



**João Medeiros Silva**

Licenciado em Bioquímica

## **G-quadruplex ligands for cancer therapy**

**Orientador:** Doutora Carla Patrícia Cruz, Professor Auxiliar Convidado,  
Faculdade de Ciências da Saúde - Universidade da Beira interior

**Co-orientador:** Doutor Eurico José Cabrita, Professor Auxiliar,  
Faculdade de Ciências e Tecnologias - Universidade Nova de Lisboa

**Júri:**

**Presidente:** Professor Doutor José Ricardo Ramos Franco Tavares  
Professor Auxiliar, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

**Vogal:** Professora Doutora Carla Patrícia Alves Freire Madeira Cruz,  
Professor Auxiliar Convidado, Faculdade de Ciências da Saúde, Universidade da Beira interior

**Vogal arguente:** Professor Doutor João Carlos dos Santos Silva e Pereira de Lima  
Professor Associado, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

**Setembro, 2015**



FACULDADE DE  
CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE NOVA DE LISBOA

## **G-quadruplex ligands for cancer therapy**

Copyright © João Medeiros Silva, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

*"If A is a success in life, then A equals x plus y plus z.  
Work is x; y is play, and z is keeping your mouth shut."*

Albert E, Quoted in *Observer*, 15 Jan 1950.





# Agradecimentos

---

Em primeiro lugar quero agradecer profundamente aos meus orientadores. Ao professor Eurico por constantemente investir e acreditar em mim. À professora Carla pelo apoio incondicional tanto a nível profissional como pessoal.

Gostaria de agradecer também ao Dr. Samuel Silvestre, Dr<sup>a</sup> Graça Baltazar e Dr. Jean-Louis Mergny a amabilidade de colaborarem com amostras para a realização dos testes com os ligandos.

Quero também agradecer aos meus pais, aos meus irmãos, aos meus padrinhos e a minha avó pelo apoio incessável e pelo estímulo que me levou até aqui.

Quero também agradecer aos meus colegas (já agora) pelo apoio no dia-a-dia, pela boa disposição que me conferiram e por serem uns chatos do caneco. Segue a longa lista:

- Sara Cegonho (obrigado pela boleia)
- Elisabete Alves
- Sandrina Maçãs
- Filipa Gomes
- Sérgio Rodrigues
- Fafe
- Mariana Matias
- Carolina Costa
- Cristina Lemos
- Josué (esse gajo)
- Micael Simões
- Diogo Poeira
- Ricardo Martinho
- Ana Dinis
- Ana Diniz
- Ana Catarina
- Inês flores
- Inês Almeida
- Inês Rosete
- Ângelo
- Ana Sofia
- Mariana Oliveira
- Tiago Paiva
- Wagner
- Tiago Páscoa
- Filipa Marcelo
- Marta Corvo
- Helena
- Maria
- André das Silvas
- Peço desculpa a quem me esqueci

E finalmente quero agradecer à Sara pelo apoio, paciência e carinho imenso todos os dias, todo o ano. Adoro-te.

This work was supported by Fundação para a Ciência e Tecnologia through project no. FCOMP-01-0124-FEDER-041068 – EXPL/QEQ-MED/1068/2013. João Silva acknowledges the fellowship BI-2-EXPL/QEQ-MED/1068/2013.

---



# Resumo

---

As moléculas de ADN apresentam um elevado polimorfismo associado às estruturas que podem adotar. As estruturas em G-quadruplex (G4) têm interesse biológico associado, uma vez que estas encontram-se nas extremidades dos telómeros e em alguns promotores de oncogenes. Consequentemente, sequências desta natureza têm sido exploradas como alvos terapêuticos na terapia do cancro, no entanto as estruturas de G4 são transientes e instáveis. Portanto, o principal objetivo deste trabalho será desenvolver ligandos específicos do G4 com potencial terapêutico.

Vários potenciais ligandos do G4 foram sintetizados e caracterizados. O processo sintético consistiu primeiramente na formação de anéis oxazole através da reação de van Leusen seguido de um acoplamento cruzado com halogenetos de piridina. Deste modo foram sintetizados derivados de fenantrolina (Phen-1, 50%; Phen-2, 20%), quinolina (Quin-1, 85%; Quin-2, 45%) e fenilo (Iso-1, 61%; Iso-2, 21%; Ter-1, 85%; Ter-2, 35%).

Foram efetuados testes preliminares com a técnicas de FRET e dicroísmo circular (DC), G4FID e DSF de modo a selecionar os ligandos com afinidade para as estruturas em G4. Estudos biofísicos qualitativos foram efetuados através de espectroscopia de fluorescência e DC. Dois dos ligandos testados, Phen-1 e Phen-2, revelaram elevada especificidade para os G4 22AG (telomérico) e *c-myc* (promotor do *c-myc*).

Verificou-se que o ligando Phen-1 estabilizou termicamente as sequências 22AG e *c-myc* em cerca de 4.1 e 4.3 °C. Por sua vez, o Phen-2 apresentou elevada afinidade para as sequências 22AG ( $K_a = 9.56 \times 10^9 M^{-1}$ ) e *c-myc* ( $K_a = 3.55 \times 10^6 M^{-1}$ ), aumentando as respetivas estabilidades térmicas em cerca de 15.0 (em K<sup>+</sup>) e 31.0 °C.

Os compostos foram biologicamente avaliados acerca do seu perfil citotóxico e antiproliferativo através do teste de MTT em três linhas celulares cancerígenas (MCF-7, LNCaP e U87) e uma linha saudável (NHDF). Dos ligandos testados destacam-se o Phen-2 e Quin-2 que apresentaram efeitos citotóxicos nas linhas celulares LNCaP (IC<sub>50</sub> = 0.40 e 39.14 μM, respetivamente) e MCF-7 (IC<sub>50</sub> = 0.64 e 4.17 μM, respetivamente). Para além disso, o ligando Quin-2 não apresentou citotoxicidade nas linhas U87 e NHDF embora o seu mecanismo antiproliferativo não esteja em princípio relacionado com o G4.

**Palavras-chave:** ADN; G-quadruplex; compostos oligoheteroarilos; reconhecimento molecular; compostos anticancerígenos; interações recetor-ligando.

---



# Abstract

---

DNA may fold into a diversity of structures and topologies such as duplexes and triplexes. Some specific guanine-rich DNA sequences may even fold into a higher order structures denominated guanine G-quadruplexes (G4). These G-quadruplex forming sequences have shown biological interest since were found in telomeres and in promoter region of oncogenes. Thus, these G4 forming sequences have been explored as therapeutic targets for cancer therapy, since G4 formation was demonstrated to inhibit RNA-polymerase and telomerase activity. However, the G4 structures are transient and are only formed under specific conditions. Hence the main objective of this work is to develop new G4-specific ligands which may potentially find applications in the therapeutic area.

Several potential G4-binding ligands were synthesized and characterized. The synthesis of these compounds consisted on a procedure based on van Leusen chemistry and a cross-coupling reaction through C-H activation, affording phenanthroline compounds (Phen-1, 50%; Phen-2, 20%), phenyl (Iso-1, 61%; Iso-2, 21%; Ter-1, 85%; Ter-2, 35%), and quinolyl (Quin-1, 85%; Quin-2, 45%) compounds.

Screening assays for selecting the potential G4 compounds were performed by FRET-melting, G4-FID, CD-melting and DSF. Qualitative biophysical studies were performed by fluorescence and CD spectroscopy. Two high-specific G-quadruplex ligands, Phen-1 and Phen-2, were found to effectively bind telomeric and *c-myc* G4 structures. Phen-1 was found to stabilize parallel telomeric 22AG and *c-myc* sequence by 4.1 and 4.3 °C, respectively. Phen-2 also displayed high affinity towards 22AG ( $K_a = 9.56 \times 10^9 M^{-1}$ ) and to *c-myc* ( $K_a = 3.55 \times 10^6 M^{-1}$ ), increasing their thermal stability by 15.0 (in K<sup>+</sup>) and 31.0 °C, respectively.

The compounds were evaluated concerning their anti-proliferative effects on three cancer cell lines (MCF-7, LNCaP and U87) and normal cell line (NHDF), by MTT assay. Phen-2 and Quin-2 displayed strong anti-proliferative effects on LNCaP (IC<sub>50</sub> = 0.40 and 39.14 μM, respectively) and MCF-7 (IC<sub>50</sub> = 0.64 and 4.17 μM, respectively) cancer cell lines. Furthermore Quin-2 did not display cytotoxic effects on U87 and normal NHDF cells.

Overall, this work explored new possibilities for finding new G4 ligands for cancer therapy.

**Keywords:** DNA; G-quadruplex; oligoheteroaryle compounds; molecular recognition; anti-cancer drugs; ligand-receptor interaction.

---



# Contents

---

AGRADECIMENTOS .....	V
RESUMO .....	VII
ABSTRACT .....	IX
CONTENTS .....	XI
LIST OF FIGURES .....	XIII
LIST OF TABLES .....	XV
NOMENCLATURE .....	XVII
<b>I.   INTRODUCTION .....</b>	<b>1</b>
1.1 Nucleic Acids: historic landmarks and general propERTies.....	1
1.2 Guanine quadruplexes.....	5
1.3 G-quadruplex sequences as therapeutic targets .....	16
1.4 G4 Ligands for cancer therapy .....	19
1.5 Objectives and outline of this thesis.....	21
<b>II.   RESULTS AND DISCUSSION .....</b>	<b>23</b>
PART A .....	25
OLIGOHETEROARYLE G-QUADRUPLEX LIGANDS: DESIGN, SYNTHESIS AND CHARACTERIZATION.....	25
<b>1 – Introduction</b> .....	25
<b>2 – Experimental</b> .....	26
<b>3 – Results and discussion</b> .....	35
<b>4 – Conclusions</b> .....	43
PART B .....	45
OLIGOHETEROARYLE G-QUADRUPLEX LIGANDS: SCREENING AND INTERACTION .....	45
<b>1 – Introduction</b> .....	45
<b>2 – Experimental</b> .....	47
<b>3 – Results and discussion</b> .....	52
<b>4 – Conclusions</b> .....	64
PART C.....	65
OLIGOHETEROARYLE G-QUADRUPLEX LIGANDS: BIOLOGICAL EVALUATION.....	65
<b>1 – Introduction</b> .....	65
<b>2 – Experimental</b> .....	67
<b>3 – Results and discussion</b> .....	69
<b>4 – Conclusions</b> .....	72
<b>III.   FINAL CONCLUSIONS .....</b>	<b>73</b>
Concluding remarks .....	73
Future perspectives.....	74
Communications related to this work .....	75
<b>IV.   BIBLIOGRAPHY .....</b>	<b>77</b>
APPENDIX A .....	95
APPENDIX B .....	115
APPENDIX C .....	121

---





# List of figures

---

<b>FIGURE I.1</b> - LEFT PANEL: NATURAL BUILDING BLOCKS OF NUCLEIC ACIDS, THE NUCLEOTIDES. A NUCLEOTIDE IS MADE UP OF A PHOSPHATE GROUP, A SUGAR (RIBOSE FOR RNA OR DEOXYRIBOSE FOR DNA), AND A NITROGENOUS BASE (A, T, C, G, U).....	1
<b>FIGURE I.2</b> - DETAILED MOLECULAR STRUCTURE OF DNA AND RNA COMPONENTS.....	2
<b>FIGURE I.3</b> – THREE DIFFERENT CONFORMATIONS OF DNA’S DOUBLE-HELICAL STRUCTURE. (A) A-DNA IS A SHORT, WIDE, RIGHT-HANDED HELIX.....	3
<b>FIGURE I.4</b> – A) ILLUSTRATION OF THE SYN-TO-ANTI CONVERSION BY APPLYING A 180° ROTATION AROUND THE GLYCOSYLIC BOND.....	4
<b>FIGURE I.5</b> – DEOXYGUANOSINE WITH ATTACHED PHOSPHOROUS ATOMS IS SHOWN IN THE TWO MAJOR NUCLEIC ACID SUGAR PUCKERS.....	5
<b>FIGURE I.6</b> – LEFT: CHEMICAL FORMULA OF CANONICAL GUANINE, REPRESENTING OF THE TWO HYDROGEN BONDING EDGES (WATSON–CRICK AND HOOGSTEN) IMPLICATED IN G-QUARTET FORMATION. ARROWS INDICATE H-BOND ACCEPTORS (YELLOW) AND DONORS (GREEN). .....	6
<b>FIGURE I.7</b> – G-QUADRUPLEXES MAY BE UNI-, BI- OR TETRAMOLECULAR STRUCTURES. EACH SEPARATE STRAND IS IDENTIFIED BY A DIFFERENT COLOR. TO CURRENT KNOWLEDGE, ONLY HOMODIMERS AND HOMOTETRAMERS WERE IDENTIFIED.....	7
<b>FIGURE I.8</b> – REPRESENTATION OF THE POSSIBLE ARRANGEMENTS FOR STRAND ORIENTATIONS FOR G <sub>4</sub> .....	8
<b>FIGURE I.9</b> – ILLUSTRATION OF THE MOST RECURRENT TYPE OF LOOPS. PROPELLER (OR DOUBLE-CHAIN-REVERSAL) LOOPS LINK PARALLEL STRANDS AND THE LATERAL AND DIAGONAL LOOPS CONNECT ANTIPARALLEL STRANDS. ....	8
<b>FIGURE I.10</b> – SCHEMATIC DRAWING ELUCIDATING THE HYPOTHESIZED MIXED-TYPE STRUCTURE (A) OF THE HIGHER-ORDER G-QUADRUPLEX OF THE TELOMERIC OVERHANG REGION.....	11
<b>FIGURE I.11</b> – (A) PROPOSED MODEL FOR DUPLEX-QUADRUPLEX EQUILIBRIUM IN SHORT SEQUENCES. THE COMPLEMENTARY STRAND DOES NOT NECESSARILY FORM AN I-MOTIF. ....	12
<b>FIGURE I.12</b> – A) AT LEFT A GENERAL EXAMPLE OF A CILIATE MICROORGANISM, EVIDENCING THE MACRONUCLEUS. AT RIGHT THE CONFOCAL MICROSCOPY IMAGE OF <i>STYLONYCHIA LEMNAE</i> ’S MACRONUCLEUS SHOWING IN SITU STAINING OF ANTIBODIES AGAINST THE ANTIPARALLEL TELOMERIC G-QUADRUPLEX STRUCTURE .....	13
<b>FIGURE I.13</b> – SUMMARY OF THE POSSIBLE DIFFERENT UNIMOLECULAR G <sub>4</sub> TOPOLOGIES.....	14
<b>FIGURE I.14</b> – SCHEMATIC VIEW OF SOME OF THE POSSIBLE DIFFERENT INTERMOLECULAR G <sub>4</sub> TOPOLOGIES IDENTIFIED FOR SMALL SEQUENCES OF TELOMERIC DNA.....	15
<b>FIGURE I.15</b> – SCHEMATIC VIEW OF SOME OF THE POSSIBLE DIFFERENT INTERMOLECULAR G <sub>4</sub> TOPOLOGIES IDENTIFIED FOR SMALL SEQUENCES OF PROTO-ONCOGENES.....	16
<b>FIGURE I.16</b> – SCHEMATIC REPRESENTATION OF THE PATHWAY OF EVENTS THAT OCCUR DURING THE LIFESPAN OF A REGULAR SOMATIC CELL, GOING FROM PROGRESSIVE SHORTENING OVER TIME UNTIL A TRANSFORMATION EVENT WHERE TELOMERASE ACTIVITY IS ENHANCED .....	18
<b>FIGURE I.17</b> – REPRESENTATION OF A LIGAND-G-QUADRUPLEX COMPLEX WITH A) EXTERNAL STACKING MODE ON THE TERMINAL G-TETRAD.....	20
<b>FIGURE I.18</b> – SCHEME OF A GENERAL G-QUADRUPLEX DRUG-DESIGN PROCEDURE. THIS DISSERTATION FOCUS ON THE SYNTHETIC CHEMISTRY, SCREENING, COMPLEX CHARACTERIZATION AND CELL GROWTH INHIBITIONS STEPS.....	21
<b>FIGURE II.1</b> – CHEMICAL STRUCTURE OF A) TELOMESTATIN AND B) PHENDC <sub>3</sub> . ....	25
<b>FIGURE II.2</b> – A) STRUCTURE OF A G-TETRAD FORM THE PARALLEL H <sub>2</sub> T G-QUADRUPLEX (PDB: 1KF1). .....	35
<b>FIGURE II.3</b> – LIPINSKI’S DRUG-LIKE PARAMETERS. ....	36

<b>FIGURE II.4</b> – SCHEMATIC REPRESENTATION OF FRET-BASED MELTING TEMPERATURE EXPERIMENTS. FÖRSTER RESONANCE ENERGY TRANSFER OCCURS BETWEEN THE FLUOROPHORE (E.G. FLUORESCEIN OR “FAM”) AND THE QUENCHER (E.G. TETRAMETHYLRHODAMINE OR “TAMRA”) .....	45
<b>FIGURE II.5</b> – SCHEMATIC REPRESENTATION OF THE G4-FID ASSAY.....	46
<b>FIGURE II.6</b> – CHEMICAL STRUCTURE OF THE LIGANDS ANALYZED IN THIS STUDY. ....	47
<b>FIGURE II.7</b> – ILLUSTRATION OF THE PLATE LAYOUT FOR A G4FID EXPERIMENT. IN THIS PARTICULAR EXAMPLE TELOMERIC 22AG SEQUENCE AND PHEN-1 AND PHEN-2 WERE APPLIED. ....	49
<b>FIGURE II.8</b> – PLOT OF THE FRET-MELTING RESULTS ON G-QUADRUPLEX STRUCTURES IN 10 mM LI CACODYLATE BUFFER WITH 10 mM KCL + 90 mM LiCl, pH 7.2. ....	52
<b>FIGURE II.9</b> – PLOT OF THE FRET-MELTING RESULTS ON DUPLEX FDXT (0.2 mM) WITH PHEN-1 AND PHEN-2 LIGANDS AT 5.0 mM AND 10.0 mM, IN 10 mM LI CACODYLATE BUFFER WITH 10 mM KCL + 90 mM LiCl, pH 7.2,.....	53
<b>FIGURE II.10</b> – DIAGRAMMATIC BAR REPRESENTATION OF G4-FID FOR PHEN-1 AND PHEN-2 PERFORMED WITH 22AG IN Na <sup>+</sup> (100 mM) AND K <sup>+</sup> (50 mM), C-MYC IN K <sup>+</sup> (50 mM) AND A DUPLEX IN K <sup>+</sup> (50 mM)..	53
<b>FIGURE II.11</b> – (A) THIAZOLE ORANGE FLUORESCENCE SIGNAL ( $\lambda_{exc} = 490$ nm, $\lambda_{emi} = 501$ nm) WITH INCREASING TEMPERATURE IN PRESENCE OF 22AG (0.25 mM), 22AG+PHEN-1 (0.25-2.5 mM). ....	55
<b>FIGURE II.12</b> – REPRESENTATION OF FLUORESCENT TITRATIONS EXPERIMENT OF PHEN-2 ( $\lambda_{exc}=350$ nm) WITH (A) 22AG AND (B) C-MYC G-QUADRUPLEX SEQUENCES, IN 30 mM PHOSPHATE WITH 50 mM K <sup>+</sup> BUFFER. PHEN-1 TITRATION IN SUP. B.3-A,B. ....	55
<b>FIGURE II.13</b> – FLUORIMETRIC TITRATION OF PHEN-1 ( $\lambda_{exc}=330$ nm) AND PHEN-2 ( $\lambda_{exc}=350$ nm) WITH 22AG AND C-MYC. A,C) .....	56
<b>FIGURE II.14</b> – (A) REPRESENTATION OF THE DIFFERENT CD PROFILES OF THE G4 TOPOLOGIES EMPLOYED IN THIS STUDY. ....	57
<b>FIGURE II.15</b> – (A) CD TITRATION SPECTRA OF PHEN-1 (0–7 MOLAR EQUIVALENTS) TO QUADRUPLEX 22AG (10 mM) IN THE PRESENCE OF K <sup>+</sup> (100 mM). (B) HILL PLOT OF THE ELLIPTICITY AT 263 nm WITH ADDITION OF LIGAND.. ....	58
<b>FIGURE II.16</b> – (A) CD TITRATION SPECTRA OF PHEN-2 (0–20 MOLAR EQUIVALENTS) TO QUADRUPLEX H22L (10 mM) IN THE PRESENCE OF K <sup>+</sup> (10 mM).....	60
<b>FIGURE II.17</b> – EFFECT OF PHEN-2 AND PHEN-1 CONCENTRATION ON MELTING TEMPERATURE INCREASE OF 22AG QUADRUPLEX IN 10 mM K <sup>+</sup> , 200 $\mu$ M K <sup>+</sup> AND 100 mM Na <sup>+</sup> AND C-MYC IN 100 mM Li <sup>+</sup> .....	61
<b>FIGURE II.18</b> – THERMODYNAMIC ANALYSIS OF THE MELTING CURVES OF 22AG IN SODIUM AND C-MYC IN LITHIUM WITH 2.0 EQUIVALENTS OF PHEN-2. THE SLOPE OF THE FITTED REGION PROVIDES $\Delta H^\circ/R$ . ....	62
<b>FIGURE II.19</b> – SCHEMATIC REPRESENTATION OF THE TYPICAL CHEMICAL SHIFTS FOR THE IMINE REGIONS OF 3 BASE-PAIRINGS. ....	62
<b>FIGURE II.20</b> – NMR TITRATION OF C-MYC G-QUADRUPLEX WITH PHEN-2 IN 30mM PHOSPHATE BUFFER, 100mM K <sup>+</sup> , AT 25°C WITH 10%D <sub>2</sub> O AT 600MHZ. THE LIGAND/QUADRUPLEX RATIOS ARE SHOWN ALONG THE SIDE OF THE SPECTRA. DNA STRAND CONCENTRATION OF 400mM. ....	63
<b>FIGURE II.21</b> – CHEMICAL STRUCTURES OF THE LIGANDS EVALUATED IN THE FOLLOWING STUDY. ....	66
<b>FIGURE II.22</b> – PLOT OF THE RELATIVE CELL PROLIFERATION OF LNCAP, MCF-7 AND U87 CELL LINES INCUBATED WITH QUIN-2 DURING 72 H EXPOSITION IN CONCENTRATIONS RANGING FROM 0 TO 100 mM, DETERMINED BY THE MTT ASSAY. ....	70
<b>FIGURE II.23</b> – PLOT OF THE RELATIVE CELL PROLIFERATION OF LNCAP, MCF-7 AND U87 CELL LINES INCUBATED WITH QUIN-2 DURING 72 H EXPOSITION IN CONCENTRATIONS RANGING FROM 0 TO 100 mM, DETERMINED BY THE MTT ASSAY. DATA ARE REPRESENTATIVE OF TWO INDEPENDENT EXPERIMENTS FOR EACH CELL LINE. *P<0.05 VERSUS THE CONTROL. ....	71

# List of tables

---

<b>TABLE I.1</b> – GENERAL FORMULAS FOR PREDICTING POTENTIAL G-QUADRUPLEXES. REFERENCES <sup>41–43</sup> .....	7
<b>TABLE I.2</b> – SUMMARY OF THE NOMENCLATURE FOR THE MAIN G-QUADRUPLEX FEATURES. ....	9
<b>TABLE II.1</b> – SUMMARY OF THE REACTION CONDITIONS AND YIELDS OF THE SYNTHESIZED COMPOUNDS. ....	40
<b>TABLE II.2</b> – MAXIMUM EMISSION AND EXCITATION WAVELENGTHS FOR THE POTENTIAL G-QUADRUPLEX LIGANDS.* .....	42
<b>TABLE II.3</b> – OLIGONUCLEIC ACID SEQUENCES EMPLOYED FOR BIOCHEMICAL ASSAYS AND PREFERRED G- QUADRUPLEX TOPOLOGIES .....	47
<b>TABLE II.4</b> - THERMAL STABILITY OF SEVERAL G4 WITH LIGANDS MEASURED BY CD MELTING. ....	61
<b>TABLE II.5</b> – MEAN RELATIVE CELL PROLIFERATION (% OF THE CONTROL) OF THE STUDIED CELLS AFTER 72 H EXPOSITION TO THE COMPOUNDS (30 μM). POTENTIAL ACTIVE COMPOUNDS WERE HIGHLIGHTED. ....	69
<b>TABLE II.6</b> – ESTIMATED IC <sub>50</sub> VALUES (μM) FOR PHEN-2 AND QUIN-2 <sup>A</sup> .....	71

---



# Nomenclature

---

<b>bp</b>	Base pair
<b>CD</b>	Circular dichroism
<b>DNA</b>	Deoxyribonucleic acid
<b>DSF</b>	Differential Scanning Fluorimetry
<b>ds</b>	Double-stranded
<b>FAM</b>	6-carboxyfluorescein
<b>FRET</b>	Fluorescence resonance energy transfer
<b>G4</b>	Guanine Quadruplex
<b>G4-FID</b>	G-quadruplex fluorescent intercalator displacement
<b>G-tetrad</b>	Guanine tetrad, cyclic array of four guanines
<b>Iso</b>	<i>meta</i> -benzene
<b>IC<sub>50</sub></b>	Half maximal inhibitory concentration
<b>K<sub>a</sub></b>	Association/Binding constant
<b>LNCaP</b>	<i>Lymph Node Carcinoma of the Prostate</i>
<b>Log P</b>	logarithm of the partition coefficient
<b>NMR</b>	Nuclear magnetic resonance
<b>NHDF</b>	<i>Normal Human Dermal Fibroblast</i>
<b>MCF-</b>	<i>7 breast cancer cells, Michigan Cancer Foundation</i>
<b>MTT</b>	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
<b>Phen</b>	(1,10)-phenanthroline
<b>Quin</b>	6-methylquinoline
<b>RNA</b>	Ribonucleic acid
<b>RT-PCR</b>	Real-time PCR
<b>ss</b>	Single-stranded
<b>TAMRA</b>	6-carboxytetramethylrhodamine
<b>Ter</b>	<i>para</i> -benzene
<b>T<sub>m</sub></b>	Melting temperature
<b>TO</b>	Thiazole Orange
<b>TosMIC</b>	p-(tolylsulfonyl)methylisocyanide
<b>U87</b>	glioma cells.
<b>θ</b>	Ellipticity

---



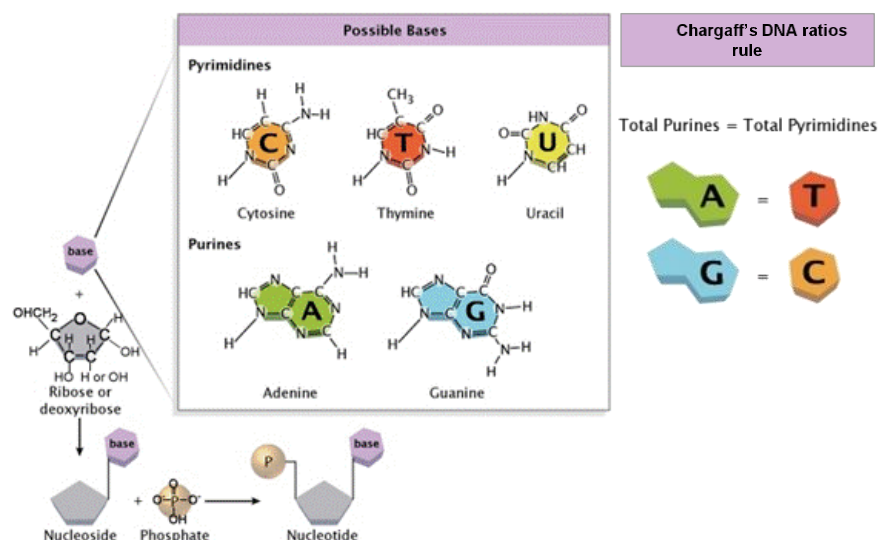
# I. INTRODUCTION

## I.1 NUCLEIC ACIDS: HISTORIC LANDMARKS AND GENERAL PROPERTIES

It was somehow by chance that the Swiss chemist Johannes Friedrich Miescher, in an attempt of characterizing the chemical components of the different organelles in white blood cells, reported in 1869 the discovery of a hitherto whole-new class of biomolecules that were named nucleins. This new class of molecules was later renamed to nucleic acids<sup>1</sup>. A scientific revolution was to follow the birth of molecular genetics was set as nucleic acids were proven to form genes, after W. Flemming identified the chromosomes in 1880 and its involvement in hereditary transmission established by Oswald Arey<sup>2-4</sup>.

However the basic structural components of nucleic acids were not to be described until early 1900s when Phoebes Levene suggested the polynucleotide model. Levene proposed that nucleic acids were composed of a string of nucleotides which were in turn composed of a 5-carbon sugar molecule, a phosphate group and only one of five nitrogen-containing bases already known at the time (fig. I.1): adenine (A), thymine (T), cytosine (C), guanine (G) and uracil (U)<sup>5</sup>. He also identified the two sugar units, the ribose and the deoxyribose that name the two nucleic acids molecules: ribonucleic acid (RNA) and deoxyribonucleic acid (DNA).

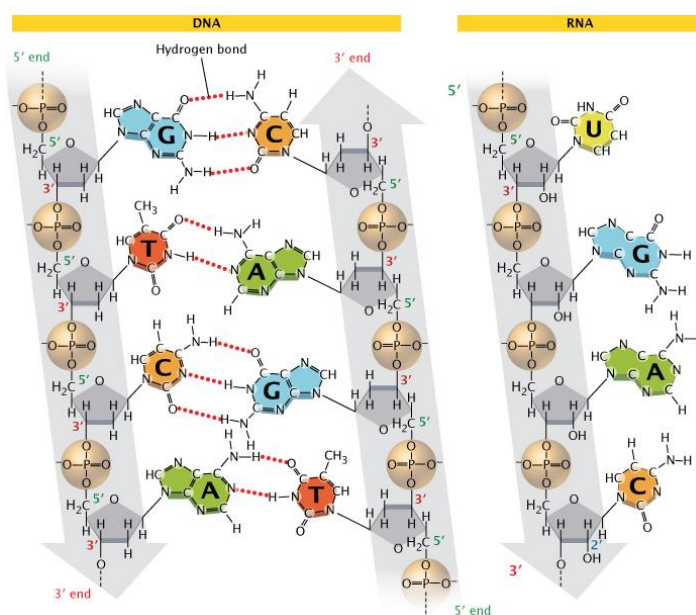
In the late 40's, Erwin Chargaff extended Levene's and Arey's. He accurately stated that the DNA content in A, T, C and G varies between different organisms wherein the A content is always equal to T and C equal to G, no matter which organism or tissue it is extracted – this was later called the Chargaff's ratios rule<sup>6</sup>.



**Figure I.1** - Left panel: Natural building blocks of nucleic acids, the nucleotides. A nucleotide is made up of a phosphate group, a sugar (ribose for RNA or deoxyribose for DNA), and a nitrogenous base (A, T, C, G, U). The chemical structures of all these components are shown. Right panel: Chargaff's ratios rule, which states that the total number of purines in a DNA molecule is equal to the total number of pyrimidines. Adapted from reference<sup>7</sup>.

Perhaps the most distinguished landmark on nucleic acids research is the discovery of the double-helix by James Watson and Francis Crick in 1953. At that time the X-ray crystallography technique was already developed and all the data were gathered by the English Rosalind Franklin and Maurice Wilkins who tried to infer the DNA structure<sup>8,9</sup>. Watson and Crick rushed to interpret their data and successfully joined all the pieces of the puzzle provided by the previous scientists<sup>10</sup>. The Watson and Crick DNA model still today remains very accurate (fig. I.2 and I.3): in its most predominant form, the DNA molecule is a two stranded helix, bound together by hydrogen bonds between nucleotide bases; the adenine bases are always paired with thymine's and cytosine with guanine's, which accounts for Chargaff's rule; the two strands are arranged in an anti-parallel fashion, i.e. the 5' end of one strand is paired with the 3' end, thus the two DNA strands are complementary of each other.

Unlike DNA, RNA molecules are composed of a ribose instead of deoxyribose and the uracil instead of thymine (fig. I.2). The resulting impact on its structure and function is quite dramatic - RNA molecules may adopt a vast range of different structures which are correlated with their function. For example, a single strand of RNA may form a double-helix hybrid with another strand of DNA. This was the first experimental evidence back on 1956 that information could be transferred from DNA to RNA (transcription phenomena that results in mRNA)<sup>11</sup>. Through the complementary bases in its strand, RNA may fold on itself into hairpins and loops that may form tRNA<sup>12</sup>, which helps incorporate aminoacids into a polypeptide chain, or either form rRNA which constitutes the ribosomes that catalyze this incorporation process (translation phenomena)<sup>13</sup>. The number of possible different structures does not stop here, there are many other classes of RNA that embrace a specific structure-function, such as snRNA (processing of pre-mRNA), snoRNA (processing and assembly of rRNA), miRNA (inhibits translation of mRNA) and siRNA (triggers degradation of other RNA molecules).



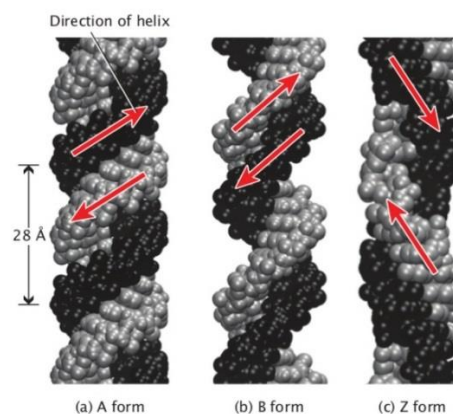
**Figure I.2** - Detailed molecular structure of DNA and RNA components. The two grey vertical arrows demonstrate that the DNA consists of two anti-parallel strands and RNA of one single strand. As illustrated, the main differences that define the two nucleic acids are the number of strands, thymine/uracil and deoxyribose/ribose content. Reproduced from reference<sup>14</sup>.



The detailed structure and properties of all the nucleic acids is quite vast and is out of the scope of this dissertation. Nonetheless, other than the basic components of nucleic acids and its sequence, there is a key factor that deeply influences the overall structure of nucleic acids which is nucleotide base-pairing.

### I.1.1 General nucleotide base-pairing properties

The double-helix structure described by Watson and Crick is not absolute. Indeed, it is the most predominant DNA structure (the B-form); nevertheless there are other conformations that DNA may adopt, such as the A-form or the Z-form. The A-form is very similar to the B-form and both structures are right-handed but in the Z-form the helix, albeit very unstable, is orientated to the left (Fig. I.3)<sup>15</sup>. This effect is a consequence of a slight difference on the base-pairing properties adopted, more particularly due to the nucleotide's R-S stoichiometric configuration.



**Figure I.3** – Three different conformations of DNA's double-helical structure. (A) A-DNA is a short, wide, right-handed helix. (B) B-DNA, the structure proposed by Watson and Crick, is the most common conformation in most living cells. (C) Z-DNA, unlike A- and B-DNA, is a left-handed helix. The different conformations result from the different ways that bases pair with each other. Reproduced from reference 7.

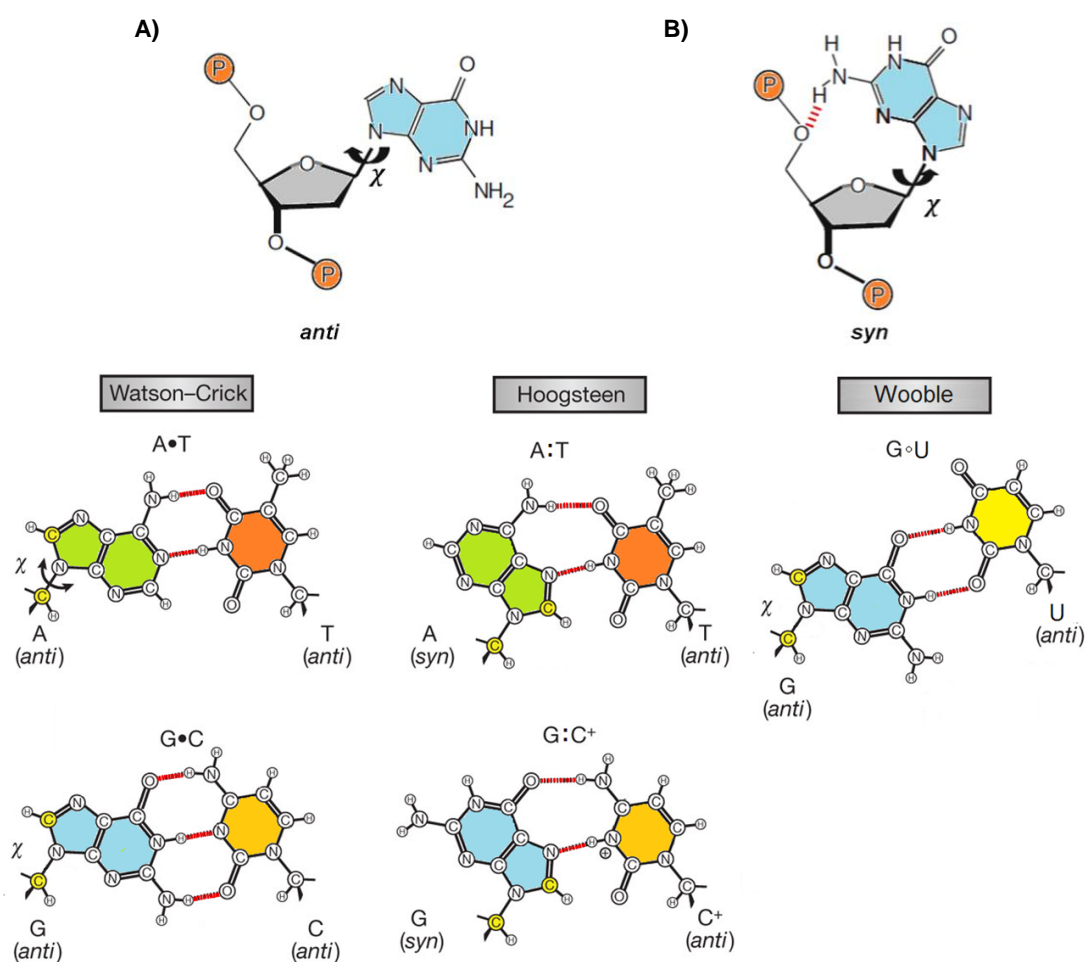
The base-pairing on the double-helix model requires A forming two hydrogen bonds with T (or U, in RNA) and C forming three hydrogen bonds with G, with all nucleotides being in the *anti* configuration, as illustrated in Fig. I.2 and I.4. These interactions are specifically called Watson-Crick pairing (WC). Nonetheless, these interactions are not the only possible arrangement for base-pairing, such as Hoogsteen and Wooble pairing.

In late 50s, Karst Hoogsteen obtained a crystallographic model of an unusual DNA complex in which he observed a markedly different A-T pairing scheme, rather than WC base-pair, wherein the purine bases (adenine) were flipped upside down<sup>16</sup>. Such flip is possible by a 180° rotation of the adenine base through the glycosidic bond, changing the base from an *anti* to a *syn* conformation as illustrated in Fig. I.4-b<sup>17</sup>. Consequently, an alteration must occur both in the sugar-ring pucker and in the topology of the hydrogen bonds, approximating the two complementary bases. The same *anti* to *syn* flip can also be verified for G-C<sup>+</sup> pairing (though in acid conditions)<sup>18</sup>. This type of base-pairing is hence called Hoogsteen pairing (HG), which are another important arrangement for base-pairing.

An additional note should be appointed for another type of base-pairing which is Wooble-pairing (WO), first described also by Francis Crick in 1966<sup>19</sup>. This type of pairing consists in a stable pairing between two non-complementary bases, such as U-G pairing. This type of pairing, however, only occurs in some special cases particularly in codon-anticodon

recognition, e.g. AAU-UUG. Regarding this last anticodon example, the 5' base (G) would not be as spatially confined as the others bases are, making it possible to “wobble” and adopt an optimum angle for pairing Uracil (Fig. I.4-c). Thus G-U pairing provides a unique array for of hydrogen bonds<sup>20</sup> where the co-planar guanosine N7 and O6 and the uridine O4 define a region of electrostatic potential, this way providing a recognition region for specific proteins or divalent metals, for example.

The impact of these base-pairing properties, such as a simple conversion from a base's *anti* to *syn* conformation is indeed quite dramatic when concerning the overall structure. For example, the peculiar shape of the Z-form (left-handed) is explained by “HG pairing in every other base and a change in the deoxyribose pucker in alternate bases, that resulted in a zig-zag arrangement of the backbone of the molecule”<sup>21</sup> (hence the name Z-DNA) while B-DNA (right-handed) practically consists in WC pairing.

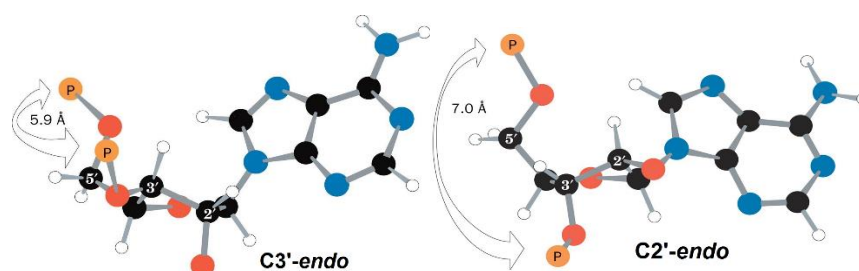


**Figure I.4** – A) Illustration of the syn-to-anti conversion by applying a  $180^\circ$  rotation around the glycosidic bond. To be noted that purines can convert from syn to anti conformation without a significant energy penalty, due to a hydrogen bond formed. B) Representation of the Watson-Crick (WC) and Hoogsteen (HG) edges for base-pairing. The HG geometry is achieved by applying the rotation  $\chi$  around the glycosidic bond. An example for Wobble base-pairing is also shown. HG pairing are represented by “:”, WC by “•” and WO by “◦”. Adapted from reference<sup>22</sup>.

The structure of the H-DNA (*triplex*) can also be explained by HG pairing, where, in particular conditions, an additional strand binds to the B-form helix of DNA forming a triple stranded helix<sup>23</sup>. In this motif each purine forms two types of pairing with the corresponding

pyridine, one through WT pairing and another trough HG pairing, where the latter corresponds to the third strand, making the third strand parallel to the purine strand. It is even possible to add one more strand into play forming a four stranded helix or a *quadruplex* which is the main subject of this dissertation. Quadruplex's most relevant structures and applications are introduced in the section to follow.

Last but not least, there is another important property that should be pointed which is the sugar pucker. Nucleotide's pentose ring is reasonably flexible, all its bonds may twist and bend so it may adopt different conformations more specifically half-chair conformation, and the type of conformation adopted has a significant impact on the overall structure. For instance, in B-DNA the lowest energy pentose conformation is the C2' *endo* conformation, whereas for A-DNA it is the C3' *endo* conformation<sup>24</sup>. In the *endo* conformations the C2' (or the C3') carbon is flipped out-of-the-plane in the same direction as C5', as represented in fig. I.5 (while in the *exo* conformations the out-of-plane atom is orientated in opposite direction as C5'). On a polynucleotide context, the C2' *endo* conformation allows a larger distance between the backbone phosphate groups hence reducing repulsion and extending the loops. In the other hand, a C3' *endo* conformation would result in much more compact and tighter helices, hence the difference in the A-form and B-form. The Z-form of DNA also illustrates this effect wherein its guanine bases are in the C3' *endo* conformation and all the other bases are in the C2' *endo* conformation.



**Figure I.5** – Deoxyguanosine with attached phosphorous atoms is shown in the two major nucleic acid sugar pucker conformations. The C2' *endo* pucker (right) is found in B-DNA, whereas the C3' *endo* pucker (left) is found in A-DNA or in RNA. The distance between successive phosphate groups is close to 7.0 Å in C2' *endo* and averages to 5.9 Å in the C3' *endo* pucker. Nucleic acids can convert from one pucker to the other, although it takes greater energy for conversion of ribonucleotides. Adapted from reference<sup>24</sup>.

## I.2 GUANINE QUADRUPLEXES

Structurally, nucleic acids are very versatile molecules and may adopt different conformations. As stated, the prominent conformation for DNA is the double-helix or duplex. Though, in special conditions it also may form a triple-stranded molecule (triplex)<sup>25</sup> or even a tetra-stranded molecule (quadruplex). The concept of a quadruplex refers to an ensemble of secondary structures that require 4 nucleotides for its formation, resulting in a quadruplex tertiary structure. Thus, the quadruplex motif does not necessarily require four independent strands, as one single strand could fold itself in a way that forms four nucleotide secondary structures. Examples of known quadruplexes are the i-Motif (i.e. cytosine-based quadruplexes)<sup>26</sup>, Holliday junctions<sup>27</sup> and guanine quadruplexes (G-quadruplexes)<sup>28</sup>.

As it will be introduced in the next section, guanine quadruplexes are structures which are implicated in biological processes and hence are considered an attractive target for drug design. It is therefore important to understand the properties and the rules that govern the stability and kinetic of these structures.

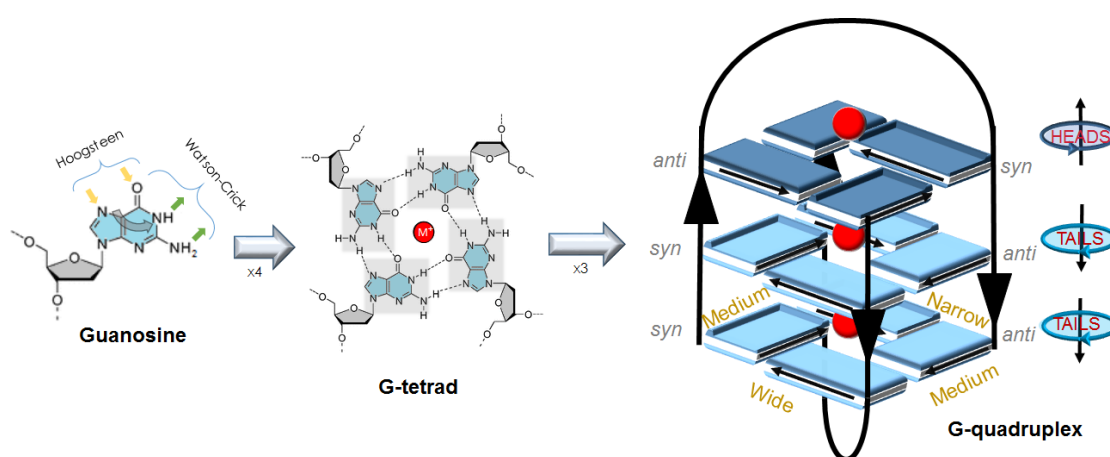
### I.2.1 Guanine Quartets

Guanine quadruplexes, abbreviated as G-quadruplexes, are tertiary structures of oligonucleotides (DNA or RNA) that comprise at least two units of guanine tetrads. Guanine tetrads or guanine quartets, also abbreviated as G<sub>4</sub>, are another type of secondary structure of nucleic acids that consist on pairing of four hydrogen-bonded guanines (fig. I.6). Although unusual, this G<sub>4</sub> arrangement is possible since each guanine base may form four hydrogen bonds through two donors (N1 and N2) and two acceptors (N7 and O6) by each pair, i.e. the WC edge pairs the HG edge, forming a cyclic tetramer of guanines<sup>29</sup>. These proprieties could origin a vast possibility of arrangements, such as triads<sup>30</sup>, quintets<sup>31</sup>, hexads<sup>32</sup>, heptets<sup>33</sup> and octets<sup>34</sup>; however the most biologically relevant arrangements are tetrads.

The G-tetrad may adopt two possible diastereotopic orientations that were called *heads* (+) and *tails* (-). This nomenclature is defined by applying the right-hand rule to the arrows pointing firm hydrogen bond donors to acceptors (fig. I.6). If the donor-to-acceptor bonds are orientated clockwise then the tetrad is "heads", and vice-versa for "tails"<sup>29</sup>.

The quartet orientation is determined by the phosphodiester backbone orientation, the number of strands and the glycosidic bond angle. For the two-stranded G-quadruplex in fig. I.6, there is one only possible arrangement for generating the G<sub>4</sub>, i.e. the glycosylic angle for the guanines in the same strand must follow 5'-*syn-syn-anti-3'* and 5'-*syn-anti-anti-3'*. Thus, *syn-syn* and *anti-anti* steps are related to *head-to-tail* orientation, and *syn-anti* related to *tail-to-tail* orientation<sup>35</sup>.

The *syn* conformation requires sugar conformation C'2 *endo* because of steric hindrance between O3 and C5, while the *anti* conformation may adopt both C'2 *endo* and C'3 *endo* pucker. Consequently, depending on the *syn/anti* conformations of the nucleotide, the G<sub>4</sub> grooves can be narrow, medium or wide (fig. I.6). All this factors taken into account also contributes to the vast conformational diversity of G-quadruplex structures<sup>36</sup>.



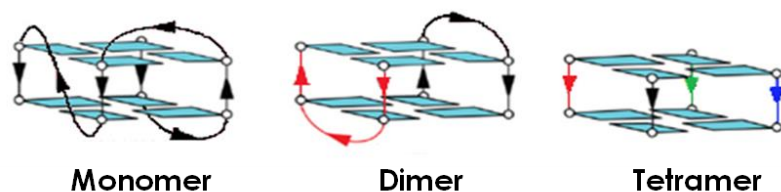
**Figure I.6** – Left: chemical formula of canonical guanine, representing of the two hydrogen bonding edges (Watson–Crick and Hoogsteen) implicated in G-quartet formation. Arrows indicate H-bond acceptors (yellow) and donors (green). Middle: G-tetrad (or G-quartet) motif, where guanines pair with each other through the WC and HG edges forming a cyclic array. In this particular case there is a donor-to-acceptors anti-clockwise orientation (tails). The red sphere represents a stabilizing cation (usually an alkali metal). Right: Example of a general quadruplex structure, with stacking of 3 G-tetrads. Strand direction along each edge of the quadruplex is indicated by large arrows from 5' to 3'. As example, the four grooves are labeled narrow, medium, or wide. Guanine bases are shown as rectangular solids; *syn* and *anti* bases are drawn with concave and full solids respectively. Dark blue squares indicates the "heads" face of a quartet.

Another aspect on the G4's conformation comes from the fact that inner G-O6 atoms are arranged in favorable conditions for metal quelation through cation-dipole interactions. Therefore, stabilization of G4 stacking is significantly increased by monovalent, bivalent and trivalent cations (depending on their ionic radii and hydration energies), as it reduces electronic repulsion between tetrads and promotes electrostatic interactions with the grooves and backbone phosphates. Hence, the stabilization of the G4 arrangement is enhanced by  $K^+ > NH_4^+ > Rb^+ > Na^+ > Cs^+ > Li^+$  monovalent ions<sup>37</sup> and  $Sr^{2+} > Ba^{2+} > Ca^{2+} > Mg^{2+}$  divalent cations<sup>38</sup>, by this order. These interactions with cations is so conserved that the successive stacked tetrads may form a central dynamic cation channel, similar to ion channels found in proteins<sup>39</sup>. The cation properties could even have a major impact on the overall G-quadruplex topology; for example, the thrombin binding aptamer and several telomeric DNA sequences adopt completely different G-quadruplex topologies in presence of  $K^+$  and  $Na^+$  media.

## I.2.2 Overview of the structural features of G-quadruplexes

G-quadruplexes are tertiary arrangements formed by DNA or RNA guanine rich sequences that should comprise at least two stacked units of guanine tetrads. The implicated guanines may be originated all from the same strand or from two or even four independent strands, thus G-quadruplexes may be uni-, bi- or tetra-molecular structures (fig I.7); three-stranded quadruplex are theoretically possible, but are not biologically relevant<sup>40</sup>. The remaining nucleotides of the strand that are not involved in the G-tetrad themselves, these form various types of loops that hold the G-tetrads together. The combination of all these features is what make G-quadruplexes so polymorphic, culminating in a vast number of different topologies that are characterized by the nucleotide sequence, number of strands, loop morphology, strand orientation and G-tetrad orientation. We should analyze some of these properties.

Starting with the sequence, it is quite straightforward that sequences composed by different nucleotides originate diverse G-quadruplexes, even if the overall topology is the same. Nonetheless the nucleotide sequence is a strong indicator for structure multiplicity and thus is used in computational methods for predicting potential G4 structures, as summarized in table I.1.



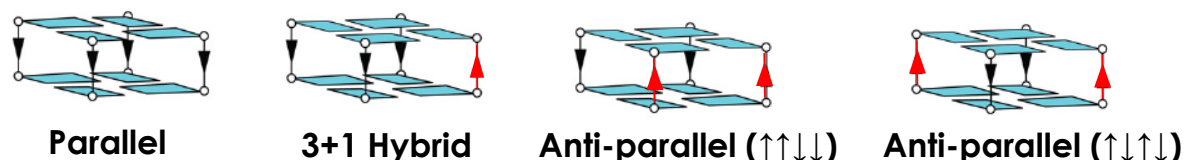
**Figure I.7** – G-quadruplexes may be uni-, bi- or tetramolecular structures. Each separate strand is identified by a different color. To current knowledge, only homodimers and homotetramers were identified.

**Table I.1** – General formulas for predicting potential G-quadruplexes. References<sup>41–43</sup>.

Type	Multiplicity	Sequence	Description
Intramolecular	Monomeric	$G_a X_m G_b Y_n G_c X_p G_d$	G represents guanine nucleotides; X represents any nucleotide except guanine; Y represents any nucleotide. $a, b$ and $c$ represents the number of G residues in each short G-tract; $3 \leq a, b, c \leq 5$ .
	Dimeric	$2 \times (X_m G_a Y_n G_b X_p)$	
Intermolecular	Tetrameric	$4 \times (X_m G_a X_n)$ or $4 \times (G_a X_m G_b)$	$m, n$ and $p$ represent the number of loop nucleotides; $1 \leq m, n, p \leq 8$ .



