UNIVERSITY OF NAPLES FEDERICO II

DEPARTMENT OF PHARMACY

PhD program in Pharmaceutical Sciences

Exploring noncanonical nucleic acid structures: Biophysical characterization and interaction with proteins and potential drugs

PhD student
SIMONA MARZANO

Coordinator
Prof. ROSARIA MELI

Tutor
Prof. BRUNO PAGANO

XXXVI CYCLE (2021-2023)

To my mum
To my dad

Thank you for all you have done,
are doing,
and will do for me.

PREFACE

During my PhD, I devoted a significant amount of time to studying non-canonical nucleic acid molecules and their interactions with various binding partners, such as proteins, peptides, and small molecules. Under the guidance of my tutor, prof. Bruno Pagano, and my co-tutor, prof. Concetta Giancola, along with prof. Jussara Amato, I have learned and applied a variety of spectroscopic techniques, including circular dichroism, fluorescence, UV-visible, and nuclear magnetic resonance. Additionally, I applied microscale thermophoresis, gel electrophoresis, and calorimetric techniques to obtain precise biophysical characterizations of these interactions. I have also expanded my knowledge of nuclear magnetic resonance spectroscopy as a powerful tool for analyzing G-quadruplex structure in the presence of G4 ligands during my research at the Slovenian NMR Center in Ljubljana (SLO) under the supervision of prof. Janez Plavec.

All the results reported here were achieved through the combination of my dedication and the support of the right people at the right time. I was honored to have outstanding mentors, prof. Bruno Pagano and prof. Jussara Amato, who gave me daily guidance, prof. Concetta Giancola and prof. Antonio Randazzo, who were significant milestones. I firmly believe that none of this would have been possible without my *friends*, Federica, Raffaele, Nunzia, Anna, Valentina, Francesca, and Mina. *Thanks*.

I dedicate this work and everything it means to me, to my family, my sister, my brother, my uncles, and my grandparents, but especially to my parents. During the storm you were my sunshine. At last, I am sincerely grateful to my beloved one, Marco, who has stood by my side since high school.

Last, but not least, I'd like to thank myself for my strength, resilience, and courage.

Simona, April 4th, 2024

TABLE OF CONTENT

ABSTRACT

Besides the canonical nucleic acid conformations, noncanonical secondary structures, such as G-quadruplexes (G4s) and i-motifs (iMs), are found in critical areas of the human genome such as telomeres and gene promoters. G4 structures play a significant role in several biological processes, including telomere maintenance, malignant transformation, and cancer development. Despite recent discussions and reviews on the structural aspects and potential biological functions of iM structures, there is still much to learn about this unique nucleic acid conformation.

In this PhD thesis, several studies have been carried out with the aim of characterizing these noncanonical secondary structures, as well as investigating their interaction with various molecules (proteins, small molecules, peptides).

CHAPTER 1 provides a general description of nucleic acids and noncanonical oligonucleotide conformations, focusing on their structural properties and biological roles.

CHAPTER 2 discusses the main biophysical methodologies employed in the studies presented in the following chapters.

CHAPTER 3 reports the study of four different DNA G4 structures as protein targets by employing a chemoproteomic approach.

CHAPTER 4 describes the identification of new G4-targeting small molecules and peptides and the characterization of their interactions by means of a combination of physicochemical, mostly spectroscopic, techniques. Particularly, Paragraph 4.2 deals with an in-depth NMR characterization of the interaction between G4s and a new hit compound, conducted as a visiting PhD student in the laboratory of prof. Janez Plavec at Slovenian NMR Center in Ljubljana (SLO).

CHAPTER 5 focuses on the investigation of the thermodynamics of folding of an iM-forming sequence, carried out with the aim of elucidating the unique features of these structures.

The last section of this PhD thesis collects the GENERAL CONCLUSIONS.

Finally, the APPENDIX reports the scientific articles published and cited throughout the thesis. PAPERS VI and VII are not included in the APPENDIX as they are still in preparation.

LIST OF PUBLICATIONS

Paper I

"On the thermodynamics of folding of an i-motif DNA in solution under favorable conditions". Amato J., D'Aria F., **Marzano S.**, Iaccarino N., Randazzo A., Giancola C., Pagano B. *Physical Chemistry Chemical Physics*, 2021, *23*(28), 15030–15037.

Paper II

"Targeting of telomeric repeat-containing RNA G-quadruplexes: from screening to biophysical and biological characterization of a new hit compound". **Marzano S.**, Pagano B., Iaccarino N., Di Porzio A., De Tito S., Vertecchi E., Salvati E., Randazzo A., Amato J. *International Journal of Molecular Sciences*, 2021, *22*(19), 10315–10335.

Paper III

"Balancing affinity, selectivity, and cytotoxicity of hydrazone-based G-quadruplex ligands for activation of interferon β genes in cancer cells". **Marzano S.**, Miglietta G., Morigi R., Marinello J., Arleo A., Procacci M., Locatelli A., Leoni A., Pagano B., Randazzo A., Amato J., Capranico G. *Journal of Medicinal Chemistry*, 2022, *65*(18), 12055–12067.

Paper IV

"Ligand-based drug repurposing strategy identified SARS-CoV-2 RNA G-quadruplex binders". Moraca F., **Marzano S.**, D'Amico F., Lupia A., Di Fonzo S., Vertecchi E., Salvati E., Di Porzio A., Catalanotti B., Randazzo A., Pagano B., Amato J. *Chemical Communications*, 2022, *58*(85), 11913–11916.

Paper V

"Unveiling the interaction between DNA G-quadruplexes and RG-rich peptides". Grasso N., Graziano R., **Marzano S.**, D'Aria F., Merlino F., Grieco P., Randazzo A., Pagano B., Amato J. *International Journal of Biological Macromolecules*, 2023, *253*(Part 3), 126749–126759.

Manuscripts in preparation:

Paper VI

"Exploiting DNA G-quadruplexes as targets for binding proteins". **Marzano S.**, *et al.*

Paper VII

"G-quadruplex DNA recognition by an anticancer benzimidazole derivative: insights from NMR spectroscopy". **Marzano S.**, *et al.*

Other publications:

Paper VIII

"Identification of effective anticancer G-quadruplex-targeting chemotypes through the exploration of a high diversity library of natural compounds". Platella C., Ghirga F., Zizza P., Pompili L., **Marzano S.**, Pagano B., Quaglio D., Vergine V., Cammarone S., Botta B., Biroccio A., Mori M., Montesarchio D. *Pharmaceutics*, 2021, *13*(10), 1611-1637.

Paper IX

"Synthesis and characterization of bis-triazolyl-pyridine derivatives as noncanonical DNA-interacting compounds". Di Porzio A., Galli U., Amato J., Zizza P., Iachettini S., Iaccarino N., **Marzano S.**, Santoro F., Brancaccio D., Carotenuto A., De Tito S., Biroccio A., Pagano B., Tron G.C., Randazzo A. *International Journal of Molecular Sciences*, 2021, *22*(21), 11959–11988.

Paper X

“Cyclic triimidazoles as stabilizers for gene promoter and human telomeric DNA G-quadruplexes”. Malpicci D., Andolina S., Di Ciolo S., Lucenti E., Cariati E., **Marzano S.**, Pagano B., Amato J., Randazzo A., Giannini C. *European Journal of Organic Chemistry*, 2022, 35, e202200718.

Paper XI

“Printed electrochemical strip for the detection of miRNA-29a: a possible biomarker related to Alzheimer's disease”. Miglione A., Raucci A., Amato J., **Marzano S.**, Pagano B., Raia T., Lucarelli M., Fuso A., Cinti S. *Analytical Chemistry*, 2022, *94*(45), 15558–15563.

Paper XII

“Theoretical and experimental studies on the interaction of biphenyl ligands with human and murine PD-L1: Up-to-date clues for drug design”. Donati G., D'Amore V.M., Russomanno P., Cerofolini L., Amato J., **Marzano S.**, Salobehaj M., Rizzo D., Assoni G., Carotenuto A., La Pietra V., Arosio D., Seneci P., Fragai M., Brancaccio D., Di Leva F.S., Marinelli L. *Computational and Structural Biotechnology Journal*, 2023, *21*, 3355–3368.

Paper XIII

“New biologic (Ab-IPL-IL-17) for IL-17-mediated diseases: identification of the bioactive sequence (nIL-17) for IL-17A/F function”. Saviano A., Manosour A.A., Raucci F., Merlino F., Marigliano N., Schettino A., Wahid M., Begum J., Filer A., Manning J.E., Casillo G.M., Piccolo M., Ferraro M.G., **Marzano S.**, Russomanno P., Bellavita R., Irace C., Amato J., Alfaifi M., Rimmer P., Iqbal T., Pieretti S., Vellecco V., Caso F., Costa L., Giacomelli R., Scarpa R., Cirino G., Bucci M., McGettrick H.M., Grieco P., Iqbal A.J., Maione F. *Annals of the Rheumatic Diseases*, 2023, *82*(11), 1415–1428.

LIST OF ABBREVIATIONS

1D	Monodimensional
2D	Bidimensional
A	Adenine
ACN	Acetonitrile
AHNAK	Neuroblast differentiation-associated protein
ALT	Alternative lengthening telomere
AMBIC	Ammonium bicarbonate
Bcl-2	B-cell lymphoma 2
C	Cytosine
c-Kit	Cellular receptor tyrosine kinase
c-Myc	Cellular myelocytomatosis
CD	Circular dichroism
COSY	Correlation spectroscopy
DAP	Death-associated protein
DDR	DNA damage response
DMSO	Dimethyl sufoxide
DNA	Deoxyribonucleic acid
DSC	Differential scanning calorimetry
DSS	Dimethyl silapentane sulfonate
DTT	Dithiothreitol
EB	Ethidium bromide
EDTA	Ethylenediamine tetraacetic acid

EGTA	Ethylene glycol-bis-tetraacetic acid tetrasodium salt
FAM	6-carboxyfluorescein
FDA	US Food and Drug Administration
FI	Fluorescence enhancement
FID	Free induction decay
FID	Fluorescence intercalator displacement
FITC	Fluorescein isothiocyanate
FRET	Fluorescence resonance energy transfer
FT	Fourier transform
G	Guanine
G4	G-quadruplex
G4RP	G4-related proteins
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
HCOOH	Formic acid
HeLa	Human cervix cancer cells
HIF-1α	Hypoxia-inducible factor 1-alpha
HMBC	Heteronuclear multiple bond correlation
HNRNP	Heteroribonucleoprotein
HPLC	High-performance liquid chromatography
HSQC	Heteronuclear single quantum coherence
IAM	Iodoacetammide
IFN-β	Interferon β
iM	i-motif
IR	Infrared
ITC	Isothermal titration calorimetry
JAZF-1	Juxtaposed with another zinc finger protein 1
KHSRP	Far upstream element-binding protein 2
KRAS	Kirsten rat sarcoma virus

LBVS	Ligand-based virtual screening
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LMNA	Prelamin A/C
LMNB1	Lamin B1
lncRNA	Long non-coding RNA
LTR	Long terminal repeat
MAZ	Myc-associated zinc finger
MD	Molecular dynamics
MNMCA1	Murine fibrosarcoma cells
mRNA	RNA messenger
MST	Microscale thermophoresis
NCL	Nucleolin
NIQI	Novel interesting quadruplex interaction motif
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
NONO	non-POU domain-containing octamer-binding protein
NPM	Nucleophosmin
PAGE	Polyacrylamide gel electrophoresis
PARP-1	Poly(ADP-ribose) polymerase 1
PCA	Principal component analysis
PDGF-A	Platelet-derived growth factor-A
PDS	Pyridostatin
POT1	Protection of telomere 1
PPIA	Peptidyl-prolyl cis-trans isomerase A
RBM3	RNA-binding protein
RET	Rearranged during transfection
RF	Radiofrequency
RG	Arginine/glycine

RMSD	Root mean square deviation
RNA	Ribonucleic acid
RT qPCR	Real time-quantitative polymerase chain reaction
SARS-CoV-2	Severe acute respiratory syndrome coronavirus-2
scrG	G-rich scrambled sequence
SI	Selectivity index
SVD	Singular value decomposition
T	Thymine
TAMRA	Carboxytetramethylrhodamine
TERRA	Telomeric repeat-containing RNA
ThT	Thioflavin T
TIF	Telomere-dysfunction-induced foci
TMS	Tetramethyl silane
TO	Thiazole orange
TOCSY	Total correlation spectroscopy
TP53	Tumor suppression protein 53
TSP	Trimethylsilyl propionic acid
U	Uracil
U2OS	Osteosarcoma cancer cell line
UTR	Untranslated regions
UVRR	UV Raman resonance
VEGF	Vascular endothelial growth factor
VIM	Vimentin
YY1	Yin Yang 1
γH2AX	S139-phosphorylated histone H2AX

CHAPTER 1
INTRODUCTION

1.1 GENERAL DESCRIPTION OF NUCLEIC ACIDS

Nucleic acids are essential biological macromolecules involved in vital biological functions such as storing and transmitting genetic information and directing protein synthesis. Nucleic acids can be described as nucleotide polymers, where each nucleotide comprises three components: a nitrogenous base, a pentose sugar, and a phosphate group (**Figure 1.1**).

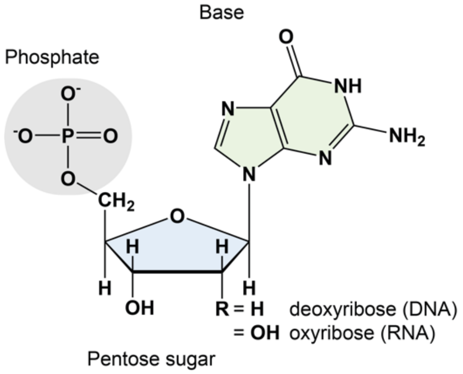


Figure 1.1 Schematic representation of nucleotide (i.e., guanine) composition.

The nitrogenous bases of nucleotides are classified as purines [adenine (A) and guanine (G)], which contain two fused carbon-nitrogen rings, or pyrimidines [cytosine (C), thymine (T), and uracil (U)] which are composed of a single six-membered nitrogen-containing ring. Each base is characterized by a unique chemical structure with its own set of functional groups (**Figure 1.2**).

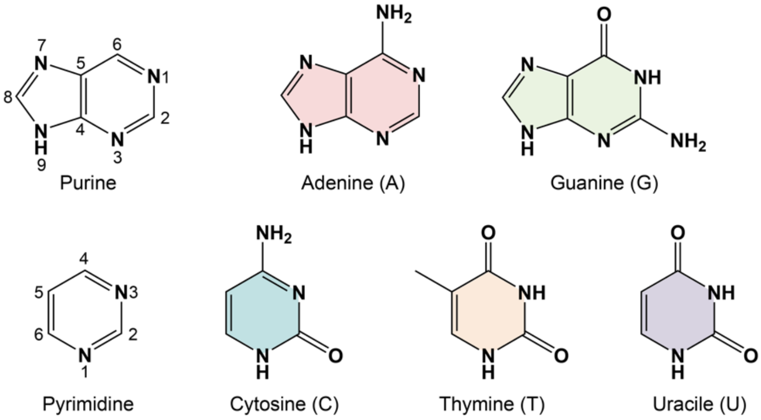


Figure 1.2 Structure of purine and pyrimidine bases.

The pentose sugar is deoxyribose or ribose in deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) polymers, respectively. DNA contains A, T, C, and G, while in the RNA, T is replaced by U. The carbon atoms of a nucleotide’s sugar moiety are numbered 1′, 2′, 3′, 4′, and 5′. The 1′ carbon is bound through an N-glycosidic bond to the N9 or N1 atom of a purine or a pyrimidine, respectively, whereas the 5′ carbon is linked to the phosphate group. When a nucleotide lacks the phosphate group, it is referred to as a nucleoside. The relative orientation of the base and sugar moieties in a nucleotide's three-dimensional structure can be in either an “*anti*” or “*syn*” conformation. In the “*anti*” conformation, the base and sugar are rotated away from each other, with the C1'-O4' bond of the sugar being *trans* to the N9-C4 bond of purines or the N1-C2 bond of pyrimidines. On the other hand, the “*syn*” conformation occurs when the C1'-O4' bond of deoxyribose or ribose is *cis* to the N9-C4 purine bond or the N1-C2 bond of pyrimidines (**Figure 1.3**).

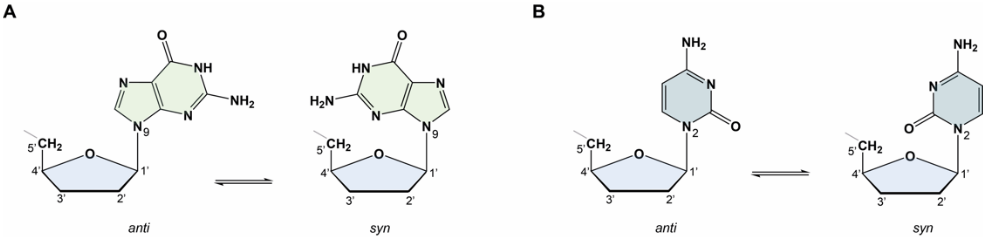


Figure 1.3 Anti and syn conformations of a nucleotide containing (**A**) a purine or (**B**) a pyrimidine base.

DNA/RNA strands have chemical polarity due to the 5'-phosphate group of one nucleotide linking with the 3'-hydroxyl group of the next nucleotide, forming a phosphodiester bond. This creates two chemically different ends: the 5'-end with a free phosphate group and the 3'-end with an unlinked hydroxyl group. In 1953, James Watson and Francis Crick discovered the double helix structure of DNA, laying the foundation of modern genetics. According to their model, the DNA molecule comprises two antiparallel strands running in opposite directions, held together in a right-handed helix (so-called B-DNA) by hydrogen bonding between the bases (Watson & Crick, 1953). When DNA strands pair up, A always pairs with T (A:T) and G with C (G:C). These so-called Watson-Crick base pairs are formed with two or three hydrogen bonds, respectively, meaning that the composition of one strand can be defined by the sequence of the other (**Figure 1.4**).

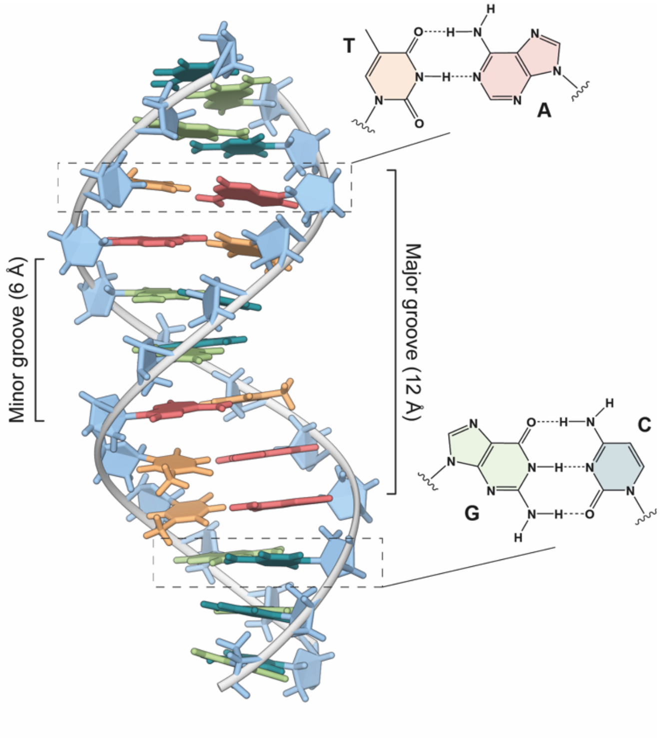


Figure 1.4 Schematic representation of DNA's three-dimensional double helix structure (B-DNA) and Watson-Crick base pairing.

Such complementary base-pairing allows the bases to be packed in the energetically most favorable arrangement within the double helix. At the same time, the hydrophilic sugar-phosphate backbone lays on the outside of the helix itself. Conversely, RNA could not form the same structure due to the close van der Waals contact caused by the ribose’s additional oxygen. The B-DNA conformation, first described by Watson and Crick, is the most common form found in living cells. The reversible inter-strand bonds found in DNA allow for easy replication and gene expression, making the information stored within highly accessible. Further, this structure features a wide major groove and a narrow minor groove that span the entire molecule, acting as binding sites for various other macromolecules, such as proteins.

1.2 NONCANONICAL SECONDARY STRUCTURES OF NUCLEIC ACIDS

Under specific physiological conditions, DNA and RNA molecules can adopt a wide range of noncanonical inter- or intramolecular structures distinct from the common double helix structure (i.e., hairpins, cruciforms, triplexes, G-quadruplexes, i-motifs) (Bansal et al., 2022). Among them, G-quadruplexes (G4s) and i-motifs (iMs) are formed due to the specific nucleotide sequences present in the nucleic acid molecule (**Figure 1.5**).

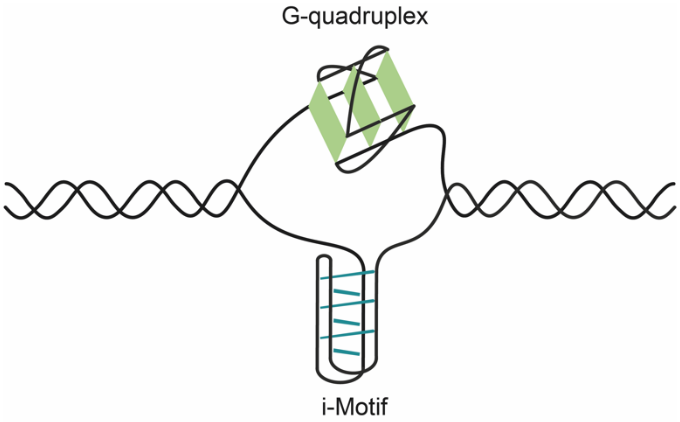


Figure 1.5 Schematic illustration of DNA fragment with G4 and iM structures.

1.2.1 G-QUADRUPLEXES

G-quadruplex structures can form in G-rich nucleic acid DNA or RNA sequences. Four guanines can be placed in a planar arrangement and held together by eight noncanonical Hoogsteen hydrogen bonds, forming a G-tetrad, which is essential for G4 formation (Huppert, 2010; Rhodes & Lipps, 2015). In each G-tetrad, N7 and O6 form H-bond with N2 and N1 of adjacent guanines, respectively (**Figure 1.6**).

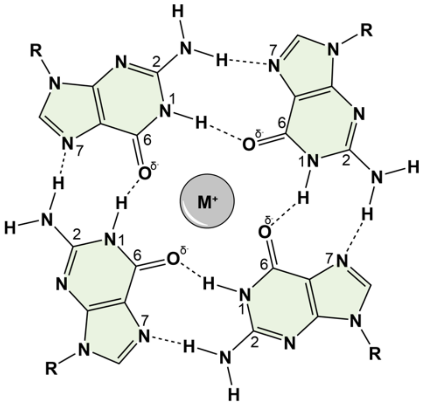


Figure 1.6 Schematic illustration of a G-tetrad with monovalent cation.

The formation of a G4 is the result of the π-π stacking interactions between two or more G-tetrads, thus creating a central channel within the structure. These G-tetrads are linked by loops that can adopt different conformations. Cations in the central channel balance guanine O6 atoms' negative electrostatic potential. Na^+ and K^+ cations effectively stabilize G4 structures, with Na^+ (ionic radius 0.95 Å) ions usually placed in the plane with the G-tetrads and K^+ (ionic radius 1.33 Å) ions equidistant between two G-tetrads. Switching between Na^+ and K^+ often induces structural changes in G4s, showing high conformational flexibility (Williamson et al., 1989; You et al., 2017). From a biological perspective, K^+ is more relevant than Na^+ due to its higher intracellular concentration (ca.140 mM) than Na^+ (5-15 mM).

The G-tetrad arrangement and the presence of cations stabilize the G4 structure. Stable G4 structures have at least three stacked G-tetrads and loops with 1-7 bases, reflecting the following motif $G_{\geq3}N_{1-7}G_{\geq3}N_{1-7}G_{\geq3}N_{1-7}G_{\geq3}N_{1-7}$, where N is any nucleotide and each $G_{\geq3}$ represent a G-tract (Bugaut & Balasubramanian, 2008; Guédin et al., 2010; Hazel et al., 2004; Todd et al., 2005).

The stability and topology of G4s depend on various factors, such as the length and composition of the G4-forming sequence, strand stoichiometry, number of G-tetrads, size of the loops, and the type of binding cations. Concerning the strand stoichiometry, G4s can exist in different inter- or intramolecular forms. Thus, they can either be unimolecular structures, where a single strand folds upon itself, or bi- or tetramolecular species, where two or four strands combine respectively (**Figure 1.7**).

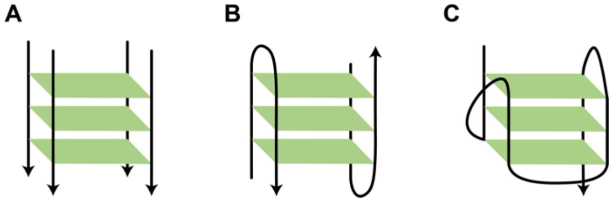


Figure 1.7 Schematic illustration of **A**) tetramolecular, **B**) bimolecular, and **C**) unimolecular G4 structures.

Depending on the relative orientation of the strands, G4s can assume different topologies (**Figure 1.8**) (Bochman et al., 2012; Huppert, 2010; Phan, 2010):

- *Parallel*, in which the four strands are oriented in the same direction;
- *Hybrid*, or *(3+1)*, in which three strands are oriented in one direction and the fourth in the opposite direction;
- *Antiparallel*, in which two strands are oriented in opposite directions; with an *up-down-up-down* or *up-up-down-down* core.

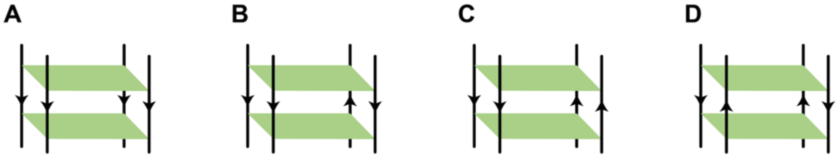


Figure 1.8 Schematic illustration of **A**) *parallel*, **B**) *hybrid*, and **C**,**D**) *antiparallel* G4 structures.

Strictly correlated to the orientation of the strand, guanines in a G-tetrad can be either *anti* or *syn*. Parallel G4s have all glycosidic angles in *anti*-conformation, while antiparallel and hybrid G4s have both *syn*- and *anti*-conformations. Finally, loops connecting G-tetrads can be classified into four groups (**Figure 1.9**) (Bugaut & Balasubramanian, 2008):

- *Lateral*, between two adjacent *antiparallel* strands;
- *Diagonal*, between two opposing *antiparallel* strands across the G-tetrad plane;
- *Double-chain-reversal* or *propeller*, between two adjacent *parallel* strands;
- *V-shaped*, linking two adjacent *antiparallel* strands, one of which a broken G-tract, through residues located on opposite outer faces of the G4.

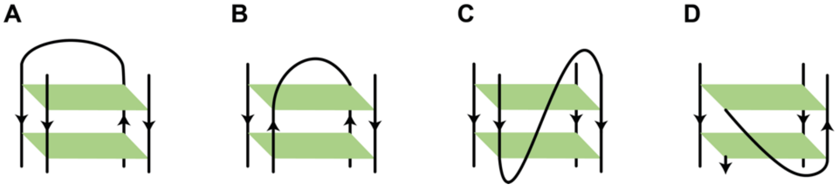


Figure 1.9 Schematic illustration of **A**) *lateral*, **B**) *diagonal*, **C**) *double-chain-reversal*, and **D**) *V-shaped* loops in G4 structures.

The topology of loops affects the dimension of G4 grooves. In the case of *lateral*/*diagonal* loops between *antiparallel* strands, wide, narrow or medium grooves are possible according to the *anti* or *syn* orientation of the glycosidic bonds. On the other hand, *propeller* loops involving *parallel* strands lead to four medium grooves, with all the glycosidic bonds in the *anti* conformation (Burge et al., 2006).

Finally, external G-tetrads from two different inter- or intramolecular G4s interfaced can stack on top of one another, assuming in three different orientations: 5' to 3' (*head-to-tail*), 5' to 5' (*head-to-head*), and 3' to 3' (*tail-to-tail*).

This stacking mode is favorable for parallel G4s (**Figure 1.10A**) (Haider et al., 2008; Smargiasso et al., 2008). Finally, *G-wires* are long, linear structures of many slipped G4 tetrameric subunits (Marsh & Henderson, 1994). They are the longest higher-order G4s and can vary in length depending on the preparation method and sequence. They can be formed by interlocked G4s connected by flanking G-rich strands or sticky ends (**Figure 1.10B**) (Ilc et al., 2013; Pavc et al., 2022).

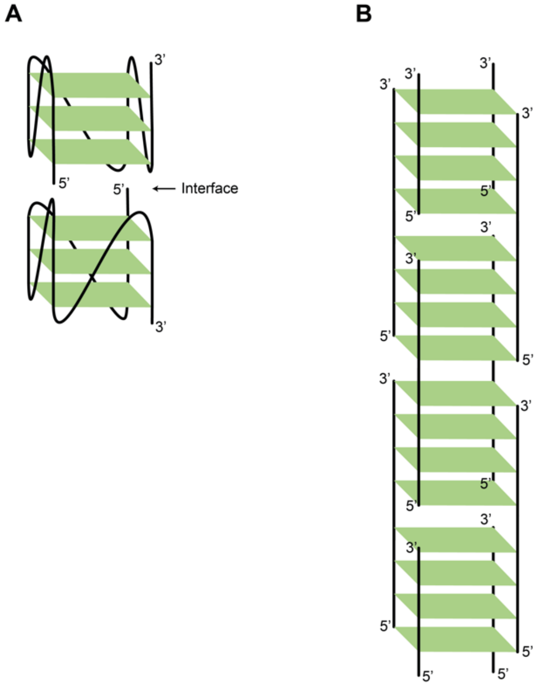


Figure 1.10 Schematic illustrations of **A**) two 5'-5' stacked parallel G4s, and **B**) a G-wire structure.

The differences in the chemical composition of DNA and RNA also affect the properties of the corresponding G4 structures. Usually, RNA G4 structures (rG4s) are more stable compared to DNA G4s. Indeed, RNA has an extra 2'-OH group, allowing for additional hydrogen bonding and intramolecular interactions. The 2'-OH group in RNA also creates steric constraints, which primarily lead to the formation of rG4s in a *parallel* topology (Banco & Ferré-D'Amaré, 2021; Joachimi et al., 2009; Malgowska et al., 2016).

1.2.2 I-MOTIFS

Under slightly acidic conditions, a four-stranded structure can form in the C-rich strand complementary to a G-rich one, the i-Motif (iM). iM structures consist of two intercalated duplexes held together by hemi-protonated C:C$^+$ base pairs, where cytosines pair via Hoogsteen-type hydrogen bonds (**Figure 1.11**) (Gehring et al., 1993).

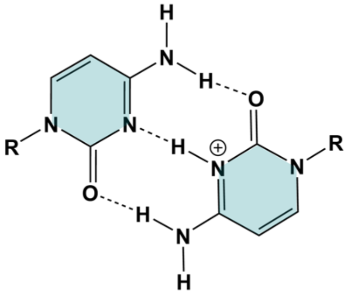


Figure 1.11 Schematic illustration of a hemi-protonated C:C$^+$ base pair.

Similar to G4 structures, iMs can exist in either inter- or intra-molecular forms. In a unimolecular structure, a single strand folds upon itself. When two stretches of cytosines are hairpin-folded and intercalated, they form bi-molecular iMs (Catasti et al., 1997). Finally, tetramolecular iMs can be formed by four stretches of cytosines (**Figure 1.12**).

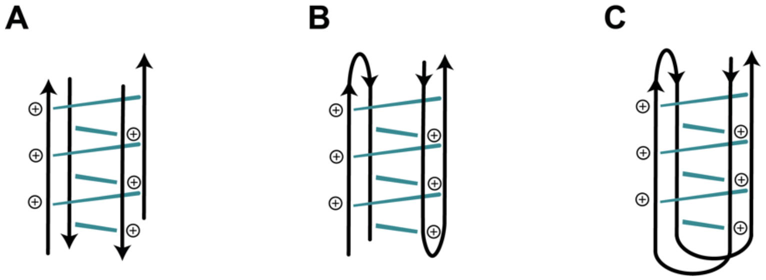


Figure 1.12 Schematic illustration of **A**) tetramolecular, **B**) bimolecular, and **C**) unimolecular iM structures.

Two distinct iM topologies are determined by the intercalation. The 3'E topology features the outermost cytosine pair from the 3' end, whereas the 5'E topology has the outer pair from the 5’ end (**Figure 1.13**) (Guéron & Leroy, 2000; Nonin-Lecomte & Leroy, 2001).

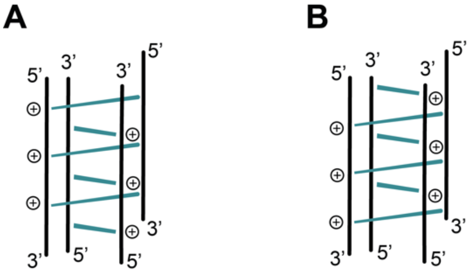


Figure 1.13 Schematic illustration of **A**) 5'E and **B**) 3'E iM topologies.

Further, there are two classes of intramolecular iMs based on the length of the loops: *class I* with short loops and *class II* with longer loops (Brooks et al., 2010). Usually, *class II* iMs are more stable due to stabilizing interactions within longer loop regions, despite the stability of *class I* iMs at neutral or nearly neutral pH (Lieblein et al., 2013; Mir et al., 2017).

Although iM structures require the protonation of cytosines, which is more likely to occur at a lower pH value than what is typically found inside living cells, scientists have been exploring their persistence in living cells. In the last decades, it has been demonstrated that the iM formation can be facilitated by the negative super-helicity induced by transcriptional activity and cell-mimicking molecular crowding conditions (Rajendran et al., 2010; Sun & Hurley, 2009). These findings are significant as they contribute to understanding the mechanisms behind iM formation and its potential implications in various biological processes.

1.3 **BIOLOGICAL RELEVANCE OF G-QUADRUPLEXES**

In 2013, G4 structures were visualized in the genomic DNA of human cells using a structure-specific antibody (BG4) (Biffi et al., 2013). There is an estimate of ~700000 potential G4-forming sequences in the human genome, playing essential roles in regulating DNA transcription and replication, RNA translation, and maintaining genome integrity (Chambers et al., 2015). These structures are highly conserved between different species and found in many areas, such as proto-oncogene promoters, origins of replication, untranslated regions (UTRs) of mRNA, and at telomeres, the ends of human chromosomes (Balasubramanian et al., 2011; Bryan, 2020; Bugaut & Balasubramanian, 2012; Capra et al., 2010; Patel et al., 2007). Further, G4 structures have also been identified in viruses such as HIV-1 and SARS-CoV-2, where they are supposed to regulate viral gene expression and replication (Perrone et al., 2013; Zhao et al., 2021).

1.3.1 G-QUADRUPLEXES AT TELOMERES

Telomeres are structures made of nucleoprotein that protect the DNA of chromosomes from degradation and prevent its natural ends from being recognized as DNA damage. In humans, telomeres comprise a repeated short G-rich sequence, d(`TTAGGG`)$_n$, which is synthesized by telomerase and ends with a 3' single-stranded DNA overhang. As normal human cells divide, telomeres progressively shorten, leading to growth arrest upon telomere uncapping, also known as replicative aging (Zakian, 1995). In contrast, in numerous cancer subtypes, telomerase is overactive, contributing to the ability of these cells to proliferate indefinitely since their chromosomal shortening is absent (Moye et al., 2015).

It has been demonstrated that truncations of d(`TTAGGG`)$_n$ motif can form G4 structures *in vitro* by X-ray crystallography and nuclear magnetic resonance (Parkinson et al., 2002; Wang & Patel, 1993). Interestingly, telomerase seems to recognize the single-stranded telomeric DNA but not the G4 structure. This has led scientists to explore the development of small molecules as selective telomeric G4 ligands to induce apoptosis in cancer cells, which could be a promising new approach to cancer therapy (Mergny et al., 2001). Moreover, telomeric DNA binds several proteins, including the *Shelterin* complex, which protects chromosome ends from being recognized as DNA breaks (De Lange, 2005). This helps maintain genome integrity, ensuring accurate transmission of genetic material between cells. Competing small molecules induce a DNA damage response, promoting senescence in cancer cells (Rodriguez et al., 2008).

The transcription of the telomeric C-rich strand, d(`CCCTAA`)$_n$, in chromosomes, produces telomeric repeat-containing RNA (*TERRA*), which has a canonical G-rich motif of sequence r(`UUAGGG`)$_n$ (Azzalin et al., 2007; Hirashima & Seimiya, 2015). *TERRA* regulates telomerase activity, protects chromosome ends from degradation, and participates in heterochromatin formation and homologous recombination (Montero et al., 2016; Redon et al., 2010). *TERRA* can form G4 structures that could be used as potential therapeutic targets. *TERRA* RNA G4s were confirmed in living cells by Xu et al.

using a pyrene probe (Xu et al., 2010). They were also detected via optical imaging and sequencing-based methods (Kwok et al., 2018). Interestingly, telomeric rG4 structures may be more valuable than DNA G4s, as they are required for telomere heterochromatin formation in all cancer cells, including those that don't use telomerase to elongate their telomeres (ALT-positive tumors) (Bryan et al., 1995). These tumors display higher *TERRA* accumulation levels that appear to play a direct role in telomere elongation (Silva et al., 2021). For this reason, the design of small molecules targeting *TERRA* G4s is attracting ever-increasing attention.

1.3.2 G-QUADRUPLEXES IN ONCOGENE PROMOTERS

Gene promoters are particularly enriched in G4 motifs, which regulate transcription in multiple genes (Huppert & Balasubramanian, 2005). Several G4 structures from gene promoters have been investigated, combining several spectroscopic and biophysical techniques, including the B-cell lymphoma 2 (*Bcl-2*) (J. Dai et al., 2006), the hypoxia-inducible factor 1-alpha (*HIF-1α*) (De Armond et al., 2005), the receptor tyrosine kinase (*c-Kit*) (Fernando et al., 2006; Kotar et al., 2019; Kuryavyi et al., 2010; Phan et al., 2007; Rankin et al., 2005), the Kirsten rat sarcoma virus (*KRAS*) (Cogoi et al., 2004), the cellular myelocytomatosis (*c-Myc*) (Ambrus et al., 2005; Siddiqui-Jain et al., 2002; Simonsson et al., 1998), the platelet-derived growth factor-A (*PDGF-A*) (Qin et al., 2007), the vascular endothelial growth factor (*VEGF*) (Sun et al., 2005), and the rearranged during transfection (*RET*) (Tong et al., 2011) genes. While human telomeres consist of tandem repeats, the G-rich sequences within gene promoters are made up of G-tracts with varying numbers of guanines and intervening bases. Thus, in solution, they can potentially form inter- or intra-molecular structures, multimeric, or even their mixtures.

