrictionMapper (www.restrictionmapper.org/) (Figure 25).

Table 2 - Primers used for pPH600 sequencing

Primers	Sequence
T7 Forward	5'-TAATACGACTCACTATAGGG-3'
T3 Reverse	5'- ATTAACCCTCACTAAAGGGA-3'

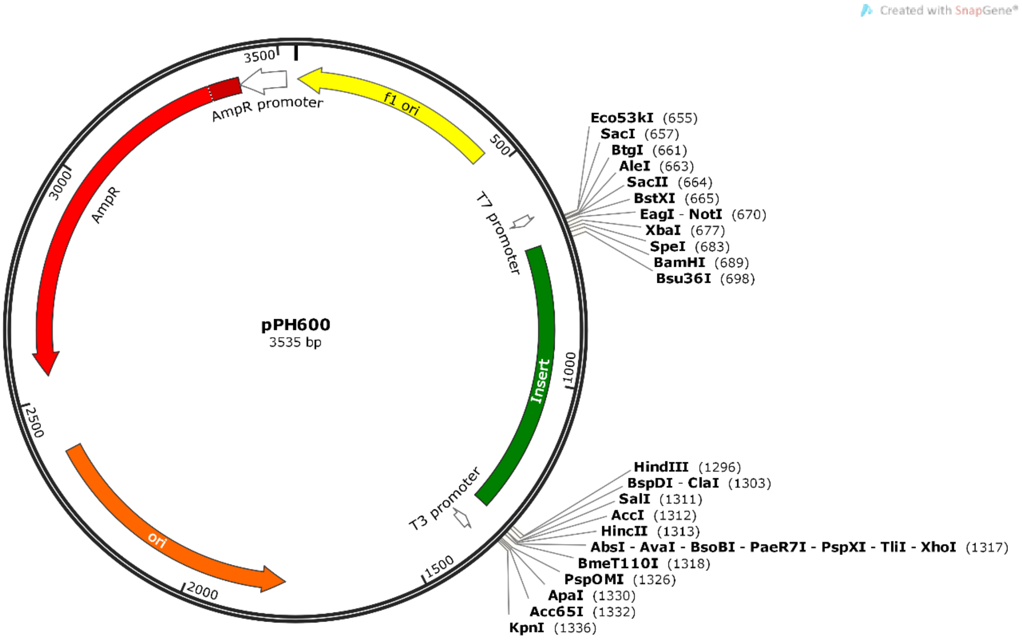

Created with SnapGene®
pPH600
3535 bp
3500
AmpR promoter
f1 ori
500
T7 promoter
Insert
1000
T3 promoter
1500
ori
2000
2500
AmpR
3000
Eco53kI (655)
SacI (657)
BtgI (661)
AleI (663)
SacII (664)
BstXI (665)
EagI - NotI (670)
XbaI (677)
SpeI (683)
BamHI (689)
Bsu36I (698)
HindIII (1296)
BspDI - ClaI (1303)
SalI (1311)
AccI (1312)
HincII (1313)
AbsI - AvaI - BsoBI - PaeR7I - PspXI - TliI - XhoI (1317)
BmeT110I (1318)
PspOMI (1326)
ApaI (1330)
Acc65I (1332)
KpnI (1336)

Figure 25 - Schematic diagram of pPH600. The plasmid was sequenced using primers T7 and T3. The restriction map was constructed using RestrictionMapper. The image was created with SnapGene®.

In order to find a putative G-quadruplex—forming sequence of interest, the DNA sequence of the 604 bp insert was submitted to a GQRS Mapper analysis, an online software that generates information on composition and distribution of putative Quadruplex forming G-Rich Sequences (QGRS) in nucleotide sequences [18]. The user can define the minimum number of tetrads, the maximum length of the G-quadruplex motif, as well as the size and composition of the loops. In this case, the search was limited to a maximum length of 40, a minimum of 2 G-quartets and no constraints relatively to the loops. The software found 16 putative G-quadruplex—forming sequences, with G-Scores ranging from 13 to 40 and different lengths, confirming the abundant occurrence of G-rich sequences within immunoglobulin switch regions (Table 3). Taking into account the availability of restriction enzymes and the G-Score, three DNA sequences were chosen for further studies: d[ATAT**GGAGACCTGGCTGGGG**AGC], d[TCGA**GGGGGGGCCCGG**TAC], and d[TGTGAATAACCTGCCTGAA**GGGCCACAGGGGAGCTGGGGCTATCAGATCACAGGG**TCC] (bold letters indicate quadruplex-forming region), excisable with the restriction enzyme combinations Eco31I/BbvCI, XhoI/KpnI and BspMI/PpuMI, respectively.

Table 3 - QGRS sequences found and the respective length and G-Scores.

Length	QGRS	G-Score
30	GGACAGATGGAGCAGTTACAGAGGAAAAGG	13
29	GGCGAATTGGAGCTCCACCGCGGTGGCGG	14
23	GGACCAGGCTGGGCAGCTACAGG	16
12	GGGGGGGCCCGG	18
15	GGTCAGGGTAGGAGG	19
29	GGTTAGTGGGAGTGTAGGGACCAGACTGG	19
29	GGCAGCTACAGGTGAGCTGGGTTGGATGG	19
27	GGGAGCTAGGGTAGGTGGAAGCATAGG	20
26	GGGTAGGAGGGAGTGTGGGGACCAGG	21
26	GGGGAGGTGGAGCTGTGGGGACCAGG	21
26	GGGTAAGTGGGAATATGGAGACCTGG	21
26	GGAGCTGAGGTAGGTGGGAACATAGG	21
29	GGTGAGCTGGGGTAGGAGGGAGTATGAGG	21
28	GGGAGCTAGGGTAAGTGAGGGTATGGGG	39
20	GGGAGCTGGGGTGGGTGGGG	40
36	GGGCCACAGGGGAGCTGGGGCTATCAGATCACAGGG	46

In order to excise the DNA fragments of interest, sequential and double digestions using the suitable restriction enzymes were undertaken. However, it was not possible to fully digest the plasmid, as no restriction fragment was produced upon enzymatic digestion. Several attempts were made, with different plasmid DNA/enzyme ratios and different incubation times, without success. The possibility of the fragments being non-detectable on agarose gels upon enzymatic reaction was also accounted and other staining methods such as ethidium bromide were used, however with the same outcome. A possible explanation is the fact that G-quadruplexes are able to resist attack by enzymes that target single-stranded or duplex DNA [60]. In fact, as seen by the QGRS analysis, there are 16 putative G-quadruplex forming sequences within the 604 bp insert, which could contribute to the stabilization of the plasmid and resultant enzyme attack resistance. In order to solve this issue, it was decided to construct a new plasmid by cloning one of the putative G-quadruplex—forming sequences into pVAX1-*lacZ*.

4.2 Construction of plasmid pVAX-G4

The chosen sequence to clone into pVAX1-*lacZ* was d[TGTGAATAACCTGCCTGAA**GGGCCACAGGGGAGCTGGGGCTATCAGATCACAGGG**TCC], a 58 bp sequence which had the highest G-score in the QGRS analysis (46) and was named 58Sγ3. The forward and reverse synthetic oligonucleotides of 58Sγ3 were purchased containing cohesive ends for BamHI and PstI restriction sites, in order to perform insertion of the sequence into pVAX1-*lacZ polylinker* (Table 4).

Table 4 - Sequences of forward and reverse oligonucleotides used for the construction of the 58Sγ3 insert.

Oligonucleotide	**Sequence**
58Sγ3 Forward	5'-GATCCTGTGAATAACCTGCCTGAAGGGCCACAGGGGAGCTGGGGCTATCAGATCACAGGGTCCCTGCA-3'
58Sγ3 Reverse	5’-GGGACCCTGTGATCTGATAGCCCCAGCTCCCCTGTGGCCCTTCAGGCAGGTTATTCACAG-3’

Prior to the molecular cloning procedures, the two oligonucleotides were annealed in order to produce the 58 bp double-stranded insert. The success of the annealing procedure was confirmed by 2% agarose gel electrophoresis (Figure 26).

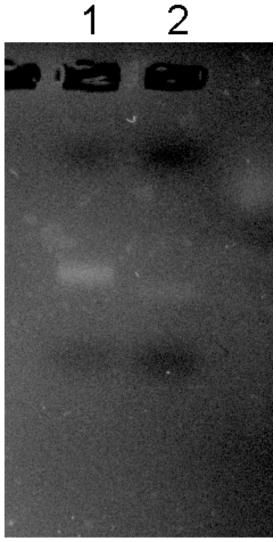


Figure 26 – 2% Agarose gel electrophoresis for the confirmation of annealing procedure. Lane 1: double-stranded oligonucleotide. Lane 2: single-stranded oligonucleotides.

The results clearly show a different pattern of migration in the agarose gel between the annealed double–stranded oligonucleotide (Lane 1) and the single-stranded oligonucleotides (Lane 2). Due to higher molecular weight, the double-stranded DNA (ds DNA) has a slower migration rate than single-stranded DNA (ss DNA), therefore the corresponding band is seen above in the gel as expected. This result allowed to proceed with the construction of the pVAX-G4 plasmid. The base plasmid pVAX1-*lacZ* was double digested with BamHI and PstI in order to create compatible cohesive ends for the DNA insert. The fact that both enzymes use the same buffer and optimal reaction temperature, enables the simultaneous use. Following enzymatic double digestion, the base plasmid was run in a 1% agarose gel and purified using GeneJET™ Gel Extraction Kit, as described in Chapter 3 section 3.2.1.4.

Once both DNA insert and base plasmid were prepared, it was possible to perform the ligation reaction with T4 DNA ligase. T4 DNA ligase catalyzes the joining of two strands of DNA between the 5’-phosphate and the 3’-hydroxyl groups of adjacent nucleotides. The ligation reaction was carried at room temperature for 3 hours and overnight at 4 °C, with different insert:vector molar ratios, 6:1 and 4:3. The ligation mixtures were transformed by thermal shock into competent *E. coli DH5a* cells and plated in kanamycin-supplemented LB agar plates, along with the negative controls. After the incubation period, single colonies containing the plasmid of interest were selected to confirm the correct plasmid construction. After purifying the plasmids using NZYMiniprep kit, these were sequenced in order to confirm the cloned insert, size and location. Automatic sequencing shown that only the plasmids constructed with the ligation reaction at room temperature harbored the desired DNA insert, by comparison with 58Sγ3 sequence. In fact, the optimal temperature of T4 DNA ligase is 25 °C, while reactions with shorter DNA inserts (less than 16 bases long) are advisable to be performed at lower temperatures, as indicated by the product's manufacturer. Cell banks of the correct plasmid constructions (Figure 27) were prepared and stored for further use.

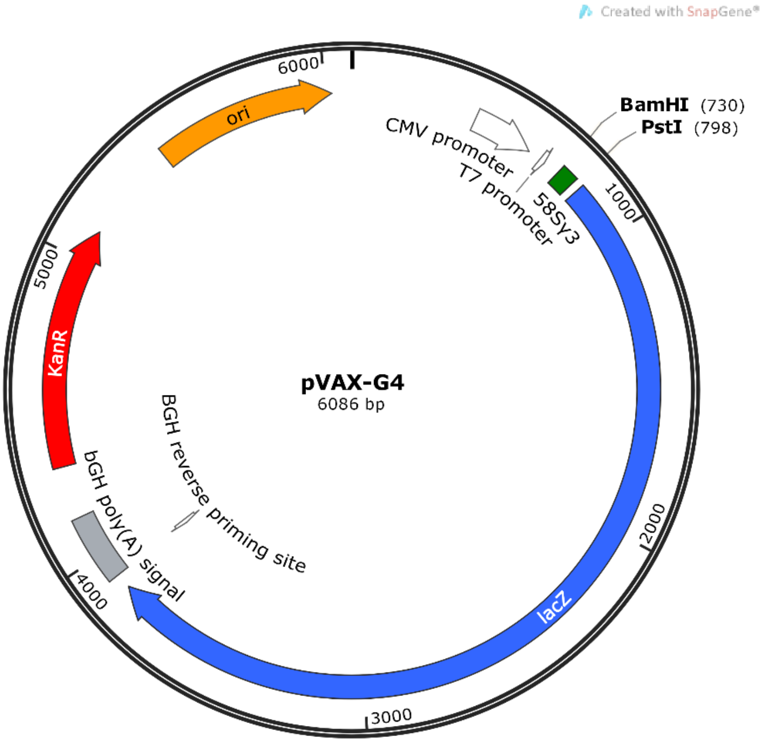


Figure 27 - Schematic diagram of the constructed plasmid pVAX-G4. The image was created with SnapGene®.

4.3 Biosynthesis of 58Sγ3 and purification

The amplification of the DNA insert 58Sγ3 was performed by autonomous replication in *E. coli DH5a* host harboring pVAX-G4 plasmid. *E. coli* has been extensively used as the cellular host for plasmid DNA amplification due to its easy transformation and fermentation, low nutritional requirements and rapid growth rate [87]. Therefore, this DNA amplification strategy has the advantage of being a relatively low cost process when compared to chemical synthesis. For this purpose, transformed *E. coli DH5a* cells were grown until late log phase (OD_{600nm}≈8) and harvested by centrifugation. The experimentally determined growth curve is shown in Figure 28.