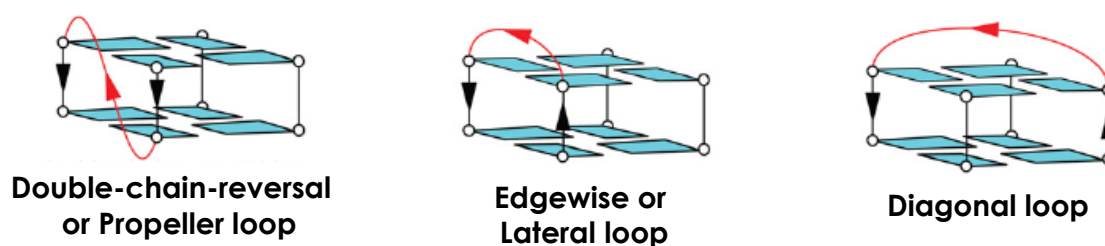
G-quadruplexes are often classified according to their relative strand orientation, such as parallel, anti-parallel or hybrid. Parallel and anti-parallel nomenclature refers to the relative orientation of the strands that form the G-quadruplex. If all strands have the same orientation, then the G-quadruplex has a *parallel topology* and the inverse concept for *antiparallel topology*, as illustrated in fig. I.8. When all strands are parallel implies that the glycosylic torsion angles must be in *anti* conformation, so the G-tetrads involved also have the same configuration. Anti-parallel quadruplexes adopt both *anti* and *syn* guanines which gives a much bigger number of possible combinations. Hybrid structures have only one strand with an inverse polarity and thus are still classified as *anti-parallel*, nonetheless these represent a special case.



**Figure I.8** – Representation of the possible arrangements for strand orientations for G-quadruplexes. When all strands have the same orientation, the structure has a parallel topology, otherwise it has an antiparallel topology. There are 3 types of anti-parallel structures: (3+1) hybrid (when one strand is anti-parallel), adjacent parallel (↓↓↑↑) and alternating parallel (↑↓↑↓).

Another feature for G4 characterization concerns the type of loops that it displays (fig. I.9). This is especially important for topology-specific drug-design. As indicated previously, loops are constituted by the nucleotides that are not involved in the tetrad formation, however are no less important. Indeed, the guanines define the G-tetrads, and the loops define the G-quadruplex. Furthermore it has strong repercussions on its stability and kinetics, for example, short loops (1-3 nucleotides) favor intermolecular quadruplexes and long loops (>7 nucleotides) are found to destabilize the G-quadruplex. An assessment of all properties of loops in G-quadruplexes are technical and would result in an extra chapter to this dissertation<sup>44-49</sup>. At this point, only the loop constitution should be the main concern.

Regarding monomeric and dimeric quadruplexes, parallel strands require an adjacent linkage within the strand, which is only possible through propeller-type loops (fig. I.9). Anti-parallel strands can be linked by diagonal or lateral loops, depending on whether the strands are adjacent or diagonally opposed (fig. I.9). For a unimolecular G-quadruplex, a minimum of three loops are required for connecting all the 3 strands, making a total of 26 possible different combinations of loops that are theoretically possible<sup>46</sup>. Tetrameric G-quadruplexes are usually absent of loops, since all strands are independent.



**Figure I.9** – Illustration of the most recurrent type of loops. Propeller (or double-chain-reversal) loops link parallel strands and the lateral and diagonal loops connect antiparallel strands.

Altogether, different combinations of all these feature originates the high polymorphism of G-quadruplex structures. Table I.2 reviews the nomenclature for the main features.

**Table I.2** – Summary of the nomenclature for the main G-quadruplex features.

Property	Description	Property	Description
<i>Strand Multiplicity</i>	Monomer; Dimer; Tetramer.	<i>G-tetrad Orientation</i>	Heads; Tails.
<i>Strand Orientation</i>	Parallel; Anti-parallel; Hybrid.	<i>Grooves</i>	Narrow; Medium; Wide.
<i>Loops</i>	Propeller; Lateral; Diagonal.	<i>Glycosylic angle</i>	<i>Anti; Syn.</i>

### I.2.3 G-quadruplex stability and assemble

At this particular point one question that might arise would probably concern the formation and stability of these G-quadruplex structures. Are they stable? Do they form spontaneously? How many conformations one could adopt; do they include duplex and other structures? And, do these structures naturally occur *in vivo*? The answer for all these questions depends on the chemical environment conditions, such as molecular crowding, pH, temperature, binding-proteins, binding nucleic acids, ions and strand concentration. Data originated from *in vitro* studies does not give the best answers to all these questions, but we should start gradually from simpler conditions to more complex as *in vivo* conditions.

First, the G-quadruplex stabilizing forces are fundamentally the same as for all other nucleic acids structures such as duplexes and triplexes. However, due to the completely different nature of the G4, namely what comes to H-bonding interactions, stacking of G-quartets, and the involvement of specific ionic interactions (direct inner spheres coordination) as well as any nonspecific electrostatic effects of ionic interactions, all the parameters weights must be adjusted. The description of the stability is best addressed in terms of free-energy as the following<sup>46,50</sup>:

$$\Delta G_{tot} = \Delta G(\text{confor}) + \Delta G(C) + \Delta G(\text{polyel}) + \Delta G(w) + \Delta G(r + t) + \Delta G(vdW) + \Delta G(HB) + \Delta G(\text{bonds}) + \Delta G(\text{electro}) \quad (\text{Equation 1})$$

where *confor* accounts for conformational differences for the strands and folded state, *C* is the Coulombic force (enthalpic), *polyel* is the polyelectrolyte effect (entropic), *w* is the hydration term (entropic), *vdW* is the van der Waals term, *HB* is the hydrogen bonding term (enthalpic), *bonds* accounts the bond lengths and angle, and finally *electro* accounts for electronic interactions such as polarization. Accordingly, the unfolding of the G-quadruplex must be entropy driven, due to the increased freedom degrees of the unfolded state and chain dissociation; enthalpy has an unfavorable contribution to the energy balance, due to loss of H-bonds, stacking interactions and cations release from the G4 cavity<sup>51,52</sup>.

By far, the most employed and practical way to measure the stability of the G-quadruplex is by temperature-induced transitions (mainly the G4 unfolding) using various spectroscopic methods such as UV, fluorescence and circular dichroism (CD). Assuming a two-state model (i.e. folded-unfolded), the melting temperature  $T_m$  corresponds to the temperature at which 50% of the transition is complete<sup>53</sup>. The only problem is that the unfolding of many G-quadruplexes, such as the telomeric, are not simple two-state transitions, there are often a third

or forth intermediate structures that must be considered in the mechanism (eq.2)<sup>54,55</sup>. Calorimetric techniques, such as differential scanning calorimetry (DSC), allow a deeper insight into the energetics of the G-quadruplex stability and may resolve the observation of intermediate transitions.



where  $U$  represents the unfolded state,  $F$  the folded state and  $I_n$  one or more intermediate states.

It is possible that a particular sequence may adopt more than one possible conformation and studies *in vitro* suggest that they in fact coexist as a mixture. The only question is one of populations and kinetics<sup>56-58</sup>. These cases must be carefully considered, since the unfolding pathway of two independent motifs could share a common intermediate and thus leading to wrong interpretations<sup>46</sup>.

For a detailed study on the kinetic of folding and unfolding of intramolecular and intermolecular G-quadruplexes *in vitro*, these references are suggested<sup>55,59,60</sup>.

- **Small sequences of G-quadruplexes**

In a saline solution, the folding/unfolding of small unimolecular quadruplex sequences is generally rapid and is thermodynamically favored, assuming no other strands are involved<sup>52</sup>. As for bimolecular and tetramolecular quadruplexes, the rate of formation is much slower and is concentration dependent, subsequently kinetic parameters should be considered when studying these type of structures.

The apparent binding of  $K^+$  and  $Na^+$  to these sequences is highly cooperative. Consequently, in absence of cations the folded fraction is much smaller, non-zero<sup>61</sup>. Additionally, in these low (or none) salt content conditions the rate of interconversion between the  $U$  and  $F$  states is usually much slower; this should be considered on the heating or cooling rates when studying G4 structures in these conditions<sup>52</sup>.

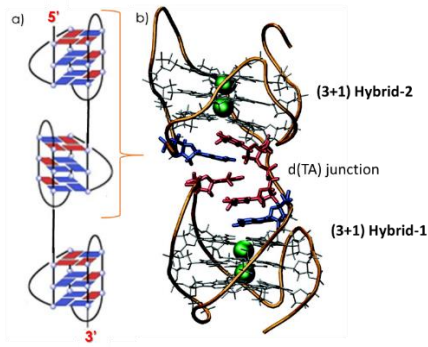
- **Long sequences of G-quadruplexes**

The studies mentioned previously were performed on short <30pb DNA sequences. There is a chance that G-quadruplex might behave differently with long flanking sequences and therefore studies were performed to investigate this topic. Again, results show that formation of G-quadruplex would depend on the chemical environment and on the G4 forming sequence. No high-resolution structures are available for these long sequences and they are difficult to study by the usual NMR and crystallographic methods. Thus biochemical assays, circular dichroism and dynamic modelling studies must rely on qualitative data.

As an example, *c-myc* promotor sequence (a sequence with G4 forming potential) indeed adopts a G-quadruplex motif in a plasmid with ~850pb long, but only when the DNA is in negative superhelicity conditions<sup>62</sup>. Still, these structures are in dynamic equilibria of unfolded and duplex forms.

The DNA telomeric sequence is even more interesting because the telomere end (called the overhang, ~200pb) consists on long repeats of 5'-TTAGGG, which form a quadruplex. Interestingly, in solution analysis suggests that this sequence forms consecutive G-quadruplex assemblies, particularly (3+1) hybrid-1 and hybrid-2 assemblies<sup>63-65</sup> (fig. I.10).





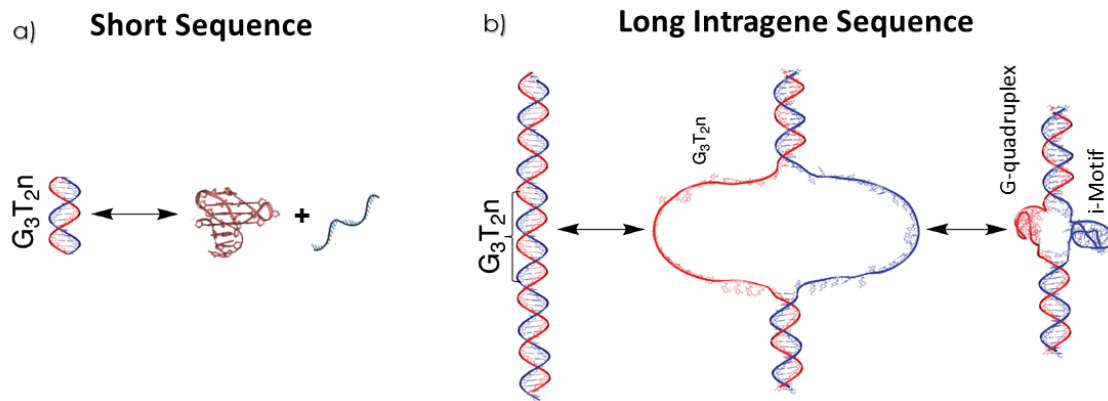
**Figure I.10** – Schematic drawing elucidating the hypothesized mixed-type structure (a) of the higher-order G-quadruplex of the telomeric overhang region. (b) Average structure from molecular dynamics simulation of the Hybrid-12 model; dT and dA residues, in the junction region between the quadruplex units, are shown in red and blue, respectively. Results adapted from reference<sup>64</sup>

However other studies demonstrate that high order multimeric quadruplex assemblies are gradually destabilized by increasing size. These studies verified that twelve 5'-TTAGGG repeats form three continuous quadruplex structures (a hybrid-1-2-1) that are energetically favorable<sup>66</sup>, more than three runs were found to be thermodynamically unstable and their existence *in vivo* is questionable. This topic is still under discussion to date.

- **Equilibria with complementary strands**

When complementary strands are presents, all the hybridizing structures are possible to occur in solution, such as duplex, quadruplex and hairpins (Fig. I.11-a). Nonetheless it is often found that the G4 topology are unexpectedly populated. It is not that G-quadruplexes are more thermodynamically favored than duplexes, loops and hairpin structures, which they are not, but is the unfolding kinetics of the G4 that is much slower than all the other structures, thus creating a build-up in population<sup>67</sup>. For telomeric sequences, the duplex dominate the overall topology at  $\text{pH} > 5.0$ <sup>68</sup>; for the short *c-myc* oncogene promotor sequence the quadruplex prevails, depending of the system and the chemical enviroment<sup>69</sup>.

In a long sequences context (>100pb) things are a little more complex. For telomeric overhang sequence this is not a problem since it's single-stranded; for other double stranded sequences such as the *c-myc*, G4 formation must overcome the energy penalty for unwinding the double-helix (fig. I.11-b), forming a G4 in one strand and leaving the complementary strand as a loop (very unfavorable) or as an *i*-motif (the *cytosine-quadruplex*). According to Lane *et al*, the quadruplex formation does not overcome the very unfavorable strand separation energy, by an amount  $>30 \text{ kcal mol}^{-1}$  in 200mM salt and a GC fraction of 0.8<sup>46</sup>. That is in the absence of any factor for stabilization of the quadruplex, which could decrease the overall free energy, such as inducing negative superhelicity or by applying G4 binding factors.

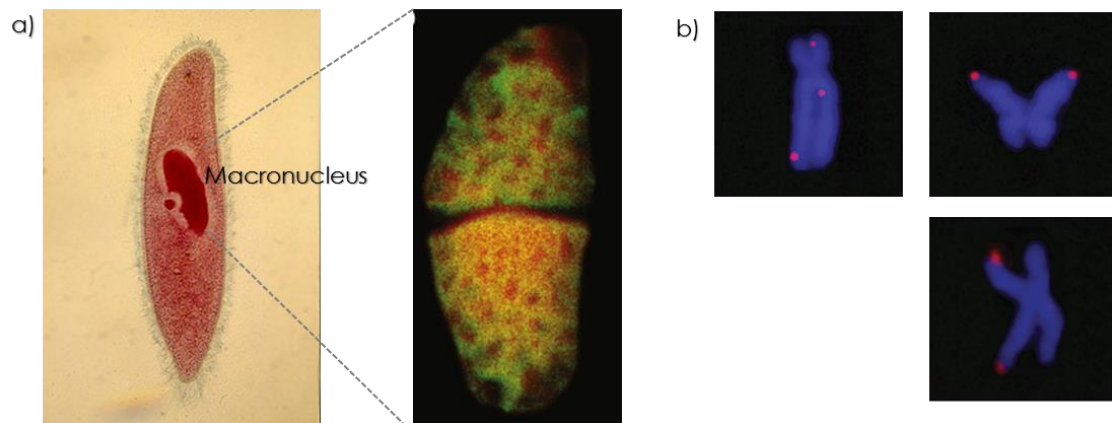


**Figure I.11** – (a) Proposed model for duplex-quadruplex equilibrium in short sequences. The complementary strand does not necessarily form an i-motif. (b) Model for formation of an internal G-quadruplex. A duplex DNA with a potential G-quadruplex sequence must unwind, redistributing histones and opening a loop. The G-quadruplex sequence may fold into an intramolecular G-quadruplex while the complementary C-rich sequence may potentially fold into an i-motif. This process is more probable under negative supercoiling induced by transcription process. Proteins or small molecule ligands may stabilize the G-quadruplex or the i-motif, locking the DNA in this conformation. Adapted from reference<sup>46</sup>.

- ***In vivo* conditions**

*In vivo* conditions are far more complex than the regular *in vitro*. First there are important crowding effects that alter the water activity and making solvation and electrostatics complex difficult to predict. Concerning molecular crowding, it is well demonstrated that it strongly influences the quadruplex-duplex equilibria and induces dramatic changes on the G4 topology<sup>70</sup>. In addition, there are countless binding factors, such as proteins and other nucleic acids that induce conformational changes on DNA. Taking in account all these factors it is difficult to extrapolate *in vivo* the structures of the G-quadruplexes. Nonetheless, what can be done is a qualitative demonstration that G-quadruplexes may in fact occur *in vivo*.

Thermodynamically, the telomere overhang is the most favorable place to find G-quadruplexes *in vivo*. There are few studies that indicated the presence of quadruplexes in the telomeres<sup>71–73</sup>, but perhaps the most remarkable was the use of highly specific fluorescent G4 antibodies (binding affinities of 0.1–1 nM)<sup>74,75</sup> that were also able to bind a specific topology (parallel or antiparallel). Because of the high concentration of telomeres in *Styloynchia lemnae*'s macronuclei, the presence of G-quadruplexes could be observed directly by *in situ* immunostaining (fig. I.12-a). Remarkably, only anti-parallel specific antibodies gave signal. Moreover, it was noticed that the signal was only present in vegetative state (i.e. not during the replication).



**Figure I.12** – a) At left a general example of a ciliate microorganism, evidencing the macronucleus. At right the confocal microscopy image of *Stylonychia lemnae*'s macronucleus showing in situ staining of antibodies against the antiparallel telomeric G-quadruplex structure (yellow) overlaid on DNA staining (red). Remarkably, no staining was observed in the replication bang. According to this study<sup>75</sup>, several controls confirmed the specificity of this method, declaring *no staining was observed with an antibody directed against the parallel telomeric G-quadruplex structure or after removal of the 3'-overhang by S1 nuclease digestion or alkali treatment before in situ analysis*. By contrast, when anti-B DNA antibody was used, immunostaining was observed in the replication band, demonstrating that this structure is accessible to the antibody. b) Immunofluorescence for a specific single-chain antibody (BG4) on metaphase chromosomes isolated from HeLa human cervical cancer cells. Discrete BG4 foci (red) were observed both within the non-telomeric regions and at the telomeres. Chromosomes are counterstained with DAPI (blue). Results from reference<sup>76</sup>

A few studies were also performed on human cells with monoclonal and structure-specific antibodies (fig.12-b) that indicated the presence of the G-quadruplex<sup>76,77</sup>. Direct detection of que G4 (“probe-free”) was accomplished by in-cell NMR studies that were performed on *Xenopus laevis* oocytes and indeed confirm presence of an inserted G-quadruplex in the cytoplasm for a certain period<sup>78</sup>. Until this moment, there is no indisputable demonstration that the G-quadruplex motif (or any other) naturally occurs *in vivo*, since the antibodies could interfere with the duplex-G-quadruplex equilibria, but there is strong evidence of it.

### I.2.3 Importance of G-quadruplexes in the human genome: telomeres and oncogene promoters

Bioinformatic analysis based on equations on table I.1, permitted to identify about 375,000 sequences of the genome that may potentially fold into G-quadruplex structures<sup>79,80</sup>. Among all these sequences there are two major categories of G-quadruplexes that deserve a brief structural description since their application was essential to this work.

- **Telomeric G-quadruplexes**

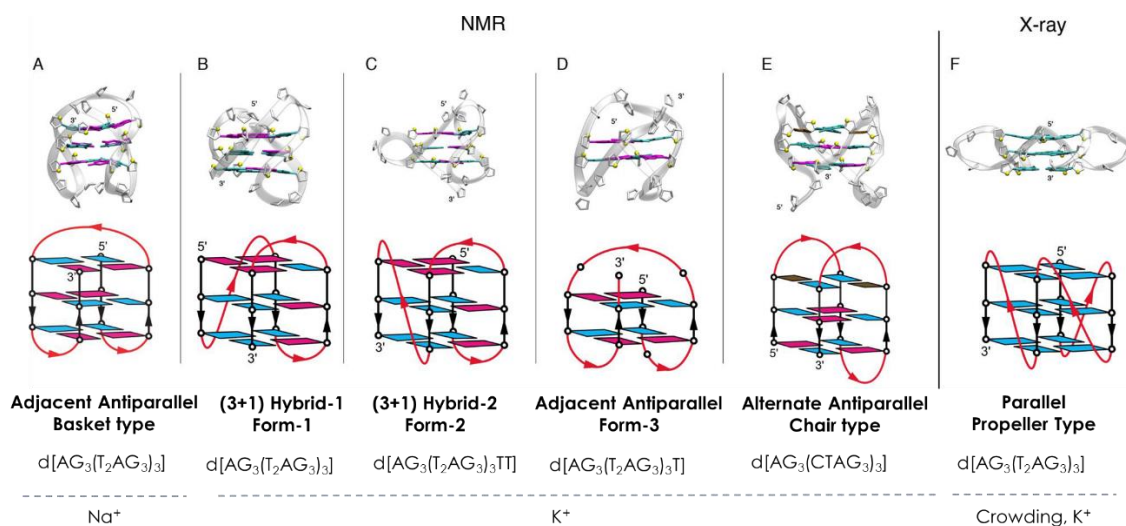
The telomere tandem repeat d(TTAGGG) is the basic sequence for all three types of unimolecular, bimolecular and tetramolecular quadruplexes<sup>81,82</sup>. The unimolecular quadruplex represent the most probable model for the single-stranded telomere overhang.

In solution, the telomeric unimolecular G4 topology is very versatile and extremely sensible to chemical conditions, especially ions, hydration and sequence. Generally, in Na<sup>+</sup> it adopts an *adjacent antiparallel* (or specifically a basket-type) topology<sup>83</sup>, with two lateral and one diagonal loop.

Remarkably, in K<sup>+</sup> media other multiple G-quadruplex topologies may be observed, depending on sequence length, concentration and crowding (fig. I.13). In dilute solution of K<sup>+</sup>

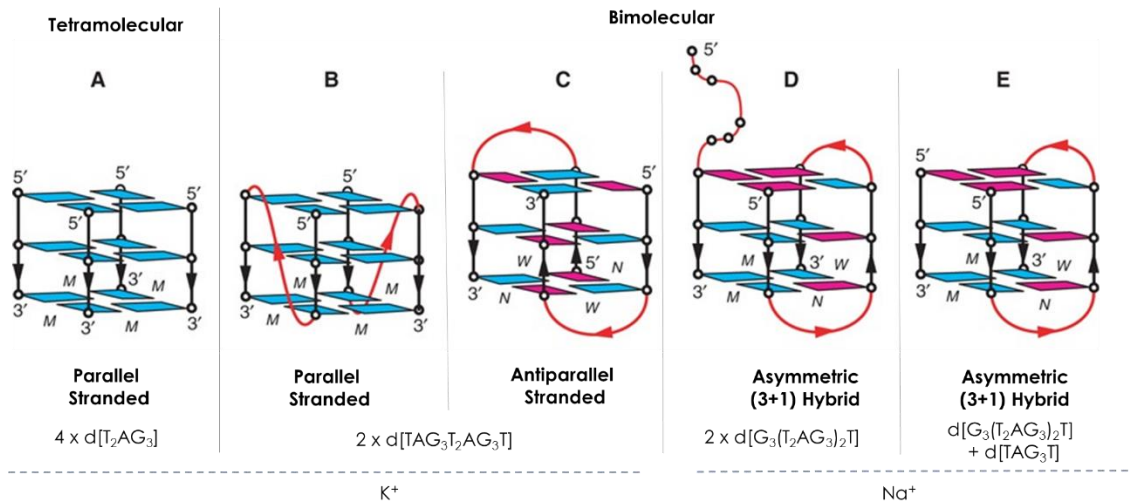
(<100 mM), the G-quadruplex is predominantly found as a (3+1) *hybrid-1*<sup>84</sup> having a propeller followed by two lateral loops. When the 3' flanking nucleotides are two thymine [i.e. (T<sub>2</sub>AG<sub>3</sub>)<sub>n</sub>TT], it was found that the predominant structure is a (3+1) *hybrid-2* topology; the difference relies only in the order of the loop arrangement with two lateral loops followed by a propeller<sup>85</sup>. In dehydrated and crowded conditions<sup>86,87</sup>, it was determined by CD and X-ray crystallography a well-defined *parallel* topology, with all propeller loops. Another more surprisingly form is found in K<sup>+</sup> media with an antiparallel basket-type topology, similar to the one formed in sodium, but with only two G-tetrads instead of three<sup>88</sup>, called the *form 3*; it is suggested that this structure represent an intermediate folding into other topologies<sup>89</sup>.

There is another quadruplex topology that this sequence may adopt being a CG-quadruplex (cytosine-guanine quadruplex). When the 5' flanking is a cytosine [i.e. (CTAG<sub>3</sub>)], another antiparallel G4 is formed, with an *alternating antiparallel* (or specifically a *chair-type*) topology<sup>90</sup>. This topology is also biologically relevant, because sequence variation in human telomeric DNA may occur<sup>91</sup>.



**Figure I.13** – Summary of the possible different unimolecular G4 topologies identified so far for telomeric DNA. At the top the 3D structures and correspondent schematic view below. The corresponded sequence is also identified at the bottom. (A) basket-type G-quadruplex in Na<sup>+</sup> solution (PDB: 143D)<sup>83</sup>; (B) Form-1 G-quadruplex in dilute K<sup>+</sup> solution (PDB: 2JSM)<sup>57</sup>; (C) Form-2 G-quadruplex in dilute K<sup>+</sup> solution (PDB: 2JSC)<sup>57</sup>; (D) Form-3 G-quadruplex in dilute K<sup>+</sup> solution (PDB: 2KF8)<sup>88</sup>; (E) Chair-type G-quadruplex in dilute K<sup>+</sup> solution (PDB: 2KM3)<sup>90</sup>; (F) Propeller-type G-quadruplex in K<sup>+</sup> containing crystal (PDB: 1KF1)<sup>87</sup>. Color code: cyan, *anti* guanines; magenta, *syn* guanines; brown, cytosines; red, TTA loops. Adapted from references<sup>86,92</sup>.

Bimolecular and tetramolecular G-quadruplexes are also biologically relevant, as the single-stranded telomeric overhangs between two chromosomes might associate with each other forming a multimeric quadruplex, thus may represent a model for chromosomal fusions and anaphase bridges<sup>81,93</sup>. Figure I.14 summarizes the some of the observed multimeric structures adopted by small sequences in solution.



**Figure I.14** – Schematic view of some of the possible different intermolecular G4 topologies identified for small sequences of telomeric DNA. The corresponded sequence is also identified at the bottom. (A) Tetrameric parallel stranded G-quadruplex in  $\text{K}^+$  solution<sup>94</sup>; (B) Dimeric parallel stranded and (C) antiparallel stranded G-quadruplexes in  $\text{K}^+$  solution<sup>95</sup>; (E) Dimeric asymmetric (3+1) hybrid G-quadruplex observed in  $\text{Na}^+$  solution<sup>96</sup>; (D) Dimeric asymmetric (3+1) hybrid hetero-stranded G-quadruplex observed in  $\text{Na}^+$  solution<sup>96</sup>. Color code: cyan, *anti* guanines; magenta, *syn* guanines; brown, cytosines; red, TTA loops. Adapted from references<sup>97</sup>.

### • Oncogene promoters

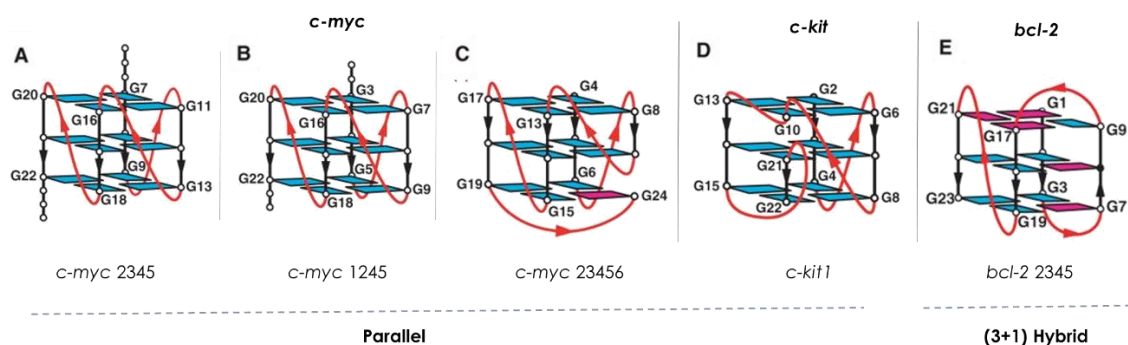
Besides telomeres, other G4 forming sequences were also identified in key regulatory regions of several genes that are strongly associated with cancer activity, such as the *c-myc*<sup>98</sup>, *c-kit*<sup>99</sup>, *bcl-2*<sup>100</sup>, *Rb*<sup>101</sup>, *VEGF*<sup>102</sup>, *KRAS*<sup>103</sup>, *HRAS*<sup>104</sup>, *hif-1a*<sup>105</sup> and *TK1*<sup>106</sup>. Unlike the telomere overhang, these sequences have a complementary strand forming the regular duplex motif, so in order to induce the G-quadruplex the strands must unwind and necessarily form an intermolecular quadruplex, as illustrated in fig. I.11.

These sequences may form hybrid and antiparallel G4 structures but the most frequent topology is parallel. Therefore, the loop arrangements are the main factor for the polymorphism in these sequences. Taking the *c-myc* as an example, its wild-type sequence (Pu27) is known to adopt a mixture of different parallel G-quadruplex structures. This sequence has six guanine-tracks that form the G-tetrads,



therefore the variation on loops is due to the alternation of the guanine-tracts upon forming the G-quadruplex. Figure I.15 exemplifies this effect, and also shows other proto-oncogene G-quadruplex structures.





**Figure I.15** – Schematic view of some of the possible different intermolecular G4 topologies identified for small sequences of proto-oncogenes. The gene sequence is identified at the top. Typically, these promoter oncogene sequences adopt a mixture of several G-quadruplexes. In order to isolate all different structures, little variations on the sequences were performed; these are named after the involvement of the corresponding guanine-track on the quadruplex. (A) Variant *c-myc* 2345 in  $K^+$  solution<sup>107</sup>; (B) Variant *c-myc* 1245 in  $K^+$  solution<sup>108</sup>; (C) Variant *c-myc* 23456 in  $K^+$  solution<sup>109</sup>; (D) *c-kit* 1 G-quadruplex structure in  $K^+$  solution<sup>110</sup>; (E) Variant *bcl-2* 2345 in  $K^+$  solution<sup>111</sup>. Color code: cyan, *anti* guanines; magenta, *syn* guanines; red, TTA loops. Adapted from reference<sup>82</sup>.

### I.3 G-QUADRUPLEX SEQUENCES AS THERAPEUTIC TARGETS

Guanine quadruplexes are not a new subject, in fact they're known since before the discovery of the double-helix (G-quartet in 1910<sup>112</sup>; G-quadruplex in 1962<sup>113</sup>). However they remained just as a scientific curiosity, without any practical applications, until the complete decoding of the human genome and after realizing that telomere overhangs had potential to form G-quadruplex structures.

G-quadruplexes may have some natural regulatory functions in the genome, after the observation that nearly half of the human genes carry one putative G-quadruplex forming sequence, particularly at the first intron on the non-template DNA strand<sup>114</sup>. Nonetheless, the natural formation of G4 in the genome must be thoroughly regulated, since it is well demonstrated that uncontrolled formation of G4 induces severe genomic instability<sup>115</sup>. For example, it is shown that a deficiency on G4 helicases (responsible for unwinding G-quadruplexes in the genome) leads to disorders associated with genomic stability such as Bloom syndrome, Werner syndrome and Fanconi anemia<sup>115</sup>.

Therefore, if somehow the G-quadruplex formation is induced on disorder-causing agents, such as cancerous cells, these agents would potentially develop genetic instability, and cellular unviability. This strategy represents an application for chemotherapy. Furthermore, the well-defined G-quadruplex makes them molecular therapeutic targets since they offer exceptional and realistic strategies for rational drug design and are easy to analyze *in vitro*. Moreover, the fact that the target (DNA) cellular concentration is very low also requires low quantities of administrated drug, which is an advantage concerning the side-effect and cytotoxicity.

There are two main approaches to therapeutically target the G-quadruplex. One approach relies on indirect telomerase inhibition using telomeric G-quadruplex sequences and the other on oncogene transcription regulation through inducing the G-quadruplex in promoters.

### I.3.1 Telomere G-quadruplex targeting

The terminal regions of linear chromosomes are called telomeres, which are essential for maintaining the genomic stability of the organism. Telomeres are guanine-rich regions that vary between species. For humans this region consists on a repetitive 5'-TTAGGG/AATCCC double-stranded sequence that may extend up to 15.000 base pairs<sup>116</sup>, followed by a 60-600 nucleotide repeats of 5'-TTAGG single-stranded terminal sequence that is commonly called telomere overhang<sup>117</sup>. The telomere overhang is usually supported by protecting proteins, such as POT1.

As previously mentioned and during the course of replication, the DNA polymerase is not able to properly replicate the last nucleotides of a template strand, which gradually leads to loss of information on the template strand. Thus, insertion of telomeres work as a non-coding expendable sequence that prevents loss of other important coding genetic information. Also the telomere overhang form a complex with POT1 that caps the telomere, masking the chromosome ends from DNA repair machinery that otherwise would endlessly extend the sequence<sup>116</sup>.

The telomeric DNA naturally shortens after each cycle of cellular replication and is responsible for the natural aging process in organisms<sup>118</sup>. After several cellular replication cycles, telomeres progressively shorten until a certain critical point is reached, the "Hayflick limit"<sup>119</sup>, where the cell halts replication (senescent state) which eventually ends-up in cellular apoptosis<sup>120</sup> (fig.14).

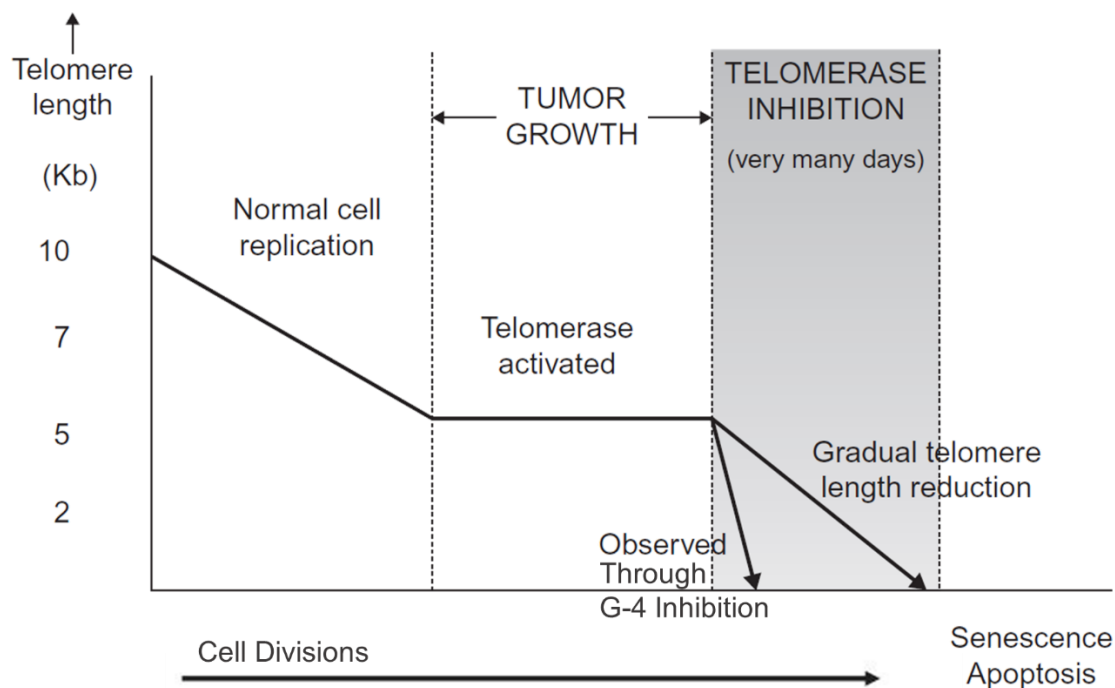
To avoid this problem and to fix the telomere shortening, a special enzymatic complex called telomerase elongates the telomeres by adding-in more telomeric repeats, thus maintaining telomere homeostasis. This enzymatic complex consists in two major domains, reverse transcriptase (TERT) and telomerase RNA (TER). The catalytic domain (TERT)<sup>121</sup> - which has intrinsic reverse transcriptase activity - catalyses the insertion of dNTPs in the DNA strand using the RNA-containing domain TER<sup>122</sup> as a template and the DNA's 3' ending as a primer sequence. Telomerase may also produce new single-stranded telomere overhangs.

In normal somatic cells, the TERT domain is expressed in extremely low levels. Remarkably, it was found that 80-85% of all human cancer cells express high levels of enzymatically active telomerase<sup>123,124</sup>. Thus, there is a direct association between telomerase and cellular immortalization. This limitless ability for replication is regarded as one of the major hallmarks of cancer<sup>125-127</sup>.

It was found that inhibition of telomerase activity by small-molecule inhibitors reverses the immortality process of cancers cells, and thus represent a tumor suppression mechanism<sup>120,128</sup>. The advantage in adopting this strategy is that it only affects cancerous cells, with little impact on somatic cells<sup>129,130</sup>, therefore can be used as a viable therapeutic strategy. The major drawback is that the process depends on the cellular replication, so it should in principle be a lengthy process (fig. I.16). For example, for a 5Kpb telomere, with a 24h cell doubling time and a subsequent loss of 100 nucleotides per round of replication, it would take about 40 days until reaching critical telomere shortening.

An alternate approach for telomerase activity inhibition relies on folding the 3'end-primer telomeric DNA into a G-quadruplex structure that inhibits binding of telomerase<sup>131</sup>. However this process require a driving force to stabilize the G-quadruplex folding. The use of small-molecule G-quadruplex ligands could act as indirect telomerase inhibitors, as demonstrated with anthraquinone derivatives<sup>132</sup>. Furthermore, it is often observed that inhibition through

the G-quadruplex approach induces senescence much faster than telomere shortening route, which is most fortunate (fig. I.16). Assuming that G-quadruplex formation on the overhang displaces POT1, the signalization for DNA damage response pathways in the cell is amplified<sup>133–135</sup>. These pathways can induce senescence on cancer cells<sup>136</sup>. Thus G-quadruplex ligands may exert anticancer activity not only through inhibiting telomere elongation but also by rapidly disrupting the capping functions of the overhang<sup>116</sup>.



**Figure I.16** – Schematic representation of the pathway of events that occur during the lifespan of a regular somatic cell, going from progressive shortening over time until a transformation event where telomerase activity is enhanced. Upon telomerase inhibition the cell is turned back to mortal, gradually shortening its telomeres through time. The actual observations through quadruplex inhibition reveals a much more rapid response, due to the overhang uncapping which activates important DNA repair machinery and antiproliferative pathways. Adapted from references<sup>124,137</sup>.

Due to limitative drug-like properties of the telomere G4 binding ligands described so far, this therapeutic approach has not yet progressed to clinical evaluation. However, these G4-binders revealed to be effective *in vitro* and hence a promising strategy for *in vivo*.

### I.3.2 Regulation of transcription by targeting promotor G-quadruplex

As stated previously, it was found that most of the human oncogenes contain putative quadruplex sequences in their promoters. This provides an opportunity for a therapeutic application and rational drug design. The mechanism would essentially consist in inducing the G-quadruplex motif in the promotor sequence through high-specific small ligands; this complex would present an impediment for RNA polymerase or other essential transcription factors, resulting on silencing of the oncogene expression. This mechanism would further render cancer cells vulnerable to other therapeutic procedures<sup>138,139</sup>.

Given the conditions prior to duplex unfolding and specific G4 stabilization (see section I.2.2, fig. I.11), the polymerase inhibition through G4 induction could reveal to be quite challenging but indeed effective.



*c-myc* silencing is one of the examples for this effect. The *c-myc* gene is a “master-controller oncogene”<sup>139</sup>; its promoter is found to be overexpressed in almost 80% of all human cancers<sup>140</sup> and makes it the most studied system for G-quadruplex targeting. A study demonstrated that induction of the G-quadruplex in the *c-myc* promoter results in overall oncogene silencing<sup>98</sup>. Consequently, application of a G-quadruplex ligand (TMPyP4, in this case) resulted in indirect inhibition of the MYC protein expression.

Clinically, quarfloxin is so far the most successful G4 ligand. It even reached phase II clinical trials for treatment of neuroendocrine/carcinoid tumors<sup>141</sup>. Unintendedly, it targets a G-quadruplex sequence found in ribosomal DNA (which is very similar to the *c-myc* quadruplex) causing the displacement of nucleolin from the nucleus to the nucleoplasm. This mechanism results in inhibition of *c-myc* expression. Moreover, the nucleolin displacement phenomenon is also a common response to cellular stress which triggers various specific pathways<sup>142</sup> resulting in apoptosis of cancer cells, but not in normal cells. Unfortunately quarfloxin was disapproved from clinical trials due to bioavailability concerns, nonetheless its toxicity profile was very encouraging for future G4 ligands applications.

## I.4 G4 LIGANDS FOR CANCER THERAPY

Perhaps the most efficient way to target the G-quadruplex motif is through monoclonal antibodies, which are capable of exhibiting outstanding binding affinities and exceptional selectivity. However their application is extremely limited, as they are unable to reach the nucleus or even the cytoplasm<sup>143</sup>. For this reason, high specific small ligands are more prospective regarding their therapeutic applications.

Concerning the design of new G4 ligands, some important structural features were identified based on previous reports. In order to achieve their target, the design of new ligands should focus on:

- Enhancing G-quadruplex affinity;
- Enhancing specific-G-quadruplex selectivity (specially over duplex and toward a particular G-quadruplex topology);
- Improving drug-like features.

The major interactions between the G-quadruplex and small ligands should be summarized. Essentially there are four major routes for building selective binders to G-quadruplex, which are summarized below.

- **G-tetrad external  $\pi$ - $\pi$  stacking.**

$\pi$ - $\pi$  stacking on the top or bottom of the G-tetrads is the most extensively exploited mode of interaction for a G-quadruplex-ligand complex. It requires larger planar aromatic systems (fig. I.17-A). The aromatic surface should have the approximate dimensions of the G-tetrad in order to enhance selectivity over duplex DNA. Well known examples are the TMPyP4<sup>144,145</sup>, BRACO-19<sup>146</sup>, DOTASQ<sup>147</sup>, daunomycin<sup>148</sup>, L2H2-6OTD-dimer<sup>149</sup> and pyridostatin<sup>150</sup>.

- **G-tetrad intercalating  $\pi$ - $\pi$  stacking.**

On the basis of the same mechanism as above, ligands with the same planar  $\pi$ - $\pi$  system can be inserted into two stacked G-tetrads (fig. I.17-B). This binding mode is less common and there is no atomic resolution structural studies that support this binding type. However, for ligands such as TMPyP4, stoichiometry assays indicate 3:1 binding modes for tetramolecular

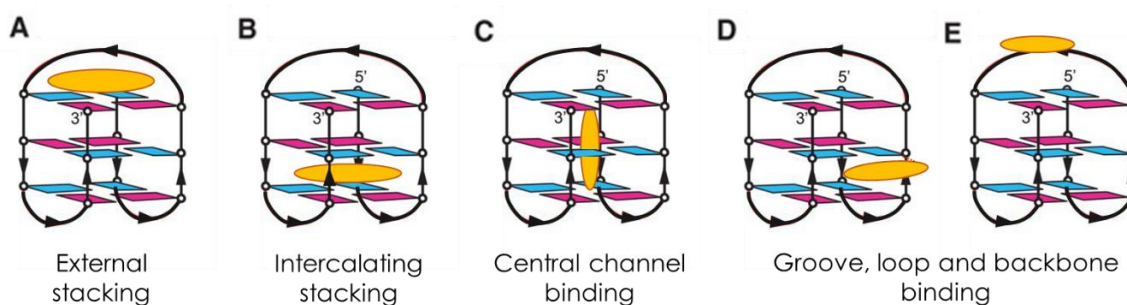
G-quadruplexes<sup>151,152</sup>. Since loops represent a physical hindrance, makes sense that this binding mode should work for these tetramolecular G-quadruplexes.

- **Central channel binding.**

As described for G-quadruplexes, the monovalent cation lies at the center of the G-tetrads stabilizing the overall structure. Using a charged aliphatic chain it is possible to thread across the tetrads serving the same purpose as the monovalent cations (fig. I.17-C). An anthracene polyammonium ligand is an example of this binding mode<sup>153</sup>. Another feature concerning the central ion channel is to develop cation coordinating groups on the  $\pi$ - $\pi$  stacking ligands so the ligands could coordinate with the central cation as well. TOxaPy is an example for this effect<sup>154</sup>.

- **Groove/loop binding.**

The loops and grooves are the features that make a G-quadruplex unique, thus ligands that are loop specific could behave as G-quadruplex topology-specific (fig. 1.17-D,E). Interactions of these nature are essentially electrostatic, thus requiring several charged and H-donor/acceptor groups. Examples include interesting alkaloid molecules as peimine<sup>155</sup>, TPE-1<sup>156</sup> and a Ru<sup>II</sup> complex Tatpp<sup>157</sup>. Interactions of these nature were structurally described for distamycin A<sup>158</sup> and an acridine derivative<sup>159</sup>.



**Figure I.17** – Representation of a ligand-G-quadruplex complex with A) external stacking mode on the terminal G-tetrad, B) intercalating mode between two stacks of G-tetrads, C) central channel binding, D) groove and E) loop binding mode, which typically both involves phosphodiester backbone interactions as well.

Since the synthesis of potential ligands involve large aromatic systems in order to cover most of the main interactions described above, serious problems concerning biocompatibility often arise, which are the most limiting factor for application of these ligands. Hence, one of the major challenges for designing G4 ligands is displaying pharmacokinetic parameters, otherwise it won't find clinical application. Therefore, upon design of new ligands the Lipinski recommendations should be considered<sup>160,161</sup>:

1. Minimization of molecular weight (Lipinski limit = 500 g/mol).
2. Optimization of permeability and uptake (modulation of log P and pK<sub>a</sub> values).
3. Avoidance of groups that are readily metabolized (such as esters, ketones and amides).
4. Avoidance of groups and elements that may be cytotoxic, such as nickel or lead metals and thiol groups.
5. Off-target toxicity (refers to adverse effects as a result of modulation of other targets that are unrelated to the target of interest<sup>162</sup>). Fortunately, G4 ligands are submitted to vast specificity assays, so this is not been a problem.

The majority of the therapeutic G4 ligands candidates fail in phase-I of the clinical trials.

## I.5 OBJECTIVES AND OUTLINE OF THIS THESIS

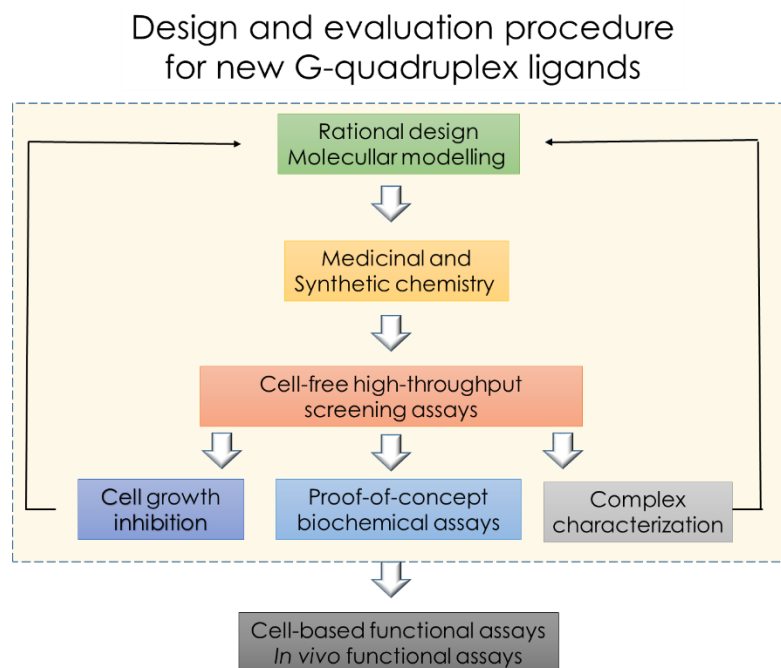
Over the last decade hundreds of G-quadruplex ligands were synthesized, some of them with outstanding affinity and specificity. Unfortunately, their biological properties were disappointing in most of the cases. Only a small minority of compounds was able to reach clinical trials stage.

In this context, the main objective of this dissertation is to create and evaluate oxazole-based phenanthroline compounds that may exhibit both biological applicability and G-quadruplex affinity. This is an iterative process that requires multiple interdisciplinary competences (fig. I.18).

The process starts with the design and synthesis of the compounds followed by identification of the binders to G4.

Subsequently, the promising compounds are selected for binding/stability detailed studies to G4 structures because it is desired to study the structural implications of ligand recognition and binding for future rational design studies. Finally, the cytotoxicity (IC<sub>50</sub>) of the ligands was evaluated in cancer and normal cells.

After refinement and optimization of the ligand properties, these would be in the right conditions to progress for pharmacological studies.

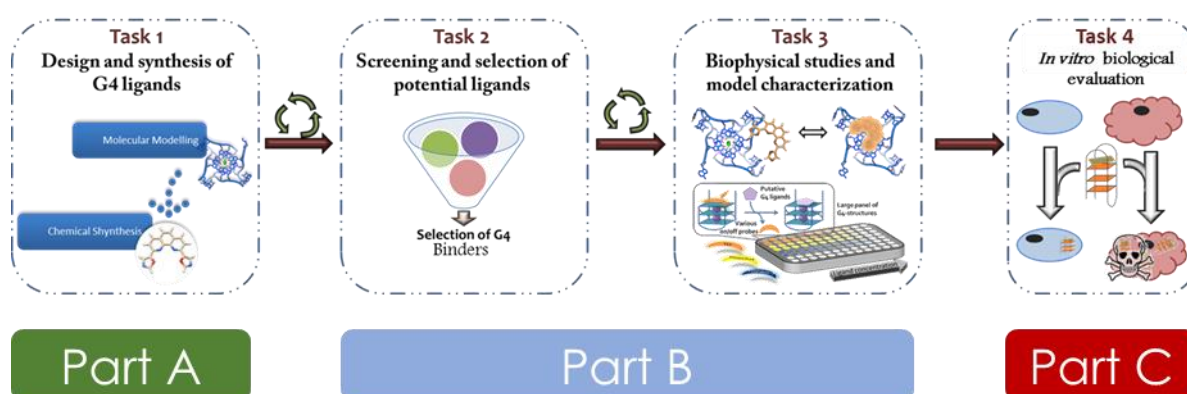


**Figure I.18** – Scheme of a general G-quadruplex drug-design procedure. This dissertation focus on the synthetic chemistry, screening, complex characterization and cell growth inhibitions steps.



## II. Results and Discussion

This chapter was divided into 3 parts: G4 ligands synthesis (part A), screening and selection of these binders for structure elucidation (part B) and finally a biological evaluation through *in vitro* assays (part C). In each part there is a small introduction, experimental section, discussion and conclusion.



**Scheme II.1** – Schematic representation of the main tasks performed in the present work.



# Part A

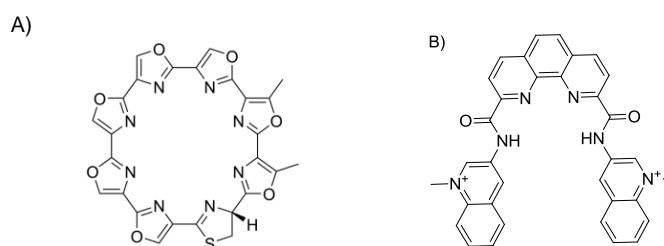
## Oligoheteroaryle G-quadruplex Ligands: design, synthesis and characterization

### 1 – INTRODUCTION

It has been demonstrated that small G-quadruplex-binding compounds can be employed as potential anticancer therapeutic agents, through indirect inhibition of telomerase activity or oncogene expression.

Some natural compounds (that curiously were found in Chinese traditional medicine), were already identified and proven to stabilize the G4 with reasonably selectivity such as berberine<sup>163</sup>, sanguerin<sup>163</sup>, isaindigotone<sup>164</sup> and telomestatin, which can be extracted from the bacteria *Streptomyces anulatus* 3533-SV4<sup>165</sup>.

Telomestatin (fig. II.1-A) is a polyoxazole macrocycle that selectively interacts with G-quadruplex by  $\pi$ - $\pi$  stacking or by hydrogen bonding as previously demonstrated by NMR and molecular dynamics studies<sup>166</sup>. Unfortunately, several concerns about drug-like properties discarded telomestatin as a therapeutic agent<sup>167</sup>. Nonetheless its discovery was undoubtedly a landmark on the search for G4 ligands, inspiring the synthesis of several oxazole-based compounds. Other classes of molecules were reported to exhibit high G4 selectivity such as Phen-DC<sub>3</sub><sup>168</sup> (fig. II.1-B), a phenanthroline bisquinolinium that revealed to fit perfectly on the G-tetrad. Thus phenanthroline and oxazole motifs could be used for designing molecules with affinity and biocompatibility to G-quadruplex.



**Figure II.1** – Chemical structure of A) Telomestatin and B) PhenDC<sub>3</sub>.

The main objective of this part of the work is to synthesise new polyoxazoles with phenanthroline unit followed by G-quadruplex binding evaluation through biochemical assays. The purity of the synthesied compounds is a major concern. The optimization of the synthetic steps in terms of yield will be performed whenever possible

## 2 – EXPERIMENTAL

### 2.1 – Materials and equipment

All solvents were purchased from Fluka® and employed without further purification. The solvents were evaporated in a Buchi vacuum system.

All reactions were monitored by thin-layer chromatography on 0.2 mm E. Merck® silica gel plates (60F-254) with UV light at 224nm and 334nm. When indicated, compounds were purified with a 50 centimeter chromatography column stacked with 60-200µm silica gel.

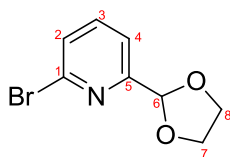
<sup>1</sup>H and <sup>13</sup>C NMR spectra were acquired in a 400MHz spectrometer operating at 400.13 MHz for proton and 100.62 MHz for carbon equipped with a BBFO probehead. The pulse programs applied were based on the corresponding templates provided Bruker pulse programs zg30 and zgpg30. Isotope labeled solvents were purchased from Cambridge Isotopes®.

Fluorescence spectra were acquired with a Perkin Elmer Precisely Luminiscence Spectrometer, model LS 45, at 25°C.

Qualitative docking studies were performed according the protocol established by G4LDB<sup>169</sup>.