G4-forming sequences in promoter regions act as transcription modulators of oncogenes by interacting or recruiting G4-related proteins (G4RP) (Sanchez-Martin, 2023). Several transcription factors have been identified as G4RPs with different selectivity. In addition to duplex DNA, Myc-associated zinc finger (MAZ) and the transcription factor Sp1 bind to *KRAS* and *c-Kit** G4s, respectively, whereas poly(ADP-ribose) polymerase 1 (PARP-1) shows specificity for *KRAS* G4 (Cogoi et al., 2010; Raiber et al., 2012). This G4 structure is also a valuable target of the ribonucleoprotein HNRNP A1, able to bind and unfold G4s in *KRAS* promoter regions (Chu et al., 2016; Ferino et al., 2021). Yin Yang 1 (YY1), another zinc finger protein, can bind two close-spaced DNA G4s and dimerize, bringing the two pieces of DNA together (Li et al., 2020). Nucleolin (NCL), a nuclear phosphoprotein, preferentially stimulates and induces the folding of G4s containing long loops (Lago et al., 2017). NCL and nucleoside diphosphate kinase (NM23-H2) act as a repressor and an activator of the *c-Myc* oncogene, respectively inducing the *c-Myc* G4 formation and promoting its unfolding (González et al., 2009; Thakur et al., 2009). On the other hand, NM23-H1 and NM23-H2 proteins have been shown to repress the transcriptional activity of the *PDGF-A* gene via functional interactions with promoter elements (Ma et al., 2002). Further, nucleophosmin (NPM) and tumor suppressor protein 53 (TP53) can bind *c-Myc* G4 (Petr et al., 2016; Scognamiglio et al., 2014). Interestingly, recently, Vimentin (VIM), an intermediate filament protein, was identified for the first time as a binder of multimeric G4s from *c-Kit* oncogene (Ceschi et al., 2022). With detailed structural information on the binding of G4s with corresponding interactors, it should be possible to confidently design a new class of small molecules that can effectively interfere with this interaction and precisely modulate the transcription of oncogenes.

1.3.3 TARGETING G-QUADRUPLEXES WITH SMALL MOLECULES

It has been found that small molecules can bind to and stabilize G4s. Therefore they can potentially be useful in the regulation of physiological and pathological processes, including carcinogenic processes, in which G4s are involved. This knowledge has led to the development of numerous small molecules studied over the last decade to evaluate their ability to stabilize G4 structures. These findings suggest that targeting G4s with small molecules could be a promising approach to treating cancer.

The natural product telomestatin, the triazine 12459, and pyridostatin (PDS) have been identified as potent telomeric G4 stabilizers, competing with *Shelterin* protein POT1 for the binding and triggering DNA damage response and cellular senescence (**Figure 1.14**) (Gomez et al., 2003, 2006; Rodriguez et al., 2008; Sun et al., 1997; Temime-Smaali et al., 2009). Interestingly, PDS cannot discriminate between RNA and DNA G4s, while the carboxy-pyridostatin derivative prefers rG4 (**Figure 1.14**) (Di Antonio et al., 2012).

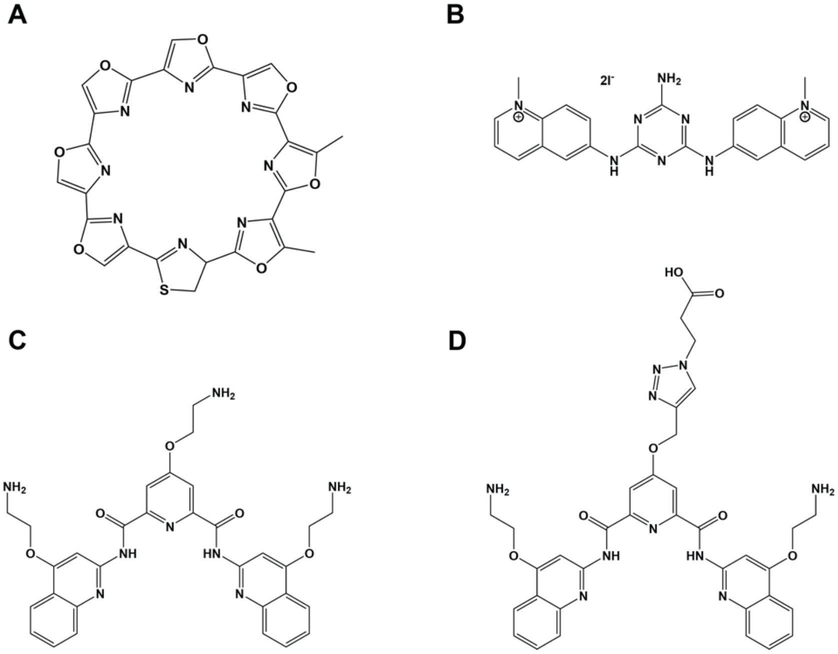


Figure 1.14 Chemical structure of **A**) telomestatin, **B**) triazine 12459, **C**) pyridostatin, and **D**) carboxy-pyridostatin.

RHPS4 (**Figure 1.15**) also induces telomere uncapping, leading to a telomere-induced DNA damage response (Leonetti et al., 2004; Phatak et al., 2007; Salvati et al., 2007).

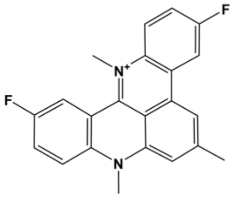


Figure 1.15 Chemical structure of RHPS4.

The tri-substituted acridine BRACO-19 (**Figure 1.16**), one of the very first promising synthetic G4-targeting small molecules, acts as a telomerase inhibitor (Gunaratnam et al., 2007).

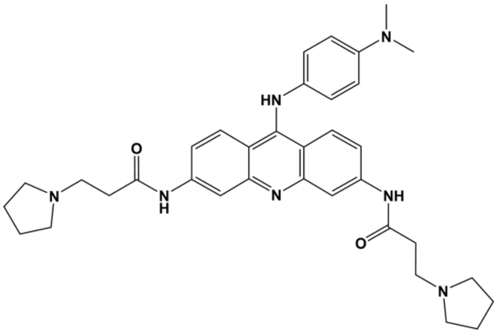


Figure 1.16 Chemical structure of BRACO-19.

Finally, a cationic porphyrin derivate TMPyP4 downregulates the transcription of *c-Myc, KRAS,* and *PDGF-A* oncogenes by stabilizing corresponding G4 structures in the promoter regions (Cogoi & Xodo, 2006; Qin et al., 2007; Siddiqui-Jain et al., 2002).

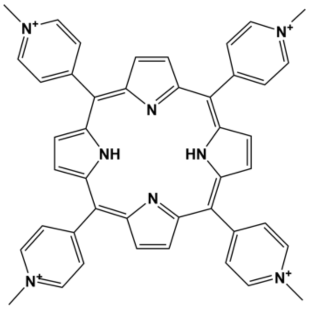


Figure 1.17 Chemical structure of TMPyP4.

1.4 BIOLOGICAL RELEVANCE OF I-MOTIFS

Despite previous assumptions that iMs are not physiologically relevant due to their supposed stability only in slightly acidic conditions, a solid foundation of data indicates that these structures are detectable *in vitro* at close to neutral pH. Recently, *in-cell* NMR experiments provided evidence for the *in vivo* formation of iMs (Dzatko et al., 2018), while a selective antibody fragment (iMab) was developed to detect iMs in cells' nuclei (Zeraati et al., 2018).

Although G4-forming and iM-forming sequences are complementary, their stability and dynamics vary significantly depending on the sequence. Initial research was primarily directed towards C-rich strands complementary to the known G4-forming promoter sequences. However, iM structures, like G4s, are conserved across species and cluster in telomeres and gene regulatory regions, thus implying their involvement in various processes, such as telomere homeostasis and transcription.

The development of carboxylated single-wall carbon nanotubes, able to interfere with telomere homeostasis in cancer cells by stabilizing the iMs (Y. Chen et al., 2012), paved the way for developing telomeric iMs DNA binders. Further, the iM formation is uppermost at the late G1 phase, when transcription levels are high (Zeraati et al., 2018). In breast cancer cell lines, it is possible to modulate the transcription of the *Bcl-2* oncogene, modulating the folding of the corresponding iM-forming sequence with two different molecular tools. In the presence of a stable iM structure, *Bcl-2* transcription levels increased, while destabilizing the iM reduced them (Kendrick et al., 2014). Further, the ribonucleoprotein HNRNP LL also binds to the *Bcl-2* iM and enhances corresponding gene expression (Kang et al., 2014). On the other hand, when the *c-Myc* iM structure is formed, it leads to lower transcriptional activity than when it is unfolded (Sutherland et al., 2016). Thus, also iM structures can regulate transcription and play different biological roles, making them ideal targets for interactor such as small molecules, inhibiting oncogenic signaling and tumor growth (Y. Zhang et al., 2018).

The iM structures also have potential applications in the field of sensors and logic devices due to the strong dependence of the structure's stability on pH (Dong et al., 2014). Nanostructures using iM are more practical for various applications as they offer superior conformational flexibility in response to chemical stimuli like changes in pH levels, in contrast to those relying on the canonical Watson-Crick base pairing that have sensitivity limitations (Debnath et al., 2019). Due to their distinctive structure and chemical properties, iMs have garnered attention as possible pH-responsive switches in nanotechnology. They also hold potential as probes for monitoring pH fluctuations in living cells and tumor microenvironments with abnormal pH levels (Alba et al., 2016; Webb et al., 2011). Given their significance, numerous research teams are devoted to comprehending the factors that impact this non-canonical fold.

CHAPTER 2
GENERAL METHODOLOGIES

2.1 CIRCULAR DICHROISM SPECTROSCOPY

Circular dichroism spectroscopy is a light absorption technique used to study the secondary structure of biomacromolecules. It measures the difference in absorbance between right- and left-circularly polarized light and is highly sensitive to conformational changes due to temperature, pH, and solvent variations. It requires low sample amounts and provides sensitive detection. Electromagnetic radiation exhibits wave-like behavior and consists of two components: the magnetic ($\vec{B}$) and electric ($\vec{E}$) field vectors, perpendicular to each other and traveling through space at the speed of light. To simplify the description, it is optimal to consider only one component, namely the E-component. This approach is justified because the two components are invariably perpendicular. A typical light source comprises various emitters with randomly oriented waves $\vec{E}$ (unpolarized light). However, it gets polarized when the light passes through a polarizer, an optical filter that allows only the light waves that vibrate in a single plane while blocking the others (**Figure 2.1**).

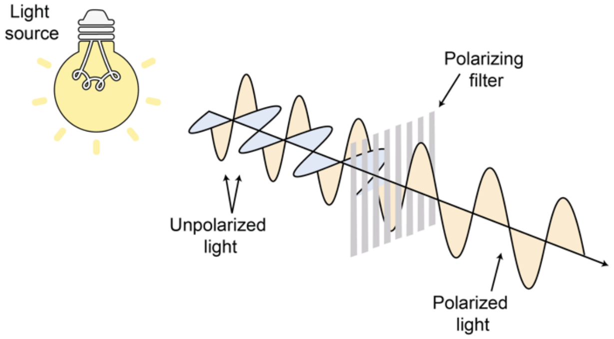


Figure 2.1 Schematic illustration of unpolarized and polarized light.

Linearly polarized light has a fixed direction of $\vec{E}$ and oscillates in only one plane perpendicular to the direction of propagation. On the other hand, circularly polarized light rotates $\vec{E}$ around the propagation axis while maintaining a constant magnitude, and it can either be left-polarized (L-CPL) or right-polarized (R-CPL) (**Figure 2.2**).

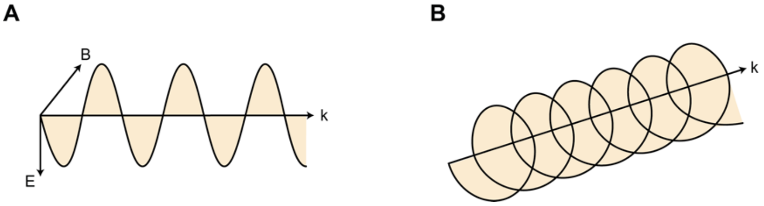


Figure 2.2 Schematic illustration of **A**) linearly polarized and **B**) circularly polarized light.

The phenomenon of differential radiation absorption, polarized in two directions concerning frequency, is called dichroism. When plane-polarized light is used, this phenomenon is called linear dichroism. Conversely, when circularly polarized light is used, the phenomenon is called circular dichroism. When light is circularly polarized, it rotates in opposite directions. If this type of light comes through an optically active material, the speed at which it passes through the material will differ for each polarization. This phenomenon is called circular birefringence or optical rotation, and it can significantly affect how light behaves in the material. At some wavelengths, right- and left-circularly polarized light will also be absorbed to different extents due to differences in extinction coefficients for the two polarized rays. This effect is known as circular dichroism (CD):

$$\mathrm{CD} = \mathrm{A_L} - \mathrm{A_R}$$

in which $\mathrm{A_L}$ and $\mathrm{A_R}$ are the absorbance of the left- and right-polarized light, respectively. The absorbance is obtained by the Lambert-Beer law:

$$\mathrm{A} = \varepsilon \cdot l \cdot \mathrm{c}$$

in which ε is the wavelength-dependent molar absorptivity coefficient ($\mathrm{M^{-1}}$ $\mathrm{cm^{-1}}$), l is the optical path length (cm), and c is the sample concentration (M). During a CD measurement, l and c are constants thus, the CD signal can be expressed as the difference between molar absorptivity coefficients:

$$\mathrm{CD} = \mathrm{A_L} - \mathrm{A_R} = (\varepsilon_\mathrm{L} - \varepsilon_\mathrm{R}) \cdot l \cdot \mathrm{c}$$

Asymmetric chromophores or symmetric chromophores in asymmetric environments interact differently with right- and left-circularly polarized light, resulting in two related phenomena. Thus, an optically active sample introduces a difference between the magnitude of the two components, making the light elliptically polarized. The CD signal is measured in units of ellipticity, defined as the tangent of the ratio of the minor to the major elliptical axis. The difference between the molar absorptivity coefficients of the two light components is directly proportional to the ellipticity, θ:

$$\theta = \frac{\pi \cdot \lambda \cdot (\varepsilon_\mathrm{L} - \varepsilon_\mathrm{R})}{l}$$

in which l is the optical path (cm), and λ (nm) is the wavelength of the incident radiation. A CD spectrum records the extent of the light's elliptical polarization about wavelength. The molar ellipticity ([θ]) can be used to compare spectra with different concentrations:

$$[\theta] = \frac{\theta}{l \cdot \mathrm{c}} = 32.98\ \Delta\varepsilon\ \text{degrees}\ \mathrm{M}^{-1}\mathrm{cm}^{-1}$$

CD experiments are conducted on a spectrometer, also known as a spectropolarimeter. It is a specialized form of an absorbance spectrophotometer, which selects linearly polarized light of a specific wavelength using a monochromator. The light then passes through a photoelastic modulator (PEM), which is a silica block attached to a piezoelectric element and is converted into circularly polarized light. An alternating drive voltage is applied to the PEM, causing the polarization state of the transmitted light to alternate between R-CPL and L-CPL (**Figure 2.3**). In the case of an optically active sample, the two components of light will be absorbed and detected at different extents. The spectrometer will measure this difference at multiple wavelengths and produce a CD spectrum.

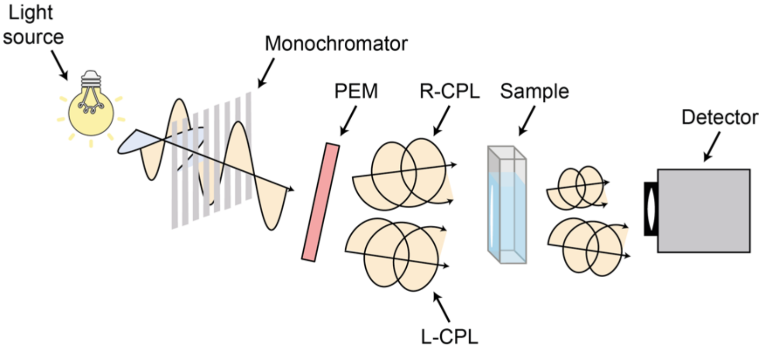


Figure 2.3 Schematic illustration of a CD spectrometer.

CD spectrum is similar to an absorption spectrum with positive and negative bands caused by differences in molar absorptivity coefficients.

Chromophores are molecules that respond to CD and have inherent dissymmetry due to their structural properties or chiral centers. The CD is applicable only to UV-Vis absorbing samples, and each sample's CD spectrum is obtained by subtracting the buffer baseline.

Herein, CD spectroscopy has been widely employed, using standard protocols, to verify the secondary structure adopted (or induced) by oligonucleotide sequences or peptides (PAPERS I, II, III, IV, V, and VI), and to check the ability of the investigated compounds to affect their folding topology or thermal stability (PAPERS II, III, IV, and V).

2.2 FLUORESCENCE SPECTROSCOPY

Fluorescence spectroscopy is an analytical technique that uses UV-Vis radiation to excite molecules in a sample, causing them to emit fluorescent light. This light is analyzed to provide information about the sample's composition and properties. Fluorescence spectroscopy has many advantages; it is a rapid, relatively inexpensive, and non-destructive technique. Although its limits of sensitivity depend on the intrinsic fluorescence of the sample, this technique is usually able to detect minimal concentrations of analytes (10^{-6} M) and has a sensitivity higher than absorption spectroscopy. Moreover, the emission of fluorescence radiation is approximately linear in a wide range of concentrations without the necessity of sample dilution.

Molecules are made up of atoms that contain electrons, which exist in different energy levels called electronic states. When an external energy source, such as electromagnetic radiation, is introduced, the electron can absorb energy and transition from its lowest energy state, known as the ground state, to a higher energy state, known as the excited state. However, the molecule's stay in the excited state is temporary, as the electron eventually relaxes back to the ground state, emitting energy in the form of electromagnetic radiation. This spontaneous emission is a fundamental process known as fluorescence. In addition to spontaneous emission, molecules can also undergo stimulated emission. This occurs when the molecule interacts with electromagnetic radiation, causing it to emit radiation in the same frequency and phase as the incident radiation. The Jablonski diagram is a graphical representation that illustrates the various electronic states of a molecule and the transitions between them, thus providing an explanation for the phenomenon of spontaneous emission. (**Figure 2.4**):

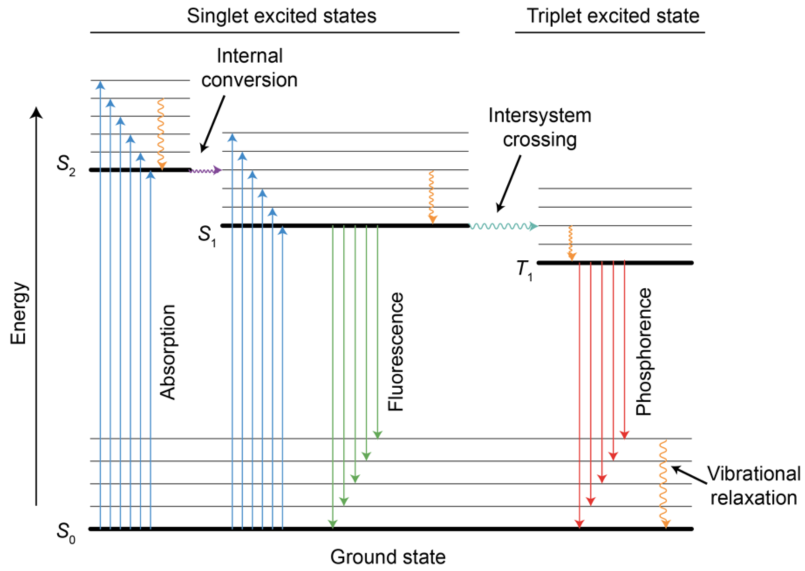


Figure 2.4 Schematic illustration of a Jablonski diagram.

When a molecule absorbs UV-Vis radiation, it most likely undergoes a transition from the singlet ground state (S_0) to the first singlet excited state (S_1) while conserving spin multiplicity. The molecule can be in any vibrational level when it reaches the excited electronic state, but it tends to return to the ground state. The first step in this process is transitioning from the excited to the ground vibrational level, known as vibrational relaxation. This process involves the dissipation of energy from the molecule to its surroundings. Fluorescence is the emission of electromagnetic radiation caused by the movement of an electron from the first excited singlet state (S_1) to the ground state (S_0). On the other hand, intersystem crossing occurs when an electron moves from the first excited singlet state (S_1) to the first excited triplet state (T_1). Phosphorescence is the emission of radiation due to the movement of an electron from the first excited triplet state (T_1) to the ground state (S_0). During this process, the molecule may return to any vibrational level, leading to a vibrational relaxation. The emission bands of a molecule have longer wavelengths compared to those of its excitation bands. This is because of the energy loss caused by non-radiative phenomena. As a result, the fluorescence spectrum is similar to the absorption spectrum, but it is shifted to higher wavelengths, and this shift is called the Stokes shift.

Fluorescence quantum yield (φF) estimates the amount of energy recovered via the fluorescence process. It's the ratio of the emitted and absorbed photon and ranges between zero and one. A high-fluorescence molecule has a ratio close to one, and a low-fluorescence molecule has a ratio close to zero. Fluorophore is the term used for a fluorescent molecule.

A fluorescence resonance energy transfer (FRET) occurs when a donor fluorophore is excited and transfers its energy to an acceptor fluorophore, resulting in the emission of light by the acceptor. Dark quenchers are alternatives to fluorescent FRET acceptors as they relax from the excited state to the ground state non-radiatively. The efficiency of energy transfer is dependent on the distance between the donor and acceptor, making FRET a valuable tool for studying molecular interactions and conformational changes.

The fluorescence spectrometer comprises three main components: an excitation source, a detection system, and a data acquisition and analysis module. The excitation source provides the light that excites the fluorescent sample.

It can be a lamp, a laser, or a light-emitting diode (LED), depending on the sample and the required wavelength of light. The detection system collects the emitted light from the sample and measures its intensity and wavelength. The most common types of detectors are photomultiplier tubes (PMTs) and charge-coupled devices (CCDs). The data acquisition and analysis module collect and analyzes the data generated by the excitation source and the detection system (**Figure 2.5**).

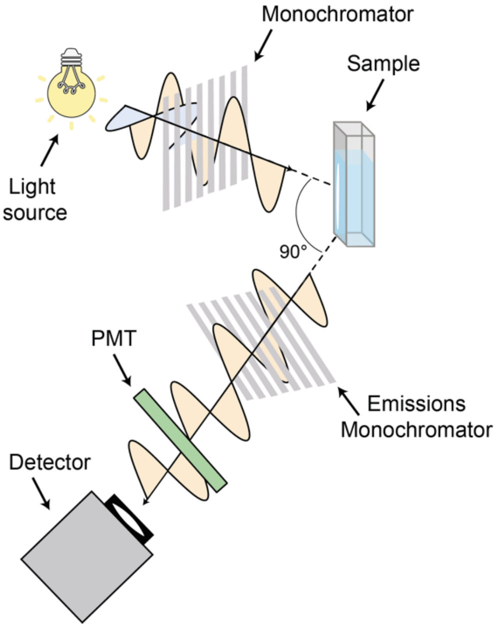


Figure 2.5 Schematic illustration of a fluorescence spectrometer.

Herein, fluorescence spectroscopy has been employed to obtain information about: *i*) the interaction between DNA/RNA G4 and ligands directly (PAPER IV), or by means of fluorescent probes (PAPERS II, III, and V); *ii*) the oligonucleotide folding using a fluorescent selective dye (PAPER VI); *iii*) G4 *vs.* duplex structures selectivity through FRET competition melting assays (PAPERS II, and III).

2.3 NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

NMR spectroscopy is a spectroscopic method that requires a strong static magnetic field. It's based on the interaction between energy and matter. Isidor Rabi described NMR in 1938 and was awarded the Nobel Prize in Physics for his work. Felix Bloch at Stanford and Edward Mills Purcell at the Massachusetts Institute of Technology demonstrated NMR in condensed matter (water and paraffin, respectively) in 1945 (Bloch et al., 1946; Purcell et al., 1946). Both Bloch and Purcell were jointly awarded the Nobel Prize for physics seven years later. The significance of NMR in chemistry was not recognized until 1950, when the chemical shift was discovered (Proctor & Yu, 1950). Chemistry has been the field with the most interest in NMR in recent years.
Nuclear magnetic resonance is a fascinating property that occurs within the nucleus of an atom. The nucleus is comprised of protons and neutrons, and it has a unique property called nuclear spin (*I*), which causes it to behave like a miniature bar magnet. Interestingly, not all nuclei have this property. Specifically, nuclei with even mass and even charge numbers do not have the nuclear spin property (I=0) and are therefore labeled as "NMR inactive" or “NMR silent”. However, specific nuclei, such as hydrogen (^{1}H), carbon (^{13}C), fluorine (^{19}F), and phosphorus (^{31}P), possess the nuclear spin property (I=½), which is essential for NMR analysis. When the nuclei are not affected by an external magnetic field (B_0), the spins are randomly oriented in all directions. However, when nuclei with *I* different from zero are placed in a magnetic field, they assume various orientations corresponding to specific energy levels (**Figure 2.6**). The number of orientations depends on the value of *I*, which equals $2I$+1.

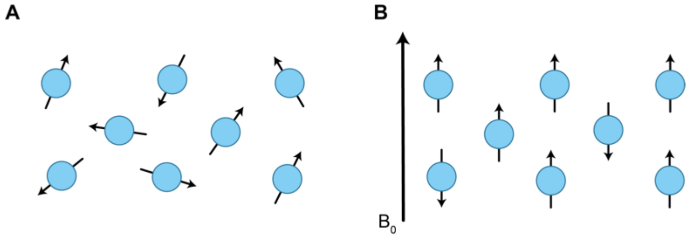


Figure 2.6 Schematic illustration of **A**) randomly oriented and **B**) aligned spins in the absence and presence of an applied magnetic field (B_0), respectively.

The most critical nuclei in biomolecular NMR studies are ^{1}H, ^{13}C, ^{15}N, and ^{31}P. Among these nuclei, ^{1}H is the most sensitive, followed by ^{31}P; the natural abundance is near 100% for both. The proton (^{1}H) is the most abundant NMR nucleus, and it has I=½; therefore, when B_0 is applied, these nuclei can assume two possible orientations, α parallel (I=+½, ↑) or β antiparallel (I=-½, ↓), each corresponding to an energy level. The difference in the energy levels is equal to:

$$\Delta E = \gamma \cdot B_0$$

where γ is the constant gyromagnetic ratio for a given nucleus ($2.6752 \cdot 10^8$ $s^{-1}T^{-1}$ for protons). This means that a stronger magnetic field corresponds to a higher gap in energy levels, resulting in greater sensitivity in the NMR experiment. Protons are distributed between the two energy states according to the Boltzmann distribution:

$$\frac{N_\alpha}{N_\beta} = \exp\left(\frac{\Delta E}{k \cdot T}\right)$$

in which N_α and N_β represent the protons populations in the lower and upper energy levels, respectively, k the Boltzmann constant, and T the temperature. The number of nuclei in the two states is not equal. A slight excess of protons will occupy the lower energy state (α) as it is more favorable energetically.

This results in the net magnetization M_0, aligned with the applied magnetic field B_0. When a radiofrequency (RF) pulse is used, the nuclei absorb energy and transition to higher energy levels:

$$\Delta E = h \cdot \nu$$

where h is the Planck constant and ν is the excitation pulse frequency that induces the transitions between the levels. This frequency is referred to as Larmor frequency and depends both on the nucleus and the magnetic field:

$$\nu = \frac{\gamma \cdot B_0}{2 \cdot \pi}$$

The NMR signal becomes detectable only when the sample is in an equilibrium state in the static magnetic field. The net magnetization vector has no component on the xy plane, where the detector coil detects the signal. The duration of the pulse is usually measured in microseconds (μs). When the pulse brings the magnetization on the xy-plane, it is called a 90° pulse. On the other hand, a 180° pulse flips the net magnetization on the negative z-axis. When the radiofrequency is switched off, the system will return to equilibrium. This return to equilibrium is referred to as relaxation, and it causes the NMR signal to decay with time, producing the observed free induction decay (FID). The NMR signal is then Fourier transformed to be converted in the frequency domain (**Figure 2.7**).

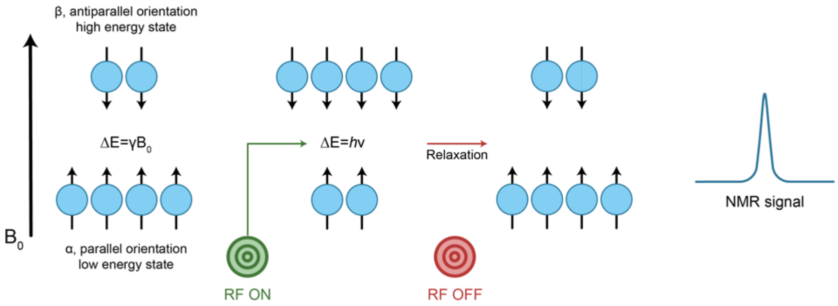


Figure 2.7 Schematic illustration of an NMR experiment.

2.3.1 CHEMICAL SHIFT AND COUPLINGS

Atoms in a molecule don't experience the same magnetic field due to their positioning, known as shielding. Electrons surrounding the nucleus rotate when an external magnetic field (B_0) is applied, creating a local magnetic field (B_{loc}) that can oppose the external field. This results in a slightly reduced field called B_{eff}. The shielding constant (σ) indicates the density and distribution of the electronic cloud around the nucleus. It ranges from 10^{-6} to 10^{-3}, depending on the nucleus's weight. The presence of functional groups near a nucleus can change the σ value. An electronegative atom can reduce electronic cloud density, causing a de-shielding effect and increasing nucleus frequency resonance. Considering that B_{loc} is equal to the product of B_0 and σ, according to the Lenz rule, B_{eff} is given by:

$$B_{eff} = B_0 - B_{loc} = B_0 - B_0\sigma = B_0(1 - \sigma)$$

Therefore, the relationship between the degree of shielding and the resulting resonance frequency is equal to:

$$\nu = \frac{\gamma \cdot B_0}{2 \cdot \pi} \cdot (1 - \sigma)$$

NMR spectroscopy is a powerful tool that reveals how protons in distinct chemical environments resonate at varying frequencies in response to magnetic fields, known as chemical shift (δ). Valuable information can be extracted from NMR experiments by analyzing the frequencies of different nuclei independently from the magnetic field utilized. Thus, a conventional way to calculate δ is in terms of ν as the frequency of the observed nucleus and ν_{ref} as the frequency of a reference compound:

$$\delta = \frac{\nu \cdot \nu_{\mathrm{ref}}}{\nu_{\mathrm{ref}}} \cdot 10^{6}$$

Chemical compounds with maximum shielding exhibit higher σ values than nuclei typically analyzed by NMR. The last equation converts chemical shift frequencies into parts per million (ppm), allowing NMR signals to be displayed on a new axis, ensuring that a given spin always shows the same value, regardless of the magnetic field used. In this system, the reference compound has δ value 0, while sample resonances have positive δ values. Most used reference compounds in NMR spectroscopy are tetramethyl silane (TMS) for organic solvents, the sodium salt of trimethylsilyl propionic acid (TSP), or dimethyl silapentane sulfonate (DSS) for aqueous solutions.

NMR experiments provide chemical information beyond just chemical shifts. Another valuable feature is the J coupling or scalar coupling. Coupling (scalar coupling) is another essential feature caused by the interaction of magnetic moments of nuclei through space (dipolar coupling) or chemical bonds (scalar coupling). The dipolar coupling can be neglected due to rapid molecular tumbling, whereas scalar coupling is visible in the NMR spectrum. NMR signal splitting occurs when one atom influences another, which can happen in both directions. However, to observe signal splitting, the interacting nuclei must be bonded in proximity or be oriented in specific optimal configurations.

2.3.2 DETECTION AND FOURIER TRANSFORM

After the nuclei are excited by a radiofrequency pulse, the magnetization flips on the xy plane (90° pulse). This causes the spins to rotate clockwise (-ν) and counter-clockwise (+ν). A detection system with two simultaneous channels on each axis often distinguishes between spins with different frequencies. This is known as quadrature detection. The nuclei then return to their equilibrium distribution between the two energy levels through relaxation, which has specific relaxation times for different nuclei. This generates the free induction decay (FID), a time-domain representation of the frequencies of all the nuclei in the sample. To interpret the FID, a mathematical operation called Fourier transform is carried out to produce the frequency spectrum (**Figure 2.8**).

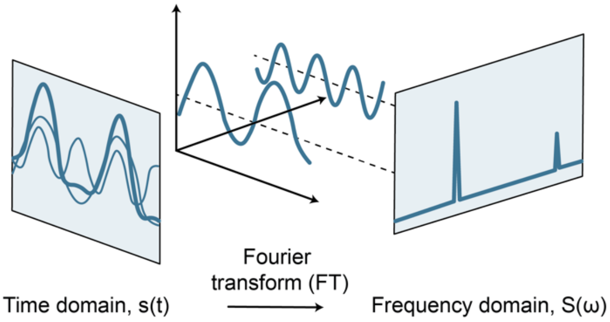


Figure 2.8 Illustration of the Fourier transformation from the time domain s(t) to the frequency domain.

2.3.3 PROCESSING TOOLS

Several processing steps need to be performed before and after the Fourier transform of the FID to improve the sensitivity and resolution of acquired data.

- Zero-filling is a technique to improve the digital resolution of NMR spectra by adding zeros to the FID data points. It does not improve the proper resolution, but only the apparent resolution. Zero-filling helps detect fine couplings that may not be visible due to low digital resolution. It is essential to avoid using zero-filling to increase digital resolution with very short acquisition time, as this may lead to baseline artifacts after the FT;
- Apodization is a spectral analysis technique that uses window functions like Lorentzian, Exponential, Gaussian, or Sine-bell to multiply the FID. Different window functions have varying effects on sensitivity, resolution, line broadening, and noise levels, depending on the analysis objectives;
- Phase correction is a process used to fix errors in spectra. Two main things cause these errors: *i*) a delay between the RF pulse and the receiver opening for FID acquisition; *ii*) the pulse may not excite all the nuclei equally due to off-resonance effects. An uncorrected spectrum shows a dispersive line shape and inverted signals. Zero-order and first-order phase corrections are used to fix this. Zero-order correction is independent of chemical shift, while the first-order correction is frequency-dependent and increases with distance from the reference signal;
- Baseline correction is crucial for accurately integrating and quantifying analytes in spectroscopic data. Distorted baselines caused by the first few data points in FID can hide small peaks, usually due to high signal amplification or incomplete RF electronics recovery. The most common correction method is fitting the baseline with a polynomial function and subtracting it from the spectrum.

2.3.4 2D NMR EXPERIMENTS

Two-dimensional (2D) NMR spectroscopy is one of the most widely used and powerful tools for studying biological macromolecules. A two-dimensional homonuclear spectrum consists of diagonal peaks where the frequency in both dimensions, ν_1 and ν_2, is the same. It also includes cross peaks outside the diagonal where the frequency differs in the two dimensions. Depending on the experiment, scalar couplings, dipolar couplings, or other information can be obtained from these cross-peaks. Two-dimensional spectra are helpful to solve problems of signal superposition that are present in 1D experiments. Homonuclear 2D NMR experiment is a technique used to study a compound's molecular structure and dynamics by measuring the interactions between nuclei of the same type. It involves the application of two radiofrequency pulses, one for excitation and the other for detection, to the sample containing the compound of interest. The experiment generates a two-dimensional spectrum that displays the frequencies of the two nuclei and their interactions. There are several 2D homonuclear experiments:

- COrrelation SpectroscopY (COSY) 2D NMR experiments can identify scalar-coupled proton pairs in a molecule. By correlating spin transitions from one nucleus to another, a 2D spectrum shows cross-peaks between scalar-coupled protons. This allows for the identification of spin systems and provides insights into the connectivity of protons in a molecule.
- TOtal Correlation SpectroscopY (TOCSY) 2D NMR spectroscopy identifies spin systems and determines connectivities between different spins in a molecule. It uses a radiofrequency pulse to transfer magnetization from a specific spin to all connected spins, resulting in a spectrum showing correlations between all spins in the molecule.

- Nuclear Overhauser Effect SpectroscopY (NOESY) 2D NMR experiment involves applying radiofrequency pulses to a sample containing a molecule of interest. The first pulse is typically a 90° pulse that excites all the spins in the sample. During the period between the first and second pulses, the nuclear Overhauser effect occurs, which results in the transfer of nuclear spin polarization between coupled nuclei. The second pulse is typically a 180° pulse that inverts the spins of the excited nuclei. The signal is detected and recorded as a function of the time delay between the first and second pulses, as well as the chemical shift of the resonances. The NOESY spectrum provides information about the proximity of different nuclei in a molecule and their spatial arrangement. Cross-peaks in the NOESY spectrum indicate that the corresponding nuclei are close to each other in space and may be involved in a particular interaction, such as a hydrogen bond or a van der Waals interaction. The cross-peak intensity is proportional to the degree of proximity between the nuclei. Overall, NOESY 2D NMR experiments are a powerful tool for studying the three-dimensional structure of molecules, particularly biomolecules such as proteins and nucleic acids.

Heteronuclear 2D NMR experiments involve using two different types of nuclei in the NMR process. Traditional NMR experiments typically observe only a kind of nuclei, usually protons. However, heteronuclear 2D NMR experiments correlate the signal from one type of nucleus with the signal from a different sort of nucleus. This allows for identifying the chemical connectivity between various nuclei in a molecule. The most used heteronuclei are ^{13}C, ^{15}N, and ^{31}P. Examples of heteronuclear 2D NMR experiments are:

- Heteronuclear Single Quantum Coherence (HSQC) 2D NMR experiments study the correlation between protons and heteronuclei directly connected with one bond. This technique provides information on the identity and location of individual atoms in a molecule, as well as the spatial arrangement of neighboring atoms;

- Heteronuclear Multiple Bond Correlation (HMBC) 2D NMR experiments determine the connectivity between protons and their attached heteronuclei (such as carbon, nitrogen, or phosphorus). This technique helps identify the long-range coupling between these nuclei, which can be challenging to detect using other NMR techniques. The resulting spectrum provides valuable information about a compound's molecular structure and connectivity.

Here, 1D ^{1}H NMR spectra were acquired for demonstrated G4-folding (PAPER VI). 1D ^{1}H NMR titrations experiments were performed to elucidate ligand binding mechanisms to DNA and RNA G4 structures (PAPERS V, and VII). Moreover, several 2D NMR homo- and heteronuclear experiments were carried out to in-depth investigation of the interaction of a G4/ligand complex (PAPER VII).

2.4 MICROSCALE THERMOPHORESIS

MicroScale Thermophoresis (MST) is a powerful approach for analyzing biomolecule interactions. It relies on the movement of molecules within a temperature gradient, a phenomenon called thermophoresis. The efficacy of this technique stems from its ability to detect even subtle variations in molecular properties such as size, charge, hydration shell, or conformation.