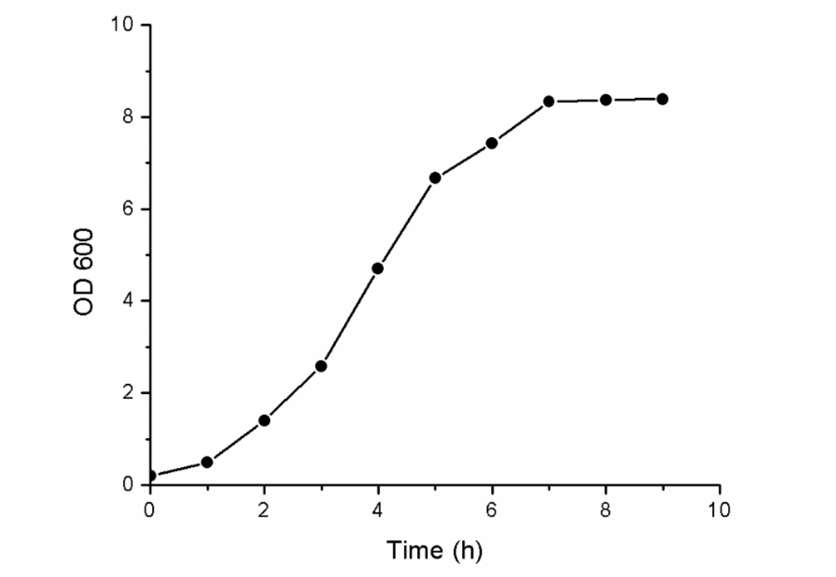


Figure 28 – Transformed *E. coli DH5a* growth curve. The latent (lag), exponential (log) and stationary phases of the bacterial growth can be identified between ≈0-1, ≈1-7 and ≈7-9 hours, respectively.

Recovered cells were disrupted by alkaline lysis and plasmid DNA was recovered using NZYMaxiprep kit. NZYTech columns are charged with a silica-based anion-exchange resin which allows the binding of plasmid DNA to the resin under appropriate low-salt and pH conditions. RNA, proteins and low-molecular–weight impurities are removed by a medium-salt wash while the plasmid DNA is finally eluted in a high-salt buffer and then concentrated and desalted by isopropanol precipitation. A representative 1% agarose gel electrophoresis resulting of a plasmid extraction is shown in Figure 29.

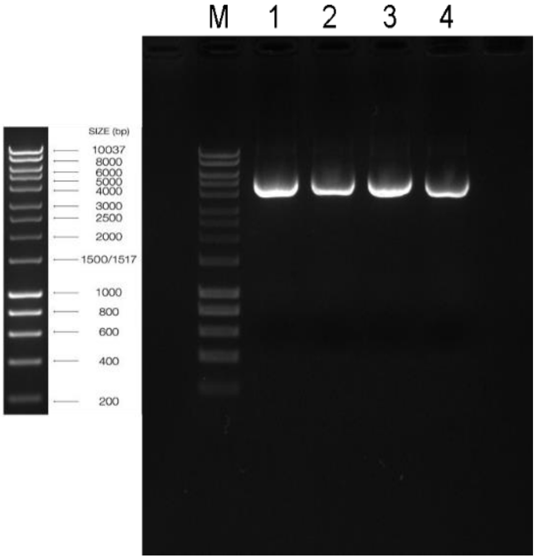


Figure 29 – Analysis of the purity and quality of pVAX-G4 by 1% agarose gel electrophoresis. Lanes 1-4 are plasmid samples obtained from the extraction procedure. Lane M: Molecular weight marker (Hyper Ladder I, Bioline). Plasmid DNA samples correspond to ≈300 µg per sample (1 mL).

Through the analysis of the gel is possible to see that the plasmid has the expected size (6086 bp) in comparison to the DNA molecular weight marker. It is important that no RNA and genomic DNA contamination results from the plasmid extraction process since it can interfere with the enzymatic reactions performed afterwards. Two different plasmid isoforms are possible to detect in lanes 1-4, namely supercoiled (sc) with traces of open-circular (oc) isoform. During the extraction process, supercoiled plasmid DNA degradation is likely to occur due to alkaline lysis, non-optimal pH and temperature and several centrifugations, so other isoforms, such as open-circular and linear, are often found in the plasmid samples. The structural conformation of the isoforms can be assessed by the agarose gel electrophoresis migration. The sc conformation migrates further in the gel than oc isoform since it has a more compact structure, which facilitates the migration through the agarose gel pores. The linear (ln) isoform, if present, would appear between the sc and oc isoforms. Following plasmid DNA extraction, pVAX-G4 had to be double digested with BamHI and PstI in order to isolate 58Sγ3 oligonucleotide. The enzymatic reaction was scaled-up in order to digest the maximum amount of plasmid DNA possible. The scale-up was accomplished by scaling the following basic reaction proportionately: one enzyme unit is defined as the amount of enzyme required to digest 1 µg of plasmid DNA in 1 hour at 37°C in a total reaction volume of 50 µL (as indicated by the restriction enzyme manufacturer). Moreover, the restriction reaction was set up in order to use the minimum enzyme concentration needed to achieve effective results. The effectiveness of each restriction digestion was assessed by horizontal agarose gel electrophoresis (Figure 30).

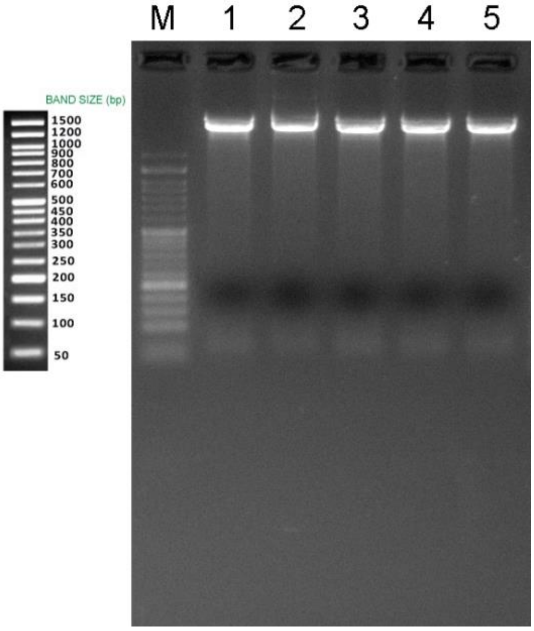


Figure 30 – Assessment of the effectiveness of the enzymatic double digestion. 2% agarose gel electrophoresis, lanes 1-5 are digested plasmid samples. Lane M: Molecular weight marker (NZYDNA Ladder VI, NZYTech).

On top, the brighter bands correspond to the plasmid backbone while the lower bands correspond to 58Sγ3 oligo, indicated by the correspondent expected molecular weight of 58 bp. Moreover, the enzymatic reaction confirmed the correct plasmid construction and the effectiveness of the enzymatic digestion reaction as no other isoform than linear is observed.

Downstream, the 58 bp oligonucleotide was separated and purified from the plasmid backbone by size-exclusion chromatography (SEC) using a pre-packed Superose 12® column, an agarose-based matrix suitable for restriction fragments [92,94,95]. All the digested samples were pooled together and loaded into the column at different flow-rates of 1 mL/min, 0.75 mL/min and 0.5 mL/min. The elution was monitored by UV absorption at 260 nm. The higher flow rates of 1 and 0.75 mL/min didn't offered the needed resolution as the restriction fragment would often elute together with the plasmid backbone (Figure 31).

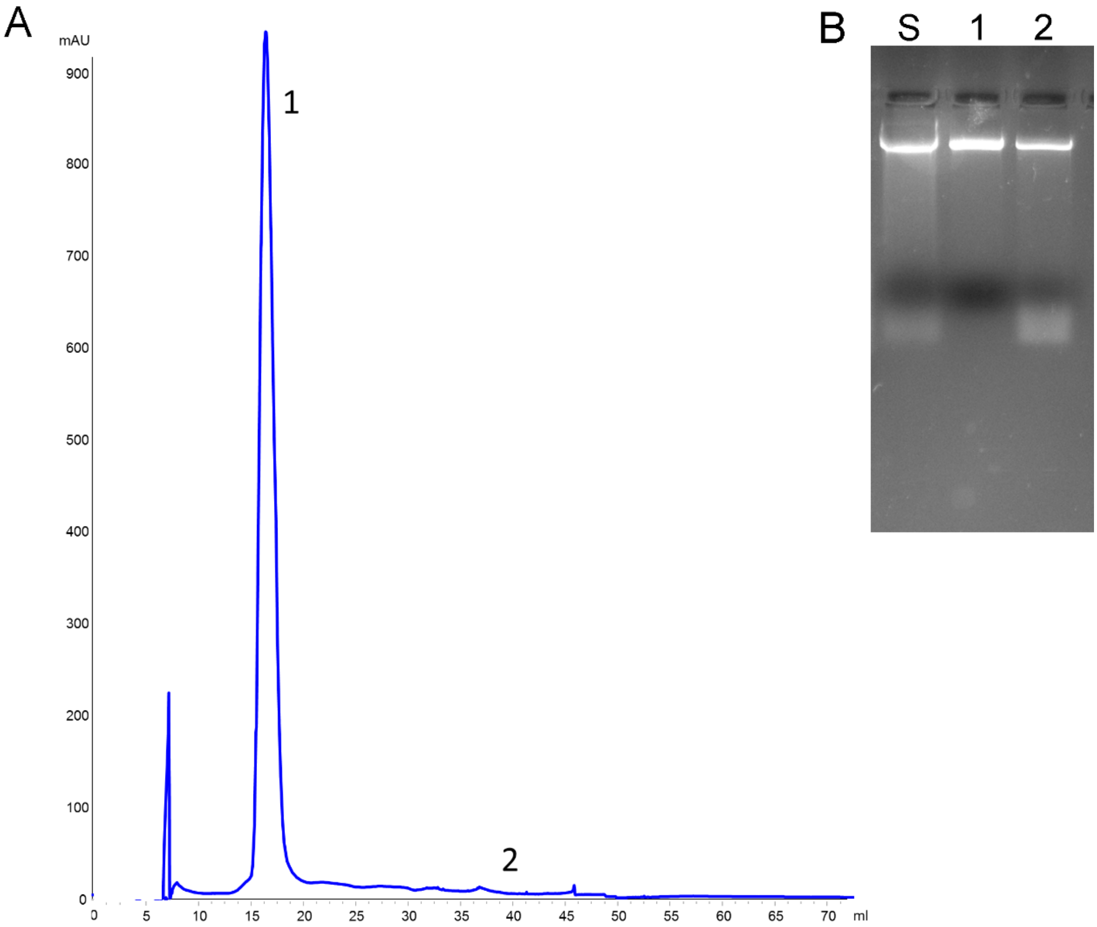


Figure 31 - Example of the chromatographic profiles obtained in the purification of 58Sγ3 at 1 mL/min and 0.75 mL/min. All experiments were performed during >1 hour, isocratically, with 150 mM NaCl in 10 mM Tris-HCl pH 8.0. (a) 1) Plasmid backbone sample; 2) 58Sγ3 sample; (b) Lane 1: Purified plasmid backbone; Lane 2: Unpurified sample of plasmid and 58Sγ3; Lane S: Injected sample.

SEC resolution is influenced by many factors: sample volume, ration sample volume/column volume, column dimensions, particle size, pore size and flow-rate [99]. Sample volume should be from 0.5-2% of the total column volume to achieve maximum resolution, while the longer columns are generally better. The resolution decreases as the flow-rate increases, despite providing faster separation, and it must be optimized to provide the necessary resolution [99]. The use of the lower flow-rate 0.5 mL/min, although increasing each chromatographic run time, offered the needed resolution for a complete separation of the two species (Figure 32).