### 2.2 – Chemical Synthesis

- **2-bromo-6-(1,3-dioxolan-2-yl)pyridine - intermediate reactant**



In a 250 mL flask, 6-bromo-2-carbaldehyde pyridine (2 g, 10.75 mmol, 1.0 eq.) and ethylene glycol (1.02 mL, 18.28 mmol, 1.7 eq.) were dissolved in benzene (50 mL). 4-methylbenzenesulfonic acid hydrate (0.10 g, 0.54 mmol, 0.05 eq.) was added and reaction mixture was refluxed under Dean-Stark conditions for 24h. After cooling to RT, mixture was quenched with a 1% aqueous Na<sub>2</sub>CO<sub>3</sub> solution. The organic phase was discarded, and the aqueous phase was extracted with CH<sub>2</sub>Cl<sub>2</sub> (5x100 mL). Organic phases were combined and concentrated under reduced pressure. The residue was purified by partition liquid chromatography Et<sub>2</sub>O/EtOH 95:5 affording 2-bromo-6-(1,3-dioxolan-2-yl)pyridine as an amber-like oil (R<sub>f</sub>≈0.75, 60% yield).

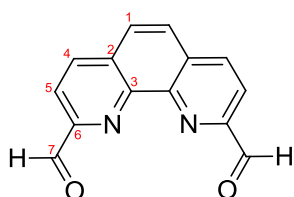
<sup>1</sup>H-NMR: (25°C, 400MHz, CDCl<sub>3</sub>) δ={7.58 (t,1H), <sup>3</sup>J =7.8 Hz; 7.47 (m,2H); 5.79 (s,1H); 4.16-4.03 (m,4H)};

<sup>13</sup>C-NMR: (25°C, 100MHz, CDCl<sub>3</sub>) δ={171.08, 1C; 141.63, 1C; 139.07, 1CH; 128.46, 1CH; 119.38, 1CH; 102.74, 1CH; 65.58 + 60.34, 2CH<sub>2</sub>}.

Corresponding spectra are displayed in the appendix A (supp. A.19).



- **1,10-phenantroline-2,9-dicarbaldehyde - "Phen-0"**



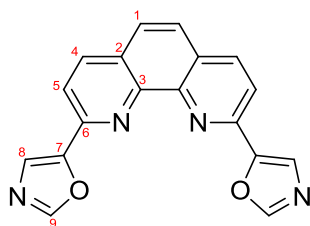
Out of a fresh batch, 4.0 g of selenium dioxide (4.5 eq, CAS# 7446-08-4, Acros Organics®) were dissolved in 100 mL of dioxane. In order to solubilize the selenium dioxide, a minimum amount of water (about 400  $\mu$ L) was added. 2.0 g of neocuproine (1eq, CAS# 484-11-7, Aldrich®) were dissolved in 60 mL of dioxane and added drop-wise to the refluxed solution. The reaction was carried out by 12 h. The selenium by-products were removed by filtration in hot conditions and the filtrate's solvent was evaporated in a vacuum assisted system to obtain a pale-yellow solid. The residue was resuspended in  $\text{CHCl}_3$  and any insoluble particles were filtered off. Chloroform was evaporated and Phen-0 was obtained as a white solid, with an approximate 60% yield.

$^1\text{H-NMR}$ : (25°C, 400MHz,  $\text{CDCl}_3$ )  $\delta$ ={10.49 (s,2H); 8.44 (d,2H),  $^3\text{J}$ =8.6 Hz; 8.31 (d,2H),  $^3\text{J}$ =8.6 Hz; 7.98 (s,2H)};

$^{13}\text{C-NMR}$ : (25°C, 100MHz,  $\text{CDCl}_3$ )  $\delta$ ={192.32, 2CH; 151.65, 2C; 144.91, 2C; 136.91, 2CH; 130.60, 2C; 128.02, 2CH}.

Corresponding spectra are displayed in appendix A (supp. A.5).

- **2,9-bis(oxazole-5-yl)-1,10-phenantroline – "Phen-1"**



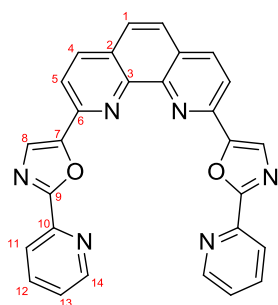
In a 100 mL round-bottomed flask, 2.34 g of potassium carbonate (4 eq, CAS# 584-08-7, anhydrous, Chem Lab®) and 1.64g of p-(tolylsulfonyl)methylisocyanide (a.k.a. TOSMIC, 2 eq, CAS# 366635-61-7, Acros Organics®) were suspended in 80 mL of methanol and stirred for 20 min at room temperature. 1.0 g of Phen-0 were stirred in 15 mL of methanol and added drop-wise to the solution in the flask. The solution was stirred for 2.5 hours at room temperature and other 3 hours at the reflux temperature, by this order. The solvent was evaporated (low pressure assisted) and the residue obtained was washed with water in order to eliminate TosMIC and salt by-products. The resulting precipitate was again washed with chloroform and methanol in order to remove other impurities. The compound was finally extracted with boiling-hot dioxane and the non-soluble impurities were removed by filtration. Dioxane was then evaporated affording 2,9-bis(oxazole-5-yl)-1,10-phenantroline (phen-1) as an orange solid (45% yield). No further purification procedures were applied.

$^1\text{H-NMR}$ : (25°C, 400MHz,  $\text{DMSO-}d_6$ )  $\delta$ ={8.46 (s,2H); 8.40 (d,2H),  $^3\text{J}$ =8.25 Hz; 7.97 (d,2H),  $^3\text{J}$ =8.25 Hz, 7.97 (s,2H); 7.80 (s,2H)};

$^{13}\text{C-NMR}$ : (25°C, 100MHz,  $\text{DMSO-}d_6$ )  $\delta$ ={153.77, 2CH; 151.22, 2C; 146.70, 2C; 145.50, 2C; 138.24, 2CH; 128.78, 2C; 128.78, 2CH; 127.23, 2CH; 120.04, 2CH}.

Corresponding spectra are displayed in the appendix A (supp. A.6).

- **2,9-bis(2-(pyridine-2-yl)oxazole-5-yl)-1,10-phenantroline – “Phen-2”**



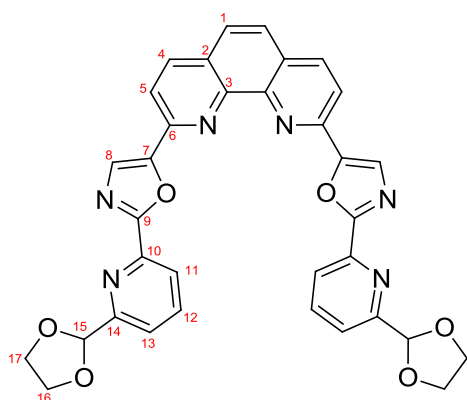
100 mg of phen-1 (1 eq), 24.2 mg of tricyclohexylphosphoniumtetrafluoroborate (0.2 eq, CAS# 58656-04-5, Acros Organics®), 186.2mg of copper iodate (3 eq, CAS# 7681.65.4, Fluka) and 30 mg of palladium(II)-acetate (0.4eq, CAS# 3375-31-3, Acros Organics®) were suspended in 5 mL of dioxane in a 50 mL argon flushed round bottomed flask. 504.40mg of cesium carbonate (4.4eq, CAS# 534-17-8, Acros Organics®) were added and the mixture was heated, then 121 $\mu$ L of 1,6-bromopyridine (4eq, CAS# 109-04-6, Aldrich®) were added and the mixture was refluxed at 125 °C for 24h in argon atmosphere. The solvent was evaporated and the black mixture was suspended in a minimum amount of water, followed by 5x50 mL extraction with chloroform. The organic phases were discarded and the blue aqueous phase was filtered. The filter containing the non-soluble residue was collected and boiled at 80°C in dioxane for a few minutes. While still hot, the solution was once again filtered, the black residue was discarded and the filtrate was evaporated affording a yellow residue. A final extraction with chloroform was applied in order to remove trace impurities affording pure phen-2 (20% yield).

<sup>1</sup>H-NMR: (25°C, 400MHz, DMSO-*d*<sub>6</sub>)  $\delta$ ={8.44 (d,2H), <sup>3</sup>J=8.50 Hz; 8.85 (d,2H), <sup>3</sup>J=4.08 Hz; 8.25 (s,2H); 8.13 (d,2H), <sup>3</sup>J=7.90 Hz; 8.80 (d,2H), <sup>3</sup>J=8.50 Hz; 8.06 (t,2H), <sup>3</sup>J=7.73 Hz; 7.89 (s,2H); 7.53 (dd,2H), <sup>3</sup>J=5.16 Hz, 1.80 Hz};

<sup>13</sup>C-NMR: (25°C, 100MHz, DMSO-*d*<sub>6</sub>)  $\delta$ ={161.10, 2C; 151.45, 2C; 149.90, 2CH; 145.73, 2C; 145.17, 2C; 144.80, 2C; 138.25, 4CH; 130.17, 2CH; 128.66, 2C; 127.35, 2CH; 126.17, 2CH; 123.21, 2CH; 120.60, 2CH}.

Corresponding spectra are displayed in the appendix A (supp. A.7).

- **2,9-bis(2-(6-(1,3-dioxolan-2-yl)pyridine-2-yl)oxazole-5-yl)-1,10-phenantroline – “Phen-3”**



300 mg of phen-1 (1eq), 105.0 mg of tricyclohexylphosphoniumtetrafluoroborate (0.3 eq, CAS# 58656-04-5, Acros Organics®), 398.0 mg of copper iodate (2.2 eq, CAS# 7681.65.4, Fluka) and 85.3 mg of palladium(II)-acetate (0.4 eq, CAS# 3375-31-3, Acros Organics®) were suspended in 15 mL of dioxane in a 50 mL argon flushed round bottomed flask. 1.47 g of cesium carbonate

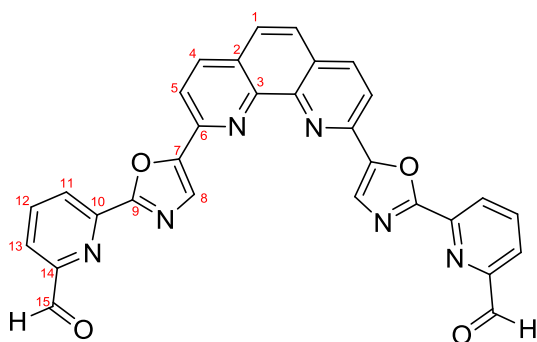
(4.4 eq, CAS# 534-17-8, Acros Organics®) were added and the mixture was heated, then 740.0 mg of 2-bromo-6-(1,3-dioxolan-2-yl)pyridine were added and the mixture was refluxed at 125 °C for 24 h in argon atmosphere. The solvent was evaporated and the dark green mixture was filtered and washed in a vacuum assisted system with ethyl acetate, followed by water too. The filter containing the residue was boiled in dioxane affording a dark-yellow solution. The solution was filtered in hot conditions, the black residue was discarded and the filtrate was evaporated affording phen-3 as a yellow residue (10% yield). No further purification procedures were applied.

$^1\text{H-NMR}$ : (25°C, 400MHz, DMSO- $d_6$ )  $\delta$ ={10.19 (s,2H); 8.78 (d,2H),  $^3\text{J}$ =8.40 Hz; 8.66 (d,2H),  $^3\text{J}$ =7.44 Hz; 8.57 (s,2H); 8.43 (d,2H),  $^3\text{J}$ =8.40 Hz; 8.38 (t,2H),  $^3\text{J}$ =7.76 Hz; 8.18 (d,2H),  $^3\text{J}$ =7.08 Hz; 8.16 (s,2H)};

$^{13}\text{C-NMR}$ : (25°C, 100MHz, DMSO- $d_6$ )  $\delta$ ={(193.35, 2C; 160.37, 2C; 153.04, 2C; 152.90, 2C; 146.21, 4C; 145.65, 2C; 139.92, 2CH; 138.41, 2CH; 129.60, 2CH; 129.02,2C; 127.49, 2CH; 127.20,2CH; 123.52, 2CH; 120.35, 2CH) }.

Corresponding spectra are displayed in the appendix A (supp. A.8).

- **6,6'-(5,5'-(1,10-phenanthroline-2,9-diyl)bis(oxazole-5,2-diyl)dipicolinaldehyde – “Phen-4”**



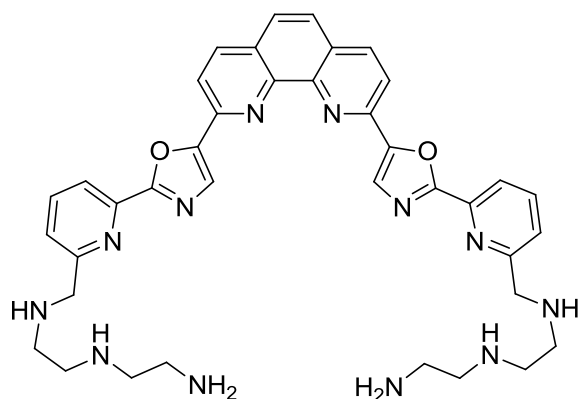
10 mL of HCl (2 M) were poured into 100 mg of Phen-3 and the solution was refluxed at 110 °C during 2 hours, forming a visible precipitate. The solution was neutralized with a saturated solution of  $\text{Na}_2\text{CO}_3$ . A brown suspension was formed. The suspension was filtered and washed with cool water and dichloromethane, affording pure Phen-4 as a dark yellow residue (80% yield).

$^1\text{H-NMR}$ : (25°C, 400MHz, DMSO- $d_6$ )  $\delta$ ={10.19 (s,2H); 8.78 (d,2H),  $^3\text{J}$ =8.40 Hz; 8.66 (d,2H),  $^3\text{J}$ =7.44 Hz; 8.57 (s,2H); 8.43 (d,2H),  $^3\text{J}$ =8.40 Hz; 8.38 (t,2H),  $^3\text{J}$ =7.76 Hz; 8.18 (d,2H),  $^3\text{J}$ =7.08 Hz; 8.16 (s,2H)};

$^{13}\text{C-NMR}$ : (25°C, 100MHz, DMSO- $d_6$ )  $\delta$ ={(193.35, 2C; 160.37, 2C; 153.04, 2C; 152.90, 2C; 146.21, 4C; 145.65, 2C; 139.92, 2CH; 138.41, 2CH; 129.60, 2CH; 129.02,2C; 127.49, 2CH; 127.20,2CH; 123.52, 2CH; 120.35, 2CH) }.

Corresponding spectra are displayed in the appendix A (supp. A.9).

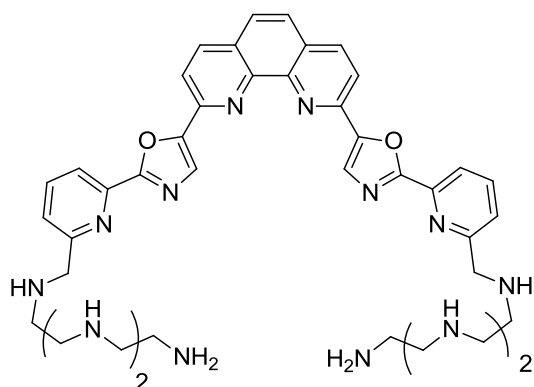
- **N1,N1'-((6,6'-(5,5'-(1,10-phenanthroline-2,9-diyl))bis(oxazole-5,2-diyl))bis(pyridine-6,2-diyl))bis(methylene))bis(N2-(2-aminoethyl)ethane-1,2-diamine)**



54  $\mu\text{L}$  of diethylenetriamine (25 eq, 99% purity, Acros Organics®) was dissolved in 50 mL of a dried acetonitrile-ethanol mixture. 10 mg of Phen-4 (1 eq) were dissolved in 50 mL of the same solvent and was added dropwise to the diethylenetriamine solution, for a period of 6h under vigorous stirring. The reaction was left for a 6h under vigorous stirring. The reaction was carried over at 10h at room temperature (25-30 °C). The imine reduction was performed with sodium borohydride (effervescence should be noticed). After 7h the solvent was evaporated under reduced pressure and extracted with water/chloroform (4 x 50 mL). Organic phases were combined and the solvent was evaporated. The obtained residue was suspended in methanol/ethanol solution and the compound was precipitated with HCl 2M, affording a mixture of 1:2 and 2:2 substituted compounds. No further purification methods were employed.

Corresponding spectra are displayed in the appendix A (supp. A.10).

- **N1,N1'-((6,6'-(5,5'-(1,10-phenanthroline-2,9-diyl))bis(oxazole-5,2-diyl))bis(pyridine-6,2-diyl))bis(methylene))bis(N2-(2-((2-aminoethyl)amino)ethyl)ethane-1,2-diamine)**

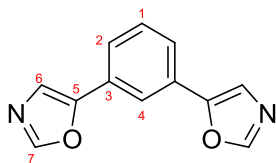


90  $\mu\text{L}$  of triethylenetetramine (30eq, 60% purity, Acros Organics®) were dissolved in 50 mL of a dried acetonitrile-ethanol mixture. 10 mg of Phen-4 (1eq) were dissolved in 50 mL of the same mixture and was added dropwise, for a period of 6h under vigorous stirring, to the triethylenetetramine solution. The reaction was carried for a 6h period under vigorous stirring. The reaction was carried over a 10h period at room temperature (25-30°C). The imine reduction was performed with sodium borohydride (effervescence should be noticed). After 7h, the solvent was evaporated under reduced pressure and extracted with water/chloroform (4 x 50 mL). Organic phases were combined and the solvent was evaporated. The obtained residue

was suspended in methanol/ethanol solution and the compound was precipitated with HCl 2M, affording a mixture of 1:2 and 2:2 compounds. No further purification methods were employed.

Corresponding spectra are displayed in the appendix A (supp. A.11).

- **1,3-bis(oxazol-5-yl)benzene – “Iso-1”**



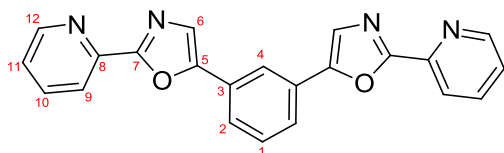
In a 100 mL round-bottomed flask, 4.12 g of potassium carbonate (4 eq, CAS# 584-08-7, anhydrous, Chem Lab®) and 2.90 g of p-(tolylsulfonyl)methylisocyanide (aka TosMIC) (2 eq, CAS# 366635-61-7, Acros Organics®) were suspended in 60 mL of methanol and stirred for 20 min at room temperature. 1.0 g of isophthalaldehyde (1 eq, CAS# 626-19-7, Sigma-Aldrich®) was stirred in 15 mL of methanol and added drop-wise to the solution in the flask. The solution was vigorously stirred for 2.5 hours at room temperature and other 3 hours at reflux temperature. A change of color should occur upon heating. The solvent was evaporated under reduced pressure and the orange residue was suspended in 2 x 60 mL of diethyl ether. Organic phases were filtered, combined and evaporated sob vacuum. The obtained orange residue was stirred in methanol resulting in a yellow precipitate and a red solution. Iso-1 was collected by filtration as a pale-yellow pure solid (65% yield).

<sup>1</sup>H-NMR: (25°C, 400MHz, CDCl<sub>3</sub>) δ={8.09 (s,2H); 7.96 (s,1H); 7.65 (d,2H), <sup>3</sup>J =7.69 Hz; 7.52 (t, 1H), <sup>3</sup>J =7.71 Hz; 7.45 (s, 2H)};

<sup>13</sup>C-NMR: (25°C, 100MHz, CDCl<sub>3</sub>) ={150.95, 2C; 150.86, 2CH; 129.66, 1CH; 127.36, 2C; 124.55, 2CH; 121.87, 1CH; 120.15, 1CH}.

Corresponding spectra are displayed in the appendix A (supp. A.12).

- **1,3-bis(2-(pyridine-2-yl)oxazol-5-yl)benzene – “Iso-2”**



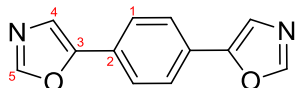
200.0mg of Iso-1 (1 eq), 71.6 mg of tricyclohexylphosphoniumtetrafluoroborate (0.2 eq, CAS# 58656-04-5, Acros Organics®), 407.2 mg of copper iodate (2.2 eq, CAS# 7681.65.4, Fluka) and 91.0 mg of palladium(II)-acetate (0.4 eq, CAS# 3375-31-3, Acros Organics®) were suspended in 6 mL of dried dioxane in 50 mL argon flushed round bottomed flask. 1.473 g of cesium carbonate (4.4 eq, CAS# 534-17-8, Acros Organics®) were added and the mixture was heated. At this point, 296.0 μL of 1,6-bromopyridine (3.3 eq, CAS# 109-04-6, Aldrich®) were added and the mixture was refluxed at 125 °C for 24 h in argon atmosphere. Additional 5 mL of dioxane were added after 12h reaction. The solvent was evaporated and the black mixture was suspended in a minimum amount of water, followed by extraction with ethyl acetate (5 x 50 mL). Organic phases were combined and evaporated given a brown oil. The compound was isolated by column liquid chromatography Et<sub>2</sub>O/propan-2-ol 90:10 affording Iso-2 as an orange residue (R<sub>f</sub>≈0.20, 20 % yield). Monosubstituted Iso-1.5 was also isolated as an orange residue (R<sub>f</sub>≈0.12, 8% yield).

$^1\text{H-NMR}$ : (25°C, 400MHz,  $\text{CDCl}_3$ )  $\delta$ ={8.72 (d,2H),  $^3\text{J}$ =4.75 Hz; 8.15 (d,2H),  $^3\text{J}$ =7.84 Hz; 8.10 (s,1H); 7.79 (t,2H),  $^3\text{J}$ =7.61 Hz; 7.72 (d,2H),  $^3\text{J}$ = 7.76 Hz; 7.57 (s,2H)  $^3\text{J}$ =; 7.46 (t,1H)  $^3\text{J}$ =7.71 Hz; 7.33 (dd,2H)  $^3\text{J}$ =5.16 Hz, 0.80 Hz};

$^{13}\text{C-NMR}$ : (25°C, 100MHz,  $\text{CDCl}_3$ )  $\delta$ ={159.26, 2C; 150.84, 2C; 149.06, 2CH; 144.98, 2C; 136.02, 2CH; 128.58, 1CH; 127.44, 2C; 123.96, 2CH; 123.71+123.69, 3CH; 121.26, 2CH; 119.58, 2CH}.

Corresponding spectra are displayed in the appendix A (supp. A.13).

- **1,4-bis(oxazol-5-yl)benzene – “Ter-1”**



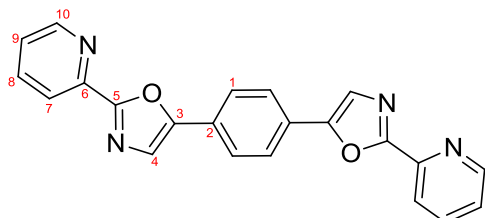
In a 100 mL round-bottomed flask, 4.12 g of potassium carbonate (4 eq, CAS# 584-08-7, anhydrous, Chem Lab®) and 2.90 g of p-toluenesulfonylmethyl isocyanide (aka TosMIC) (2 eq, CAS# 366635-61-7, Acros Organics®) were suspended in 60 mL of methanol and stirred for 20 min at room temperature. 1.0 g of terephthaldehyde (1 eq, CAS# 626-19-7, Sigma-Aldrich®) was stirred in 15 mL of methanol and added drop-wise to the solution in the flask. The solution was vigorously stirred for 2.5 hours at room temperature and other 3 hours at reflux temperature. A change of color should occur upon heating. The solvent was evaporated under reduced pressure and the orange residue was extracted with diethyl ether (2 x 60 mL). Organic phases were filtered, combined and evaporated in vacuum. Ter-1 was collected by filtration as a pale-yellow solid (70% yield).

$^1\text{H-NMR}$ : (25°C, 400MHz,  $\text{CDCl}_3$ )  $\delta$ = {7.86 (s,2H); 7.63 (s,4H); 7.33 (s,2H)};

$^{13}\text{C-NMR}$ : (25°C, 100MHz,  $\text{CDCl}_3$ )  $\delta$ = {149.89, 2C; 149.89, 2CH; 126.75, 2C, 123.84, 4CH; 121.11, 2CH}.

Corresponding spectra are displayed in the appendix A (supp. A.14).

- **1,4-bis(2-(pyridine-2-yl)oxazol-5-yl)benzene – “ter-2”**



100.0 mg of ter-1 (1 eq), 35.5mg of tricyclohexylphosphoniumtetrafluoroborate (0.2 eq, CAS# 58656-04-5, Acros Organics®), 202.0 mg of copper iodate (2.2 eq, CAS# 7681.65.4, Fluka) and 45.0 mg of palladium(II)-acetate (0.4 eq, CAS# 3375-31-3, Acros Organics®) were suspended in 4 mL of dried dioxane in a 50 mL argon flushed round bottomed flask. 741.0 mg of cesium carbonate (4.4 eq, CAS# 534-17-8, Acros Organics®) were added and the mixture was heated. At this point, 150.0  $\mu\text{L}$  of 1,6-bromopyridine (3.3 eq, CAS# 109-04-6, Aldrich®) were added and the mixture was refluxed at 125 °C for 24 h in argon atmosphere. Additional 5 mL of dioxane were added after 12 h reaction. The solvent was evaporated and the black mixture was suspended in a minimum amount of water, followed by extraction with ethyl acetate (5 x 50mL). Organic phases were combined and evaporated affording a brown oil. The compound was separated by column chromatography Et<sub>2</sub>O/propan-2-ol 90:10 ( $R_f$ ≈0.23). The obtained residue was washed with cyclohexane, dissolved in dichloromethane and the blue suspension was discarded. Dichloromethane was evaporated affording ter-2 as an orange residue (35% yield). Monosubstituted ter-1.5 was also isolated as a red residue ( $R_f$ ≈0.12, 1.5% yield).

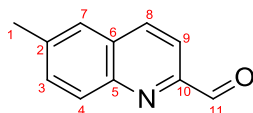


$^1\text{H-NMR}$ : (25°C, 400MHz,  $\text{CDCl}_3$ )  $\delta$ ={8.72 (d,2H),  $^3J$ =4.75 Hz; 8.13 (d,2H),  $^3J$ =7.33 Hz; 7.80 (s,4H); 7.78 (t,2H),  $^3J$ =7.54 Hz; 7.52 (s,2H); 7.32 (d,2H),  $^3J$ =5.56 Hz};

$^{13}\text{C-NMR}$ : (25°C, 100MHz,  $\text{CDCl}_3$ )  $\delta$ ={159.23, 2C; 150.90, 2C; 149.11, 2CH; 144.99, 2C; 135.99, 4CH; 126.73, 2C; 124.06 2CH; 123.67 2CH; 123.52, 2CH; 121.23, 2CH}.

Corresponding spectra are displayed in the appendix A (supp. A.15).

• **6-methylquinoline-2-carbaldehyde – “quin-0”**



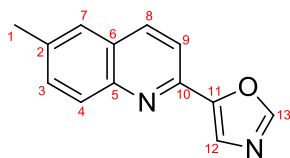
Freshly opened 5.9 g of selenium dioxide (4.5 eq, CAS# 7446-08-4, Acros Organics®) were dissolved in 100 mL of dioxane (a minimum amount of water, about 400  $\mu\text{L}$ , was added in order to solubilize the selenium dioxide). 2.0 g of 2,6-dimethylquinoline (1 eq, CAS# 877-43-0, Aldrich®) were dissolved in 60 mL of dioxane and added drop-wise to the refluxed solution. The reaction was carried out by 18 h. The selenium by-products were filtered and dioxane was evaporated in a vacuum system resulting a dark-brown solid. The residue was resuspended in  $\text{CHCl}_3$  and the insoluble particles were filtered off. The solvent was evaporated and Quin-0 was obtained as a brown solid, with an approximate 80% yield.

$^1\text{H-NMR}$ : (25°C, 400MHz,  $\text{CDCl}_3$ )  $\delta$ ={9.13 (s,1H); 8.12 (d,1H),  $^3J$ =8.62 Hz; 8.05 (d,1H),  $^3J$ =8.62 Hz; 7.92 (s,1H); 7.89 (d,1H),  $^3J$ =8.62 Hz; 7.58 (m,2H); 2.5 (s,9H)};

$^{13}\text{C-NMR}$ : (25°C, 100MHz,  $\text{CDCl}_3$ )  $\delta$ ={192.50, 1CH; 150.91, 1C; 145.46, 1C; 138.72, 1C; 135.51, 1CH; 131.88, 1CH; 129.11, 1C; 128.93, 1CH; 125.70, 1CH; 116.23, 1CH; 20.85, 1CH $_3$ }.

Corresponding spectra are displayed in the appendix A (supp. A.16).

• **6-methyl-(2-oxazol-5-yl)quinoline – “quin-1”**



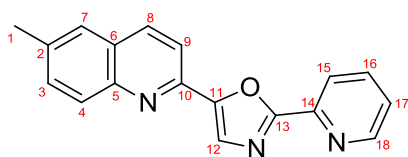
In a 50 mL round-bottomed flask, 1.65 g of potassium carbonate (2 eq, CAS# 584-08-7, anhydrous, Chem Lab®) and 1.17 g of p-toluenesulfonylmethyl isocyanide (aka TOSMIC) (1eq, CAS# 366635-61-7, Acros Organics®) were suspended in 30 mL of methanol and stirred for 20 min at room temperature. 1.17 g of Quin-0 (1 eq, CAS# 626-19-7, Sigma-Aldrich®) were stirred in 10 mL of methanol and added drop-wise to the solution in the flask. The solution was stirred for 2.5 hours at room temperature and other 2.5 hours at the reflux temperature, by this order. The solvent was evaporated under reduced pressure and the product was dissolved in a minimum amount of cool water (about 10 mL). Quin-1 was extracted from the aqueous phase by liquid extraction with ether (5 x 50 mL) and it was obtained as a dark orange solid (85% yield) after solvent evaporation. No further purification procedures were applied.

$^1\text{H-NMR}$ : (25°C, 400MHz,  $\text{CDCl}_3$ )  $\delta$ ={8.56 (d, 1H),  $^3J$ =8.06 Hz; 7.96 (s, 1H); 7.95 (d, 1H),  $^3J$ =8.62 Hz; 7.75, (s, 1H); 7.66 (d, 1H),  $^3J$ =8.62 Hz; 7.49+7.48, (m, 2H); 2.45, (s, 3H); 2.45 (s, 9H)};

$^{13}\text{C-NMR}$ : (25°C, 100MHz,  $\text{CDCl}_3$ )  $\delta$ ={151.55, 1CH; 151.35, 1C; 146.59, 1C; 145.93, 1C; 137.10, 1C; 136.46, 1CH; 132.62, 1CH; 129.10, 1CH; 127.58, 1C; 126.57, 1CH; 125.74, 1CH; 117.63, 1CH; 21.63, 3CH $_3$ }.

Corresponding spectra are displayed in the appendix A (supp. A.17).

- **6-methyl-(2-(pyridine-2-yl)oxazol-5-yl)quinoline–“quin-2”**



150.0 mg of quin-1 (1 eq), 34.9 mg of tricyclohexylphosphoniumtetrafluoroborate (0.2 eq, CAS# 58656-04-5, Acros Organics®), 200.0 mg of copper iodate (2.2 eq, CAS# 7681.65.4, Fluka) and 44.6 mg of palladium(II)-acetate (0.4 eq, CAS# 3375-31-3, Acros Organics®) were suspended in 4 mL of dried dioxane in a 50 mL argon flushed round bottomed flask. 746.0 mg of cesium carbonate (4.4 eq, CAS# 534-17-8, Acros Organics®) were added and the mixture was heated. At this point, 75.0  $\mu$ L of 1,6-bromopyridine (3.3 eq, CAS# 109-04-6, Aldrich®) were added and the mixture was refluxed at 125  $^{\circ}$ C for 24 h in argon atmosphere. Additional 5 mL of dioxane were added after 12 h reaction. The solvent was evaporated and the black mixture was suspended in a minimum amount of water, followed by extraction with ethyl acetate (5 x 50 mL). Organic phases were combined and evaporated affording a brown oil. The compound was isolated by partition liquid chromatography Et<sub>2</sub>O/propan-2-ol 90:10 affording quin-2 ( $R_f \approx 0.25$ , 43%) as a brown residue.

<sup>1</sup>H-NMR: (25 $^{\circ}$ C, 400MHz, CDCl<sub>3</sub>)  $\delta$ ={8.80 (d,1H), <sup>3</sup>J=4.05 Hz; 8.22 (d,2H), <sup>3</sup>J=7.87 Hz; 8.12 (d,1H) <sup>3</sup>J=8.60 Hz; 8.02 (s,1H); 8.00 (s,1H); 7.92 (d,2H), <sup>3</sup>J=8.56 Hz; 7.84 (t,1H) <sup>3</sup>J= 6.91 Hz; 7.56+7.54, (t+s,2H); 7.39, (dd,1H), <sup>3</sup>J= 5.30, 1.8Hz; 2.52 (s,9H)};

<sup>13</sup>C-NMR: (25 $^{\circ}$ C, 100MHz, CDCl<sub>3</sub>)  $\delta$ ={160.83, 1C; 152.34, 1C; 150.11, 1CH; 146.72, 1C; 145.99, 1C; 145.92, 1C; 136.96, 1CH, 163.93, 1C; 136.25, 1CH; 132.48, 1CH; 129.06, 1CH; 127.78, 1CH; 127.64, 1C; 126.53, 1CH; 124.83, 1CH, 122.48, 1CH, 117.84, 1CH, 21.62, 3CH<sub>3</sub>}

Corresponding spectra are displayed in the appendix A (supp. A.18).



## 3 – RESULTS AND DISCUSSION

### 3.1 – Molecular design

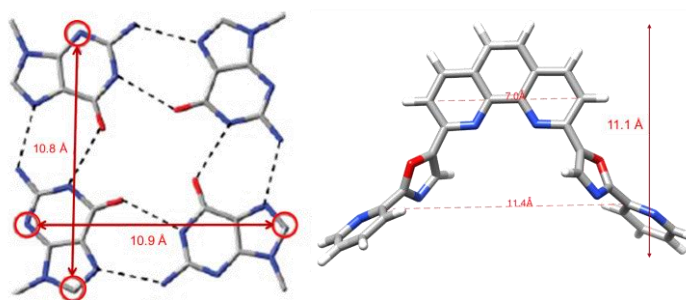
As mentioned previously, telomestatin is a natural cyclic polyoxazole that revealed affinity and selectivity towards G-quadruplex structures. Its large cyclic shape results in a high circular surface area that might be the main factor for absence of affinity towards duplex structures. Moreover this natural compound exhibited impressive antitumor activity, but its physico-chemical properties failed to present good drug like properties, especially concerns about its very low aqueous solubility.

The high selectivity and affinity of compound telomestatin towards G-quadruplexes in addition to its potent anti-proliferative effects has prompted drug designers to develop several molecular analogues. Macrocyclic ligands L1H1-7OTD<sup>170</sup>, S2T1-6OTD<sup>171</sup>, HXDV<sup>172</sup> and 6OTD-dimer<sup>173</sup> are some of these examples (for structural information see supp. A.1). Telomestatin also inspired synthesis of other polyoxazole ligands but as acyclic compounds, since it could mimic the macrocyclic shape of telomestatin in a horse-shoe shape. Some of these acyclic examples are two phenyl bis-oxazole derivatives<sup>174</sup>, TOxaPy and BOxaPy<sup>154</sup>.

It was demonstrated that oxazole-containing macrocycles represent a promising class of G-quadruplex-targeting anticancer agents. Oxazole-based acyclic motif also demonstrated to have the same potential, as verified with BOxaPy and TOxaPy compounds. The main advantage for this acyclic motif is the increasing in flexibility and adaptability upon binding to G-tetrads.

Phenanthroline scaffolds also defined other landmark on G-quadruplex ligand design. Phenanthroline architecture is very convenient, its rings are spaced in the optimal manner for  $\pi$ - $\pi$  stacking and nitrogen atoms are in position to perform hydrogen bonds with the guanine tetrad or with the cations. We believe that this scaffold is essential for G-tetrad stacking and indubitably increases selectivity against duplex secondary structures.

Considering these facts, the combination of phenanthroline moiety and oxazole groups with an acyclic crescent shape like BOxaPy could result in an interesting planar surface for  $\pi$ - $\pi$  stacking interactions that could in principle fit precisely in the G-quartet (fig. II.2). Moreover, the heteroatoms would be placed in favorable position for central-cation coordination and the nitrogen atoms could in principle perform hydrogen bonds.

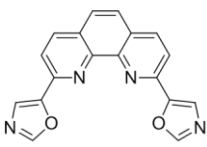


**Figure II.2**– Structure of a G-tetrad form the parallel h22t G-quadruplex (PDB: 1KF1) compared with structure of the energy minimized ligand Phen-2, compared to scale. Noteworthy, Phen-2 dimensions correspond to the G-tetrad dimensions. A directed qualitative docking procedure indicates interaction potential between the G-tetrad and the designed ligand (supp. A.2).

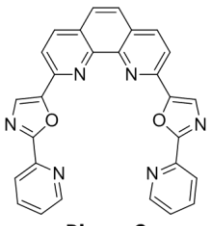
Considering Lipinski's recommendations for drug-design, some basic physicochemical properties were investigated for these ligands (fig. II.3). The two designed ligands (Phen-1 and

Phen-2, fig. II.3) fulfill the basic requirements for drug-like properties, however we can anticipate that the main drawback for these ligand will probably concern aqueous solubility.

Calculated parameters		
4.05	Log P	4.05
77.85	Polar Surface Area (Å <sup>2</sup> )	103.63
314.30	Molecular Weight (g/mol)	468.47
2	Rotatable Bonds	4



**Phen-1**



**Phen-2**

**Figure II.3** – Lipinski's drug-like parameters calculated using the service from molinspiration.com web-site<sup>175</sup>. Polar surface area and Log P were respectively calculated using the methodology reported for TPSA<sup>176</sup> and miLogP<sup>175</sup>.

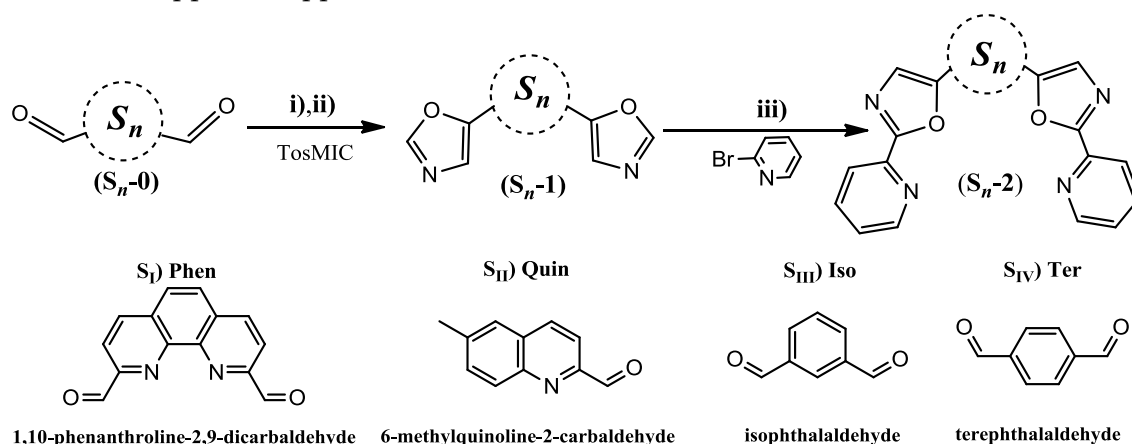
In order to enhance selectivity and solubility, our scaffold may be further functionalized with additional aliphatic amines or charged groups, as similarly applied to other G-quadruplex ligands. The spacing between the inserted aliphatic amine groups could play an important role in adaptation of the ligand to the quadruplex phosphate backbone through hydrogen bonding, which could result in specific binding to G-quadruplex topologies.

Overall the designed ligands could represent a basic motif for specific G-quadruplex recognition. These also exhibit the recommended basic drug-like physicochemical properties, but we are still in a very preliminary stage for assessing biological applicability.

For drug design comparison, it would be interesting to perform an extended study based on importance of the main scaffold on the overall recognition to G4, such as size and symmetry. For this reason, other three scaffolds available in the laboratory were also selected; these were isophthalaldehyde, terephthalaldehyde and 6-methylquinoline-2-carboxaldehyde.

### 3.2 – Molecular Synthesis

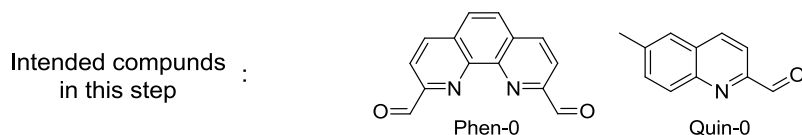
In order to synthesize the proposed ligands, a convergent procedure based on van Leusen chemistry and direct cross-coupling of activated C-H with heteroarene halides was employed with a similar approach applied in recent studies<sup>154,177</sup>.



**Scheme II.2** – Synthetic pathway for the synthesis of the proposed potential G4 ligands. i) p-toluenesulfonylmethyl isocyanide (TosMIC), K<sub>2</sub>CO<sub>3</sub>, MeOH, 2h low temperature; ii) 4h reflux temperature; iii) P(cyclohexyl)•3HBF<sub>4</sub>, Pd(OAc)<sub>2</sub>, CuI, cesium carbonate, 2-bromo-pyridine, dioxane, 130°C 24h.

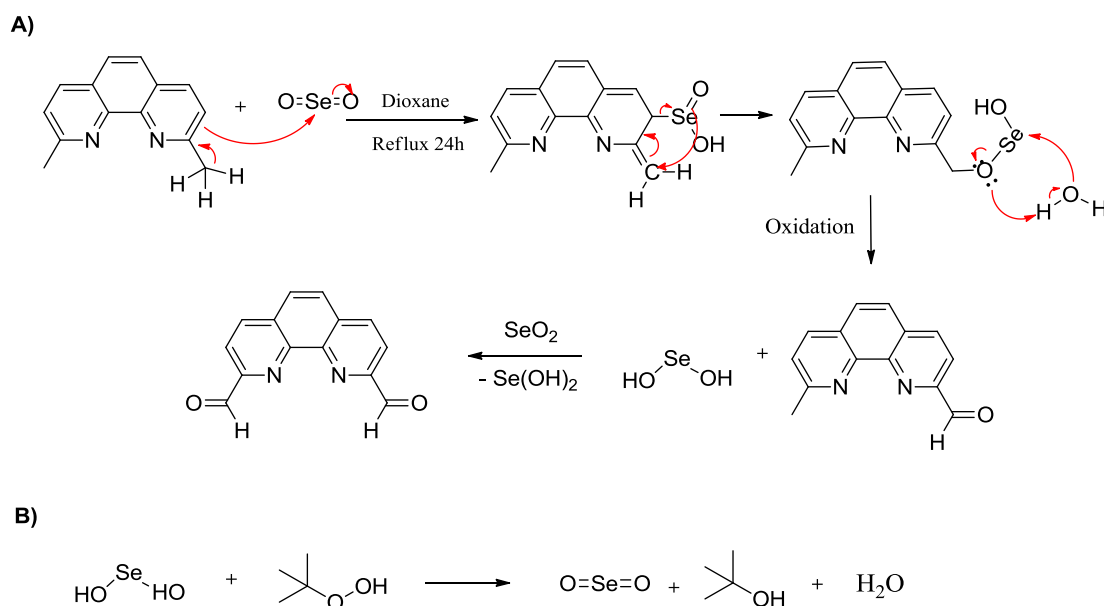
The procedure starts with aldehyde oxidation with TosMIC; but before that, a preliminary step is required in order to synthesize the dialdehyde form. Isophthalaldehyde and terephthalaldehyde are already commercially available; 1,10-phenanthroline-2,9-dicarbaldehyde and 6-methyl-quinoline-2-carbaldehyde were obtained through allylic reduction of neocuproine and 2,6-dimethylquinoline, respectively, with selenium dioxide.

- **Synthesis of the starting materials - allylic reduction with selenium dioxide**



**Scheme II.3** – Illustration of the intended compounds in this topic.

Selenium dioxide and its by-products are usually hazardous materials and very difficult to get rid-of. Therefore, in our synthetic approach it is desired to apply the minimum possible quantity of  $\text{SeO}_2$ . Several approaches were tested in order to establish the minimum quantity of selenium (Table II.1) however the higher yields were obtained with an excess of  $\text{SeO}_2$ . According to the mechanism displayed in scheme II.4-A,B, using a strong oxidant agent could in principle re-oxidize  $\text{Se}(\text{OH})_2$ . Therefore, a catalytic approach was also tested, as suggested in the literature<sup>178–180</sup>, but also without success.

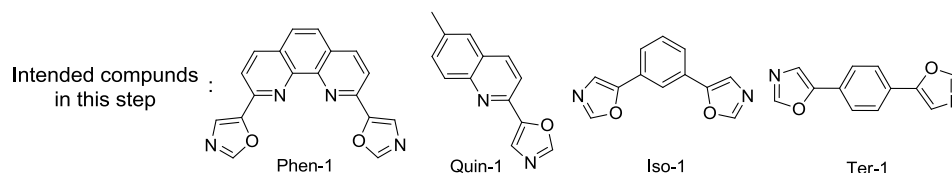


**Scheme II.4** – A) Proposed mechanism for the allylic oxidation mechanism with neocuproine and selenium dioxide. The ene CH group reacts with selenium dioxide in a [4,2] cycloaddition process, forming an intermediate of selenic acid. The intermediate undergoes in a [2,3] allylic rearrangement. Further oxidation reactions convert alcohol to the pretended aldehyde group. The aldehyde form undergoes the same mechanism, resulting the dialdehyde form. B) Catalytic approach with the addition of *tert*-butyl hydroperoxide (TBHP) as a strong oxidizing agent in order to lower the amount of  $\text{SeO}_2$  applied in the reaction and resulting toxic residues obtained.

Although phenanthroline-dicarbaldehyde product was initially obtained with success, its yields were considerably lower when compared to previous reports (50-80%)<sup>181,182</sup>. Apparently, the air oxidizes selenium dioxide, lowering its net purity in the supply batch<sup>183,184</sup>. Hence, the use of fresh batches of  $\text{SeO}_2$  resulted in a boost in final yield for phenanthroline-dicarbaldehyde (61%) and quinoline-carbaldehyde (87%), which now are in agreement with

previous reports that used the same conditions<sup>185,186</sup>. Thereby, a preparative sublimation step of SeO<sub>2</sub> ( $T_{sub} = 340\text{ }^{\circ}\text{C}$ ) is advised before application. Based on preparative TLC and <sup>1</sup>H-NMR spectrum up to 3 phenanthroline by-products were identified but no further efforts were employed for their characterization.

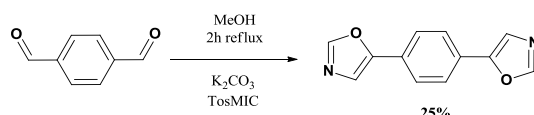
### • Synthesis of oxazoles with TosMIC



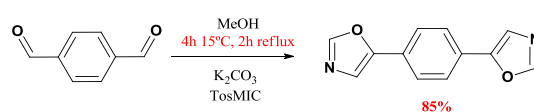
**Scheme II.5** – Illustration of the intended compounds in this topic.

Compounds S<sub>(I-IV)</sub>-0 were submitted to van Leusen conditions in order to form the 4,5-disubstituted oxazole rings. Van Leusen reaction with TosMIC is well described concerning monoaldehyde<sup>187,188</sup> and, up until recently, dialdehyde molecules<sup>154</sup>. Following the established procedure, Quin-1 and Iso-1 were obtained in good-to-high yield (85 and 50%, respectively), however, only trace amounts were obtained for Phen-1 and Ter-1. This led us to adjust the recurrent procedure. The accepted mechanism for TosMIC oxazole synthesis from aldehydes consists of two stages where the first is the formation of a 4-tosyloxazoline intermediate and the second is the 1,2-elimination of tosyl leaving group upon heating<sup>189</sup> (scheme II.3-b). Considering this fact, reactions were first submitted to low temperature conditions (10-20 °C) for a few hours in order to accomplish complete substitution of bis-4-tosyloxazoline (scheme II.6-a,b). Oxazole formation was then achieved by heating the reaction to the reflux temperature. This preliminary step enabled the formation of Phen-1 (50%) and Ter-1 (85%), whereas increased the yield of Iso-1 (65%) and had no effect for Quin-1 yield (85%) (Table II.1).

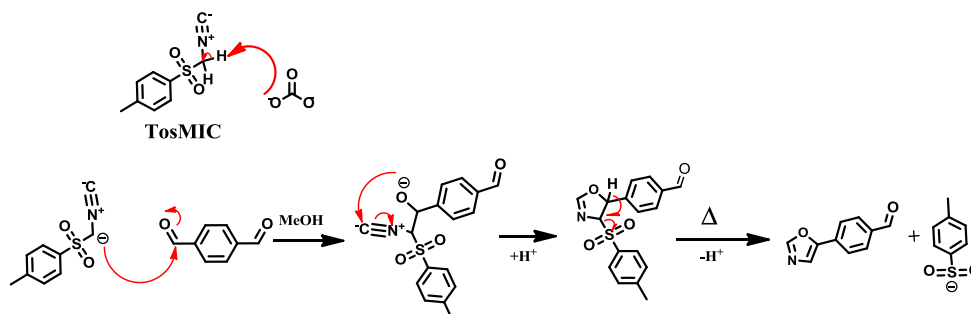
#### A - Recurrent Procedure



#### B - Modified Procedure

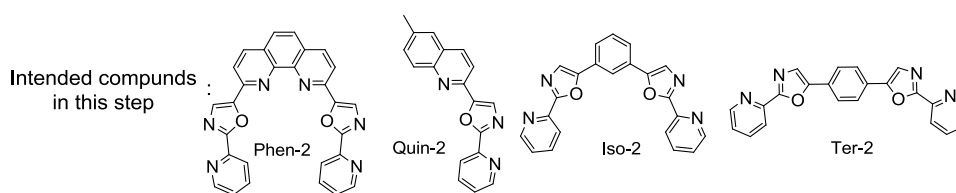


#### C - Reaction Mechanism



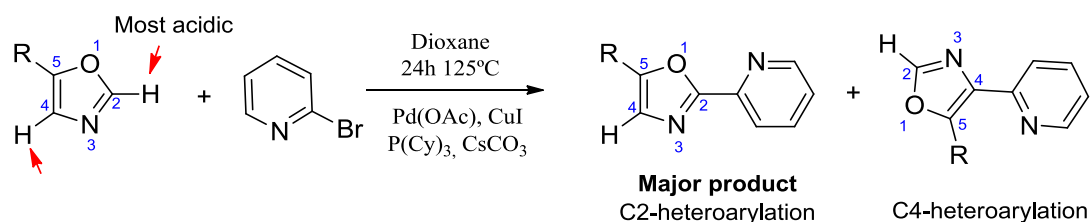
**Scheme II.6** – Oxazole synthesis through van Leusen chemistry employing (A) the traditional procedure and (B) the modified procedure developed in this work. (C) Mechanism for the reaction of oxazole formation through aldehyde reduction with TosMIC.

- Oxazole-pyridine C-C coupling



**Scheme II.7** – Illustration of the intended compounds in this topic.

Coupling reaction of 2-bromopyridine with  $(S_{I-IV})$ -1 provided the oligoheteroaryle form of the ligands  $S_{(I-IV)}$ -2. The procedure was accomplished through Sonogashira reaction conditions, with  $Pd^0$  and  $Cu^I$  as catalysts and tricyclophosphine. Despite the wide use of these cross-coupling reactions, their applications is often based on trial-and-error approach, since its mechanism is not completely elucidated (proposed mechanism displayed in supp. A.3 and A.4). Nonetheless, in principle the C-C coupling should mainly occur on the most activated C-H bond, which belongs to the oxazole group (scheme II.8). Even so, other by-products are expected to form.



**Scheme II.8** – Oxazole activated CH direct coupling with halides. In this reaction conditions, the coupling occurs according to the proton acidity (C2>C4). For mechanism see sup. A.3 and A.4.

Overall, Quin-2 was obtained in admissible yield (45%). However Phen-2, Iso-2 and Ter-2 were obtained with low yields (20, 21 and 35%, respectively). NMR spectra and TLC indicated that at least two more adduct species were formed but were not fully characterized (probably the oxazole C4 and {C2 + C4} heteroarylation product).

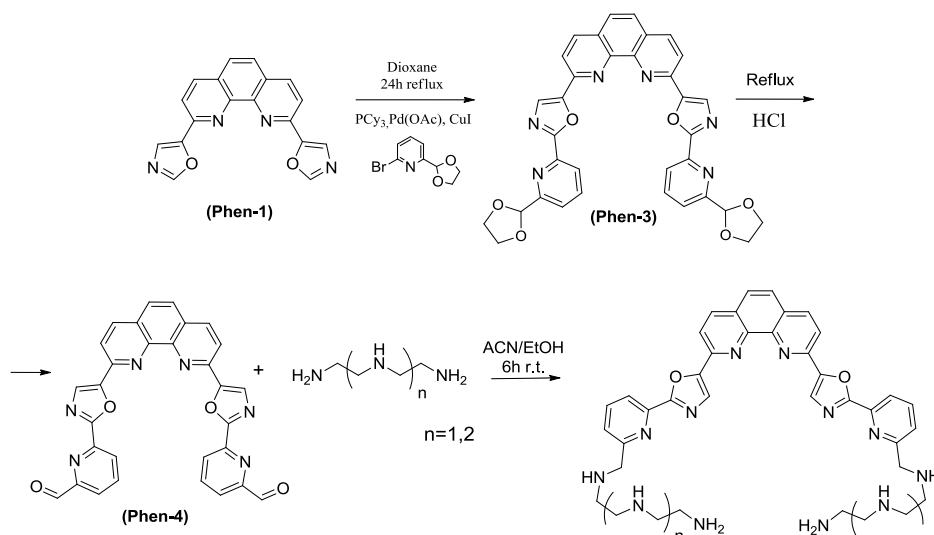
In order to increase the overall yields, a specific study concerning other catalyst systems and regioselective ligands should be performed, as has been specifically demonstrated for oxazole molecules<sup>190</sup>. Furthermore, phenanthroline is well known for its quelant properties; for this reason copper might not be the best metal to be employed as co-catalyst, perhaps a system with Pd and pivalic acid would best addressed in this case<sup>191</sup>. Naturally, an optimization study is inevitably an arduous time-consuming procedure that would fall out of the scope of this dissertation. This optimization step was not performed and the remaining studies proceeded since only a few milligrams of the pure product is required.