During an MST experiment, an infrared (IR) laser generates a microscopic temperature gradient. This results in the movement of molecules via thermophoresis, which can be detected and measured using either covalently attached or intrinsic fluorophores. By merging the accuracy of fluorescence detection with the flexibility and sensitivity of thermophoresis, MST serves as a versatile and dependable method for precisely analyzing binding events. This technique enables the precise examination of binding events within a small solution (i.e., µL) of nearly any molecule, regardless of size or physical properties (El Deeb et al., 2022; Jerabek-Willemsen et al., 2011). Therefore, MST can detect various events, such as the binding of small molecules to proteins, substrates to enzymes, or ligands to liposomes.

The interface between a molecule and solvent causes thermophoretic depletion. The Soret coefficient, S_T, measures the depletion of solvated biomolecules in a region of higher temperature created by a temperature difference, ΔT:

$$\frac{C_{hot}}{C_{cold}} = \exp(-S_T \Delta T)$$

Thermophoresis is a precise method for evaluating molecules' size, charge, and solvation entropy while maintaining a consistent buffer environment. When assessing the thermophoresis of a protein versus a protein-ligand complex, notable distinctions are often observed because of variations in size, charge, and solvation energy that arise during the binding process. Even if protein size or charge is only marginally modified by the binding event, MST can still identify it by measuring alterations in the solvation entropy of the molecules (Seidel et al., 2013).

The experimental setup involves coupling an IR laser to the fluorescence excitation/emission path with an IR dichroic mirror. The IR laser focuses on the sample using the same optics used for fluorescence detection. This setup enables precise observation of thermophoresis since the IR and visible optics are well-aligned (**Figure 2.9**). The IR laser enables localized, accurate, and reproducible solution heating. This is especially useful for analyzing a serial dilution containing 10 to 16 samples.

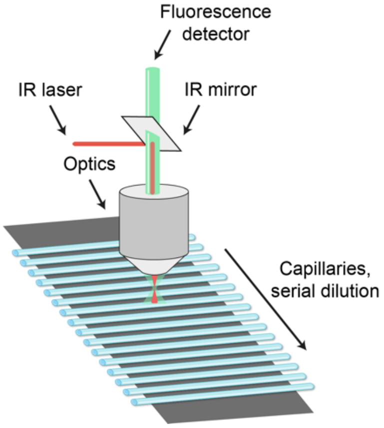


Figure 2.9 Schematic illustration of the instrumental setup for MST.

The movement of molecules is mainly limited by diffusion. By confining the temperature distribution to the millimeter scale, changes in thermophoretic properties can be measured in less than 30 s. Turning on the IR laser for less than a second results in heat dissipation reaching equilibrium, leading to a steady-state temperature increase (typically between 2 and 6 K). This temperature increase, known as the MST temperature jump (T-jump), triggers thermophoresis, causing a concentration gradient in the solution.

At the start, when the IR laser is turned off, the molecules are uniformly spread out, and a steady initial fluorescence reading is recorded. As soon as the IR laser is switched on, a T-jump takes place, causing a rapid shift in the characteristics of the fluorescent molecules. Then, the movement of the fluorescently labeled molecules out of the heated sample region can be monitored through thermophoresis. Upon deactivating the IR laser, an inverse T-jump occurs, followed by the back-diffusion of molecules, which is solely driven by mass diffusion (**Figure 2.10**).

In a typical MST experiment, the concentration of a fluorescently labeled molecule is kept constant. In contrast, the concentration of an unlabeled binding partner is varied, starting with a concentration of at least about ten times above the expected dissociation constant down to sub-stoichiometric concentration concerning the labeled molecule. The MST signal will detect the binding by quantifying the change in the normalized fluorescence (i.e., the amplitude of the MST signal). Changing fluorescence signal depending on the concentration of the target, expressed as the fraction of labeled molecules bound, x:

$$\mathrm{F_{norm}} = (1 - x) \cdot \mathrm{F_{norm}}(\mathrm{unbound}) + x \cdot \mathrm{F_{norm}}(\mathrm{bound})$$

where $\mathrm{F_{norm}}$(unbound) and $\mathrm{F_{norm}}$(bound) are the normalized fluorescence of unbound and bound to their targets labeled molecules, respectively. Differences in normalized fluorescence of the bound and unbound state, defined as the ratio $\mathrm{F_{hot}}/\mathrm{F_{cold}}$ (F-values correspond to average fluorescence values between defined "hot" and "cold" areas, respectively), will allow the determination of the fraction bound and thus, the equilibrium dissociation constant (K_D). Typically, $\mathrm{F_{norm}}$ values are multiplied by a factor of 1000, yielding a relative fluorescence change per thousand.

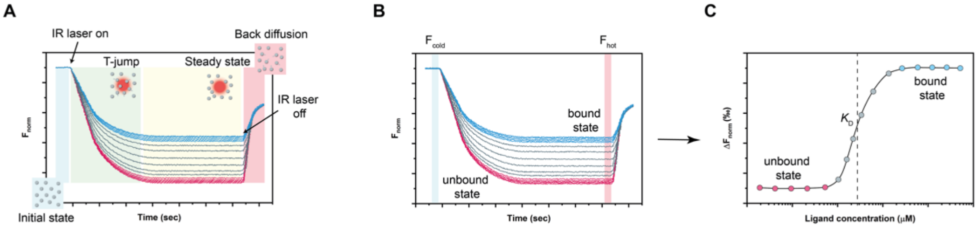


Figure 2.10 Schematic illustration of **A**) MST movement profiles, **B**) multiple MST traces recorded for different mixture ratios of target and ligand molecules, and **C**) the corresponding dose-response curve.

MST was employed to analyze oligonucleotide interactions (Baaske et al., 2010), protein–DNA interactions (Zillner et al., 2013), protein–protein interactions (Keren-Kaplan et al., 2013), and protein–small molecule (Gaffarogullari et al., 2013; Wienken et al., 2010).

Here, MST experiments were performed to obtain K_D values for the interaction of both small molecules and peptides for G4 structures (PAPERS II, III, and V).

2.5 GEL ELECTROPHORESIS

Electrophoresis is a process in which charged particles are separated and moved through a solution by an electric field. Many biological molecules such as amino acids, peptides, proteins, nucleotides, and nucleic acids contain ionizable groups, which means they can exist as charged species in solution, depending on the pH.

Gel electrophoresis is a widely practiced method for sorting, refining, and recognizing charged macromolecules in a solution. The electrophoretic setup consists of two electrodes, an anode, and a cathode, which are united by an electrolyte solution, encompassing the electrophoresis gel. The sample is deposited into suitable wells, and the molecules travel towards the appropriate electrode, determined by their charge. The electric force (F_e) is given by:

$$F_e = E \cdot q$$

in which E is the electric field, and q is the charge. A friction force, F_f, opposes to the motion generated by the electric field:

$$F_f = \gamma \cdot \nu$$

in which γ is friction coefficient and ν migration velocity. At constant speed:

$$F_e = F_f, \qquad \nu = \frac{E \cdot q}{\gamma}$$

The electrophoretic mobility (m) is defined as:

$$m = \frac{\nu}{E} = \frac{q}{\gamma}$$

It exclusively depends on the particle; in the case of a globular molecule, γ is equal to:

$$\gamma = 6 \cdot \pi \cdot \eta \cdot r$$

in which η is the medium viscosity and r being the molecule radius. Then, m is:

$$m = \frac{q}{6 \cdot \pi \cdot \eta \cdot r}$$

Thus, electrophoretic mobility of a molecule is determined by its charge, dimensions, shape, medium viscosity, and gel pore size. Electrophoresis separates molecules from a mixture based on their charge and size.

Here, native polyacrylamide gel electrophoresis (PAGE) was performed to discriminate between inter- or intramolecular iM structures (PAPER I).

CHAPTER 3
IDENTIFICATION OF NOVEL G-QUADRUPLEX-BINDING PROTEINS

With the aim of further investigating the biological functions of DNA G4s and their interaction network and then exploiting these interactions for the development of new potential therapeutic agents, in the study described in this CHAPTER we performed a spectroscopic characterization of four different G4-forming DNA sequences from oncogene promoters, which were then employed as baits in a proteomic "*fishing-for-partners*" approach to identify interacting proteins. Our protocol succeeded in the identification of some new DNA G4-binding proteins, with different selectivity among different G4s and/or versus unstructured G-rich strand (PAPER VI).

3.1 **DNA G-QUADRUPLEXES AS TARGETS FOR PROTEIN BINDING**

3.1.1 INTRODUCTION

As discussed above, the G4 structures can exert their biological functions in cells by interacting with several proteins, such as helicases or transcription factors (see **Paragraph 1.3**). Some proteins promote the formation and/or stabilization of G4 structures, while some other proteins exhibit unwinding activity (Brázda et al., 2014; Meier-Stephenson, 2022; Sanchez-Martin, 2023). From a drug discovery perspective, many of these G4RPs and/or their complexes with G4s are potential drug targets (Shu et al., 2022). Thus, it is crucial to discover as many as possible G4RPs in order to better elucidate the potential involvement of G4s in a range of biological processes.

Mostly of the already known RNA/DNA G4RPs have been discovered thanks to the development of different affinity enrichment protocols from nuclear or cytoplasmatic extracts, as well as native chromatin (Y. Dai et al., 2023). In a typical proteomic approach, already employed by our research group (Pagano et al., 2015), biotin-tagged G4 structures are employed as baits in solution and incubated with nuclear or cellular extracts. Streptavidin-coated magnetic beads isolate the labelled G4 structure together with its interactors, which can be eluted in denaturing conditions, digested, and identified by LC-MS/MS tandem analysis (González et al., 2009).

The death-associated protein 1 (DAP1), hypoxia-inducible factor 1α subunit (HIF-1α), juxtaposed with another zinc finger protein 1 (JAZF-1), and platelet-derived growth factor (PDGF-A) proteins are correlated with numerous cancer processes, and characterized by the presence of G-/C-rich complementary strands in their promoter regions. While C-rich strands have been deeply characterized in their ability to fold into iM secondary structures at neutral pH (Brazier et al., 2012; Dzatko et al., 2018; Wright et al., 2017), the complementary G-rich strands have received less attention except those of *HIF-1α* and *PDGF-A* (De Armond et al., 2005; Qin et al., 2007).

HIF-1α is one of the two subunits of the HIF-1 transcription factor responsible for the regulation of over 60 genes involved in many processes, including tumorigenic activity, mostly in hypoxia conditions (Erler et al., 2004). Its polypurine/polypyrimidine tract in the promoter region (-65 to -85) is adjacent to several putative transcription factor binding sites. The G-rich tract in solution is able to fold into a mixture of G4 structures with three stacked guanine tetrads and a parallel orientation of the strands in high-potassium conditions.

Mutagenesis of this region results in lower basal *HIF-1α* expression (De Armond et al., 2005).

The PDGF-A signaling pathway leads to proliferation, migration, angiogenesis, and metastasis and is particularly critical for pancreatic cancer progression (Hoffmann et al., 2008; Liu et al., 2011). Its gene proximal 5'-flanking region contains one nuclease hypersensitive element (NHE) that is critical for *PDGF-A* gene transcription and has a G4-forming sequence composed of five G-tracts in the -82 to -47 region. This can fold in two stable intramolecular parallel G4 structures that are potentially involved in transcriptional regulation. The ligand-mediated stabilization of the major G4 structure can silence *PDGF-A* expression (Qin et al., 2007).

DAP1 is a highly conserved phosphoprotein, associated with a pro-apoptotic function (Jia et al., 2014). It is correlated with disease progression and long-term survival of patients with colorectal cancer, while a low *DAP1* expression is associated with clinicopathological parameters of human breast cancer (Jia et al., 2014; Lehmann et al., 2002).

Finally, JAZF-1 is a zinc finger protein associated with tumor progression and type 2 diabetes (Liao et al., 2019; Prokunina-Olsson et al., 2010; Sung et al., 2017). Its overexpression is correlated with enhanced prostate cancer cell proliferation, migration, and invasion via regulating JNK/Slug signaling, even if its molecular mechanism has not yet been clarified (Sung et al., 2017).

In this study, *DAP*, *HIF-1α*, *JAZF-1*, and *PDGF-A* G-rich sequences from the corresponding oncogene promoters were investigated *in vitro* for their ability to form G4 structures. Then, their biotin-labeled forms were used as baits in solution to identify putative G4RPs from the nuclear extracts of osteosarcoma cancer cells (U2OS) with a "*fishing-for-partners*" classical proteomic approach. This protocol allowed us to identify 60 putative G4RPs with different selectivity for G4s over an unfolded G-rich oligonucleotide (used as a control in the *fishing* protocol), and/or for different G4s.

3.1.2 RESULTS AND DISCUSSION

BIOPHYSICAL INVESTIGATION OF THE SELECTED G4-FORMING SEQUENCES. To demonstrate that *HIF-1α*, *PDGF-A*, *JAZF-1*, and *DAP* G-rich sequences fold into G4 structures under the solution conditions employed in pull-down assays (i.e., 20 mM HEPES buffer at pH 6.6 containing 150 mM KCl, and 1 mM EDTA), a combination of spectroscopic techniques (CD, fluorescence, and NMR) has been employed. Furthermore, a scrambled G-rich sequence (*scrG*) used as a negative control has also been characterized (**Table 3.1**).

Table 3.1 List of here investigated G-rich sequences.

Name	Sequence (5'-3')	Length
DAP	GGGGGCGGGGGCGGGGGCGGGGGCGGGG	29-mer
HIF-1α	GCGCGGGGAGGGGAGAGGGGGCGGGAGCGCG	31-mer
JAZF-1	GGGGGGAGGGCGGGGGCGGGGGCGGGGGGGG	31-mer
PDGF-A	GGGGGGGGGGGGGCGGGGGCGGGGGCGGGGGAGGGGCGCGG	41-mer
scrG	GTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	30-mer

The Thioflavin T (ThT), a small molecule employed as fluorescence dye, is currently used as a selective G4 probe for their recognition. Indeed, when ThT binds G4 structures in solution, a severe fluorescence enhancement (FI) at 487 nm is observed compared to its unbounded form (FI_0) (De La Faverie et al., 2014). To avoid ambiguities, we also employed two different well-known G4-forming sequences, *AS1411* and *c-Kit2*, and a *hairpin* duplex as positive and negative controls, respectively. *AS1411* is characterized by a highly structural polymorphism in solution, with at least eight different G4 conformations (Bates et al., 2017). On the contrary, *c-Kit2* G4, here standing for a mutant *c-Kit2-G12T/G21T* sequence, is characterized by a unique monomolecular parallel topology in low-potassium solution conditions (Kuryavyi et al., 2010). The induced fluorescence of ThT in the presence of G4s is very sensitive to the stoichiometry of the interaction as well as the topology of the G4s. In a ThT/G4 ratio of 1:2, *AS1411* and *c-Kit2* G4s induced a 60- and 145-fold fluorescence enhancement of ThT, respectively. When in the presence of *DAP*, *HIF-1α*, *JAZF-1*, and *PDGF-A*, ThT fluoresced with 72-, 197-, 131-, and 94-fold higher intensities, respectively (**Figure 3.1**). This indicated that all four investigated G-rich sequences are effectively able to fold into G4 structures. As expected, the *scrG* sequence and *hairpin* duplex structure showed less than a 15-fold increase in ThT fluorescence increase.

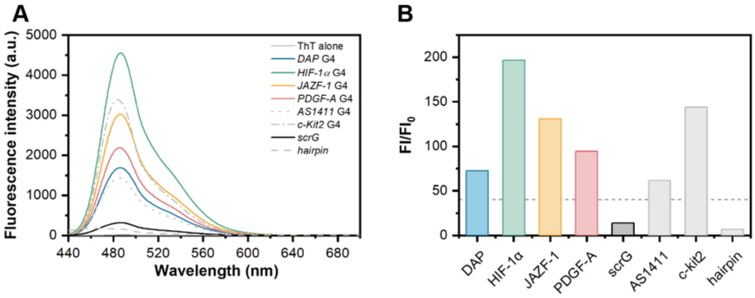


Figure 3.1 Characterization of *DAP*, *HIF-1α*, *JAZF-1*, *PDGF-A*, and *scrG* G-rich sequences. **A**) Fluorescence emission spectra of ThT in the absence and presence of *DAP*, *HIF-1α*, *JAZF-1*, *PDGF-A*, *scrG*, *AS1411*, *c-Kit2*, or *hairpin*. **B**) Bar graph of fluorescence enhancement of ThT in the presence of the investigated DNAs.

To corroborate this result, the CD spectra of *DAP*, *HIF-1α*, *JAZF-1*, *PDGF-A* G4s, and *scrG* sequence were acquired. The CD spectra of *JAZF-1* and *PDGF-A* G4s recorded at 20 °C are characterized by an intense absorbance maximum at 264 nm and a minimum at around 240 nm, indicating parallel G4 conformations (**Figure 3.2**). On the other hand, the CD spectra of *HIF-1α* and *DAP* G4s showed a less intense maximum at 264 nm and a bump at 295 nm, slightly more pronounced in the case of *DAP* G4, suggesting a parallel main conformation beyond the possible minor contribution of other topologies with antiparallel-oriented strands (**Figure 3.2**). Not surprisingly, the CD spectrum of *scrG* showed typical features of an unstructured oligonucleotide, with a weak positive peak at 280 nm (Clark et al., 1997). Different dichroic signal intensities could be ascribed to a different number of stacked G-tetrad involved in G4s formation. Indeed, *HIF-1α* and *DAP* G-rich sequences are characterized by four G-tracts, while the one of *JAZF-1* and *PDGF-A*, exhibit five G-tracts.

The thermal stability of these G4-forming sequences was investigated through CD melting experiments. All DNA structures showed very high thermal stability in solution in the presence of 150 mM potassium ion, with melting temperature vales certainly higher than 80 °C (**Figure 3.2**).

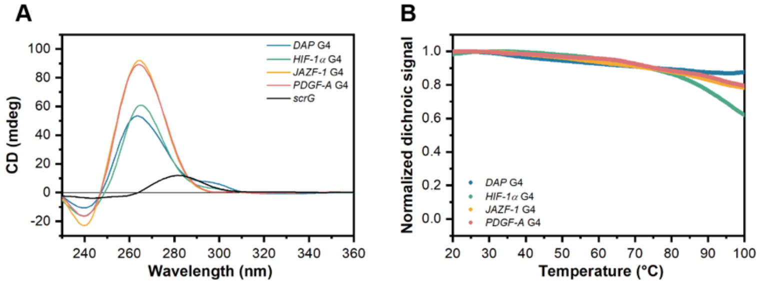


Figure 3.2 A) CD spectra at 20 °C of *DAP*, *HIF-1α*, *JAZF-1*, *PDGF-A* G4s, and *scrG* and **B**) G4s normalized CD melting curves.

The CD spectra of biotinylated oligonucleotide sequences were acquired in the same conditions as the unlabeled forms, showing no significant differences (data not shown). To further demonstrate the formation of *DAP*, *HIF-1α*, *JAZF-1*, and *PDGF-A* G4 structures, 1D ^{1}H NMR spectra were acquired for each sequence. The presence of humped and unresolved groups of signals in the imino proton region (10.0-12.5 ppm) confirms the presence of mixtures of G4 structures in the employed solution conditions.

PROFILING OF DNA G4-INTERACTING PROTEINS. Aiming to identify G4RPs from U2OS cancer cell lines, biotinylated *DAP*, *HIF-1α*, *JAZF-1*, and *PDGF-A* G4-forming sequences were targeted in a classical proteomic approach. To evaluate the selectivity of putative G4RPs for each secondary structure over the unstructured G-rich strand, nuclear lysates were first incubated with the biotinylated *scrG*. The recovery of the oligonucleotide probe was done by the addition of streptavidin-modified beads, taking advantage of the strong and specific affinity between streptavidin and the biotin moiety. Thus, the washed unbound fractions, representing the most promising selective G4 interactors, were incubated with the biotinylated G4 structure. The bound fractions, both from the *scrG* and the G4s were eluted in denaturing conditions (**Figure 3.3**).

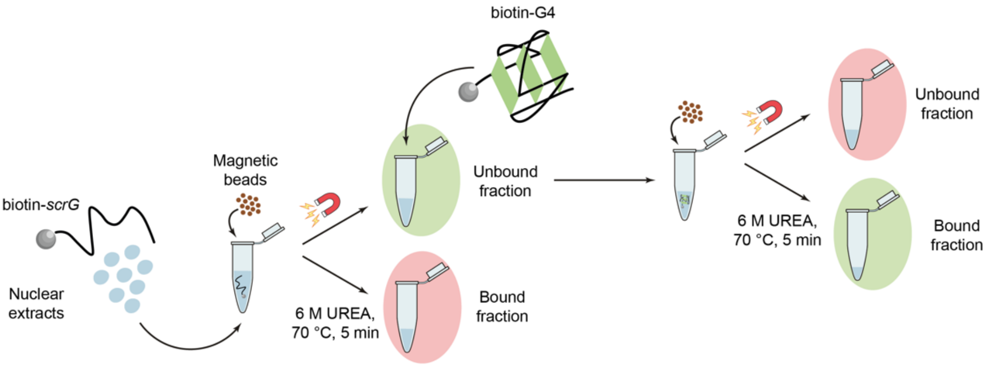


Figure 3.3 Schematic overview of chemical proteomics workflow for the identification of G4 DNA interactors.

All the collected fractions were digested by an enzymatic reaction in solution with trypsin, and subsequent peptide mixtures were analyzed by LC-MS/MS. Label-free quantification mass analysis resulted in 60 promising G4RPs. All of them were classified according to their biological activities and molecular functions using the DAVID database (D. W. Huang et al., 2009; Sherman et al., 2022) (**Figure 3.4**). The analysis of the annotated biological processes revealed that this set of G4RPs is implicated in various nuclear processes, most of them in mRNA processing or splicing, and more than 25% are involved in transcriptional events. Moreover, 45 out of 60 are molecular recognized as RNA- and/or DNA-binding, while 18 out of 60 were previously reported as RNA and/or DNA G4 interactors.

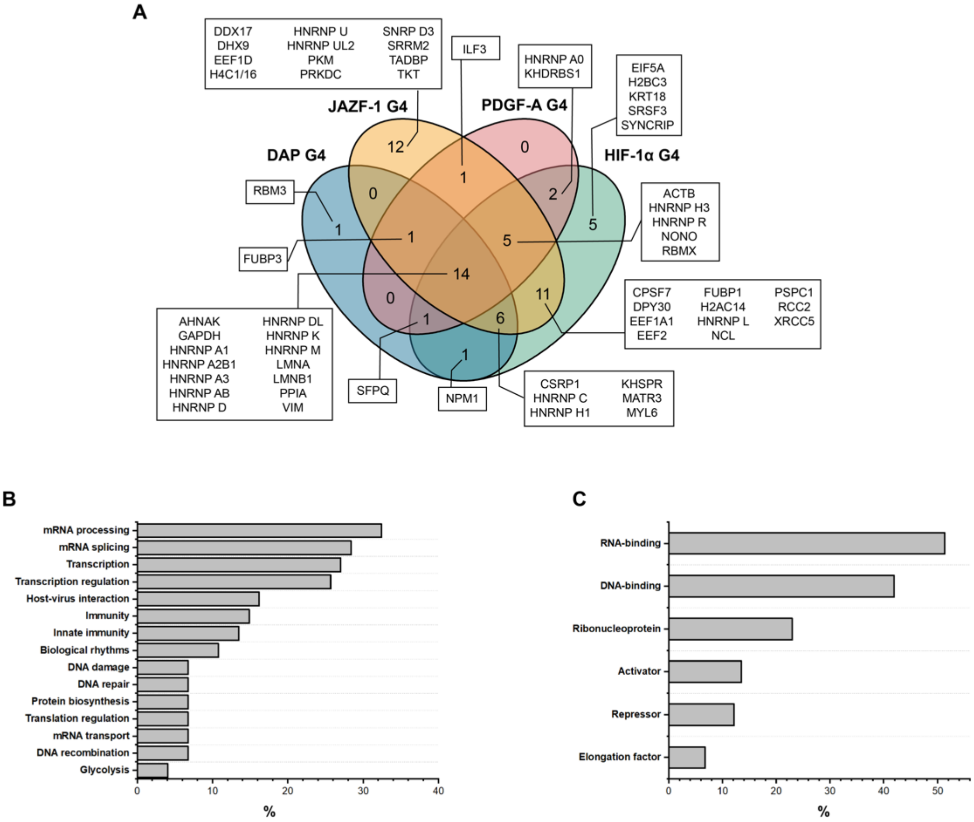


Figure 3.4 A) Overlap between enriched *DAP*, *HIF-1α*, *JAZF-1*, and *PDGF-A* G4RPs obtained by pull-down experiments; **B**, **C**) Percentage distribution for biological process and molecular function, respectively, of all the putative G4RPs (60) obtained by DAVID software.

Even if our protocol did not allow quantitative analysis of binding fractions, a selectivity index (SI) can be defined considering the ratio between the number of times each protein bound to *DAP*, *HIF-1α*, *JAZF-1*, or *PDGF-A* G4s and *scrG*. Thus, the following classification can be made: *i*) n. 17 proteins able to bind with a certain selectivity the DNA G4s, with SI > 1; *ii*) n. 43 proteins that showed no selectivity for DNA G4s, with SI ≤ 1. Among the proteins belonging to the latter group, we can identify 20 proteins with SI < 1, which slightly preferred the unstructured G-rich sequence, and 23 proteins that equally interacted with *scrG* and the DNA G4 structures (**Figure 3.5**).

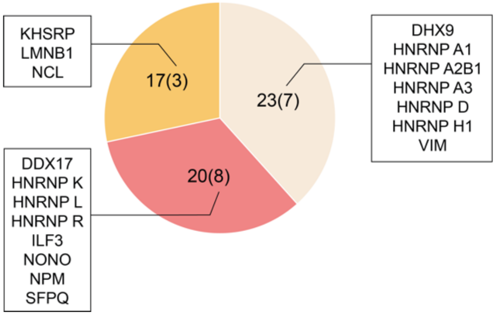


Figure 3.5 Pie chart for the distribution of the 60 putative G4RPs (orange, beige, and red for S > 1, S = 1, S < 1). In each slice, the total number of proteins is reported and the number of those already known in the literature as G4 ligands is indicated in brackets.

Among 40 proteins with SI ≥ 1, 10 of them have been already discovered and investigated as G4 binders. In agreement with our results, 5 out of 7 known G4RPs with SI = 1 are already well characterized for their unwinding activity or their unspecific binding to unstructured G-rich sequences. In particular, the DHX9 helicase has been investigated for its unwinding activity of RNA G4s in 5'-UTR (Chakraborty & Grosse, 2011; Murat et al., 2018). On the other hand, the ribonucleoprotein HNRNP A1 has shown helicase activity on both DNA and RNA G4s, as well as HNRNP A2B1 with unfolded DNA G4 in the long terminal repeat (LTR) promoter of HIV-1 (Krüger et al., 2010; Paramasivam et al., 2009). Also, another ribonucleoprotein, HNRNP D, is involved in the regulation of the telomere 3′-overhang through destabilization of the corresponding G4 structure, while HNRNP H1 binds G-rich sequences irrespective of whether in a non-G4 or G4 state (Enokizono et al., 2005; Vo et al., 2022). Finally, HNRNP A3 is only known to bind G4-forming RNA hexanucleotide `GGGGCC` repeats in *C9orf72* mutant gene (Mori et al., 2013). On the other hand, it was recently discovered that VIM can bind to two adjacent G4 DNA motifs from both telomere and oncogene promoters (Ceschi et al., 2022). Interestingly, according to our results, VIM seems to interact with all the G4s investigated here. As for the far upstream element-binding protein 2, (also called KHSRP), the lamin B1 (LMB1), and NCL, they are the only three known G4RPs that show a certain selectivity for G4s here. In our opinion, this result supports the effectiveness of our protocol since NCL is widely considered an excellent and selective G4RP, acting on both RNA and DNA G4s (Ghosh et al., 2023; González et al., 2009; Lago et al., 2017; Santos, Miranda, et al., 2021; Tosoni et al., 2015). Concerning KHSRP and LMNB1, we have here shown their ability to also bind G4 structures from oncogene promoters, not only from telomere as already found by our research group (Pagano et al., 2015).

Finally, considering the 8 proteins with SI < 1 already known to bind to G4s, they are mostly known as mostly RNA G4 interactors. DDX17 acts on RNA G4s in 5’-UTR (Herdy et al., 2018), while nucleophosmin (NPM), and the non-POU domain-containing octamer-binding protein (NONO), interact with lncRNA MALAT1 G4 (Ghosh et al., 2022, 2023). Regarding HNRNP K, HNRNP L, HNRNP R, ILF3, and SFPQ, they were fished out as interactors of G4 RNA in *C9orf72* mutant gene (Mori et al., 2013). Therefore, further investigation should be carried out to evaluate RNA over DNA G4s selectivity. The helicase activity of DDX17 (SI < 1) and DHX9 (SI =1) could explain the lack of selectivity for the DNA G4s here investigated.

Considering the four investigated G4-forming sequences, *JAZF-1* and *HIF-1α* G4s are those for which a greater number of interactors have been identified. Altogether, they have 50 and 45 putative binders, respectively (**Figure**

3.4), and share 36 G4RPs, 15 of which characterized by SI > 1. As for the *DAP* and *PDGF-A* G4s, 24 G4RPs were fished out, of which 16 in common (**Figure 3.4**). Interestingly, 18 proteins interact with only one G4 structure, particularly: *i*) 12 proteins with *JAZF-1* G4; *ii*) 5 proteins with *HIF-1α* G4; *iii*) 1 protein with *DAP* G4. As for *PDGF-A* G4, no potentially specific interactors were found.

Overall, the most interesting set of G4RPs is the one shared by all four G4 structures investigated, which contain 14 proteins. Among them, 6 proteins are characterized by SI > 1 (AHNAK, HNRNP K, HNRNP M, LMNA, LMNB1, and PPIA), while 8 with S = 1 (GADPH, HNRNP A1, HNRNP A2B1, HNRNP A3, HNRNP AB, HNRNP D, HNRNP DL, VIM). Among the 6 selective proteins, LMNB1, a major structural component of the nucleus involved in the regulation of many nuclear functions, was already found by our research group as an interactor of telomeric G4 (Pagano et al., 2015). On the other hand, the prelamin A/C (LMNA), which also plays an important role in nuclear assembly, chromatin organization, and telomere dynamics, has been identified here for the first time as a promising G4RP. Remarkably, a high molecular weight protein as the neuroblast differentiation-associated protein AHNAK (M.W. 629,101 Da), required for neuronal cell differentiation, and a small one as peptidyl-prolyl cis-trans isomerase A (PPIA, M.W. 18,012 Da), which catalyzes the cis-trans isomerization of proline imidic peptide bonds in oligopeptides, were for the first time identified as putative G4 interactors. Further, another enzyme, the glyceraldehyde-3-phosphate dehydrogenase (GAPDH), involved in several nuclear functions, including transcription, was identified for the first time as a potential binder of G4s, even if without selectivity for this type of structure. As for VIM, according to its preferential G4 binding mode described in the literature (Ceschi et al., 2022), the interaction of the protein with the G4s investigated here suggests the possibility that it might accommodate at the interface of stacked G-tetrads in higher-order G4 structures. Notably, 8 out of the 17 total ribonucleoproteins fished out act as G4RPs of all the four investigated G4 structures. In particular, HNRNP AB, HNRNP DL, and HNRNP M, were determined as G4RPs for the first time.

Overall, we have identified for the first time at least 7 new proteins capable of binding the G4 structures from the promoter regions of four oncogenes (*DAP*, *HIF-1α*, *JAZF-1*, and *PDGF-A*), namely AHNAK, GAPDH, HNRNP AB, HNRNP DL, HNRNP M, LMNA, and PPIA. Several other promising proteins showed different preferences for each G4, and/or preference for an unstructured G-rich strand. In the near future, validation of direct interactions and detailed biophysical characterization of the most promising G4/protein systems will be carried out.

3.1.3 CONCLUSIONS

G4 structures represent key elements in the regulation of several biological processes, including gene expression. In particular, identifying G4-binding proteins is essential to clarify their actual role in such processes. Inspired by the literature, in this study, we employed a classical "*fishing-for-partners*" proteomic approach in solution to identify putative interactors of G4 structures in promoter regions of *DAP*, *HIF-1α*, *JAZF-1*, and *PDGF-A* oncogenes. The G4 motifs of *HIF-1α* and *PDGF-A* were already investigated in literature (De Armond et al., 2005; Qin et al., 2007). While here, by employing a combination of spectroscopic techniques, we demonstrated that also *DAP* and *JAZF-1* G-rich sequences are able to fold into G4 structures *in vitro*, representing new target G4 structures. Then, biotin-labeled G4 structures were used as baits in solution to identify putative G4RPs in U2OS cancer cells with the proteomic approach. Noteworthy, 60 G4RPs, most of them involved in mRNA splicing and processing or in transcriptional activities, have been identified. 18 of the 60 proteins were previously reported as RNA and/or DNA G4 interactors. The LMB1 and KHSRP proteins, already discovered as telomeric G4 interactors from our research group (Pagano et al., 2015), were found to interact also with G4 in oncogene promoter region. The *JAZF-1* and *HIF-1α* G4s led to the identification of more than 40 proteins, while *DAP* and *PDGF-A* G4s were less targetable, counting 24 G4RPs. Interestingly, 14 proteins (8 of them were ribonucleoproteins) potentially interact with all the four investigated G4 structures. Among them, 7 were identified as G4RPs for the first time, AHNAK, GAPDH, HNRNP AB, HNRNP DL, HNRNP M, LMNA, and PPIA. In our perspective, targeting mixtures of G4s could be a promising strategy to obtain a better depiction of intricated in-cell mechanisms. Once validated the direct interactions of these proteins with the G4 structures, new potential G4 binders could be designed as competitors, to try to modulate the subsequent biological effects.

3.1.4 EXPERIMENTAL SECTION

OLIGONUCLEOTIDE SYNTHESIS AND SAMPLE PREPARATION. DNA sequences were synthesized on an ABI 394 DNA/RNA synthesizer (Applied Biosystem, Foster City, CA, USA) using standard β-cyanoethyl phosphoramidite solid phase chemistry at 1 μmol synthesis scale. Deprotection and detachment were performed by using a concentrated NH_4OH aqueous solution at 55 °C for 12 h. DNA filtrates and washings were combined and concentrated under reduced pressure, solubilized in water, and then purified by high-performance liquid chromatography using an anionic exchange column (Nucleogel SAX, Macherey-Nagel, 1000-8/46) and eluting with a linear gradient as previously reported (Amato et al., 2014). The purified fractions of the oligomers were then desalted by using Sep-Pak cartridges (C-18). DNA samples were prepared in 20 mM HEPES buffer (pH 6.6) containing 150 mM KCl, and 1 mM EDTA. Their concentration was measured by UV adsorption at 90 °C using the appropriate molar extinction coefficient values, ε (λ = 260 nm), calculated by the nearest-neighbor model (Cantor et al., 1970). Thus, they were annealed at 90 °C per 5 min, and then slowly cooled to room temperature and stored at 4 °C.

CD EXPERIMENTS. CD experiments were carried out on a Jasco J-815 spectropolarimeter (JASCO Inc., Tokyo, Japan) equipped with a PTC-423S/15 Peltier temperature controller using a quartz cuvette with a path length of 1.0 cm. Biotin-labeled and unlabeled oligonucleotides were prepared at 50 μM as described above. CD spectra were recorded between 230-360 nm at 100 nm/min scan speed, 1.0 bandwidth 0.5 response, and 3 accumulations. CD melting experiments were carried out in the 20-100 °C temperature range at a 1 °C min^{-1} heating rate by following changes of the CD signal at the wavelengths of the maximal CD intensity, i.e., 264 nm. Since the exceptional G4 thermal stability, the melting process was not completed even at 100 °C, and the corresponding curves were normalized by dividing only by the maximum.

THT FLUORESCENCE ASSAY. Experiments were carried out in a 1 cm pathlength cell at 20 °C on a Jasco FP-8300 spectrofluorometer (Jasco, Easton, MD, USA) equipped with a PCT-818 Peltier cell holder. For the experiments, a stock solution (ca. 300 μM) of Thioflavin T (ThT) was prepared in water and the concentration was determined using the molar extinction coefficient of 36,000 M^{-1} cm^{-1} at 412 nm (De La Faverie et al., 2014). Typically, a solution of DNA/ThT in a 2:1 ratio was prepared and allowed to equilibrate in the dark for 30 min at room temperature. Fluorescence emission spectra were recorded in the 440-700 nm range at 100 nm min^{-1} scan speed, using the excitation wavelength of 420 nm and setting both the excitation and emission slits at 5 nm. The results were reported as fluorescence intensity enhancement (FI/FI_0) of ThT at 487 nm, where FI is the fluorescence of ThT in the presence of DNA and FI_0 is the background fluorescence of ThT alone, after subtraction in both cases of the buffer fluorescence. All measurements were performed in triplicate.

NMR EXPERIMENTS. 1D ^1H NMR spectra were recorded on a Bruker Advance NEO NMR spectrometer (Bruker BioSpin, Rheinstetten, Germany), operating at 600 MHz (^1H Larmor frequency) and equipped with a 5-mm QCI cryo-probe set and a cooled SampleJet autosampler. DNA samples (50 µM) were transferred into 3-mm NMR tubes, and 10% D_2O was added. All spectra were acquired at 4 °C and 20 °C, using excitation sculpting with gradients for water suppression (Hwang & Shaka, 1995). All the experiments were performed using 2048 scans per spectrum, with a recovery delay of 1.5 s. Before Fourier transformation, free induction decays were multiplied by an exponential function equivalent to a 0.3-Hz line-broadening factor. The transformed spectra were phase-adjusted, baseline-corrected, and calibrated against the DSS signal as an external reference (Markley et al., 1998). Spectra were processed and analyzed with the Bruker TopSpin 4.3.0 software package.

G4RPS FISHING PROTOCOL. Nuclear lysates (200 µg) of U2OS cancer cells were stored at -80°C in 20 mM HEPES, 150 mM NaCl, 1 mM EDTA, and 1 mM EGTA (pH 7.9). Before use, the pH was adjusted to 6.6 value, then they were incubated with biotin-*scrG* (0.2 mg ml^{-1}) for 20 min at 4 °C. Then, 1 mg of Dynabeads™ M-280 Streptavidin (10 mg ml^{-1}, Thermo Fisher) was added to the mixtures and incubated for 30 min at 4 °C. The supernatants, enriched with putative G4 selective proteins, were washed away and further incubated first with a biotin-G4 structure (0.2 mg ml^{-1}) up to 30 minutes at 4 °C, and then with 1 mg of beads at same conditions. The supernatant, containing the unbound proteins, was washed away. Both G-rich strand and G4s bound protein fractions were eluted in denaturing conditions in 6 M urea, 300 mM Tris, and 10 mM EDTA buffer for 5 minutes at 70 °C.