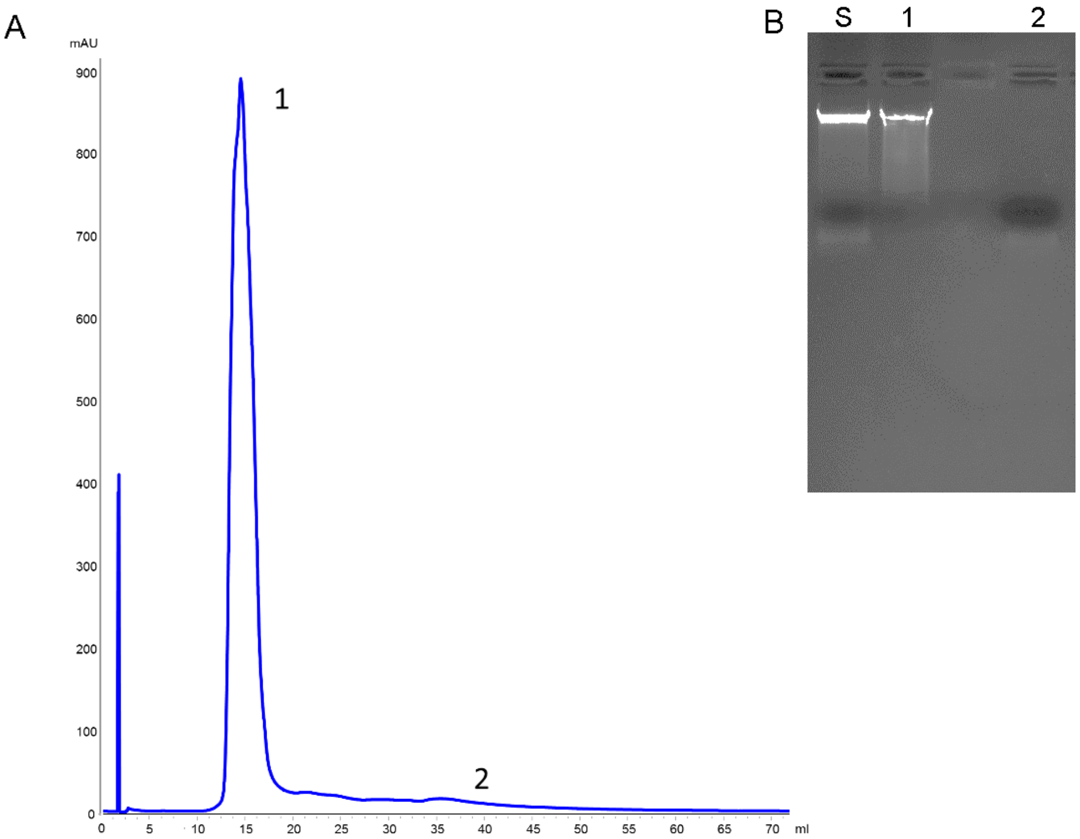


Figure 32 - Example of the chromatographic profiles obtained in the purification of 58Sγ3 at 0.5 mL/min. All experiments were performed during >1 hour, isocratically, with 150 mM NaCl in 10 mM Tris-HCl pH 8.0. (a) 1) Plasmid backbone sample; 2) 58Sγ3 sample; (b) Lane 1: Purified plasmid backbone; Lane 2: purified 58Sγ3; Lane S: Injected sample.

The use of 150 mM NaCl in the elution buffer prevents the DNA from interacting with the agarose matrix through ionic interactions, being the separation based strictly on molecular weight and size differences, which governs the access of the analytes to the pore volume inside the column packing particles [92,94,95]. As shown in Figure 31, the plasmid backbone elutes first while the restriction fragment 58Sγ3 elutes last. This result is in accordance with the expected since the molecules with higher molecular weight elute first. This is due to the fact that, with the increase of size, it becomes increasingly difficult for large DNA molecules to penetrate the matrix agarose gel pores that would be available for more compact molecules. Therefore, small DNA molecules will virtually penetrate each pore, eluting later than the large DNA molecules. The use of low flow-rates increases the resolution but has the advantage of increasing the experiment's duration, being the average time of each chromatographic run 1 hour.

4.4 58Sγ3 G-quadruplex formation and structure assessment

Immunoglobulin switch region Sγ3 was already reported as a G-quadruplex-forming region [11,15]; however there seem to be no other studies regarding its structure or interaction with ligands. Due to the limited time remaining to perform the proposed tasks of this work, an oligonucleotide stock was purchased in advance for the remaining studies. The oligonucleotide

58Sγ3 was purchased from Stab Vida Inc., dissolved in 30 mM phosphate buffer (15 mM KH_2PO_4, 15 mM K_2HPO_4, pH 7.1) containing 100 mM KCl and stored at -20 °C.

CD spectroscopy has been extensively used to study the polymorphism of G-quadruplexes, most notably to distinguish parallel structures from antiparallel structures by comparing their spectra with those of quadruplex DNAs of known structure [71]. It requires very little sample (μM concentration range) and is appropriate to examine a wide range of solution conditions and their influence on quadruplex formation [19]. G-quadruplex formation is induced by monovalent cations such as K^+ and Na^+, being K^+ considered more biologically relevant due to its higher intracellular concentration [25]. The formation of 58Sγ3 G-quadruplex and its topology were assessed by CD by performing a KCl titration. In K^+ solution it's expected that 58Sγ3 spontaneously fold into a stable G-quadruplex. Increasing concentrations of KCl (100 to 550 mM) were added to a 58Sγ3 solution and the CD spectra were acquired (Figure 33).

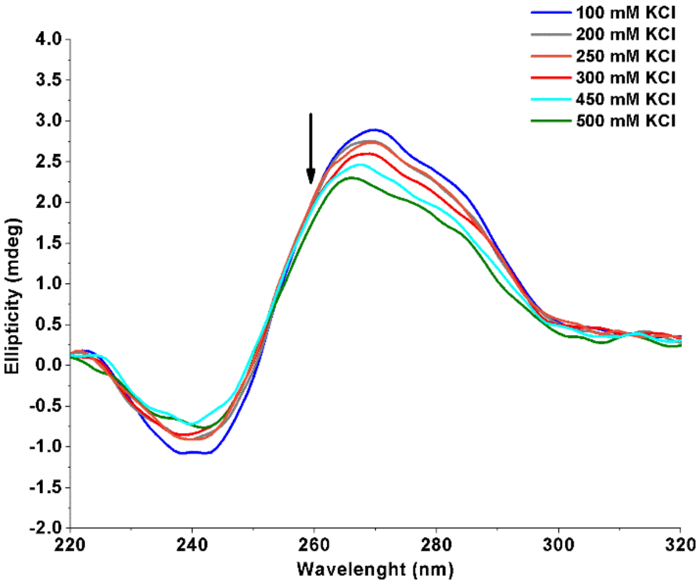


Figure 33 - CD titration spectra of 58Sγ3 G-quadruplex at 25 °C with increasing concentrations of KCl. Arrow indicates the increasing salt concentration.

The analysis of the spectra indicates that 58Sγ3 adopts a parallel-stranded conformation displaying the maximum ellipticity at around 265 nm and the minimum at around 240 nm, characteristic of parallel G-quadruplexes [100]. As the concentration of KCl increased, the positive band shifted into 265 nm. These results also indicate that 500 mM KCl is necessary for the complete folding of the DNA strands. The existence of a shoulder peak at around 285 nm may also be indicative of a mixed parallel/antiparallel conformation as previously reported for other G-quadruplexes, such as *bcl2* and d[TAGGGUTAGGGT], however it is not clear [71,101]. The higher KCl concentration needed in this case relatively to other sequences may be due to the strands length, requiring a higher ionic strength to fold into G-quadruplex structure. A similar experiment was performed with hTel49 oligonucleotide $d[AG_3T_2]_8$ and the same concentration of KCl was needed to fold into a stable G-quadruplex structure, supporting the hypothesis regarding the strands length [102]. The fact that 58Sγ3 sequence comprises four consecutive uneven G-tracts lead to the assumption that the parallel-stranded conformation is unimolecular but probably not the only conformation possible in terms of loop arrangement [19]. In order to confirm the formation of a parallel G-quadruplex structure, the well described porphyrin TMPyP4, which is known to induce the folding of DNA into quadruplex, was used [103,104]. The CD spectra were collected at differing TMPyP4/58Sγ3 molar ratios (from 0.1 to 4 eq.) in phosphate buffer without KCl (Figure 34).

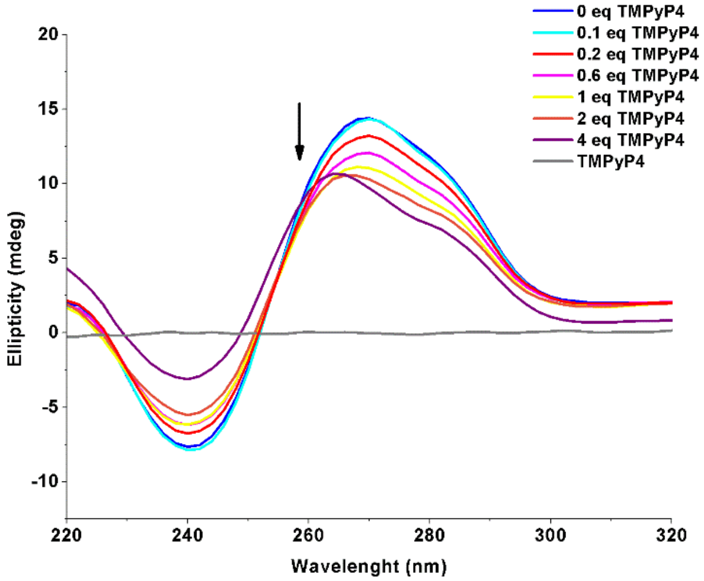


Figure 34 - CD titration spectra of 58Sγ3 G-quadruplex at 25 °C with increasing concentrations of TMPyP4. Arrow indicates the increasing ligand concentration.

The results clearly indicate the DNA folding in a parallel-stranded conformation with 4 molar equivalents of TMPyP4 which confirms the assumption relatively to quadruplex structure. It should be noted that CD results cannot be used to unambiguously determine the G-quadruplex structural type, and more potent techniques such as NMR and X-ray crystallography should be employed in the future.

4.5 Binding and stabilization of 58Sγ3 quadruplex with ligands

The development of naphthalene and quinoline-based ligands that exhibit both hydrophobic and hydrophilic characteristics presents itself as a promising strategy in G-quadruplex targeting. While the aromatic naphthalene and quinoline are capable of π-stacking with the G-tetrads, the protonated amine arms improve both the water solubility and DNA-binding by electrostatic interactions [2]. Four novel quadruplex-ligands were used in this study along with the porphyrin TMPyP4 (Figure 35) for control reasons.

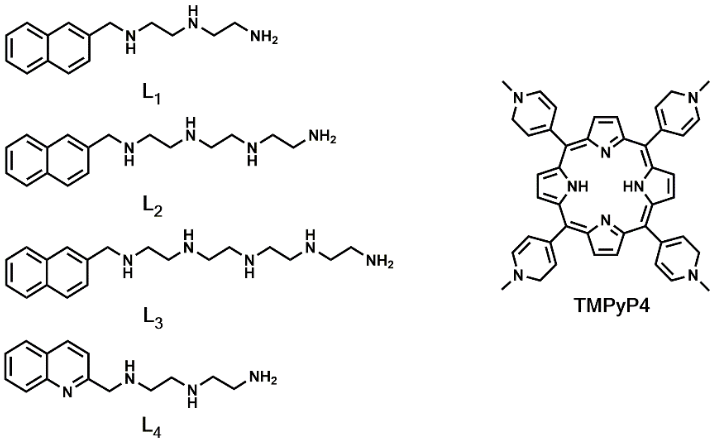


Figure 35 - Chemical structures of the ligands used in this work L1, L2, L3, L4 and TMPyP4.

Ligands L_1, L_2 and L_3 are naphthalene-derivatives with variable size amine arms, diethylenetriamine, triethylenetetramine and tetraethylenepentamine, with three, four and five protonable N, respectively. Ligand L_4 is a quinoline-derivative functionalized with a diethylenetriamine side arm. CD titrations and melting studies with the ligands were performed in order to explore the binding and stabilization of the quadruplex structure. The ligands were added to 58Sγ3 at different ligand/DNA molar ratio (from 0.1 to 8 eq.) in 500 mM KCl and the spectra were acquired. Upon addition of ligands, changes in the CD spectra were observed suggesting interaction of these molecules with the DNA (Figure 36).