**Table II.1** – Summary of the reaction conditions and yields of the synthesized compounds. \*eq=equivalents

Compound	Reactant	Quantity (eq*)	Conditions/Agent added	Yield (%)	
<b>Phen-o</b>	Selenium dioxide (stored batch)	Catalytic (<2%)	Dioxane, 18h, reflux ButOOH (4eq)	Residual	
		2.1	Dioxane, 6h, reflux Water (400mL)	19%	
		4.0	Dioxane, 6h, reflux	38%	
		4.0	Dioxane, 6h, reflux Water (400mL)	51%	
		4.0	Dioxane, 6h, reflux Water (400uL)	61%	
<b>Phen-o</b>	Selenium dioxide (fresh batch)	4.0	Dioxane, 6h, reflux Water (400uL)	61%	
<b>Quin-o</b>		4.0	Dioxane, 6h, reflux Water (500uL)	87%	
<b>Phen-1</b>	TosMIC	2.1	Methanol, 4h, reflux	<5%	
			Methanol, 3h, 20°C +3h, reflux	50%	
Methanol, 4h, reflux			50%		
<b>Iso-1</b>			Methanol, 3h, 20°C +3h, reflux	61%	
			Methanol, 4h, reflux	20%	
<b>Ter-1</b>			Methanol, 3h, 20°C +3h, reflux	85%	
			Methanol, 4h, reflux	85%	
<b>Quin-1</b>			1.1	Methanol, 3h, 20°C +3h, reflux	85%
			<b>Phen-2</b>	4.0	Dioxane, 24h, 130°C
Pd(Ac), PPh, CuI					21%
<b>Ter-2</b>	2-Bromopyridine	35%			
	<b>Quin-2</b>	2.0	Dioxane, 24h, 130°C Pd(Ac), PPh, CuI	45%	

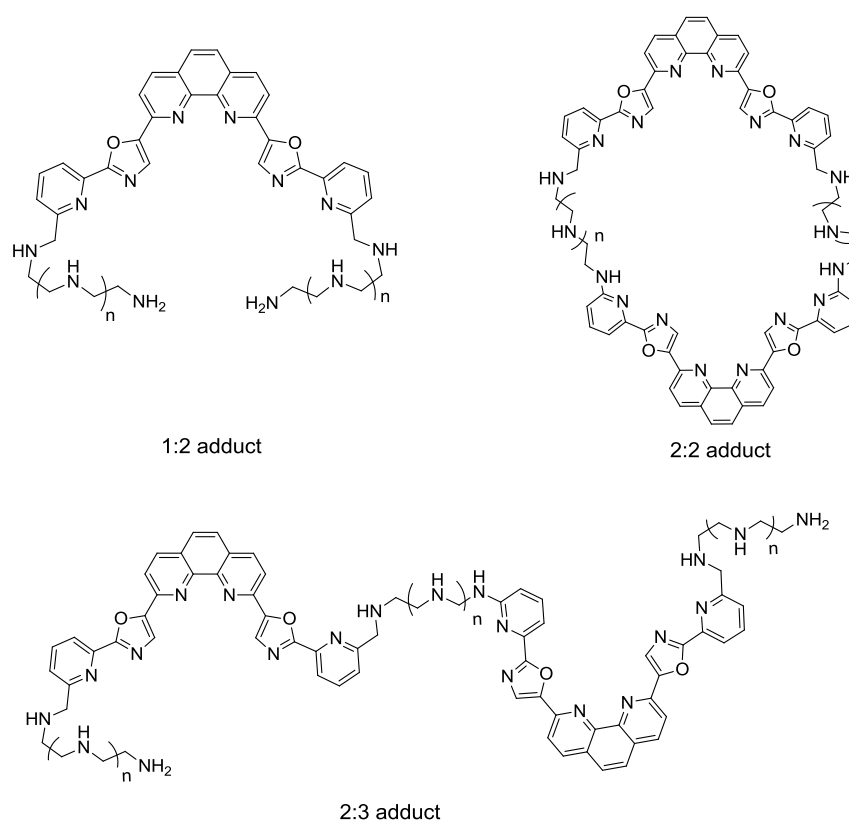
- Improving solubility of Phen-2 with aliphatic amines

The compound Phen-2 is the best candidate to G-quadruplex binding and stabilization. However its low solubility will provide a challenge in further biological assays, similarly to what happened to telomestatin. For this reason we propose further functionalizing of Phen-2 ligand with polyammonium cations that are capable of forming hydrogen bonds. Aliphatic chains with amines were also demonstrated to interact with the G-quadruplex, so these will be employed.



**Scheme II.9** – Synthetic route for coupling aliphatic amines (or other nucleophilic groups) to phen-4.

In order to couple an aliphatic amine group, our strategy consists of a nucleophilic addition of the amine to an aldehyde group. In this approach (scheme II.9) it was employed the same procedure for direct heteroarylation coupling but with 2-bromo-6-dioxolanyl-pyridine to afford Phen-3 bisacetal precursor, which after deprotection in acidic conditions afforded Phen-4. This was reacted with triethylenetetramine (TETA) or diethylenetriamine (DETA) to afford the correspondent amine-substituted compound. However, TLC and NMR spectra revealed a mixture of compounds, most probably the 1:2 and 2:2 adducts (scheme. II.10), even when the reaction was performed with a kinetic control (low temperatures).



**Scheme II.10** – Mixture of products that are probably formed upon condensation reaction of TETA of DETA.

The proceeding G-quadruplex binding studies and the biological assays require high pure compounds; hence other synthetic procedure or purification steps in order to obtain high pure functionaized-Phen-2 must be developed. A procedure like this would reveal as very time consuming and thus out of the scope of this dissertation.

### 3.2 - Spectrophotometric characterization

The synthesized compounds have extensive planar aromatic systems that may exhibit some potential to be applied as fluorescent probes. For this reason, their emission and excitation spectra were recorded (full spectra in appendix A). Table II.2 summarizes the maximums of emission and excitation. Overall, most of the compounds do not display intrinsic fluorescence in the visible-range except Phen-2 and Ter-2 which fluoresce at 425 and 415 nm; still, their application as fluorescent probes seems limited. Nonetheless, unexpected photophysical events may occur upon interaction with DNA, so their potential behavior as fluorescent probes is reserved.

**Table II.2** – Maximum emission and excitation wavelengths for the potential G-quadruplex ligands.\*

<i>Compound</i>	<i>Excitation/Emission maximum (nm)</i>	<i>Compound</i>	<i>Excitation/Emission maximum (nm)</i>
<b>Phen-1</b>	330/388	<b>Ter-1</b>	320/351
<b>Phen-2</b>	350/415	<b>Ter-2</b>	360/425
<b>Iso-1</b>	310/350	<b>Quin-1</b>	340/372
<b>Iso-2</b>	325/390	<b>Quin-2</b>	355/395

\*Spectra recorded in 20%DMSO, 80% phosphate buffer 30mM with 50mM KCl.



## 4 – CONCLUSIONS

Considering the reported phenanthroline and polyoxazoid G4 ligands as models, eight new potential G-quadruplex ligands were designed, synthesized and characterized, affording two phenanthroline (Phen-1 and Phen-2), four phenyl (Iso-2, Iso-2, Ter-1, Ter-2) and two quinoline (Quin-1 and Quin-2) oligoheteroaryle compounds.

The synthesis of these compounds started with a procedure based on van Leusen, which lead to oxazole formation. This reaction was optimized for dialdehydes, increasing their yield up to 10 fold, granting Phen-1 (50%), Iso-1 (61%), Ter-1 (85%) and Quin-1 (85%). A cross coupling reaction through C-H activation with pyridine bromide afforded Phen-2 (20%), Iso-2 (21%), Ter-2 (35%) and Quin (45%). Phen-1 and Phen-2 ligands accounts for Lipinski rule of five, thus these may exhibit therapeutic potential regardless of their low solubility. Other ligands based on Phen-2 motif are being developed in order to improve its solubility.



## Part B

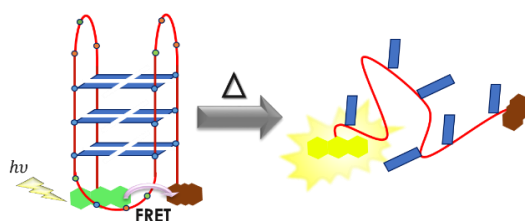
# Oligoheteroaryle G-quadruplex Ligands: screening and interaction

### 1 – INTRODUCTION

As presented previously several potential G-quadruplex ligands were synthesized. The obvious next step consists on screening which ligands display affinity for G-quadruplex structures. Positive hits will then undergo further studies for assessing duplex/quadruplex specificity. The goal of the following work also aims at identifying potential molecular determinants for recognition and interaction of the G-quadruplex, hence the characterization of the G4-ligand interaction will be performed as detailed as possible. Identification of these molecular determinants could provide important hints for future design of new ligands.

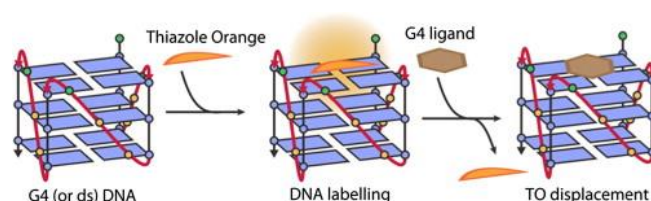
Some of the techniques explored in this work, especially the screening assays, were specially developed for G-quadruplex-ligand interactions, therefore a brief presentation about these techniques should be appointed.

The screening procedure starts with a FRET-melting assay, which is perhaps the most typically employed high-throughput technique<sup>192,53,193</sup>. Particularly to G-quadruplexes this approach uses truncated fluorescent-labeled oligonucleotides, which are non-fluorescent when the quadruplex is folded; disruption of the quadruplex structure results on fluorescence emission, thus enabling melting temperature measurement (Figure II.4). Considering this mechanism, if a compound binds and stabilizes the G-quadruplex the melting temperature should increase. Since this assay may be performed on conventional multi-well RT-PCR apparatus, this technique allows a rapid and easy screening of a large pool of different compounds with minimal expense of DNA.



**Figure II.4** – Schematic representation of FRET-based melting temperature experiments. Förster resonance energy transfer occurs between the fluorophore (e.g. fluorescein or “FAM”) and the quencher (e.g. tetramethylrhodamine or “Tamra”), due to spatial proximity between the two. As the structure is denatured with temperature (melting) the two species are moved apart and fluorescence is restored.

G4FID is another typical employed high-throughput screening assay based on competition between the ligand and an intercalating fluorescent probe that lights up upon binding to DNA<sup>194–197</sup> (Figure II.5). If the ligand competes with this probe, the net fluorescence signal decreases. Typically the probe employed is Thiazole Orange (TO) whose quantum yield increases by a factor up to 1000 after binding to DNA<sup>198</sup>. The fluorescence quenching is hence related to ligand G4 affinity. Just as FRET-melting assay this technique is rapid and easy, the big advantage is that it doesn't require fluorescent labeled oligonucleotides. The major downside is that the results are not directly correlated to ligand-G4 interaction, as it requires displacement of TO from the G-quadruplex; if the ligand behaves as a groove binder, there might occur low displacement even if a strong interaction is taking place.



**Figure II.5** – Schematic representation of the G4-FID assay. Thiazole Orange intercalates with the G4, greatly enhancing fluorescence signal. Application of a competitor (G4 ligand) displaces TO from DNA diminishing the signal. Scheme from reference<sup>199</sup>.

After selecting the quadruplex interacting ligands, their interaction with the quadruplex can be characterized by circular dichroism (CD) studies and nuclear magnetic resonance (NMR) techniques could further elucidate a molecular model.

CD spectra are extremely sensitive to the nucleic acid topology and conformation. This sensibility arises from the nucleotide's glycosylic angle, thus a CD spectra may distinguish a G-quadruplex from a duplex and a single-strand nucleic acid<sup>200</sup>; it can even distinguish between G4 topologies due to the G-tetrad orientation combination (heads-to-tails / tails-to-tails, etc.)<sup>201–203</sup>. Thus, conformational modifications and the structural stability induced by ligands may be studied by monitoring the CD profile<sup>204</sup>.

NMR spectroscopy techniques provide very powerful tools for assessing interactions and provide a dynamic and structural insight at an atomic level. The complete structural determination by NMR would be the climax for describing the interaction, but this is a challenging long and rigorous process that requires several weeks (months) of commitment. Nevertheless, NOESY correlation signals between the nucleus from the ligand and G-quadruplex implies that these are in spatial proximity (up to 5–6 Å); this data allows to locate the ligand on the G-quadruplex and might be used as spatial constraints to apply on molecular docking and modelling calculations.

As for the G-quadruplex sequences applied in the following studies, two representative G4 sequences were employed: a human telomeric sequence (entry: 22AG) and a mutated fragment from the promotor sequence of the oncogene *c-myc* (entry: *c-myc*). Both sequences consist on a single stranded of 22bp DNA sequence with biological relevance. The telomeric sequence has a very high structural polymorphism associated, so it is interesting to observe what kind of transformations ligands may induce. The *c-myc* sequence adopts a very rigid and well defined G-quadruplex, so this sequence may act as a structural model. Both sequences are extensively studied and described in the literature<sup>54,97,205,89,87,108,206</sup>.

## 2 – EXPERIMENTAL

### 2.1 – Oligonucleotides and sample preparation

The oligonucleotides used in this work are described in table II.3.

**Table II.3** – Oligonucleic acid sequences employed for biochemical assays and preferred G-quadruplex topologies

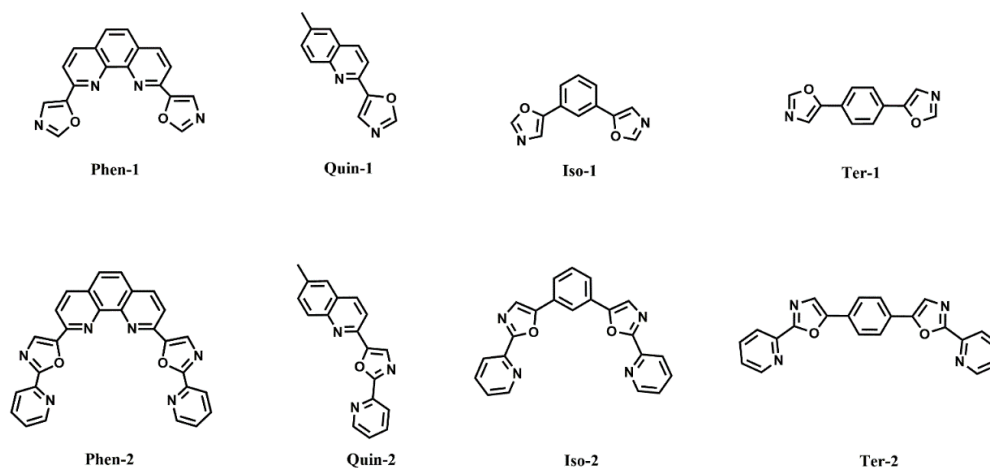
<i>Oligo Entry name</i>	<i>Sequence</i>	<i>Topology</i>
22AG	d [AG <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	Antiparallel
c-myc	d [TGA(G <sub>3</sub> TG <sub>3</sub> TA) <sub>2</sub> A]	parallel
ds42	d [(AG) <sub>3</sub> (T <sub>2</sub> AG) <sub>9</sub> ] + d [(TC) <sub>3</sub> (A <sub>2</sub> TC) <sub>9</sub> ]	Intermolecular duplex
F21T	FAM - [d G <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ] - Tamra	Antiparallel
FmycT	FAM- d [T <sub>2</sub> GA(G <sub>3</sub> TG <sub>3</sub> TA) <sub>2</sub> A] -Tamra	Parallel
FdxT	FAM- d [(TA) <sub>2</sub> GC(TA) <sub>2</sub> -hexaethyleneglycol-(TA) <sub>2</sub> GC(TA) <sub>2</sub> ] -Tamra	Intramolecular duplex

Oligonucleotides were dissolved in 30mM phosphate buffer, the salt content is specified along the manuscript. Stock solutions of approximately 0.5 mM strand concentration were prepared with Milli-Q ultrapure water and were stored at -20 °C. Oligonucleotides were purchased from STAB VIDA laboratories.

Labeled nucleotides were provided by Dr. Jean-Luis Mergny Laboratory (INSERM) for FRET-melting assays, which correspond to F21T, FmycT and FdxT. The 5' and 3' ends of these oligonucleotides were modified with fluorescein (FAM) and rhodamine (TAMRA), respectively.

In order to induce the G-quadruplex motif, samples were heated to 95°C and gradually cooled-down to room temperature over a period of 60 min, as described for annealing intramolecular G-quadruplexes<sup>199,207,208</sup>.

Purified ligand stocks of 5 mM were prepared in pure DMSO and diluted directly on the DNA working buffer. Ligands employed in this study are exhibited on fig. II.6.



**Figure II.6** – Chemical structure of the ligands analyzed in this study.

## 2.2 – Data processing of melting experiments

- **$T_m$  determination**

All temperature-monitored experimental data was normalized and plotted against correspondent temperature, defining a melting curve. Data are normalized between [0,1] according to equation

$$Norm(x) = \frac{x - a_{min}}{b_{max} - a_{min}},$$

where  $a_{min}$  and  $b_{max}$  are the mean minimum and maximum values, respectively.

The melting temperature ( $T_m$ ) is determined from melting curves and it is defined as  $T_{1/2}$  of the folded-unfolded transition<sup>209</sup>. Concerning the applied intermolecular G-quadruplex equilibria, no hysteresis phenomena was reported, thus this approximation is valid.

Melting curves were fitted with a Boltzmann function,

$$y = \frac{A_1 - A_2}{1 + e^{(x-x_0)/dx}} + A_2$$

with Origin 9.0 software.

- **Thermodynamic parameters**

In a properly normalized simple two-state equilibria, the fraction folded ( $\alpha$ ) is calculated at each temperature giving the corresponding equilibrium constant  $K_{eq} = \frac{\alpha}{1-\alpha}$ .

From this data may result,

$$K_T = K_0 e^{-\frac{\Delta H}{RT}}$$

where  $\Delta H$  may be obtained using an Arrhenius plot from the slope of the plot  $\ln K_T$  vs  $1/T$ , given that

$$\ln K_T = \ln K_0 - \frac{\Delta H}{RT}$$

where  $R$  is the gas constant. This approach is valid for a unimolecular two-state process where  $\Delta H$  is not dependent of temperature<sup>209</sup>.

At the melting temperature  $T_{1/2}$  (i.e.  $K_{eq} = 1$ )  $\Delta G = 0$ , therefore from Gibbs equation

$$\Delta G = \Delta H - T\Delta S$$

results  $\Delta S = \frac{\Delta H}{T_{1/2}}$ .

## 2.3 – Fluorescence Resonance Energy Transfer (FRET) experiments

A Stratagene Mx3005P instrument was used to carry out the FRET melting experiments in 96-well plates as previously described<sup>53</sup>. The excitation and detection wavelengths were 492 nm and 516 nm, respectively. After an initial incubation at 25 °C for 5 min, the temperature was increased by 1 °C every minute until 95 °C was reached. The experiments were performed with samples containing 0.2 mM oligonucleotide, 10 mM lithium cacodylate (pH 7.2), and indicated concentrations of KCl, NaCl and LiCl (for a total salt concentration of 100 mM). For the measurements in the presence of a ligand, the concentration of ligand was 5 and 10  $\mu$ M. Each experimental condition was tested in duplicate on at least three separate plates.

FRET experiments were performed in collaboration with Dr. Jean-Louis Mergny at Laboratoire ARNA at INSERM, IECB, Pessac (France).

## 2.4 – G4 Fluorescent Intercalator Displacement assay

Thiazole Orange (TO) was selected as probe. Its intrinsic fluorescence in solution is irrelevant, however it behaves as a switch-on fluorescent probe upon binding to tertiary structures of DNA; also it is a well-described DNA intercalator. More properties about TO-DNA interaction are presented in cited reference<sup>198</sup>. Stock solutions of Thiazole Orange (CAS# 107091-89-4, Aldrich) were prepared in DMSO at 1 mM concentration, and kept in 4°C for a maximum period of 3 weeks.

High-throughput G4-FID assay was performed in a Bio-Rad CFX RT-PCR fluorimeter. The software was properly calibrated for Thiazole Orange as was indicated by the manufacturer. A temperature of 20 °C was kept constant in thermostatic plate holders. Experiments were performed in conventional hard-shell 96-well PCR plate, in 30 mM phosphate buffer pH 7.2 with 50 mM KCl or 100 mM NaCl depending on the experiments, in a total volume of 25 µL. The plate layout is exemplified in fig. II.7.

		1	2	3	4	5	6	7	8	9	10	11	12
<b>A</b>	Entry	Buffer	Buffer		22AG	22AG	22AG						
	Proportion Description	Blank	Blank		2eq TO FA0	2eq TO FA0	2eq TO FA0						
<b>B</b>	Entry	TO	TO										
	Proportion Description	Buffer+TO	Buffer+TO										
<b>C</b>	Entry	Phen1	Phen1	22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG
	Proportion Description	Fb control	Fb control	0,2eq Phen1 2 eq TO	0,4eq Phen1 2 eq TO	0,6eq Phen1 2 eq TO	0,8eq Phen1 2 eq TO	1eq Phen1 2 eq TO	2eq Phen1 2 eq TO	4eq Phen1 2 eq TO	6eq Phen1 2 eq TO	8eq Phen1 2 eq TO	10eq Phen1 2 eq TO
<b>D</b>	Entry	Phen2	Phen2	22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG
	Proportion Description	Fb control	Fb control	0,2eq Phen2 2 eq TO	0,4eq Phen2 2 eq TO	0,6eq Phen2 2 eq TO	0,8eq Phen2 2 eq TO	1eq Phen2 2 eq TO	2eq Phen2 2 eq TO	4eq Phen2 2 eq TO	6eq Phen2 2 eq TO	8eq Phen2 2 eq TO	10eq Phen2 2 eq TO
<b>E</b>	Entry			22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG
	Proportion Description			0,2eq Phen1 2 eq TO	0,4eq Phen1 2 eq TO	0,6eq Phen1 2 eq TO	0,8eq Phen1 2 eq TO	1eq Phen1 2 eq TO	2eq Phen1 2 eq TO	4eq Phen1 2 eq TO	6eq Phen1 2 eq TO	8eq Phen1 2 eq TO	10eq Phen1 2 eq TO
<b>F</b>	Entry			22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG
	Proportion Description			0,2eq Phen2 2 eq TO	0,4eq Phen2 2 eq TO	0,6eq Phen2 2 eq TO	0,8eq Phen2 2 eq TO	1eq Phen2 2 eq TO	2eq Phen2 2 eq TO	4eq Phen2 2 eq TO	6eq Phen2 2 eq TO	8eq Phen2 2 eq TO	10eq Phen2 2 eq TO
<b>G</b>	Entry			22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG
	Proportion Description			0,2eq Phen1 2 eq TO	0,4eq Phen1 2 eq TO	0,6eq Phen1 2 eq TO	0,8eq Phen1 2 eq TO	1eq Phen1 2 eq TO	2eq Phen1 2 eq TO	4eq Phen1 2 eq TO	6eq Phen1 2 eq TO	8eq Phen1 2 eq TO	10eq Phen1 2 eq TO
<b>H</b>	Entry			22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG	22AG
	Proportion Description			0,2eq Phen2 2 eq TO	0,4eq Phen2 2 eq TO	0,6eq Phen2 2 eq TO	0,8eq Phen2 2 eq TO	1eq Phen2 2 eq TO	2eq Phen2 2 eq TO	4eq Phen2 2 eq TO	6eq Phen2 2 eq TO	8eq Phen2 2 eq TO	10eq Phen2 2 eq TO

**Figure II.7** – Illustration of the plate layout for a G4FID experiment. In this particular example telomeric 22AG sequence and Phen-1 and Phen-2 were applied.

The procedure for G4-FID assay consists in applying each well the indicated reagents in fig. II.7. The ratio 2:1 TO:DNA was the optimum ratio for the maximum fluorescence value (see sup. B.1). Solutions were properly mixed and equilibrated for a minimum of 5 min period after the fluorescence (excitation of 490, emission 500 nm) is recorded.

The percentage of displacement is calculated by

$$TO \text{ Displacement (\%)} = 100 - ((FA/FA_0) \times 100),$$

where  $FA_0$  is the fluorescence area (or maximum value) of DNA-TO complex without ligand and  $FA$  is the fluorescence area of DNA-TO complex in the presence of the ligand. FID curves were obtained by plotting TO displacement percentage versus concentration of ligand used.

## 2.5 – Differential Scanning Fluorimetry

Data was recorded on the same solutions for the G4FID assay on the same RT-PCR apparatus. Temperature was ranged from 20-95 °C, data pitch of 0.5 °C with 500ms stabilization.

The transition temperature correspond to TO disassociation, therefore it does not directly correspond to the DNA exact melting temperature. Furthermore TO slightly increases the  $T_m$ , as it binds DNA<sup>198</sup>. Study on the influence of [TO] on melting temperature as displayed on sup. B.1.

## 2.6 – Circular Dichroism spectroscopy

Circular dichroism spectra were acquired on a Jasco J-815 spectropolarizer equipped with a Peltier CDF-426S temperature controller. Measurements were performed in a 1mm light-path quartz cell. CD spectra were recorded with 1nm bandwidth at a scanning speed of 50nm/min and averaged over 3 recordings per scan, with a 1s DIT response. Buffer spectra was subtracted to all spectra. CD melting studies were performed by monitoring the correspondent wavelength ranging temperatures from 25-110°C with  $1,0\pm 0,1$  nm data pitch, ramp rate 4°C/min and with a 4 seconds temperature stabilization. DNA samples were recorded at a concentration [5-10  $\mu$ M] depending on the sequence in phosphate buffer. Salt content is specified along the manuscript.

The thermal denaturation experiments were monitored at the wavelengths of maximum CD intensity.

The stock solutions of Phen-1 and Phen-2 were prepared in DMSO (5 mM) and diluted in the required buffer prior to use. The study of the effects of addition of DMSO on the G4 topology is presented in the appendix sup. B.5, where no major alterations on the CD profile were verified.

## 2.7 – Fluorimetry

Fluorescence spectra were obtained with a Perkin Elmer Precisely Luminiscence Spectrometer model LS 45, at 25°C. Data were set with a 10nm slit width and a scanning speed of 1000 nm/min

Measurements were performed in 30mM phosphate buffer with consecutive additions of DNA to a ligand sample, with a 3 min equilibration time between measurements. Total ligand concentrations was fixed at 0.5  $\mu$ M and DNA was gradually added from a concentrated stock solution (500 $\mu$ M). Excitation and emission wavelength values correspond to the ligand's maximum, signal contribution from DNA is insignificant.

$F_0$  and  $F$  correspond to the emission band area of free ligand and ligand + DNA respectively. The  $K_a$  values were determined from the fitted titration curves.



- **Affinity constant and binding sites**

In principle, fluorescence signal (or any other signal measurement) may be considered as the sum of the weighted contributions from free and bound ligand:

$$F = (1 - \alpha_b)F_0 + \alpha_b F_b$$

where  $F$  is the value of the experimental variable at each titrant concentration,  $F_0$  and  $F_b$  are the values of the variable at the initial and final states of titration, and  $\alpha_b$  is the mole fraction of ligand in bound form. Assuming 1:1 stoichiometry for the interaction and no intermediate states, it can be shown that

$$[L_0] - \alpha_b^2 - \left( [L_0] + [G4] + \frac{1}{K_a} \right) \alpha_b + [G4] = 0$$

where  $K_a$  is the binding constant,  $[L_0]$  is the total ligand concentration, and  $[G4]$  is the added DNA concentration. From combination of previous equation and considering

$$\alpha_b = \frac{F - F_0}{F_b - F_0}$$

it can be shown that:

$$\Delta F = \left( \frac{\Delta F_{max}}{[L_0]} \right) \frac{\left\{ \left( [L_0] + n[G4] + \frac{1}{K_a} \right) - \sqrt{\left( [L_0] + n[G4] + \frac{1}{K_a} \right)^2 - 4([L_0]n[G4])} \right\}}{2}$$

where  $\Delta F = F - F_0$  and  $\Delta F_{max} = F_{min} - F_0$ .  $F_{min}$  is the minimum value obtained for through the quenching process and  $n$  is the binding sites. Worth mention that this is applied for a quenching effect; an enhancement effect  $\Delta F_{max} = F_{max} - F_0$ . For a deeper data processing, see cited references<sup>51,210,211</sup>.

In order to quickly calculate the binding sites, a simple Stern-Volmer linearization yield the estimated binding sites through the equation,

$$\log\left(\frac{F_0 - F}{F}\right) = \log(K_{SV}) + n \log([G4])$$

where  $K_{SV} = K_{eq} + K_{Quen}$  and  $K_{Quen}$  corresponds to dynamic quenching parameter, if applicable<sup>212</sup>.

## 2.8 – Nuclear Magnetic Resonance

NMR spectra were recorded on a Bruker Avance III (600MHz), equipped with a cryoprobe, at temperature of 25°C. 1D proton spectra were recorded using an excitation sculpting scheme with pulse-field gradients (Bruker pulse program ZGESGP) for water signal suppression.

DNA samples were prepared at a concentration of 400µM strand concentration in (10:90 D<sub>2</sub>O/H<sub>2</sub>O) 30mM phosphate buffer solution with 100mM KCl, pH 7.2. Phen-2 ligand was prepared at a 5 mM DMSO-d<sub>6</sub> solution and added directly to DNA samples.

### 3 – RESULTS AND DISCUSSION

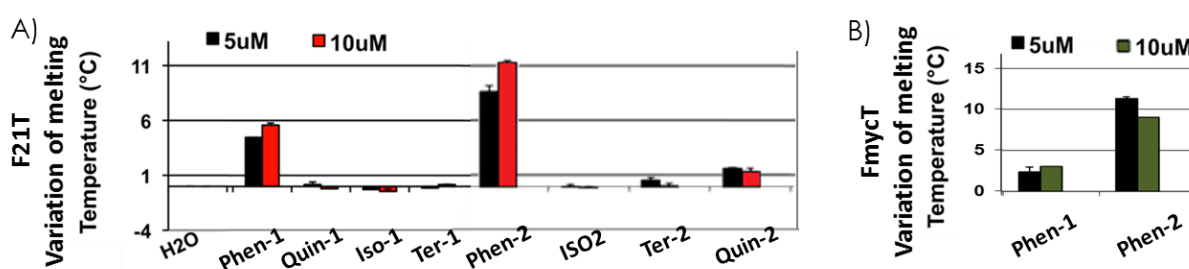
#### 3.1 – Förster Resonance Energy Transfer (FRET) experiments

The ability of the new synthesized compounds (displayed in fig. II.6) for interacting with the G-quadruplex was assessed through high-throughput FRET-melting assays using a labeled telomeric G-quadruplex sequence (*FAM-G<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>-Tamra*). In the applied conditions, this sequence demonstrated to adopt antiparallel quadruplex structures, particularly a mixture of (3+1) hybrid-1 and (3+1) hybrid-2 topologies<sup>193</sup>.

FRET-melting results are displayed on fig. II.8 as the variation of melting temperature in the presence and absence of ligand. Fig. II.8-A clearly indicates positive effects for Phen-1 and Phen-2 ligands, with approximately  $\Delta T_{1/2}$  of 6 °C and 11 °C, respectively. Other ligands failed to induce any considerable thermal stabilization. This underlines the importance of the phenanthroline motif as it provides a wider gap between oxazole and pyridine substituents, improving stacking interaction; Phen-1 and Phen-2 have essentially the same contour as Iso-1 and Iso-2 however the referred gap is tighter in the latter, which might be the reason for absence of any interaction in these ligands. Also, comparison between Phen-1, Phen-2 and Quin-2 emphasizes the importance for symmetric shapes, if our hypothesis for G-tetrad stacking is correct.

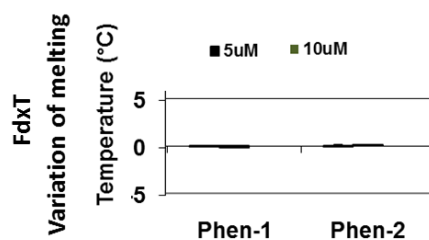
The stabilizing ability of Phen-1 and Phen-2 was also investigated for other topologies (fig. II.8-B). Thermal stabilization induced by Phen-1 was much smaller for FmycT (parallel topology), thus suggesting smaller affinity. On the other hand, with Phen-2 the  $\Delta T_{1/2}$  remained approximately with the same magnitude; curiously the  $\Delta T_{1/2}$  decreased with increasing concentration, which in principle should not be related to solubility given that the same did not happen with F21T.

Iso, Quin and Ter compounds will have to be analyzed against other topologies in order to better rationalize these results.



**Figure II.8** – Plot of the FRET-melting results on G-quadruplex structures in 10 mM Li cacodylate buffer with 10 mM KCl + 90 mM LiCl, pH 7.2. A) Stabilization of F21T (0.2 μM) with corresponding ligands at 5.0 μM and 10.0 μM. (B) Stabilization of FmycT (0.2 μM) with Phen-1 and Phen-2 ligands at 5.0 μM and 10.0 μM.

Phen-1 and Phen-2 ligands were selected for melting studies with a duplex sequence (fig. II.9) and  $\Delta T_{1/2} = 0$  for both ligands. Thus, these ligands specifically stabilize G-quadruplex structures.



**Figure II.9** – Plot of the FRET-melting results on duplex FdxT (0.2  $\mu\text{M}$ ) with Phen-1 and Phen-2 ligands at 5.0  $\mu\text{M}$  and 10.0  $\mu\text{M}$ , in 10 mM Li cacodylate buffer with 10 mM KCl + 90 mM LiCl, pH 7.2,

Overall, Phen-2 revealed to be the most effective binder. It appears that the chemical modification of Phen-1 by coupling with 2-bromo-6-(1,3-dioxolan-2-yl)pyridine to afford Phen-2, leads to an increase in G-quadruplex stabilization, and the shape phenanthroline-bisoxazole motif appears to be important for selective binding of G-quadruplexes. Excepting Phen-1 and Phen-2, all other ligands were discarded from further biochemical assays.

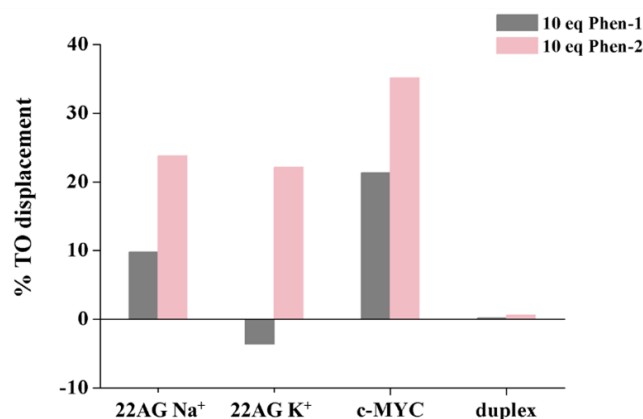
### 3.2 – G4 Fluorescent Intercalator Displacement assay

The capability of Phen-1 and Phen-2 ligands for selectively binding quadruplexes was also investigated with G4-FID assay. This assay does not require labeled oligonucleotides but consists on competition between the ligand and a fluorescent probe for binding the nucleic acid. FRET and G4FID are considered complementary screening techniques.

The assay was also performed with two quadruplex sequences, c-myc and 22AG, and a duplex sequence. Telomeric sequence 22AG was studied both in sodium and potassium phosphate buffer, once it adopts a mixture of (3+1) hybrid structures in potassium and an antiparallel structure in sodium medium.

The intrinsic fluorescence of Phen-1 and Phen-2 was previously evaluated in absence of DNA (table II.2 and sup. A.20) and its excitation and emission maxima do not overlap to TO's.

Surprisingly, Phen-1 and Phen-2 did not induce a decrease in TO fluorescence intensity; in fact it did quite the opposite, it enhanced the fluorescence with 22AG. In the presence of a large excess (10 equivalents) of Phen-1 and Phen-2, only partial displacement of TO was verified and the standard 50% threshold was not reached (fig. II.10). Noteworthy, no TO-displacement was verified for the duplex sequence, reinforcing the conclusion concerning the specificity of the ligands.



**Figure II.10** – Diagrammatic bar representation of G4-FID for Phen-1 and Phen-2 performed with 22AG in Na<sup>+</sup> (100 mM) and K<sup>+</sup> (50 mM), c-myc in K<sup>+</sup> (50 mM) and a duplex in K<sup>+</sup> (50 mM).

This partial displacement is somewhat surprising given the relatively high stabilizations found for Phen-1 and Phen-2 in the FRET assay.

In order to the TO displacement occur the ligand must compete with the TO probe. FRET results already indicated that interaction should indeed occur, therefore this suggest that the ligand does not bind at the same binding-site of TO. It can be also speculated about TO-displacing ability of the ligands that seems to have direct correlation with the number of positive charges of the ligands; the highly cationic molecules are the most efficient TO-displacers<sup>197</sup> and Phen-1 and Phen-2 at pH 7.2 are not charged. But still, a minimum of 50% displacement should have been reached, as was reported for similar polyoxazoids<sup>154</sup>.

Therefore it should be concluded that the response observed for Phen-1 and Phen-2 with 22AG and c-myc might result from interaction with the loops or grooves or the ligands bind secondary sites before (or instead of) displacing the probe.

### 3.3 – Differential scanning fluorimetry

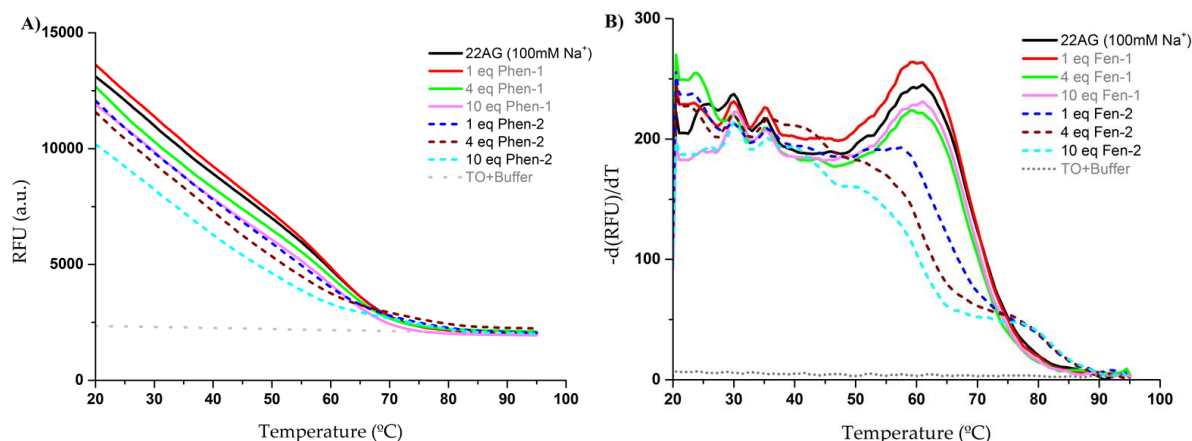
Since our ligands cannot be evaluated by G4-FID assay, we tried to explore the TO light-up fluorescence from another perspective in order to access the hypothesis of groove-binding ligands.

Regarding the TO fluorescence signal of a typical TO-G4 complex, when the TO's dissociation temperature is reached (such as the G4 melting temperature) the fluorescence signal is rapidly quenched. We hypothesized that if the ligand interacts and stabilizes the G-quadruplex then the TO's (fluorescence intensity vs temperature) profile should change. This is a similar approach employed in typical thermal shift assay, where the denaturation of the target biomolecule causes the probe to light-up<sup>213</sup>; in this study we are using the opposite mechanism.

The fluorescence of TO + 22AG decreases linearly with increasing temperature (due to decrease of fluorescence quantum yield) until it declines rapidly after reaching 60°C (fig. II.11, black line). This value matches the melting temperature of 22AG measured previously in 100 mM Na<sup>+</sup> by CD melting. Worth mention that TO slightly increases DNA  $T_{1/2}$  (sup. B.1).

Addition of Phen-1 did not significantly affect the fluorescence intensity nor the TO fluorescence profile. However, addition of Phen-2 clearly altered this profile in a concentration-dependent manner (fig. II.11). Furthermore it increased the transition temperature in about 15°C, which is in the same range of the stabilization temperature obtained by FRET melting studies. This result indicates that in fact there is an interaction of Phen-2 with DNA even at low concentrations, although the % TO displacement was very low. This implies that Phen-2 does not compete with TO and should interact by groove-binding or an unusual stacking mechanism instead.

This study was not performed with c-myc due to its high melting temperatures (>80°C); at this temperature the fluorescence quantum yield of TO is too low. Working with lithium instead of potassium phosphate should decrease the melting temperature to measurable values and this problem would be solved, so this approach will be considered in future experiments.



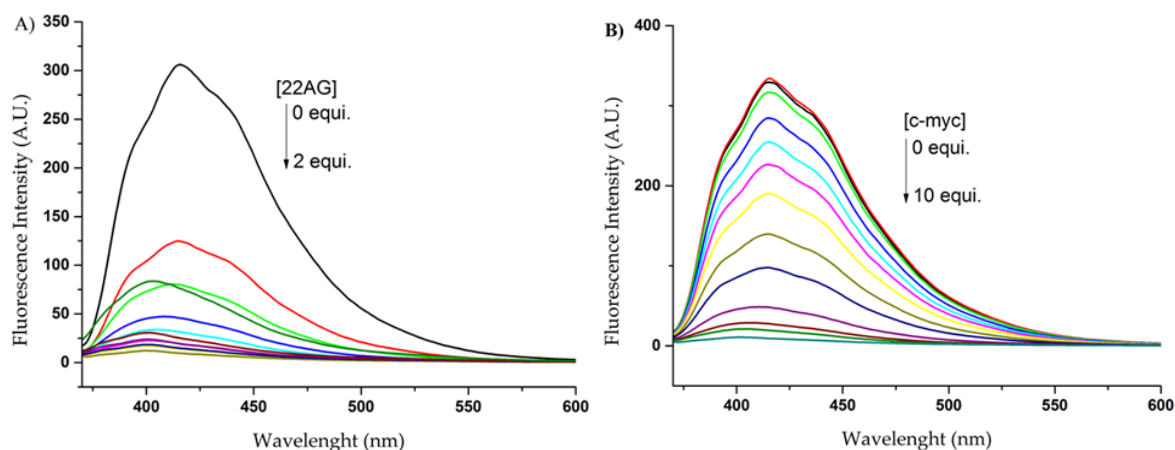
**Figure II.11** – (A) Thiazole Orange fluorescence signal ( $\lambda_{exc} = 490$  nm,  $\lambda_{emi} = 501$  nm) with increasing temperature in presence of 22AG (0.25  $\mu$ M), 22AG+Phen-1 (0.25-2.5  $\mu$ M) and 22AG+Phen-2 (0.25-2.5  $\mu$ M) in 30 mM phosphate with 100 mM K<sup>+</sup> buffer condition. (B) Plot of the negative of the first derivative of the curves presented in (A).

### 3.4 – Fluorimetry

The fluorescent properties of Phen-1 and Phen-2 may be affected by the presence of G4. As shown previously, Phen-1 and Phen-2 exhibit an emission band centered at 388 and 415 nm respectively (part A table II.2, spectra on sup. A.20).

Addition of 22AG to Phen-2 in K<sup>+</sup> led to a strong (>90%) and concentration-dependent decrease in fluorescence intensity (fig. II.12-A). A much milder effect was observed for Phen-1. The fitted curves for Phen-2 show that 0.5 equivalents of 22AG were enough to quench 80% of the signal (fig. II.13).

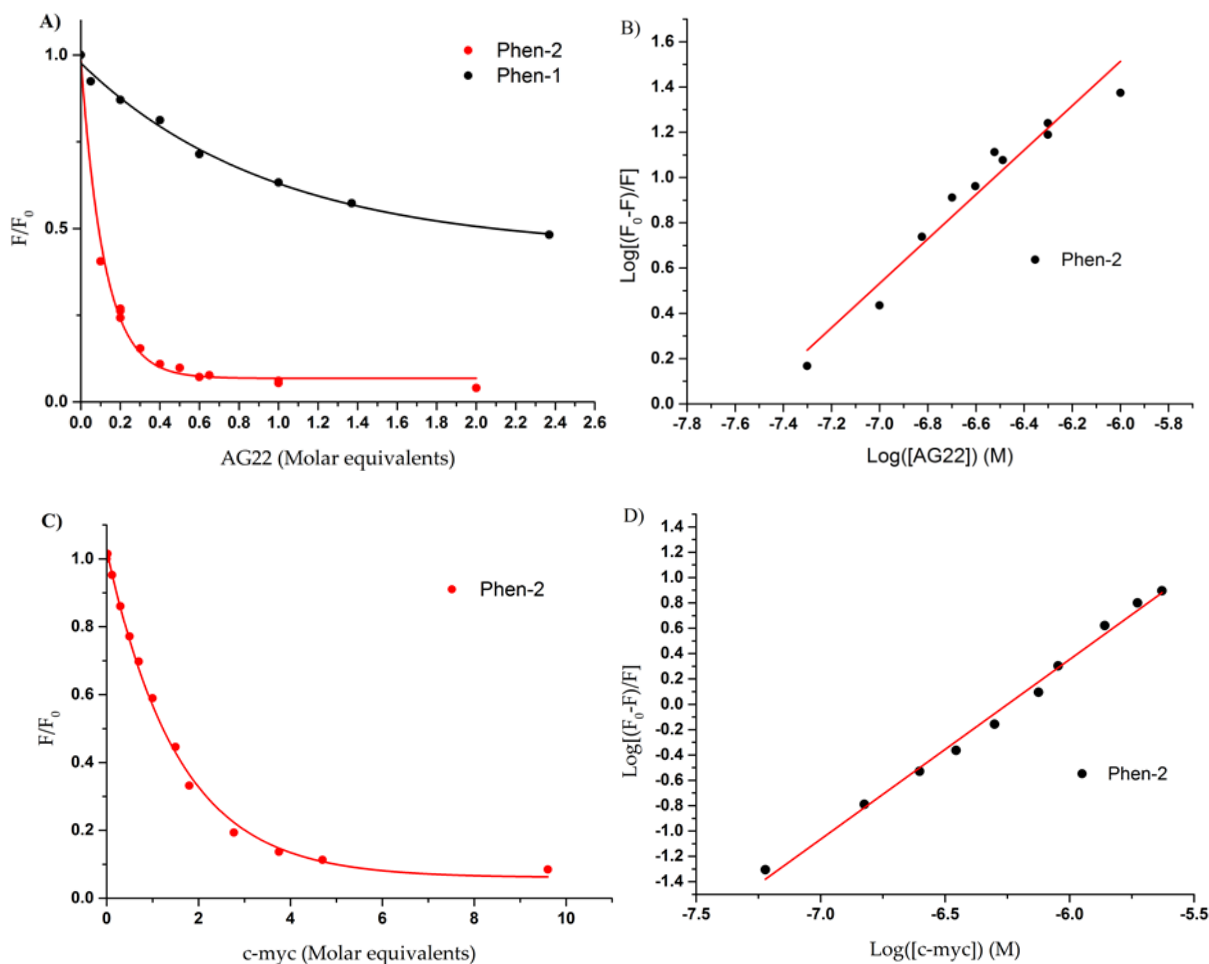
Addition of c-myc to Phen-2 in K<sup>+</sup> resulted in the same strong quenching effect but in a smoother manner (fig. II.12-B). The fact that this effect extends until addition of several equivalents (>2 eq) may indicate that hydrophobic quenching interactions may be not as strong in c-myc as with telomeric quadruplex, hence interaction with the c-myc quadruplex probably is not solely related to G-tetrad stacking.



**Figure II.12** – Representation of fluorescent titrations experiment of Phen-2 ( $\lambda_{exc}=350$  nm) with (A) 22AG and (B) c-myc G-quadruplex sequences, in 30 mM phosphate with 50 mM K<sup>+</sup> buffer. Phen-1 titration in sup. B.3-a,b.

Figure II.13 represents the binding curves for the effect represented above. Fitting of these curves provided affinity constants for Phen-2-quadruplex binding. Admitting a 1:1 stoichiometry, the  $K_a = 9.56 \times 10^9 M^{-1}$  for 22AG quadruplex and  $K_a = 3.55 \times 10^6 M^{-1}$  for c-myc. The affinity of Phen-2 to 22AG is an outstanding value compared to other ligands and even with other phenanthroline compounds, such as the hyped ligand PhenDC3<sup>214</sup> that exhibits a  $K_a = 9.8 \times 10^8 M^{-1}$  (although the telomeric quadruplex sequence was not the same). Stern-Volmer plot gave an approximate value of  $n=1.02$  for binding stoichiometry. If this is correct, Phen-2 most probably interacts with one of the terminal G-tetrads. Obviously, calorimetric techniques such as Isothermal Titration Calorimetry should confirm this binding stoichiometry.

As for c-myc, the binding constant is very similar to what is found on porphyrinic compounds ( $10^6 M^{-1}$ )<sup>211</sup>, which are described to interact both through groove binding and G-tetrad stacking. Calculated number of binding sites for c-myc indicates that the fitting may not be accurate; since  $n=1.63$  this suggests that maybe there are more than 1 binding sites, as already suspected, hence the fitting equation should account a 1:2 binding stoichiometry. Nonetheless, for this particular case other calorimetric techniques should be employed.



**Figure II.13** – Fluorimetric titration of Phen-1 ( $\lambda_{exc}=330$  nm) and Phen-2 ( $\lambda_{exc}=350$  nm) with 22AG and c-myc. A,C) Binding curves of fluorescent titration experiment: plot of the fluorescence area enhancement  $F/F_0$  (interval area 340–600 nm for Phen-1 and 370–650 for Phen-2) as a function of added 22AG/c-myc concentration in 30 mM phosphate with 50 mM  $K^+$  buffer. Total concentration of ligand was maintained at 0.5  $\mu$ M. B,D) Stern-Volmer linearization of the values obtained in A,C.



The parameters obtained above were not calculated for Phen-1. While mixing Phen-1+DNA the fluorescence signals obtained are completely different after and before annealing of DNA; this suggests that binding of Phen-1 is not dynamic, therefore calculating an equilibrium constant is not appropriate to this case.

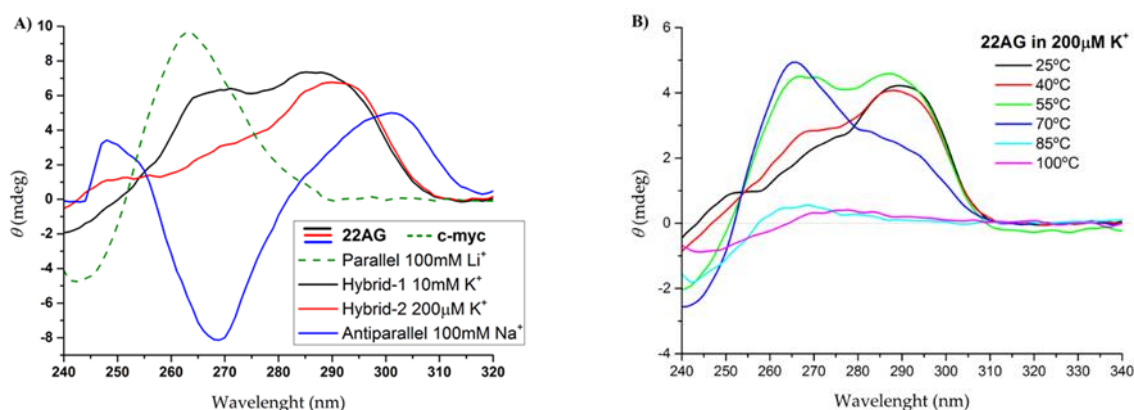
Overall, these results show us that Phen-1 and Phen-2 cannot be used as G4 fluorescent light-up probes. Nonetheless Phen-2 revealed very strong affinity for their targets.

### 3.5 – Circular Dichroism spectroscopy

CD studies were performed in order to further assess the binding of Phen-1 and Phen-2 ligands to G-quadruplex; these include thermal denaturation studies (CD-melting) with non-labeled quadruplexes. CD studies also enable to explore the conformational changes induced by the ligands.

The mutated oncogene promotor c-MYC and human telomeric 22AG quadruplex sequences were selected for the studies as their CD spectra are well characterized. As introduced, these G4 structures especially the telomeric 22AG are sensible to the concentration and nature of the salt applied (section I, fig. I.13). For example, in Na<sup>+</sup>, 22AG adopts an antiparallel topology in contrast to its parallel or hybrid antiparallel topologies formed in K<sup>+</sup> solution<sup>205,89,87</sup>. On the other hand, the mutated c-MYC adopts a well-defined parallel topology, which was studied in a Li<sup>+</sup> solution in order to displace its melting temperature to practical values<sup>107,215</sup>. The DNA-stabilizing properties of Phen-1 and Phen-2 were evaluated by measuring the ligand-induced change in melting temperature ( $\Delta T_m$ ) of 22AG and c-MYC, as well for a representative duplex-forming sequence, in CD melting experiments.

In 10 mM K<sup>+</sup> solution, the hybrid-1 telomeric 22AG quadruplex exhibits a strong positive peak at 290 nm with a shoulder at 265 nm, a smaller positive peak at 250 nm and a small negative peak at 235 nm (fig. II.14-A). This G4 sequence experiences at least two structural transitions before denaturizing with temperature (fig. II.14-B)<sup>55</sup>. In Na<sup>+</sup> solution antiparallel 22AG displays a broad positive signal at 290 nm and a sharp negative peak at 270 nm (fig. II.14-A). In contrast, this sequence unfolds in a single-step process. The folding-unfolding process for both sequences was found to be completely reversible, as was also reported<sup>61</sup>.