IN SOLUTION DIGESTION. For the proteomics analysis, eluates recovered from G4s and *scrG* fishing experiments were subjected to a digestion protocol in solution. The cysteines were reduced with 20 mM dithiothreitol (DTT) dissolved in 25 ammonium bicarbonate (AMBIC) to the protein suspension and incubated at 60 °C for 1 hour. After cooling the samples, the reduced cysteines were alkylated with 40 mM iodoacetamide (IAM, dissolved in 25 AMBIC) and incubated in the dark for 45 min at room temperature. A solution of 100 mM DTT was added to the reaction to quench the alkylation process for 1 hour at room temperature. The protein mixture was then diluted with 50 mM AMBIC to achieve a concentration of 1 M urea before enzymatic digestion. An aliquot of 10 µL of trypsin 0.1 µg μL^{-1} was added to each sample in a ratio enzyme:substrate = 1:50. The enzymatic hydrolysis was performed overnight at 37 °C in a thermostatic bath. Each sample was treated with the same desalting procedure using stage tips containing 3 layers of 3M Empore C18 membrane. Stage tips were washed with 100 µL 0.1% formic acid (HCOOH) and peptides were eluted with 50 µL of 50% acetonitrile (ACN) and, subsequently, with 80% ACN, both acidified with 0.2% HCOOH. The peptide mixture was dried in a vacuum Speed-Vac, and re-suspended in 50 µL 5% ACN, 0.2% HCOOH.

LC-MS/MS analysis and proteomic data analysis were performed in collaboration with Dr. Gabriella Pinto and Prof. Angela Amoresano from the Department of Chemistry of the University of Naples Federico II.

CHAPTER 4
TARGETING OF G-QUADRUPLEXES BY SMALL MOLECULES AND PEPTIDES

The key role of noncanonical nucleic acid structures as potential therapeutic targets was extensively discussed in CHAPTER 1, as well as DNA G4 structures interaction with nuclear proteins in CHAPTER 3. In this CHAPTER, two different studies are reported in which we identified (PAPER II) or rationally designed and synthesized (PAPER III) new RNA/DNA G4-targeting small molecules as potential anti-cancer agents. A drug repurposing approach was also carried out to identify SARS-CoV-2 RNA G4 binders (PAPER IV). Not only small molecules but also peptides were investigated as potential G4 ligands (PAPER IV). Of note, an in-depth NMR structural characterization of the interaction between a hit compound from PAPER II and a target DNA G4 was also reported (PAPER VII). This study was performed in close collaboration with Prof. Janez Plavec's research group at NMR Slovenian Center in Ljubljana (SLO), where I spent several months as a visiting PhD student.

4.1 TARGETING OF TELOMERIC REPEAT-CONTAINING RNA G-QUADRUPLEXES: FROM SCREENING TO BIOPHYSICAL AND BIOLOGICAL CHARACTERIZATION OF A NEW HIT COMPOUND (PAPER II)

4.1.1 INTRODUCTION

G4-forming sequences have been extensively studied as potential therapeutic targets, especially when located at the telomere level (Bryan, 2020). The transcription of the telomeric C-rich strand in chromosomes produces telomeric repeat-containing RNA (*TERRA*), characterized by a G4-forming sequence from repeated r(UUAGGG) motifs (Azzalin et al., 2007; Hirashima & Seimiya, 2015). *TERRA* G4 structures are potentially more valuable therapeutic targets than their DNA counterparts since *TERRA* also takes part in heterochromatin formation and homologous recombination (Montero et al., 2016; Redon et al., 2010), even in those even in cells that do not rely on telomerase for telomere elongation that do not require telomerase to elongate their telomeres (ALT-positive cells) (Bryan et al., 1995). A relevant percentage of human tumors (around 15%) possess ALT mechanisms for telomere elongation that correlate with high *TERRA* expression. These tumors are of mesenchymal origin and are characterized by high genetic instability, usually leading to a worse prognosis (Mackenzie et al., 2021). In this regard, targeting of *TERRA* G4 could represent an effective pharmacological strategy to hit this class of tumors.

The presence of the 2’-hydroxyl (2’-OH) group in the ribose sugar makes RNA G4s more compact and thermally stable than their DNA counterparts (D. H. Zhang et al., 2010). This also induces guanine bases to adopt the *anti* conformation in the G4 structure, which thereby can only be parallel (G. W. Collie et al., 2010). In addition, the presence of the 2’-OH groups may interfere with the interaction of ligands with the loops of RNA G4 by reducing their depth and width (G. Collie et al., 2009), and/or affecting the π–π stacking surface of the external G-tetrads. Modifications to a generic G4 binder can enable selective RNA *vs.* DNA G4 targeting. For instance, pyridostatin lacks the ability to distinguish between RNA and DNA G4s, while the carboxypyridostatin derivative prefers RNA G4s (Di Antonio et al., 2012).

By employing a high-throughput in silico screening strategy on a large number of compounds from a commercially available database, new molecular scaffolds able to target *TERRA* G4 (**Figure 4.1**) were identified. Interestingly, among 103 molecules, N-[3-1H-1,3-benzodiazol-2-yl)phenyl]-1H-1,3-benzodiazol-2-amine (namely, **BPBA**, **Figure 4.1**) emerged as the most promising hit compound during experimental validation. Consequently, it underwent further biophysical investigation to assess its affinity and selectivity for RNA *vs*. DNA G4s, as well as its selectivity for G4 over duplex structures. Moreover, biological investigations were carried out in cancer cells with high *TERRA* expression to assess the effects of **BPBA** binding to *TERRA* and its potential displacement of *TERRA* from telomeres.

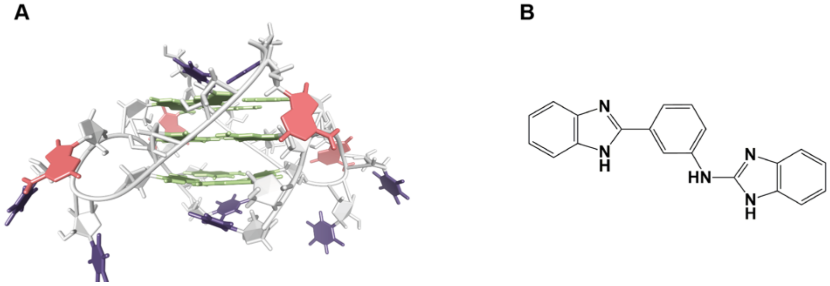


Figure 4.1 A) Bimolecular G4 structure formed by *TERRA* RNA (guanine, green; adenine, red; uracil, violet); **B**) Chemical structure of compound **BPBA**.

4.1.2 RESULTS AND DISCUSSION

Virtual-screening calculations were performed to identify drug-like molecules capable of binding to the three-dimensional *TERRA* G4 structure formed by the 12-mer r(`UAGGGUUAGGGU`) sequence (**Table 4.1**)

(Martadinata & Phan, 2009). A receptor-based virtual-screening approach based on the identification of druggable RNA hot spots and molecular docking was employed. First, aiming to identify accessible regions on the target RNA G4 structure capable of forming strong polar and nonpolar interactions with a putative ligand, some preliminary calculations were performed by docking calculations using the AutoDock Vina tool integrated into Mcule (https://mcule.com/, accessed in 2017). This involved employing a set of small solvent molecules with different polarity (Brenke et al., 2009; Trott & Olson, 2010). Once RNA hot spot targetable regions were identified, docking calculations were carried out specifically within these identified binding sites using Auto-Dock Vina. A diverse set of 58,870 commercially available compounds served as the screening library, resulting in the selection of 103 drug-like compounds for further experimental investigations. These selected compounds were then purchased for subsequent analysis.

CD EXPERIMENTS. In order to identify true hits, the 103 computationally selected small molecules were experimentally validated for their capacity to thermally stabilize *TERRA* G4, employing a CD melting assay (Santos, Salgado, et al., 2021).

In the absence of compounds, the CD spectrum of *TERRA* G4 in K^+ buffer matched that of the dimeric propeller-type parallel conformation reported in literature (Martadinata & Phan, 2009), showing a positive band at around 265 nm and a negative one at around 245 nm. Interestingly, no significant variations in the CD profile of *TERRA* G4 were detected in the presence of any investigated compounds (at a 1:10 ratio), clearly suggesting they did not alter the parallel topology *TERRA* G4 topology (see APPENDIX, and **Figure 4.2**).

Then, the stabilizing properties of each compound were assessed through CD melting experiments, measuring the ligand-induced change in the apparent melting temperature (ΔT_m) of *TERRA* G4, by monitoring the CD signal at the wavelength of maximal intensity (265 nm). Noteworthy, among the 103 tested compounds, only **BPBA** significantly increased the T_m of *TERRA* G4 (ΔT_m > 4.0 °C, **Figure 4.2**). Consequently, it was selected for further biophysical investigation. Additional CD melting experiments of *TERRA* G4 in the presence of different **BPBA** concentrations were recorded (**Figure 4.2**). The corresponding thermal-shift curve exhibited a dose-response pattern (**Figure 4.2**), suggesting a specific interaction between **BPBA** and *TERRA* G4.

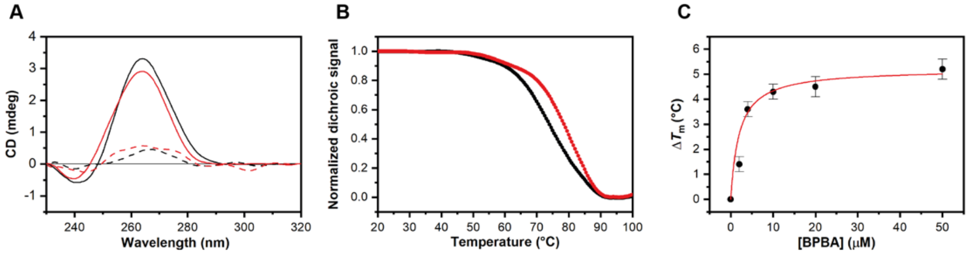


Figure 4.2 A) CD spectra of *TERRA* G4 in the absence (black line) and presence (red line) of **BPBA** (at a 1:10 ratio) recorded at 20 and 100 °C (solid and dashed lines, respectively); **B**) Normalized melting curves of *TERRA* G4 in the absence and presence (black and red dots, respectively) of **BPBA** (at a 1:10 ratio); **C**) CD stabilization curve for *TERRA* G4 with **BPBA**.

The evaluation of BPBA's interaction capabilities included also analogous parallel G4 structures formed by other RNA G4-forming sequences (i.e., *GSEC* and *Bcl-2* RNA G4s) and DNA G4-forming sequences (i.e., *c-Kit2* and *c-Myc* DNA G4s), as well as DNA G-rich sequences able to fold in other topologies, for example, the 23-mer telomeric truncation which form a hybrid G4 conformation (Tel_{23} G4, **Table 4.1**). The proper folding adopted by each of these G4-forming sequences was first confirmed by CD spectra: *i*) parallel RNA (*GSEC* and *Bcl-2*) and DNA (*c-Kit2* and *c-Myc*) G4s displayed a positive band at around 265 nm and a negative one around 240 nm; *ii*) the (3+1) hybrid G4 conformation adopted by Tel_{23} showed a positive band at 289 nm with a shoulder at ca. 268 nm and a weak negative band at around 240 nm.

As in the case of *TERRA* G4, no significant variations in the CD profiles were observed for any of these G4 structures in the presence of **BPBA** (at a 1:10 ratio), suggesting an overall preservation of the original G4 architectures (**Figure 4.3**).

Table 4.1 List of here investigated G4-forming DNA and RNA sequences.

Name	Sequence (5'-3')	Region
Parallel DNA G4s		
c-*Kit2*	`CGGGCGGGCGCTAGGGAGGGT`	Oncogene promoter
c-Myc	`TGAGGGTGGGTAGGGTGGGTAA`	Oncogene promoter
Hybrid DNA G4		
Tel$_{23}$	`TAGGGTTAGGGTTAGGGTTAGGG`	Human telomere
Parallel RNA G4s		
Bcl-2	`GGGCCGUGGGGUGGGAGCUGGG`	mRNA
GSEC	`GGGGUGGAGGAGGGGGAAGGGCGGGGG`	lncRNA
TERRA	`UAGGGUUAGGGU`	Human telomere

Since binding to B-DNA duplex may lead to undesired toxicity effects, it is crucial to determine the selectivity of the ligand towards G4 over duplex structures before proceeding with a more in-depth characterization of its binding properties. A 20-mer hairpin-forming sequence (*Hrp*$_{20}$) was selected as a suitable duplex model. Its CD spectrum, featuring a positive band at around 280 nm and a negative one at 250 nm, confirmed duplex formation and remained unaltered upon the addition of **BPBA** (at a 1:10 ratio) (**Figure 4.3**).

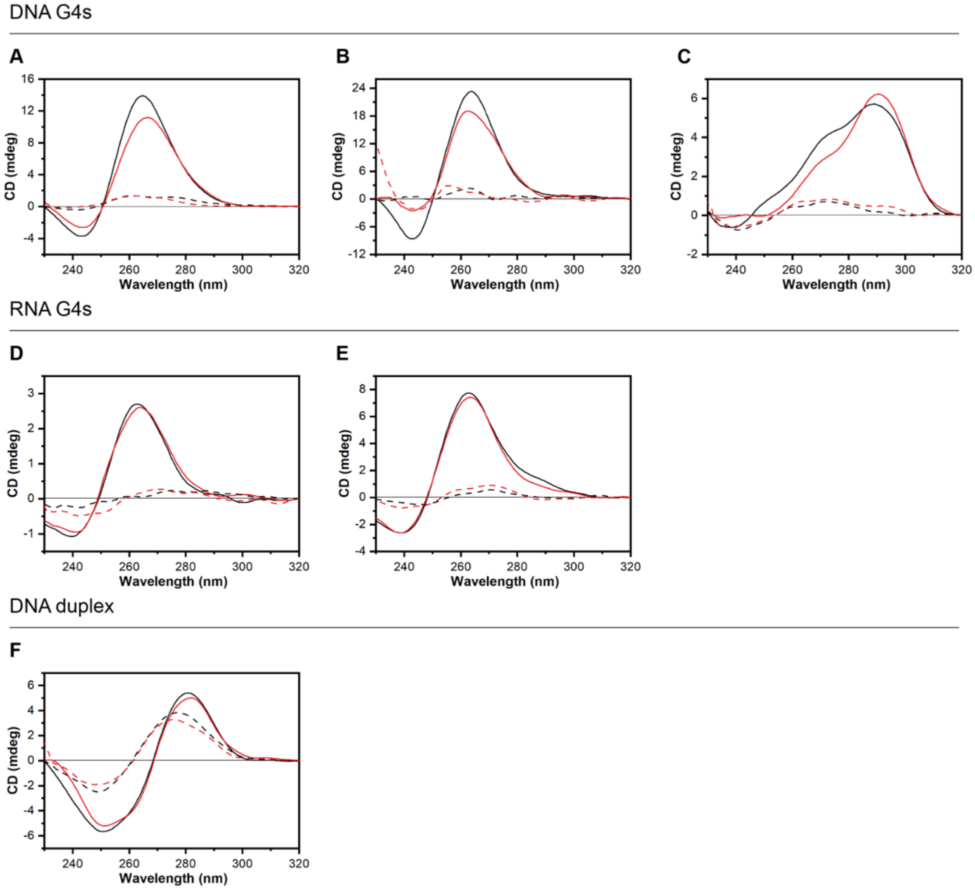


Figure 4.3 CD spectra of **A**) *c-Kit2*, **B**) *c-Myc*, **C**) *Tel*$_{23}$ DNA G4s, **D**) *Bcl-2*, **E**) *GSEC* RNA G4s, and **F**) *Hrp*$_{20}$ DNA duplex in the absence (black line) and presence (red line) of 10 molar equiv of **BPBA** recorded at 20 and 100 °C (solid and dashed lines, respectively).

The results of CD melting experiments revealed that **BPBA** can also bind to and stabilize the parallel DNA G4s formed by *c-Kit2* and *c-Myc*. Conversely, no relevant ligand-induced thermal shift was observed in the case of the DNA *Tel*$_{23}$ hybrid G4, as well as the parallel *GSEC* and *Bcl-2* RNA G4s (**Figure 4.4**, and **Table 4.2**). Furthermore, no significant change in T_m was observed in the case of *Hrp*$_{20}$ DNA duplex structure, indicating that the ligand selectively stabilizes the G4 conformation over duplex DNA (**Figure 4.4**, and **Table 4.2**). Finally, even though the stabilization imparted by the ligand would naturally be more pronounced in intrinsically fewer stable oligonucleotides (Rocca et al., 2017), the CD results suggest that **BPBA** could preferentially stabilize *TERRA* G4 over other RNA G4s and discriminate between parallel and non-parallel DNA G4 conformations, with a clear preference for the former.

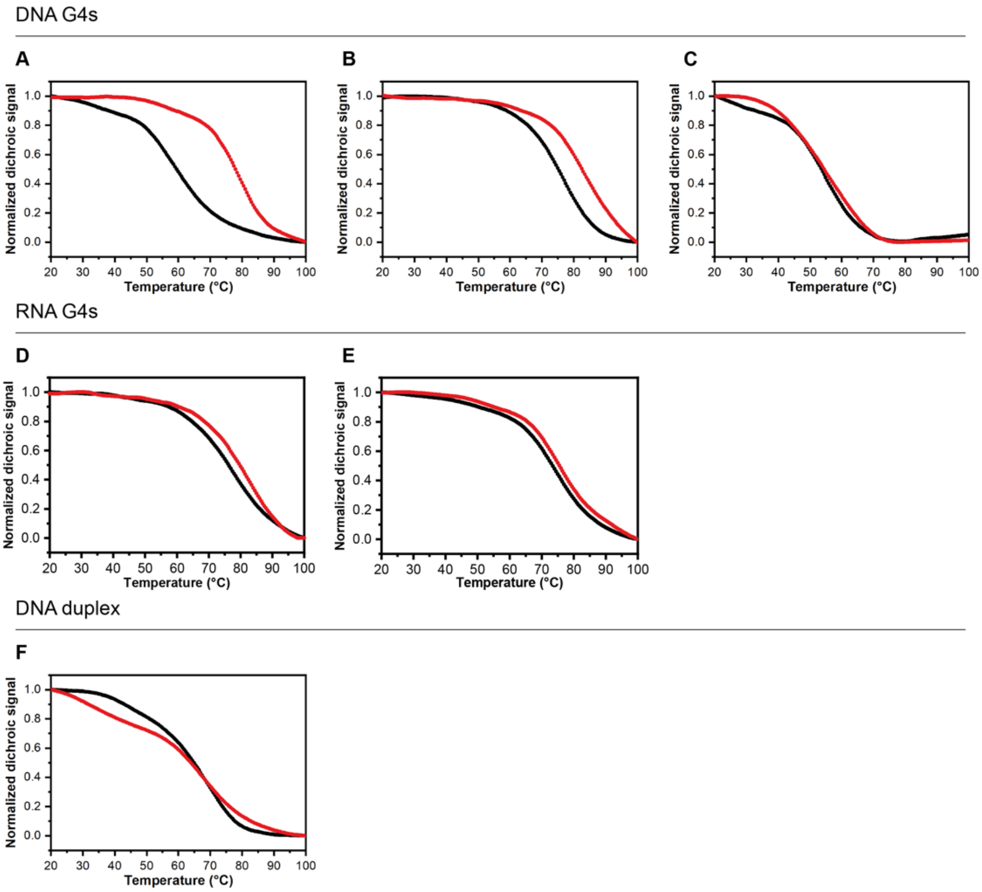


Figure 4.4 Normalized CD melting curves of **A**) *c-Kit2*, **B**) *c-Myc*, **C**) *Tel$_{23}$* DNA G4s, **D**) *Bcl-2*, **E**) *GSEC* RNA G4s, and **F**) *Hrp$_{20}$* DNA duplex in the absence (black line) and presence (red line) of 10 molar equiv of **BPBA**.

Table 4.2 BPBA-induced thermal stabilization of DNA and RNA G4s, and duplex DNA measured by CD melting experiments.

Oligo	Sequence (5'-3')	ΔT_m (°C)[a]
Parallel DNA G4s		
c-*Kit2*	`CGGGCGGGCGCTAGGGAGGGT`	18.7 (±0.3)
c-Myc	`TGAGGGTGGGTAGGGTGGGTAA`	9.4 (±0.3)
Hybrid DNA G4		
Tel_{23}	`TAGGGTTAGGGTTAGGGTTAGGG`	0.7 (±0.2)
Parallel RNA G4s		
Bcl-2	`GGGCCGUGGGGUGGGAGCUGGG`	2.9 (±0.3)
GSEC	`GGGGUGGAGGAGGGGGAAGGGCGGGGG`	1.7 (±0.3)
TERRA	`UAGGGUUAGGGU`	4.5 (±0.4)

[a] ΔT_m represents the difference in melting temperature [$\Delta T_m = T_m$ (oligo + 10 **BPBA** equiv) - T_m (oligo)]. The T_m values of oligonucleotides alone are: *TERRA* G4 = 74.3 (±0.1) °C, *c-Kit2* G4 = 59.7 (±0.1) °C, *c-Myc* G4 = 75.5 (±0.1) °C, Tel_{23} G4 = 53.7 (±0.1) °C, *GSEC* G4 = 78.8 (±0.1) °C, *Bcl-2* G4 = 74.7 (±0.2) °C, Hrp_{20} = 65.5 (±0.2) °C.

FRET MELTING EXPERIMENTS. The stabilizing properties of **BPBA** on *TERRA* G4 were further investigated using a FRET melting assay. For this assay a G4-forming telomeric RNA sequence labeled at both the 5′- and 3′-end with a 6-carboxyfluorescein (FAM, F) and a carboxytetramethylrhodamine (TAMRA, T), respectively (referred to as F-*TERRA*-T) (Giancola & Pagano, 2012) was employed. Additionally, to further confirm the ligand's selectivity for G4 over the duplex, a competition FRET melting experiment was carried out in the presence of a large excess of a duplex model (i.e., a 27-mer hairpin duplex-forming DNA, Hrp_{27}) (De Cian et al., 2007). The FRET melting curves confirmed that **BPBA** is capable of stabilizing the G4 structure formed by telomeric RNA without being affected by the presence of the duplex competitor (**Figure 4.5**, and **Table 4.3**). This suggests that this compound preferably binds to G4s.

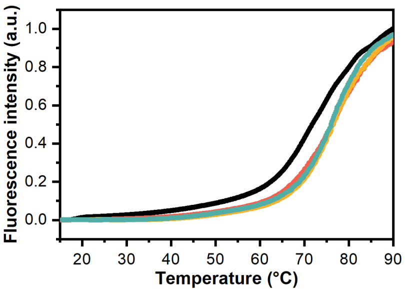


Figure 4.5 Normalized FRET melting curves for F-*TERRA*-T G4 in the absence (black lines) and presence (red lines) of **BPBA** and large excesses of *Hrp27* (1:30 and 1:100 ratio, yellow and green lines, respectively).

Table 4.3 G4-selectivity of **BPBA** measured by FRET melting competition experiments.

ΔT_m (°C)[a]

Compound	**F-*TERRA*-T G4**	**F-*TERRA*-T G4 + *Hrp27* (1:30)**	**F-*TERRA*-T G4 + *Hrp27* (1:100)**
BPBA	25.4 (±0.5)	3.7 (±0.2)	3.1 (±0.2)

[a] ΔT_m represents the difference in melting temperature [$\Delta T_m = T_m$ (F-*TERRA*-T G4 + 10 **BPBA** equiv, or *Hrp27* (1:30, or 1:100)) − T_m (F-*TERRA*-T G4)]. The T_m value of F-*TERRA*-T G4 alone is 73.5 (±0.1) °C. Differences in CD- and FRET-determined T_m could be explained with modified sequence and/or experimental conditions.

FLUORESCENCE INTERCALATOR DISPLACEMENT (FID) ASSAYS. To gain insight into the binding affinity of **BPBA** for various RNA/DNA G4s, fluorescent intercalator displacement (FID) experiments were carried out. This assay is based on the competitive displacement of a light-up fluorescent probe, thiazole orange (TO), from the oligonucleotide upon the addition of increasing amounts of a candidate ligand (Monchaud et al., 2006). The ligand-induced TO displacement decreases TO fluorescence, thus allowing for the determination of the relative binding affinity of the ligand for the structure under examination. In this study, TO displacement by **BPBA** was investigated for the G4s from *TERRA*, *c-Kit2*, *c-Myc*, and *Tel23*, as well as for the *Hrp27* duplex model. **BPBA** concentrations required to give 50% TO displacement (DC_{50}

values) were calculated from the dose–response curves (**Figure 4.6**, and **Table 4.4**). The lowest DC_{50} value was obtained for *TERRA* G4, suggesting a very good affinity of **BPBA** for this G4, followed by c-*Myc* and *c-Kit2* G4s. In the case of *Tel*$_{23}$ G4, it was not possible to reach a 50% displacement of TO even after addition of a large excess of the binder, clearly suggesting weaker ligand interactions.

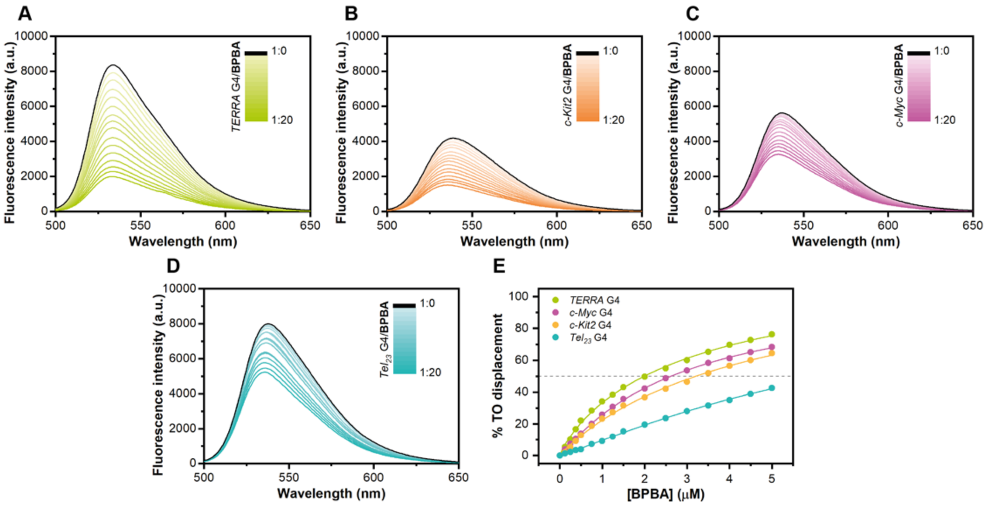


Figure 4.6 TO displacement titrations for **A**) *TERRA*, **B**) *c-Kit2*, **C**) *c-Myc*, and **D**) *Tel*$_{23}$ G4s, upon addition of increasing amounts of **BPBA**; **E**) Dose-response curves from FID experiments.

Table 4.4 BPBA DC_{50} values for DNA and RNA G4s and duplex structure determined by G4-FID assays.

Oligo	**Sequence (5’-3’)**	**DC_{50} (µM)**
Parallel DNA G4s		
c-*Kit2*	`CGGGCGGGCGCTAGGGAGGGT`	3.8 (±0.6)
c-Myc	`TGAGGGTGGGTAGGGTGGGTAA`	n.d.[a]
Hybrid DNA G4		
Tel$_{23}$	`TAGGGTTAGGGTTAGGGTTAGGG`	n.d.[a]
Parallel RNA G4s		
TERRA	`UAGGGUUAGGGU`	2.4 (±0.4)

[a] n.d. = not determined

Moreover, to clarify the binding mode of **BPBA** to *TERRA* G4, an ethidium bromide (EB) displacement assay was performed using fluorescence spectroscopy. EB binds to duplex DNA through intercalation, and to G4 DNA through π–π stacking on the external G-tetrads (Guo et al., 1992). Its fluorescence at 595 nm is strongly enhanced upon association with G4s as a consequence of the hydrophobic environment experienced upon its binding to the nucleic acid (Sengupta et al., 2013). Therefore, the addition of a G4 ligand decreases EB fluorescence intensity when it binds to G4 via end-stacking mode, displacing EB. Displacement titrations were performed by adding increasing amounts of **BPBA** to the *TERRA* G4/EB complex, resulting in a substantial decrease in the fluorescence intensity of EB (**Figure 4.7**). This suggests an end-stacking binding mode for **BPBA** to *TERRA* G4 (Roy et al., 2020). The DC_{50} value of 1.3 (±0.4) μM, calculated from dose–response curves obtained by plotting the percentage of EB displacement against ligand concentration (**Figure 4.7**), confirmed once again the strong interaction between **BPBA** and this G4 motif.

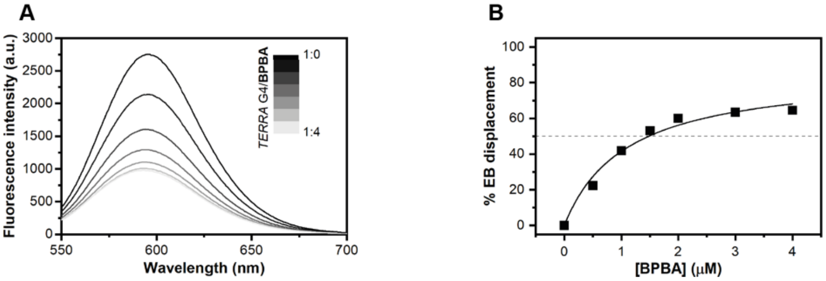


Figure 4.7 A) EB displacement titration for *TERRA* G4 upon addition of increasing amounts of **BPBA**; **B**) Dose-response curves from FID experiments.

MST ASSAYS. Quantitative data on the binding affinity of **BPBA** for the investigated G4s, were obtained by means of MST experiments. Serial dilutions of **BPBA** were prepared, mixed with a constant concentration of Cy5.5-labeled oligonucleotides (*TERRA*, *c-Kit2*, *c-Myc*, *Tel*$_{23}$ G4s, or *Hrp*$_{20}$), loaded into capillaries, and analyzed by MST. Results of MST binding curves confirmed that **BPBA** was able to bind to the parallel-stranded G4 structures, showing the lower K_D for *TERRA* G4, followed by *c-Kit2* and *c-Myc* G4s (**Figure 4.8**, and **Table 4.5**). On the other hand, no significant change in the thermophoretic signal was observed for *Tel*$_{23}$ G4 and *Hrp*$_{20}$ hairpin-duplex, clearly indicating the absence of a significant interaction in these cases.

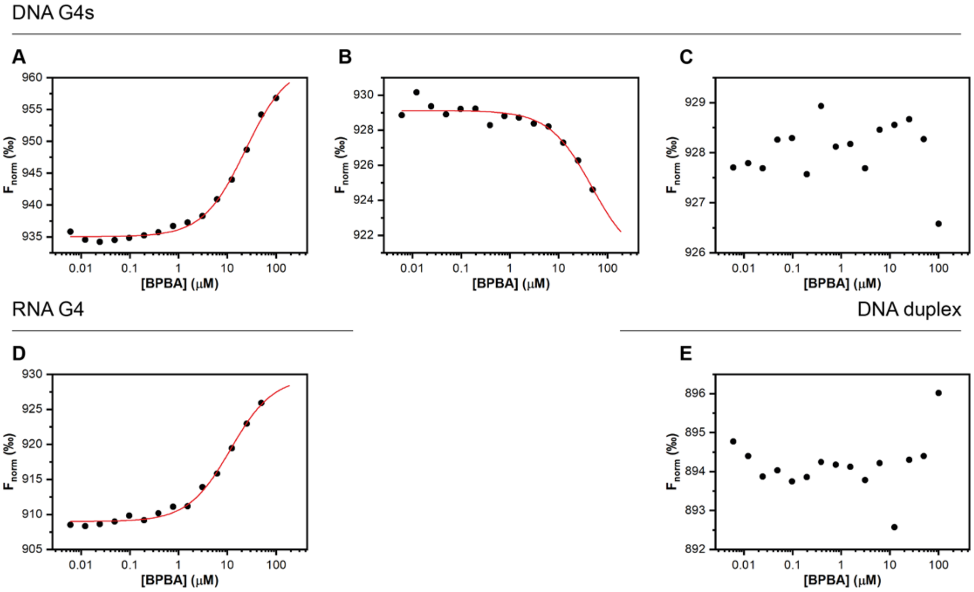


Figure 4.8 Dose-response curves from MST experiments for the interaction of **BPBA** with **A**) Cy5.5-*c-Kit2*, **B**) Cy5.5-*c-Myc*, **C**) Cy5.5-*Tel*$_{23}$, **D**) Cy5.5-*TERRA* G4s, and **E**) Cy5.5-*Hrp*$_{20}$.

MST results agree with those obtained with other techniques and confirm the preferential binding of **BPBA** to parallel over antiparallel-stranded G4 topologies, as well as its selectivity for the G4 over duplex form. Additionally, although **BPBA** showed a propensity to bind to both *TERRA* G4 and the parallel *c-Myc* and *c-Kit2* G4s, it showed a higher binding affinity for the former.

CYTOTOXICITY ASSAYS. G4 binders have an established antiproliferative effect in cancer cells depending on their ability to induce DNA damage response (DDR) or to inhibit the expression of cellular oncogenes. Given the high affinity exhibited by **BPBA** for *TERRA* G4, its antiproliferative effect was investigated in correlation with *TERRA* expression in both human cervix cancer cells (HeLa), characterized by telomerase activity and low *TERRA* expression, and human osteosarcoma cells (U2OS), lacking telomerase activity and expressing high levels of *TERRA* (Arora et al., 2014; Pompili et al., 2017). In these cell lines, *TERRA* expression was measured by RT qPCR assay with primers targeting some of the most active *TERRA* promoters located at subtelomeres of chromosomes 10q, XqYq, and XpYp, revealing a substantial difference in expression between the two cell lines. The cell viability, assessed by crystal violet assay, demonstrated that U2OS cells were significantly more sensitive to **BPBA** (IC_{50} = 8.1 (±1.0) μM) compared to HeLa cells (IC_{50} >> 10 μM).

TERRA G4 STABILIZATION AND DNA DAMAGE ASSAYS. To understand the mechanism underlying the different biological effects of **BPBA** on high *vs.* low *TERRA*-expressing cells, **BPBA** ability to bind and stabilize *TERRA* G4 in cellulo was assessed by performing RT qPCR analysis of the relative TERRA expression upon treatment with the drug. Both HeLa and U2OS cell lines were exposed to different concentrations of **BPBA**, then RNA was extracted and processed for *TERRA* analysis. Interestingly, **BPBA** induced a stabilization of *TERRA* expression only in high *TERRA*-expressing U2OS cells (**Figure 4.9**). This resulted in the accumulation of the RNA within the cell, presumably due to the binding of **BPBA** to *TERRA* G4 molecules and their protection from the degradation complexes that regulate the physiological turnover of lncRNA.

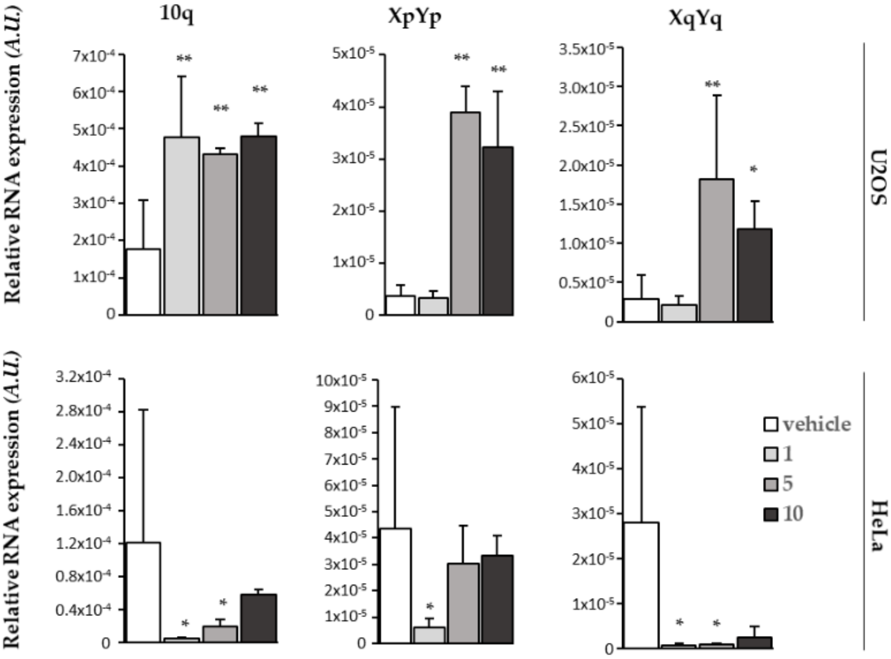


Figure 4.9 RT qPCR showing relative expression of specific *TERRA* RNA at different subtelomeric loci of HeLa and U2OS cells treated with the vehicle or indicated doses (1, 5, or 10 μM) of **BPBA** for 72 h. Mean of three independent experiments is shown. Error bars are SD; * = $p < 0.5$; ** = $p < 0.1$.

To better understand the mechanisms underlying the cytotoxicity of **BPBA** in U2OS cells, we investigated the possibility that the induced *TERRA* stabilization could stimulate a DDR at telomeres (a marker for telomere dysfunction), leading to cell death. Both U2OS and HeLa cells were exposed to **BPBA**, and DDR activation was measured as the percentage of cells displaying γH2AX histone phosphorylation, a marker of DDR. The HeLa-treated cells showed a negligible DDR induction; conversely, **BPBA** was able to induce DDR in a significant fraction of U2OS cells even in the first 24 h, also persistent at 72 h in 50% of the U2OS cells treated with 10 μM **BPBA** (**Figure 4.10**), in agreement with the calculated IC_{50} dose. Lastly, to ascertain if the activated DDR overlapped with telomeric loci, the latter U2OS-treated cells were employed to measure telomere-dysfunction-induced foci (TIF) activation. TIF-positive cells were significantly increased in treated samples compared to control (**Figure 4.10**), confirming our hypothesis of DDR localization at telomeric level.