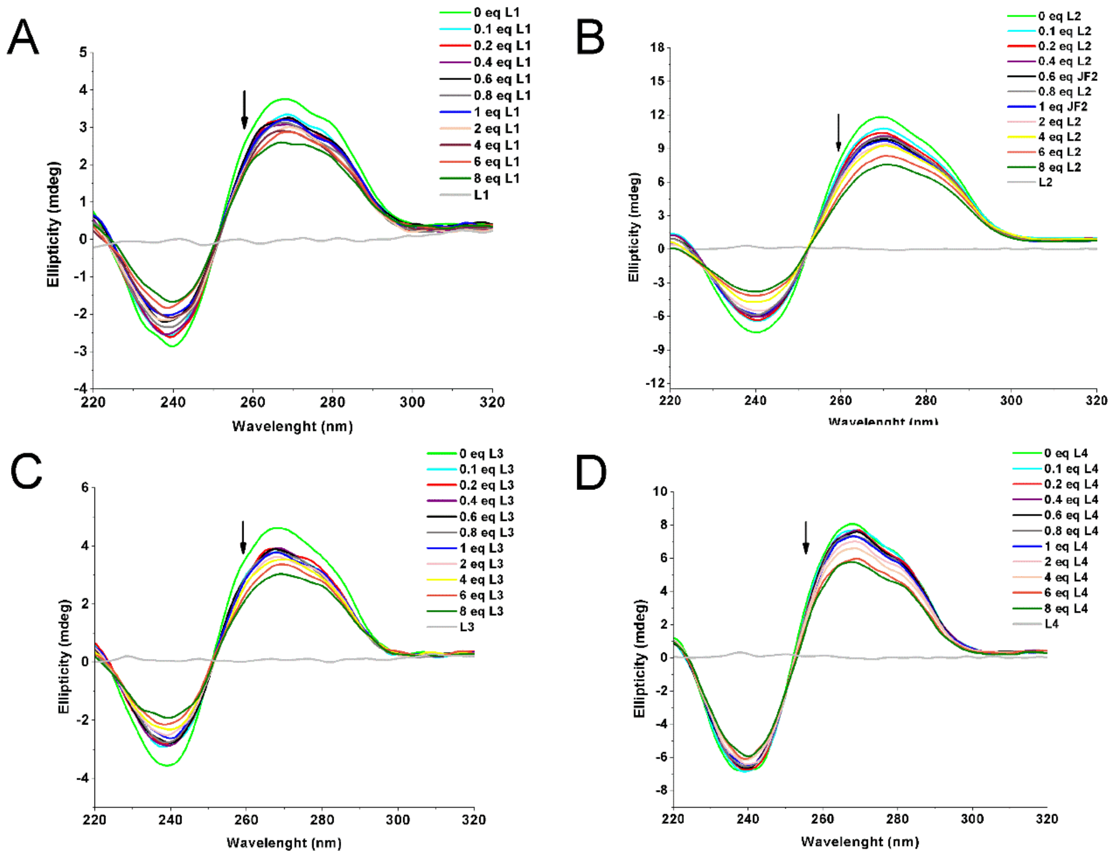


Figure 36 - CD titration spectrum of 58Sγ3 G-quadruplex at 25 °C in 30 mM phosphate buffer (500 mM KCl) with increasing concentrations of (a) L_1, (b) L_2, (c) L_3 and (d) L_4. Arrow indicates the increasing ligand concentration.

As the concentration of ligand increases, a slight decrease of ellipticity was observed; however the characteristic bands for parallel-stranded topology were maintained, demonstrating binding of the ligands and no disruption of the structure, maintaining the parallel topology. The degree of variation in the CD spectra is the same for all ligands with no induced CD band present which demonstrates that there is no structure interconversion induced by the ligands. CD spectra of the ligands was also recorded to demonstrate that the ligand had no CD absorption whatsoever in the wavelength range used (L_1, L_2, L_3 and L_4 spectra light-gray lines).

The thermal stabilization of 58Sγ3 quadruplex was further studied by measuring the ligand-induced change in the melting temperature (T_m) using CD melting analysis. Thermal denaturation of the quadruplex was monitored at 265 nm in the presence of K^+ and ligands when required (Figure 37).

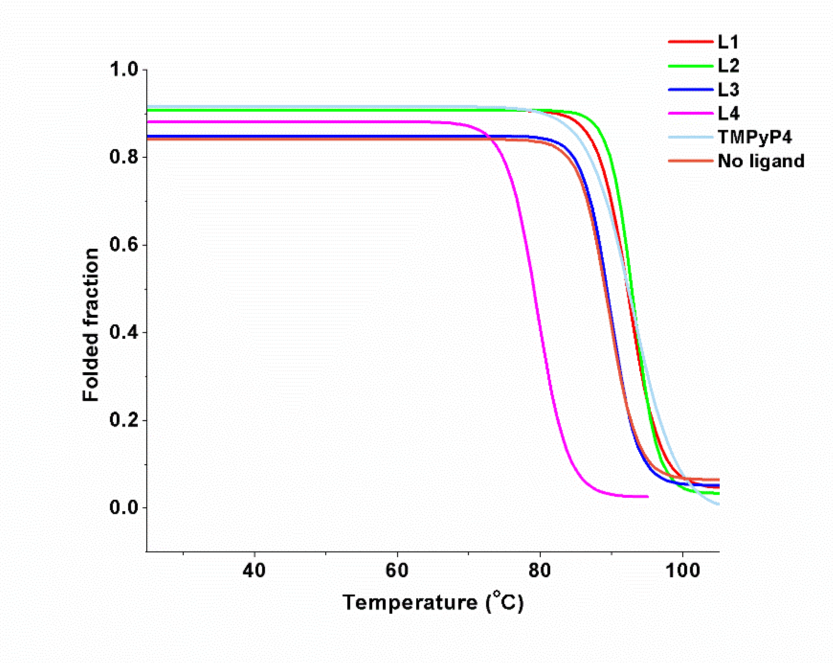


Figure 37 - CD melting curves of 58Sγ3 G-quadruplex in the absence and in the presence of 8 molar equivalents of ligands. Data was converted into folded fraction plots and fit to a Boltzmann distribution (OriginPro 8).

In order to determine T_m, the CD signal is normalized between 0 and 1, and the T_m is defined as the temperature for which the normalized CD is 0.5. In order to apply this method, the oligonucleotide must be (quasi)completely folded at the lowest recording temperature and unfolded at the highest temperature, otherwise the T_m will be overestimated. The determined melting temperature of 58Sγ3 quadruplex was 89.5 °C in 500 mM KCl. Upon addition of the ligands the stability of the quadruplex was enhanced with a clear increase of 58Sγ3 T_m values for ligands L_1, L_2 and slightly for L_3 while L_4 led to a decrease of $T_{m.}$ (Table 5).

Table 5 - Thermal Stability of 58Sγ3 with Ligands Measured by CD Melting Experiments.

Ligand	**T_m (°C)**	**ΔT_m (°C)**
No ligand	89.5 ± 0.38	-
TMPyP4	92.9 ± 0.37	3.4
L1	92.4 ± 0.29	2.9
L2	93.6 ± 0.33	4.1
L3	89.8 ± 0.27	0.3
L4	79.6 ± 0.27	-9.9

Ligand L_1 promoted an increase of around 3 °C in the T_m, similarly to the reference compound TMPyP4, while L_2 enhanced the stability of the quadruplex, reflected in the 4 °C increase in T_m. In turn, ligand L_3 led only to a slight increase of T_m. The length of the amine arm seems to play an important role in the stabilization of the G-quadruplex by naphthalene ligands; diethylenetriamine and triethylenetetramine arms promoted an increase in T_m when compared to the ligand with tetraethylenepentamine arm. On the contrary, the quinoline-based ligand L_4 with diethylenetriamine arm seems to destabilize the quadruplex resulting in a decrease of 10 °C in the T_m. This result indicates that, besides the amine arms, the naphthalene ring plays a major role in the stabilization of G-quadruplex, probably by π-stacking with the G-quartets.

In order to confirm the complete thermal denaturation of the G-quadruplex structure during the CD melting experiment and validate the applicability of the T_m determination method, a CD spectrum of 58Sγ3 at 95 °C was recorded (Figure 38).

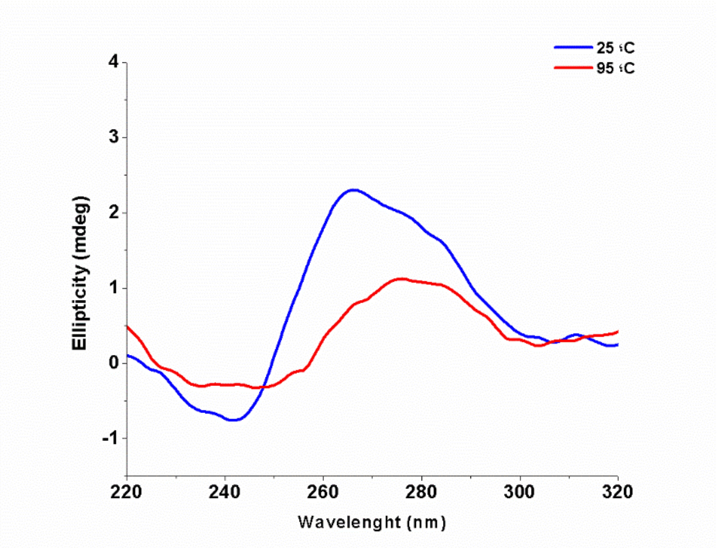


Figure 38 – CD spectra of the comparison between the folded (25 °C) and unfolded (95 °C) 58Sγ3 G-quadruplex.

The CD spectrum of 58Sγ3 at 25 °C shows a characteristic CD signature of a parallel G-quadruplex structure. Notably, the dominant CD band at 265 is weakened at 95 °C and shifted to the characteristic CD spectrum of an unstructured single strand [105,106]. The calculated T_m values by this method can therefore be considered as accurate.

4.6 Ligands block *Taq* DNA polymerase in a concentration-dependent manner

The stabilization of 58Sγ3 quadruplex was further investigated by PCR-stop assay. In this experiment, primer extension of DNA is inhibited by the stabilization of the quadruplex structure and inhibition of hybridization of the complementary sequence. The assay was carried out in the presence of 500 mM KCl. Despite quadruplex formation was induced by the presence of K^+, primer extension was not inhibited in absence of ligands and the full amplification product is detected. This indicates that *Taq* DNA polymerase is capable of unfolding the quadruplex easily leading to the extension of the full product (Figure 39).

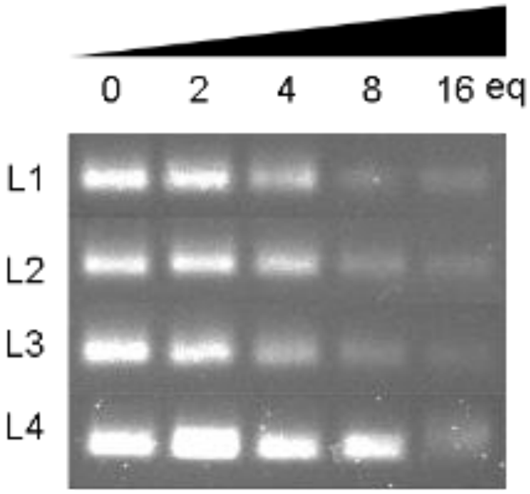


Figure 39 - PCR-stop assay and the effect of the ligands on the 58 bp double-stranded PCR product. 58Sγ3 was incubated with increasing concentrations of the ligands (0-16 eq) prior to PCR reaction.

The addition of ligands L_1, L_2 and L_3 led to a concentration-dependent decrease in the 58 bp amplification product. These results are in agreement with the stabilization observed in the CD melting experiments and indicate that the ligands are capable of inducing or stabilizing G-quadruplex formation in biological systems. The ligands demonstrated considerable amplification inhibition with 8 molar equivalents and with 16 equivalents there was almost a complete inhibition of *Taq* DNA polymerase. On the other hand, ligand L_4 showed dissimilar results, inhibiting *Taq* DNA polymerase with 16 equivalents but not in a concentration-depended manner as the results with 2, 4 and 8 equivalents are not concise. This may arise from the G-quadruplex destabilization upon ligand binding indicated by the CD melting results and the fact that the ligand appears to bind the duplex portion of 58Sγ3 in a similar extent as the quadruplex, inhibiting *Taq* DNA polymerase in an unspecific manner (discussed below in section 4.7). The fluorescence of each amplification product was scanned and quantified with ImageJ software and plotted for a better visual representation of the ligand effect (Figure 40).

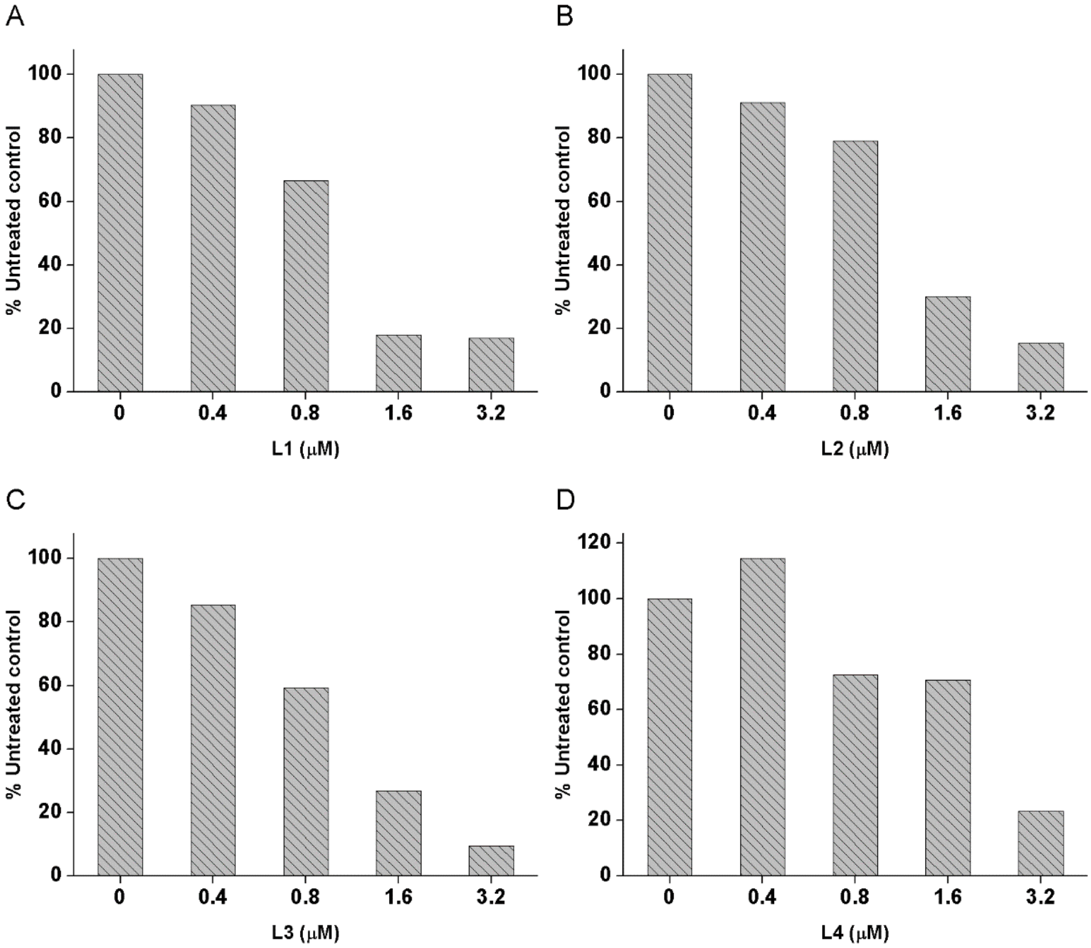


Figure 40 – Fluorescence quantification of the PCR-stop assay gel bands. The fluorescence is expressed in percentage of product amplification relatively to the untreated control.