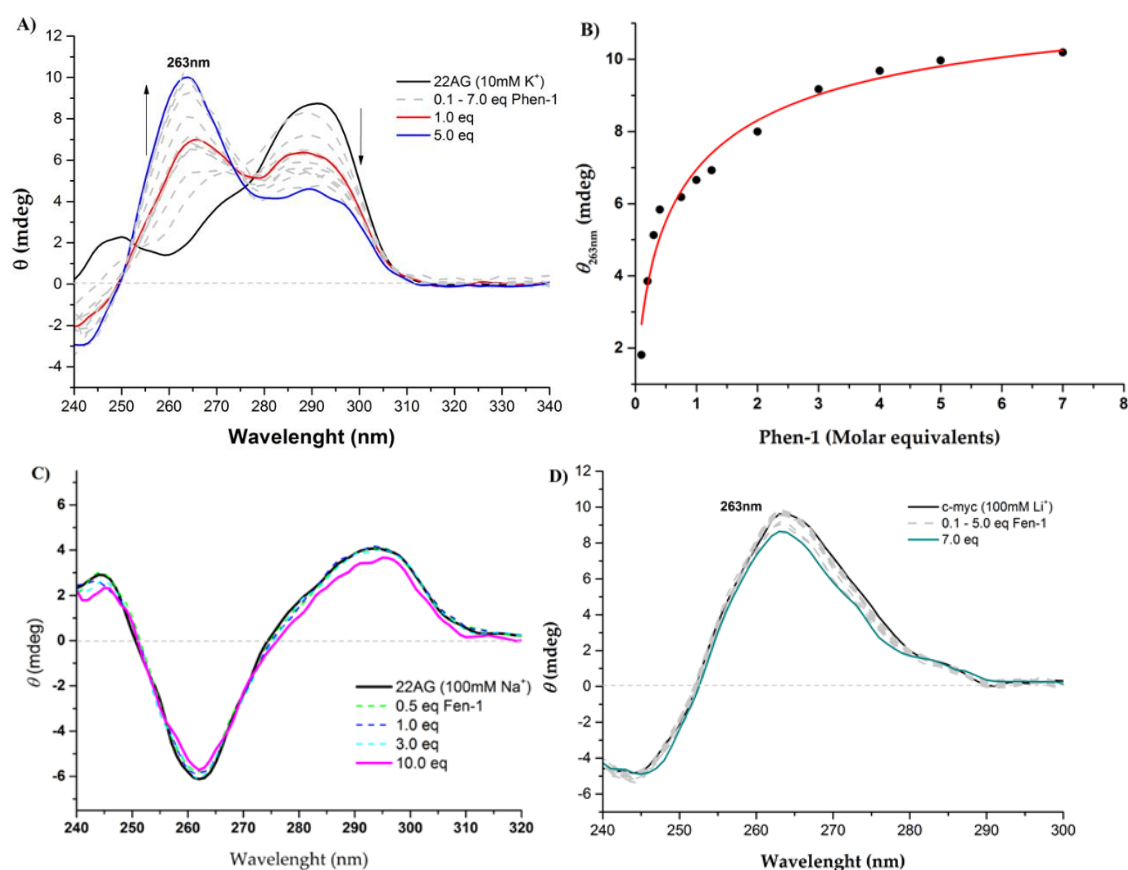


**Figure II.14** – (A) Representation of the different CD profiles of the G4 topologies employed in this study. (B) Demonstration of the structural transformation of 22AG in potassium media that occurs with increasing temperature. To be noticed that ellipticity values are approximately zero when the DNA is melted (linear single strain).

After addition of 0.5-1.25 equivalents of Phen-1 on 22AG in 10mM K<sup>+</sup>, the positive peak at 250 nm disappeared while a positive peak at 265 nm appeared accompanied by a weak negative band around 240 nm. These data are consistent with a hybrid-1 quadruplex folding topology. After addition of an excess of Phen-1 (5 equivalents), the CD spectrum showed a change of conformation into a parallel structure, which has a characteristic positive peak centered around 260 nm and a negative peak at 240 nm (fig. II.15-A). This profile is also similar to the transition that naturally occurs upon heating at approximately 70°C (fig. II.14-b) in ligand-free K<sup>+</sup> media. Also this effect has been correlated to partial dehydration of the telomeric G-quadruplex<sup>216</sup>. A Hill-plot of  $\Delta\theta_{263\text{nm}}$  suggests a 1:1 stoichiometry with possible groove binding interactions (fig. II.15-B). Phen-1 enhanced the stability of this parallel telomeric quadruplex by some modest 4.1°C.

Titration of antiparallel 22AG with Phen-1 in 100 mM Na<sup>+</sup> showed no relevant variations on DNA ellipticity, suggesting an overall conservation of the unimolecular antiparallel quadruplex topology (fig. II.15-C). Even when increasing the Phen-1 concentration by 10 times the G4 concentration, only a slight increase (up to 2.5°C) of thermal stability was observed.

The c-myc sequence predominantly adopts a parallel G-quadruplex topology, regardless of the cation present (fig. II.14-A). When Phen-1 was added to c-MYC, the parallel topology was retained (fig. II.15-D); nevertheless, parallel c-MYC structure was further stabilized by 4°C.



**Figure II.15** – (A) CD titration spectra of Phen-1 (0–7 molar equivalents) to quadruplex 22AG (10  $\mu\text{M}$ ) in the presence of K<sup>+</sup> (100 mM). (B) Hill plot of the ellipticity at 263 nm with addition of ligand. (C) CD titration spectra of Phen-1 (0–10 molar equivalents) to quadruplex 22AG (10  $\mu\text{M}$ ) in the presence of Na<sup>+</sup> (100 mM). (D) CD titration spectra of Phen-1 (0–7 molar equivalents) to quadruplex c-myc (5  $\mu\text{M}$ ) in Li<sup>+</sup> (100 mM). All DNA samples were suspended in phosphate buffer (30 mM, pH 7.2).



Furthermore, duplex DNA was also monitored by CD dichroism and CD melting studies to examine the effect of ligands on duplex interaction and stabilization. The CD spectra of an intermolecular duplex with and without Phen-1 (10.0 equivalents) were practically superimposable (sup. B.6) and no change in melting temperature was observed, thus suggesting that Phen-1 did not interact with the duplex.

These results suggest that Phen-1 favors telomeric parallel quadruplex topologies and selectively stabilize quadruplex over duplex DNA structures, which are in agreement with the FRET-melting data.

Phen-2 is our most promising ligand, as already indicated by the previous studies. Upon addition of 1.0 molar equivalent of Phen-2 to 22AG in 10mM K<sup>+</sup> occurs a structural transition from hybrid-1 quadruplex to hybrid-2 (fig. II.16-A), as indicated by the sharp positive peak at 290 nm with a shoulder and a small positive peak at 250 nm. This effect has been also verified for other phenanthroline compounds<sup>217</sup>. Controversially, addition of 2.0 molar equivalents of Phen-2 results in disappearance of the hybrid-2 peak profile and appearance of other positive peak at 268 nm, which indicates another structural transformation (fig. II.16-A). Phen-2 led to an increase in thermal stability of 22AG of 13°C and 15°C at 1.0 and 10.0 equivalents, respectively. The thermal stabilization behaved concentration-dependent until approximately 1 eq of Phen-2, afterwards the addition of ligand almost had no additional stabilizing effect (fig. II.17). This peculiar effect may indicate a G-quadruplex saturation point at 1.0 eq, which supports the data results from fluorimetric titrations.

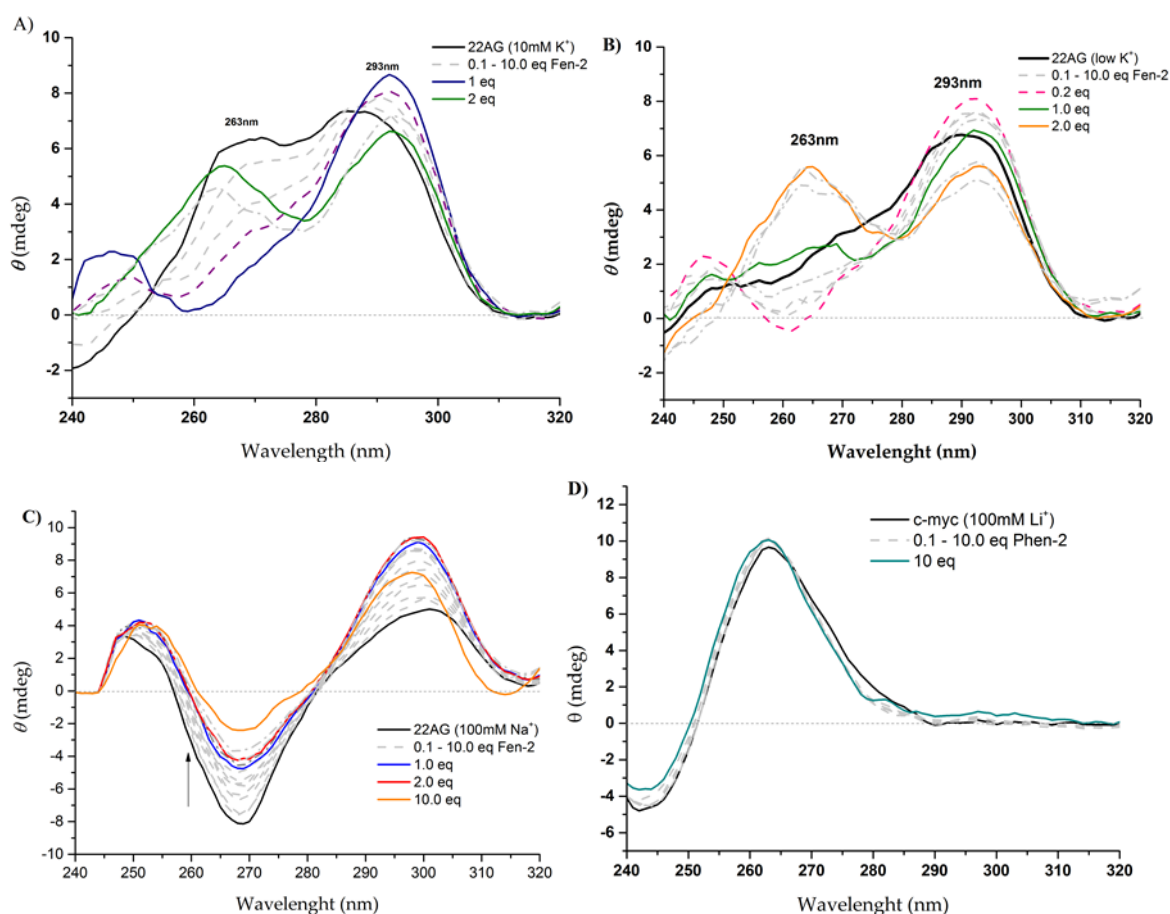
The effect of Phen-2 in 22AG was also studied in low salt conditions (the only salt contributions were those of the phosphate buffer). In low salt conditions, 22AG exists in a hybrid-2 quadruplex structure, since the CD spectrum contains two positive bands at 290 and 268 nm and a weak negative band at around 240 nm (fig. II.14-A). Addition of Phen-2 induced again structural modifications from hybrid-2 (0-0.5 equivalents of Phen-2) to hybrid-1 (1-10 equivalents) (fig. II.16-B). A similar increase in thermal stability (up to 12.8 °C) was obtained under those conditions, after addition of 10.0 equivalents of Phen-2. The concentration-dependence behavior was similar to the previous result (curves displayed on fig. II.17).

Likewise, we tested the influence of Na<sup>+</sup> on the observed topological switching and thermal stability. Phen-2 stabilizes the antiparallel topology of 22AG in Na<sup>+</sup> by 12.7 °C and enhanced the antiparallel profile ellipticity at 10.0 equivalents, but did not induce major structural modifications (fig. II.16-C). Again, the concentration-dependence behavior was similar to the previous two.

Adding an excess of Phen-2 (10 equivalents) to c-MYC in 100 mM Li<sup>+</sup> did not induce any change in c-myc ellipticity, suggesting that c-MYC retains a parallel topology (fig. II.16-D). Interestingly, Phen-2 enhanced the stability of c-MYC quadruplex by 31°C upon addition of 10 equivalents. The  $\Delta T_m$  values of Phen-1 and Phen-2 are listed in Table II.4.

These results are somewhat unexpected given that Phen-2 induced hybrid antiparallel topologies but greatly stabilizes parallel conformation. In fact, a study shows that the parallel form of telomeric quadruplex may not be biologically relevant<sup>92</sup>, and therefore, ligands that could induce antiparallel quadruplex topology of human telomeric DNA may find use in the therapeutic arena. Moreover, these results also demonstrated that Phen-2 potentially stabilizes the *c-myc* promoter quadruplex much better than human telomeric quadruplex.

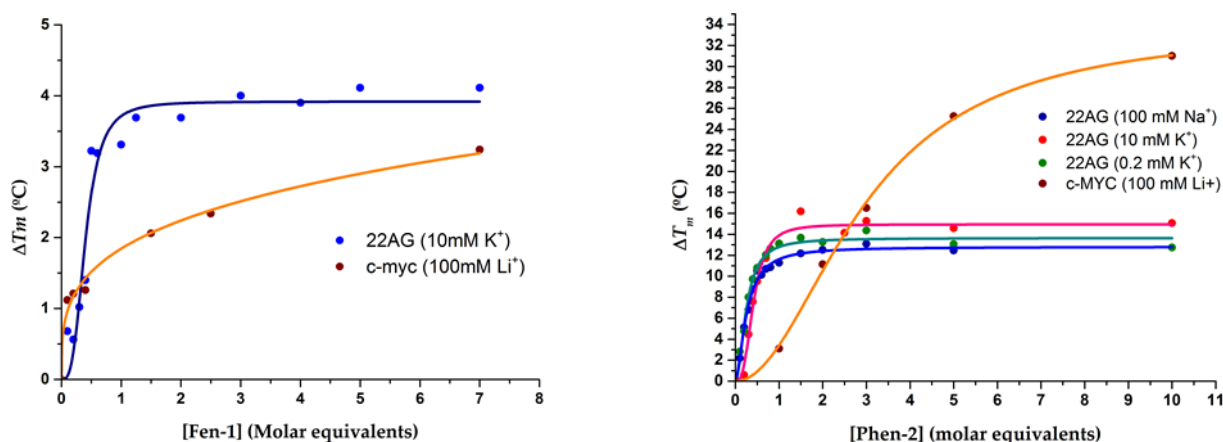
Effects of Phen-2 were also investigated on a duplex sequence (sup. B.6) and, as expected, no major modifications on the CD profile and no significant change in the thermal stability of the double helix were observed, in agreement with FRET-melting studies.



**Figure II.16** – (A) CD titration spectra of Phen-2 (0–20 molar equivalents) to quadruplex h22L (10  $\mu$ M) in the presence of  $K^+$  (10 mM). (B) CD titration spectra of Phen-2 (0–10 molar equivalents) to quadruplex 22AG (10  $\mu$ M) low salt content (0.2 mM  $K^+$ ). (C) CD titration spectra of Phen-2 (0–10 equivalents) to 22AG (10  $\mu$ M) in the presence of  $Na^+$  (100 mM). (D) CD titration spectra of Phen-2 (0–25 molar equivalents) to quadruplex c-myc (5  $\mu$ M) in  $Li^+$  (100 mM). All DNA samples were suspended in phosphate buffer (30 mM, pH 7.2).

Figure II.17 plots the calculated  $\Delta T_m$  of the sequences in the conditions mentioned above with the concentration (in molar equivalents) of Phen-1 and Phen-2, which are summarized in table II.4. Surprisingly, the ligand concentration dependency was very different for the two G4 investigated. For telomeric 22AG, a plateau is reached below two molar equivalents of Phen-2. In contrast, c-myc's  $T_m$  gradually increased with Phen-2 concentration which may point to groove-binding interactions. If true, this would support the results for weak TO displacement obtained earlier. Nevertheless, these results suggest two distinct binding modes for Phen-2 depending on quadruplex topology.

The cation effect does not seem to have a major impact on the binding event; although it is verified a larger stabilization in  $K^+$  solution than in  $Na^+$ , most probably due to the coordination effects on oxazole rings. This underscores the importance of this moiety in quadruplex topology recognition.



**Figure II.17** – Effect of Phen-2 and Phen-1 concentration on melting temperature increase of 22AG quadruplex in 10 mM K<sup>+</sup>, 200 μM K<sup>+</sup> and 100 mM Na<sup>+</sup> and c-myc in 100 mM Li<sup>+</sup>.

**Table II.4** - Thermal stability of several G4 with ligands measured by CD melting.

Ligands	$\Delta T_m$ (°C) <sup>a</sup>			
	ds46	c-MYC	22AG <sup>b</sup>	22AG <sup>c</sup>
Phen-1	0.32 ± 1.4	4.34 ± 0.6	4.11 ± 0.1	2.46 ± 0.3
Phen-2	-2.80 ± 1.5	31.01 ± 0.2	15.06 ± 0.2	12.74 ± 0.3

<sup>a</sup>  $\Delta T_m$  represents difference in melting temperature [ $\Delta T_m = T_m(\text{DNA} + 10 \text{ molar eq. ligand}) - T_m(\text{DNA})$ ]. The buffer used was 30 mM phosphate buffer, pH 7.2; Measured  $T_m$  for G-quadruplexes in absence of ligand: 50.1 ± 0.1 °C [*c-myc* (5 μM) in 100 mM LiCl]; 60.9 ± 0.2 °C [hybrid type telomeric DNA (10 μM) in 10 mM KCl]; 61.1 ± 0.1 °C [antiparallel telomeric DNA (10 μM) in 100 mM NaCl]. All experiments were done in triplicate, and the values reported are average of 3 measurements with the estimated standard deviation. Curves are presented in the appendix B.

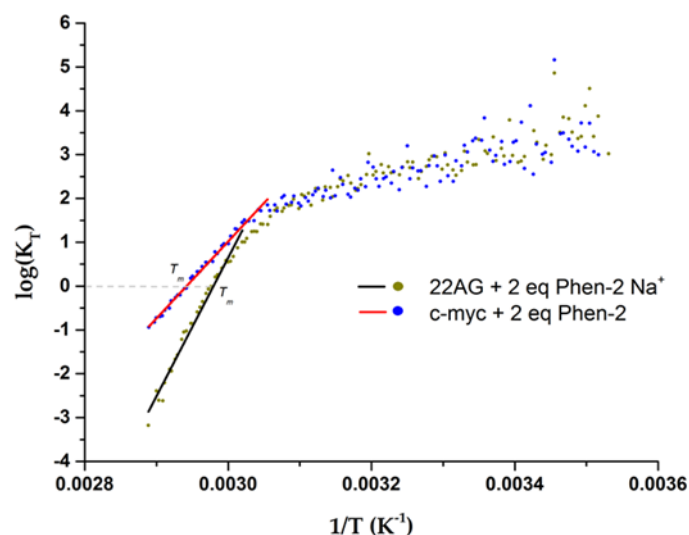
<sup>b</sup> Hybrid type parallel-antiparallel topology of 22AG in 10 mM K<sup>+</sup> buffer.

<sup>c</sup> Antiparallel topology of 22AG in 100 mM Na<sup>+</sup> buffer.

Interpretation of thermal denaturation curves may unveil the energetics of the ligand-G4 interaction. By application of Arrhenius plot ( $\ln(K_T)$  vs  $1/T$ ) it is possible to determine the van't Hoff enthalpy  $\Delta H^\circ$  and entropy  $\Delta S^\circ$  parameters (see experimental section). This study is limited to 22AG in Na<sup>+</sup> and *c-myc* experiments, given that denaturation of 22AG in K<sup>+</sup> is not a simple two-step mechanism. Also, the curves employed will correspond to 2.0 equivalents of Phen-2 added.

The Arrhenius plot is displayed in Figure II.18. It can be readily perceived that the data obtained are not entirely linear, which implies that  $\Delta H^\circ$  dependency with temperature is not constant. Thus this approach is not the most appropriated for calculating these parameters and other calorimetric techniques should be applied. Nonetheless an approximation may in principle be performed by applying a regression into the linear region that corresponds to the melting transition (fig. II.18, red and black lines) which provide an approximate enthalpic value. From the fitted lines yields  $\Delta H^\circ = -145.79 \text{ KJ mol}^{-1}$  for *c-myc* and  $\Delta H^\circ = -262.60 \text{ KJ mol}^{-1}$  for 22AG, which are comparable to other good G4 ligands of the same nature. Interestingly, in presence of Phen-2 the  $\Delta H^\circ$  of *c-myc* was considerably lower than 22AG's.

When ligands interact preferentially through hydrophobic interactions (G-tetrad stacking) the process is mostly enthalpic driven<sup>51</sup>. In contrast, groove-binding ligands generally display low values of  $\Delta H^\circ$ . As comparison, the groove-binder distamycin showed  $\Delta H^\circ = -14.0 \text{ KJ mol}^{-1}$  with d(TGGGGG)<sub>4</sub><sup>218</sup>, thus the binding process should be entropically driven. For this reason, the obtained  $\Delta H^\circ$  values may indicate that Phen-2 ligands interact preferentially with 22AG through G-tetrad binding while for c-myc quadruplex it may present more than one type of interactions.

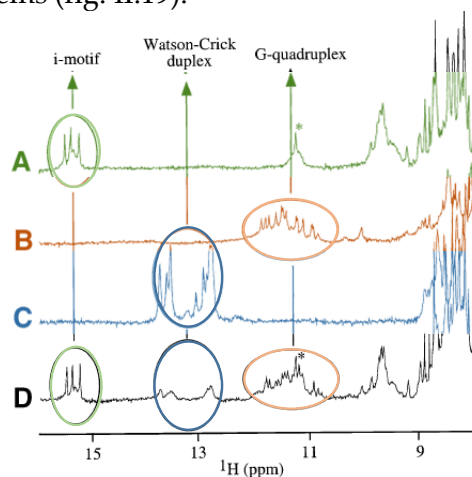


**Figure II.18** – Thermodynamic analysis of the melting curves of 22AG in sodium and c-myc in lithium with 2.0 equivalents of Phen-2. The slope of the fitted region provides  $\Delta H^\circ/R$ .

Overall, these studies showed that Phen-2 induces antiparallel topologies on 22AG but also greatly stabilizes c-myc parallel G-quadruplex; in addition the binding mechanism should be different between two sequences. In contrast, Phen-1 favored parallel topologies but barely stabilized G-quadruplex structures.

### 3.5 – Nuclear Magnetic Resonance Spectroscopy

NMR techniques may provide important structural information at the atomic level. In a <sup>1</sup>H-NMR, each signal corresponds to one (equivalent) proton nuclei of the molecule analyzed, so the signals correspondent to certain G-tetrad protons may be monitored and distinguished from other base-pairing systems (fig. II.19).

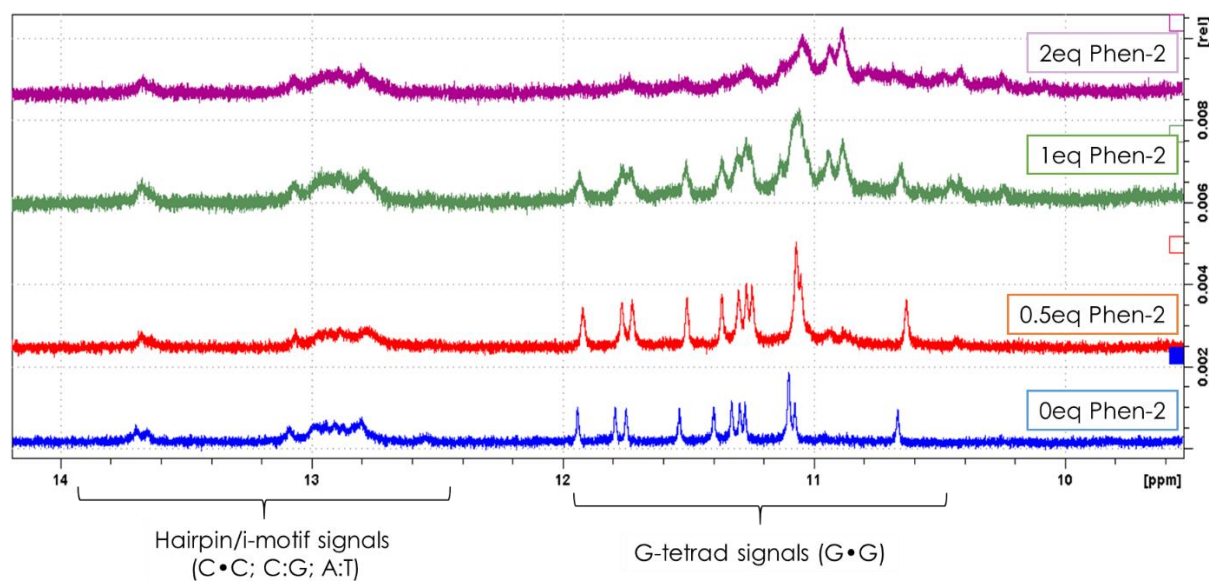


**Figure II.19** – Schematic representation of the typical chemical shifts for the imine regions of 3 base-pairings.

In contrast to 22AG, the c-myc G-quadruplex is very rigid and well characterized by NMR, thus this sequence is perfectly suited as a model for structural characterization. In figure II.20 is displayed a partial spectrum (full spectrum in sup. B.11, B.13) of the c-myc G-quadruplex sequence. In this sample the complementary strand of c-myc was also added in order to study how the ligand will behave and how it will affect the duplex-quadruplex equilibria.

In the absence of ligand, the c-myc sequence predominantly adopted a G-quadruplex motif, as can be verified from the signals in the region between 10.5-12 ppm (fig. II.19 - blue spectra). These correspond to the imino protons and are characteristic of the G-tetrad due to the NH...O hydrogen bonds that result from the Hoogsteen alignments. The other downfield signals (12.5-13.5 ppm) correspond to imino signals that result from Watson-Crick pairing, thus either the c-myc is partially on a duplex arrangement or the complementary sequence folded on itself forming an i-motif, or simply a hairpin.

Addition of Phen-2 (0,5-2.0 molar equivalents) to the c-myc sample in 100mM K<sup>+</sup> resulted in gradual shifting and linebroadening of the G-tetrad signals. This is a clear indication of close interaction with the Phen-2 ligand. Furthermore, two new signals appeared around 10.9 ppm which indicate a new chemical environment in their surroundings. The Watson-Crick signals remained unchanged, which emphasizes the Phen-2 ligand specificity.



**Figure II.20** – NMR titration of c-myc G-quadruplex with Phen-2 in 30mM phosphate buffer, 100mM K<sup>+</sup>, at 25°C with 10%D<sub>2</sub>O at 600MHz. The ligand/quadruplex ratios are shown along the side of the spectra. DNA strand concentration of 400μM.

NOESY data could provide the definitive prove and reveal the site of the interaction between Phen-2 with the c-myc. These experiments are currently being processed.

## 4 – CONCLUSIONS

From the potential ligands synthesized, two G4 ligands Phen-1 and Phen-2, were effectively identified with a high-throughput FRET-melting assay. These ligands also displayed very high G-quadruplex over duplex selectivity. The other ligands failed to display any significant thermal stabilization with the G4 sequences employed.

The interaction of the ligands with G4 sequences were studied by several techniques such as fluorescence spectrometry, circular dichroism, G4FID assay, DSF and NMR titrations. DSF and G4FID assays combined showed that Phen-2 does not compete with TO in order to bind the G4. Therefore this ligand in principle should interact through G-tetrad stacking

CD studies show that Phen-2 binds and induces telomeric antiparallel topologies in 22AG quadruplex ( $K_a = 9.56 \times 10^9 M^{-1}$ ) and also binds c-myc promotor ( $K_a = 3.55 \times 10^6 M^{-1}$ ), increasing their  $T_m$  in about 12°C and 30°C respectively. Overall, the results show that Phen-2 binds more effectively to telomeric antiparallel quadruplex but displays higher ability for stabilizing parallel quadruplexes. CD studies also indicate two different binding Phen-2 ways depending on the G-quadruplex topology. In contrast, Phen-1 induces telomeric parallel topologies in 22AG and binds c-myc, resulting in low thermal stabilization of about 4°C for both sequences.

Although the ligand affinity towards the G4 is well demonstrated in solution, its ability of inducing and stabilizing the G-quadruplex in long crowded DNA sequences was not entirely demonstrated. NMR titration studies showed that Phen-2 did not bind other motifs than the G-quadruplex. Still, a proof-of-concept biochemical assay (such as Taq polymerase stop-assay) should be performed in order to demonstrate *in vitro* that Phen-2 is capable of inhibiting telomerase/polymerase activity through the G-quadruplex mechanism.



## Part C

# Oligoheteroaryle G-quadruplex Ligands: biological evaluation

### 1 – INTRODUCTION

Preliminary biological studies were performed in order to evaluate ligands cytotoxicity and anti-proliferative effects in cancerous and normal cells.

Other proof-of-concept methodologies, such as fluorescence microscopy, will be performed in future investigation; for example, if the ligand exhibit anti-cancer properties it should be demonstrated that the ligand effectively reach the nucleus in order to bind telomeres or oncogene promoters.

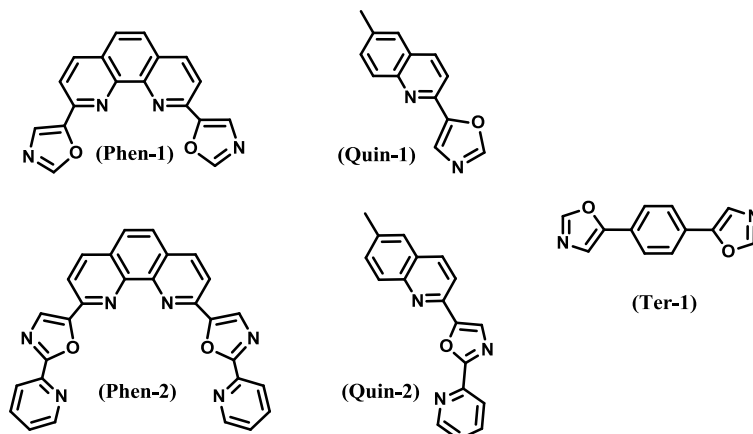
In the following studies three different human cancer cell lines were employed as experimental models (MCF-7, LNCaP and U87) and a normal (i.e. non-cancer) human cell line (NHDF) as selectivity control model.

- MCF-7 cell line was derived from a pleural effusion taken from a patient with metastatic breast cancer<sup>219</sup>. Thus these cell represent a model for breast cancer. The sample was taken from a 69 –year-old Caucasian woman in the *Michigan Cancer Foundation* (MCF).
- LNCaP cell line was established from a metastatic deposit in a lymph node from a 50-year-old Caucasian male<sup>220</sup>. LNCaP stands for *Lymph Node Carcinoma of the Prostate*.
- U87 is a human primary epithelial glioblastoma cell line from a 44-year-old Caucasian male<sup>221</sup>. Thus these cell represent a model for brain cancer.
- NHDF stands for *Normal Human Dermal Fibroblast* are healthy cells from adult human skin.

The anti-proliferative and cytotoxic effects of the ligands were investigated through MTT assay. The MTT assay is a rapid colorimetric assay for assessing the cellular metabolic activity<sup>222</sup>. Briefly, it consists on quantification of formazan ( a water insoluble purple compound) that results from reduction of a tetrazolium compound (MTT) by NADH and other oxidoreductases<sup>223</sup>. Thus the formazan formation is correlated to cellular metabolism and, consequently, to cellular viability.

In the following experiments the Quin, Ter and Phen synthesized compounds were evaluated (fig. II.21), even if they did not display any G-quadruplex affinity. The reason for

this approach is an attempt of exploring unpredicted effects of the compounds synthesized. If by chance a compound exhibit anti-cancer properties, it may be considered and studied in future investigations in order to identify the mechanism responsible for this effects.



**Figure II.21** – Chemical structures of the ligands evaluated in the following study.



## 2 – EXPERIMENTAL

### 2.1 – General

The experimental work was performed in asepsis conditions on a laminar flow hood (Vertical NuAire, Classel) during cell samples manipulation. Cellular samples were incubated on a Nuair DHD Autoflow CO<sub>2</sub> Air-Jacked at 37°C. Plates were read with a Bio-Rad Xmark spectrophotometer.

All aqueous solutions were prepared with high-pure Milli-Q (Millipore) water and sterilized either through temperature or filtration.

### 2.2 – Preparation of compounds stock solutions

The stock solution of compounds were prepared in DMSO (>99% purity, Sigma-Aldrich) at 5 mM (Phen-1 and Phen-2) and 10 mM (Ter-1, Iso-1, Iso-2, Quin-1 and Quin-2) concentration. Compounds were added directly to the cell culture with the adequate dilution prior to each experiment. The maximum safe DMSO final concentration in the cell culture is set at 1%.

### 2.3 – Cell cultures

MCF-7, LNCaP, U87 and NHDF cells were obtained from American Type Culture Collection (ATCC; Manassas, VA, USA) and were cultured in 75 or 175 cm<sup>2</sup> culture flasks and stored at 37 °C in a humidified air incubator with 5% CO<sub>2</sub>.

MCF-7 and U87 cells were cultured in the high-glucose Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS; Biochrom AG..) and with 1% of an antibiotic/antimycotic solution (10,000 U/mL penicillin G, 100 mg/mL streptomycin and 25 µg/mL anfotericin B) (*Ab*; Sigma-Aldrich).

LNCaP cells, were cultured in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 10% FBS and 1% of the antibiotic mixture of 10,000 U/mL penicillin G and 100 mg/mL of streptomycin.

NHDF cells were cultured in RPMI 1640 medium supplemented with 10% FBS, 2 mM L-glutamine, 10 mM HEPES, 1 mM sodium pyruvate (final concentration) and 1% *Ab*.

The cellular culture medium was changed every 2-3 days once cells reach 90-95% confluence. In this stage, cells were suspended by gentle trypsinization, viable cells were counted by trypan-blue (0.4%, Merck) exclusion assay and were properly diluted into a new culture medium.

### 2.4 – Incubation of compounds with cells

MCF-7, LNCaP, U87 and NHDF cells were seeded in 48-well culture plates (Nunc, Apogent, Denmark) with 250µL of cell suspension on each well (cellular density  $2 \times 10^4$  cells/mL, counted by trypan-blue exclusion assay). Cells were left to adhere for 48 h and after this period the cellular medium was replaced by the drug-containing medium at the final concentration described along the text (30 µM for screening and 0.01, 0.1, 1.0, 10.0, 50.0 and 100.0 µM for IC<sub>50</sub> determination). Cells were incubated with the ligand for a period of 72 h prior to evaluation. Each sample/well conditions were replicated in quintuplicate. A ligand-free control replicate was reproduced in each plate.

Each plate was reproduced at least in two independent experiments.

## 2.5 – MTT assay

The anti-proliferative effect of the ligands were measured indirectly through MTT assay. After 72 h compound incubation the cellular medium was removed and each well was washed with 200  $\mu$ L of a saline 10mM sodium phosphate buffer solution (PBS, 130 mM NaCl, 2.7 mM KCl, pH 7.4). 250  $\mu$ L of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT, CAS#298-93-1, Amresco) at 5 mg/mL concentration in PBS was added to each well followed by a 4 h incubation at 37 °C. After that, the MTT solution was carefully removed from the wells, keeping the formazan crystals. 225  $\mu$ L of DMSO was added to each well in order to dissolve the formazan crystals. Plates were transferred to a plate reader where the absorbance at 570nm was recorded.

Cell viability values were normalized and represented as percentage relatively to the ligand-free control wells. Blank plate values were subtracted. The results were presented as a mean $\pm$ standart deviation (SD) of the five well replicates. The difference between replicates was considered statically relevant when  $p < 0.05$ . IC<sub>50</sub> values were estimated from fitted sigmoidal curves considering a 95% confidence interval.

### 3 – RESULTS AND DISCUSSION

#### 3.1 – Anti-proliferative analysis

The anti-proliferative effect of the compounds were firstly examined through a MTT assay after 72 h exposition of each compound at 30  $\mu$ M. These conditions were chosen in order to screen which ligands are cytotoxic for further in depth studies. The overall mean results are displayed in table II.5.

**Table II.5** – Mean relative cell proliferation (% of the control) of the studied cells after 72 h exposition to the compounds (30  $\mu$ M). Potential active compounds were highlighted.

Cell lines	MCF-7	LNCaP	U87	NHDF
Ter-1	97.1 $\pm$ 11.2	66.9 $\pm$ 7.9	96.3 $\pm$ 4.4	98.2 $\pm$ 10.0
Quin-1	70.1 $\pm$ 8.2	42.9 $\pm$ 16.9	81.7 $\pm$ 0.0	-
Quin-2	32.2 $\pm$ 0.4	27.0 $\pm$ 10.7	114.2 $\pm$ 2.7	85.0 $\pm$ 5.4
Phen-1	78.9 $\pm$ 35.4	65.0 $\pm$ 11.5	77.8 $\pm$ 10.3	100.6 $\pm$ 7.5
Phen-2	3.6 $\pm$ 2.0	1.0 $\pm$ 2.7	1.2 $\pm$ 2.3	7.6 $\pm$ 0.7*

Detailed plots on appendix C. MCF-7 – breast cancer; LNCaP – prostate cancer; U87 – glioma (brain cancer); NHDF – normal fibroblasts. \*Duplicate discarded as internal results were not consistent.

Ter-1, Quin-1 and Phen-1 did not display any interesting anti-proliferative effect. There is an exception for Quin-1 on LNCaP cells that resulted on 57% inhibition (43% proliferation), but still, this is considered a mild inhibition result. On the other hand, Quin-2 showed cell inhibition of 68% (32% proliferation) and 73% (27% proliferation) for breast and prostate cancer cells, respectively. Curiously, on brain cancer cells (U87) Quin-2 showed an opposite effect as it even slightly increased cancer cells proliferation. However in normal cells (NHDF) Quin-2 had a negligible effect in cell proliferation, indicating selectivity on the anti-proliferative mechanism. These results were unexpected since Quin-2 didn't interact or stabilize G-quadruplex structures; therefore Quin-2 should in principle be involved in other mechanism. Furthermore, this mechanism is only being active or somehow affected on prostate and breast cancer, with little impact on normal cells and glioma cells.

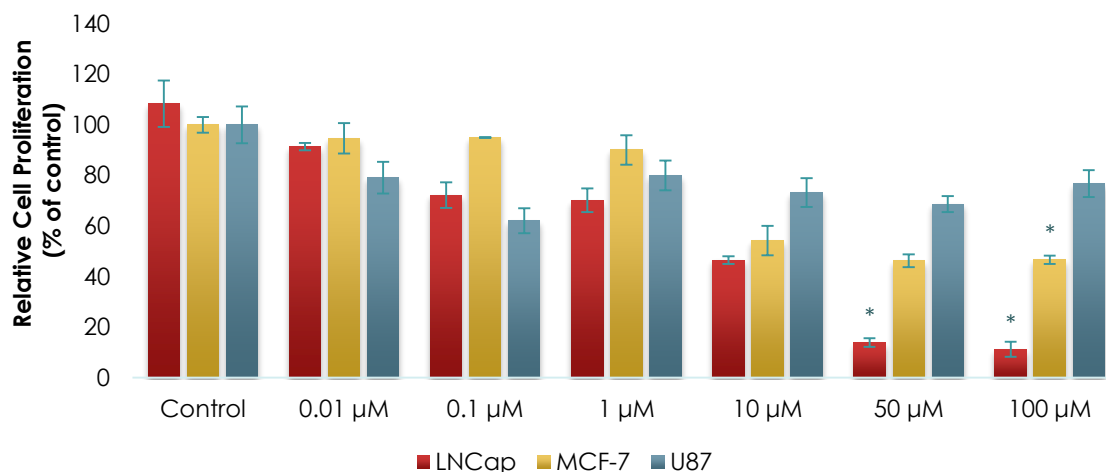
Concerning to the promising G-quadruplex binder, Phen-2 also exhibited an outstanding anti-proliferative effect with 95% of inhibition in all cells lines studied. Unfortunately, Phen-2 displayed a cytotoxic effect equally on the normal cells, which might raise some questions about the Phen-2 mechanism. It might be noticed that Phen-2 symmetrically has the same groups as Quin-2, and therefore there is a possibility that the two compounds might be interfering with the same mechanism, but this is speculative.

The cytotoxic effect of Phen-2 in normal cells (NHDF) still need confirmation by doing more replicates.

Based on results of table II.5, the most anti-proliferative compounds (Phen-2 and Quin-2) were selected for determination of concentration-response ( $IC_{50}$  parameter). In this assay, cells were exposed to 0.01, 0.1, 1, 10 50 and 100  $\mu$ M concentrations of these compounds during 72 h. Estimated  $IC_{50}$  are summarized on table II.6 and dose-response profile are displayed on fig. II.22 for Quin-2 and II.23 for Phen-2.

Quin-2 did not exhibit any significant anti-proliferative effect for U87 glioma cells, as verified earlier, and its dose-response profile was inconsistent. In contrast, for MCF-7 breast cells it did present very interesting values of half inhibitory concentration,  $IC_{50} = 39.14 \mu\text{M}$ , which are on the same range of values of diosgenin, a described breast cancer anti-proliferative steroid compound<sup>224</sup>. Curiously, the anti-proliferative effect intensity reaches a plateau at  $10 \mu\text{M}$ . Very potent inhibitory effects were also obtained for LNCaP prostate cancer cells, with a  $IC_{50} = 4.17 \mu\text{M}$ . Furthermore the response was linear with increasing concentrations of Quin-2.

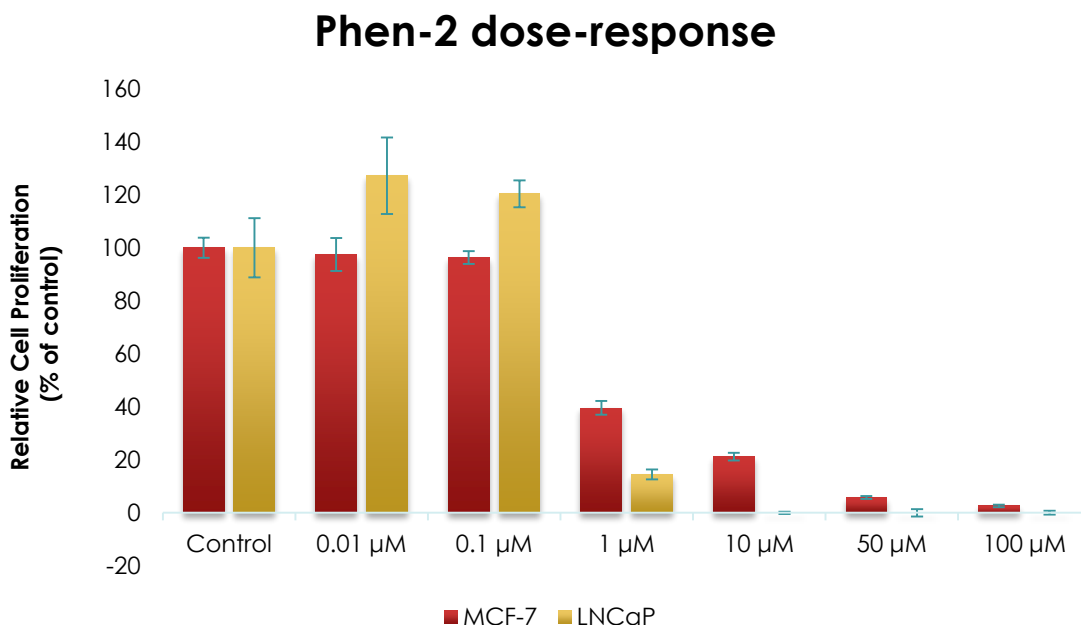
### Quin-2 dose-response



**Figure II.22** – Plot of the relative cell proliferation of LNCaP, MCF-7 and U87 cell lines incubated with Quin-2 during 72 h exposition in concentrations ranging from 0 to  $100 \mu\text{M}$ , determined by the MTT assay. Data are representative of two independent experiments for each cell line. \* $p < 0.05$  versus the control.

As expected, Phen-2 exhibited a strong anti-proliferation effect. For MCF-7 cells the anti-proliferative response was linear and had a low  $IC_{50}$  of  $0.64 \mu\text{M}$ . For LNCaP cells a peculiar effect was observed, at low concentrations ( $0.01$ - $0.1 \mu\text{M}$ ) Phen-2 increased cellular proliferation; however as the concentration of Phen-2 increases ( $0.1$ - $10 \mu\text{M}$ ) a cytotoxic mechanism is triggered leading to total proliferation inhibition or cellular death ( $IC_{50} = 0.40 \mu\text{M}$ ).

This results accentuate the idea of Phen-2 as a cytotoxic compound, so until further studies on normal epidermal cells, extra precautions must be adopted while manipulating it.



**Figure II.23** – Plot of the relative cell proliferation of LNCaP, MCF-7 and U87 cell lines incubated with Quin-2 during 72 h exposition in concentrations ranging from 0 to 100  $\mu\text{M}$ , determined by the MTT assay. Data are representative of two independent experiments for each cell line. \* $p < 0.05$  versus the control.

**Table II.6** – Estimated  $\text{IC}_{50}$  values ( $\mu\text{M}$ ) for Phen-2 and Quin-2<sup>a</sup>

Cell lines	MCF-7		LNCaP		U87	
	$\text{IC}_{50}$ ( $\mu\text{M}$ )	$R^2$	$\text{IC}_{50}$ ( $\mu\text{M}$ )	$R^2$	$\text{IC}_{50}$ ( $\mu\text{M}$ )	$R^2$
<b>Quin-2</b>	39.14	0.93	4.17	0.99	NR	
<b>Phen-2</b>	0.64	0.95	0.40	0.99	ND	

<sup>a</sup> $\text{IC}_{50}$  were fitted by dose-response sigmoidal fitting. NR = not reached 50% in the concentration range; ND = not determined.

Overall, Phen-2 and Quin-2 displayed effective anti-proliferative effects. One of this might be related to the G-quadruplex mechanism (Phen-2) and the other probably triggers other unassigned mechanism (Quin-2). Furthermore Quin-2 had negligible cytotoxic effects on normal cells, thus this compound could potentially find application on the therapeutic area. Conclusions about Phen-2 selectivity are dubious at this moment as the only assay lacks replicates. Nevertheless, these are still very preliminary results that need to be explored. Also, co-localization studies using confocal microscopy must be performed in order to investigate the localization of Phen-2 and Quin-2 to suggest if the mechanism might be DNA related or not.

## 4 – CONCLUSIONS

The synthesized compounds Ter-1, Phen-1, Phen-2, Quin-1 and Quin-2 were evaluated regarding their cytotoxic effect on MCF-7, LNCaP and U87 cancer cells and NHDF normal cells. Phen-2 and Quin-2 exhibited strong anti-proliferative effect while the others failed to display a significant response.

Phen-2 revealed an extreme cytotoxic effect towards cancer cells, during 72 h exposure and displayed a  $IC_{50}$  of 0.40  $\mu$ M for prostate cancer cells,  $IC_{50}$  = 0.64  $\mu$ M for breast cancer cells and 98% of inhibition in glioma cells at 30  $\mu$ M. The relation of this effect with G-quadruplex is not confirmed, and its selectivity vs normal cell is still unclear.

Quin-2 also displayed strong anti-proliferative effects, although in principle these are not related to the G-quadruplex. Quin-2 revealed a good  $IC_{50}$  = 39.14  $\mu$ M for breast cancer cells and a very good  $IC_{50}$  = 4.17  $\mu$ M for prostate cancer cells, whereas its effects on glioma cells were negligible. The fact that this ligand did not entirely affect normal fibroblast cells provides insights for exploring new anti-cancer mechanisms.

# III. Final Conclusions

## CONCLUDING REMARKS

Eight new potential G4-binding acyclic oligoheteroaryle compounds were synthesized and characterized. The synthesis of these compounds were based on van Leusen oxazole formation from aldehydes by reaction with tosylmethyl isocyanide (TosMIC), followed by cross-coupling reaction through C-H activation; this strategy afforded phenanthroline-derived compounds Phen-1 (50%) and Phen-2 (20%), phenyl Iso-1 (61%), Iso-2 (21%), Ter-1 (85%) and Ter-2 (35%), and quinolyl Quin-1 (85%) and Quin-2 (45%) compounds. The van Leusen reaction procedure was modified and optimized, increasing obtained yields up to 10 fold. The synthesized compounds were characterized by NMR and fluorescence spectroscopy.

FRET-melting and FID assays in human telomeric DNA (F21T and 22AG), *c-myc* oncogene promotor (FmycT and c-myc) and a duplex sequence showed that Phen-1 and Phen-2 discriminated among various quadruplex topologies and exhibited high selectivity for G-quadruplexes over duplexes.

CD melting and titration studies revealed that Phen-2 induces antiparallel topologies for quadruplex 22AG whereas it did not induce conformational changes on the *c-myc* quadruplex. Phen-2 increased the thermal stabilities of *c-myc* by 31 °C and 22AG by 12.7 °C (in Na<sup>+</sup>) and 15.0 °C (in K<sup>+</sup>). In contrast, Phen-1 favored telomeric parallel topologies but barely stabilized the G-quadruplex motif, increasing  $\Delta T_{1/2}$  of *c-myc* by 4.3 °C and 22AG by 4.1 °C in K<sup>+</sup>, whereas no significant stabilization was verified for 22AG in Na<sup>+</sup>. Nonetheless, this part of the study underscores the importance of the central aromatic scaffold phenanthroline which could be further fine-tuned with other substituents in order to enhance binding and stabilization to G-quadruplex structures, as occurred with Phen-2.

Fluorescence titration studies revealed of a strong quenching effect of Phen-2 fluorescent signal upon interaction with G-quadruplex structures. These revealed a strong binding affinity towards 22AG ( $K_a = 9.56 \times 10^9 M^{-1}$ ) and *c-myc* ( $K_a = 3.55 \times 10^6 M^{-1}$ ) sequences, admitting a 1:1 stoichiometry. A much milder quenching effect was verified for Phen-1 ligand.

A novel DSF assay based on G4-FID was developed in order explore and evaluate the binding mechanism of Phen-1 and Phen-2 to c-MYC and 22AG. Combining these results with thermodynamic data, it is suggested that Phen-2 might interact both through G-tetrad end-stacking and groove binding mechanisms, while Phen-1 seems to preferentially interact through groove binding.

The ligands were biologically evaluated concerning their cytotoxic and anti-proliferative effects by MTT assay on three cancer cell lines (MCF-7, LNCaP and U87) and one healthy cell line (NHDF). Solely Phen-2 and Quin-2 exhibited significant cytotoxic effects. Phen-2 and Quin-2 ligands displayed strong anti-proliferative effects on LNCaP ( $IC_{50} = 0.40$  and  $39.14 \mu M$ , respectively) and MCF-7 ( $IC_{50} = 0.64$  and  $4.17 \mu M$ , respectively) cancer cell lines. Phen-2 also revealed cytotoxic effects on U87 and normal NHDF cells while Quin-2 surprisingly did not.

Thus, Quin-2 could potentially be explored as a selective agent for breast (MCF-7) and prostate (LNCaP) cancer.

Overall, new G4 ligands were synthesized which exhibited specific and strong interactions with G-quadruplex structures; one of these G4 ligands exhibited strong anti-proliferative effects on cancer cells but the involvement of the G4 mechanism is yet to be confirmed. The quest for the perfect G4 ligand is far from over but we definitely took a step closer.

## FUTURE PERSPECTIVES

The low water solubility of Phen-2 ligand is clearly a disadvantage. Thus, new strategies with aliphatic polyamines are being developed in order to improve its solubility and further enhance affinity towards the G4, as these amine groups could engage in preferential electrostatic backbone interactions.

The binding studies of the compounds should be extended to other G-quadruplex sequences; it has been verified that some ligands induce and stabilize other types of G-quadruplex topologies. Also proof-of-concept biochemical assays, such as the PCR-stop assay, should be applied in order to assess if the ligands are effectively able to inhibit polymerase through G-quadruplex binding.

Structural molecular insights of the G4-Phen-2 complex needs to be elucidated by NMR experiments to provide information of the recognition process and interactions involved.

Phen-2 exhibited potent anti-proliferative effect, however needs to be confirmed if this effect is related to G-quadruplex inducing and stabilization mechanism. In a preliminary approach, it should be investigated *in vitro* by confocal microscopy if Phen-2 effectively reaches the nucleus of the cells.

Quin-2 anti-proliferative effect is in principle not related to G-quadruplex target, hence an extensive *in vitro* investigation should be employed in order to identify the cellular mechanism of this compound.



## COMMUNICATIONS RELATED TO THIS WORK

### Oral communications

- **Phenantroline dioxazole as potential G-quadruplex ligand for cancer therapy**, João Medeiros-Silva, Eurico J. Cabrita and Carla Cruz, *International Symposium on Synthesis and Catalysis*, Évora, Portugal, September 2015.
- **Phenanthroline polyoxazoles as G-Quadruplex ligands for cancer therapy**, João Medeiros-Silva, Eurico J. Cabrita and Carla Cruz, *X Annual CICS-UBI Symposium*, Covilhã, Portugal, 2015.
- **G4 binders for cancer therapy**, *Chemistry and Life Sciences Department Annual Master's Dissertations Symposium*, Monte de Caparica, Portugal, 2015

### Poster communications

- **G-quadruplex production and stabilization**, João Medeiros-Silva, Josué Carvalho, Eurico J. Cabrita and Carla Cruz; *5th International Meeting on Quadruplex Nucleic Acids*, Bordeaux, France, 2015
- **Synthesis and biological evaluation of steroid and polyoxazole derivatives for G-quadruplex binding**, João Medeiros-Silva, Sara Cegonho, Elisabete Alves, Filipa Ramilho-Gomes, Sandrina Maçãs, Graça Baltazar, Eurico J. Cabrita, Samuel Silvestre and Carla Cruz, *International Symposium on Synthesis and Catalysis*, Évora, Portugal, 2015.
- **Phenantroline-based polyoxazoids as potent G-quadruplex ligands**, João Medeiros-Silva, Eurico J. Cabrita and Carla Cruz; *UBIQuímica Chemistry and Biochemistry Symposium*, Covilhã, Portugal, 2015



## IV. Bibliography

- (1) Dahm, R. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum. Genet.* **2008**, *122*, 565–81.
- (2) Flemming, W. Zur Kenntnis der Zelle in ihrer Teilung-Erscheinungen. *Schriften Naturwiss. Vereins Schl.-Holsk* **1878**, 23–27.
- (3) Paweletz, N. Walther Flemming: pioneer of mitosis research. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 72–5.
- (4) Avery, O. T. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* **1944**, *79*, 137–158.
- (5) Levene, P. A. The structure of yeast nucleic acid. Iv. Ammonia hydrolysis. *J. Biol. Chem.* **1919**, *40*, 415–424.
- (6) Kresge, N.; Simoni, R. D.; Hill, R. L. Chargaff's Rules: the Work of Erwin Chargaff. *J. Biol. Chem.* **2005**, *280*, e21–.
- (7) Pray, L. Discovery of DNA structure and function: Watson and Crick. *Nat. Educ.* **2008**, *1*, 100.
- (8) Franklin, R. E.; Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **1953**, *171*, 740–741.
- (9) Wilkins, M. H. F.; Strokes, A. R.; Wilson, H. R. Molecular structure of deoxypentose nucleic acids. 1953. *Nature* **2003**, *421*, 398–400; discussion 396.
- (10) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738.
- (11) Rhuland, L. E.; Bannister, B. The biosynthesis of  $\alpha$ ,[unk]-diaminopimelic acid. i. isolation of an intermediate, active for a diaminopimelic acid-requiring E. Coli Mutant. *J. Am. Chem. Soc.* **1956**, *78*, 3548–3548.
- (12) Holley, R. W.; Apgar, J.; Everett, G. A.; Madison, J. T.; Marquisee, M.; Merrill, S. H.; Penswick, J. R.; Zamir, A. Structure of a ribonucleic acid. *Science* **1965**, *147*, 1462–5.