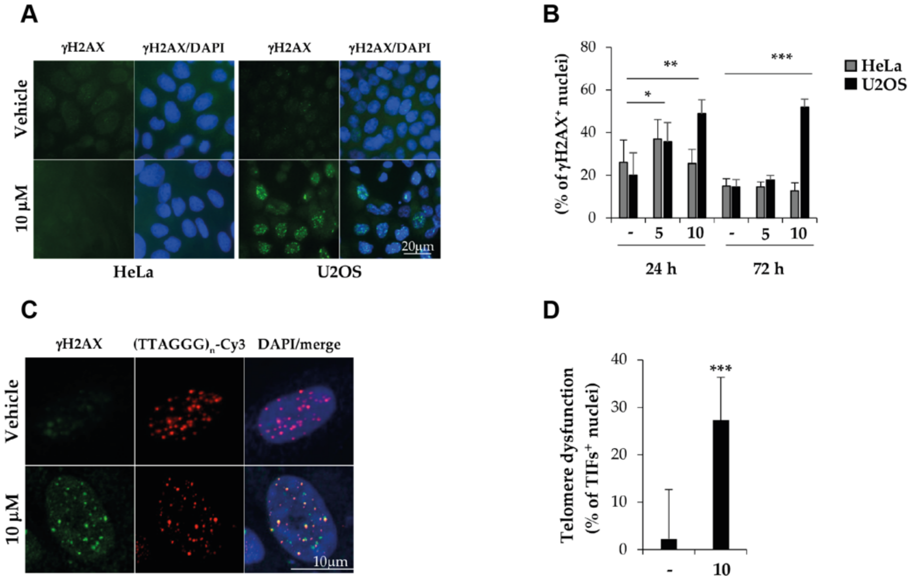


Figure 4.10 HeLa and U2OS cells treated with indicated doses of **BPBA**, processed for immunofluorescence with the γH2AX mAb primary antibody, followed by goat-antimouse 488 secondary antibody **A**,**B**) alone or **C**,**D**) in combination with FISH using the telomere-specific (`TTAGGG`)$_n$-Cy3 PNA probe. Fluorescence signals acquired with a Leica DMIRE deconvolution microscope (representative images at 72 h shown in **A**)) or a Zeiss LMS confocal scanner **B**) at 63× magnification. Percentage of cells displaying γH2AX signals or TIFs (> 4 (`TTAGGG`)$_n$-Cy3/γH2AX colocalizing spots) was scored and reported in histograms (**B**,**D**, respectively). Histograms report the mean of at least 6 different fields per sample ($_n$ > 150). Error bars are SD; * = $p < 0.5$; ** = $p < 0.1$; *** = $p < 0.05$

4.1.3 CONCLUSIONS

The application of a virtual screening approach in tandem with experimental screening via CD melting assay succeeded in the identification of a new hit compound (**BPBA**) as a binder of *TERRA* G4. The *in vitro* G4 binding properties of **BPBA** were characterized by several biophysical methodologies

(CD, FRET, FID, and MST). CD and FRET melting assays, as well as MST experiments, revealed that **BPBA** features high selectivity for G4s over duplex DNA. Furthermore, TO-FID and MST experiments showed that **BPBA** has enhanced binding affinity towards *TERRA* G4 *vs.* other G4-forming DNA sequences present along the human genome. The examination of the molecular structure of **BPBA**, compared to the other 102 screened compounds, suggests that its preferable binding properties may be due to the presence of two benzoimidazole units connected by an aniline residue. This unique molecular arrangement, exclusive to **BPBA**, imparts extensive planarity to the molecule and likely facilitates the optimal distribution of polar groups for interaction with *TERRA* G4.

Biological characterization demonstrated that **BPBA** can bind and stabilize *TERRA* lncRNAs in cellulo, likely by sequestering them from the physiological turnover cell machinery. Moreover, *TERRA* stabilization induced a DDR, potentially by displacing *TERRA* from telomeric DNA. Indeed, TERRA *physically* interacts with telomeric chromatin by forming DNA:RNA hybrids that are required for telomere homeostasis, especially in ALT-positive cells such as U2OS, where *TERRA* downregulation leads the formation of TIFs (Deng et al., 2012). In line with this, DDR activation predominantly occurs in U2OS cells, where **BPBA** also exerts the highest cytotoxic effect.

Overall, this study demonstrates the feasibility of identifying *TERRA* G4 binders with potential pharmacological effects, thereby paving the way for the search for new drug candidates targeting RNA.

4.1.4 EXPERIMENTAL SECTION

OLIGONUCLEOTIDE SYNTHESIS AND SAMPLE PREPARATION. RNA/DNA sequences were synthesized on an ABI 394 DNA/RNA synthesizer (Applied Biosystem, Foster City, CA, USA) using standard β-cyanoethyl phosphoramidite solid phase chemistry at 1-μmol synthesis scale. Regarding RNA synthesis, 5-benzylthio-1-H-tetrazole (BTT) instead of 4,5-dicyanoimidazole (DCI) was used as activator reagent, and coupling steps were prolonged of 5 min. Another difference concerned the deprotection of bases and phosphates. A concentrated NH_4OH/EtOH (3:1, v/v) solution was used in the case of RNA, and the reaction was left at r.t. for 12 h. For DNA sequences, deprotection and detachment were performed by using a concentrated NH_4OH aqueous solution at 55 °C for 12 h. Both RNA and DNA sequences were purified by high-performance liquid chromatography (HPLC) on a Nucleogel SAX column (1000-8/46, Macherey-Nagel, GmbH & Co. KG, Dueren, Germany), as previously reported (Amato et al., 2014). The fractions of the oligomers were collected and successively desalted by Sep-Pak cartridges (C-18). Lastly, 2′-TBDMS groups in RNA were removed by Et3N·$_3$HF/DMF (1:3, v/v) at r.t. for 12 h. The reaction was quenched with 0.1 M TEAA buffer (pH 7.0) and again desalted on a Sep-pak (C-18) cartridge. All oligonucleotides were proven to be > 98% pure by NMR. The list of synthesized oligonucleotides is reported in **Table 4.1**. A 20-mer hairpin duplex-forming sequence d(`CGAATTCGTTTTCGAATTCG`) (*Hrp*$_{20}$), and a 27-mer hairpin duplex-forming sequence d(`CGCGAATTCGCGTTTCGCGAATTCGCG`) (*Hrp*$_{27}$) were also included in the study. Oligonucleotide solutions were prepared in the appropriate buffer, and their concentration was measured by UV adsorption at 90 °C using the appropriate molar extinction coefficient values, ε (λ = 260 nm), calculated by the nearest-neighbor model (Cantor et al., 1970).

CD EXPERIMENTS. CD experiments were performed on a Jasco J-815 spectropolarimeter (Jasco, Easton, MD, USA) equipped with a PTC-423S/15 Peltier temperature controller. All spectra were recorded at 20 and 100 °C in the wavelength range of 230-320 nm and averaged over three scans. A scan rate of 100 nm min^{-1} with a 1 s response time and 1 nm bandwidth were used. The buffer baseline was subtracted from each spectrum. Concentrations of 2 and 4 μM oligonucleotides were used for G4s and *Hrp$_{20}$*, respectively. RNA G4s were prepared in 20 mM KH_2PO_4/K_2HPO_4 buffer (pH 7.0) containing 70 mM KCl, while a buffer solution consisting of 5 mM KH_2PO_4/K_2HPO_4 (pH 7.0) containing 20 mM KCl was used for all DNA samples. All oligonucleotide samples were annealed by heating at 90 °C for 5 min, followed by a slow cooling to room temperature overnight. CD spectra of oligonucleotide/ligand mixtures were obtained by adding 10 molar equiv of each putative ligand (stock solutions of ligands were 10 mM in DMSO). CD melting experiments were carried out in the 20-100 °C temperature range at a 1 °C min^{-1} heating rate by following changes of the CD signal at the wavelengths of the maximal CD intensity (i.e., 264 nm for *TERRA*, *GSEC*, *Bcl-2*, *c-Kit2*, and *c-Myc* G4s; 287 nm for *Tel$_{23}$* G4; 280 nm for *Hrp$_{20}$*). CD melting experiments were performed in the absence and presence of ligands (10 molar equiv) added to the folded nucleic acid structures. The apparent melting temperatures (T_m) were determined from a curve fit using Origin 7.0 software (OriginLab Corp., MA, USA). ΔT_m values were determined as the difference in the T_m values of the nucleic acid structures in the presence and absence of ligands. All experiments were performed in triplicate, and the reported values are the average of the three independent measurements.

FRET EXPERIMENTS. Measurements were carried out on a FP-8300 spectrofluorometer (Jasco, Easton, MD, USA) equipped with a Peltier temperature controller system (Jasco PCT-818) using the dual-labeled G4-forming telomeric RNA sequence FAM-[r(GGGUAAGGGUAAGGGUAAGGG)]-TAMRA (F-*TERRA*-T), provided from Biomers (Ulm, Germany). The oligonucleotide was dissolved in water at a concentration of 1 mM, diluted at 1 μM using 5 mM potassium phosphate buffer (pH 7.0) containing 20 mM KCl, and then annealed by heating to 90 °C for 5 min, followed by cooling to room temperature over-night, and storage at 4 °C for 24 h before data acquisition. Experiments were performed in sealed quartz cuvettes with a path length of 1 cm by using 0.2 μM of prefolded F-*TERRA*-T G4 target, the ligand at 2 μM, and the *Hrp*$_{27}$ duplex competitor at final concentrations of 0, 6, and 20 μM. In addition, a blank with no compound or competitor was also analyzed. Fluorescence spectra were acquired before (at 15 °C) and after (at 90 °C) FRET melting assay. The dual-labeled oligonucleotide was excited at 492 nm, and emission spectra were recorded between 500 and 650 nm by using 100 nm s^{-1} scan speed. Excitation and emission slit widths were both set at 5 nm. FRET melting was performed by monitoring the emission of FAM at 522 nm (upon excitation at 492 nm), using a heating gradient of 1 °C min^{-1} over the range 15-90 °C. Emission of FAM was normalized between 0 and 1. Final analysis of the data was carried out using Origin 7.0 software (OriginLab Corp., MA, USA).

TO G4-FID ASSAYS. A solution containing 0.25 μM of prefolded RNA (*TERRA* G4) or DNA G4 target and two equivalents of TO (0.5 μM) in 20 mM KH_2PO_4/K_2HPO_4 (pH 7.0) and 70 mM KCl was prepared in a 1 cm path-length cell, and the corresponding fluorescence spectrum was acquired in the absence and presence of increasing concentrations of **BPBA** (10 mM stock solution in pure DMSO). Each ligand addition (from 0.5 to 20 molar equiv) was followed by a 3 min equilibration time before spectrum acquisition. The FID experiment was also extended to a duplex DNA model (*Hrp*$_{27}$). In this case, three equivalents of TO (0.75 μM) were added to the oligonucleotide solution (0.25 μM). Measurements were run at 20 °C on a FP-8300 spectrofluorometer (Jasco, Easton, MD, USA) equipped with a Peltier cell holder (Jasco PCT-818), using an excitation wavelength of 485 nm and recording the emission in the 500-650 nm wavelength range. Both excitation and emission slits were set at 5 nm. The percentage of TO displacement was calculated as follows:

$$\text{TO displacement (\%)} = 100 - \left[\left(\frac{\mathrm{F}}{\mathrm{F_0}}\right) \cdot 100\right]$$

where F_0 is the fluorescence in the absence of a ligand and F is the fluorescence after each ligand addition. The percentage of displacement was then plotted as a function of the ligand concentration, and DC_{50} was calculated as the required concentration to displace 50% TO from each investigated DNA. Each titration was repeated at least in triplicate.

EB G4-FID ASSAYS. A solution containing 10 µM of prefolded *TERRA* G4 and 5 µM of EB in 20 mM potassium buffer (pH 7.0) containing 70 mM KCl was prepared in a 1 cm path-length cell. The fluorescence spectrum of the EB/*TERRA* G4 complex in the absence of **BPBA** was first recorded. Then, increasing concentrations of **BPBA** (10 mM stock solution in pure DMSO) were added to the EB/*TERRA* G4 complex, and spectra were recorded 3 min after each ligand addition. Measurements were run at 20 °C on a FP-8300 spectrofluorometer (Jasco, Easton, MD, USA) equipped with a Peltier cell holder (Jasco PCT-818), using an excitation wavelength of 510 nm, and recording the emission spectra in the 550-700 nm wavelength range. Both excitation and emission slits were set at 5 nm. Experiments were performed in duplicate.

MST EXPERIMENTS. MST measurements were performed on a Monolith NT.115 (Nanotemper Technologies, Munich, Germany). The Cy5.5-fluorescently labeled oligonucleotides (from Biomers, Ulm, Germany) were prepared at a concentration of 10-20 µM in 20 mM KH_2PO_4 buffer (pH 7.0) containing 70 mM KCl and annealed as described above. Nucleic acid samples were then diluted using the same phosphate buffer supplemented with 0.1% Tween. The ligand stock solution was 2 mM in pure DMSO. For the MST experiments, the concentration of the labeled G4s was kept constant at 50 nM, while a serial dilution of the ligand (1:2 from 0.4 mM) in the same buffer supplemented with 20% DMSO was prepared and mixed with the oligonucleotide solution with a volume ratio of 1:1. All samples containing 10% DMSO as the final concentration were loaded into standard capillaries (NanoTemper Technologies, Munich, Germany). Measurements were performed at 20 °C using autotune LED power and medium MST power. MST data analysis was performed by employing the MO. Affinity Analysis software (v2.3) provided with the instrument.

Virtual screening was performed by Mcule. Crystal violet assays, immunofluorescence, and RT qPCR analysis were performed in collaboration with Eleonora Vertecchi and Erica Salvati from the Institute of Molecular Biology and Pathology at the National Research Council in Rome, Italy. For further information, see APPENDIX.

4.2 G-QUADRUPLEX DNA RECOGNITION BY AN ANTICANCER BENZIMIDAZOLE DERIVATIVE: INSIGHTS FROM NMR SPECTROSCOPY

4.2.1 INTRODUCTION

As previously discussed (refer to **Paragraph 4.1**), a novel compound, named **BPBA**, has been identified as an interesting G4 binder, also characterized by a good selectivity for G4 over duplex DNA. This compound is characterized by a central phenyl core substituted with two benzimidazole rings in a relative meta position, one of which is connected via an -NH group. The affinity of **BPBA** for various G4-forming sequences has been explored. Notably, this compound has exhibited the highest affinity for *TERRA* RNA G4, followed by two parallel DNA G4s, *c-Kit2* and *c-Myc*. Interestingly, **BPBA** has displayed a preferential cytotoxic effect on U2OS cancer cells that are characterized by high levels of *TERRA*. In these cells, **BPBA** induced a DNA damage response at the telomere level (Marzano et al., 2021), suggesting a potential mechanism of action for **BPBA** in targeting and affecting cancer cells, particularly through interactions with G4 structures and induction of DNA damage response.

Encouraged by these promising results, we considered it crucial to undertake an in-depth study on the interaction between **BPBA** and G4 structures, with the aim of elucidating the structural details of this robust and selective binding. Additionally, the outcomes of this investigation hold the potential to guide the rational design of optimized derivatives of **BPBA**, characterized by enhanced affinity, selectivity, and cytotoxicity effects.

We started this investigation by performing mono-dimensional proton NMR (1D ^{1}H NMR) titration experiments to examine the interaction of **BPBA** with parallel G4s formed by *TERRA*, *c-Kit2* and *c-Myc* in phosphate buffer (5 mM KH_2PO_4/K_2HPO_4, pH 7.0, containing 20 mM KCl). The results indicated that *c-Kit2* G4/**BPBA** complex was the most feasible system for a detailed NMR structural investigation. Therefore, a combination of 1D and 2D homo- and hetero-nuclear experiments was used to in-depth characterize the interaction between **BPBA** and the *c-Kit2* G4.

4.2.2 RESULTS AND DISCUSSION

BPBA CHARACTERIZATION. Before studying the interaction of **BPBA** with G4, the ligand alone was analyzed in solution by NMR. Due to the relatively low solubility of **BPBA** in aqueous solution (logP = 4.85), and to slow down the prototropic tautomerism that could affect the two benzimidazole rings

(**Figure 4.11**) (Nieto et al., 2014), the NMR characterization was performed in an aprotic solvent, DMSO-*d₆*.

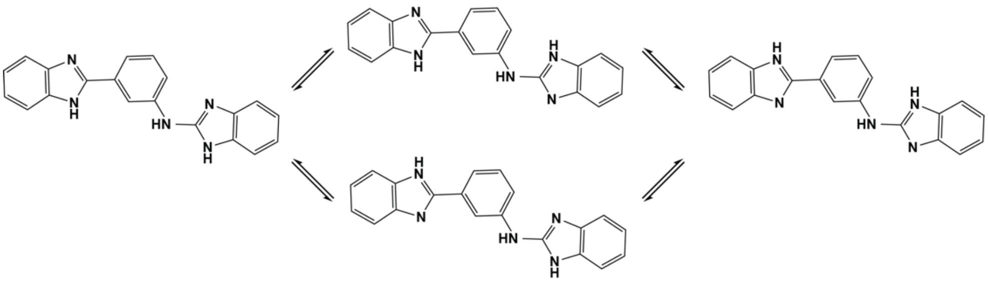


Figure 4.11 Schematic illustration of **BPBA** tautomers.

The ^{1}H NMR signals of **BPBA** were assigned by means of 2D homo- and hetero-nuclear NMR experiments. Under the employed conditions, **BPBA** is likely to be in its tricationic form, with protonation at N5, N10, and N11 (**Figure 4.12A**). In the aromatic region (6.8-8.6 ppm), a characteristic pattern of multiplet signals was observed, with the exception of one singlet peak at 8.47 ppm. The 1D ^{1}H NMR spectrum also showed three sharp singlet peaks, corresponding to each -NH group, in the 9.6-13.2 ppm chemical shift (δ) range (**Figure 4.12B**). The assignment of these was confirmed through 2D ^{15}N HSQC (**Figure 4.12C**) and NOESY experiments (mixing time of 750 ms). The chemical shift of H10, the most up-field peak resonating at 9.63 ppm, is consistent with that of a secondary aliphatic amine. On the other hand, the chemical shifts of H11 and H5, located at 11.07 and 12.88 ppm respectively, can be attributed to "pyrrole-like" nitrogen atoms. Regarding the two benzimidazole rings and the phenyl group, heteronuclear ^{13}C HSQC and ^{13}C H2BC spectra were recorded with the aim of correlating carbons and neighboring protons (two bonds). Noteworthy, this type of 2D experiments allows us to observe only protonated carbons; therefore, a ^{13}C HMBC spectrum was recorded to also verify non-protonated carbons (**Figure 4.12D**). As a result, the complete correlation map between carbons and protons in a long-range coupling (two or three bonds) was obtained.

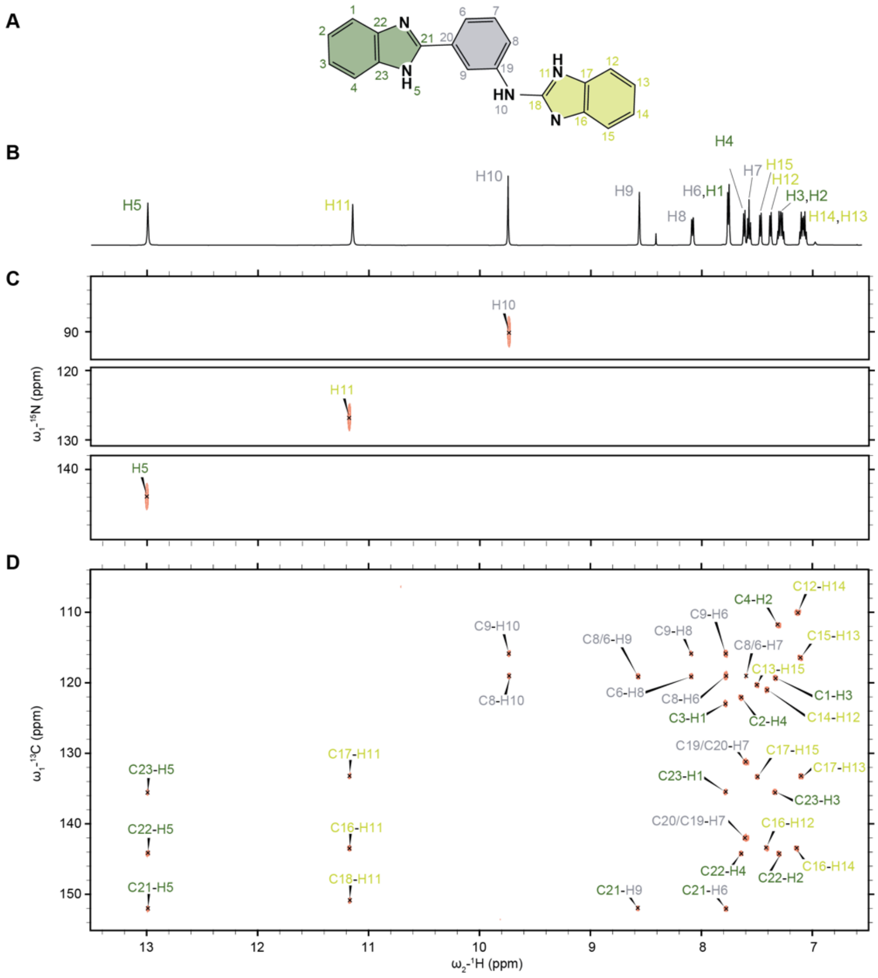


Figure 4.12 NMR characterization of **BPBA** 0.3 mM in DMSO-d_6, at 25 °C on a 600 MHz spectrometer. **A**) Schematic illustration of **BPBA** with color-coded rings and labeled protons; **B**) 1D ^{1}H NMR spectrum with corresponding peaks assignment; **C**,**D**) ^{15}N HSQC and ^{13}C HMBC 2D NMR spectra, respectively.

Homonuclear correlation experiments, such as COSY and TOCSY at different mixing times, were performed to correlate the protons of each individual ring (**Figure 4.13**). Thanks to the combination of 1D and 2D NMR experiments, the complete assignment of protons from each ring was possible.

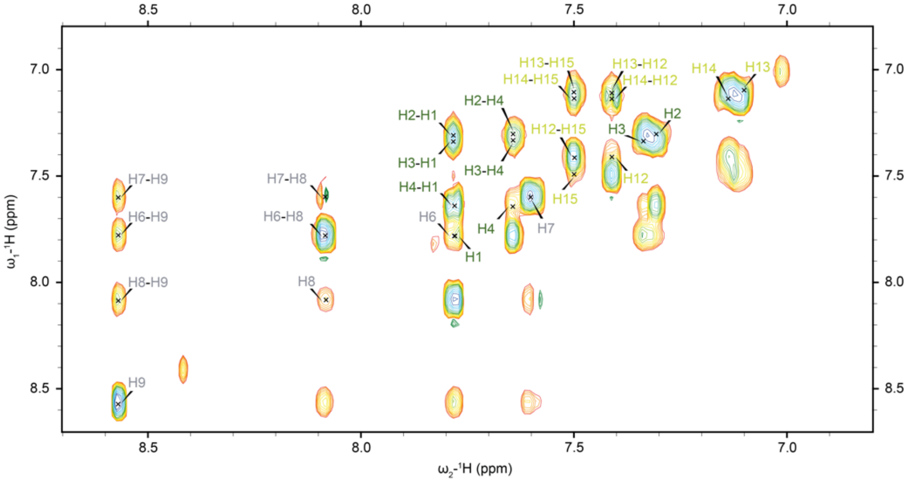


Figure 4.13 TOCSY NMR spectrum of **BPBA** (0.3 mM in 100% DMSO-d_6) recorded at 25 °C using a mixing time of 80 ms on a 600 MHz spectrometer. For reasons of clarity, the assignment of only the inter-residue cross-peaks on one side of the diagonal is reported.

The 1D ^{1}H NMR spectra of **BPBA** did not show any detectable change if recorded on freshly prepared or aged samples (up to 3 months), thus confirming the chemical stability of **BPBA** in DMSO over time.

Notably, when **BPBA** is diluted at a concentration of 0.05 mM in 5 mM KH_2PO_4/K_2HPO_4 (pH 7.0) buffer containing 20 mM KCl and 10% DMSO-d_6, it still maintained good ligand solubility, allowing for the acquisition of 1D and 2D NMR spectra. In protic solvents, prototropic tautomerism is favored, causing the signals of the two benzimidazole rings to merge in pairs, resulting in an average signal for each pair (Nieto et al., 2014). Therefore, it was not possible to assign all protons of **BPBA** under aqueous conditions. Nonetheless, a characteristic pattern could be discerned in the TOCSY spectrum acquired at 40 °C and a mixing time of 80 ms (**Figure 4.14**). For each benzimidazole ring (hereafter labeled "A" and "B"), only two sets of peaks on the diagonal were identified (A1, A2, and B1, B2), connected by symmetric off-diagonal cross-peaks. However, the assignment of each set remains ambiguous. Furthermore, due to the pseudo-symmetry of **BPBA** (where the amine linker is the only dissymmetry element in the molecule), the assignment of peaks to the corresponding benzimidazole ring remains unclear. Conversely, a preliminary assignment of the phenyl ring was possible. Ongoing experiments aim to provide a more comprehensive understanding of the assignment of **BPBA** in aqueous buffer.

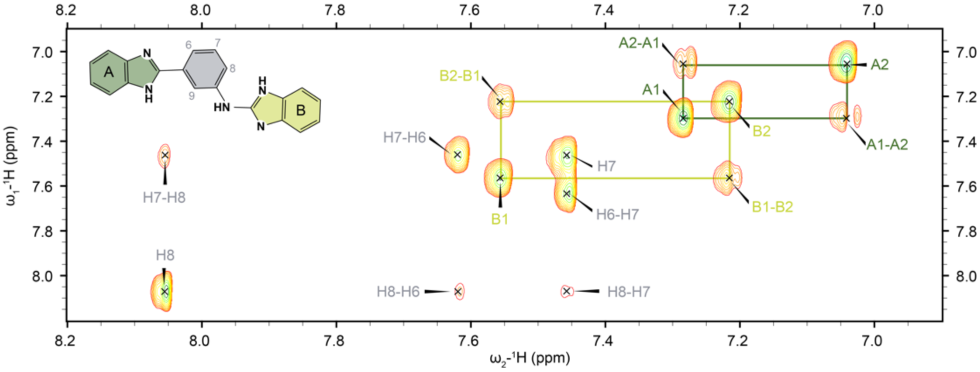


Figure 4.14 TOCSY NMR spectrum of **BPBA** (0.05 mM) in phosphate buffer containing 10% DMSO-d$_6$ recorded at 40 °C using a mixing time of 80 ms on a 600 MHz NMR spectrometer.

BPBA INTERACTION WITH G4S. The interaction of **BPBA** with *TERRA*, *c-Kit2*, and *c-Myc* G4s was monitored through 1D ^{1}H NMR titration experiments. The proper folding of the investigated G4 structures under the experimental conditions (5 mM potassium phosphate buffer, pH 7.0, containing 20 mM KCl) in the absence of the ligand was verified. Because the wild-type G4 sequences of *c-Kit2* and *c-Myc* are not suitable for study by NMR (their ^{1}H NMR spectra are affected by the presence of minor conformations), the well-characterized modified sequences (*c-Kit2-G12T/G21T* and *Myc22-G14T/G23T*) containing two G-to-T mutations at positions 12 and 21 for *c-Kit2,* and 14 and 23 for *c-Myc*, were used.

According to the literature, all three G4s exhibited high-quality ^{1}H NMR spectra featuring twelve narrow and well-resolved imino proton signals, indicative of the formation of three G-quartets (Kuryavyi et al., 2010; Ambrus et al., 2005 Martadinata & Phan, 2009). As for *TERRA,* it adopts a parallel dimeric G4 conformation characterized by high symmetry, resulting in the observation of six imino proton signals, each integrating for two protons (**Figure 4.15 A**). Upon the addition of 0.5 and 1 molar equivalents of **BPBA**, a broadening of all the ^{1}H NMR imino proton signals was observed for all G4s. At a **BPBA**/G4 ratio of 2:1, a new set of up-fielded imino proton peaks emerged for both *TERRA* and *c-Kit2* G4s, indicating the formation of the corresponding G4/ligand complexes, while the signals of free G4s were no longer detected. In addition, no new signals in the imino proton regions of *TERRA* and *c-Kit2* spectra were observed at a **BPBA**/G4 ratio of 4:1, suggesting that the titration was virtually complete at a ligand/DNA ratio of 2:1. Overall, these results suggest a possible involvement of both external G-tetrads of *TERRA* and *c-Kit2* G4s in ligand binding (**Figure 4.15**). In contrast, during the titration of *c-Myc* G4 with **BPBA**, broadening of the imino proton signals was observed up to the 1:2 molar ratio. Only at the highest ratio (1:4), a new set of twelve well-resolved up-fielded imino peaks emerged, indicating complex formation. No signals corresponding to free G4 were observed at this ratio (**Figure 4.15**). An analogous behavior to imino protons is observed for all the other protons located in the aromatic, anomeric, and methyl proton regions of the spectrum. The different binding behavior of **BPBA** to the selected G4 structures observed in NMR titrations could be explained considering the relatively lower affinity of **BPBA** for *c-Myc* G4 compared to *c-Kit2* and *TERRA* G4s (refer to **Paragraph 4.1**, or Marzano et al., 2021).

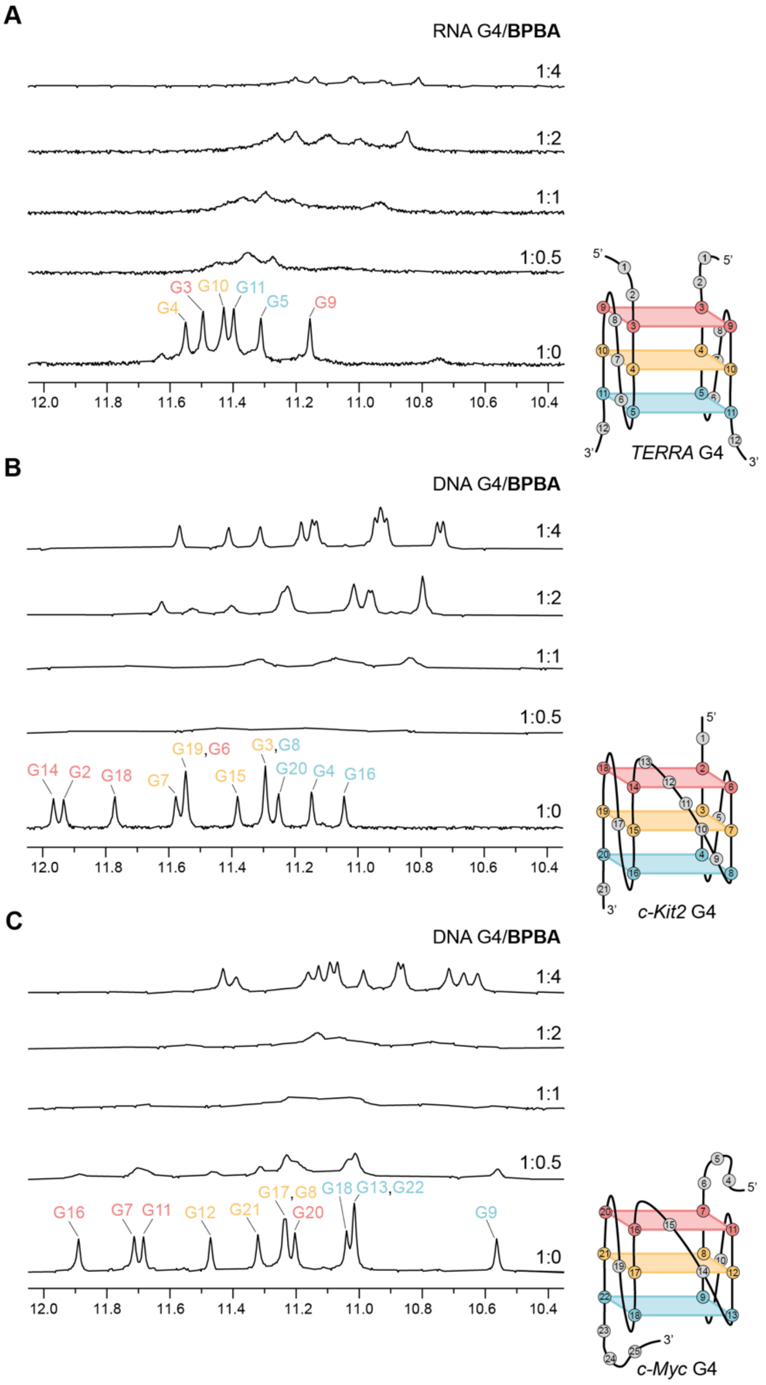


Figure 4.15 Imino proton regions of the ^{1}H NMR spectra for **A**) *TERRA*, **B**) *c-Kit2*, and **C**) *c-Myc* G4s upon titration with **BPBA** (ranging from 0.5 to 4 molar equiv). Spectra were recorded at 25 °C on a 600 MHz NMR spectrometer, with each G4 at a concentration of 0.05 mM in a 90%/10% H_2O/D_2O phosphate buffer.

To elucidate the binding mode of **BPBA**, we started by focusing our attention on its interaction with *c-Kit2* DNA G4. The NMR titration results indeed revealed that the **BPBA**/*c-Kit2* complex, formed at 2:1 ratio, represented the most suitable system for a comprehensive NMR structural investigation. This choice was driven by the sharpness of the peaks and the absence of any turbidity issues associated with the relatively low solubility of an excess of unbound **BPBA** in the solution. Consequently, a combination of 1D and 2D homo- and hetero-nuclear experiments was employed to thoroughly characterize the interaction between **BPBA** and *c-Kit2* G4.

BPBA/*C-KIT2* G4 COMPLEX CHARACTERIZATION. To unambiguously assign ^{1}H NMR resonances for the *c-Kit2* G4/**BPBA** complex (1:2 ratio), a residue-specific ^{15}N- and ^{13}C-labelled *c-Kit2* oligonucleotide was synthesized and analyzed by ^{15}N- and ^{13}C-HSQC experiments (**Figures 4.16**). Further assignments were established based on exchange cross-peaks in NOESY spectra at different mixing times (**Figures 4.17** and **4.18**, and **Table 4.5**). Similar fingerprints in the NOESY spectra of the *c-Kit2* G4 in the free and bound states suggest that the parallel G4 topology remained unchanged upon ligand binding, aligning with the results of the CD spectra reported above (**Figure 4.3A** in **Paragraph 4.1**), (Marzano et al., 2021).

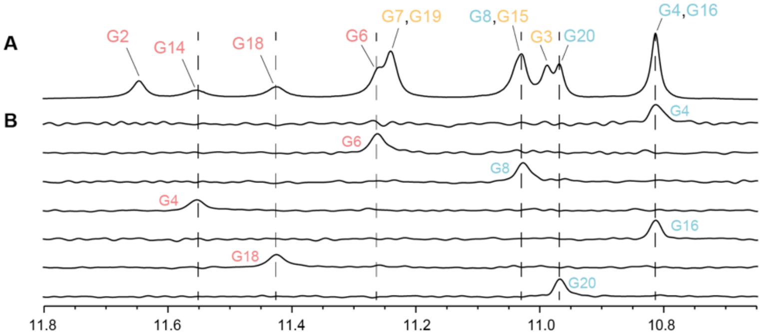


Figure 4.16 Expanded **A**) 1D ^{1}H NMR and **B**) ^{15}N-filtered 1D HSQC spectra illustrating the imino proton region of the *c-Kit2* G4/**BPBA** complex (at a 1:2 ratio). For the ^{15}N-filtered 1D HSQC spectra, samples of *c-Kit2* G4 containing 10% uniformly-^{15}N,^{13}C-labeled guanine (G) at the indicated positions, were used. All spectra were acquired at 25 °C on a 600 MHz NMR spectrometer. DNA samples were prepared at a concentration of 0.05 mM in 90%/10% H_2O/D_2O phosphate buffer.

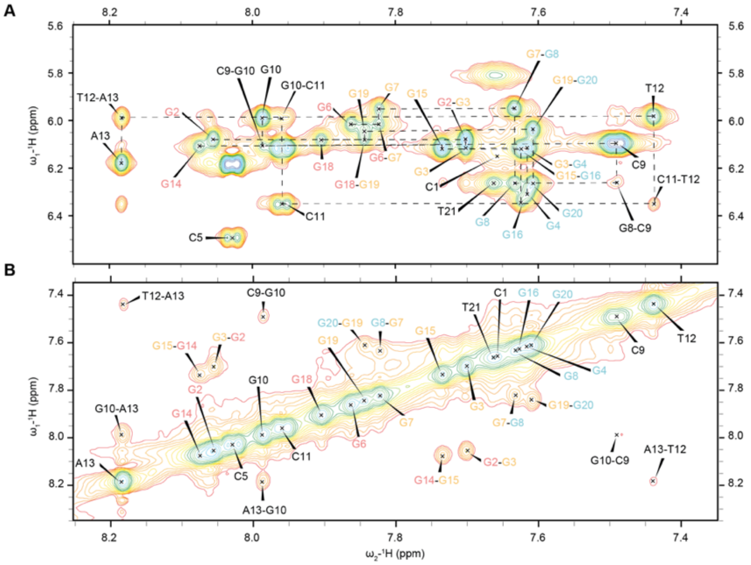


Figure 4.17 Expanded **A**) H1'-H8, and **B**) H8-H8 regions of the NOESY spectrum of *c-Kit2* G4/**BPBA** complex (1:2 ratio). The complete sequential assignment pathway is shown (dashed line). The NOESY experiment was recorded at 25 °C using a mixing time of 300 ms on a 600 MHz spectrometer. DNA was prepared at a concentration of 0.4 mM in phosphate buffer, and subsequently exchanged with D_2O (99.97%).

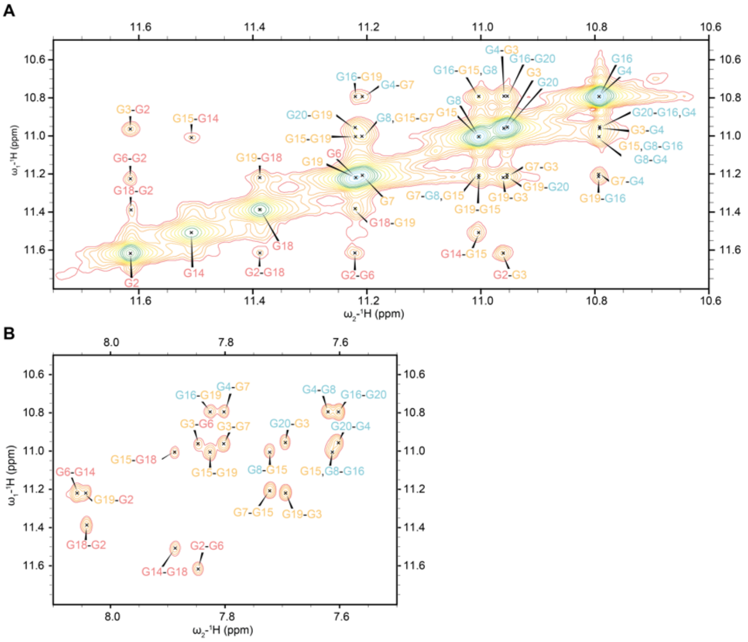


Figure 4.18 Expanded **A**) H1-H1, and **B**) H1-H8 regions of the NOESY spectrum of *c-Kit2* G4/**BPBA** complex (1:2 ratio) in 90%/10% H_2O/D_2O phosphate buffer. Spectra were recorded at 25 °C using a mixing time of 300 ms on an 800 MHz spectrometer.