The analysis of the plots reinforces the visual analysis of the agarose gels of Figure 39. As stated above, ligands L_1, L_2 and L_3 decreased the PCR products in a concentration-dependent manner, showing around 90% inhibition at the highest concentration of 3.2 μM of ligand. The highest values of inhibition at the maximum concentration used were observed for ligand L_3; this may be due to the fact that L_3 seems to bind selectively the G-quartets as was demonstrated by G4-FID assay and discussed below. To further demonstrate that the inhibition of the DNA amplification was due to G-quadruplex stabilization, a similar experiment was performed with a mutated oligonucleotide in one of the G-tracts with the substitution of two guanines by two adenines (58Sy3mu, 5'-TGTGAATAACCTGCCTGAAGGGCCACAGAAGAGCTGGGGCTATCAGATCACA-GGGTCC-3'). The mutated oligo does not fold into a G-quadruplex structure since one of the G-tracts has been disrupted and can be used to discriminate whether the *Taq* inhibition is due to G-quadruplex—ligand interaction or to unspecific interactions. In this case, no inhibition of the DNA amplification was observed at the higher concentration tested (Figure 41), which demonstrates that the ligands bound specifically to the G-quadruplex structure. A similar analysis using ImageJ was also performed.

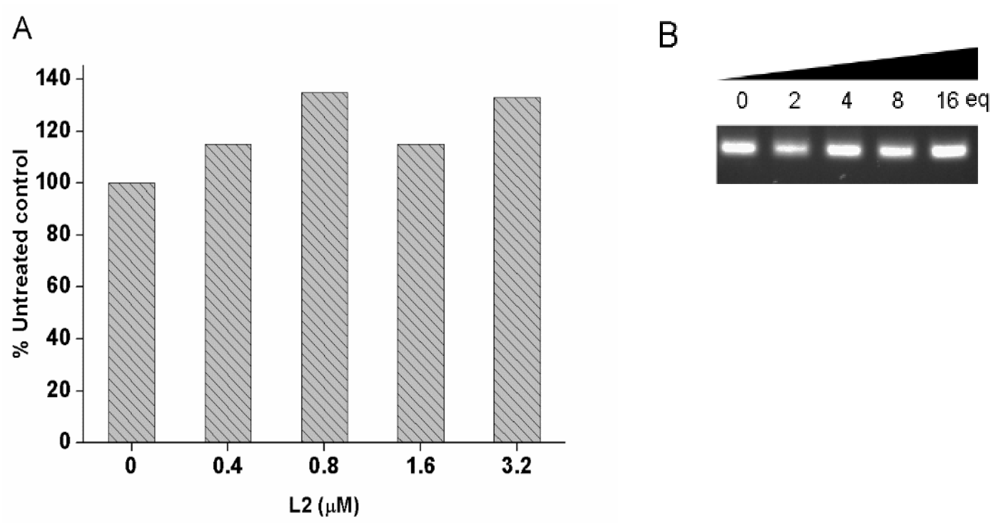


Figure 41 – Effect of the ligands on the 58 bp double-stranded PCR product with a control mutated 58Sy3mu oligonucleotide. (a) Fluorescence quantification plot and the (b) 58 bp double-stranded PCR product (representative agarose gel image of ligand L2 shown).

The existence of values above 100% is possible since, in this case, the amplification reaction depends only on the availability reagents and *Taq*'s activity and the results may differ from the PCR-tubes. From the fluorescence quantification plots is possible to determine the IC_{50} values which indicate the concentration of ligand required to achieve 50% inhibition of the amplification reaction [107]. Moreover, from the ratio IC_{50} 58Sy3mu/ IC_{50} 58Sy3 is possible to estimate a selectivity index. The values are listed in Table 6.

Table 6 - Effect of the ligands on 58Sy3 measured by the PCR-stop assay.

Ligand	**IC_{50} (μM)**		**Selectivity index**
	58Sy3[a]	58Sy3mu[a]	
L1	1.11	>3.2	>2.88
L2	0.89	>3.2	>3.59
L3	0.94	>3.2	>3.40
L4	-	-	-

[a] Mean of duplicates

All the ligands presented IC_{50} around 1 μM and selectivity indexes ranging from 2.88 to 3.40. Ligand L_2 presented the highest selectivity index and the lower IC_{50} when compared to the other ligands, which is in agreement with the results obtained previously. For L_4 it was not possible to determine IC_{50} or selective index since the results are not conclusive and the fitting to a dose-response model is inadequate. It should be noted, that the method here employed to determine IC_{50} values through fluorescence quantification is subjected to potential bias, since some subjective judgment is needed to decide where the gel bands end and the background noise begins. Therefore the values are qualitative and used to support the discussion.

4.7 Ligands affinity towards G-quadruplex over duplex DNA

To provide additional information regarding the interaction and specificity of the ligands towards the G-quadruplex, a fluorescence intercalator displacement (G4-FID) assay was performed using thiazole orange (TO) as a fluorophore. The assay is based on the competitive displacement of TO from quadruplex or duplex DNA when the ligand binds to the DNA. Since TO is totally quenched when free in solution but strongly fluorescent when bound to DNA (≈500- to 3000-fold exaltation), one can monitor the decrease of fluorescence induced by the ligand binding [80]. Therefore, the quadruplex-affinity of a candidate ligand can be determined through its ability to displace TO from the G-quadruplex DNA. Additionally, the selectivity of the ligands can be assessed by comparing the ability of the ligand to displace TO from quadruplex and duplex DNA [80]. The fluorescence is measured using two equivalents of TO as this is the preferential stoichiometry of the assay [79].

The analysis of the TO displacement plots (Figure 42) indicates that ligands L_1 and L_2 bind strongly to 58Sy3 G-quadruplex (TO displacement around 60%) while ligand L_3 binds more moderately with TO displacement around 40%. In the case of ligand L_4, it appears to bind weakly to G-quadruplex as the TO displacement is only 10%. The results are in agreement with the previous data obtained from CD melting curves with L_1 and L_2 strongly stabilizing the G-quadruplex structure. Regarding the specificity of the ligands, a similar assay was performed in the presence of duplex DNA and the results indicate that ligands L_1, L_2 and L_3 are quadruplex-specific as the TO displacement in those cases is significantly lower, in the case of L_3 is close to 0. Ligand L_4 seems to bind to duplex DNA in a similar extent as to G-quadruplex, being therefore non-specific. The results support the data obtained from the PCR-stop assay.

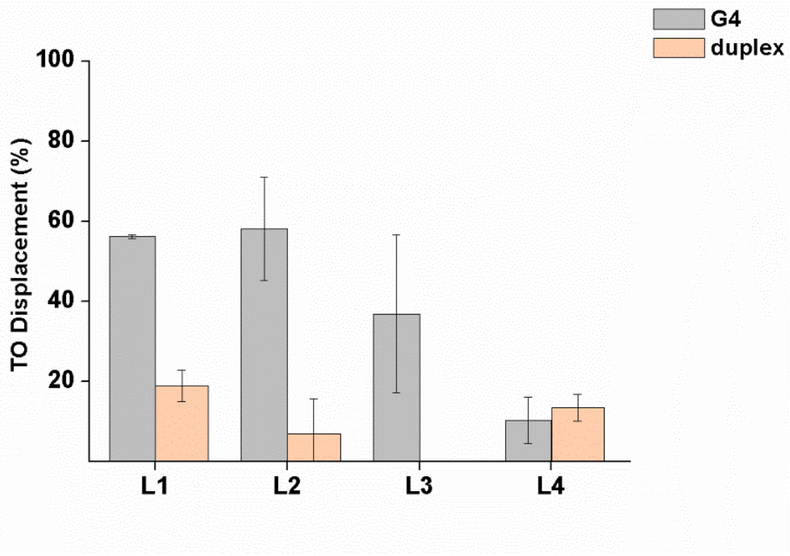


Figure 42 - TO displacement plot with the percentage of displacement of TO for each ligand with G-quadruplex and duplex DNA. For ligands L1, L2 and L3 TO displacement is higher in the case of G4 DNA relatively to duplex, indicating that the ligands are quadruplex-specific and bind more strongly to this DNA topology.

List of publications related to this thesis.

- T. Santos*, **J. Carvalho***, C. Feijó, M.C. Corvo, E.J. Cabrita, J.A. Queiroz, C. Cruz, Binding affinity between L-tryptophan containing dipeptide derivatives and nucleic acids, Int. J. Biol. Macromol. *(under review after major revision)*

*Equal contribution

- **J. Carvalho**, J. Ferreira, P. Pereira, E. Coutinho, F. Sousa, J.A. Queiroz, F. Sousa, E.J. Cabrita, C. Cruz, Biosynthesis of a G-quadruplex forming sequence and its stabilization by naphthalene-based ligands, Bioorg. Med. Chem. Lett. *(under review)*

Poster in national and international conferences.

- **J. Carvalho**, J. Ferreira, E. Coutinho, F. Sousa, E.J. Cabrita, J.A. Queiroz, C. Cruz (May 2015). G-quadruplex—forming sequence production and stabilization. III Jornadas de Química e Bioquímica, Covilhã, Portugal. *(“Best poster communication” award)* (Appendix A)

- J. Silva*, **J. Carvalho***, E.J. Cabrita, C. Cruz (May 2015). G-quadruplex production and stabilization. 5th International Meeting on Quadruplex Nucleic Acids, Bordeaux, France. (Appendix B)

*Equal contribution

Chapter 5

Conclusions and Future Perspectives

Since the discovery of the G-quadruplex, these high order structures have been proposed to have roles in many biological processes such as transcription, translation, recombination and genome stability. The biological relevance is reinforced by the presence of putative G-quadruplex—forming sequences in human genome promoters, oncogenic promoters for instance, and telomeric regions. The unique structural features of G-quadruplex DNA present a remarkable opportunity for the targeting of DNA in a structure-specific manner, leading to increased selectivity for quadruplex over duplex DNA. Moreover, current research focuses on the targeting of a particular G-quadruplex structure over other quadruplexes, increasing the selectivity. Such selectivity is important in order to reduce cytotoxicity from duplex-binding and to increase the bioavailability of the compounds, since the drug loss from non-specific binding is prevented.

The main objectives of this work were: i) firstly, the production of a novel G-quadruplex forming sequence and the assessment of its effective folding into a stable G-quadruplex structure; finally, the study of the interaction and stabilization of the G-quadruplex structure by novel acyclic quadruplex-ligands by a set biophysical and biochemical techniques.