- (13) Schweet, R.; Heintz, R. Protein Synthesis. *Annu. Rev. Biochem.* **1966**, *35*, 723–758.
- (14) Clancy, S. Chemical structure of RNA. *Nat. Educ.* **2008**, *1*, 60.
- (15) Rich, A.; Zhang, S. Timeline: Z-DNA: the long road to biological function. *Nat. Rev. Genet.* **2003**, *4*, 566–72.
- (16) Hoogsteen, K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **1963**, *16*, 907–916.
- (17) Nikolova, E. N.; Zhou, H.; Gottardo, F. L.; Alvey, H. S.; Kimsey, I. J.; Al-Hashimi, H. M. A historical account of Hoogsteen base-pairs in duplex DNA. *Biopolymers* **2013**, *99*, 955–68.
- (18) Courtois, Y.; Fromageot, P.; Guschlbauer, W. Protonated polynucleotide structures. 3. An optical rotatory dispersion study of the protonation of DNA. *Eur. J. Biochem.* **1968**, *6*, 493–501.
- (19) Crick, F. H. C. Codon–anticodon pairing: The wobble hypothesis. *J. Mol. Biol.* **1966**, *19*, 548–555.
- (20) Varani, G.; McClain, W. H. The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* **2000**, *1*, 18–23.
- (21) Rich, A.; Zhang, S. Z-DNA: the long road to biological function. *Nat. Rev. Genet.* **2003**, *4*, 566–572.
- (22) Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **2011**, *470*, 498–502.
- (23) Frank-Kamenetskii, M. D.; Mirkin, S. M. Triplex DNA structures. *Annu. Rev. Biochem.* **1995**, *64*, 65–95.
- (24) Rich, A. The double helix: a tale of two puckers. *Nat. Struct. Biol.* **2003**, *10*, 247–249.
- (25) Radhakrishnan, I.; Patel, D. J. Solution structure of a purine.purine.pyrimidine DNA triplex containing G.GC and T.AT triples. *Structure* **1993**, *1*, 135–52.
- (26) Weil, J.; Min, T.; Yang, C.; Wang, S.; Sutherland, C.; Sinha, N.; Kang, C. Stabilization of the i-motif by intramolecular adenine-adenine-thymine base triple in the structure of d(ACCCT). *Acta Crystallogr. D. Biol. Crystallogr.* **1999**, *55*, 422–9.
- (27) Holliday, R. A mechanism for gene conversion in fungi. *Genet. Res.* **2009**, *5*, 282.
- (28) Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **1989**, *59*, 871–80.

- (29) Davis, J. T. G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 668–98.
- (30) Limongelli, V.; De Tito, S.; Cerofolini, L.; Fragai, M.; Pagano, B.; Trotta, R.; Cosconati, S.; Marinelli, L.; Novellino, E.; Bertini, I.; Randazzo, A.; Luchinat, C.; Parrinello, M. The G-triplex DNA. *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 2269–73.
- (31) Chaput, J. C.; Switzer, C. A DNA pentaplex incorporating nucleobase quintets. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 10614–9.
- (32) Lipay, J. M.; Mihailescu, M.-R. NMR spectroscopy and kinetic studies of the quadruplex forming RNA r(UGGAGGU). *Mol. Biosyst.* **2009**, *5*, 1347–55.
- (33) Matsugami, A.; Okuizumi, T.; Uesugi, S.; Katahira, M. Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA. *J. Biol. Chem.* **2003**, *278*, 28147–53.
- (34) Lieberman, O. J.; DeStefano, J. J.; Lee, V. T. Detection of cyclic diguanylate G-octaplex assembly and interaction with proteins. *PLoS One* **2013**, *8*, e53689.
- (35) Smith, F. W.; Lau, F. W.; Feigon, J. d(G3T4G3) forms an asymmetric diagonally looped dimeric quadruplex with guanosine 5'-syn-syn-anti and 5'-syn-anti-anti N-glycosidic conformations. *Proc. Natl. Acad. Sci.* **1994**, *91*, 10546–10550.
- (36) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2006**, *34*, 5402–15.
- (37) Wong, A.; Wu, G. Selective binding of monovalent cations to the stacking G-quartet structure formed by guanosine 5'-monophosphate: a solid-state NMR study. *J. Am. Chem. Soc.* **2003**, *125*, 13895–905.
- (38) Venczel, E. A.; Sen, D. Parallel and antiparallel G-DNA structures from a complex telomeric sequence. *Biochemistry* **1993**, *32*, 6220–6228.
- (39) Sket, P.; Plavec, J. Not all G-quadruplexes exhibit ion-channel-like properties: NMR study of ammonium ion (non)movement within the d(G(3)T(4)G(4))(2) quadruplex. *J. Am. Chem. Soc.* **2007**, *129*, 8794–800.
- (40) Zhou, J.; Bourdoncle, A.; Rosu, F.; Gabelica, V.; Mergny, J.-L. Tri-G-quadruplex: controlled assembly of a G-quadruplex structure from three G-rich strands. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 11002–5.
- (41) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.* **2006**, *34*, 5402–5415.

- (42) Ou, T.; Lu, Y.; Tan, J.; Huang, Z.; Wong, K.; Gu, L. G-quadruplexes: targets in anticancer drug design. *ChemMedChem* **2008**, *3*, 690–713.
- (43) Islam, B.; D’Atri, V.; Sgobba, M.; Husby, J.; Haider, S. *Guanine Quartets*; Spindler, L.; Fritzsche, W., Eds.; Royal Society of Chemistry: Cambridge, 2012.
- (44) Tippana, R.; Xiao, W.; Myong, S. G-quadruplex conformation and dynamics are determined by loop length and sequence. *Nucleic Acids Res.* **2014**, *42*, 8106–14.
- (45) Guédin, A.; Gros, J.; Alberti, P.; Mergny, J.-L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.* **2010**, *38*, 7858–68.
- (46) Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **2008**, *36*, 5482–515.
- (47) Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S. Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.* **2004**, *126*, 16405–15.
- (48) Rachwal, P. A.; Brown, T.; Fox, K. R. Sequence effects of single base loops in intramolecular quadruplex DNA. *FEBS Lett.* **2007**, *581*, 1657–60.
- (49) Silva, M. W. da; Karsisiotis, A. I. *Guanine Quartets*; Spindler, L.; Fritzsche, W., Eds.; Royal Society of Chemistry: Cambridge, 2012.
- (50) Sponer, J.; Spacková, N. Molecular dynamics simulations and their application to four-stranded DNA. *Methods* **2007**, *43*, 278–90.
- (51) Giancola, C.; Pagano, B. Energetics of ligand binding to G-quadruplexes. *Quadruplex Nucleic Acids* **2013**, 211–242.
- (52) Giancola, C. *Guanine Quartets*; Spindler, L.; Fritzsche, W., Eds.; Royal Society of Chemistry: Cambridge, 2012.
- (53) De Cian, A.; Guittat, L.; Kaiser, M.; Saccà, B.; Amrane, S.; Bourdoncle, A.; Alberti, P.; Teulade-Fichou, M.-P.; Lacroix, L.; Mergny, J.-L. Fluorescence-based melting assays for studying quadruplex ligands. *Methods* **2007**, *42*, 183–95.
- (54) Bončina, M.; Hamon, F.; Islam, B.; Teulade-Fichou, M.-P.; Vesnaver, G.; Haider, S.; Lah, J. Dominant Driving Forces in Human Telomere Quadruplex Binding-Induced Structural Alterations. *Biophys. J.* **2015**, *108*, 2903–2911.
- (55) Mashimo, T.; Yagi, H.; Sannohe, Y.; Rajendran, A.; Sugiyama, H. Folding Pathways of Human Telomeric Type-1 and Type-2 G-Quadruplex Structures. *J. Am. Chem. Soc.* **2010**, *132*, 14910–14918.

- (56) Freyer, M. W.; Buscaglia, R.; Kaplan, K.; Cashman, D.; Hurley, L. H.; Lewis, E. A. Biophysical Studies of the c-MYC NHE III1 Promoter: Model Quadruplex Interactions with a Cationic Porphyrin. *Biophys. J.* **2007**, *92*, 2007–2015.
- (57) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K<sup>+</sup> solution. *Nucleic Acids Res.* **2007**, *35*, 6517–25.
- (58) Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K<sup>+</sup> solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.* **2007**, *35*, 4927–40.
- (59) Mergny, J.-L.; De Cian, A.; Ghelab, A.; Saccà, B.; Lacroix, L. Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res.* **2005**, *33*, 81–94.
- (60) Tran, P. L. T.; De Cian, A.; Gros, J.; Moriyama, R.; Mergny, J.-L. Tetramolecular quadruplex stability and assembly. *Top. Curr. Chem.* **2013**, *330*, 243–73.
- (61) Gray, R. D.; Chaires, J. B. Kinetics and mechanism of K<sup>+</sup>- and Na<sup>+</sup>-induced folding of models of human telomeric DNA into G-quadruplex structures. *Nucleic Acids Res.* **2008**, *36*, 4191–203.
- (62) Sun, D.; Hurley, L. H. The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J. Med. Chem.* **2009**, *52*, 2863–74.
- (63) Renciuik, D.; Kejnovská, I.; Skoláková, P.; Bednářová, K.; Motlová, J.; Vorlícková, M. Arrangements of human telomere DNA quadruplex in physiologically relevant K<sup>+</sup> solutions. *Nucleic Acids Res.* **2009**, *37*, 6625–34.
- (64) Petraccone, L.; Trent, J. O.; Chaires, J. B. The tail of the telomere. *J. Am. Chem. Soc.* **2008**, *130*, 16530–2.
- (65) Xu, Y.; Ishizuka, T.; Kurabayashi, K.; Komiyama, M. Consecutive formation of G-quadruplexes in human telomeric-overhang DNA: a protective capping structure for telomere ends. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 7833–6.
- (66) Petraccone, L.; Spink, C.; Trent, J. O.; Garbett, N. C.; Mekmaysy, C. S.; Giancola, C.; Chaires, J. B. Structure and Stability of Higher-Order Human Telomeric Quadruplexes. *J. Am. Chem. Soc.* **2011**, *133*, 20951–20961.
- (67) Li, W.; Wu, P.; Ohmichi, T.; Sugimoto, N. Characterization and thermodynamic properties of quadruplex/duplex competition. *FEBS Lett.* **2002**, *526*, 77–81.
- (68) Phan, A. T.; Mergny, J.-L. Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res.* **2002**, *30*, 4618–25.

- (69) Risitano, A.; Fox, K. R. Stability of Intramolecular DNA Quadruplexes: Comparison with DNA Duplexes †. *Biochemistry* **2003**, *42*, 6507–6513.
- (70) Miyoshi, D.; Fujimoto, T.; Sugimoto, N. Molecular crowding and hydration regulating of G-quadruplex formation. *Top. Curr. Chem.* **2013**, *330*, 87–110.
- (71) Granotier, C.; Pennarun, G.; Riou, L.; Hoffschir, F.; Gauthier, L. R.; De Cian, A.; Gomez, D.; Mandine, E.; Riou, J.-F.; Mergny, J.-L.; Mailliet, P.; Dutrillaux, B.; Boussin, F. D. Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res.* **2005**, *33*, 4182–90.
- (72) Paeschke, K.; Juranek, S.; Rhodes, D.; Lipps, H. J. Cell cycle-dependent regulation of telomere tethering in the nucleus. *Chromosome Res.* **2008**, *16*, 721–8.
- (73) Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.* **2005**, *12*, 847–54.
- (74) Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Plückthun, A. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 8572–7.
- (75) Lipps, H. J.; Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.* **2009**, *19*, 414–22.
- (76) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **2013**, *5*, 182–6.
- (77) Henderson, A.; Wu, Y.; Huang, Y. C.; Chavez, E. A.; Platt, J.; Johnson, F. B.; Brosh, R. M.; Sen, D.; Lansdorp, P. M. Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.* **2014**, *42*, 860–9.
- (78) Salgado, G. F.; Cazenave, C.; Kerkour, A.; Mergny, J.-L. G-quadruplex DNA and ligand interaction in living cells using NMR spectroscopy. *Chem. Sci.* **2015**, *6*, 3314–3320.
- (79) Huppert, J. L.; Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **2007**, *35*, 406–13.
- (80) Todd, A. K.; Johnston, M.; Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **2005**, *33*, 2901–7.
- (81) Neidle, S. The Structures of Human Telomeric DNA Quadruplexes. In *Therapeutic Applications of Quadruplex Nucleic Acids*; Elsevier, 2012; pp. 43–66.
- (82) Patel, D. J.; Phan, A. T.; Kuryavyi, V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.* **2007**, *35*, 7429–55.



- (83) Wang, Y.; Patel, D. J. Solution structure of the human telomeric repeat d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] G-tetraplex. *Structure* **1993**, *1*, 263–82.
- (84) Xu, Y.; Noguchi, Y.; Sugiyama, H. The new models of the human telomere d[AGGG(TTAGGG)<sub>3</sub>] in K<sup>+</sup> solution. *Bioorg. Med. Chem.* **2006**, *14*, 5584–91.
- (85) Phan, A. T.; Luu, K. N.; Patel, D. J. Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K<sup>+</sup> solution. *Nucleic Acids Res.* **2006**, *34*, 5715–9.
- (86) Heddi, B.; Phan, A. T. Structure of human telomeric DNA in crowded solution. *J. Am. Chem. Soc.* **2011**, *133*, 9824–33.
- (87) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **2002**, *417*, 876–80.
- (88) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. Structure of the human telomere in K<sup>+</sup> solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.* **2009**, *131*, 4301–9.
- (89) Zhang, Z.; Dai, J.; Veliath, E.; Jones, R. A.; Yang, D. Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in K<sup>+</sup> solution: insights into the interconversion of human telomeric G-quadruplex structures. *Nucleic Acids Res.* **2010**, *38*, 1009–21.
- (90) Lim, K. W.; Alberti, P.; Guédin, A.; Lacroix, L.; Riou, J.-F.; Royle, N. J.; Mergny, J.-L.; Phan, A. T. Sequence variant (CTAGGG)<sub>n</sub> in the human telomere favors a G-quadruplex structure containing a G.C.G.C tetrad. *Nucleic Acids Res.* **2009**, *37*, 6239–48.
- (91) Allshire, R. C.; Dempster, M.; Hastie, N. D. Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucleic Acids Res.* **1989**, *17*, 4611–27.
- (92) Hänsel, R.; Löhr, F.; Foldynová-Trantírková, S.; Bamberg, E.; Trantírek, L.; Dötsch, V. The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Res.* **2011**, *39*, 5768–75.
- (93) Víglaský, V.; Bauer, L.; Tluczková, K. Structural features of intra- and intermolecular G-quadruplexes derived from telomeric repeats. *Biochemistry* **2010**, *49*, 2110–20.
- (94) Wang, Y.; Patel, D. J. Guanine residues in d(T<sub>2</sub>AG<sub>3</sub>) and d(T<sub>2</sub>G<sub>4</sub>) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. *Biochemistry* **1992**, *31*, 8112–8119.
- (95) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **2002**, *417*, 876–80.

- (96) Zhang, N.; Phan, A. T.; Patel, D. J. (3 + 1) Assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex. *J. Am. Chem. Soc.* **2005**, *127*, 17277–85.
- (97) Phan, A. T. Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* **2010**, *277*, 1107–17.
- (98) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11593–8.
- (99) Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M.; Parkinson, G. N.; Todd, A. K.; Ladame, S.; Balasubramanian, S.; Neidle, S. Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.* **2005**, *127*, 10584–9.
- (100) Dexheimer, T. S.; Sun, D.; Hurley, L. H. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.* **2006**, *128*, 5404–15.
- (101) Xu, Y.; Sugiyama, H. Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res.* **2006**, *34*, 949–54.
- (102) Sun, D.; Guo, K.; Rusche, J. J.; Hurley, L. H. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.* **2005**, *33*, 6070–80.
- (103) Cogoi, S.; Xodo, L. E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* **2006**, *34*, 2536–49.
- (104) Membrino, A.; Cogoi, S.; Pedersen, E. B.; Xodo, L. E. G4-DNA formation in the HRAS promoter and rational design of decoy oligonucleotides for cancer therapy. *PLoS One* **2011**, *6*, e24421.
- (105) De Armond, R.; Wood, S.; Sun, D.; Hurley, L. H.; Ebbinghaus, S. W. Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1 $\alpha$  promoter. *Biochemistry* **2005**, *44*, 16341–50.
- (106) Basundra, R.; Kumar, A.; Amrane, S.; Verma, A.; Phan, A. T.; Chowdhury, S. A novel G-quadruplex motif modulates promoter activity of human thymidine kinase 1. *FEBS J.* **2010**, *277*, 4254–64.
- (107) Phan, A. T.; Modi, Y. S.; Patel, D. J. Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.* **2004**, *126*, 8710–6.
- (108) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* **2005**, *44*, 2048–58.

- (109) Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol.* **2005**, *1*, 167–73.
- (110) Phan, A. T.; Kuryavyi, V.; Burge, S.; Neidle, S.; Patel, D. J. Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc.* **2007**, *129*, 4386–92.
- (111) Dai, J.; Chen, D.; Jones, R. A.; Hurley, L. H.; Yang, D. NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.* **2006**, *34*, 5133–44.
- (112) Bang, I. Untersuchungen über die Guanylsäure. *Biochemische Zeitschrift* **1910**, 293–311.
- (113) GELLERT, M.; LIPSETT, M. N.; DAVIES, D. R. Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48*, 2013–8.
- (114) Eddy, J.; Maizels, N. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.* **2008**, *36*, 1321–33.
- (115) Maizels, N. Biological Functions of G-quadruplexes. In *Guanine Quartets*; Royal Society of Chemistry: Cambridge; pp. 215–222.
- (116) Moon, I. K. The human telomere and its relationship to human disease, therapy, and tissue engineering. *Front. Biosci.* **2007**, *12*, 4595.
- (117) Wright, W. E.; Tesmer, V. M.; Huffman, K. E.; Levene, S. D.; Shay, J. W. Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev.* **1997**, *11*, 2801–9.
- (118) Geserick, C.; Blasco, M. A. Novel roles for telomerase in aging. *Mech. Ageing Dev.* **2006**, *127*, 579–83.
- (119) Hayflick, L.; Moorhead, P. S. The serial cultivation of human diploid cell strains. *Exp. Cell Res.* **1961**, *25*, 585–621.
- (120) Zhang, X.; Mar, V.; Zhou, W.; Harrington, L.; Robinson, M. O. Telomere shortening and apoptosis in telomerase-inhibited human tumor cells. *Genes Dev.* **1999**, *13*, 2388–99.
- (121) Nakamura, T. M.; Morin, G. B.; Chapman, K. B.; Weinrich, S. L.; Andrews, W. H.; Lingner, J.; Harley, C. B.; Cech, T. R. Telomerase catalytic subunit homologs from fission yeast and human. *Science* **1997**, *277*, 955–9.
- (122) Feng, J.; Funk, W. D.; Wang, S. S.; Weinrich, S. L.; Avilion, A. A.; Chiu, C. P.; Adams, R. R.; Chang, E.; Allsopp, R. C.; Yu, J. The RNA component of human telomerase. *Science* **1995**, *269*, 1236–41.

- (123) Shay, J. W.; Bacchetti, S. A survey of telomerase activity in human cancer. *Eur. J. Cancer* **1997**, *33*, 787–91.
- (124) Shay, J. W.; Wright, W. E. Telomeres and telomerase in normal and cancer stem cells. *FEBS Lett.* **2010**, *584*, 3819–25.
- (125) Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W. Specific association of human telomerase activity with immortal cells and cancer. *Science* **1994**, *266*, 2011–5.
- (126) Hanahan, D.; Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **2011**, *144*, 646–74.
- (127) Hanahan, D.; Weinberg, R. A. The hallmarks of cancer. *Cell* **2000**, *100*, 57–70.
- (128) Herbert, B.; Pitts, A. E.; Baker, S. I.; Hamilton, S. E.; Wright, W. E.; Shay, J. W.; Corey, D. R. Inhibition of human telomerase in immortal human cells leads to progressive telomere shortening and cell death. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 14276–81.
- (129) Kosciolk, B. A.; Kalantidis, K.; Tabler, M.; Rowley, P. T. Inhibition of telomerase activity in human cancer cells by RNA interference. *Mol. Cancer Ther.* **2003**, *2*, 209–16.
- (130) Hahn, W. C.; Stewart, S. A.; Brooks, M. W.; York, S. G.; Eaton, E.; Kurachi, A.; Beijersbergen, R. L.; Knoll, J. H.; Meyerson, M.; Weinberg, R. A. Inhibition of telomerase limits the growth of human cancer cells. *Nat. Med.* **1999**, *5*, 1164–70.
- (131) Wang, Q.; Liu, J.; Chen, Z.; Zheng, K.; Chen, C.; Hao, Y.-H.; Tan, Z. G-quadruplex formation at the 3' end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic Acids Res.* **2011**, *39*, 6229–37.
- (132) Sun, D.; Thompson, B.; Cathers, B. E.; Salazar, M.; Kerwin, S. M.; Trent, J. O.; Jenkins, T. C.; Neidle, S.; Hurley, L. H. Inhibition of human telomerase by a G-quadruplex-interactive compound. *J. Med. Chem.* **1997**, *40*, 2113–6.
- (133) Salvati, E.; Scarsella, M.; Porru, M.; Rizzo, A.; Iachettini, S.; Tentori, L.; Graziani, G.; D'Incalci, M.; Stevens, M. F. G.; Orlandi, A.; Passeri, D.; Gilson, E.; Zupi, G.; Leonetti, C.; Biroccio, A. PARP1 is activated at telomeres upon G4 stabilization: possible target for telomere-based therapy. *Oncogene* **2010**, *29*, 6280–93.
- (134) Pennarun, G.; Granotier, C.; Hoffschir, F.; Mandine, E.; Biard, D.; Gauthier, L. R.; Boussin, F. D. Role of ATM in the telomere response to the G-quadruplex ligand 360A. *Nucleic Acids Res.* **2008**, *36*, 1741–54.
- (135) Rizzo, A.; Salvati, E.; Porru, M.; D'Angelo, C.; Stevens, M. F.; D'Incalci, M.; Leonetti, C.; Gilson, E.; Zupi, G.; Biroccio, A. Stabilization of quadruplex DNA perturbs telomere replication leading to the activation of an ATR-dependent ATM signaling pathway. *Nucleic Acids Res.* **2009**, *37*, 5353–64.

- (136) Neidle, S. The Biology and Pharmacology of Telomeric Quadruplex Ligands. In *Therapeutic Applications of Quadruplex Nucleic Acids*; Elsevier, 2012; pp. 109–117.
- (137) Neidle, S. Introduction: Quadruplexes and their Biology. In *Therapeutic Applications of Quadruplex Nucleic Acids*; Elsevier, 2012; pp. 1–20.
- (138) Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.* **2011**, *10*, 261–75.
- (139) Neidle, S. Genomic Quadruplexes as Therapeutic Targets. In *Therapeutic Applications of Quadruplex Nucleic Acids*; Elsevier, 2012; pp. 119–138.
- (140) Dang, C. V Enigmatic MYC Conducts an Unfolding Systems Biology Symphony. *Genes Cancer* **2010**, *1*, 526–531.
- (141) Drygin, D.; Siddiqui-Jain, A.; O'Brien, S.; Schwaebe, M.; Lin, A.; Bliesath, J.; Ho, C. B.; Proffitt, C.; Trent, K.; Whitten, J. P.; Lim, J. K. C.; Von Hoff, D.; Anderes, K.; Rice, W. G. Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Res.* **2009**, *69*, 7653–61.
- (142) Storck, S.; Shukla, M.; Dimitrov, S.; Bouvet, P. Functions of the histone chaperone nucleolin in diseases. *Subcell. Biochem.* **2007**, *41*, 125–44.
- (143) Parkinson, G. N. *The Reality of Quadruplex Nucleic Acids as a Therapeutic Targets*; Spindler, L.; Fritzsche, W., Eds.; Royal Society of Chemistry: Cambridge, 2012.
- (144) Martino, L.; Pagano, B.; Fotticchia, I.; Neidle, S.; Giancola, C. Shedding light on the interaction between TMPyP4 and human telomeric quadruplexes. *J. Phys. Chem. B* **2009**, *113*, 14779–86.
- (145) Izbicka, E.; Wheelhouse, R. T.; Raymond, E.; Davidson, K. K.; Lawrence, R. A.; Sun, D.; Windle, B. E.; Hurley, L. H.; Von Hoff, D. D. Effects of Cationic Porphyrins as G-Quadruplex Interactive Agents in Human Tumor Cells. *Cancer Res.* **1999**, *59*, 639–644.
- (146) Campbell, N. H.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. Structural basis of DNA quadruplex recognition by an acridine drug. *J. Am. Chem. Soc.* **2008**, *130*, 6722–4.
- (147) Haudecoeur, R.; Stefan, L.; Denat, F.; Monchaud, D. A model of smart G-quadruplex ligand. *J. Am. Chem. Soc.* **2013**, *135*, 550–3.
- (148) Clark, G. R.; Pytel, P. D.; Squire, C. J.; Neidle, S. Structure of the first parallel DNA quadruplex-drug complex. *J. Am. Chem. Soc.* **2003**, *125*, 4066–7.
- (149) Iida, K.; Majima, S.; Nakamura, T.; Seimiya, H.; Nagasawa, K. Evaluation of the interaction between long telomeric DNA and macrocyclic hexaoxazole (6OTD) dimer of a G-quadruplex ligand. *Molecules* **2013**, *18*, 4328–41.

- (150) Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **2012**, *8*, 301–10.
- (151) Lubitz, I.; Borovok, N.; Kotlyar, A. Interaction of monomolecular G4-DNA nanowires with TMPyP: evidence for intercalation. *Biochemistry* **2007**, *46*, 12925–9.
- (152) Haq, I.; Trent, J. O.; Chowdhry, B. Z.; Jenkins, T. C. Intercalative G-Tetraplex Stabilization of Telomeric DNA by a Cationic Porphyrin 1. *J. Am. Chem. Soc.* **1999**, *121*, 1768–1779.
- (153) Rodriguez, R.; Pantoş, G. D.; Gonçalves, D. P. N.; Sanders, J. K. M.; Balasubramanian, S. Ligand-driven G-quadruplex conformational switching by using an unusual mode of interaction. *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 5405–7.
- (154) Hamon, F.; Largy, E.; Guédin-Beaurepaire, A.; Rouchon-Dagois, M.; Sidibe, A.; Monchaud, D.; Mergny, J.-L.; Riou, J.-F.; Nguyen, C.-H.; Teulade-Fichou, M.-P. An acyclic oligoheteroaryle that discriminates strongly between diverse G-quadruplex topologies. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 8745–9.
- (155) Li, Q.; Xiang, J.; Li, X.; Chen, L.; Xu, X.; Tang, Y.; Zhou, Q.; Li, L.; Zhang, H.; Sun, H.; Guan, A.; Yang, Q.; Yang, S.; Xu, G. Stabilizing parallel G-quadruplex DNA by a new class of ligands: two non-planar alkaloids through interaction in lateral grooves. *Biochimie* **2009**, *91*, 811–9.
- (156) Hong, Y.; Xiong, H.; Lam, J. W. Y.; Häußler, M.; Liu, J.; Yu, Y.; Zhong, Y.; Sung, H. H. Y.; Williams, I. D.; Wong, K. S.; Tang, B. Z. Fluorescent bioprobes: structural matching in the docking processes of aggregation-induced emission fluorogens on DNA surfaces. *Chemistry* **2010**, *16*, 1232–45.
- (157) Rajput, C.; Rutkaite, R.; Swanson, L.; Haq, I.; Thomas, J. A. Dinuclear monointercalating RuII complexes that display high affinity binding to duplex and quadruplex DNA. *Chemistry* **2006**, *12*, 4611–9.
- (158) Martino, L.; Virno, A.; Pagano, B.; Virgilio, A.; Di Micco, S.; Galeone, A.; Giancola, C.; Bifulco, G.; Mayol, L.; Randazzo, A. Structural and thermodynamic studies of the interaction of distamycin A with the parallel quadruplex structure [d(TGGGGT)]<sub>4</sub>. *J. Am. Chem. Soc.* **2007**, *129*, 16048–56.
- (159) Campbell, N. H.; Patel, M.; Tofa, A. B.; Ghosh, R.; Parkinson, G. N.; Neidle, S. Selectivity in ligand recognition of G-quadruplex loops. *Biochemistry* **2009**, *48*, 1675–80.
- (160) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26.



- (161) Neidle, S. Design Principles for Quadruplex-binding Small Molecules. In *Therapeutic Applications of Quadruplex Nucleic Acids*; Elsevier, 2012; pp. 151–174.
- (162) Rudmann, D. G. On-target and off-target-based toxicologic effects. *Toxicol. Pathol.* **2013**, *41*, 310–4.
- (163) Bessi, I.; Bazzicalupi, C.; Richter, C.; Jonker, H. R. A.; Saxena, K.; Sissi, C.; Chioccioli, M.; Bianco, S.; Bilia, A. R.; Schwalbe, H.; Gratteri, P. Spectroscopic, Molecular Modeling, and NMR-Spectroscopic Investigation of the Binding Mode of the Natural Alkaloids Berberine and Sanguinarine to Human Telomeric G-Quadruplex DNA. *ACS Chem. Biol.* **2012**, *7*, 1109–1119.
- (164) Tan, J.-H.; Ou, T.-M.; Hou, J.-Q.; Lu, Y.-J.; Huang, S.-L.; Luo, H.-B.; Wu, J.-Y.; Huang, Z.-S.; Wong, K.-Y.; Gu, L.-Q. Isaindigotone Derivatives: A New Class of Highly Selective Ligands for Telomeric G-Quadruplex DNA. *J. Med. Chem.* **2009**, *52*, 2825–2835.
- (165) Shin-ya, K.; Wierzba, K.; Matsuo, K.; Ohtani, T.; Yamada, Y.; Furihata, K.; Hayakawa, Y.; Seto, H. Telomestatin, a Novel Telomerase Inhibitor from *Streptomyces anulatus*. *J. Am. Chem. Soc.* **2001**, *123*, 1262–1263.
- (166) Kim, M.-Y.; Vankayalapati, H.; Shin-ya, K.; Wierzba, K.; Hurley, L. H. Telomestatin, a Potent Telomerase Inhibitor That Interacts Quite Specifically with the Human Telomeric Intramolecular G-Quadruplex. *J. Am. Chem. Soc.* **2002**, *124*, 2098–2099.
- (167) Neidle, S. Telomeric Quadruplex Ligands II. In *Therapeutic Applications of Quadruplex Nucleic Acids*; Elsevier, 2012; pp. 93–107.
- (168) Chung, W. J.; Heddi, B.; Hamon, F.; Teulade-Fichou, M.-P.; Phan, A. T. Solution structure of a G-quadruplex bound to the bisquinolinium compound Phen-DC(3). *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 999–1002.
- (169) Li, Q.; Xiang, J.-F.; Yang, Q.-F.; Sun, H.-X.; Guan, A.-J.; Tang, Y.-L. G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res.* **2013**, *41*, D1115–D1123.
- (170) Tera, M.; Iida, K.; Ishizuka, H.; Takagi, M.; Suganuma, M.; Doi, T.; Shin-ya, K.; Nagasawa, K. Synthesis of a potent G-quadruplex-binding macrocyclic heptaoxazole. *Chembiochem* **2009**, *10*, 431–5.
- (171) Shalaby, T.; von Bueren, A. O.; Hürlimann, M.-L.; Fiaschetti, G.; Castelletti, D.; Masayuki, T.; Nagasawa, K.; Arcaro, A.; Jelesarov, I.; Shin-ya, K.; Grotzer, M. Disabling c-Myc in childhood medulloblastoma and atypical teratoid/rhabdoid tumor cells by the potent G-quadruplex interactive agent S2T1-6OTD. *Mol. Cancer Ther.* **2010**, *9*, 167–79.
- (172) Barbieri, C. M.; Srinivasan, A. R.; Rzuczek, S. G.; Rice, J. E.; LaVoie, E. J.; Pilch, D. S. Defining the mode, energetics and specificity with which a macrocyclic hexaoxazole binds to human telomeric G-quadruplex DNA. *Nucleic Acids Res.* **2007**, *35*, 3272–86.

- (173) Iida, K.; Tera, M.; Hirokawa, T.; Shin-ya, K.; Nagasawa, K. G-quadruplex recognition by macrocyclic hexaoxazole (6OTD) dimer: greater selectivity than monomer. *Chem. Commun. (Camb)*. **2009**, 6481–3.
- (174) Ohnmacht, S. A.; Micco, M.; Petrucci, V.; Todd, A. K.; Reszka, A. P.; Gunaratnam, M.; Carvalho, M. A.; Zloh, M.; Neidle, S. Sequences in the HSP90 promoter form G-quadruplex structures with selectivity for disubstituted phenyl bis-oxazole derivatives. *Bioorg. Med. Chem. Lett.* **2012**, 22, 5930–5.
- (175) Molinspiration [www.molinspiration.com](http://www.molinspiration.com) (accessed Sep 1, 2011).
- (176) Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, 43, 3714–3717.
- (177) Besselièvre, F.; Mahuteau-Betzer, F.; Grierson, D. S.; Piguel, S. Ligandless microwave-assisted Pd/Cu-catalyzed direct arylation of oxazoles. *J. Org. Chem.* **2008**, 73, 3278–80.
- (178) Goto, Y.; Tagawa, Y.; Yamashita, K.; Higuchi, Y. Improved Oxidation of Active Methyl Group of N-Heteroaromatic Compounds by Selenium Dioxide in the Presence of tert-Butyl Hydroperoxide. *Heterocycles* **2003**, 60, 953.
- (179) Umbreit, M. A.; Sharpless, K. B. Allylic oxidation of olefins by catalytic and stoichiometric selenium dioxide with tert-butyl hydroperoxide. *J. Am. Chem. Soc.* **1977**, 99, 5526–5528.
- (180) McNally, J. J.; Jackson, Y. A.; Downer-Riley, N. K. Selenium(IV) Oxide–tert-butyl Hydroperoxide.
- (181) Adam, R.; Ballesteros-Garrido, R.; Vallcorba, O.; Abarca, B.; Ballesteros, R.; Leroux, F. R.; Colobert, F.; Amigó, J. M.; Rius, J. Synthesis and structural properties of hexaaza[5]helicene containing two [1,2,3]triazolo[1,5-a]pyridine moieties. *Tetrahedron Lett.* **2013**, 54, 4316–4319.
- (182) M. Harwood, L.; M. Laventine, D.; Afsar, A.; J. Hudson, M. Tuning the Solubilities of Bis-triazinylphenanthroline Ligands (BTPHens) and Their Complexes. *Heterocycles* **2012**, 86, 1419.
- (183) Rabjohn, N. Selenium Dioxide Oxidation. In *Organic Reactions*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2011; pp. 331–386.
- (184) Achremowicz, L. A New Approach to the Oxidation of Methylquinolines with Selenium Dioxide. *Synth. Commun.* **2006**.
- (185) SAKAMOTO, T.; SAKASAI, T.; YAMANAKA, H. Studies on pyrimidine derivatives. XXII. Site-selective oxidation of di-methylpyrimidines with selenium dioxide to pyrimidine-monoaldehydes. *Chem. Pharm. Bull. (Tokyo)*. **1981**, 29, 2485–2490.



- (186) Chandler, C. J.; Deady, L. W.; Reiss, J. A. Synthesis of some 2,9-disubstituted-1,10-phenanthrolines. *J. Heterocycl. Chem.* **1981**, *18*, 599–601.
- (187) Leusen, D. Van; Leusen, A. M. Van Synthetic Uses of Tosylmethyl Isocyanide (TosMIC). In *Organic Reactions*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2004.
- (188) Sisko, J.; Kassick, A. J.; Mellinger, M.; Filan, J. J.; Allen, A.; Olsen, M. A. An Investigation of Imidazole and Oxazole Syntheses Using Aryl-Substituted TosMIC Reagents 1. *J. Org. Chem.* **2000**, *65*, 1516–1524.
- (189) Leusen, D. Van; Leusen, A. M. Van Synthetic Uses of Tosylmethyl Isocyanide (TosMIC). In *Organic Reactions*; Overman, L. E., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2004; pp. 420–432.
- (190) Verrier, C.; Lassalas, P.; Théveau, L.; Quéguiner, G.; Trécourt, F.; Marsais, F.; Hoarau, C. Recent advances in direct C–H arylation: Methodology, selectivity and mechanism in oxazole series. *Beilstein J. Org. Chem.* **2011**, *7*, 1584–1601.
- (191) Théveau, L.; Verrier, C.; Lassalas, P.; Martin, T.; Dupas, G.; Querolle, O.; Van Hijfte, L.; Marsais, F.; Hoarau, C. Mechanism Selection for Regiocontrol in Base-Assisted, Palladium-Catalysed Direct C<sup>o</sup>H Coupling with Halides: First Approach for Oxazole- and Thiazole-4-Carboxylates. *Chem. - A Eur. J.* **2011**, *17*, 14450–14463.
- (192) Mergny, J. L.; Maurizot, J. C. Fluorescence resonance energy transfer as a probe for G-quartet formation by a telomeric repeat. *Chembiochem* **2001**, *2*, 124–32.
- (193) De Rache, A.; Mergny, J.-L. Assessment of selectivity of G-quadruplex ligands via an optimised FRET melting assay. *Biochimie* **2015**, *115*, 194–202.
- (194) Monchaud, D.; Teulade-Fichou, M.-P. G4-FID: a fluorescent DNA probe displacement assay for rapid evaluation of quadruplex ligands. *Methods Mol. Biol.* **2010**, *608*, 257–71.
- (195) Largy, E.; Hamon, F.; Teulade-Fichou, M.-P. Development of a high-throughput G4-FID assay for screening and evaluation of small molecules binding quadruplex nucleic acid structures. *Anal. Bioanal. Chem.* **2011**, *400*, 3419–3427.
- (196) Valeur, B. *Molecule Detection in Solution . Methods and Applications*; 2001; Vol. 8.
- (197) Monchaud, D.; Allain, C.; Bertrand, H.; Smargiasso, N.; Rosu, F.; Gabelica, V.; De Cian, A.; Mergny, J.-L.; Teulade-Fichou, M.-P. Ligands playing musical chairs with G-quadruplex DNA: a rapid and simple displacement assay for identifying selective G-quadruplex binders. *Biochimie* **2008**, *90*, 1207–23.
- (198) Nygren, J.; Svanvik, N.; Kubista, M. The interactions between the fluorescent dye thiazole orange and DNA. *Biopolymers* **1998**, *46*, 39–51.

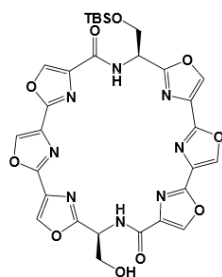
- (199) Tran, P. L. T.; Largy, E.; Hamon, F.; Teulade-Fichou, M.-P.; Mergny, J.-L. Fluorescence intercalator displacement assay for screening G4 ligands towards a variety of G-quadruplex structures. *Biochimie* **2011**, *93*, 1288–1296.
- (200) Vorlíčková, M.; Kejnovská, I.; Bednářová, K.; Renčiuk, D.; Kypr, J. Circular Dichroism Spectroscopy of DNA: From Duplexes to Quadruplexes. *Chirality* **2012**, *24*, 691–698.
- (201) Paramasivan, S.; Rujan, I.; Bolton, P. H. Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding. *Methods* **2007**, *43*, 324–331.
- (202) Karsisiotis, A. I.; Hessari, N. M.; Novellino, E.; Spada, G. P.; Randazzo, A.; Webba da Silva, M. Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 10645–8.
- (203) Masiero, S.; Trotta, R.; Pieraccini, S.; De Tito, S.; Perone, R.; Randazzo, A.; Spada, G. P. A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Org. Biomol. Chem.* **2010**, *8*, 2683.
- (204) Garbett, N. C.; Ragazzon, P. A.; Chaires, J. B. Circular dichroism to determine binding mode and affinity of ligand–DNA interactions. *Nat. Protoc.* **2007**, *2*, 3166–3172.
- (205) Sattin, G.; Artese, A.; Nadai, M.; Costa, G.; Parrotta, L.; Alcaro, S.; Palumbo, M.; Richter, S. N. Conformation and stability of intramolecular telomeric G-quadruplexes: sequence effects in the loops. *PLoS One* **2013**, *8*, e84113.
- (206) Yang, D.; Hurley, L. H. Structure of the Biologically Relevant G-Quadruplex in The c-MYC Promoter. *Nucleosides, Nucleotides and Nucleic Acids* **2006**, *25*, 951–968.
- (207) Le, H. T.; Miller, M. C.; Buscaglia, R.; Dean, W. L.; Holt, P. A.; Chaires, J. B.; Trent, J. O. Not all G-quadruplexes are created equally: an investigation of the structural polymorphism of the c-Myc G-quadruplex-forming sequence and its interaction with the porphyrin TMPyP4. *Org. Biomol. Chem.* **2012**, *10*, 9393.
- (208) Rachwal, P. a; Fox, K. R. Quadruplex melting. *Methods* **2007**, *43*, 291–301.
- (209) Mergny, J.-L.; Lacroix, L. Analysis of thermal melting curves. *Oligonucleotides* **2003**, *13*, 515–37.
- (210) Arora, A.; Balasubramanian, C.; Kumar, N.; Agrawal, S.; Ojha, R. P.; Maiti, S. Binding of berberine to human telomeric quadruplex - spectroscopic, calorimetric and molecular modeling studies. *FEBS J.* **2008**, *275*, 3971–3983.
- (211) Nicoludis, J. M.; Barrett, S. P.; Mergny, J.-L.; Yatsunyk, L. A. Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Res.* **2012**, *40*, 5432–5447.

- (212) Dash, S. P.; Panda, A. K.; Pasayat, S.; Dinda, R.; Biswas, A.; Tiekink, E. R. T.; Patil, Y. P.; Nethaji, M.; Kaminsky, W.; Mukhopadhyay, S.; Bhutia, S. K. Syntheses and structural investigation of some alkali metal ion-mediated LVVO<sub>2</sub><sup>-</sup> (L<sub>2</sub><sup>-</sup> = tridentate ONO ligands) species: DNA binding, photo-induced DNA cleavage and cytotoxic activities. *Dalt. Trans.* **2014**, *43*, 10139.
- (213) Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. A.; Lane, P.; Salemme, F. R. High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery. *J. Biomol. Screen.* **2001**, *6*, 429–440.
- (214) Halder, K.; Largy, E.; Benzler, M.; Teulade-Fichou, M.-P.; Hartig, J. S. Efficient Suppression of Gene Expression by Targeting 5'-UTR-Based RNA Quadruplexes with Bisquinolinium Compounds. *ChemBioChem* **2011**, *12*, 1663–1668.
- (215) Mathad, R. I.; Hatzakis, E.; Dai, J.; Yang, D. c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability. *Nucleic Acids Res.* **2011**, *39*, 9023–33.
- (216) Petraccone, L.; Pagano, B.; Giancola, C. Studying the effect of crowding and dehydration on DNA G-quadruplexes. *Methods* **2012**, *57*, 76–83.
- (217) Marchand, A.; Granzhan, A.; Iida, K.; Tsushima, Y.; Ma, Y.; Nagasawa, K.; Teulade-Fichou, M.-P.; Gabelica, V. Ligand-induced conformational changes with cation ejection upon binding to human telomeric DNA G-quadruplexes. *J. Am. Chem. Soc.* **2015**, *137*, 750–6.
- (218) Pagano, B.; Virno, A.; Mattia, C. A.; Mayol, L.; Randazzo, A.; Giancola, C. Targeting DNA quadruplexes with distamycin A and its derivatives: An ITC and NMR study. *Biochimie* **2008**, *90*, 1224–1232.
- (219) Levenson, A. S.; Jordan, V. C. MCF-7: The first hormone-responsive breast cancer cell line. *Cancer Res.* **1997**, *57*, 3071–3078.
- (220) Russell, P. J.; Kingsley, E. A. Human Prostate Cancer Cell Lines. In *Prostate Cancer Methods and Protocols*; Humana Press: New Jersey; pp. 21–40.
- (221) Clark, M. J.; Homer, N.; O'Connor, B. D.; Chen, Z.; Eskin, A.; Lee, H.; Merriman, B.; Nelson, S. F. U87MG Decoded: The Genomic Sequence of a Cytogenetically Aberrant Human Cancer Cell Line. *PLoS Genet.* **2010**, *6*, e1000832.
- (222) Mosmann, T. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J. Immunol. Methods* **1983**, *65*, 55–63.
- (223) Berridge, M. V.; Herst, P. M.; Tan, A. S. Tetrazolium dyes as tools in cell biology: New insights into their cellular reduction. In; 2005; pp. 127–152.

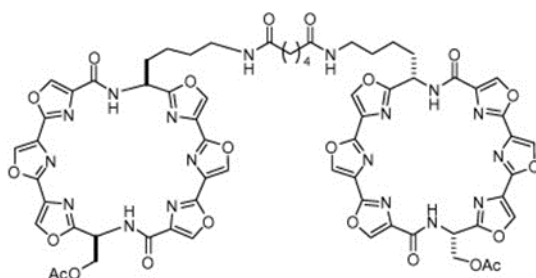
- (224) Selim, S.; Al Jaouni, S. Anticancer and apoptotic effects on cell proliferation of diosgenin isolated from *Costus speciosus* (Koen.) Sm. *BMC Complement. Altern. Med.* **2015**, *15*, 301.

# Appendix A

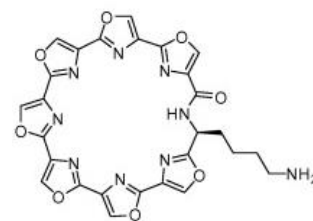
## Miscellaneous



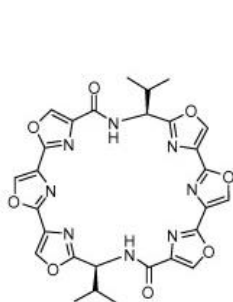
S2T1-6OTD



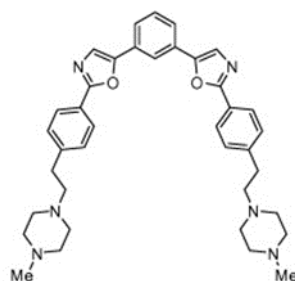
6OTD-dimer



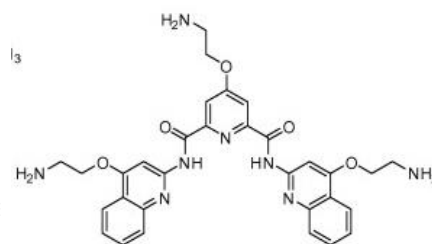
L1H1-7OTD



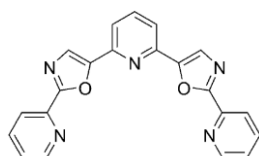
HXDV



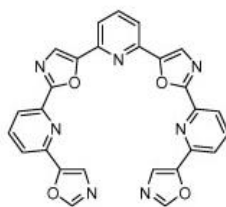
Bis-oxazole r



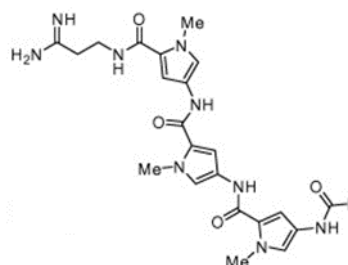
Pyridostatin



BOxaPy

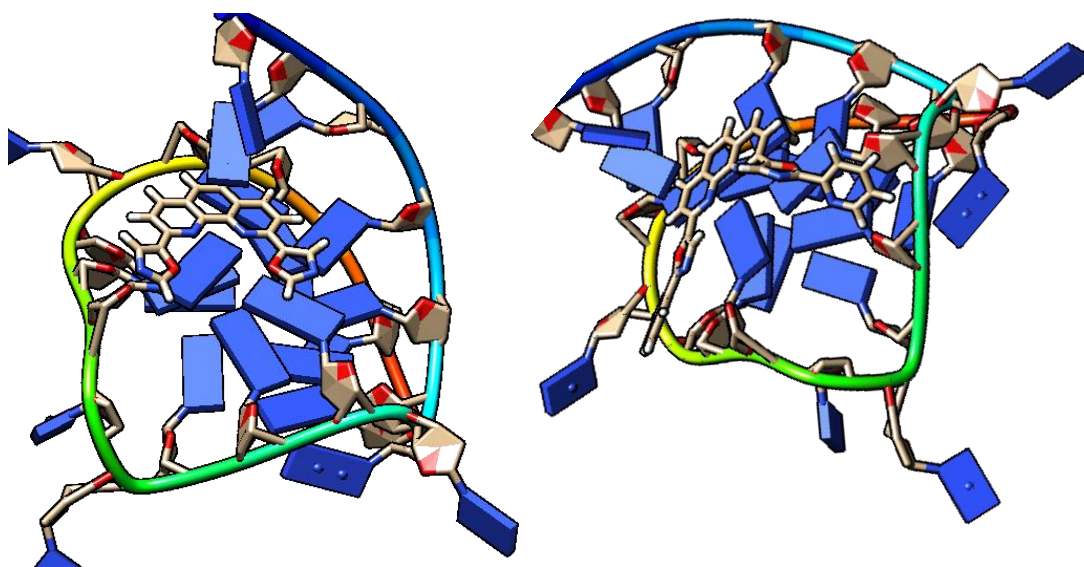


TOxaPy



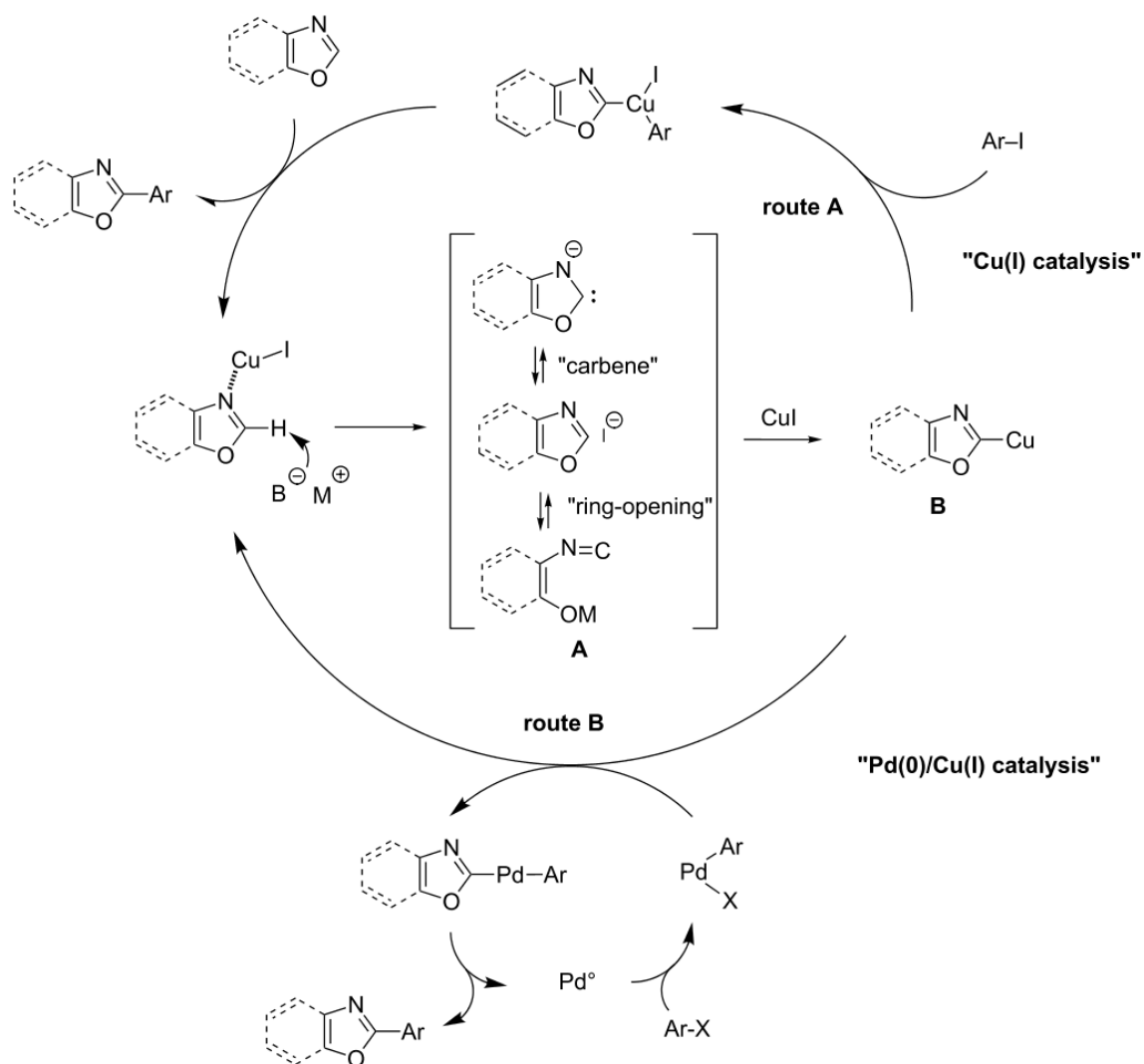
Distamycin A

**Supplementary A.1** – Chemical structure of relevant G4 ligands referred in the main text.

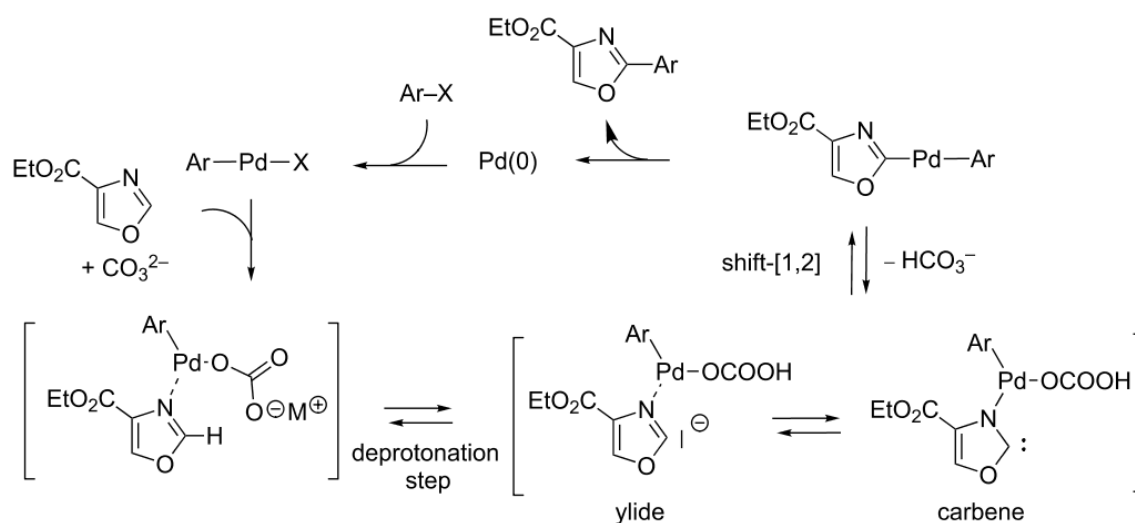


**Supplementary A.2** – Qualitative directed docking of Phen-1 (left) and Phen-2 (right) into a c-Myc G-quadruplex structure (PDB 2L7V). Phen-1 and Phen-2 show potential for interacting with the G-quadruplex.