The addition of 2 molar equivalents of **BPBA** caused changes in ^{1}H NMR δ of *c-Kit2* G4. The most prominently affected imino protons ($|\Delta\delta| > 0.3$ ppm) were those of G14 and G18 belonging to the 5’-end G-tetrad, along with G3, G7, G15 and G19 of the middle G-tetrad, and G4 of the 3’-end G-tetrad (**Figure 4.19A**). Regarding the aromatic protons, the overall chemical shift changes were smaller than that observed for the imino ones. The most affected ($|\Delta\delta| > 0.1$ ppm) protons included H8 from the guanines of the 5’-end G-quartet (G2, G6, G14, and G18), as well as those of G7, G15, and G19 of the middle G-tetrad, and from G4, G8, and G16 belonging to the 3’-end one (**Figure 4.19B**). Notably, the highest Δδ(H6/H8) values were exhibited by T21 and G16 both located at the 3’-end side of the G4, along with C1 at the 5’-end. Overall, these data are consistent with binding of **BPBA** to both the outer G-tetrads. Moreover, residues in loops seemed to be not significantly involved in the interaction, except for T12 which is close to 5'-end. To be noted, imino and aromatics chemical shifts of the G4/**BPBA** were almost all up-fielded with respect to the free *c-Kit2* G4, with the exception of H6/H8 for C1, C5, C11, A17 and T21.

Table 4.5 Chemical shifts (δ) of selected signals of *c-Kit2* G4 monitored upon addition of **BPBA** (2 mol equiv).

Residue/ Proton	H6/H8	H1/Me H2/H5	H1'	H2'/H2''	H3'
C1	7.660	5.811	6.151	2.156/2.546	4.750
G2	8.055	11.61	6.078	2.730/2.997	4.993
G3	7.701	10.96	6.117	2.655/2.914	5.037
G4	7.616	10.79	6.304	2.752/2.751	5.139
C5	7.888	6.184	6.494	2.401/2.727	5.082
G6	7.862	11.22	6.013	2.357/2.797	5.081
G7	7.823	11.21	5.949	2.608/2.727	4.993
G8	7.634	11.00	6.260	2.667/2.724	4.899
C9	7.491	6.096	6.103	1.709/2.322	4.643
G10	7.986	-	5.988	2.841/2.734	5.084
C11	7.958	6.115	6.352	2.335/2.668	4.894
T12	7.439	1.894	5.985	1.545/2.024	4.730
A13	8.183	8.061	6.179	2.898/2.707	5.049
G14	8.074	11.51	6.110	2.815/3.088	5.077
G15	7.735	11.00	6.119	2.669/2.927	5.084
G16	7.621	10.79	6.339	2.793/2.731	5.177
A17	8.551	8.314	6.686	2.919/2.921	5.227
G18	7.904	11.39	6.081	2.397/2.859	5.138
G19	7.844	11.22	6.040	2.659/2.769	5.087
G20	7.607	10.95	6.208	2.667/2.734	5.084
T21	7.663	1.841	6.262	2.320/2.365	4.604

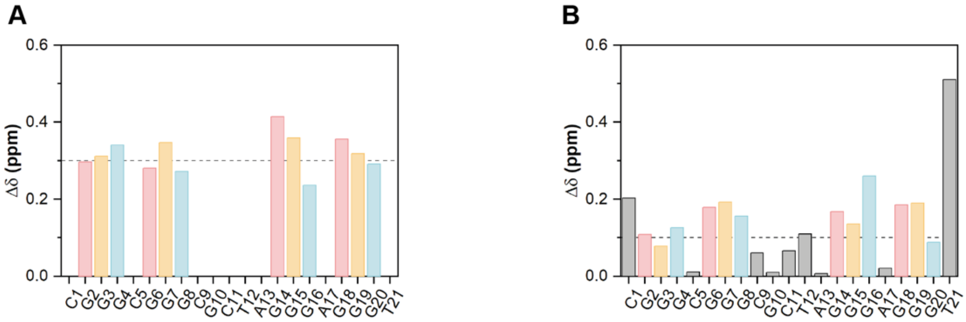


Figure 4.19 Plot of chemical shift perturbation of **A**) imino and **B**) aromatic protons of *c-Kit2* G4 alone and with twofold molar excess ***BPBA***. To facilitate comparison, Δδ values are presented in absolute terms, with associated errors within ±0.05 ppm.

In order to better characterize the interaction of 1:2 *c-Kit2* G4/**BPBA** complex, 2D NMR NOESY and TOCSY experiments were also performed on samples both in 90%/10% H_2O/D_2O and in exchanged D_2O, both at 25 and 40 °C. Indeed, a large part of the cross-peaks correlating **BPBA** and aromatic protons of the G4 overlap with the DNA fast-exchangeable signals between 7.0 and 6.6 ppm. Thus, solvent exchange in D_2O allowed us to obtain a better resolution for 1:2 *c-Kit2* G4/**BPBA** complex cross-peaks in that region. Interestingly, the similar pattern observed for **BPBA** alone aromatic signals in an aqueous solution (**Figure 4.14**) was conserved in 1:2 *c-Kit2* G4/**BPBA** complex at 40 °C, even if slightly up-fielded (**Figure 4.20**).

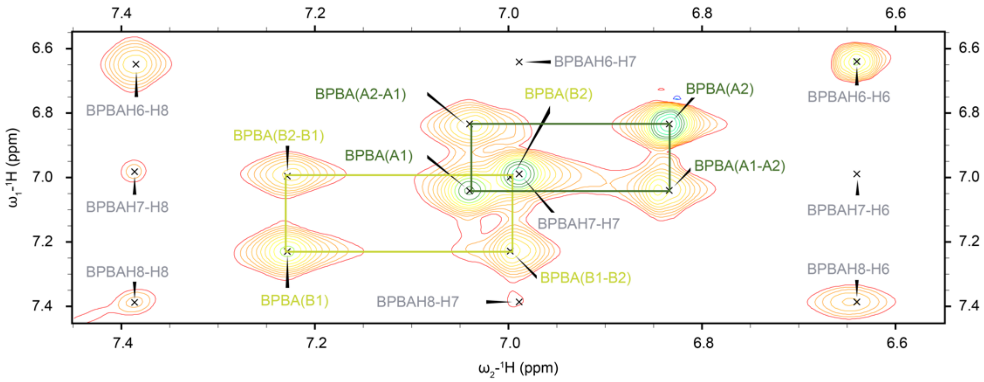


Figure 4.20 Expanded region of the TOCSY NMR spectrum acquired at 40 °C on *c-Kit2* G4/**BPBA** (at a 1:2 ratio) in 90%/10% H_2O/D_2O phosphate buffer. The experiment was conducted using a mixing time of 80 ms on a 600 MHz NMR spectrometer.

Although the **BPBA** proton signals were only partially assigned, it was still possible to identify the region of the ligand most involved in the interaction with *c-Kit2* G4. Several intermolecular NOE cross-peaks were observed between DNA protons H1 and H8 and **BPBA** (**Figures 4.21** and **4.22**). In particular, as for the DNA H1-**BPBA** NOE contacts, which are characterized by low intensity, a better resolution was obtained at 40 °C compared to 25 °C and using higher mixing times (500 ms) (**Figure 4.22**). The greatest number of NOE contacts were observed between the imino protons of all guanines involved in the G-tetrad at the 3'-end and the two benzimidazole rings of the ligand, while only G2, G6, and G18 belonging to the 5'-end G-tetrad, seem involved in the interaction. Surprisingly, no intermolecular cross-peaks were observed for **BPBA** and guanines from the middle G-tetrad, which exhibited significant Δδ values (**Figure 4.19**), suggesting they are not directly involved in ligand binding. Overall, our preliminary results suggest that two **BPBA** molecules bind to the *c-Kit2* G4 by π-π stacking on the 5' and 3' outer G-tetrads, forming a 2:1 complex, similar to a previously characterized complex involving *c-Myc* G4 and a quindoline derivative (J. Dai et al., 2011).

Furthermore, **BPBA** stacking at the 5’-end also caused the displacement of the reverse non-canonical C1:A13 base pair, which normally stacks on the G14:G18 segment of the upper G-tetrad (Kuryavyi et al., 2010). To validate this hypothesis regarding **BPBA** binding, additional analysis of 2D NOESY and TOCSY NMR spectra at different mixing times and temperatures is currently underway. Concurrently, he calculation of NMR restrains is in progress. These results are expected to be valuable for the molecular dynamics simulation of the *c-Kit2* G4/**BPBA** 1:2 complex.

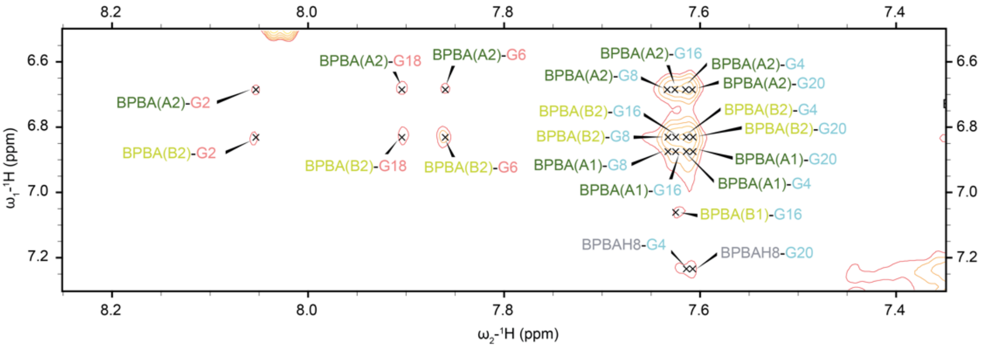
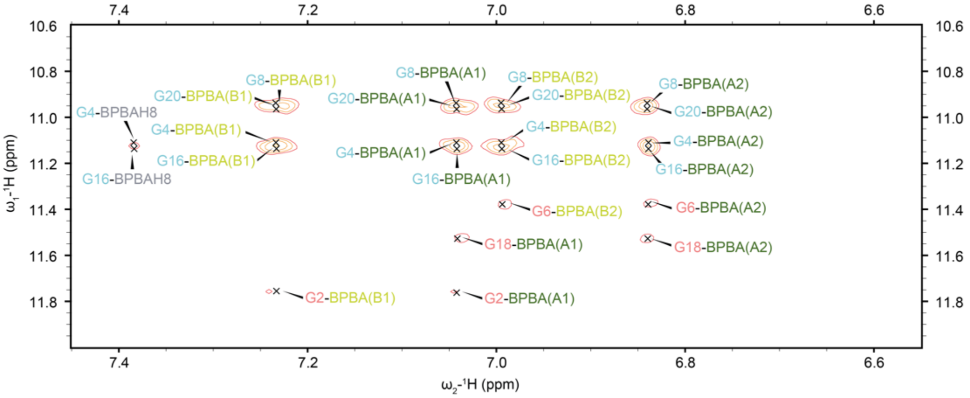


Figure 4.21 Intermolecular NOE cross-peaks between **BPBA** and *c-Kit2* G4 (H8 region). The NOESY experiment was recorded at 25 °C on *the c-Kit2* G4/**BPBA** complex (at a1:2 ratio), using a mixing time of 300 ms on a 600 MHz spectrometer. DNA was prepared at a concentration of 0.4 mM in phosphate buffer, and subsequently exchanged with D_2O (99.97%).



.

Figure 4.22 Intermolecular NOE cross-peaks between **BPBA** and *c-Kit2* G4 (H1 region). The NOESY spectrum was recorded at 40 °C on the *c-Kit2* G4/**BPBA** complex (at a 1:2 ratio) using a mixing time of 500 ms on a 600 MHz spectrometer. DNA was prepared at a concentration of 0.4 mM in 90%/10% H_2O/D_2O buffer.

4.2.3 CONCLUSIONS

To better elucidate the interaction between **BPBA**, a newly identified G4 binder, and its RNA and DNA targets (*TERRA*, *c-Kit2* and *c-Myc* G4s, respectively), a detailed structural characterization was undertaken by using a combination of 1D and 2D NMR experiments. The 1D ^{1}H NMR titration of each investigated G4 with increasing amounts of **BPBA** well agree with the corresponding ligand affinity reported in our precedent study (Marzano et al., 2021). We observed a new set of twelve well-resolved imino peaks for the G4/**BPBA** complexes at a 1:2 DNA/ligand ratio for *TERRA* and *c-Kit2* G4s, and a 1:4 ratio for *c-Myc* G4. *c-Kit2* G4/**BPBA** complex (at a 1:2 ratio) was selected as the most feasible system for a detailed structural investigation.

BPBA was structurally characterized in an aprotic solvent, DMSO-d_6, where it exhibits complete solubility. In contrast, in an aqueous buffer, prototropic tautomerism and partial solubility affected the quality of NMR spectra. Subsequently, a comprehensive assignment of the *c-Kit2* G4 in the **BPBA** bound form was accomplished. This involved synthesizing residue-specific ^{15}N- and

^{13}C-labeled *c-Kit2* G4 samples and acquiring ^{15}N- and ^{13}C-HSQC spectra of the labeled *c-Kit2* G4/**BPBA** (at a 1:2 ratio), coupled with the analysis of exchange cross-peaks in NOESY spectra at different mixing times. Indeed, we conducted NOESY and TOCSY experiments under various experimental conditions, including different temperatures, mixing times, and solvents) to elucidate the binding mode of **BPBA**. Despite some ambiguity in assigning specific ligand residues involved in the interaction with *c-Kit2* G4, we were able possible to hypothesize an end-stacking binding mode of **BPBA** to G4. This binding mode likely exhibits different affinities for 5’- and 3’-end G-tetrads. A more detailed analysis of the 2D NMR spectra, followed by NMR restraint calculations are currently underway. Subsequently, molecular dynamics simulations will be performed to validate the hypothesis regarding the end-stacking binding mode of **BPBA** to *c-Kit2* G4, considering potential differences in affinities for the 5’- and 3’-end G-tetrads.

4.2.4 EXPERIMENTAL SECTION

SAMPLE PREPARATION. Unlabeled, and residue-specific low-enrichment (10%) ^{15}N- and ^{13}C-labelled DNA oligonucleotides were chemically synthesized on a H-8 synthesizer (K&A LaborGerëate) using the standard phosphoramidite chemistry on solid-phase, followed by detachment from the resin and deprotection with aqueous ammonia. Purification and desalting of DNA oligonucleotides were performed by means of Amicon-15 centrifuge filter with 3.0 kDa MWCO. The oligonucleotide concentration was determined at 90 °C by measuring the absorbance at 260 nm, using the appropriate molar extinction coefficients. The **BPBA** stock solution for NMR studies was prepared by dissolving the compound in pure DMSO-d_6 at a concentration of 40 mM. Subsequently, it was further appropriately diluted either in pure DMSO-

d6 or in a phosphate buffer (5 mM KH_2PO_4/K_2HPO_4, pH 7.0, 20 mM KCl) containing 10% DMSO-*d6*.

NMR EXPERIMENTS. NMR data were collected over a temperature range of 5-40 °C using Bruker AVANCE NEO spectrometers operating at 600 and 800 MHz. Oligonucleotide samples were prepared at G4 concentrations of 0.05 or 0.4 mM in 90%/10% H_2O/D_2O phosphate buffer (5 mM KH_2PO_4/K_2HPO_4, pH 7.0, and20 mM KCl). NMR spectra were acquired using the DPFGSE method for water suppression. Unambiguous assignment of imino, aromatic, and methyl ^1H NMR resonances was done with the help of ^{15}N- and ^{13}C-HSQC spectra. NOESY spectra were acquired with mixing times ranging from 80 and 750 ms. The TOCSY spectrum was acquired at mixing times of 30 and 60 ms. 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) was used as a reference to calibrate the chemical shifts, assuming that the DSS resonates at 0.0 ppm. NMR spectra were processed and analyzed with the use of TopSpin 4.3.0 (Bruker), MestReNova, and Sparky (UCSF) software.

4.3 BALANCING AFFINITY, SELECTIVITY, AND CYTOTOXICITY OF HYDRAZONE-BASED G-QUADRUPLEX LIGANDS FOR ACTIVATION OF INTERFERON β GENES IN CANCER CELLS (PAPER III)

4.3.1 INTRODUCTION

G4 ligands have been explored as potentially anticancer drugs, but their effectiveness in patients has been limited, and only a few have advanced clinical trials. Recently, Miglietta *et al.* demonstrated that two G4 binders, PDS and PhenDC3, can elicit an innate immune gene response, activating interferon β (IFN-β)-dependent pathways in human cancer cells at non-cytotoxic concentrations, through micronuclei accumulation (Miglietta et al., 2021). Non-cytotoxic immune modulators could enhance the effectiveness of immunotherapy in unresponsive cancers. G4 ligands could be employed as cytostatic immune-stimulating agents in combination with anticancer immunotherapy (Miglietta et al., 2022). In 2016, our research group identified highly selective diimidazo[1,2-a:1,2-c]pyrimidine-based compounds, namely **FG** and **FIM**, capable of preferentially binding to the parallel G4 topology and stabilizing G4s in living cancer cells (Amato et al., 2016). In the present study, the relationship between G4 affinity/selectivity, cell-killing potency, and immune gene activation was investigated for a series of diimidazo[1,2-a:1,2-c]-pyrimidine derivatives of **FG** and **FIM** (**Figure 4.23**, and **Table 4.6**).

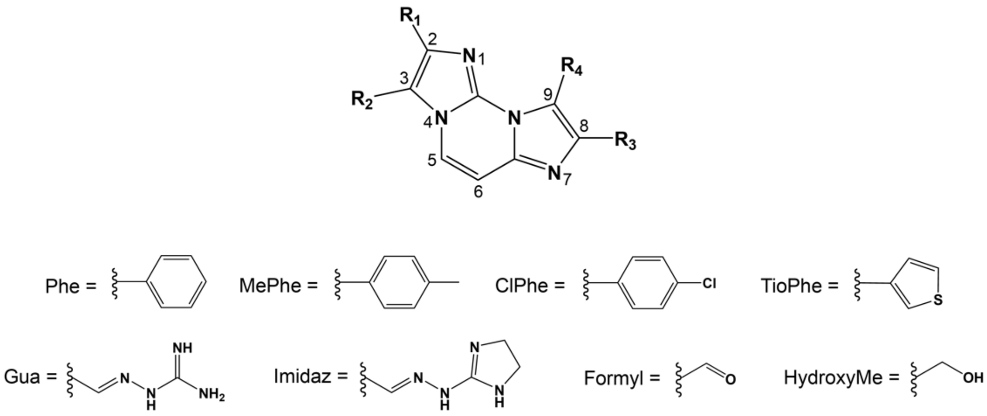


Figure 4.23 Chemical structure of diimidazo[1,2-a:1,2-c]-pyrimidine derivatives scaffold.

Table 4.6 List of substituents for each derivative on diimidazo[1,2-a:1,2-c]pyrimidine scaffold.

Compound	R_1	R_2	R_3	R_4
FG	Phe	Gua	Phe	Gua
1b	Phe	Gua	MePhe	Gua
2b	Phe	Imidaz	MePhe	Imidaz
3b	MePhe	Gua	Phe	Gua
4b	MePhe	Imidaz	Phe	Imidaz
5b	MePhe	Gua	MePhe	Gua
6b	MePhe	Imidaz	MePhe	Imidaz
7b	Phe	Gua	ClPhe	Gua
8b	Phe	Imidaz	ClPhe	Imidaz
9b	ClPhe	Gua	Phe	Gua
10b	ClPhe	Imidaz	Phe	Imidaz
11b	ClPhe	Gua	ClPhe	Gua
12b	ClPhe	Imidaz	ClPhe	Imidaz
13b	TioPhe	Imidaz	TioPhe	Imidaz
FIM	Phe	Formyl	Phe	Imidaz
14b	MePhe	Formyl	MePhe	Imidaz
15b	Phe	Formyl	Phe	Gua
16b	TioPhe	Formyl	TioPhe	Gua
17b	MePhe	Formyl	Phe	Gua
18b	MePhe	Formyl	MePhe	Gua
19b	Phe	HydroxyMe	Phe	Imidaz
20b	Phe	HydroxyMe	Phe	Gua

4.3.2 RESULTS AND DISCUSSION

To improve affinity and selectivity toward G4 structures and finely tune the biological effects of close **FG** and **FIM** analogues, a new series of molecules having different electron distribution and similar steric hindrance was designed and synthesized. The diimidazo[1,2-a:1,2-c]pyrimidine core of **FG** and **FIM** was maintained unaltered and a chlorine (electron-withdrawing) or a methyl group (electron-donating) was inserted at the para position of one or both the pending phenyl rings. Both guanidine (Gua) and imidazoline

(Imidaz) moieties were considered as positively charged chains, either to obtain FG analogues (compounds **1b-12b**, characterized by two positively charged side chains), and **FIM** analogues (compounds **13b**-**20b**, having only one positively charged side chain). In addition, since an FG analogue bearing thiophenes instead of phenyl groups (compound **13b**) proved to be a good G4 binder (Amato et al., 2016), we also considered this kind of modification along with the replacement of the Gua chains with the Imidaz ones. Finally, the formyl group of **FIM** was substituted with a primary alcohol group able to either accept or donate hydrogen bonds (compounds **19b** and **20b**). The binding ability and stabilizing effects of compounds **1b**-**20b** were investigated on the G4 structures from the NHE region of the c-Kit (*c-Kit1* and *c-Kit2*) and c-Myc (*c-Myc*) gene promoters, as well as from a 26-mer truncation of the human telomeric sequence (*Tel*$_{26}$).

CD EXPERIMENTS. According to literature, CD spectra of *c-Kit1*, *c-Kit2*, and *c-Myc* G4s showed a positive band at 264 nm and a negative one around 240 nm, suggesting they adopt parallel arrangements, whereas the CD spectrum of *Tel*$_{26}$ showed a positive band at 290 with a shoulder at ca. 268 nm and a weak negative band at around 240 nm, confirming it forms a hybrid G4 structure as the main conformation (Randazzo et al., 2013). A 27-residue hairpin oligonucleotide (*Hrp*$_{27}$) was also used as a negative control. The CD spectrum of *Hrp*$_{27}$ confirmed duplex formation, as indicated by the presence of positive and negative bands at around 280 nm and 250 nm, respectively. Next, CD spectra were recorded to test if compounds **1b**-**20b** affect the folding of G4s. Each compound was added to folded G4 or hairpin structures (in a 2:1 ligand/DNA ratio). No significant changes were observed in the CD signal for any of the analyzed DNA structures (**Figure 4.24**), indicating no alteration in DNA secondary structures upon the addition of compounds.

CD melting experiments of DNA in the absence and presence of each compound were recorded to assess the stabilizing effects of **1b**-**20b**. CD-melting curves were obtained by measuring changes in CD signals at specific wavelengths (264, 290, and 252 nm corresponding to parallel G4s, hybrid G4, and duplex structure, respectively), as a function of temperature. All compounds, except for **14b**, effectively stabilize parallel G4s over hybrid ones. As expected, molecules with a higher number of positive charges have a stronger stabilizing effect on G4s. However, compound interactions with dsDNA (*hairpin*) were also slightly increased by positive charges (**Figure 4.25**, and **Table 4.7**).

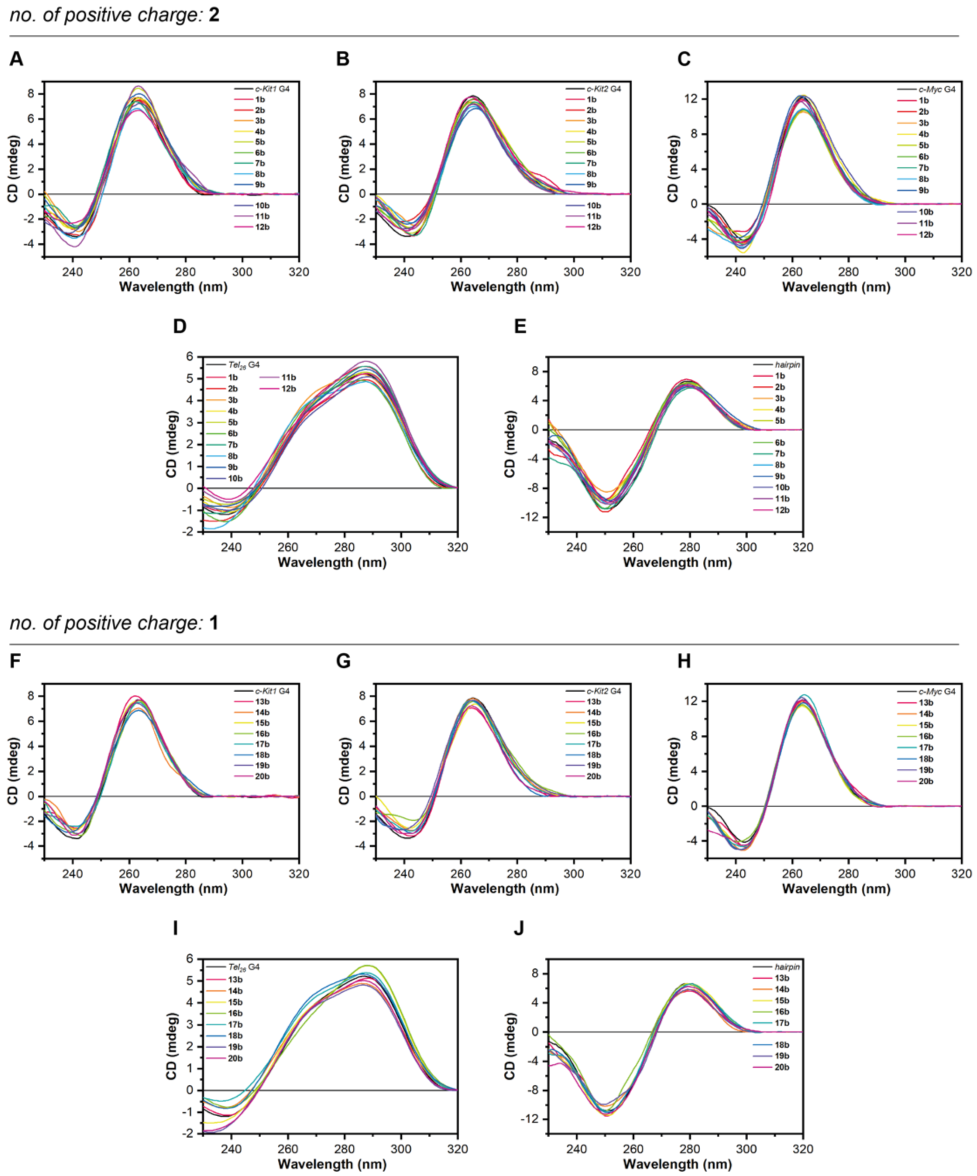


Figure 4.24 CD spectra of **A**,**F**) *c-Kit1*, **B**,**G**) *c-Kit2*, **C**,**H**) *c-Myc*, **D**,**I**) *Tel*$_{26}$ G4s, and **E**,**J**) *hairpin* in the absence (black lines) and presence (color-coded lines) of 2 molar equiv of compounds **A**,**E**) **1b-12b**, and **F**,**J**) **13b-20b**.

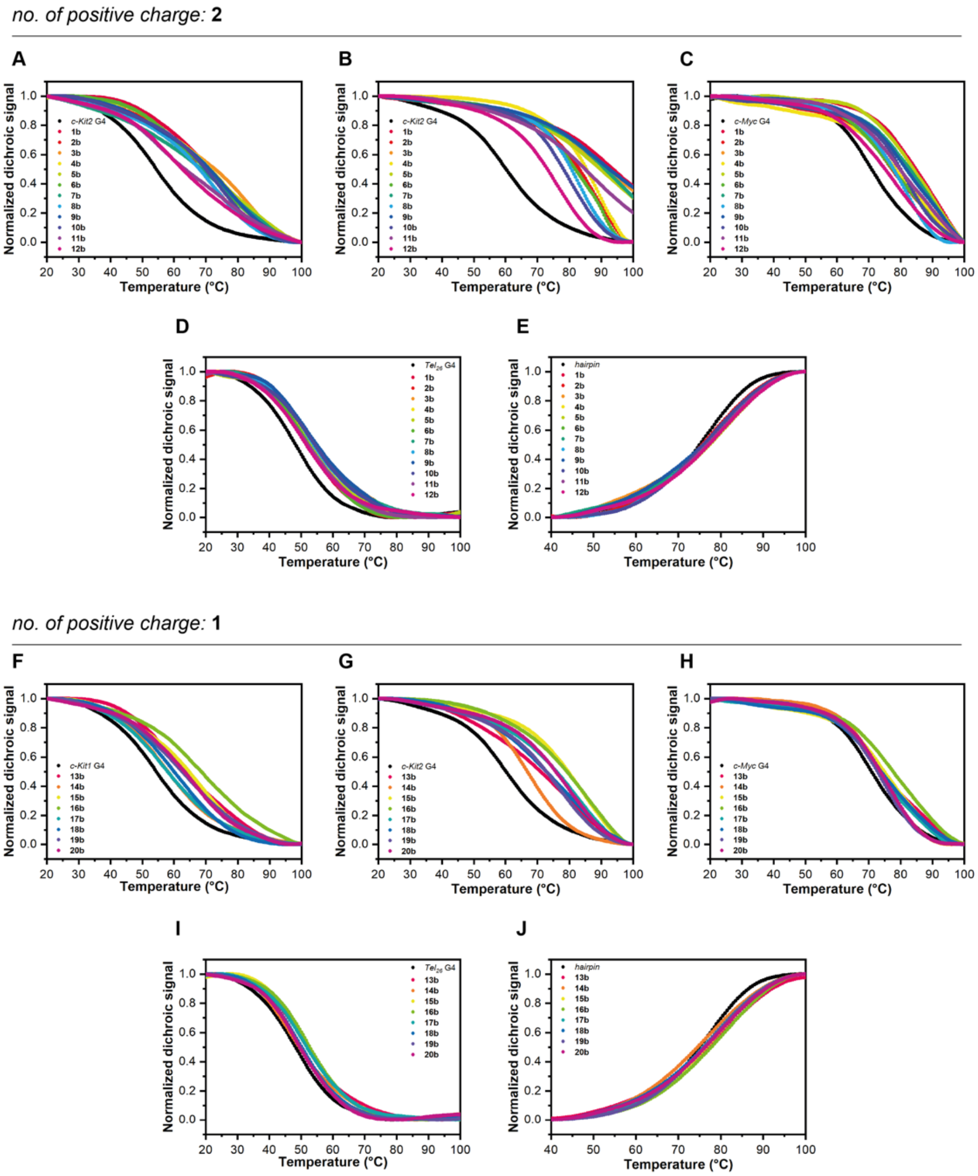


Figure 4.25 Normalized CD melting curves of **A**,**F**) *c-Kit1*, **B**,**G**) *c-Kit2*, **C**,**H**) *c-Myc*, **D**,**I**) *Tel*$_{26}$ G4s, and **E**,**J**) *hairpin* in the absence (black lines) and presence (color-coded lines) of 2 molar equiv of compounds **A**,**E**) **1b-12b**, and **F**,**J**) **13b**.

Table 4.7 Compound-induced thermal stabilization of G4 and *hairpin* structures measured by CD melting experiments.

$\Delta T_{1/2}$ (°C)[a]

Compound	*Tel*$_{26}$ G4	*c-Kit1* G4	*c-Kit2* G4	*c-Myc* G4	*hairpin*
FG[b]	-4.5	> 15.0	> 20	> 20	0.1[c]
1b	4.8 (±0.2)	18.2 (±0.2)	> 30[d]	16.9 (±0.2)	1.3 (±0.3)
2b	5.3 (±0.2)	15.2 (±0.2)	> 30[d]	9.1 (±0.3)	1.1 (±0.4)
3b	6.3 (±0.4)	25.2 (±0.5)	> 30[d]	16.9 (±0.4)	3.2 (±0.3)
4b	6.2 (±0.2)	21.4 (±0.4)	> 30[d]	13.1 (±0.6)	2.2 (±0.2)
5b	5.5 (±0.2)	18.5 (±0.3)	> 30[d]	15.6 (±0.2)	4.1 (±0.3)
6b	4.5 (±0.3)	17.1 (±0.2)	26.5 (±0.4)	8.5 (±0.2)	2.4 (±0.2)
7b	5.0 (±0.2)	21.4 (±0.4)	> 30[d]	17.6 (±0.4)	2.3 (±0.2)
8b	6.3 (±0.2)	15.5 (±0.3)	22.0 (±0.3)	7.4 (±0.3)	1.2 (±0.4)
9b	8.3 (±0.2)	20.1 (±0.3)	> 30[d]	14.3 (±0.2)	2.6 (±0.2)
10b	6.6 (±0.2)	17.7 (±0.3)	19.2 (±0.3)	10.6 (±0.3)	1.2 (±0.2)
11b	4.1 (±0.4)	11.1 (±0.2)	> 30[d]	11.4 (±0.2)	2.9 (±0.2)
12b	3.3 (±0.2)	9.6 (±0.2)	13.7 (±0.3)	5.5 (±0.2)	3.5 (±0.2)
13b	3.8 (±0.3)	10.3 (±0.2)	14.9 (±0.3)	4.7 (±0.2)	2.0 (±0.2)
FIM[b]	-3.0	2.7	9.5	> 20	-0.8[c]
14b	1.3 (±0.2)	3.8 (±0.2)	6.8 (±0.2)	1.5 (±0.2)	0.0 (±0.3)
15b	4.3 (±0.2)	11.2 (±0.2)	24.2 (±0.4)	5.8 (±0.3)	1.7 (±0.2)
16b	4.8 (±0.3)	16.1 (±0.2)	25.2 (±0.3)	9.1 (±0.3)	2.7 (±0.2)
17b	3.8 (±0.2)	6.6 (±0.3)	20.8 (±0.5)	3.1 (±0.2)	1.2 (±0.2)
18b	2.1 (±0.2)	6.1 (±0.2)	16.2 (±0.2)	4.6 (±0.2)	2.3 (±0.3)
19b	2.0 (±0.2)	7.1 (±0.2)	14.7 (±0.3)	1.7 (±0.2)	1.3 (±0.2)
20b	1.8 (±0.3)	9.7 (±0.2)	18.6 (±0.4)	2.8 (±0.2)	2.3 (±0.3)

[a] $\Delta T_{1/2}$ represents the difference in melting temperature [$\Delta T_{1/2} = T_{1/2}$ (DNA + ligand (1:2)) − $T_{1/2}$ (DNA)]. The $T_{1/2}$ values of DNA alone are: *c-Kit1* G4 = 54.0 (±0.5) °C, *c-Kit2* G4 = 61.5 (±0.5) °C, *c-Myc* G4 = 72.0 (±0.5) °C, *Tel*$_{26}$ G4 = 47.9 (±0.5) °C, and *hairpin* = 75.4 (±0.2) °C.
[b] Data from Amato *et al.*, 2018.
[c] A self-complementary 12-mer duplex-forming sequence was used as a duplex model.
[d] $\Delta T_{1/2}$ could not be determined as the compound increases hugely the thermal stability of *c-Kit2* G4.

Therefore, to further assess the selectivity for G4 structures of these compounds, the analogs showing a strong stabilizing effect on at least two G4s and a negligible effect on the *hairpin* ($\Delta T_{1/2} < 2.0$ °C) were selected (i.e. compounds **1b**, **2b**, **8b**, and **10b** among those with two positively charged side chains, and **15b**, **19b**, and **20b** among those with a positive charge only).

FRET MELTING EXPERIMENTS. FRET melting competition assays were used to further evaluate G4-stabilizing properties and G4 *vs.* duplex selectivity of designated putative ligands. The G4-forming *c-Kit1* oligonucleotide was labeled with FAM and TAMRA at the 5′ and 3′ ends, respectively, was employed (F-*c-Kit1*-T). Among the G4s more stabilized by these compounds, *c-Kit1* has the lowest $T_{1/2}$ value, allowing for a more accurate estimation of the stabilizing properties of different ligands and evaluating their ability to distinguish between G4 and duplex structures. The ligand selectivity for G4 was evaluated by adding an excess of an unlabeled competitor (*hairpin*) to the labeled G4 (F-*c-Kit1*-T). The effect of varying competitor concentrations on the $\Delta T_{1/2}$ of G4 in the presence of 2 molar equiv of each ligand was measured to assess the compounds' ability to selectively stabilize G4. The results confirmed that the selected compounds are efficient G4 stabilizers. However, compounds **1b**, **2b**, **8b**, and **10b** (with two positively charged side chains) faced challenges in the presence of the excess *hairpin* sequence. In contrast, compounds **15b**, **19b**, and **20b** (with a positive charge only) appeared to be more selective for G4 (**Figure 4.26**, and **Table 4.8**).

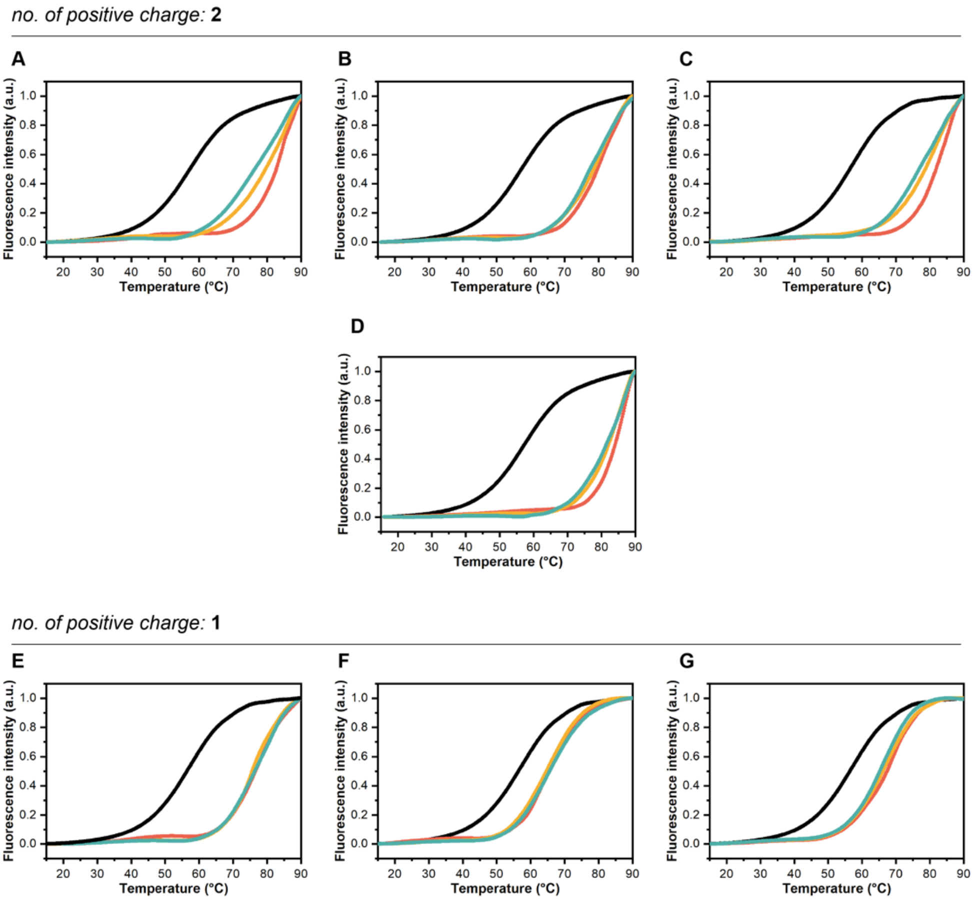


Figure 4.26 Normalized FRET melting curves for F-*c-Kit1*-T G4 in the absence (black lines) and presence (red lines) of both **A**) **1b**, **B**) **2b**, **C**) **8b**, **D**) **10b**, **E**) **15b**, **F**) **19b**, and **G**) **20b** (1:2 ratio), and large excesses of *hairpin* (1:15 and 1:50 ratio, yellow and green lines, respectively).