The first approach for the production of a putative G-quadruplex—forming sequence shown to be unproductive, since the enzymatic digestion of the plasmid pPH600 was unsuccessful despite all the attempts made with different conditions, enzymes and approaches. The existence of several G-quadruplex motifs within the 604 bp insert of pPH600, possibly offered a higher degree of stabilization to the plasmid and resistance to enzymatic attack as was previously reported for G-quadruplex structures. To overcome such problems, a new plasmid based on pVAX1-*lacZ* was constructed, harboring a 58 bp putative G-quadruplex—forming sequence 58Sγ3, previously selected based on bioinformatics data. The plasmid construction and production were successful implemented and a restriction fragment purification strategy was employed. The use of plasmid amplification and size-exclusion chromatography purification strategy was found to be a less expensive and labor-intensive approach when compared to chemical synthesis of DNA sequences, although being slower and with lower yields. Circular dichroism studies with 58Sγ3 suggested that the sequence folded into a parallel-stranded G-quadruplex in 500 mM KCl buffer, with the characteristic signals at ≈240 nm and ≈265 nm, confirming the putative formation of DNA G-quadruplexes within the immunoglobulin switch regions. The same results were obtained in the presence of the G-quadruplex—inducing porphyrin TMPyP4. Moreover, four ligands derived from naphthalene and quinoline were tested to interact and stabilize the G-quadruplex structure. The naphthalene-based ligands L_1, L_2 and

L_3 were found to bind and stabilize the 58Sγ3 G-quadruplex structure improving its thermal stability in 2.9, 4.1 and 0.3 °C, respectively. However, the quinoline-based ligand L_4 seems to destabilize the structure, decreasing its melting temperature in 9.9 °C. PCR-stop assay shown that the naphthalene-based ligands are capable of inhibiting *Taq* DNA polymerase in a concentration-dependent manner. Ligand L_4 produced dissimilar results, probably due to the DNA structure destabilization, being the amplification inhibition at 16 eq possibly due to unspecific inhibition. The use of the control mutated oligo 58Sγ3mu made it possible to discriminate whether the inhibition of *Taq* polymerase was due to quadruplex-binding. Ligands L_1, L_2 and L_3 presented good results as no inhibition was detected, suggesting that the ligands were quadruplex-specific binders. The IC_{50} for ligands L_1, L_2 and L_3 was 1.11, 0.89 and 0.94, respectively. Finally, using G4-FID assay it was further shown that ligands L_1, L_2 and L_3 bound strongly and specifically quadruplex over duplex DNA. L_3 was found to be the more selective ligand as no TO displacement was produced for duplex DNA. On the other hand, L_4 bound in a similar extent to duplex DNA as to quadruplex, and with low affinity. The naphthalene-based ligands may therefore be potentially used as therapeutic agents to inhibit genes transcription in tumor cells.

In the future, the newly discovered G-quadruplex needs to be validated in terms of structure and biological role (linking the G-quadruplex formation in the immunoglobulin S region and immunodeficiency, meiosis was already proposed as one of the involved processes as well), with high resolution techniques such as crystallography and NMR and *in vivo* assays, respectively. A comprehensive picture of ligand binding interactions can also be provided by the crystallographic techniques. The quadruplex ligands should be tested with other G-quadruplex sequences such as telomeric or promoter quadruplexes and its cytotoxicity should be assessed. Moreover, the 58Sγ3 biosynthesis should be further optimized in order to obtain higher yields, probably through the use of new bacterial strains and plasmid combinations that offer higher production rates and easier DNA recovery. Relatively to the G-quadruplex field, still much effort is needed in the future, in order to clarify questions such as: what is the impact of the G-quadruplexes on gene expression? How can we discriminate its existence *in vivo*? Do the different topologies have an influence on the formation and regulation of G-quadruplex? Much work is still to be done.

References

[1] J.D. Watson, F.H.C. Crick, Molecular structure of nucleic acids, Nature. 171 (1953) 737–738.

[2] J.L. Huppert, Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes., Chem. Soc. Rev. 37 (2008) 1375–1384.

[3] A. Rich, DNA comes in many forms., Gene. 135 (1993) 99–109.

[4] P. Belmont, J.F. Constant, M. Demeunynck, Nucleic acid conformation diversity: from structure to function and regulation, Chem. Soc. Rev. 30 (2001) 70–81.

[5] A. Bacolla, R.D. Wells, Non-B DNA conformations, genomic rearrangements, and human disease, J. Biol. Chem. 279 (2004) 47411–47414.

[6] I. Bang, Examination of the guanyle acid., Biochem. Zeitschrif. 26 (1910) 293–311.

[7] M. Gellert, M. Lipsett, D. Davies, Helix formation by guanylic acid, Proc. Natl. Acad. Sci. U. S. A. 48 (1962) 2013–2018.

[8] S. Neidle, Therapeutic Applications of Quadruplex Nucleic Acids., 1st edition, Elsevier, 2012.

[9] H.T. Miles, J. Frazier, Poly(I) helix formation. Dependence on size-specific complexing to alkali metal ions, J. Am. Chem. Soc. 100 (1978) 8037–8038.

[10] E. Henderson, C.C. Hardin, S.K. Walk, I. Tinoco, E.H. Blackburn, Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs., Cell. 51 (1987) 899–908.

[11] D. Sen, W. Gilbert, Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis, Nature. 334 (1988) 364–366.

[12] C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H.J. Lipps, A. Plückthun, In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei., Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 8572–7.

[13] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, Quantitative visualization of DNA G-quadruplex structures in human cells., Nat. Chem. 5 (2013) 182–6.

[14] T. Simonsson, P. Pecinka, M. Kubista, DNA tetraplex formation in the control region of c-myc., Nucleic Acids Res. 26 (1998) 1167–72.

[15] A.I. Murchie, D.M. Lilley, Retinoblastoma susceptibility genes contain 5’ sequences with a high propensity to form guanine-tetrad structures., Nucleic Acids Res. 20 (1992) 49–53.

[16] T. Simonsson, G-quadruplex DNA structures - Variations on a theme, Biol. Chem. 382 (2001) 621–628.

[17] N. Maizels, L.T. Gray, The G4 Genome, PLoS Genet. 9 (2013) e1003468.

[18] O. Kikin, L. D'Antonio, P.S. Bagga, QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences, Nucleic Acids Res. 34 (2006) W676–682.

[19] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, Quadruplex DNA: Sequence, topology and structure, Nucleic Acids Res. 34 (2006) 5402–5415.

[20] J.L. Huppert, S. Balasubramanian, Prevalence of quadruplexes in the human genome, Nucleic Acids Res. 33 (2005) 2908–2916.

[21] A.K. Todd, M. Johnston, S. Neidle, Highly prevalent putative quadruplex sequence motifs in human DNA., Nucleic Acids Res. 33 (2005) 2901–7.

[22] J.L. Huppert, Hunting G-quadruplexes, Biochimie. 90 (2008) 1140–1148.

[23] T.M. Ou, Y.J. Lu, J.H. Tan, Z.S. Huang, K.Y. Wong, L.Q. Gu, G-quadruplexes: Targets in anticancer drug design, ChemMedChem. 3 (2008) 690–713.

[24] S. Neidle, The structures of quadruplex nucleic acids and their drug complexes, Curr. Opin. Struct. Biol. 19 (2009) 239–250.

[25] J. Dai, M. Carver, D. Yang, Polymorphism of human telomeric quadruplex structures, Biochimie. 90 (2008) 1172–1183.

[26] S. Zhang, Y. Wu, W. Zhang, G-quadruplex structures and their interaction diversity with ligands, ChemMedChem. 9 (2014) 899–911.

[27] J.T. Davis, G-Quartets 40 Years Later: From 5'-GMP to Molecular Biology and Supramolecular Chemistry, Angew. Chemie - Int. Ed. 43 (2004) 668–698.

[28] D.J. Patel, A.T. Phan, V. Kuryavyi, Human telomere, oncogenic promoter and 5′-UTR G-quadruplexes: Diverse higher order DNA and RNA targets for cancer therapeutics, Nucleic Acids Res. 35 (2007) 7429–7455.

[29] J. Zhou, A. Bourdoncle, F. Rosu, V. Gabelica, J.L. Mergny, Tri-G-quadruplex: Controlled assembly of a G-quadruplex structure from three G-rich strands, Angew. Chemie - Int. Ed. 51 (2012) 11002–11005.

[30] A.T. Phan, Human telomeric G-quadruplex: Structures of DNA and RNA sequences, FEBS J. 277 (2010) 1107–1117.

[31] M. Webba Da Silva, Geometric formalism for DNA quadruplex folding, Chem. - A Eur. J. 13 (2007) 9738–9745.

[32] M. Marusic, P. Sket, L. Bauer, V. Viglasky, J. Plavec, Solution-state structure of an intramolecular G-quadruplex with propeller, diagonal and edgewise loops., Nucleic Acids Res. 40 (2012) 6946–56.

[33] T.K. Raouf, G-quadruplexes: Kinetic stability and effects on the c-KIT promoter, University of Southampton, 2013.

[34] H. Yaku, T. Fujimoto, T. Murashima, D. Miyoshi, N. Sugimoto, Phthalocyanines: a new class of G-quadruplex-ligands with many potential applications, Chem. Commun. (Camb). 48 (2012) 6203–6216.

[35] S.M. Kerwin, G-Quadruplex DNA as a target for drug design., Curr. Pharm. Des. 6 (2000) 441–478.

[36] G.N. Parkinson, M.P.H. Lee, S. Neidle, Crystal structure of parallel quadruplexes from human telomeric DNA., Nature. 417 (2002) 876–880.

[37] Y. Wang, D.J. Patel, Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex., Structure. 1 (1993) 263–282.

[38] D.-H. Zhang, G.-Y. Zhi, Structure monomorphism of RNA G-quadruplex that is independent of surrounding condition., J. Biotechnol. 150 (2010) 6–10.

[39] P. Schultze, N. V Hud, F.W. Smith, J. Feigon, The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the Oxytricha nova telomere repeat oligonucleotide d(G(4)T(4)G(4))., Nucleic Acids Res. 27 (1999) 3018–28.

[40] 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.

[41] A. Henderson, Y. Wu, Y.C. Huang, E. a. Chavez, J. Platt, F.B. Johnson, et al., Detection of G-quadruplex DNA in mammalian cells, Nucleic Acids Res. 42 (2014) 860–869.

[42] S. Balasubramanian, S. Neidle, G-quadruplex nucleic acids as therapeutic targets, Curr. Opin. Chem. Biol. 13 (2009) 345–353.

[43] Y. Wang, D.J. Patel, Guanine residues in d(T2AG3) and d(T2G4) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution., Biochemistry. 31 (1992) 8112–9.

[44] A.T. Phan, D.J. Patel, Two-Repeat Human Telomeric d(TAGGGTTAGGGT) Sequence Forms Interconverting Parallel and Antiparallel G-Quadruplexes in Solution: Distinct Topologies, Thermodynamic Properties, and Folding/Unfolding Kinetics, J. Am. Chem. Soc. 125 (2003) 15021–15027.

[45] N. Zhang, A.T. Phan, D.J. Patel, (3 + 1) Assembly of Three Human Telomeric Repeats into an Asymmetric Dimeric G-Quadruplex NMR Spectra of Three-Repeat Human Telomeric Se-, J. Am. Chem. Soc. 127 (2005) 17277–17285.

[46] A.T. Phan, K.N. Luu, D.J. Patel, Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K+ solution, Nucleic Acids Res. 34 (2006) 5715–5719.

[47] K.N. Luu, A.T. Phan, V. Kuryavyi, L. Lacroix, D.J. Patel, Structure of the human telomere in K+ solution: An intramolecular (3 + 1) G-quadruplex scaffold, J. Am. Chem. Soc. 128 (2006) 9963–9970.

[48] K.W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D.J. Patel, et al., Structure of the human telomere in K + solution: A stable basket-type G-quadruplex with only two G-tetrad layers, J. Am. Chem. Soc. 131 (2009) 4301–4309.

[49] A. Larsen, H. Weintraub, An altered DNA conformation detected by S1 nuclease occurs at specific regions in active chick globin chromatin, Cell. 29 (1982) 609–622.

[50] S. Balasubramanian, L.H. Hurley, S. Neidle, Targeting G-quadruplexes in gene promoters : a novel anticancer strategy?, Nat. Rev. Drug Discov. 10 (2011) 261–275.

[51] V. González, L.H. Hurley, The c-MYC NHE III(1): function and regulation., Annu. Rev. Pharmacol. Toxicol. 50 (2010) 111–129.

[52] J. Seenisamy, E.M. Rezler, T.J. Powell, D. Tye, V. Gokhale, C.S. Joshi, et al., The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4, J. Am. Chem. Soc. 126 (2004) 8702–8709.

[53] G. Collie, G. Parkinson, The application of DNA and RNA G-quadruplexes to therapeutic medicines., Chem. Soc. Rev. 40 (2011) 5867–92.

[54] D. Yang, L.H. Hurley, Structure of the biologically relevant G-quadruplex in the c-MYC promoter., Nucleosides. Nucleotides Nucleic Acids. 25 (2006) 951–968.

[55] S. Rankin, A.P. Reszka, J. Huppert, M. Zloh, G.N. Parkinson, A.K. Todd, et al., Putative DNA quadruplex formation within the human c-kit oncogene, J. Am. Chem. Soc. 127 (2005) 10584–10589.

[56] H. Sun, J. Xiang, Y. Shi, Q. Yang, A. Guan, Q. Li, et al., A newly identified G-quadruplex as a potential target regulating Bcl-2 expression., Biochim. Biophys. Acta. 1840 (2014) 3052–7.

[57] J. Dai, T.S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R.A. Jones, et al., An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution., J. Am. Chem. Soc. 128 (2006) 1096–8.

[58] N. Maizels, Immunoglobulin gene diversification., Annu. Rev. Genet. 39 (2005) 23–46.

[59] J. Charles A Janeway, P. Travers, M. Walport, M.J. Shlomchik, Immunobiology, 6th edition, Garland Science, 2005.