## Proposed mechanism for CH activated coupling

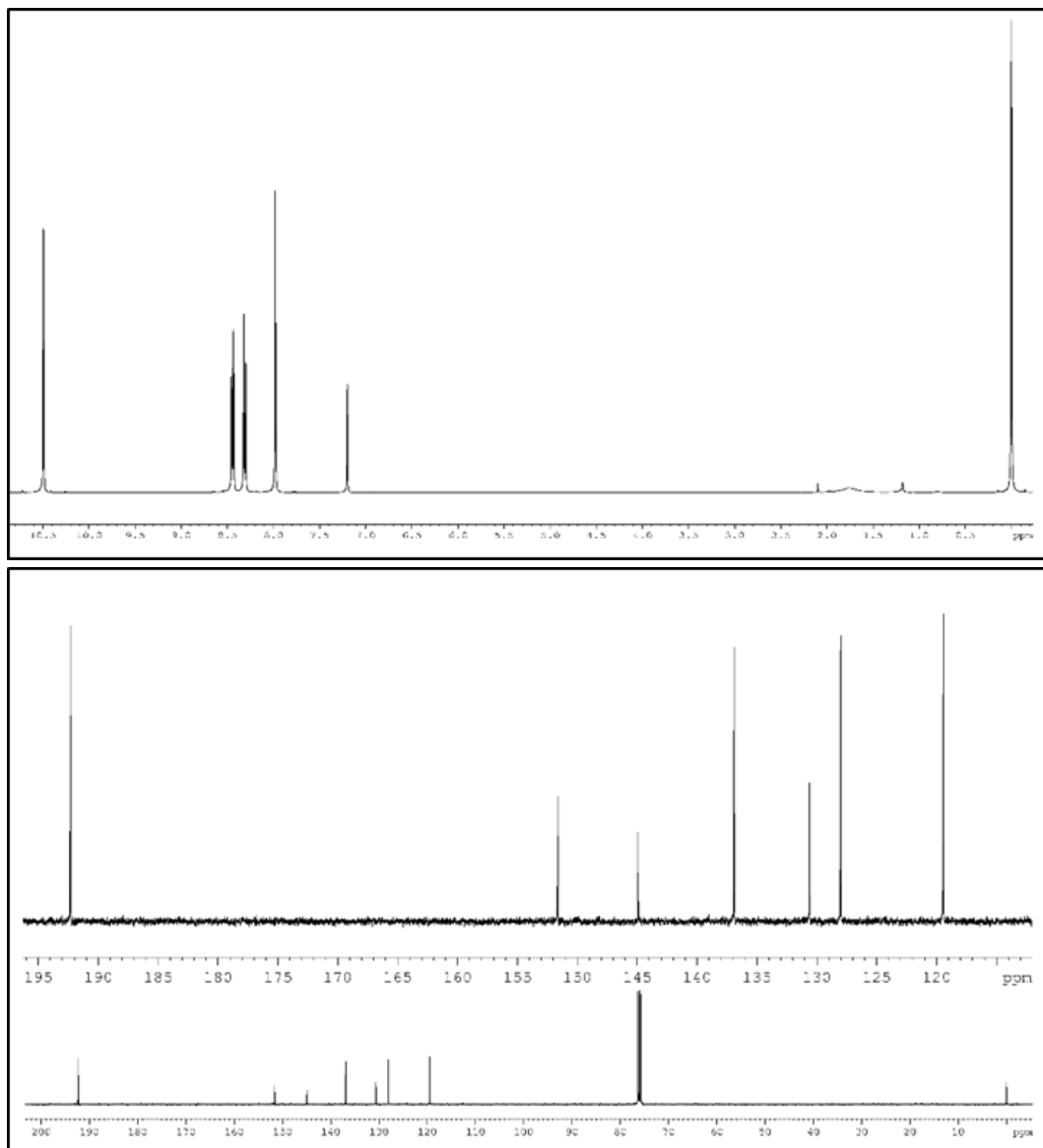
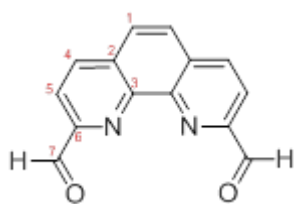


**Supplementary A.3** – Example of proposed catalytic cycles for copper catalyzed (route A) and Pd(0)/Cu(I) catalyzed (route B) direct arylation of oxazoles with halides. Adapted from reference<sup>29</sup>.



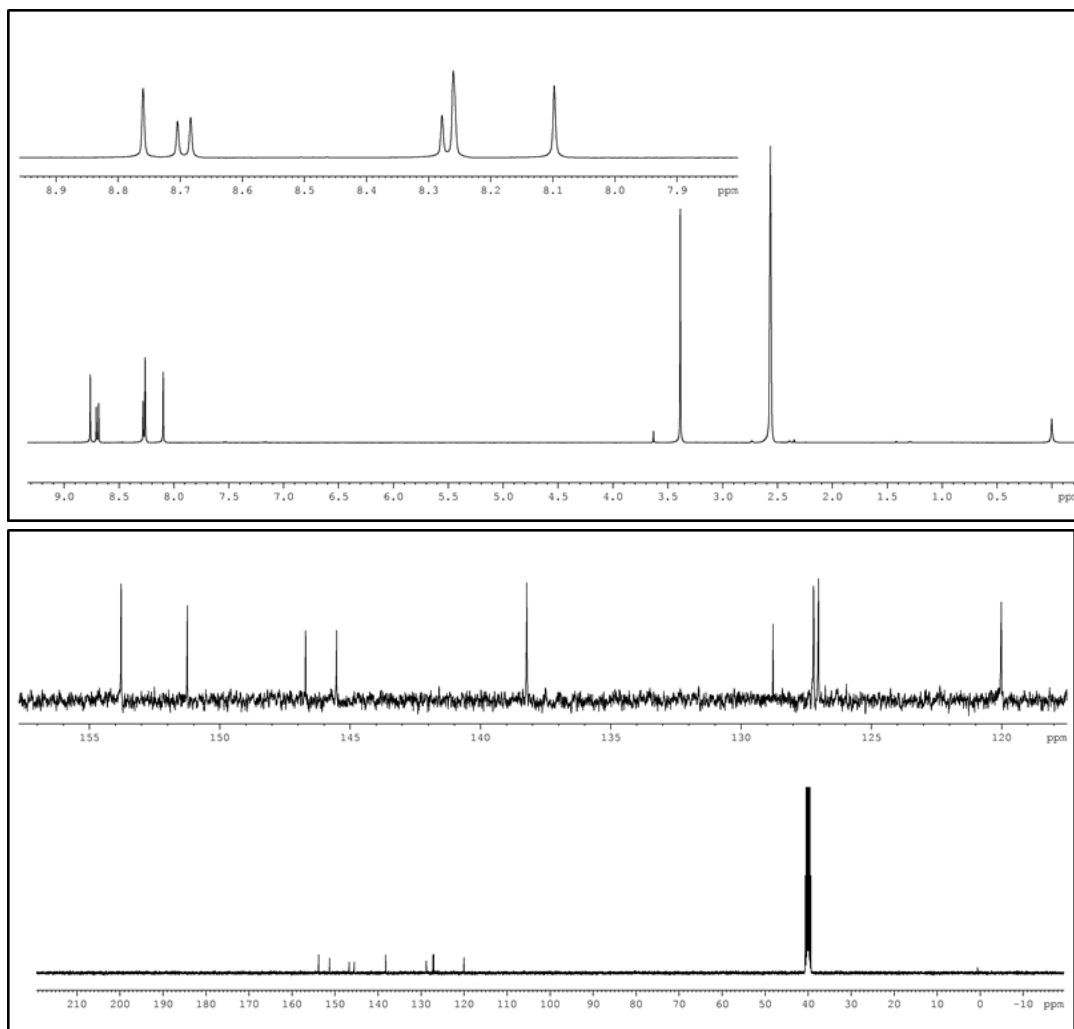
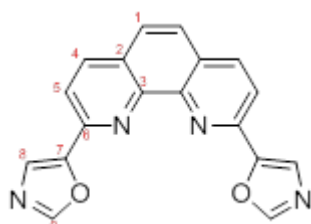
**Supplementary A.4** – Example of proposed catalytic cycles for Pd(0) catalyzed direct arylation of oxazoles with halides. Adapted from reference<sup>29</sup>.

## NMR spectra of the synthesized compounds

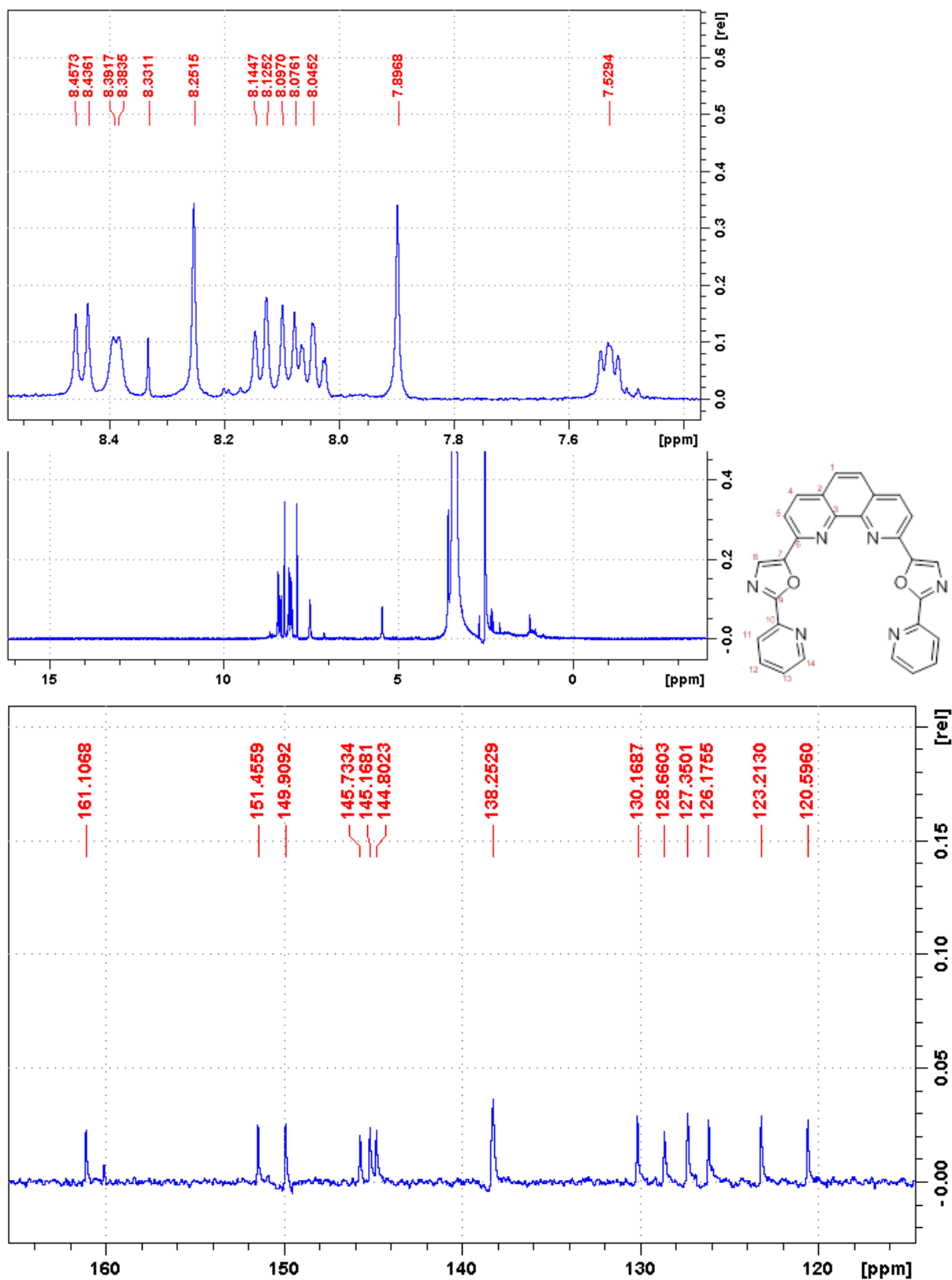


**Supplementary A.5** – NMR spectra of 1,10-phenanthroline-2,9-dicarbaldehyde (Phen-0) in CDCl<sub>3</sub> at 25°C. Top <sup>1</sup>H-NMR spectrum; bottom <sup>13</sup>C-NMR spectrum.

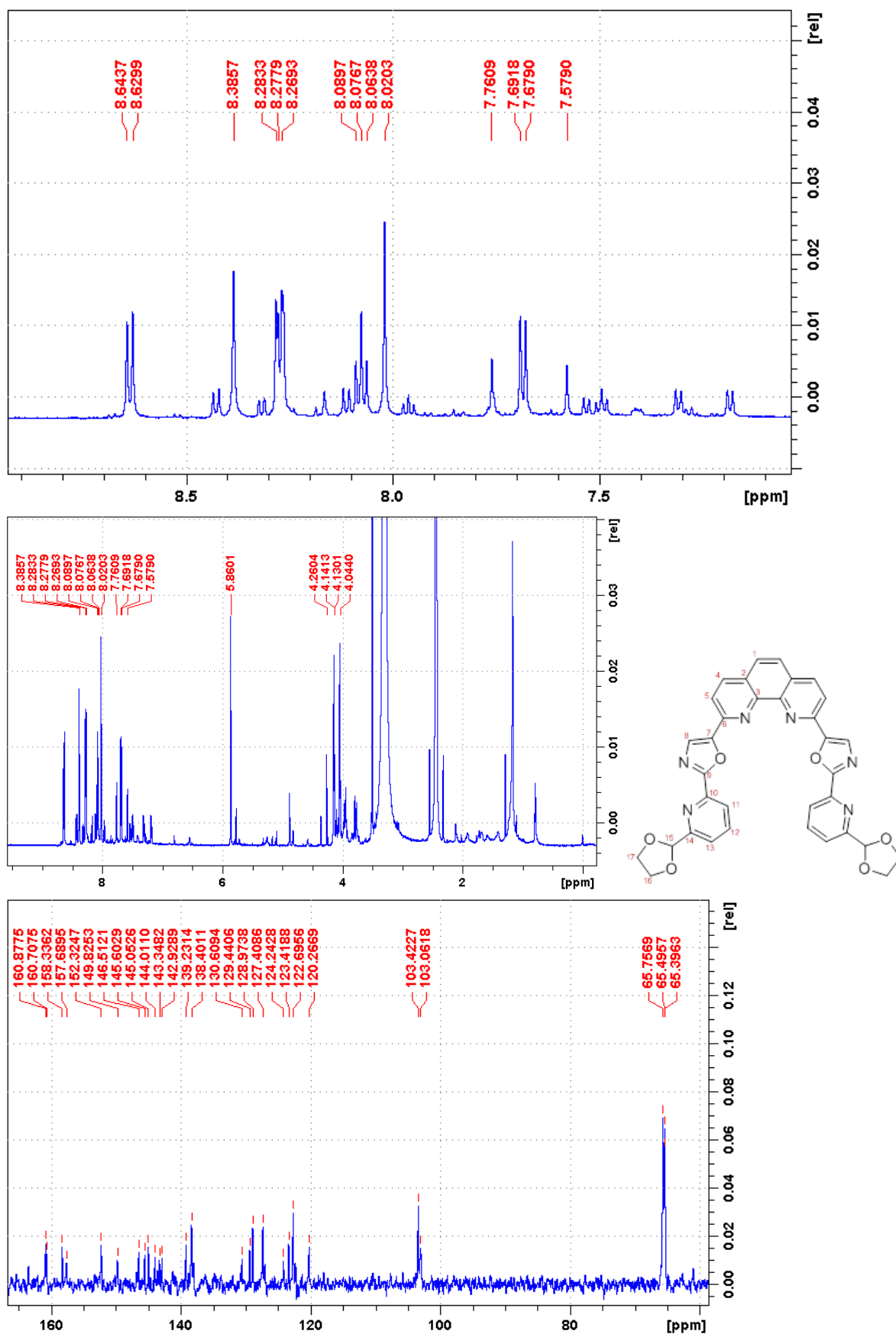




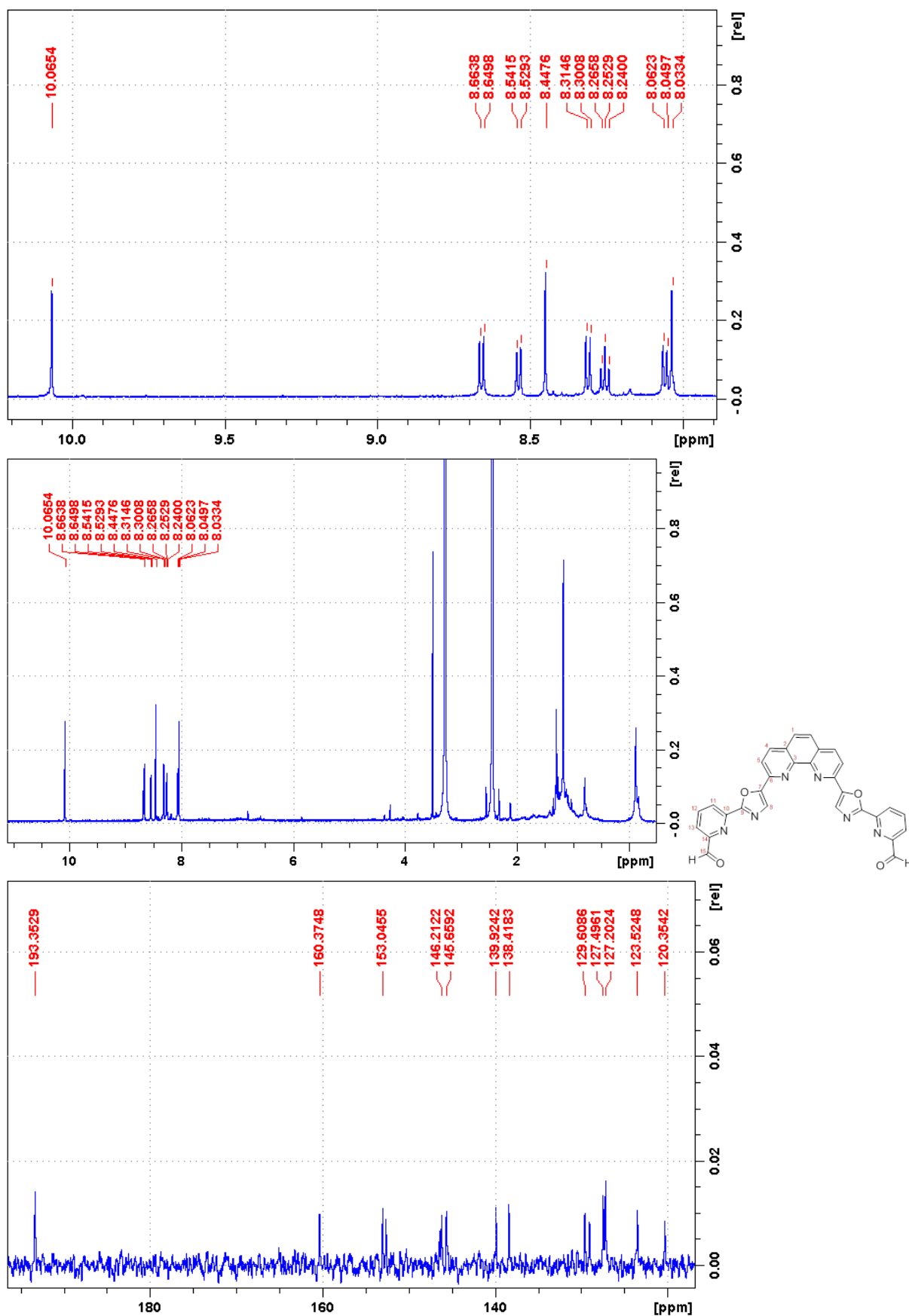
**Supplementary A.6** - NMR spectra of 2,9-bis(oxazole-5-yl)-1,10-phenanthroline (Phen-1) in DMSO-*d*<sub>6</sub> at 25°C. Top <sup>1</sup>H-NMR spectrum; bottom <sup>13</sup>C-NMR spectrum.



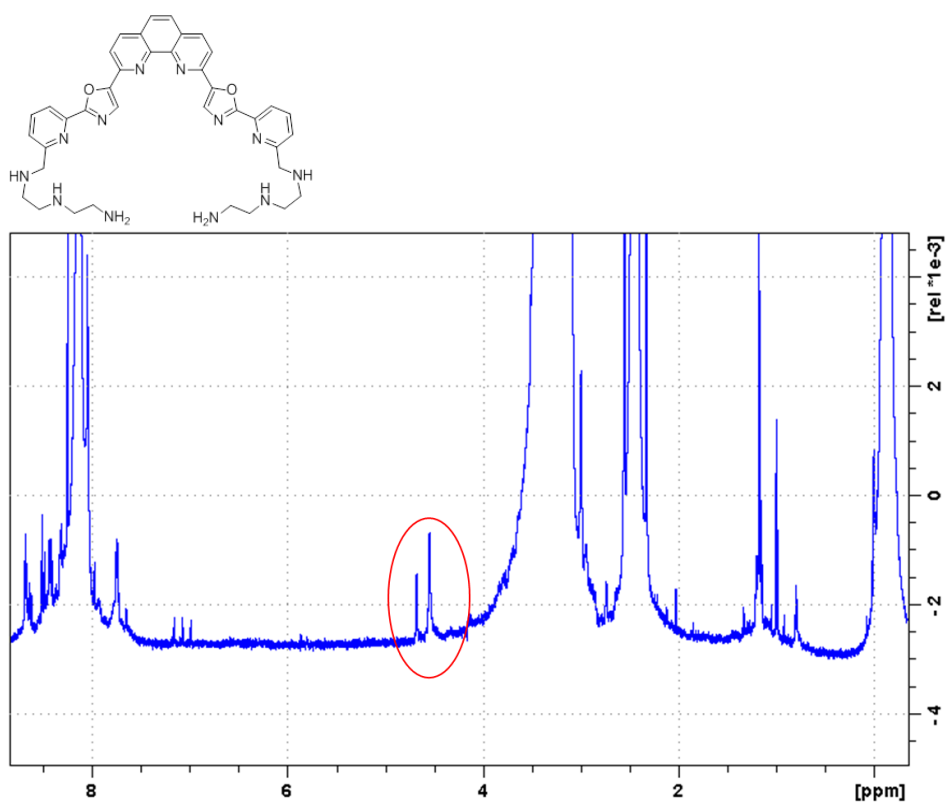
**Supplementary A.7** - NMR spectra of 2,9-bis(2-(pyridine-2-yl)oxazole-5-yl)-1,10-phenanthroline (Phen-2) in DMSO-*d*<sub>6</sub> at 25°C. Top <sup>1</sup>H-NMR spectrum; bottom <sup>13</sup>C-NMR spectrum.



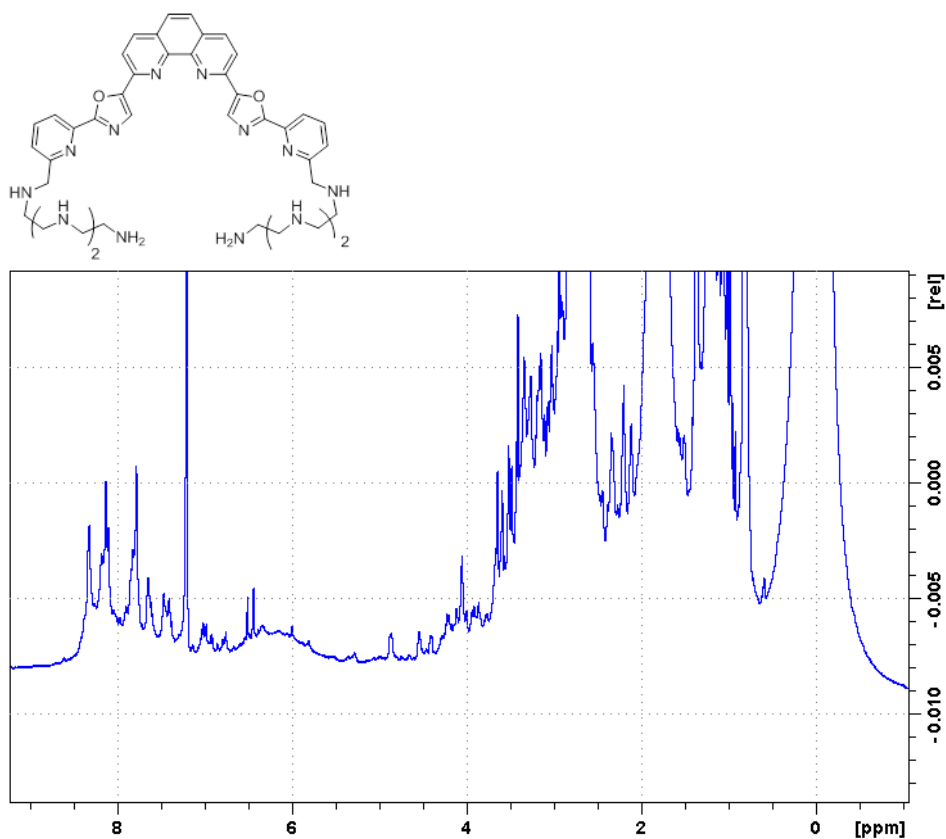
**Supplementary A.8** - NMR spectra of 2,9-bis(2-(6-(1,3-dioxolan-2-yl)pyridine-2-yl)oxazole-5-yl)-1,10-phenanthroline (Phen-3) in DMSO- $d_6$  at 25°C. Top  $^1\text{H}$ -NMR spectrum; bottom  $^{13}\text{C}$ -NMR spectrum.



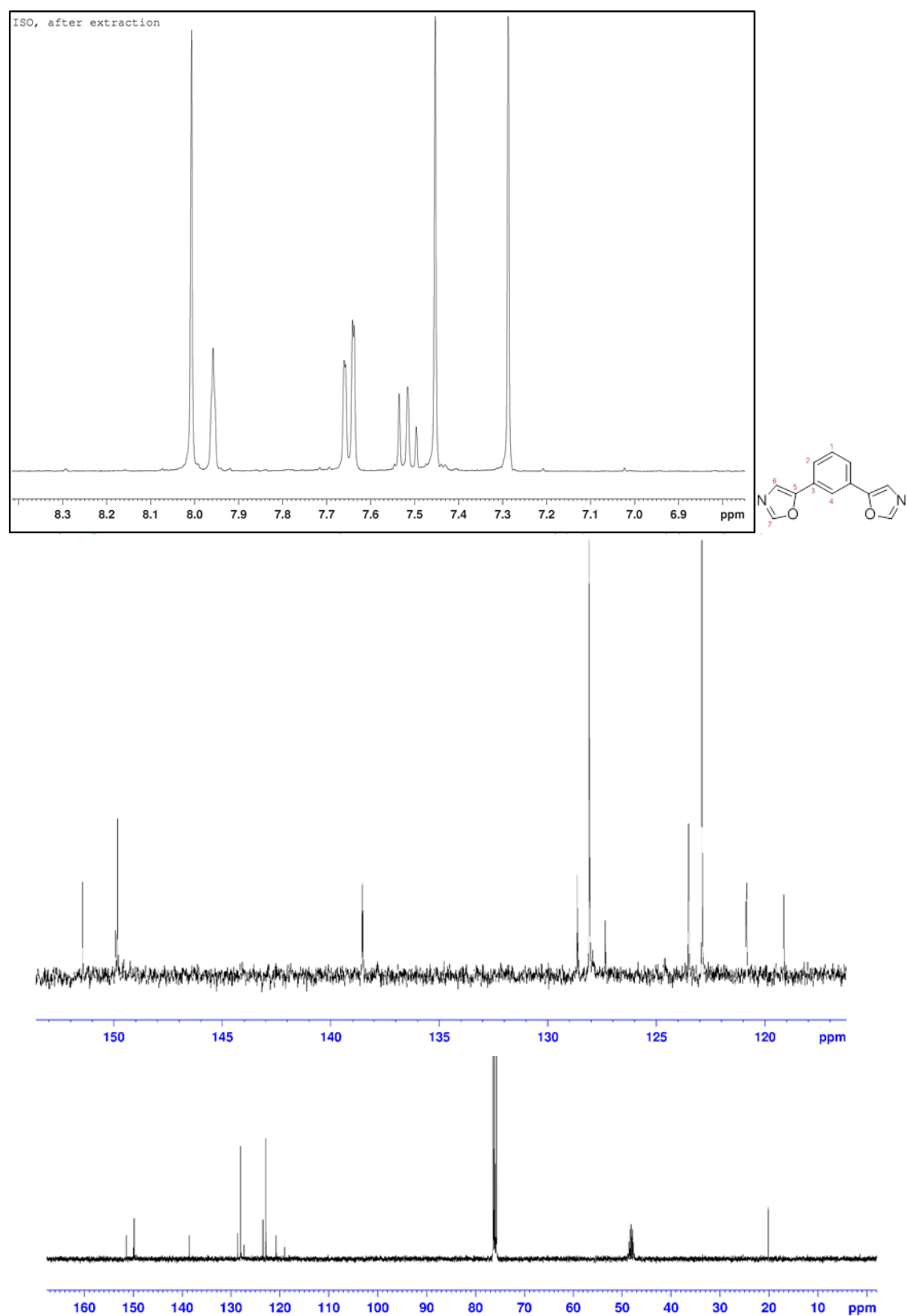
**Supplementary A.9** - NMR spectra of 6,6'-(5,5'-(1,10-phenanthroline-2,9-diyl)bis(oxazole-5,2-diyl)dipicolinaldehyde (Phen-4) in DMSO-*d*<sub>6</sub> at 25°C. Top <sup>1</sup>H-NMR spectrum; bottom <sup>13</sup>C-NMR spectrum.



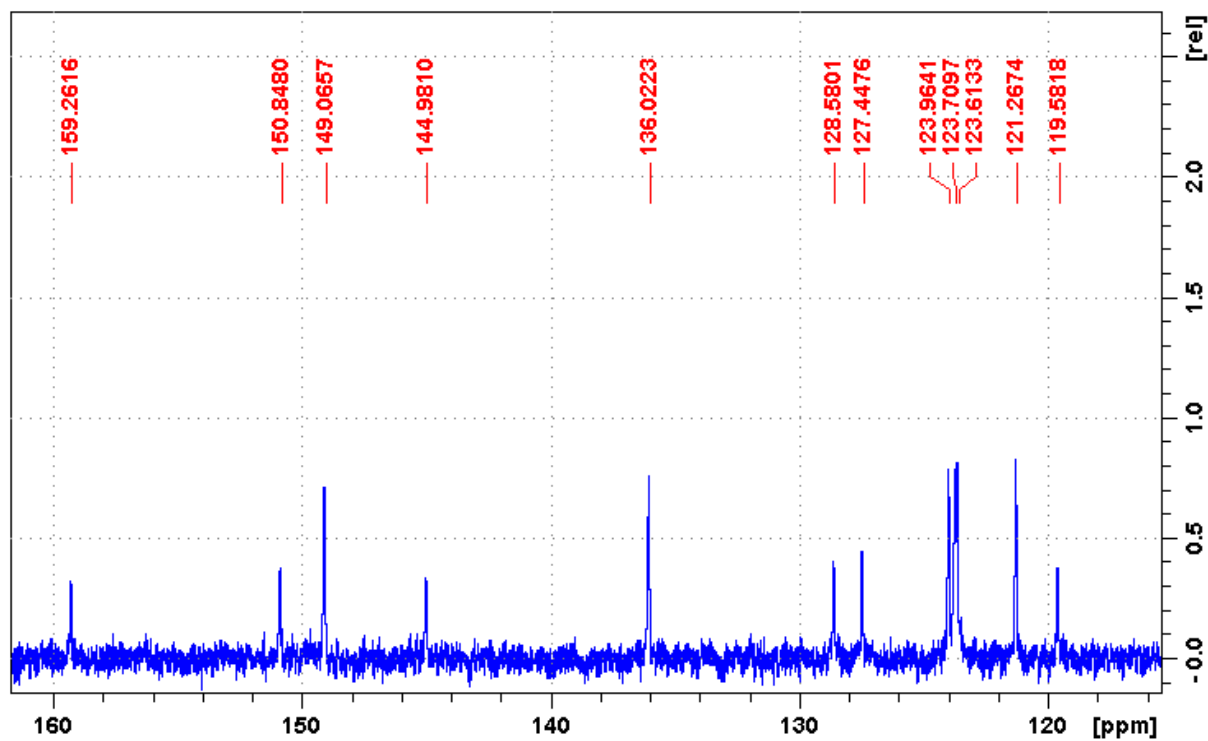
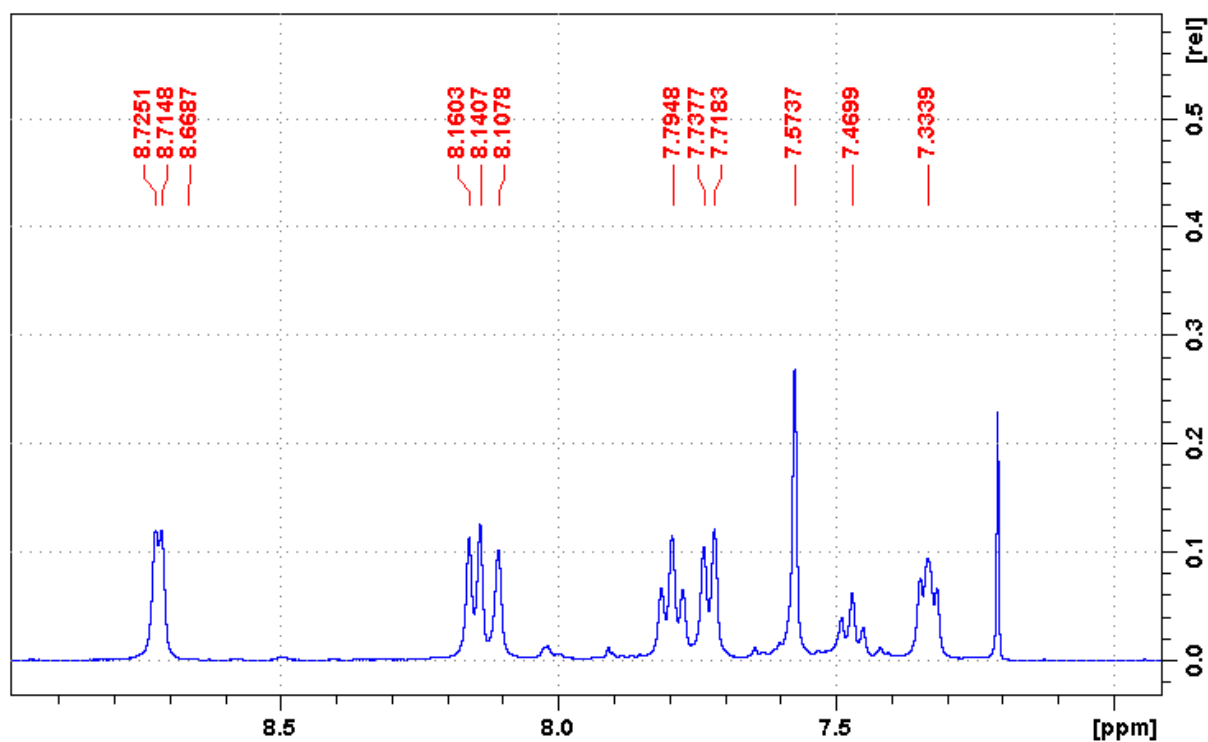
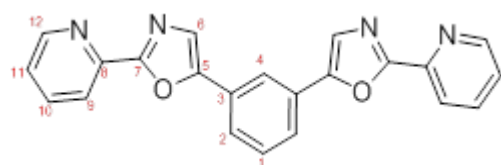
**Supplementary A.10** -  $^1\text{H-NMR}$  spectrum of impure  $\text{N1,N1}'\text{-}((6,6'\text{-}(5,5'\text{-}(1,10\text{-phenanthroline-2,9-diyl)bis(oxazole-5,2-diyl))bis(pyridine-6,2-diyl))bis(methylene))bis(\text{N2-(2-aminoethyl)ethane-1,2-diamine})$  in  $\text{DMSO-}d_6$  at  $25^\circ\text{C}$ . Top  $^1\text{H-NMR}$  spectrum; Circle highlights the typical chemical shift of  $\text{-C-NH-}$  bond.



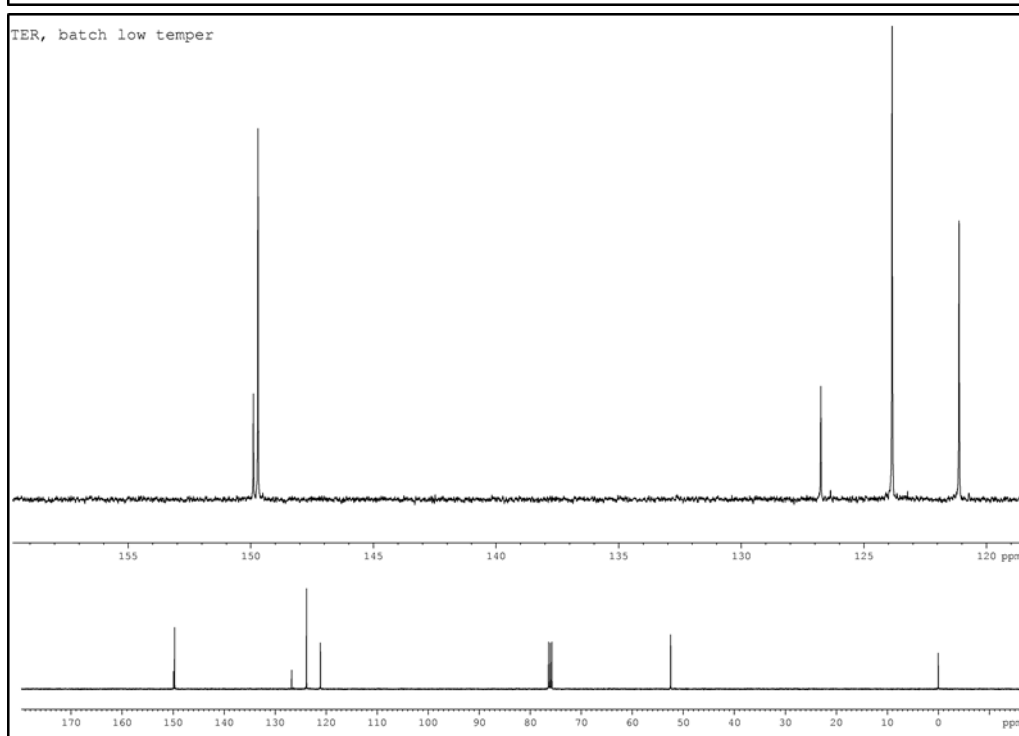
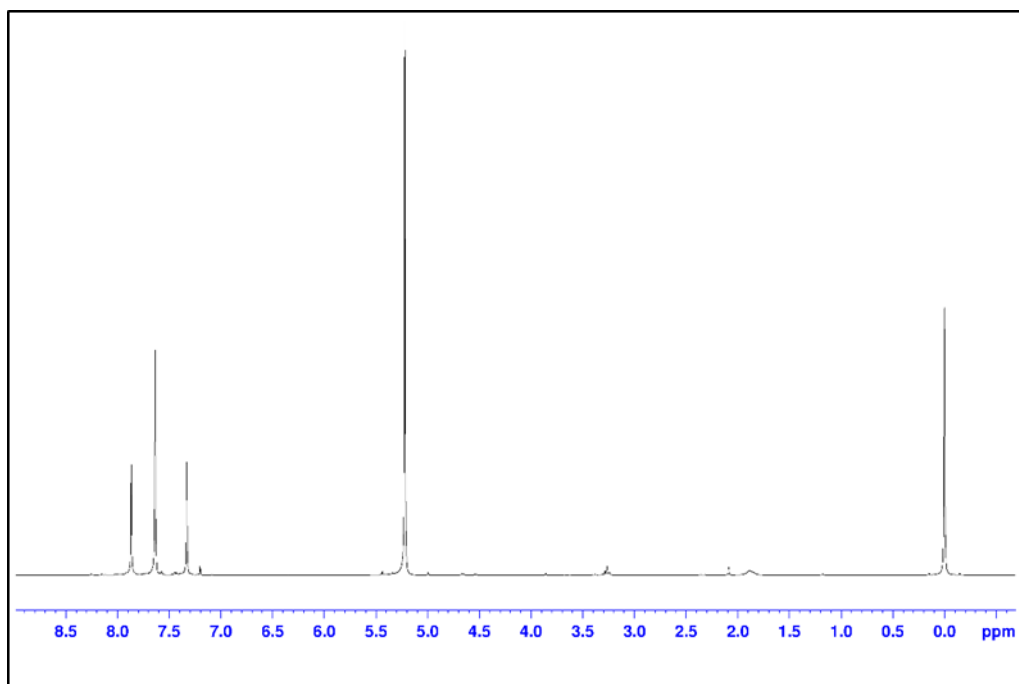
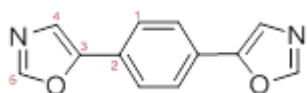
**Supplementary A.11** -  $^1\text{H-NMR}$  spectrum of impure  $\text{N1,N1}'\text{-}((6,6'\text{-}(5,5'\text{-}(1,10\text{-phenanthroline-2,9-diyl)bis(oxazole-5,2-diyl))bis(pyridine-6,2-diyl))bis(methylene))bis(\text{N2-(2-((2-aminoethyl)amino)ethyl)ethane-1,2-diamine})$  in  $\text{DMSO-}d_6$  at  $25^\circ\text{C}$ .



**Supplementary A.12** - NMR spectra of 1,3-bis(oxazol-5-yl)benzene (Iso-1) in  $\text{CDCl}_3$  at  $25^\circ\text{C}$ . Top  $^1\text{H-NMR}$  spectrum; bottom  $^{13}\text{C-NMR}$  spectrum.

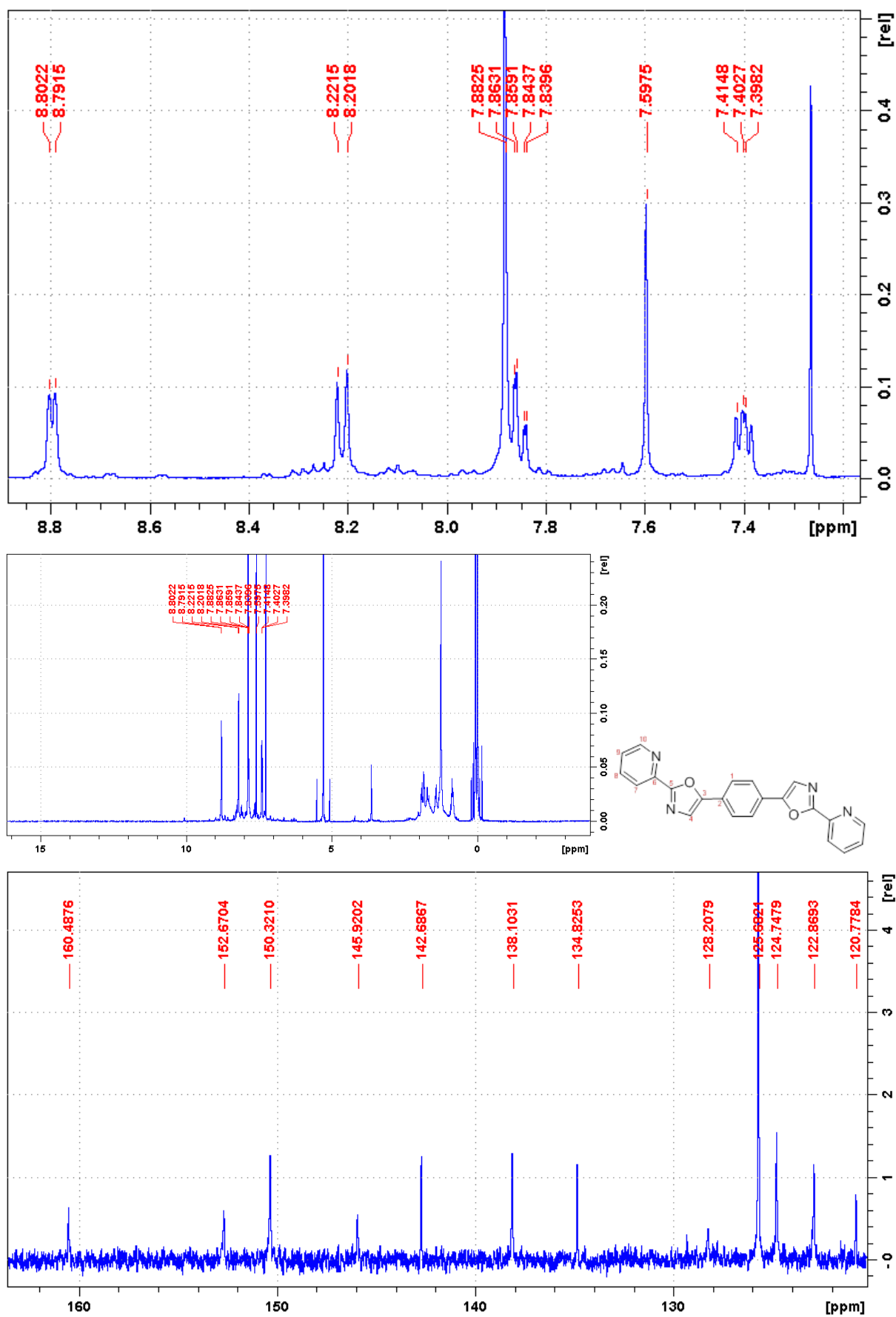


**Supplementary A.13** - NMR spectra of 1,3-bis(2-(pyridine-2-yl)oxazol-5-yl)benzene (Iso-2) in  $\text{CDCl}_3$  at  $25^\circ\text{C}$ . Top  $^1\text{H}$ -NMR spectrum; bottom  $^{13}\text{C}$ -NMR spectrum.

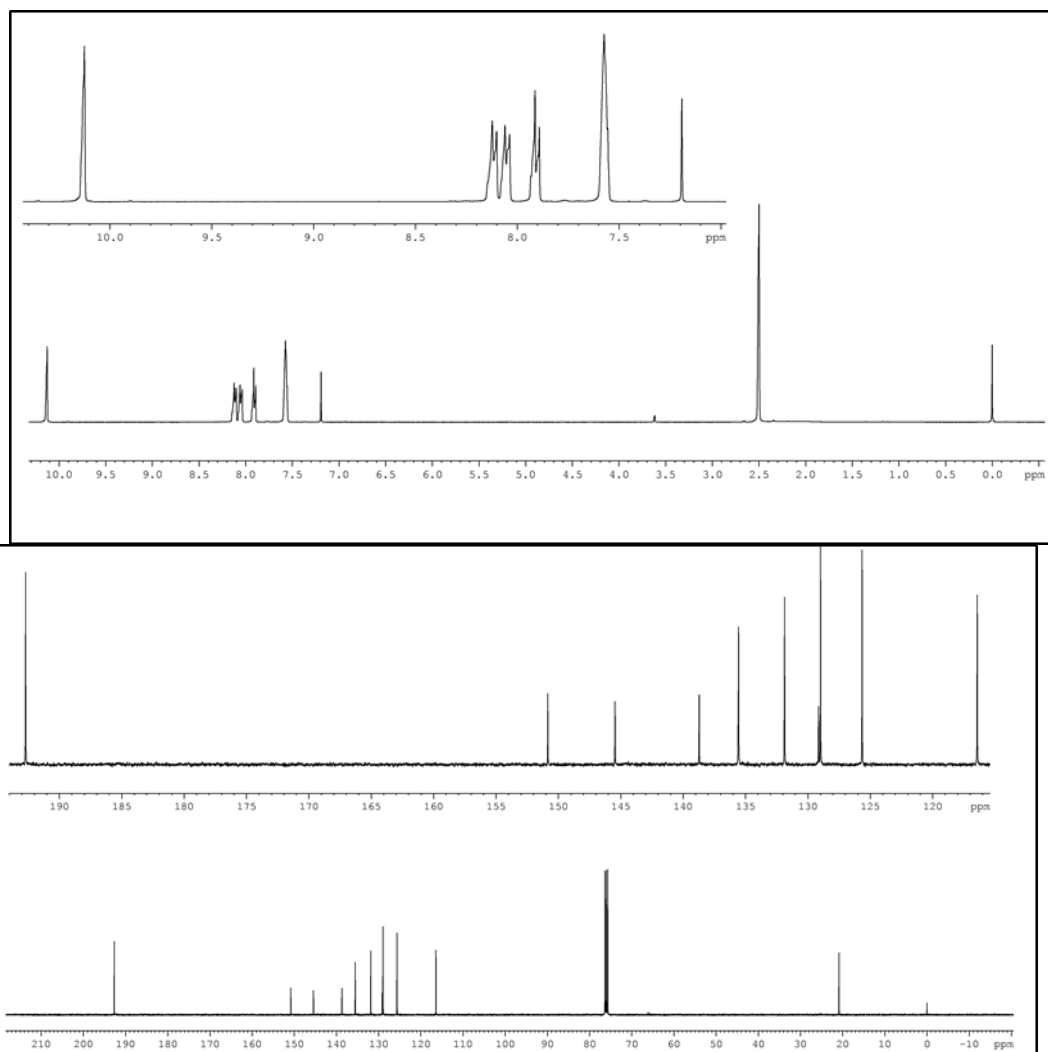
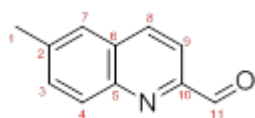


**Supplementary A.14-** NMR spectra 1,4-bis(oxazol-5-yl)benzene (Ter-1) in  $\text{CDCl}_3$  at  $25^\circ\text{C}$ . Top  $^1\text{H}$ -NMR spectrum; bottom  $^{13}\text{C}$ -NMR spectrum.