Table 4.8 G4-selectivity of the selected compounds measured by FRET melting competition experiments.

Compound	$\Delta T_{1/2}$ (°C)[a] F-*c-Kit1*-T G4	F-*c-Kit1*-T G4 + *hairpin* (1:15)	F-*c-Kit1*-T G4 + *hairpin* (1:50)
1b	25.4 (±0.5)	22.2 (±0.5)	19.2 (±0.5)
2b	22.4 (±0.5)	21.4 (±0.5)	20.2 (±0.5)
8b	24.6 (±0.5)	21.4 (±0.5)	19.6 (±0.5)
10b	26.6 (±0.5)	24.8 (±0.5)	23.8 (±0.5)
15b	22.0 (±1.0)	21.6 (±1.0)	22.2 (±1.0)
19b	8.0 (±0.4)	7.8 (±0.4)	8.4 (±0.4)
20b	10.0 (±0.4)	11.1 (±0.5)	12.1 (±0.5)

[a] $\Delta T_{1/2}$ represents the difference in melting temperature [$\Delta T_{1/2} = T_{1/2}$ (DNA + ligand (1:2), or *hairpin* (1:15, or 1:50)) − $T_{1/2}$ (DNA)]. The $T_{1/2}$ values of DNA alone is 57.4 (±0.2) °C. Different CD- and FRET- determined $T_{1/2}$ could be explained with modified sequence and/or experimental conditions.

FID ASSAYS. G4-FID (Fluorescent Intercalator Displacement) experiments with the fluorescent probe thiazole orange (TO) were used to measure the affinity of selected compounds for G4s (Monchaud et al., 2006). The displacement of TO from DNA by candidate ligands was monitored, allowing the determination of their relative binding affinity to the most thermally stabilized G4 structures, specifically, *c-Kit1*, *c-Kit2*, and *c-Myc* G4s. Dose-response curves were used to plot the percentage of TO displacement against compound concentration. The concentrations at which 50% displacement was achieved, known as DC_{50}, give an indication of DNA relative affinity, with lower values suggesting higher affinity. Compounds **1b**, **2b**, **8b**, and **10b** exhibited good TO displacement ability for almost all the investigated parallel G4s, indicating that ligand/G4 interaction is not sequence specific. On the other hand, compounds **15b**, **19b**, and **20b** were not able to reach 50% displacement in any case (**Figure 4.27**, and **Table 4.9**). This suggests that the presence of positive charges on ligands is correlated with their ability to displace TO, with highly cationic molecules being the most efficient. The observed apparent discrepancy between the G4-FID assay and melting experiments concerning **15b** may be attributed to the possibility that this ligand binds to G4s without directly competing with TO (Zuffo et al., 2018).

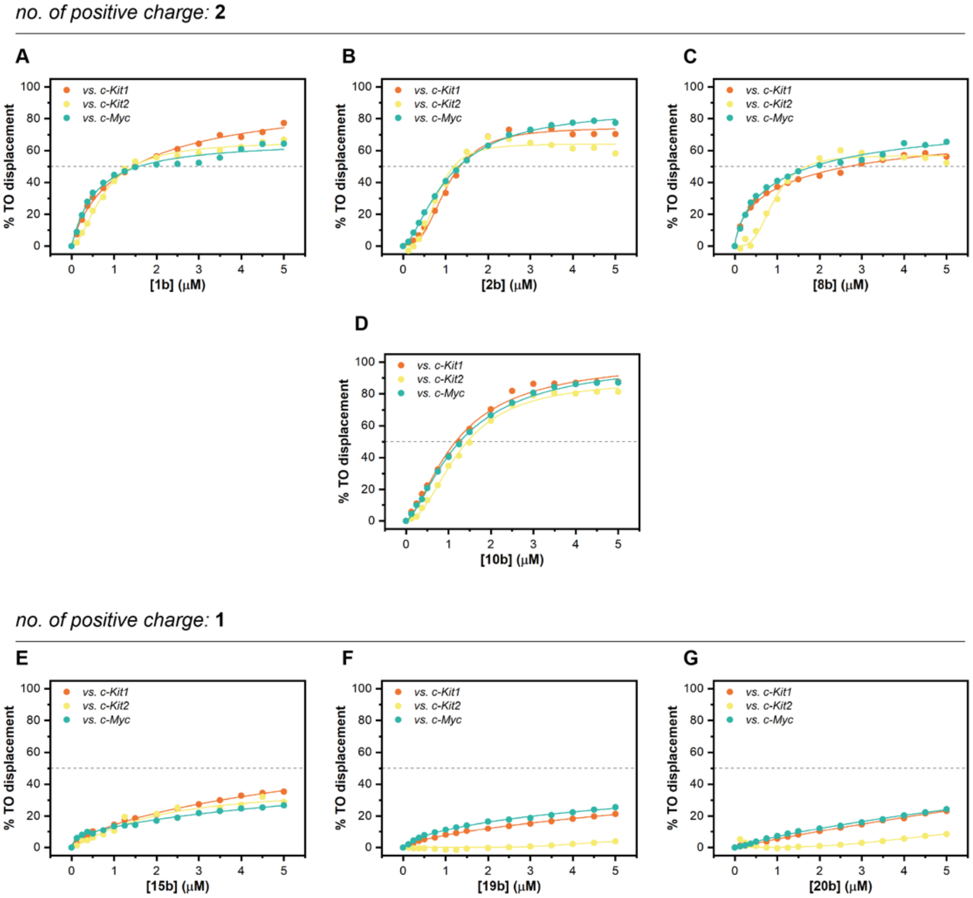


Figure 4.27 Dose-response curves from G4-FID assays for *c-Kit1*, *c-Kit2*, and *c-Myc* G4s with **A) 1b**, **B) 2b**, **C) 8b**, **D) 10b**, **E) 15b**, **F) 19b**, and **G) 20b**.

Table 4.9 Ligand DC_{50} values for *c-Kit1*, *c-Kit2*, and *c-Myc* G4s determined by G4-FID assays.

	DC_{50} (μM)[a]		
Compound	***c-Kit1* G4**	***c-Kit2* G4**	***c-Myc G4***
1b	1.4	1.4	1.6
2b	1.3	1.2	1.3
8b	2.7	1.7	1.8
10b	1.2	1.4	1.3

[a] The error in DC_{50} values is ±5%.

MST ASSAYS. MST was used to obtain quantitative data on the binding affinity of **1b**, **2b**, **8b**, **10b**, **15b**, **19b**, and **20b** for the selected G4s. Serial dilutions of ligands were prepared, mixed with a constant concentration of Cy5.5-labeled G4s (*c-Kit1*, *c-Kit2*, and *c-Myc*), and analyzed by MST. Results of the binding curves indicated that the compounds exhibited higher affinity for *c-Kit1* and *c-Kit2* G4s than for *c-Myc* G4, with a slight preference for *c-Kit1* over *c-Kit2*, except for **19b** and **20b** (**Figures 4.28**, **4.29**, and **4.30**, and **Table 4.10**). Noteworthy, compound **1b** exhibited the strongest binding affinity in the series, while compounds **19b** and **20b** had the weakest.

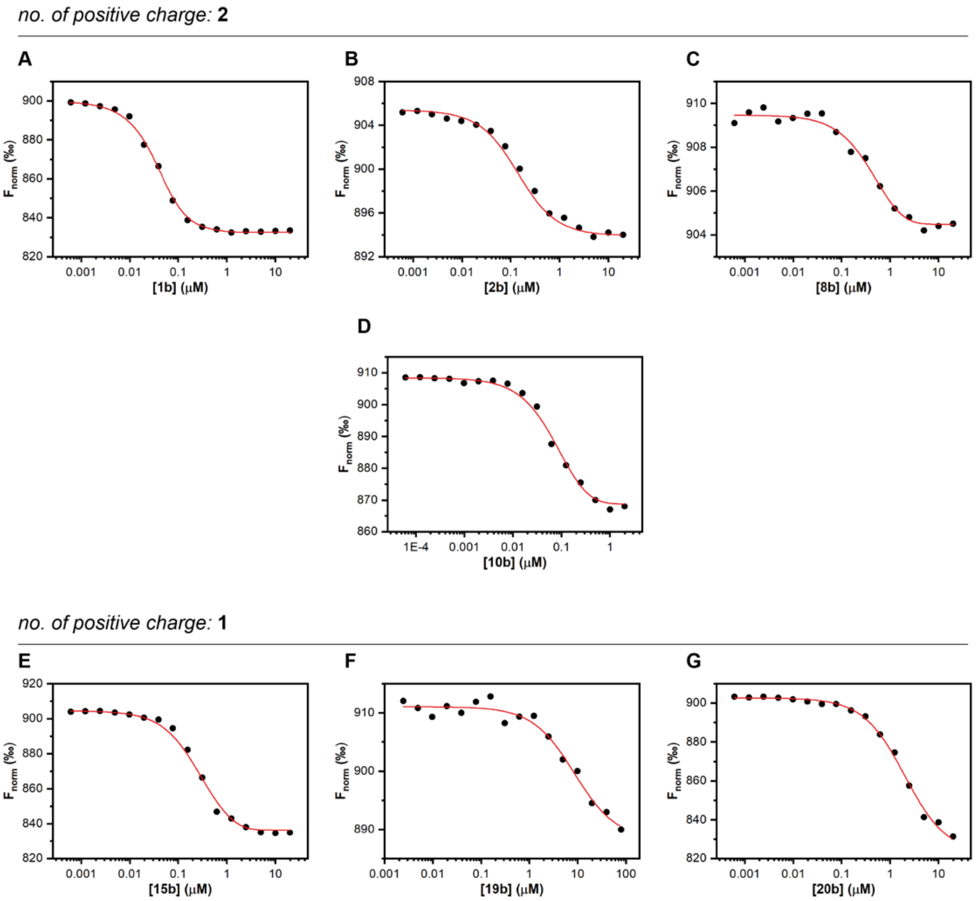
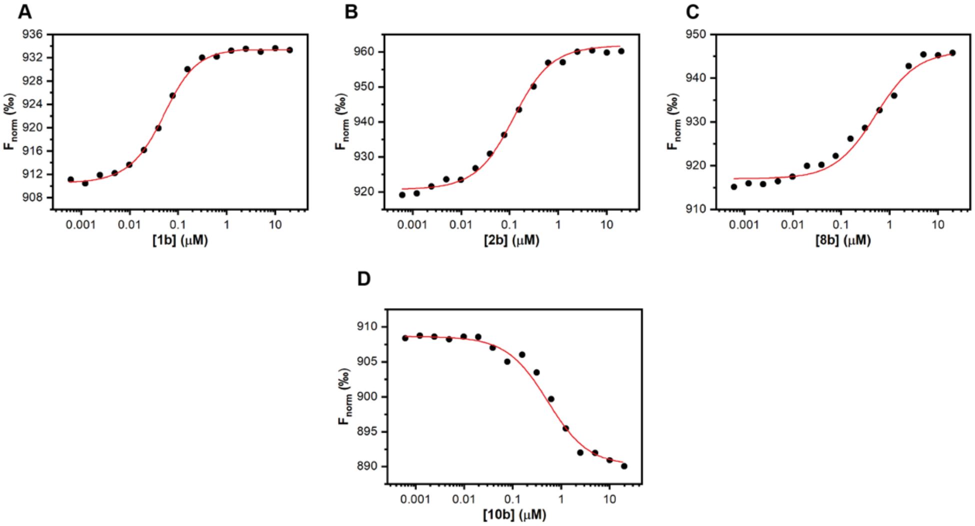


Figure 4.28 Dose-response curves from MST experiments for the interaction of **A**) **1b**, **B**) **2b**, **C**) **8b**, **D**) **10b**, **E**) **15b**, **F**) **19b**, and **G**) **20b** with Cy5.5-*c-Kit1* G4.

no. of positive charge: **2**



no. of positive charge: **1**

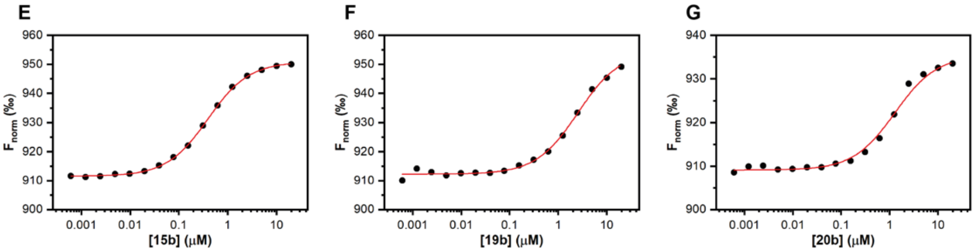


Figure 4.29 Dose-response curves from MST experiments for the interaction of **A**) **1b**, **B**) **2b**, **C**) **8b**, **D**) **10b**, **E**) **15b**, **F**) **19b**, and **G**) **20b** with Cy5.5-*c-Kit2* G4.

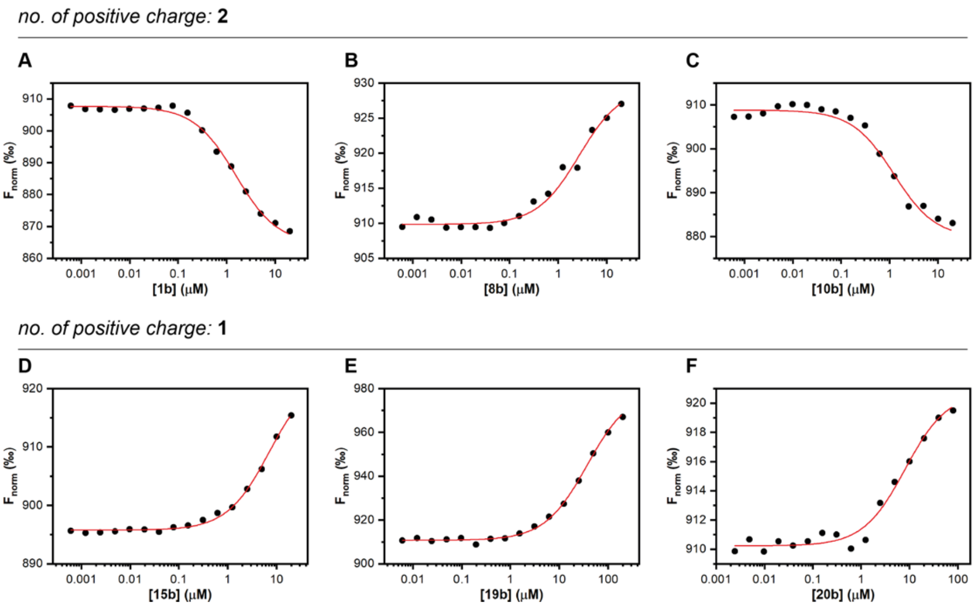


Figure 4.30 Dose-response curves from MST experiments for the interaction of **A) 1b**, **B) 2b**, **C) 8b**, **D) 10b**, **E) 15b**, **F) 19b**, and **G) 20b** with Cy5.5-*c-Myc* G4.

Table 4.10 K_D values for ligands to *c-Kit1*, *c-Kit2*, and *c-Myc* G4s determined by MST experiments.

		K_D (µM)	
Compound	***c-Kit1* G4**	***c-Kit2* G4**	***c-Myc G4***
1b	0.03 (±0.01)	0.04 (±0.01)	1.5 (±0.4)
2b	0.12 (±0.02)	0.13 (±0.02)	n.d.[a]
8b	0.38 (±0.09)	0.49 (±0.09)	3.0 (±1.0)
10b	0.07 (±0.02)	0.50 (±0.04)	1.2 (±0.4)
15b	0.30 (±0.04)	0.37 (±0.03)	7.0 (±1.0)
19b	8.0 (±2.0)	2.5 (±0.3)	37 (±1)
20b	2.0 (±0.2)	1.3 (±0.2)	8.0 (±3.1)

[a] n.d. = not determined.

Based on the entire set of biophysical data, compounds were classified into three groups: *i*) strong binders (compounds **1b**, **2b**, and **10b**); *ii*) moderate binders (**8b** and **15b**); *iii*) modest binders (**19b** and **20b**).

CYTOTOXICITY ASSAYS. The cytotoxic potencies of compounds **1b**, **2b**, **8b**, **15b**, **19b**, and **20b**, together with their precursor, **FIM,** and **FG**, were determined. The assessments were conducted in human osteosarcoma U2OS and murine fibrosarcoma MNMCA1 cells, known for producing high levels of IFN-β. **PDS** was used as a reference compound (Nanni et al., 1992). **FIM** and **15b** displayed higher IC_{50} values compared to **19b** and **20b**, respectively, suggesting that an aldehyde moiety imparts greater cytotoxicity than an alcohol group. Furthermore, compound **8b** exhibited higher cytotoxic than **1b** and **2b** (**Table 4.11**). The tested analogs demonstrated similar cytotoxic potency in normal human MRC5 fibroblasts, as expected.

Table 4.11 Cytotoxic potency of selected hydrazone derivatives in U2OS and MNMCA1 cells.

IC_{50} (µM)[a]

Compound	Human U2OS	Murine MNMCA1
FG	15.9 (±1.2) [c]	n.d. [a]
1b	46.8 (±12.7)	38.5 (±5.8)
2b	108.3 (±35.4)	28.6 (±8.4)
8b	20.2 (±1.0)	23.5 (±5.1)
FIM	4.0 (±0.33)	2.3 (±0.92)
15b	2.6 (±0.87)	1.9 (±0.33)
19b	24.5 (±1.2)	12.1 (±2.5)
20b	14.3 (±9.3)	35.7 (±0.67)
PDS	> 50 [c]	27.0 (±16.0)

[a] Treatments were for 24 h in exponentially growing cells. Cell survival was evaluated with the MTT test after 48 h of cell recovery in drug-free medium.
[b] n.d. = not determined.
[c] See *ref.* (De Magis et al., 2019)

G4 STABILIZATION, AND DNA DAMAGE ASSAYS. Immunofluorescence (IF) assays were conducted to determine the induced stabilization of G4 structures in U2OS cells using the BG4 antibody, with **PDS** serving as the positive control (De Magis et al., 2019). Compounds **1b** and **2b**, which induce higher thermal stabilization of *c-Kit2* and *c-Myc G4s*, can stabilize G4 structures in living cells similarly to **PDS**. In contrast, compound **8b** was ineffective, indicating a different type of interaction (**Figure 4.31A**). Even though compounds **FIM**, **15b**, **19b**, and **20b** showed a significantly reduced affinity for the tested G4 structures compared to **1b**, **2b**, and **8b**, their cytotoxic potencies are either equal or higher. Moreover, they all effectively stabilized G4 structures in living cells. In particular, the "Gua" moiety in compounds **15b** and **20b** exhibits better interactions with G4 compared to the "Imidaz" one in **FIM** and **19b**. This preference leads to a somewhat higher stabilization, likely attributed to its increased flexibility (**Figure 4.32A**). The evaluation of DNA damage, indicated by the levels of S139-phosphorylated histone H2AX (γH2AX), was conducted in U2OS cancer cells, with **PDS** used as a reference. Compound **8b** was found to induce DNA damage even in the absence of stabilization in the cell. On the other hand, **2b** induced both G4 stabilization and DNA damage, suggesting a distinct cytotoxicity mechanism (**Figure 4.31B**). Compounds with only one positive charge in their side chains increased γH2AX levels at comparable levels in cancer cells, with **15b** showing slightly less effectiveness (**Figure 4.32B**).

MICRONUCLEI AND IFN-β ACTIVATION ASSAYS. Given that non-cytotoxic doses of G4 binders, such as **PDS** and **PhenDC3**, can activate IFN-β-dependent pathways through micronuclei induction in human cancer cells, we evaluated the ability of the new analogs to induce micronuclei formation. At similar cytotoxic concentrations, compounds **1b**, **2b**, and **8b** induced almost the same levels of micronuclei in MNMCA1 cells, but less than those induced by **PDS**, which also resulted in higher IFN-β levels compared to the tested analogs. Notably, **1b** was more effective in inducing IFN-β, even at non-cytotoxic concentrations, than **2b**, whereas **8b** was completely ineffective, even at higher concentrations. Therefore, compound **1b**, which promotes *in vivo* G4 stabilization without causing DNA damage, can induce the expression of

IFN-β gene at non-cytotoxic concentrations. Differences in the activation of the autophagic pathway might elucidate the variations in IFN-β production between analog **1b** and the other investigated derivatives (**Figure 4.31C,D**). As for **FIM**, **15b**, **19b**, and **20b**, they did not affect significantly IFN-β expression, showing a low, if any, with a maximum of less than two-fold change for **FIM**. No difference was observed between derivatives bearing "Gua" or "Imidaz" groups as chains (**Figure 4.32C,D**). The IFN-β activation was independent of the induced micronuclei level, suggesting that cytosolic DNA from micronuclei was necessary but not sufficient for immune gene expression (Crowl et al., 2017; Pilger et al., 2021). However, a high ligand affinity for G4 may still be required for immune gene activation.

In conclusion, compounds with one positively charged side chain exhibit less affinity for G4s but are more selective binders compared to those with two positively charged side chains. Compound **8b** is more cytotoxic than **1b** and **2b**; however, it minimally stabilizes G4 in nuclei and does not trigger IFN-β production. On the contrary, **1b** shows high G4 stabilization *in vivo* and the least cytotoxic potency in murine cells, where it triggers a good activation of IFN-β genes. Conversely, **FIM** analogs bearing the chemically reactive aldehyde group (**FIM** and **15b**) or the hydroxymethyl group (**19b** and **20b**) exhibit greater cytotoxic potency than **FG** derivatives, with compounds **FIM** and **15b** being the most cytotoxic compounds of the series. In addition, they can stabilize G4 in cells but are not able to trigger IFN-β activation.

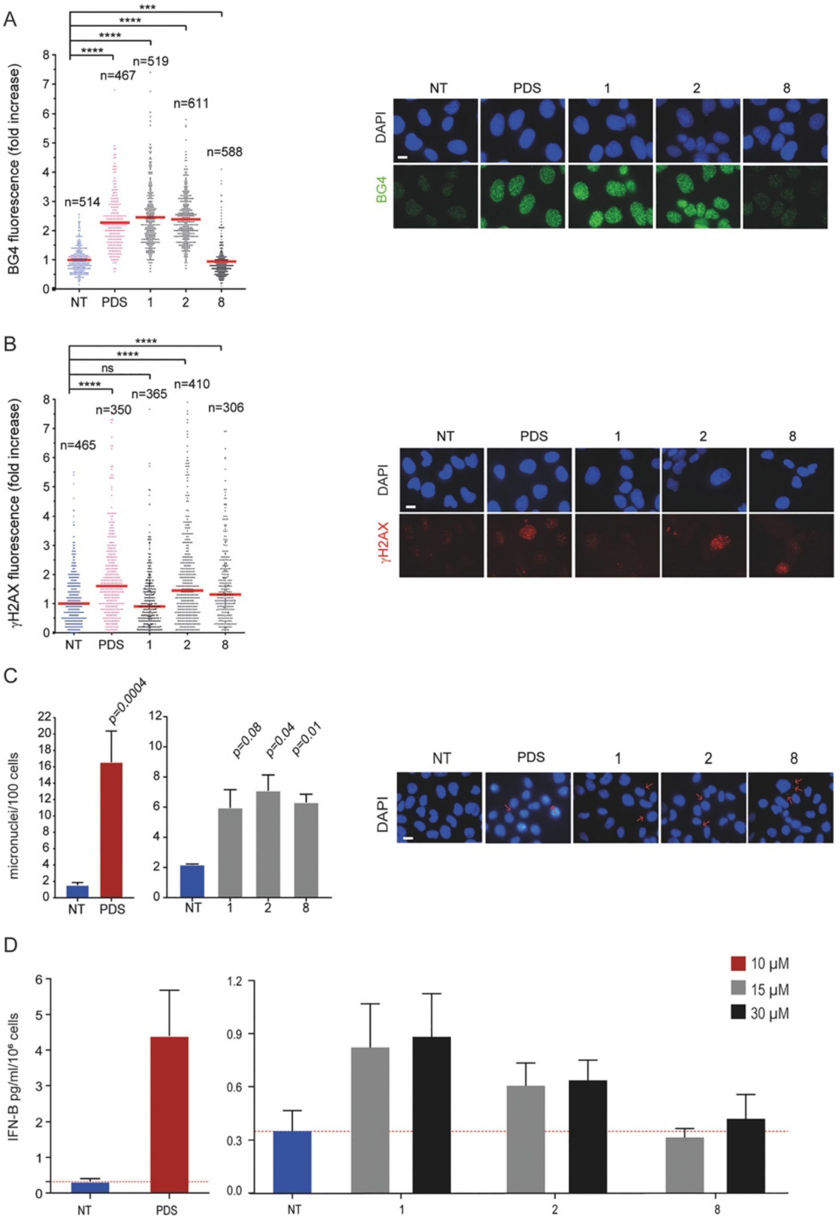


Figure 4.31 G4 stabilization, DNA damage, and IFN-β stimulation induced by **PDS**, **1b**, **2b**, and **8b**. **A**) Quantification of fluorescence signals of BG4 foci in U2OS cells being treated with compounds. **B**) Quantification of fluorescence signals of γH2AX in U2OS cells being treated with compounds at IC_{50} concentrations. **C**) Micronuclei quantification by DAPI staining in MNMCA1 cells treated (24 h) of treatment followed by 24 h of drug-free recovery. The scale bar is 10 µm. **D**) Quantification of IFN-β produced by MNMCA1 cells treated with compounds at different concentrations. Significance in all the graphs was calculated by the Mann-Witney test (*p < 0.05, **p > 0.01, ***p > 0.001, and ****p < 0.0001).

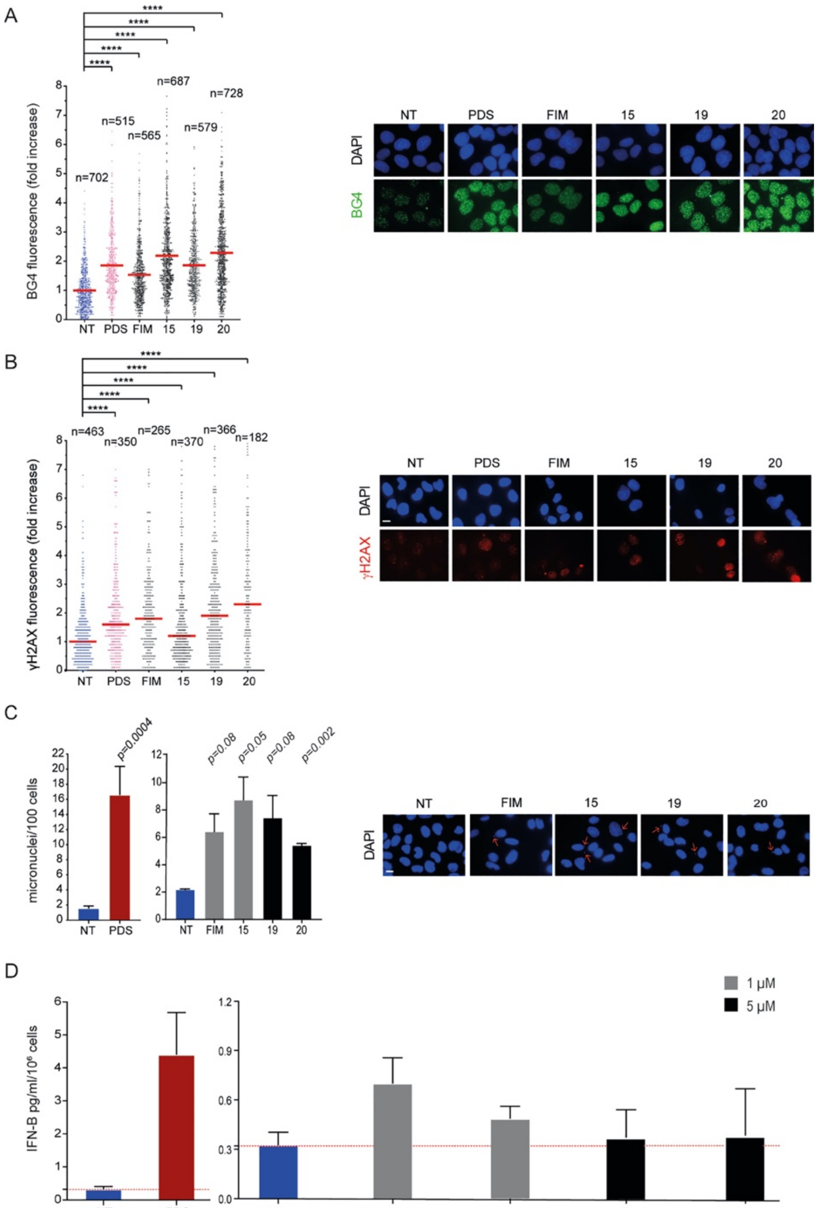


Figure 4.32 G4 stabilization, DNA damage, and IFN-β stimulation induced by **PDS**, **FIM**, **15b**, **19b**, and **20b**. **A**) Quantification of fluorescence signals of BG4 foci in U2OS cells being treated with compounds. **B**) Quantification of fluorescence signals of γH2AX in U2OS cells being treated with compounds at IC_{50} concentrations. **C**) Micronuclei quantification by DAPI staining in MNMCA1 cells treated (24 h) of treatment followed by 24 h of drug-free recovery. The scale bar is 10 μm. **D**) Quantification of IFN-β produced by MNMCA1 cells treated with compounds at different concentrations. Significance in all the graphs was calculated by the Mann-Witney test (*p < 0.05, **p > 0.01, ***p > 0.001, and ****p < 0.0001).

4.3.3 CONCLUSIONS

By comparing closely related G4 binders, we have been able to explain the significance of maintaining a delicate equilibrium between G4 affinity/selectivity and cytotoxicity for the activation of immune genes. Specifically, achieving a high G4 affinity and maintaining a relatively low cytotoxic potency is essential for a G4 ligand to effectively activate immune genes. Therefore, we propose a new approach, emphasizing low cell-killing potency and high G4 affinity, for the discovery of effective anticancer G4 ligands with immune-stimulation activity.

4.3.4 EXPERIMENTAL SECTION

OLIGONUCLEOTIDE SYNTHESIS AND SAMPLE PREPARATION. The following oligonucleotides sequences, employed in this investigation, were chemically synthesized on the 1-μmol scale on an ABI 394 DNA/RNA synthesizer (Applied Biosystems, CA, USA) by using the standard β-cyanoethyl phosphoramidite solid-phase chemistry, as described elsewhere (Amato et al., 2020):

- *c-Kit1* d(AGGGAGGGCGCTGGGAGGAGGG);
- *c-Kit2* d(CGGGCGGGCGCTAGGGAGGGT);
- *c-Myc* d(TGAGGGTGGGTAGGGTGGGTAA);
- *Tel*$_{26}$ d(TTAGGGTTAGGGTTAGGGTTAGGGTT);
- *Hairpin* d(CGCGAATTCGCGTTTCGCGAATTCGCG).

After synthesis, oligonucleotides were detached from the support and deprotected by treating with an aqueous solution of concentrated ammonia at 55 °C, for 17 h. The filtrates and washings, after being combined and concentrated under reduced pressure, were solubilized in water and purified using a high-performance liquid chromatography system equipped with a Nucleogel

SAX column (Macherey-Nagel, 1000-8/46), using a 30 min linear gradient from 100% (20 mM KH_2PO_4/K_2HPO_4 aqueous solution (pH 7.0), 20% (v/v) CH_3CN) to 100% buffer B (20 mM KH_2PO_4/K_2HPO_4 aqueous solution (pH 7.0), 20% (v/v) CH_3CN containing 1.0 M KCl) at a flow rate of 1 mL min^{-1}. The purified fractions were then desalted by means of C-18 cartridges (Sep-Pak). The purity of the isolated oligomers was checked by NMR and proved to be higher than 98%. All oligonucleotides were dissolved in a buffer solution consisting of 5 mM KH_2PO_4/K_2HPO_4 (pH 7.0) and 20 mM KCl (or LiCl in the case of *c-Myc* because of its high thermal stability). The concentration of each oligonucleotide was verified by measuring the UV absorption at 90 °C, considering the appropriate molar extinction coefficient values ε (λ = 260 nm) calculated using the nearest-neighbor model (Cantor et al., 1970). Finally, to achieve the correct folding of the DNA sequences, oligonucleotide solutions were annealed by heating at 95 °C for 5 min followed by a slow cooling to room temperature and storage overnight at 4 °C.

CD EXPERIMENTS. CD experiments were performed on a Jasco J-815 spectropolarimeter equipped with a PTC-423S/15 Peltier temperature controller. All the spectra were recorded at 20 and 100 °C in the wavelength range of 230-320 nm and averaged over three scans. A scan rate of 100 nm/min, with a 0.5 s response time and 1 nm bandwidth, was used. The buffer baseline was subtracted from each spectrum. For the CD experiments, 10 μM G4 and 15 μM duplex DNA in the absence or presence of 2 molar equiv of ligand were used. CD spectra were recorded 10 min after ligand addition. Ligand stock solutions were 10 mM in DMSO. CD melting experiments were carried out in the 20-100 °C temperature range at a 1 °C/min heating rate and recorded both in the absence and presence of compounds (2 molar equiv) added to the folded nucleic acid structures. The apparent melting temperatures ($T_{1/2}$) were determined from a curve fit using OriginPro 2021 software (OriginLab Corp., MA, USA). Normalization of melting curves between 0 and 1 was performed to better compare the results. In cases where the melting process was not completed even at 100 °C, due to an exceptional ligand-induced G4 thermal stabilization, the relative melting curves were normalized by dividing only by the maximum.

FRET MELTING EXPERIMENTS. Measurements were carried out on a Jasco FP-8300 spectrofluorometer equipped with a Peltier temperature controller system (PCT-818) using a dual-labeled G4-forming sequence FAM-[d(`AGGGAGGGCGCTGGGAGGAGGG`)]-TAMRA (F-*c-Kit1*-T). The oligonucleotide was dissolved in water at 1 mM, diluted at 1 μM in the buffer reported above, and annealed by heating to 90 °C for 5 min, followed by slow cooling to room temperature overnight and storage at 4 °C for 24 h before data acquisition. Experiments were performed in sealed quartz cuvettes with a path length of 1 cm by using 0.2 μM prefolded F-*c-Kit1*-T G4 target, in the absence and presence of 2 molar equiv of the ligand and of the duplex competitor at 3 and 10 μM final concentrations. In addition, an experiment in the absence of compounds and competitors was also performed. Fluorescence spectra were acquired before and after the melting, at 15 and 90 °C, respectively. The dual-labeled oligonucleotide was excited at 492 nm, and emission spectra were recorded between 500 and 650 nm by using a 100 nm s^{-1} scan speed. Excitation and emission slit widths were both set to 5 nm. FRET melting experiments were performed by monitoring the emission of FAM at 520 nm (upon excitation at 492 nm), using a heating gradient of 0.2 °C min^{-1} over the range of 15-90 °C. Emission of FAM was normalized between 0 and 1. Final analysis of the data was carried out using OriginPro 2021 software.

TO G4-FID ASSAYS. A solution containing 0.25 μM G4 DNA (*c-Kit1*, *c-Kit2*, or *c-Myc*) and 0.5 μM TO in the above reported buffer was prepared in a 1 cm-path length cell, and the corresponding fluorescence spectrum was acquired in the absence and presence of increasing concentrations of selected compounds (1 mM stock solution in DMSO). Each ligand addition (from 0.5 to 20 molar equiv) was followed by a 3 min equilibration time before spectrum acquisition. Measurements were run at 20 °C on a Jasco FP-8300 spectrofluorometer equipped with a Peltier cell holder (PCT-818), using an excitation wavelength of 485 nm and recording the emission spectra in the 500-650 nm wavelength range. Both excitation and emission slits were set at 5 nm. Final analysis of the data was carried out using OriginPro 2021 software. The percentage of TO displacement was calculated as described in Paragraph 4.1.4. The percentage of displacement was then plotted as a function of the ligand concentration, and DC_{50} was calculated as the required concentration to displace 50% TO. Each titration was performed in duplicate.

MST EXPERIMENTS. MST measurements were performed using a Monolith NT.115 instrument (NanoTemper Technologies). The Cy5.5 fluorescently labeled oligonucleotides (*c-Kit1*, *c-Kit2*, or *c-Myc*) were prepared and annealed at 1 μM in the above described conditions. DNA samples were then diluted using the same phosphate buffer supplemented with 0.1% Tween. For the MST experiments, the concentration of the labeled oligonucleotides was kept constant at 20 nM, while a serial dilution of the ligand (1:2 from 5.0, 40, 160, or 400 μM ligand stock solution) in the same buffer used for DNAs was prepared and mixed with the oligonucleotide solution with a volume ratio of 1:1. All the samples, containing 20% DMSO as the final concentration, were loaded into standard capillaries (NanoTemper Technologies). Measurements were performed and analyzed as previously reported (Amato et al., 2018).

The compound library was synthesized in collaboration with Rita Morigi, Alessandra Locatelli, and Alberto Leoni from the Department of Pharmacy and Biotechnology of the University of Bologna (Italy). Cell lines treatments, MTT cell proliferation assays, IF microscopy measurements, and IFN-B ELISA assays were performed in collaboration with Giovanni Capranico, Giulia Miglietta, Jessica Marinello, Andrea Arleo, and Monica Procacci from the Department of Pharmacy and Biotechnology of the University of Bologna (Italy). For further information, see APPENDIX.

4.4 **LIGAND-BASED DRUG REPURPOSING STRATEGY IDENTIFIED SARS-COV-2 RNA G-QUADRUPLEX BINDERS** (PAPER IV)

4.4.1 INTRODUCTION

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is the highly contagious virus responsible for recent COVID-19 pandemic (F. Wu et al., 2020). Despite the effectiveness of SARS-CoV-2 vaccination, finding drugs to treat the viral infection remains an urgent task (Pouwels et al., 2021). The conventional approach to discovering new drugs has been proven to be a lengthy and often unsuccessful process. However, the repurposing of pre-approved medications has emerged as a more promising alternative. By taking advantage of de-risked compounds, this strategy has the potential to significantly reduce development expenses and timelines, making it a more confident and effective approach to drug discovery (Pushpakom et al., 2019). As many drugs have already been successfully repurposed to treat various diseases, including viral infections, scientists are confidently exploring repurposed therapeutic molecules to further enhance the arsenal against COVID-19 (Ciliberto & Cardone, 2020; Jang et al., 2021; Ng et al., 2021).