[60] 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 Dev. 18 (2004) 1618–1629.

[61] D. Monchaud, M.-P. Teulade-Fichou, A hitchhiker's guide to G-quadruplex ligands., Org. Biomol. Chem. 6 (2008) 627–636.

[62] A.M. Burger, F. Dai, C.M. Schultes, A.P. Reszka, M.J. Moore, J.A. Double, et al., The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function., Cancer Res. 65 (2005) 1489–96.

[63] M.-Y. Kim, H. Vankayalapati, K. Shin-Ya, K. Wierzba, L.H. Hurley, Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular g-quadruplex., J. Am. Chem. Soc. 124 (2002) 2098–9.

[64] H. Zhang, X. Wang, P. Wang, S. Pang, X. Ai, J. Zhang, Interactions between meso-tetrakis(4-(N-methylpyridiumyl))porphyrin TMPyP4 and DNA G-quadruplex of telomeric repeated sequence TTAGGG, Sci. China, Ser. B Chem. 51 (2008) 452–456.

[65] D. Yang, K. Okamoto, Structural insights into G-quadruplexes: towards new anticancer drugs., Future Med. Chem. 2 (2010) 619–646.

[66] T. Paris, J. Vigneron, J. Lehn, Molecular Recognition of anionic substrates. Crystal structures of the supramolecular inclusion complexes of terephthalate and isophthalate dianions with a bis-intercaland receptor molecule, J. Incl. Phenom. Macrocycl. Chem. 33 (1999) 191–202.

[67] S.L. Cree, M. a. Kennedy, Relevance of G-quadruplex structures to pharmacogenetics, Front. Pharmacol. 5 (2014) 4–11.

[68] D. Drygin, A. Siddiqui-Jain, S. O'Brien, M. Schwaebe, A. Lin, J. Bliesath, et al., Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis., Cancer Res. 69 (2009) 7653–61.

[69] P.J. Bates, D. a. Laber, D.M. Miller, S.D. Thomas, J.O. Trent, Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer, Exp. Mol. Pathol. 86 (2009) 151–164.

[70] J.E. Rosenberg, R.M. Bambury, E.M. Van Allen, H. a. Drabkin, P.N. Lara, A.L. Harzstark, et al., A phase II trial of AS1411 (a novel nucleolin-targeted DNA aptamer) in metastatic renal cell carcinoma, Invest. New Drugs. 32 (2014) 178–187.

[71] S. Paramasivan, I. Rujan, P.H. Bolton, Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding, Methods. 43 (2007) 324–331.

[72] J. Amato, N. Iaccarino, B. Pagano, R. Morigi, A. Locatelli, A. Leoni, et al., Bis-indole derivatives with antitumor activity turn out to be specific ligands of human telomeric G-quadruplex., Front. Chem. 2 (2014) 54.

[73] P. a. Rachwal, K.R. Fox, Quadruplex melting, Methods. 43 (2007) 291–301.

[74] J.-L. Mergny, A.-T. Phan, L. Lacroix, Following G-quartet formation by UV-spectroscopy, FEBS Lett. 435 (1998) 74–78.

[75] J. Jaumot, R. Gargallo, Experimental Methods for Studying the Interactions between G-Quadruplex Structures and Ligands, Curr. Pharm. Des. 18 (2012) 1900–1916.

[76] D.S.H. Chan, H. Yang, M.H.T. Kwan, Z. Cheng, P. Lee, L.P. Bai, et al., Structure-based optimization of FDA-approved drug methylene blue as a c-myc G-quadruplex DNA stabilizer, Biochimie. 93 (2011) 1055–1064.

[77] A. De Cian, L. Guittat, M. Kaiser, B. Saccà, S. Amrane, A. Bourdoncle, et al., Fluorescence-based melting assays for studying quadruplex ligands, Methods. 42 (2007) 183–195.

[78] D. Monchaud, C. Allain, M.-P. Teulade-Fichou, Development of a fluorescent intercalator displacement assay (G4-FID) for establishing quadruplex-DNA affinity and selectivity of putative ligands, Bioorganic Med. Chem. Lett. 16 (2006) 4842–4845.

[79] D. Monchaud, C. Allain, H. Bertrand, N. Smargiasso, F. Rosu, V. Gabelica, et al., Ligands playing musical chairs with G-quadruplex DNA: A rapid and simple displacement assay for identifying selective G-quadruplex binders, Biochimie. 90 (2008) 1207–1223.

[80] P.L.T. Tran, E. Largy, F. Hamon, M.-P. Teulade-Fichou, J.L. Mergny, Fluorescence intercalator displacement assay for screening G4 ligands towards a variety of G-quadruplex structures, Biochimie. 93 (2011) 1288–1296.

[81] S. Di Micco, C. Bassarello, G. Bifulco, R. Riccio, L. Gomez-Paloma, Differential-frequency saturation transfer difference NMR spectroscopy allows the detection of different ligand-DNA binding modes, Angew. Chemie - Int. Ed. 45 (2005) 224–228.

[82] B. Meyer, T. Peters, NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors, Angew. Chem. Int. Ed. 42 (2003) 864-890.

[83] M. Mayer, B. Meyer, Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor, J. Am. Chem. Soc. 123 (2001) 6108–6117.

[84] H. Han, L.H. Hurley, M. Salazar, A DNA polymerase stop assay for G-quadruplex-interactive compounds, Nucleic Acids Res. 27 (1999) 537–542.

[85] H. Lodish, A. Berk, C.A. Kaiser, M. Krieger, M.P. Scott, A. Bretscher, et al., DNA Cloning with Plasmid Vectors, in: Mol. Cell Biol., 6th edition, W. H. Freeman, New York, 2007.

[86] S.N. Cohen, A.C. Chang, H.W. Boyer, R.B. Helling, Construction of biologically functional bacterial plasmids *in vitro*, Proc. Natl. Acad. Sci. U. S. A. 70 (1973) 3240–4.

[87] J.M. Berg, J.L. Tymoczko, L. Stryer, Manipulating the Genes of Eukaryotes, in: Biochemistry, 5th edition, W H Freeman, New York, 2002: pp. 256–266.

[88] Z. Zhong, L. Shiue, S. Kaplan, T. de Lange, A mammalian factor that binds telomeric TTAGGG repeats in vitro., Mol. Cell. Biol. 12 (1992) 4834–4843.

[89] B. Lv, D. Li, H. Zhang, J.Y. Lee, T. Li, DNA gyrase-driven generation of a G-quadruplex from plasmid DNA., Chem. Commun. (Camb). 49 (2013) 8317–9.

[90] M.D. Huber, M.L. Duquette, J.C. Shiels, N. Maizels, A Conserved G4 DNA Binding Domain in RecQ Family Helicases, J. Mol. Biol. 358 (2006) 1071–1080.

[91] J.M. Schmitter, Y. Mechulam, G. Fayat, M. Anselme, Rapid purification of DNA fragments by high-performance size-exclusion chromatography., J. Chromatogr. 378 (1986) 462–466.

[92] K. Kasai, Size-dependent chromatographic separation of nucleic acids, J. Chromatogr. - Biomed. Appl. 618 (1993) 203–221.

[93] S.C. Hardies, R.D. Wells, Preparative fractionation of DNA restriction fragments by reverse phase column chromatography ., Proc. Natl. Acad. Sci. U. S. A. 73 (1976) 3117–3121.

[94] H. Ellegren, T. Låås, Size-exclusion chromatography of DNA restriction fragments. Fragment length determinations and a comparison with the behaviour of proteins in size-exclusion chromatography., J. Chromatogr. 467 (1989) 217–226.

[95] C.M.C. Tellez, K.D.K. Cole, Method for the characterization of size-exclusion chromatography media for preparative purification of DNA restriction fragments, Biotechnol. Tech. 13 (1999) 395–401.

[96] M.D. Abràmoff, P.J. Magalhães, S.J. Ram, Image processing with imageJ, Biophotonics Int. 11 (2004) 36–41.

[97] R. Siegel, J. Ma, Z. Zou, A. Jemal, Cancer Statistics , 2014, CA. Cancer J. Clin. 64 (2014) 9–29.

[98] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool., J. Mol. Biol. 215 (1990) 403–10.

[99] Amersham Biosciences, Gel Filtration: Principles and Methods, 2002.

[100] Q.J. Li, X.J. Tong, Y.M. Duan, J.Q. Zhou, Characterization of the intramolecular G-quadruplex promoting activity of Est1, FEBS Lett. 587 (2013) 659–665.

[101] W. Tan, H. Chen, J. Zhou, M. Xu, G. Yuan, Induction of formation and conformational conversion of DNA G-quadruplexes by fangchinoline, Anal. Methods. 6 (2014) 8476-8481.

[102] P. Groves, M. Webba Da Silva, Rapid stoichiometric analysis of G-quadruplexes in solution, Chem. - A Eur. J. 16 (2010) 6451–6453.

[103] R.D. Gray, J. Li, J.B. Chaires, Energetics and Kinetics of a Conformational Switch in G-Quadruplex DNA, J. Phys. Chem. B. 113 (2009) 2676–2683.

[104] L. Martino, B. Pagano, I. Fotticchia, S. Neidle, C. Giancola, Shedding light on the interaction between TMPyP4 and human telomeric quadruplexes, J. Phys. Chem. B. 113 (2009) 14779–14786.

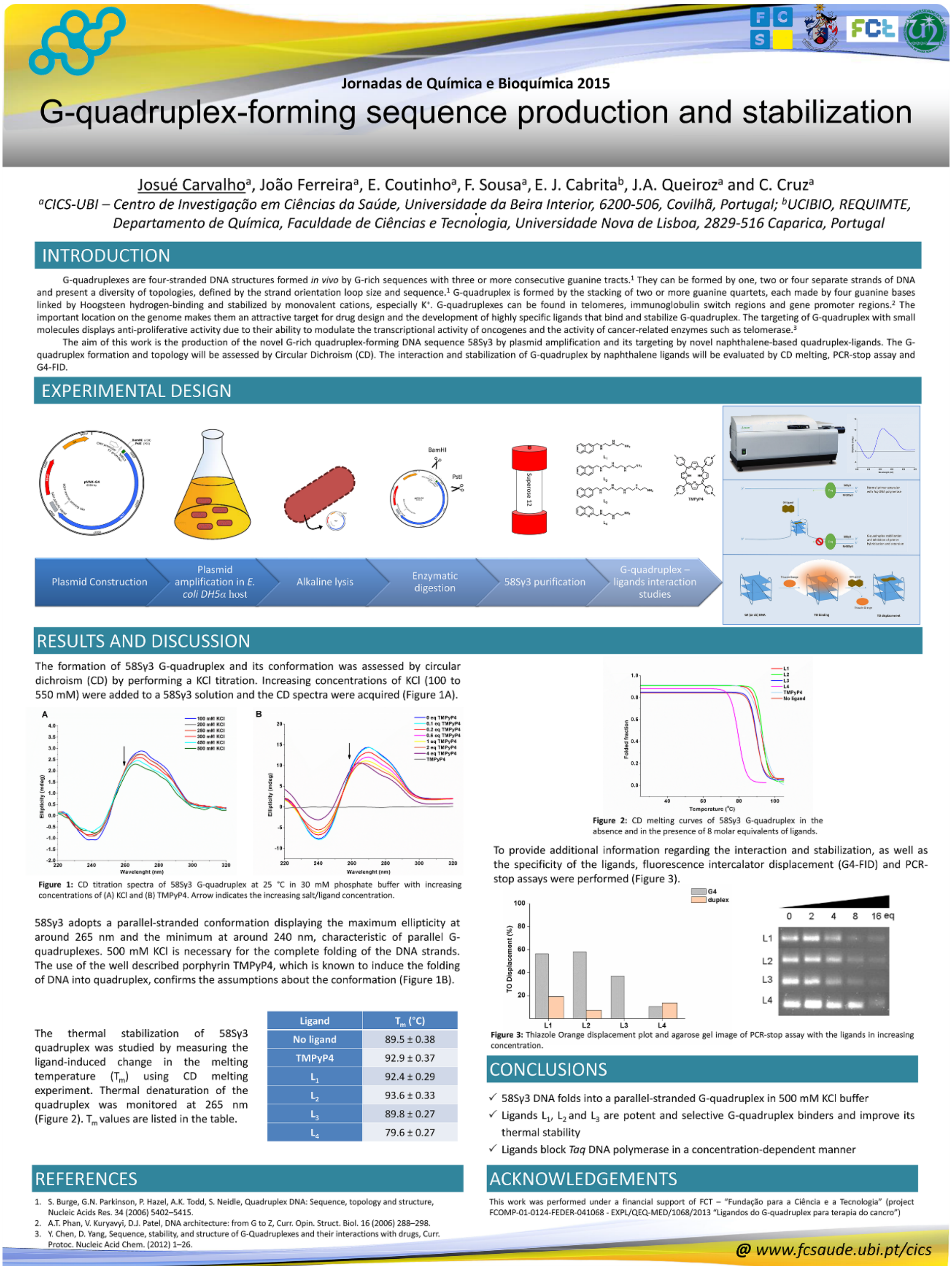
[105] Y.-P. Xing, C. Liu, X.-H. Zhou, H.-C. Shi, Label-free detection of kanamycin based on a G-quadruplex DNA aptamer-based fluorescent intercalator displacement assay, Sci. Rep. 5 (2015) 8125.