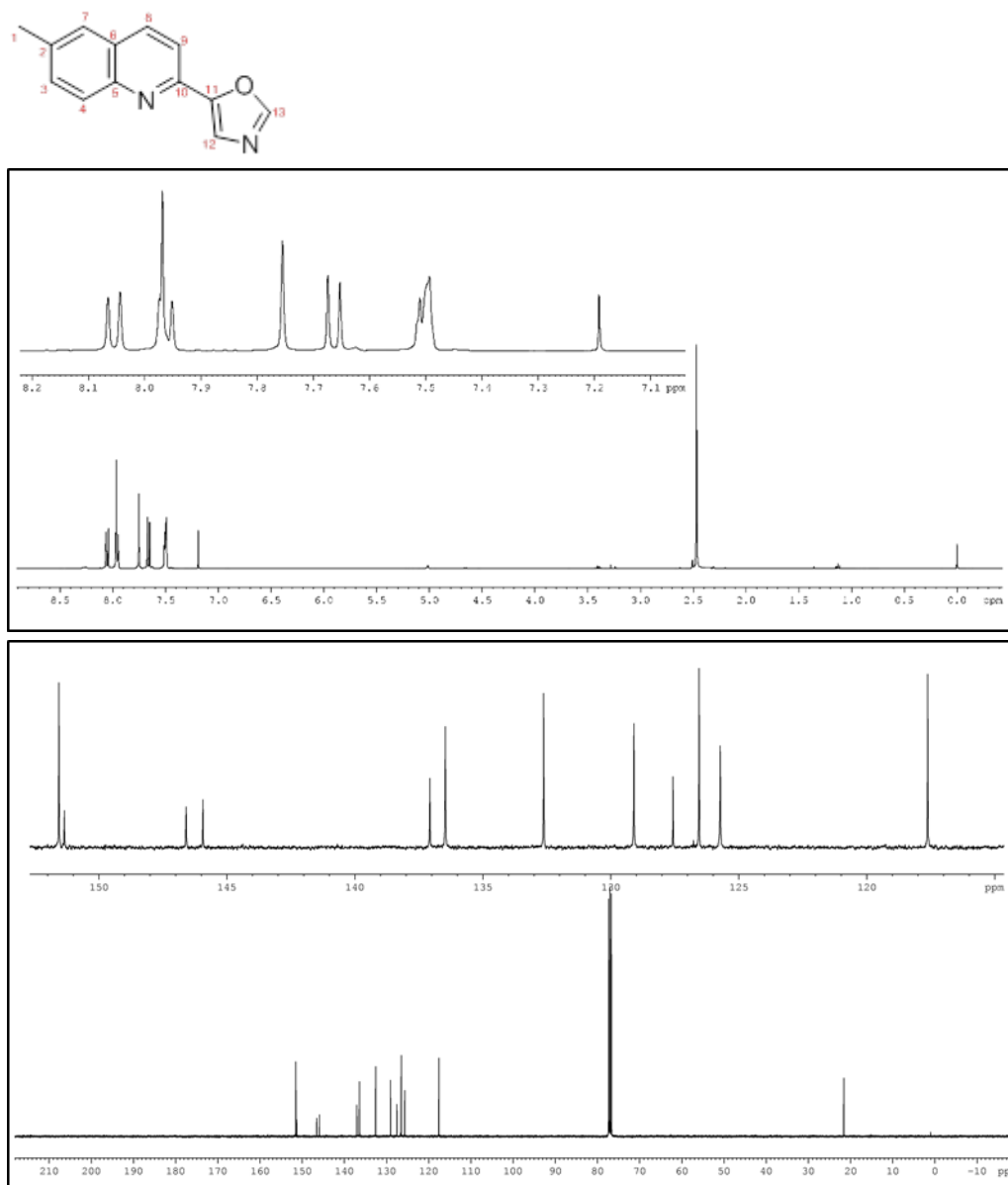




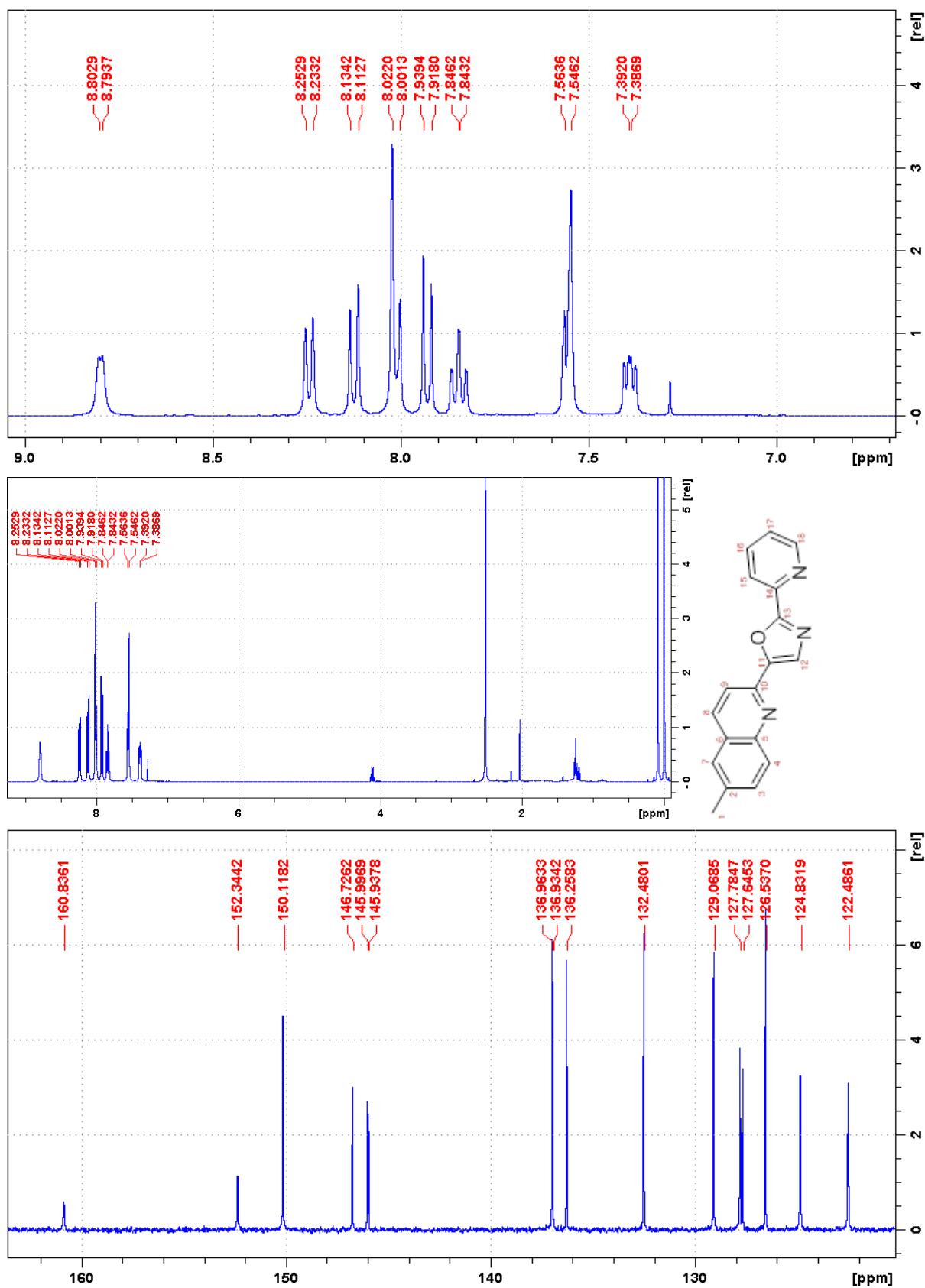
**Supplementary A.15** - NMR spectra of 1,4-bis(2-(pyridine-2-yl)oxazol-5-yl)benzene (Ter-2) in CDCl<sub>3</sub> at 25°C. Top <sup>1</sup>H-NMR spectrum; bottom <sup>13</sup>C-NMR spectrum.



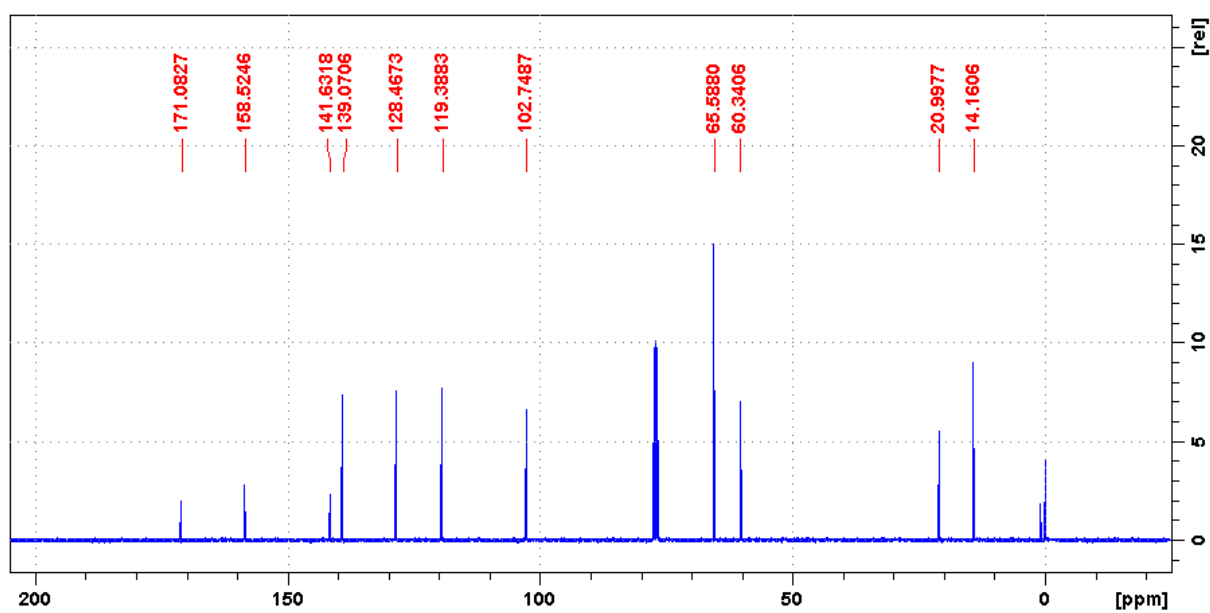
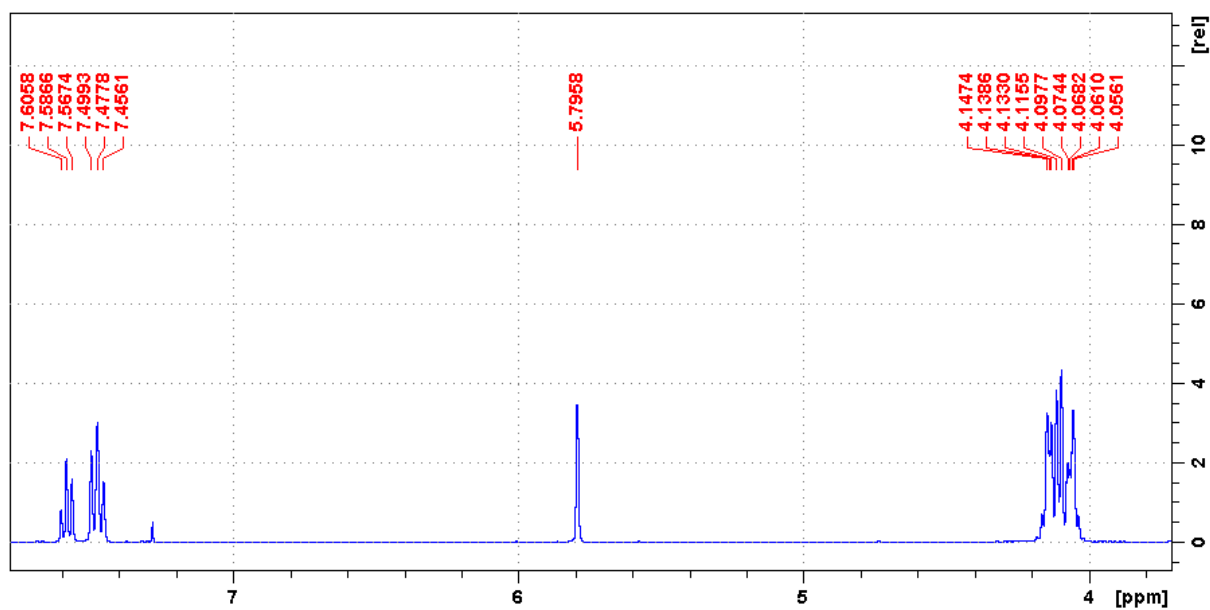
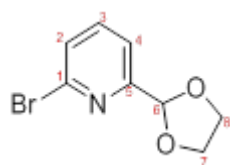
**Supplementary A.16** - NMR spectra of 6-methylquinoline-2-carbaldehyde (Quin-0) in  $\text{CDCl}_3$  at  $25^\circ\text{C}$ . Top  $^1\text{H}$ -NMR spectrum; bottom  $^{13}\text{C}$ -NMR spectrum.



**Supplementary A.17** - NMR spectra of 6-methyl-(2-oxazol-5-yl)quinoline (Quin-1) in  $\text{CDCl}_3$  at  $25^\circ\text{C}$ . Top  $^1\text{H}$ -NMR spectrum; bottom  $^{13}\text{C}$ -NMR spectrum.



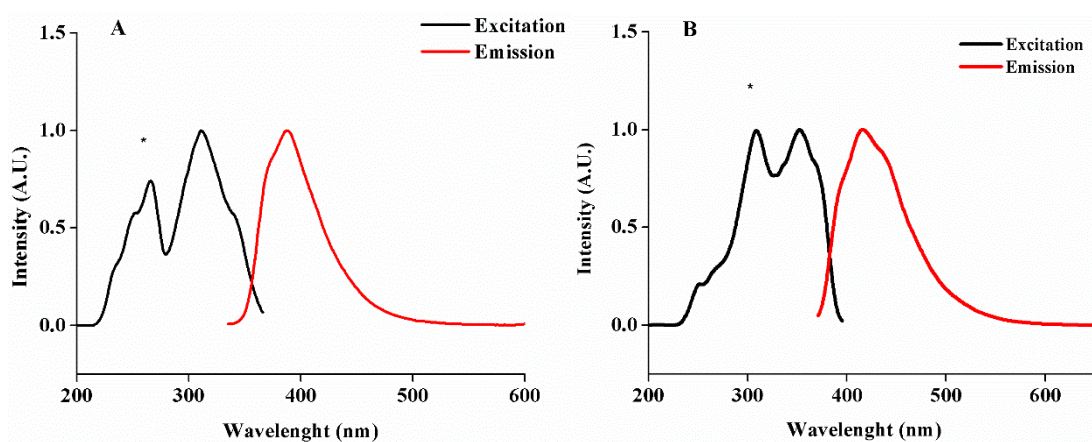
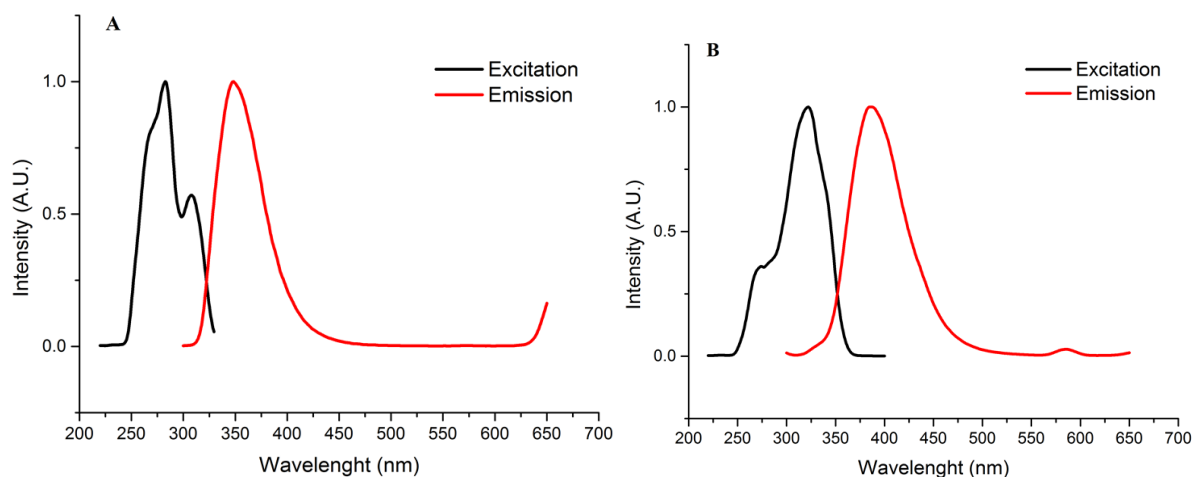
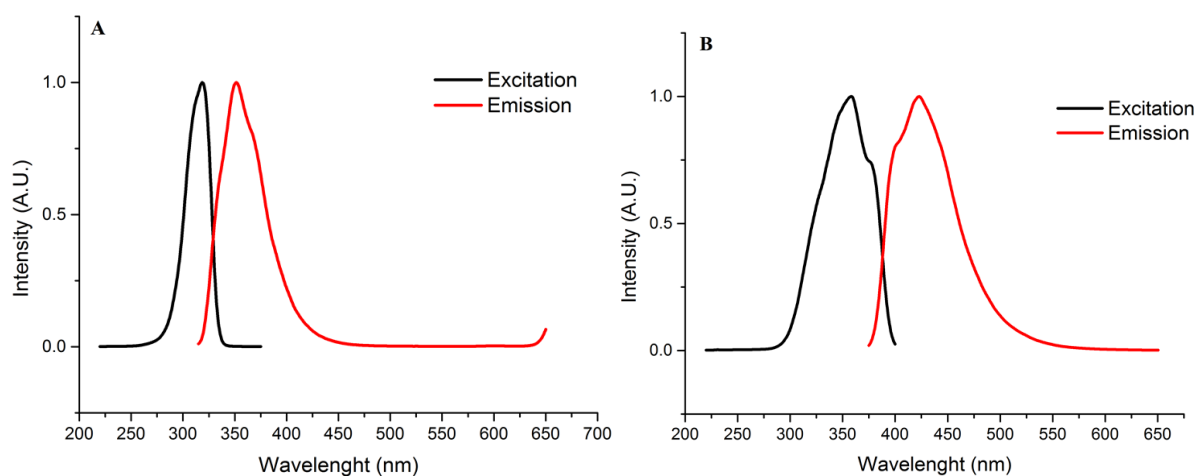
**Supplementary A.18** - NMR spectra of 1,4-bis(2-(pyridine-2-yl)oxazol-5-yl)benzene (Quin-2) in CDCl<sub>3</sub> at 25°C. Top <sup>1</sup>H-NMR spectrum; bottom <sup>13</sup>C-NMR spectrum.

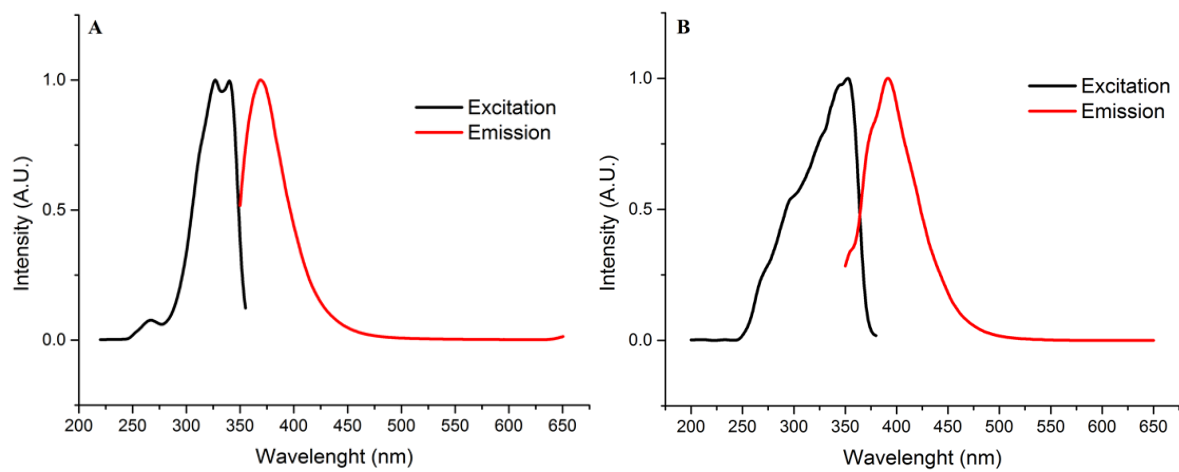


**Supplementary A.19** - NMR spectra of 2-bromo-6-(1,3-dioxolan-2-yl)pyridine (intermediate reactant) in CDCl<sub>3</sub> at 25°C. Top <sup>1</sup>H-NMR spectrum; bottom <sup>13</sup>C-NMR spectrum.

**Excitation / Emission spectra of the synthesized potential G4 ligands**

Note: \* indicates water Ramman signal.

**Supplementary A.20** - Excitation and emission spectrum of (A) Phen-1 and (B) Phen-2.**Supplementary A.21** - Excitation and emission spectrum of (A) Iso-1 and (B) Iso-2.**Supplementary A.22** - Excitation and emission spectrum of (A) Ter-1 and (B) Ter-2.



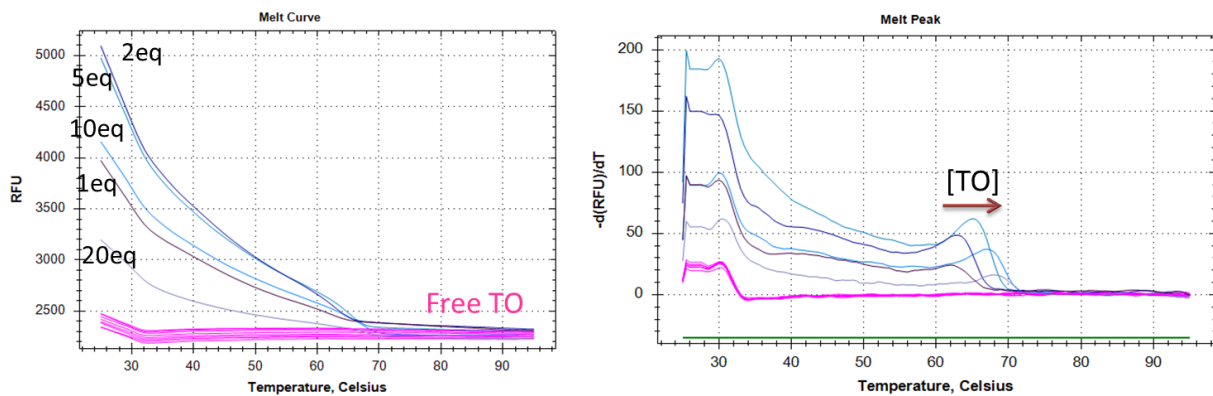
**Supplementary A.23** - Excitation and emission spectrum of (A) Quin-1 and (B) Quin-2.



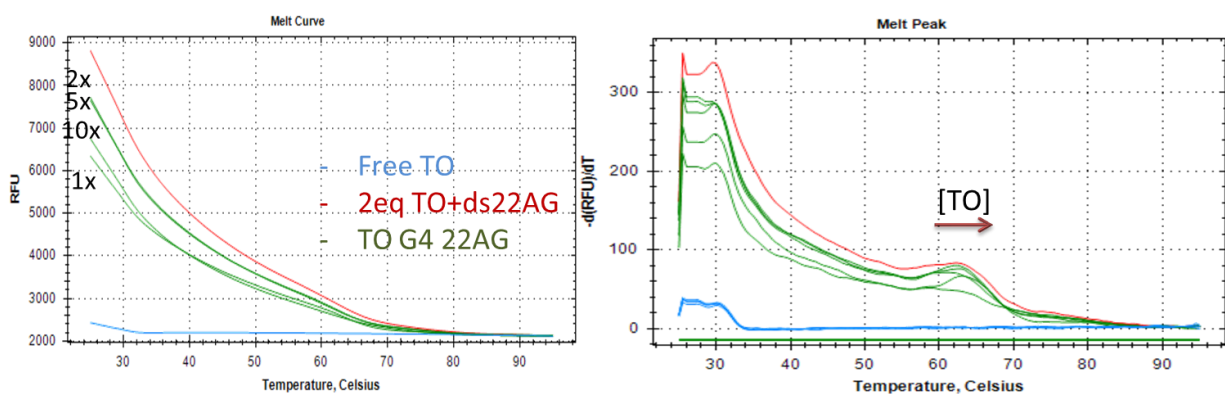


## Appendix B

## TO:DNA ratio fluorescence optimization

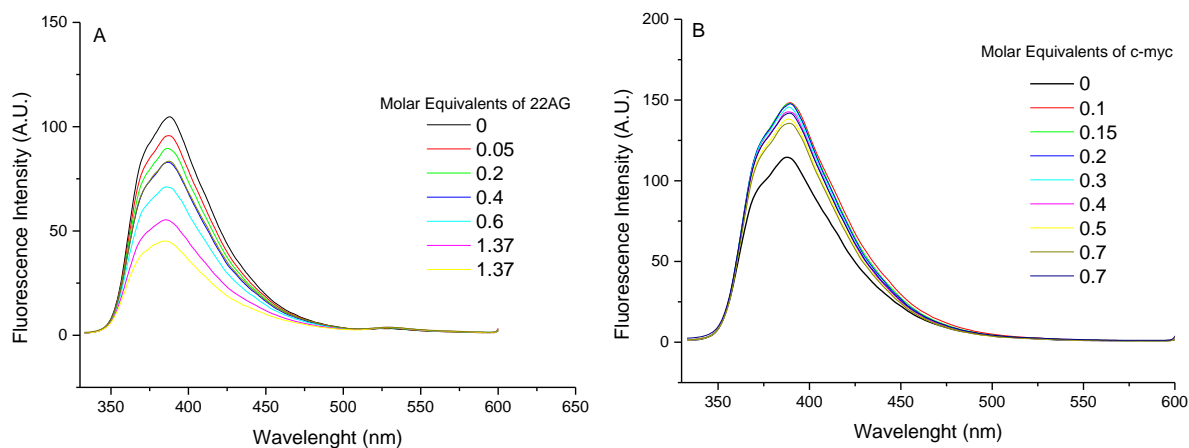


**Supplementary B.1** – Left: Thiazole Orange fluorescence signal ( $\lambda_{exc} = 490 \text{ nm}$ ,  $\lambda_{emi} = 501 \text{ nm}$ ) with increasing temperature in presence of c-myc ( $0.25 \mu\text{M}$ ) with increasing [TO] in 30 mM phosphate with 100 mM  $\text{K}^+$  buffer condition. (B) Plot of the negative of the first derivative of the curves presented in (A).

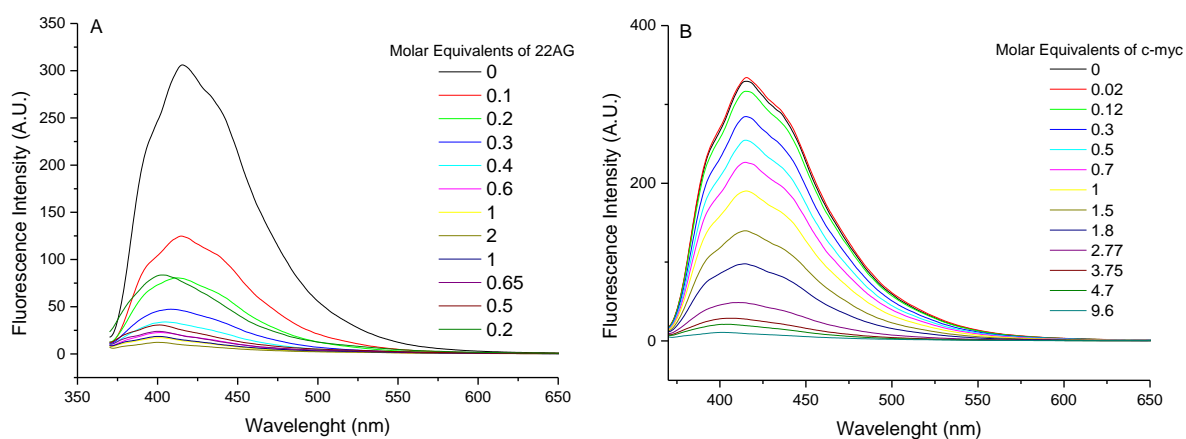


**Supplementary B.2** – Left: Thiazole Orange fluorescence signal ( $\lambda_{exc} = 490 \text{ nm}$ ,  $\lambda_{emi} = 501 \text{ nm}$ ) with increasing temperature in presence of 22AG ( $0.25 \mu\text{M}$ ) with increasing [TO] in 30 mM phosphate with 100 mM  $\text{K}^+$  buffer condition. (B) Plot of the negative of the first derivative of the curves presented in (A).

## Fluorescence Titration

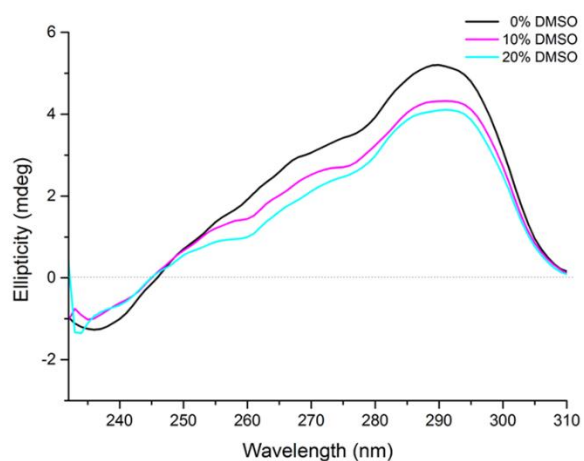


**Supplementary B.3** - Fluorescent titrations experiment of Phen-1 ( $\lambda_{exc}=350$  nm) with (A) 22AG and (B) c-myc G-quadruplex sequences, in 30 mM phosphate with 50 mM  $K^+$  buffer.

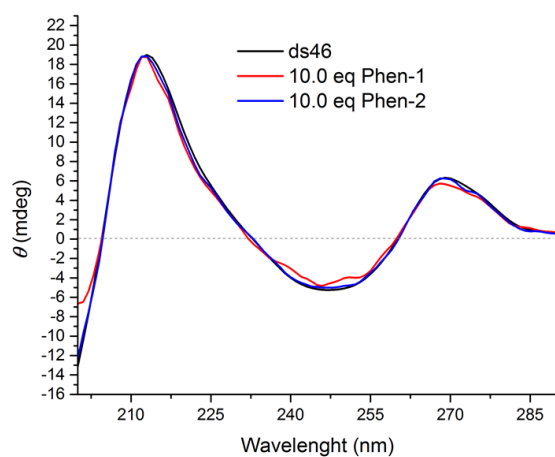


**Supplementary B.4** - Fluorescent titrations experiment of Phen-2 ( $\lambda_{exc}=350$  nm) with (A) 22AG and (B) c-myc G-quadruplex sequences, in 30 mM phosphate with 50 mM  $K^+$  buffer.

## Relevant circular dichroism spectra

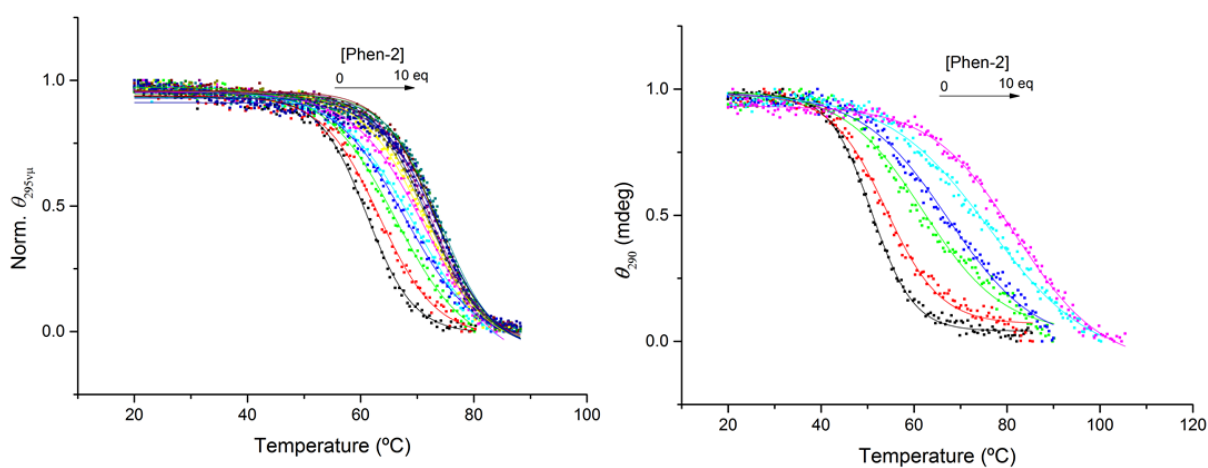


**Supplementary B.5** – CD spectra of 22AG in 10mM potassium with increasing %DMSO.

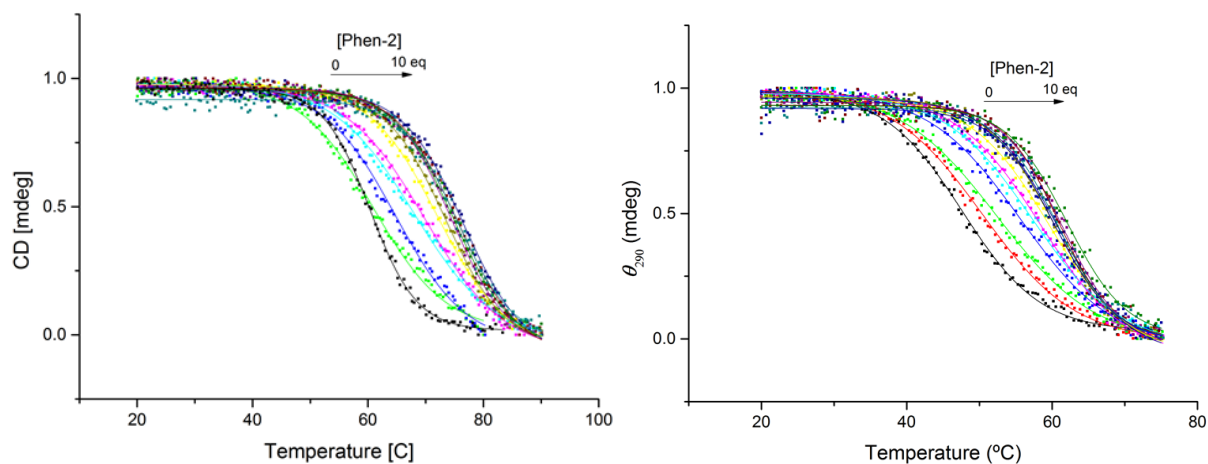


**Supplementary B.6** - CD spectra of an intermolecular duplex in 10mM potassium with 10 eq of Phen-2 and Phen-1.

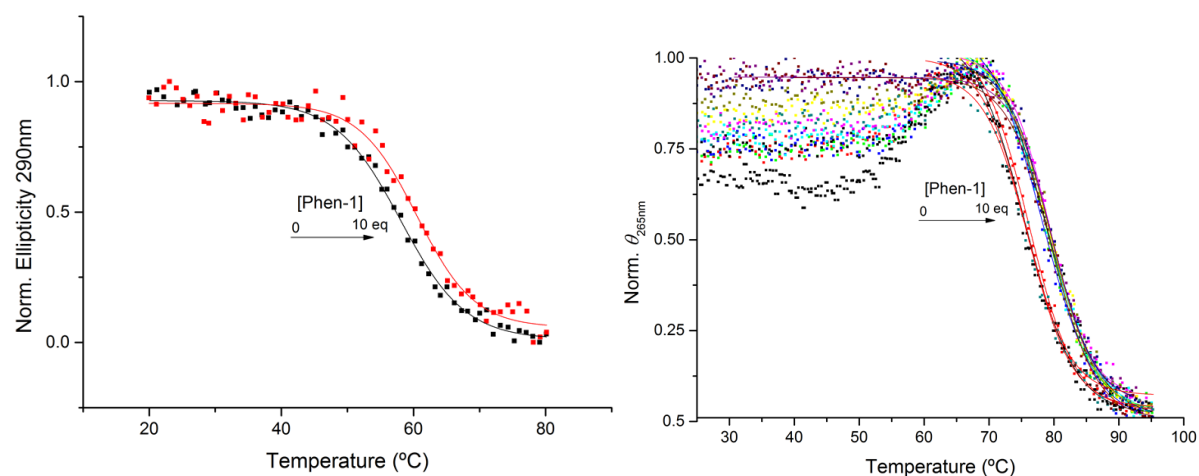
## CD-melting Curves



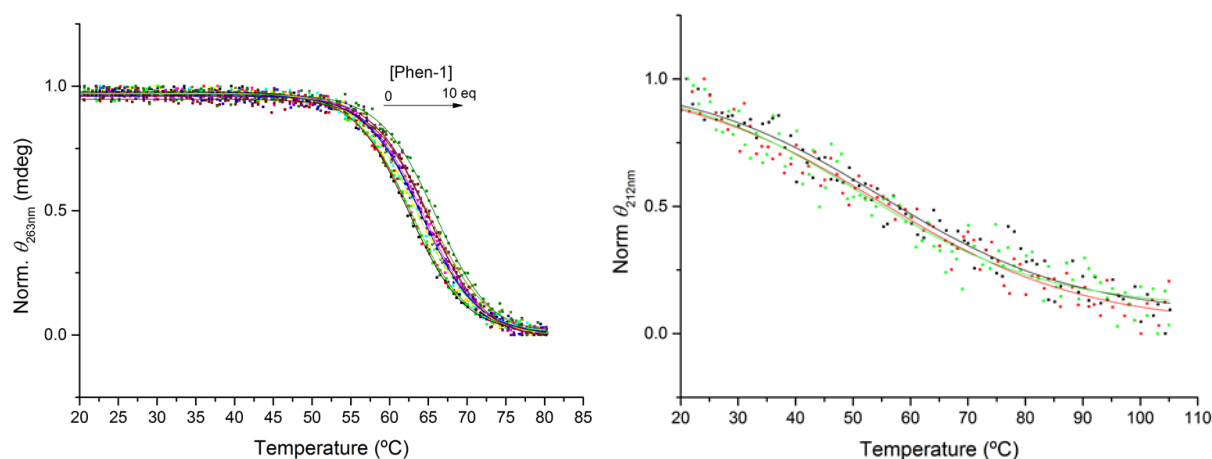
**Supplementary B.7** – Normalized CD-melting curves for 22AG in 100mM Sodium (left) and c-myc 100mM Lithium (right) with increasing concentrations of Phen-2.



**Supplementary B.8** - Normalized CD-melting curves for 22AG in 10mM (left) and 0,2 mM (right) potassium with increasing concentrations of Phen-2.

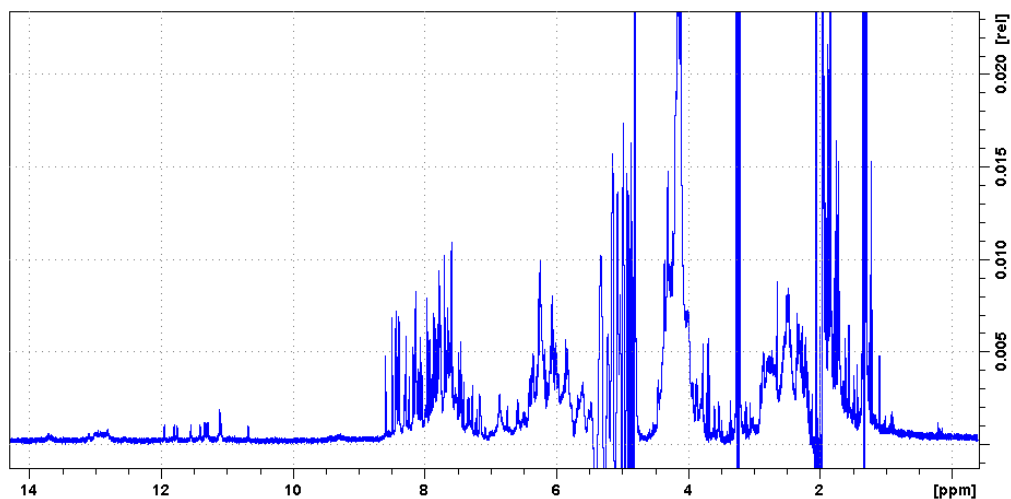


**Supplementary B.9** - Normalized CD-melting curves for 22AG in 100mM Sodium (left) and 100mM potassium (right) with increasing concentrations of Phen-2.

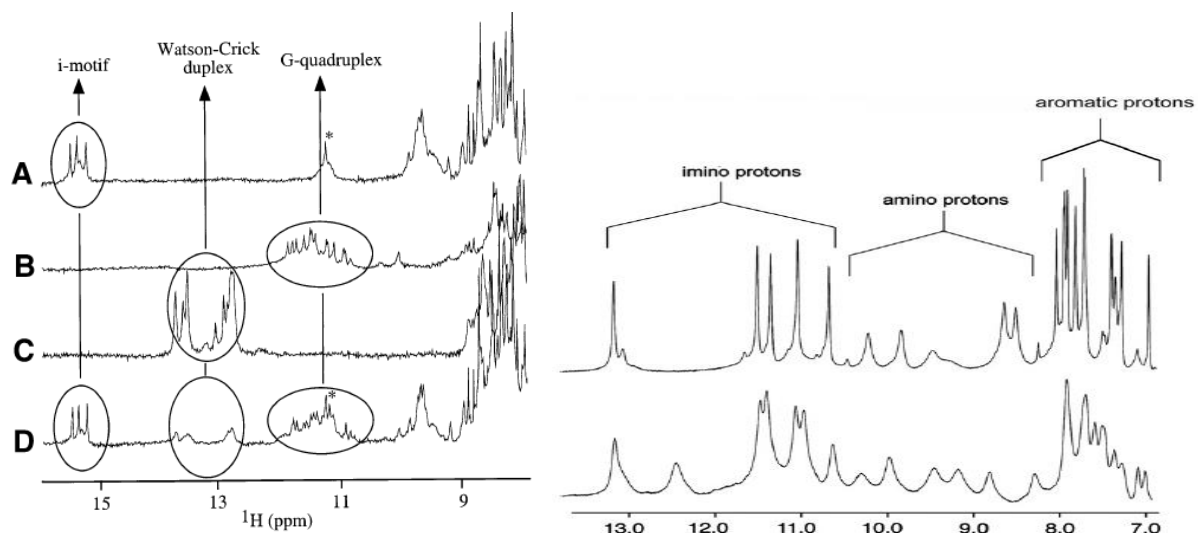


**Supplementary B.10** - Normalized CD-melting curves for 22AG in 100mM Sodium (left) with increasing concentrations of Phen-2. Normalized CD-melting curves for intermolecular duplex in 10mM potassium (right) with 10 equivalents of Phen-2 (green) and Phen-1 (red).

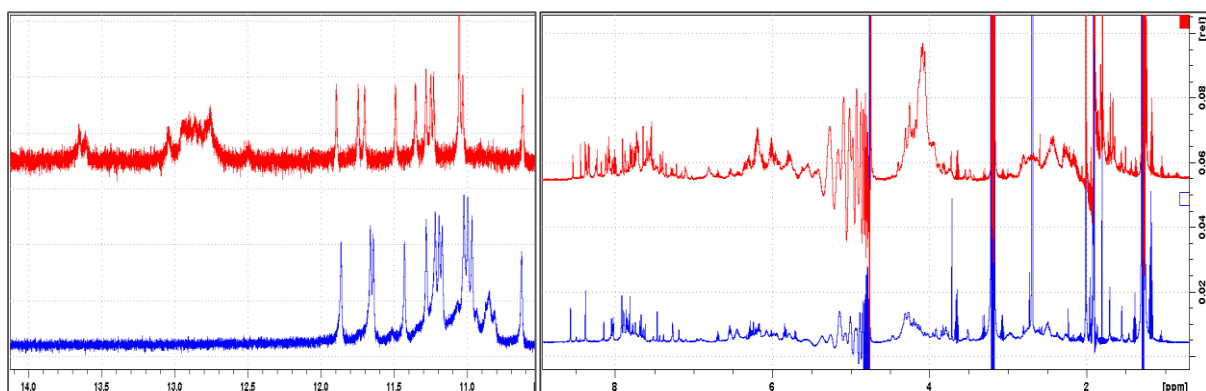
## NMR Spectra



**Supplementary B.11** –  $^1\text{H}$ -NMR spectra of c-myc sequence + complementary sequence, 500 $\mu\text{M}$  strand concentration, in 10mM phosphate buffer, 100mM  $\text{K}^+$ , 10%D $_2\text{O}$  at 600MHz at 25 $^\circ\text{C}$ .



**Supplementary B.12** – Representation of signature signals for i-motif, G-quadruplex and WC duplex base, showing the base-pairing H bond signals.

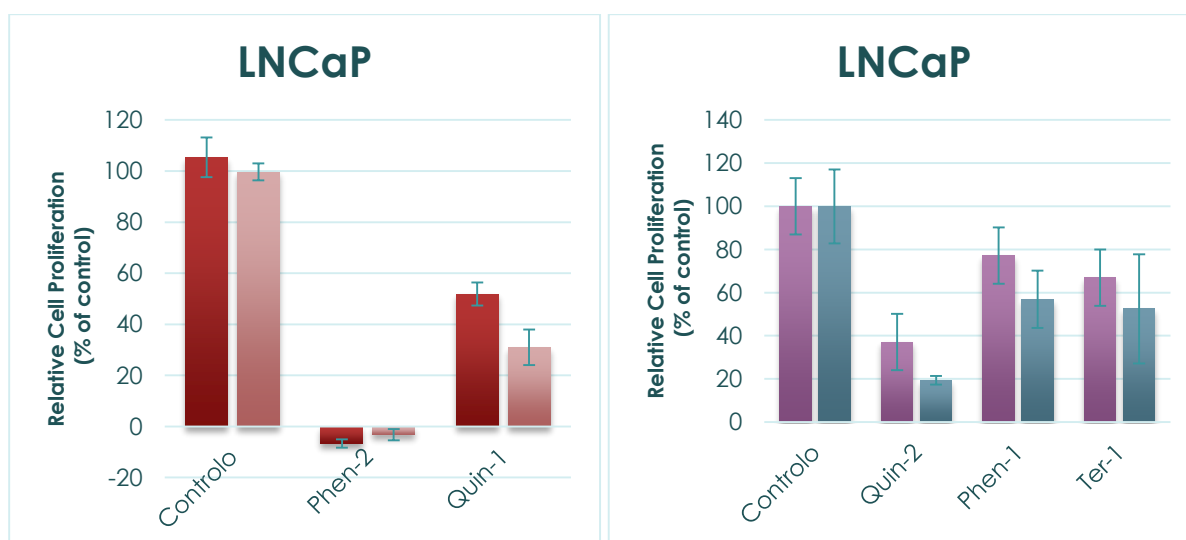


**Supplementary B.13** – Comparison of  $^1\text{H}$ -NMR spectrum of c-myc sequence alone (blue) and with its complementary strand (red).

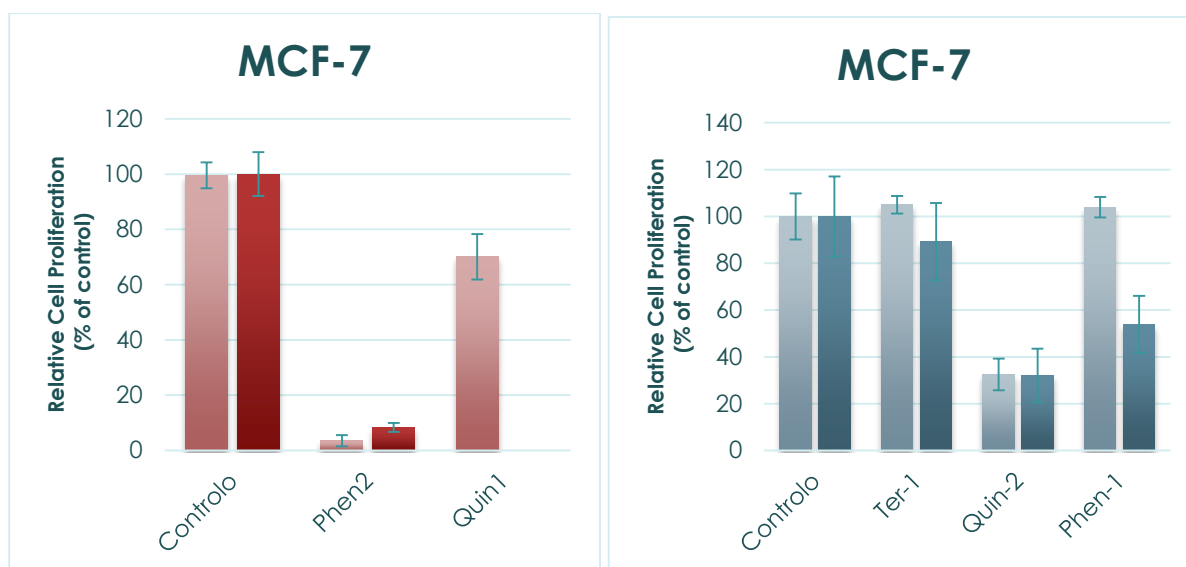


## Appendix C

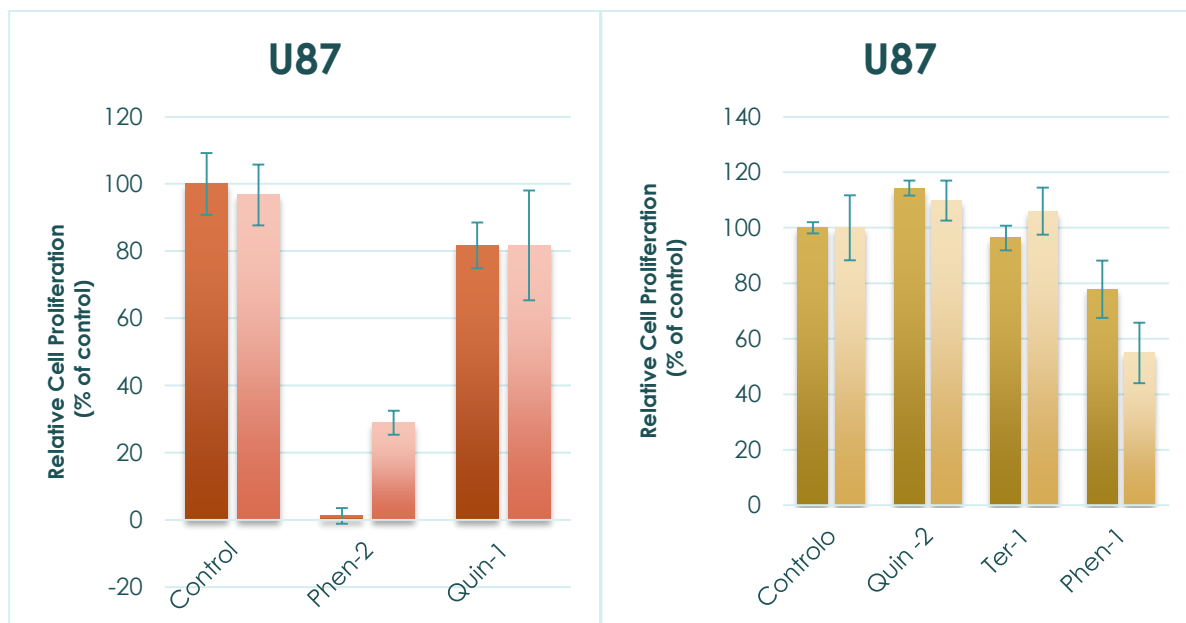
### MTT screening assay replicates



**Supplementary C.1** – Plot of the relative cell proliferation of LNCaP incubated with described compounds during 72 h exposition at 30.0  $\mu$ M, determined by the MTT assay. Each color represent a duplicate and each graph correspond to experiments on separate plates.

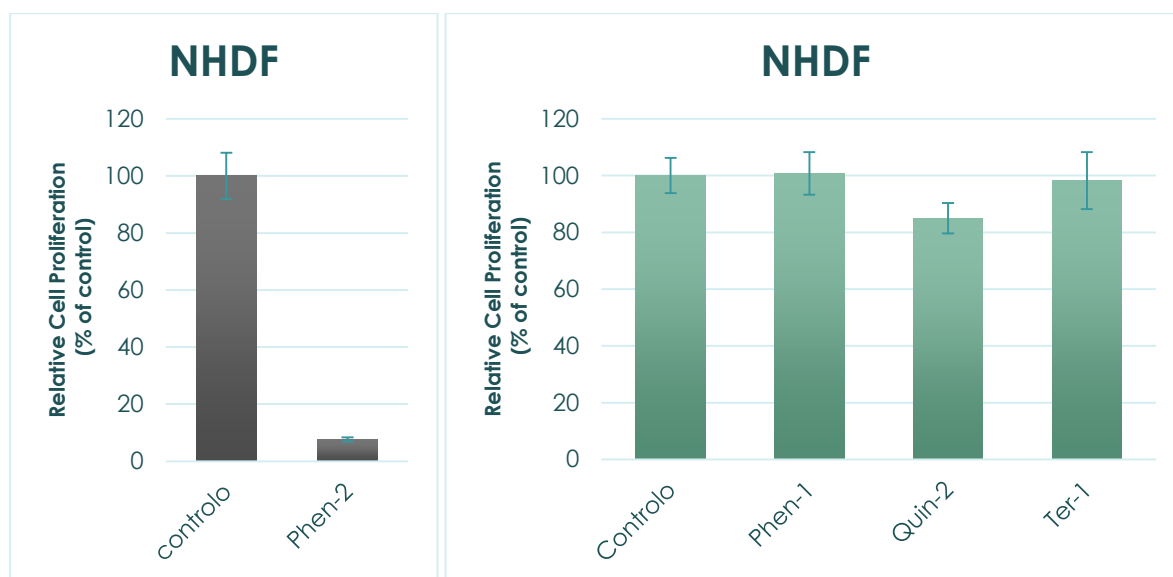


**Supplementary C.2** – Plot of the relative cell proliferation of MCF-7 incubated with described compounds during 72 h exposition at 30.0  $\mu\text{M}$ , determined by the MTT assay. Each color represent a duplicate and each graph correspond to experiments on separate plates.



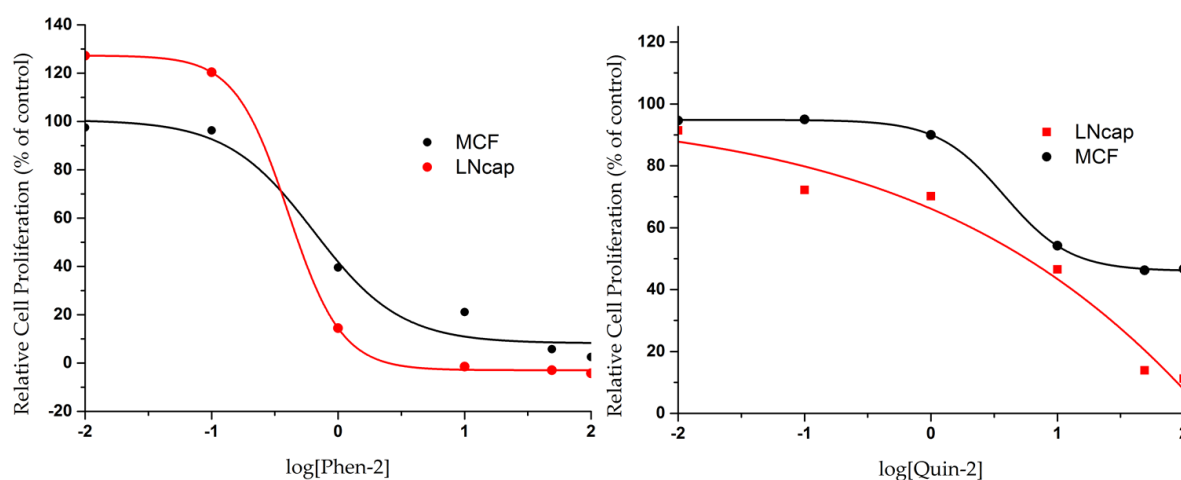
**Supplementary C.3** – Plot of the relative cell proliferation of U87 incubated with described compounds during 72 h exposition at 30.0  $\mu\text{M}$ , determined by the MTT assay. Each color represent a duplicate and each graph correspond to experiments on separate plates.





**Supplementary C.4** – Plot of the relative cell proliferation of NHDF incubated with described compounds during 72 h exposition at 30.0  $\mu\text{M}$ , determined by the MTT assay. Each graph correspond to experiments on separate plates.

### Concentration-response fitting curves



**Supplementary C.5** – Plot of the relative cell proliferation of LNCaP, MCF-7 and U87 cell lines with log [Quin-2] and [Phen-2]. Cells were incubated with Quin-2 and Phen-2 during 72 h exposition at 30.0  $\mu\text{M}$ , determined by the MTT assay. Curves were fitted with a curve dose-response.