[106] D.M. Gray, R.L. Ratliff, M.R. Vaughan, Circular dichroism spectroscopy of DNA, Methods Enzymol. 211 (1992) 389–406.

[107] T. Lemarteleur, D. Gomez, R. Paterski, E. Mandine, P. Mailliet, J.F. Riou, Stabilization of the c-myc gene promoter quadruplex by specific ligands’ inhibitors of telomerase, Biochem. Biophys. Res. Commun. 323 (2004) 802–808.

Appendices

Appendix A



Jornadas de Química e Bioquímica 2015

G-quadruplex-forming sequence production and stabilization

Josué Carvalho[a], João Ferreira[a], E. Coutinho[a], F. Sousa[a], E. J. Cabrita[b], J.A. Queiroz[a] and C. Cruz[a]
[a]*CICS-UBI – Centro de Investigação em Ciências da Saúde, Universidade da Beira Interior, 6200-506, Covilhã, Portugal;* [b]*UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal*

INTRODUCTION

G-quadruplexes are four-stranded DNA structures formed *in vivo* by G-rich sequences with three or more consecutive guanine tracts.[1] They can be formed by one, two or four separate strands of DNA and present a diversity of topologies, defined by the strand orientation loop size and sequence.[1] G-quadruplex is formed by the stacking of two or more guanine quartets, each made by four guanine bases linked by Hoogsteen hydrogen-binding and stabilized by monovalent cations, especially K$^+$. G-quadruplexes can be found in telomeres, immunoglobulin switch regions and gene promoter regions.[2] The important location on the genome makes them an attractive target for drug design and the development of highly specific ligands that bind and stabilize G-quadruplex. The targeting of G-quadruplex with small molecules displays anti-proliferative activity due to their ability to modulate the transcriptional activity of oncogenes and the activity of cancer-related enzymes such as telomerase.[3]

The aim of this work is the production of the novel G-rich quadruplex-forming DNA sequence 58Sγ3 by plasmid amplification and its targeting by novel naphthalene-based quadruplex-ligands. The G-quadruplex formation and topology will be assessed by Circular Dichroism (CD). The interaction and stabilization of G-quadruplex by naphthalene ligands will be evaluated by CD melting, PCR-stop assay and G4-FID.

EXPERIMENTAL DESIGN

Plasmid Construction → Plasmid amplification in *E. coli DH5α* host → Alkaline lysis → Enzymatic digestion → 58Sγ3 purification → G-quadruplex – ligands interaction studies

RESULTS AND DISCUSSION

The formation of 58Sγ3 G-quadruplex and its conformation was assessed by circular dichroism (CD) by performing a KCl titration. Increasing concentrations of KCl (100 to 550 mM) were added to a 58Sγ3 solution and the CD spectra were acquired (Figure 1A).

Figure 1: CD titration spectra of 58Sγ3 G-quadruplex at 25 °C in 30 mM phosphate buffer with increasing concentrations of (A) KCl and (B) TMPyP4. Arrow indicates the increasing salt/ligand concentration.

58Sγ3 adopts a parallel-stranded conformation displaying the maximum ellipticity at around 265 nm and the minimum at around 240 nm, characteristic of parallel G-quadruplexes. 500 mM KCl is necessary for the complete folding of the DNA strands. The use of the well described porphyrin TMPyP4, which is known to induce the folding of DNA into quadruplex, confirms the assumptions about the conformation (Figure 1B).

The thermal stabilization of 58Sγ3 quadruplex was studied by measuring the ligand-induced change in the melting temperature (T_m) using CD melting experiment. Thermal denaturation of the quadruplex was monitored at 265 nm (Figure 2). T_m values are listed in the table.

Ligand	T_m (°C)
No ligand	89.5 ± 0.38
TMPyP4	92.9 ± 0.37
L_1	92.4 ± 0.29
L_2	93.6 ± 0.33
L_3	89.8 ± 0.27
L_4	79.6 ± 0.27

Figure 2: CD melting curves of 58Sγ3 G-quadruplex in the absence and in the presence of 8 molar equivalents of ligands.

To provide additional information regarding the interaction and stabilization, as well as the specificity of the ligands, fluorescence intercalator displacement (G4-FID) and PCR-stop assays were performed (Figure 3).

Figure 3: Thiazole Orange displacement plot and agarose gel image of PCR-stop assay with the ligands in increasing concentration.

CONCLUSIONS

- ✓ 58Sγ3 DNA folds into a parallel-stranded G-quadruplex in 500 mM KCl buffer
- ✓ Ligands L_1, L_2 and L_3 are potent and selective G-quadruplex binders and improve its thermal stability
- ✓ Ligands block *Taq* DNA polymerase in a concentration-dependent manner

REFERENCES

1. S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, Quadruplex DNA: Sequence, topology and structure, Nucleic Acids Res. 34 (2006) 5402–5415.
2. A.T. Phan, V. Kuryavyi, D.J. Patel, DNA architecture: from G to Z, Curr. Opin. Struct. Biol. 16 (2006) 288–298.
3. Y. Chen, D. Yang, Sequence, stability, and structure of G-Quadruplexes and their interactions with drugs, Curr. Protoc. Nucleic Acid Chem. (2012) 1–26.

ACKNOWLEDGEMENTS

This work was performed under a financial support of FCT – "Fundação para a Ciência e a Tecnologia" (project FCOMP-01-0124-FEDER-041068 - EXPL/QEQ-MED/1068/2013 "Ligandos do G-quadruplex para terapia do cancro")

@ www.fcsaude.ubi.pt/cics

Appendix B

5th International Meeting on Quadruplex Nucleic Acids

G-quadruplex production and stabilization

João Medeiros-Silva[a,b †], Josué Carvalho[a †], Eurico J. Cabrita[b] and Carla Cruz[a]
[a]*CICS-UBI – Centro de Investigação em Ciências da Saúde, Universidade da Beira Interior, 6200-506, Covilhã, Portugal;*
[b]*UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal.*

Abstract

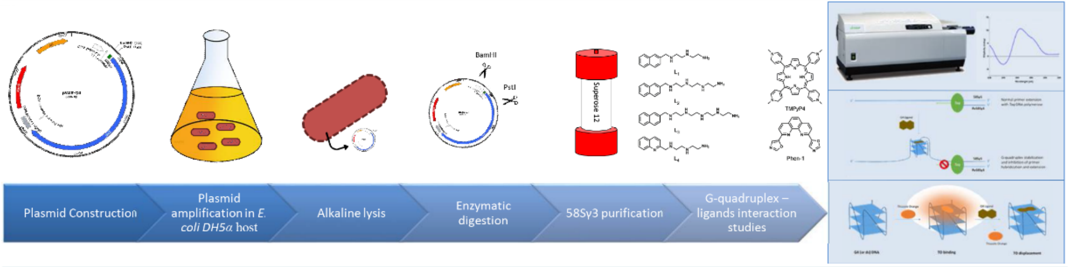
The aim of this work is the production of a novel G-rich quadruplex-forming DNA sequence 58Sγ3 through plasmid amplification and its targeting by novel phenanthroline and naphthalene-based quadruplex-ligands.

The putative G-quadruplex forming sequence 58Sγ3 was selected based on QGRS mapper[1] software results and was produced by cloning the sequence into pVAX-*LacZ* plasmid and carrying autonomous replication in *E. coli DH5α* host.

The G-quadruplex formation for the sequence 58Sγ3 was assessed by Circular Dichroism (CD), witch indicates a parallel topology. A further study with TMPyP4 supports these assumptions.

The interaction and stabilization of this G-quadruplex with novel naphthalene and phenanthroline-based ligands were evaluated by CD melting, PCR-stop assay and G4-FID. Two naphthalene ligands were found to stabilize the G-4 topology up to 4ºC. In it's turn quinoline-based ligands, in the same conditions, showed to destabilize the 58Sγ3 adopted structure while phenantroline(1,10)dioxazole did not interact. These results enhance the importance of the disposition of the aliphatic chain and amine groups.

EXPERIMENTAL DESIGN



Plasmid Construction → Plasmid amplification in *E. coli DH5α* host → Alkaline lysis → Enzymatic digestion → 58Sγ3 purification → G-quadruplex – ligands interaction studies

RESULTS AND DISCUSSION

The formation of 58Sγ3 G-quadruplex and its conformation was assessed by CD through KCl titration. Increasing concentrations of KCl (100 to 500 mM) were added to a 58Sγ3 solution and the CD spectra were acquired (Figure 1A).

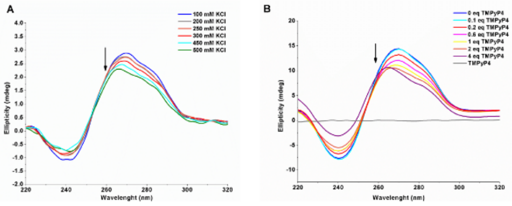


Figure 1: CD titration spectra of 58Sγ3 G-quadruplex at 25 °C in 30 mM phosphate buffer with increasing concentrations of (A) KCl and (B) TMPyP4, 0 mM KCl. Arrow indicates the increasing salt/ligand concentration.

58Sγ3 appears to adopt a parallel-stranded conformation as its spectra displays an ellipticity maximum at 265 nm and a minimum at 240 nm, characteristic of parallel G-quadruplexes[2]. 500 mM KCl is necessary for the complete folding of the DNA strands. The use of the porphyrin TMPyP4, which is known to induce the folding of DNA into quadruplex[3], confirms the assumptions about the conformation of G4 (Figure 1B).

The thermal stabilization of 58Sγ3 quadruplex was studied by measuring the ligand-induced change in the melting temperature (T_m) using CD melting experiment. Thermal denaturation of the quadruplex was monitored at 265 nm (Figure 2). T_m values are listed in the table. Due to the poor solubility of Phen-1 in buffer, the experiments were conducted in 20% DMSO.

Ligand	T_m (°C)	ΔT_m (°C)
No ligand	89.5 ± 0.38	-
TMPyP4	92.9 ± 0.37	3.4
L_1	92.4 ± 0.29	2.9
L_2	93.6 ± 0.33	4.1
L_3	89.8 ± 0.27	0.3
L_4	79.6 ± 0.27	-9.9
No ligand + 20% DMSO	81.9 ± 0.13	-
Phen-1	80.9 ± 0.16	-1.0

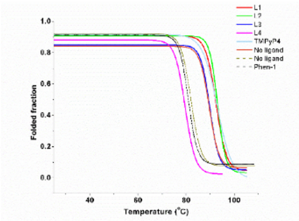


Figure 2: CD melting curves of 58Sγ3 G-quadruplex in the absence and in the presence of 8 molar equivalents of ligands. Dash lines indicate the experiments was performed in 20% DMSO.

To provide additional information regarding the interaction and stabilization, as well as the specificity of the ligands over duplex, fluorescence intercalator displacement (G4-FID) and PCR-stop assays were performed (Figure 3)

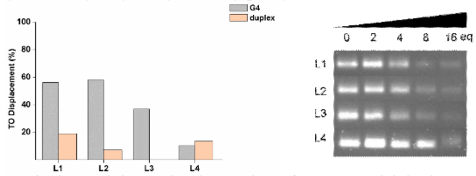


Figure 3: Thiazole Orange displacement plot and agarose gel image of PCR-stop assay with the ligands in increasing concentration of ligand. Phen-1 was not included in this study given that no interaction was observed.

CONCLUSIONS

- ✓ Present studies indicate that 58Sγ3 DNA folds into a parallel-stranded G-quadruplex in 500 mM KCl buffer;
- ✓ Ligands L_1, L_2 and L_3 are potent and selective G-quadruplex binders and improve 58Sγ3's G-4 thermal stability, while Phen-1 doesn't interact and L_4 revealed to destabilize it;
- ✓ Ligands block *Taq* DNA polymerase in a concentration-dependent manner.

REFERENCES

1. Kikin, O., D'Antonio, L. and Bagga, S.; QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences., Nucleic Acids Research, (2006) Vol. 34, w676-w682;
2. Huppert, J. L., Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes, Chem Soc Rev. 2008 Jul;37(7):1375-84;
3. Gray, R. D., Li, J. and Chaires , J. B., Energetics and Kinetics of a Conformational Switch in G-Quadruplex DNA, J. Phys. Chem. B, 2009, 113 (9), pp 2676–2683.

ACKNOWLEDGEMENTS

This work was performed under a financial support of FCT – "Fundação para a Ciência e a Tecnologia" (project FCOMP-01-0124-FEDER-041068 - EXPL/QEQ-MED/1068/2013 "Ligandos do G-quadruplex para terapia do cancro")
João Medeiros Silva acknowledge the fellowship BI-2-EXPL/QEQ-MED/1068/2013.