Most antiviral therapies target viral proteins (Minenkova et al., 2022), but alternative approaches, such as targeting viral RNA and its secondary structures, can be explored to combat the threat of SARS-CoV-2 infection

(Sreeramulu et al., 2021; Zhao et al., 2021). RNA plays a significant role in virus replication and propagation, and targeting its secondary structures can potentially disrupt the virus's life cycle. This approach could be a significant breakthrough in the fight against SARS-CoV-2 and other viral infections.

G4 structures are known to play a crucial role in various viruses, including single-stranded RNA viruses. Promisingly, certain G4-targeting compounds have already demonstrated antiviral activity, pointing toward the potential use of G4 ligands as potent antiviral agents (Abiri et al., 2021; Lavezzo et al., 2018). Several potential G4-forming sequences have been identified within the genome of SARS-CoV-2. *In vitro* experiments have confirmed the formation of G4s by some of these sequences (Belmonte-Reche et al., 2021; Cui & Zhang, 2020; Ji et al., 2021; Zhao et al., 2021).

In this study, we used a fast and low-expensive screening process called ligand-based virtual screening (LBVS) to identify FDA-approved drugs that could potentially stabilize SARS-CoV-2 G4 RNAs. To experimentally validate the G4-binding properties of the identified drugs, three biologically relevant G-rich sequences of SARS-CoV-2 found in the coding sequence regions of nucleocapsid protein, non-structural protein 10, and non-structural protein 3 (*G4-1*, *G4-2*, and *G4-3*, respectively) were employed as potential targets (**Table 4.12**) (Belmonte-Reche et al., 2021; Zhao et al., 2021). Further *in silico* and *in vitro* assays identified three drugs as emerging SARS-CoV-2 RNA G4 binders and provided a plausible mechanism of action for these molecules at the molecular level.

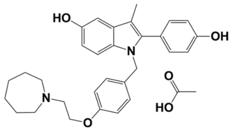
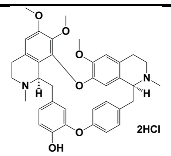
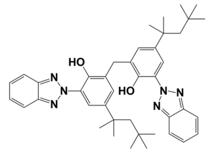
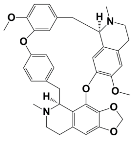
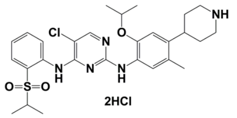
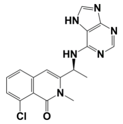
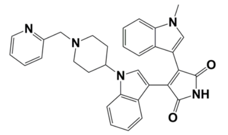
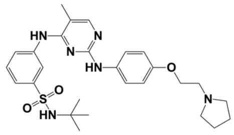
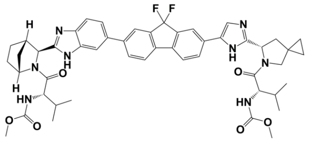
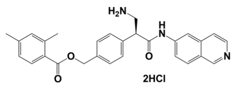
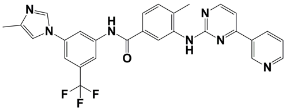
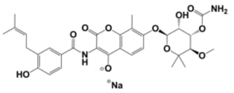
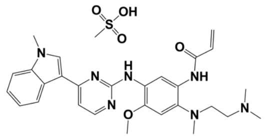
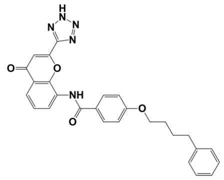
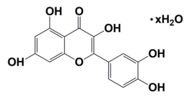
Table 4.12 List of here employed G4-forming RNA sequences in SARS-CoV-2.

Name	Sequence (5'-3')	Region
G4-1	GGCUGGCAAUGGCGG	Nucleocapsid protein
G4-2	GGUAUGUGGAAAGGUUAUGG	Non-structural protein 10
G4-3	UGGAGGAGGUGUUGCAGGA	Non-structural protein 3

4.4.2 RESULTS AND DISCUSSION

LIGAND-BASED VIRTUAL SCREENING (LBVS). Molecules that share similar structures exhibit comparable properties and functions. Thus, drugs matching the structural and geometrical features of typical G4 binders should have analogous G4-binding properties. Consequently, in the context of LBVS, a pharmacophore model may be developed to precisely determine the structural characteristics necessary to bind a target and elicit biological activity (Maruca et al., 2019). Starting from known active RNA G4 ligands from the literature (*training set*), 3D pharmacophore models were generated and validated before performing the drug repurposing LBVS. 3D ligand-based pharmacophores were built on the 2D Fingerprint clustered *training set* generating further 50 conformers to increase the chances that ligands matched the pharmacophores of the known G4 binders. A library of decoys was used to test each hypothesis, leading to five best-performing pharmacophore models. All five models exhibited some typical features of G4 binders: three aromatic rings and a hydrogen bond donor or a positively ionizable group but placed in different 3D spatial arrangements, allowing to retrieve of different RNA G4 binders from the *training set* on the basis of the specific model used (**Figure 4.33**). Around 3,000 FDA-approved drugs were screened using the selected pharmacophore models, and the best-matching drugs were selected based on the fitness score of the *training set* (**Table 4.13**). Notably, among the 15 putative G4 ligands identified, 5 have been already evaluated as potential anti-COVID-19 agents.

Table 4.13 Name of the selected drugs together with the corresponding 2D chemical structures.

Name	Chemical structure	Name	Chemical structure
Bazedoxifene		**Berbamine** *(L. Huang et al., 2021)	
Bisoctrizole		**Cepharanthine** *(Ohashi et al., 2021)	
Ceritinib		**Duvelisib**	
Enzastaurin		**Fedratinib**	
Ledipasvir *(Nourian et al., 2020)		**Netarsudil**	
Nilotinib		**Novobiocin**	
Osimertinib		**Pranlukast** *(Barré et al., 2020)	
Quercetin *(Derosa et al., 2021)			

* Already evaluated as promising anti COVID-19 agents and relative references

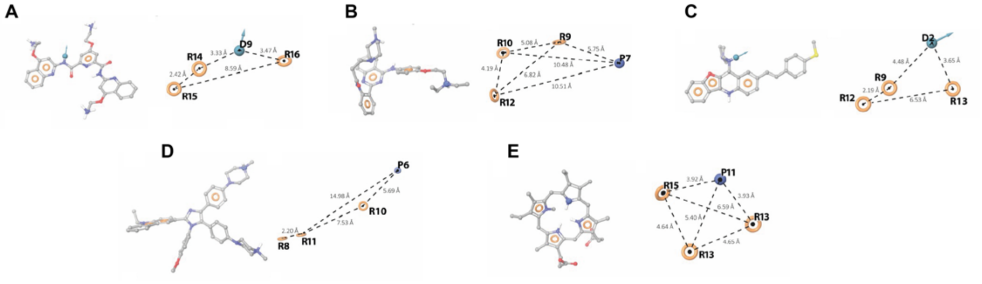


Figure 4.33 Ligand-based 3D pharmacophore models **A-E**) (*Right*) Hy1, Hy3, Hy5, Hy8, and Hy9 overlapping the reference ligands PDP, 19, 4a-10, IZCZ-3, and NMM-IX, respectively; (*Left*) pharmacophore hypotheses and distances between the pharmacophoric sites (dashed lines). Aromatic rings, hydrogen bond donor and positive ionizable groups are labelled as R, D, and P, respectively.

FLUORESCENCE INTERCALATOR DISPLACEMENT (FID) ASSAYS. The ability of the 15 selected drug candidates to bind to *G4-1*, *G4-2*, and *G4-3* was assessed by G4-FID assay, using TO as a light-up fluorescent probe and **PDS** as a positive control (Monchaud et al., 2006). Only a select few compounds, including **Netarsudil**, **Quercetin**, and **PDS**, were successful in displacing TO from all three RNA G4 structures. **Pranlukast** proved to be a strong TO competitor in the case of *G4-1*, while **Ledipasvir** and **Osimertinib** showed some level of TO displacement, but only from *G4-1* and *G4-3*, respectively. Based on DC_{50} values, the drugs were ranked as follows: *i*) drugs with low to null affinity for G4s ($DC_{50} > 10.0$ μM), *ii*) drugs with moderate affinity ($5.0 < DC_{50} < 10.0$ μM) (**Ledipasvir** for *G4-1*, **Osimertinib** and **Quercetin** for *G4-3*), and *iii*) good G4 binders ($DC_{50} < 5.0$ μM) (**Netarsudil**, **Pranlukast** for *G4-1*, **Quercetin** for *G4-1* and *G4-2*) (**Figure 4.34**, and **Table 4.14**).

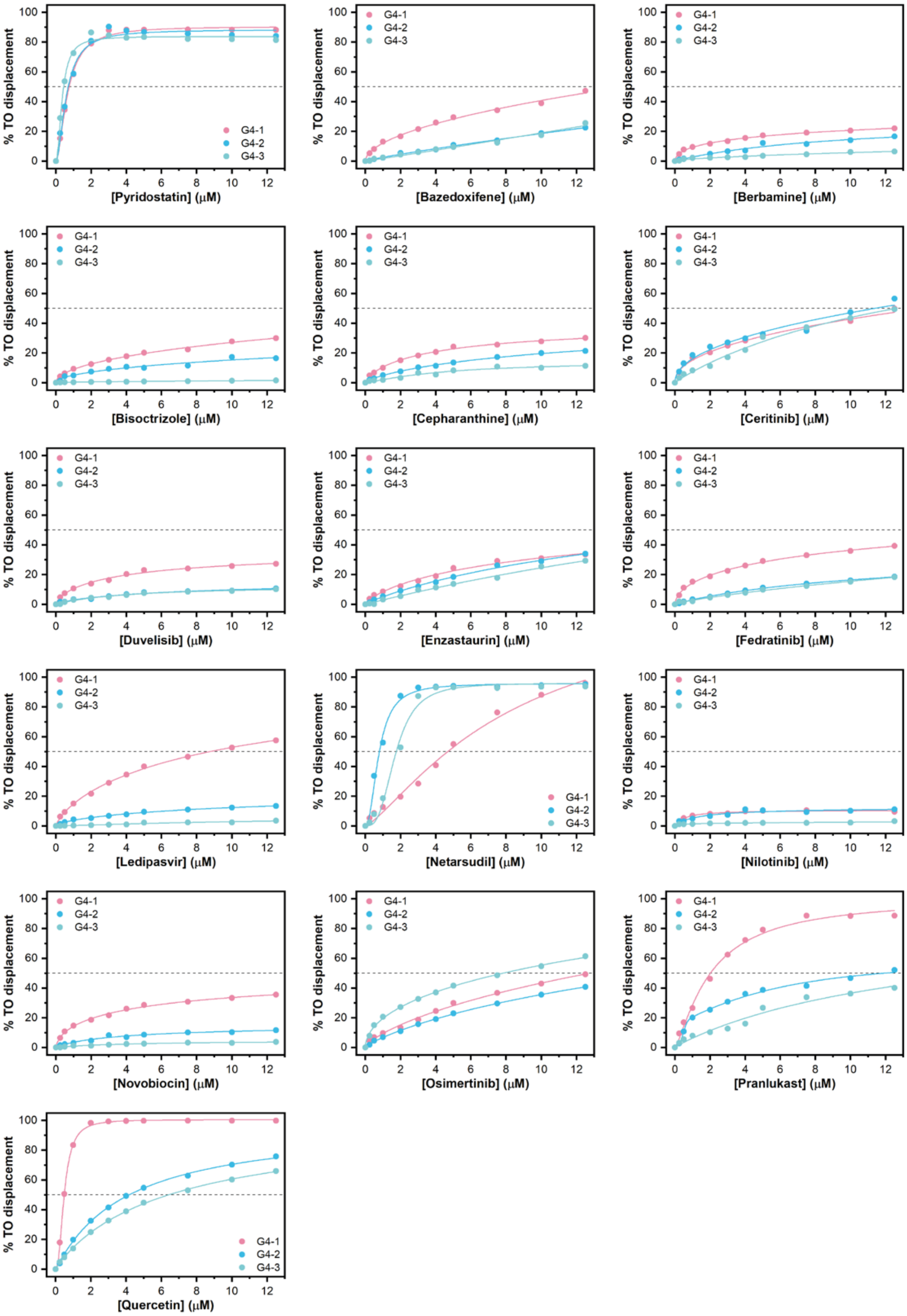


Figure 4.34 Dose-response curves from G4-FID assay of RNA G4s with selected drugs.

Table 4.14 Drugs DC$_{50}$ values for *G4-1*, *G4-2*, and *G4-3* determined by G4-FID assays.

	DC$_{50}$ (μM)^{a}		
Compound	***G4-1***	***G4-2***	***G4-3***
Pyridostatin	0.7	0.8	0.4
Ceritinib	12.5	11.4	12.5
Ledipasvir	8.8	n.d.^{b}	n.d.^{b}
Netarsudil	4.7	0.8	1.7
Osimertinib	n.d.^{b}	n.d.^{b}	7.8
Pranlukast	2.0	12.2	n.d.^{b}
Quercetin	0.5	4.2	6.4

[a] The error in DC$_{50}$ values is ±5%. Drugs with low to null affinity for all the three investigated G4s and DC$_{50}$ > 12.5 μM were not reported.
[b] n.d. = not determined, because DC50 > 12.5 μM.

CD EXPERIMENTS. CD spectra were recorded to verify the G4 structures adopted by *G4-1*, *G4-2*, and *G4-3*, which showed a maximum around 265 nm and a minimum around 240 nm in agreement with the formation of parallel-stranded G4 topology (**Figure 4.35**). Melting and annealing profiles of G4s were collected by following the changes in CD signal at the wavelength of maximum intensity. The profiles were recorded at 0.5 and 1.0 °C min^{-1}. The melting and annealing curves are identical for *G4-1*, which means that the unfolding and folding processes are at thermodynamic equilibrium, allowing an accurate determination of the T_{m} value. Conversely, hysteresis was observed for *G4-2* and *G4-3*, indicating that they are affected by the kinetics of the process, and suggesting the presence of intermolecular G4 species in solution or higher order structures through self-association between G4 units in solution or higher order structures through self-association between G4 units (**Figure 4.35**, and **Table 4.15**) (Belmonte-Reche et al., 2021). Therefore, subsequent studies focused only on *G4-1*.

Table 4.15 Melting (T_m) and annealing temperature (T_a) values for *G4-1*, *G4-2*, and *G4-3* determined by CD experiments.

	G4-1	*G4-2*	*G4-3*
T_m *	47.4 (±0.4)	42.8 (±0.4)	48.0 (±0.4)
T_a *	48.3 (±0.4)	33.6 (±0.4)	37.3 (±0.4)

* Data reported for 0.5 °C min^{-1} scan rate.

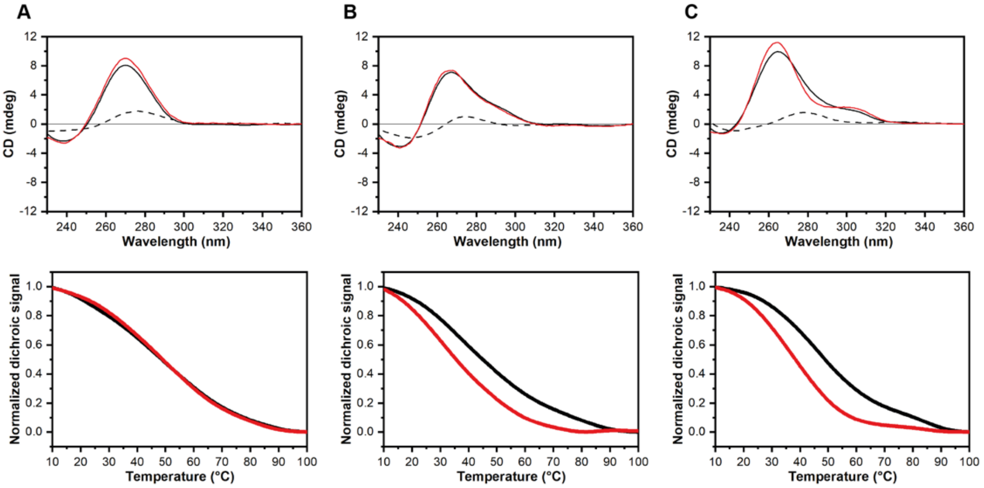


Figure 4.35 CD experiments for **A**) *G4-1*, **B**) *G4-2*, and **C**) *G4-3*. (*Top*) CD spectra at 10 before and after controlled annealing (black and red solid lines, respectively) and 100 °C (dashed lines); and (*Bottom*) normalized CD melting and annealing curves (0.5 °C/min scan rate) (black and red dotted lines, respectively).

The selected drugs significantly increased the thermal stability of *G4-1* (Fig. S4, ESI†). Interestingly, the best effects were found for drugs that showed higher affinity for *G4-1* in the G4-FID assay (**Netarsudil**, **Pranlukast**, and **Quercetin**). No changes in the CD spectrum of *G4-1* were observed upon interacting with **Ledipasvir**, **Osimertinib**, and **Quercetin**, indicating that it kept its parallel G4 structure. Conversely, **Netarsudil** and **Pranlukast** seem to alter the native conformation of *G4-1*. Actually, these two drugs exhibit CD signal in the wavelength region of G4 (**Figure 4.36**, and **Table 4.16**). Nevertheless, the experimental spectra of the two G4/drug mixtures differ significantly from the corresponding spectra resulting from the arithmetic sum of the single spectra, indicating once again their interactions.

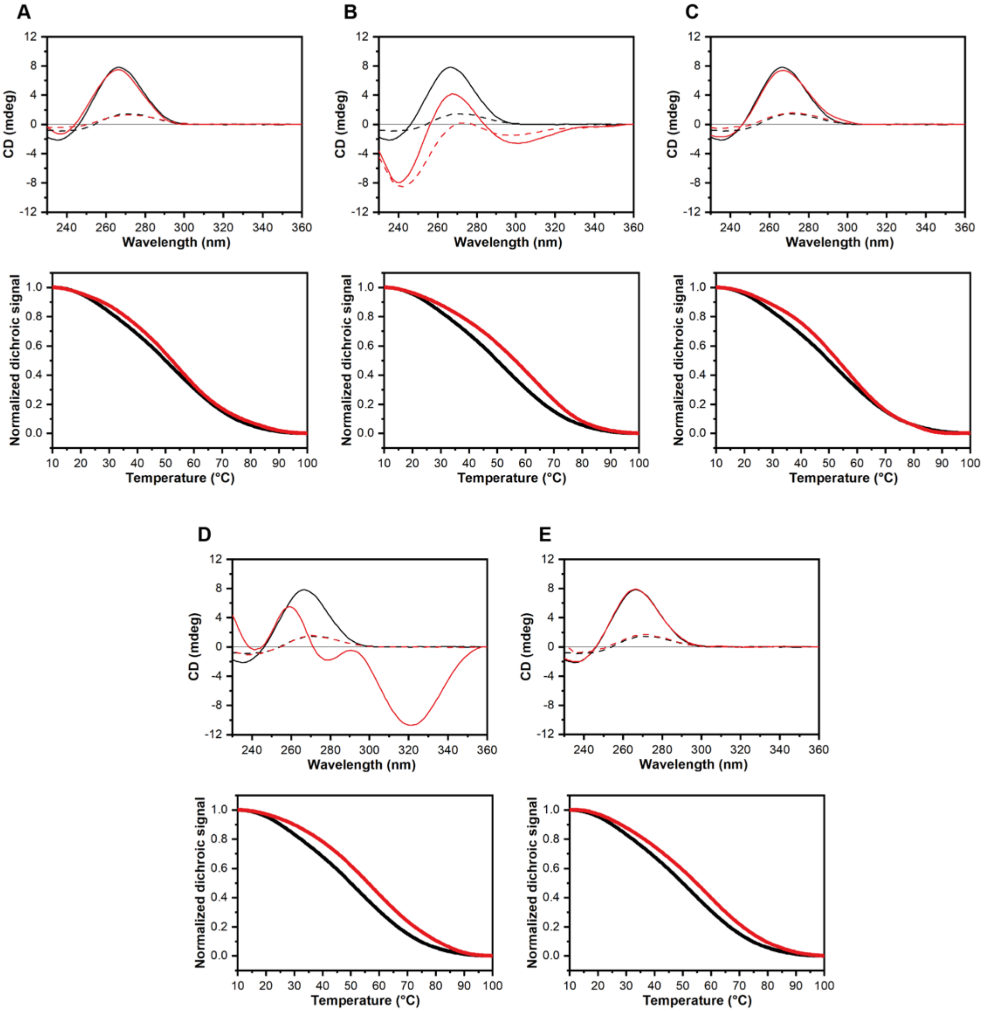


Figure 4.36 CD experiments for *G4-1* with **A**) **Ledipasvir**, **B**) **Netarsudil**, **C**) **Osimertinib**, **D**) **Pranlukast**, and **E**) **Quercetin**. (*Top*) CD spectra of *G4-1* in the absence (black line) and presence (red line) of 5 molar equiv of drug recorded at 10 and 100 °C (solid and dashed lines, respectively); and (*Bottom*) normalized CD melting curves (1.0 °C/min scan rate) in the absence (black dotted lines) and presence (red dotted lines) of drugs.

Table 4.16 Drug-induced thermal stabilization of *G4-1* measured by CD melting experiments.

Compound	**ΔT_m (°C)[a]** ***G4-1***
Ledipasvir	3.0 (±0.4)
Netarsudil	7.9 (±0.4)
Osimertinib	4.4 (±0.4)
Pranlukast	7.7 (±0.4)
Quercetin	5.6 (±0.4)

[a] The T_m value for *G4-1* at 1°C min^{-1} scan rate was 48.8 (±0.4) °C.

To get some information on the selectivity of the drugs, we evaluated their ability to stabilize the G4s derived from the 5′ untranslated region of *Bcl-2* (*Bcl2* G4) and the long non-coding RNA *GSEC* (*GSEC* G4). Results of CD melting experiments indicate that they do not significantly increase the thermal stability of *GSEC* G4 ($\Delta T_m \leq 2$ °C). On the other hand, **Netarsudil** has only a marginal stabilizing effect on *Bcl2* G4 ($\Delta T_m = 2.6$ °C), while **Pranlukast** and **Quercetin** exhibit a moderate stabilizing capacity (ΔT_m = 5.3 °C and 5.4 °C, respectively), although lower than that observed for *G4-1*.

FLUORESCENCE TITRATION EXPERIMENTS. Quantitative data on the affinity of **Netarsudil**, **Pranlukast**, and **Quercetin** for *G4-1* were obtained from fluorescence titration experiments. Fluorescence emission spectra of the drugs in the absence and presence of increasing amounts of G4 were recorded. Upon the addition of RNA, **Netarsudil** showed significant fluorescence quenching, whereas **Quercetin** showed enhancement. As for **Pranlukast**, which is intrinsically weakly fluorescent, no significant changes in fluorescence intensity were observed, making impossible to quantify its affinity to RNA. By plotting fluorescence changes against G4 concentration, binding isotherms were obtained for **Netarsudil** and **Quercetin** and the fitted curves gave K_D values of 0.8 (±0.2) and 13 (±5) µM, respectively (**Figure 4.37**).

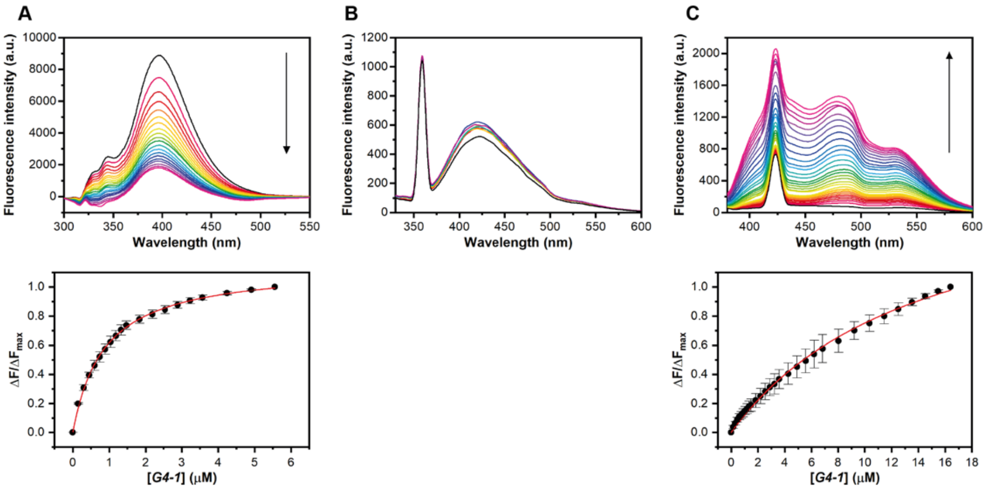


Figure 4.37 (*Top*) Fluorescence emission spectra of **A**) **Netarsudil**, **B**) **Pranlukast**, and **C**) **Quercetin**, in the absence and presence of increasing concentrations of *G4-1* at 20 °C. (*Bottom*) Binding isotherm plots to determine corresponding K_D values.

UV RESONANCE RAMAN (UVRR) EXPERIMENTS. The UVRR spectra at 266 nm of *G4-1*, drugs, and corresponding complexes were recorded to get insights into the binding mode (Di Fonzo et al., 2020). A change in the position or intensity of the bands in the spectrum indicates that an interaction is occurring and suggests the structural moieties involved (**Figure 4.38**)

As for **Netarsudil**, most of its bands overlap with those of *G4-1*, hindering a clear identification of the RNA regions that bind the drug. The only exceptions are the spectral variations associated to the adenine ring vibrations at 1334, 1480, 1510, and 1579 cm^{-1}. Compared to the sum, the experimental spectrum of the RNA/Netarsudil complex shows an intensity increase of these bands, indicating that adenine residues of loop are involved in drug binding. On the other hand, **Netarsudil**'s strong spectral contribution enabled us to evaluate the parts of drug involved in the interaction. The shift of the peaks at 1652 and 1414 cm^{-1}, attributed to the stretching of the two rings of the isoquinoline group and the corresponding C–C/C–N stretching modes, respectively, suggests that this region of the drug is mainly involved in RNA binding. **Pranlukast** and **Quercetin** showed negligible UVRR spectral contributions compared to *G4-1*, so binding-induced spectral perturbations could be easily correlated with the band changes of nucleotides. Upon RNA/**Quercetin** interaction, a decrease in intensity of the difference spectrum at the wavelengths of 1231, 1293, 1393, 1530 and 1625 cm^{-1} related to bands of U and C residues, suggests their involvement in the interaction. Further, an increase in intensity and redshift of the band at 1483 cm^{-1} which is mainly related to guanines indicate that these residues take part in the interaction (Di Fonzo et al., 2020). This finding, together with the downshift of the **Quercetin** peak (ring stretching vibration from 1620 to 1604 cm^{-1}), suggests that end-stacking is the favored binding mode for this drug. As for **Pranlukast**, the difference spectrum shows changes associated to A, G, and U residues, although of lower intensity than those observed for **Quercetin**. Even if it is a clear indication that **Pranlukast** binds to *G4-1*, this is not enough to obtain precise information on the molecular regions involved in the interaction.

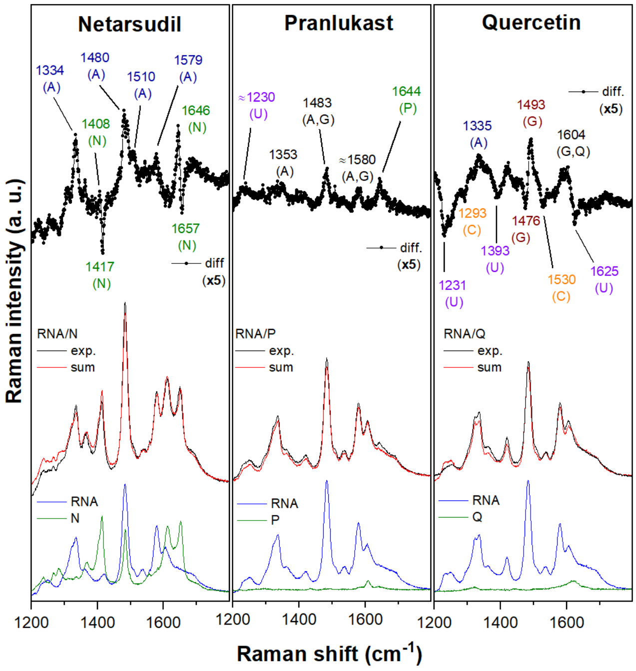


Figure 4.38 UVRR spectra to study the binding of **Netarsudil** (N), **Pranlukast** (P), and **Quercetin** (Q) to *G4-1* RNA. From bottom to top: RNA (blue); drug (green); RNA/drug complex (black); arithmetic sum of RNA and drug spectra (red); normalized difference between the spectra of the complex and the arithmetic sum (diff.). All the spectra have been normalized with the RNA spectrum intensity.

DOCKING AND MOLECULAR DYNAMICS SIMULATIONS (MDS). To elucidate the binding poses of the three drugs to *G4-1*, docking and molecular dynamics simulations (MDs) were performed. The best docking poses showed that, as suggested by experimental data, **Quercetin** and **Pranlukast** preferentially bind to the 5’-end G-tetrad making stacking interactions with G11 and G14 and H-bonds with U4 and C13. Conversely, Netarsudil was found to interact with the loop on the 3’-end side, also in good agreement with UVRR results. To further verify the docking results and the geometrical stability of *G4-1* upon binding, each complex was submitted to 500 ns of MDs. The most populated poses extracted after MDs are shown in **Figure 4.39**. Root Mean Square Deviation (RMSD) analysis revealed that **Quercetin** adopts a stable binding conformation, while **Pranlukast** slightly changed its binding more due to flexibility of the phenylbutoxy group. On the other hand, **Netarsudil** undergoes a deeper change of its binding mode during MDs, due to the `CAAU` loop, which fluctuates more than in the other complexes. Changes in the *G4-1* RNA geometry were also analyzed by calculating some basic G4 parameters (Tsvetkov et al., 2016). Looking at the trend of these parameters, no significant variations in the position and planarity of G-tetrads and in G4 structural integrity were observed, confirming no major structural rearrangements in *G4-1*.

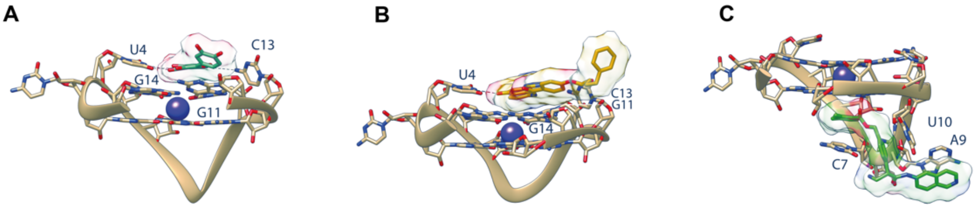


Figure 4.39 Most populated binding poses obtained after 500 ns of MDs for **A**) *G4-1*/**Quercetin** (97%), **B**) *G4-1*/**Pranlukast** (95%), and **C**) *G4-1*/**Netarsudil** (84%) complexes. **Quercetin**, **Pranlukast**, and **Netarsudil** are shown as cyan, yellow, and dark green sticks, respectively. K$^+$ ion is depicted as a violet sphere. Hydrogen atoms are not displayed for clarity reasons.

REVERSE TRANSCRIPTION INHIBITION ASSAY. To evaluate if **Netarsudil**, **Pranlukast**, and **Quercetin** could interfere with biological functions such as viral RNA replication, we investigated whether their binding to G4 affected reverse transcription. A synthetic SARS-CoV-2 genomic RNA template was reverse transcribed using specific primers flanking the *G4-1* or a non-G4-forming control region in the presence of growing concentrations of drugs. Quantitative RT-PCR was used to assay the levels of reverse transcription. The drugs significantly reduced *G4-1* reverse transcription compared to untreated samples when normalized with the control region, thus showing that they may be able to interfere with viral replication (**Figure 4.40**).

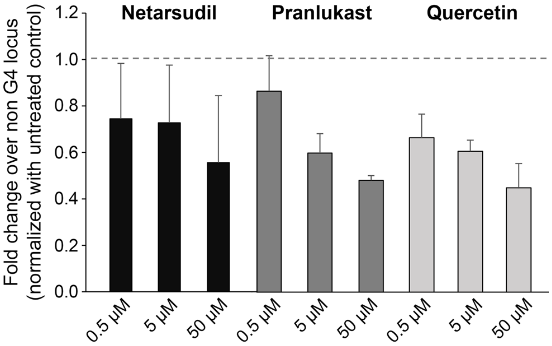


Figure 4.40 Fold change in reverse transcription of the *G4-1* region over a non-G4-forming control region from a synthetic SARS-CoV-2 RNA in the presence of the indicated concentrations of drugs. Untreated conditions were used as a control.

4.4.3 CONCLUSIONS

Three FDA approved drugs, **Netarsudil**, **Pranlukast** and **Quercetin** were identified as effective SARS-CoV-2 G4 binders through a ligand-based drug repurposing strategy. The results of this study lay the basis for further studies aiming to evaluate the antiviral activity of such drugs, while the methodological approach employed will certainly impact medicinal chemistry approaches for targeting of viral RNA G4s, even beyond SARS-CoV-2.

4.4.4 EXPERIMENTAL SECTION

OLIGONUCLEOTIDE SYNTHESIS AND SAMPLE PREPARATION. The investigated SARS-CoV-2 RNA sequences are reported in **Table 4.12**. The G4-forming RNA sequence from the lncRNA GSEC r(GGGGUGGAGGAGGGG-GAAGGGCGGGGG) (*GSEC* G4) and that from the 5′-UTR of *Bcl*-2 r(GGGCCGUGGGGUGGGAGCUGGG) (*Bcl-2* G4) were also tested. The oligonucleotides were chemically synthesized and purified as reported in **Paragraph 4.1.4**. All oligonucleotides were proven to be > 98% pure by NMR, then lyophilized and stored at -20 °C until use. *G4-1*, *G4-2*, and *G4-3* samples were prepared in 10 mM Tris-HCl buffer (pH 7.0) containing 100 mM KCl, while GSEC and *Bcl-2* G4s were prepared in 10 mM Tris-HCl buffer (pH 7.0) containing 1 mM KCl and 99 mM LiCl, and 20 mM KCl and 80 mM LiCl, respectively. Oligonucleotide concentration was verified by UV absorption at 90 °C using the appropriate molar extinction coefficient value ε (λ = 260 nm), calculated by the nearest-neighbor model (Cantor et al., 1970). Oligonucleotide solutions were then annealed by heating at 90 °C for 5 min followed by a slow cooling to r.t., and storage overnight at 4 °C.

TO G4-FID ASSAYS. G4-FID experiments were performed at 20 °C on a Jasco FP-8300 spectrofluorometer (Jasco, Easton, MD, USA) equipped with a Peltier cell holder (PCT-818) using an excitation wavelength of 485 nm and recording the emission in the 500-650 nm wavelength range at 100 nm min^{-1} scan speed. Both excitation and emission slits were set at 5 nm. Typically, a solution of 0.25 μM of pre-folded RNA G4 and 0.5 μM of TO in 10 mM Tris-HCl buffer (pH 7.0) containing 100 mM KCl was prepared, placed in a 1 cm path-length cell, and allowed to equilibrate for 15 min. After equilibration, fluorescence spectrum was acquired in the absence and presence of increasing concentrations of drugs (10 mM stock solution in pure DMSO). Each drug addition was followed by a 3 min equilibration time before spectrum acquisition. The percentage of TO displacement was calculated as described in Paragraph 4.1.4. The percentage of displacement was then plotted as a function of the drug concentration and, when possible, DC_{50} was calculated as the required concentration to displace 50% TO. Experiments were performed in duplicate. Data analysis was performed using OriginPro 2021 software (OriginLab Corp., MA, USA).

CD EXPERIMENTS. CD experiments were performed on a Jasco J-815 spectropolarimeter equipped with a Peltier temperature controller (PTC-423S/15). CD spectra were recorded at 10 and 100 °C, in the 230-360 nm wavelength range, with a scan rate of 100 nm/min, 0.5 s response time, and 1 nm bandwidth using 1 cm path-length cuvettes. Spectra were baseline subtracted and averaged over three scans. The RNA concentration used for each experiment was 2 μM, except for *GSEC* G4 that was used at 1 μM. Measurements were run in the absence or presence of 5 molar equiv of drugs. RNA/drug mixtures were allowed to equilibrate for 10 min before spectra acquisition. CD melting and annealing experiments were carried out in the 10-100 °C temperature range at 1.0 or 0.5 °C min^{-1} scan rate. CD experiments of RNA in the presence of selected drugs (5 molar equiv) were also performed by using the same parameters. CD melting curves were normalized between 0 and 1 to better compare the results and fitted to a Boltzmann distribution using OriginPro 2021 software to obtain the apparent melting temperatures (T_m). ΔT_m values were determined as the difference in the T_m values of the RNA in the presence and absence of drugs.

FLUORESCENCE TITRATIONS EXPERIMENTS. Fluorescence titration experiments were performed at 25 °C on a Jasco FP-8300 spectrofluorometer equipped with a Peltier temperature controller (PCT-818) by using a sealed quartz cuvette with 1 cm path-length. Titrations were carried out by stepwise additions (2-10 μL) of G4-1 (75 μM) to a solution of drug (at constant concentration of about 1 μM) in 10 mM Tris-HCl buffer (pH 7.0) containing 100 mM KCl. Excitation wavelengths were set at 290 and 370 nm for **Netarsudil** and **Quercetin**, respectively, and the corresponding emission spectra were recorded in the wavelength ranges of 300–550 and 380–600 nm. Both excitation and emission slit widths were set at 5 nm. After each RNA addition, the solutions were stirred and allowed to equilibrate for 3 min before spectrum acquisition. For **Netarsudil**, since an excitation wavelength of 290 nm was used, which falls into the absorption band of RNA, blank titration experiments were carried out by stepwise additions of RNA to the buffer solution (under the same experimental conditions used for drug solution). The obtained fluorescence emission spectra were then subtracted from the corresponding spectra of the **Netarsudil** titration, thus obtaining the binding isotherm to determine the K_D. K_D values were calculated from the plot of $\Delta F/\Delta F_{max}$ at 397 nm for **Netarsudil** and 535 nm for Quercetin versus concentration of RNA using the following equation:

$$\Delta\mathrm{F}/\Delta\mathrm{F}_{\mathrm{max}} = \frac{B_{\mathrm{max}}[\mathrm{RNA}]}{K_{\mathrm{D}} + [\mathrm{RNA}]}$$

where K_D is the dissociation constant, ΔF is the change in emission fluorescence after each addition of RNA, ΔF_{max} is the change in emission spectrum when the ligand is totally bound to DNA, and B_{max} is the maximum specific binding, in the same units as $\Delta F/\Delta F_{max}$ (Bhattacharjee et al., 2017). Experiments were performed in triplicate, and the K_D values reported are the average of three measurements ± SD.

Dataset generation, common pharmacophore hypothesis generation, pharmacophore virtual screening, docking simulations, and molecular dynamics simulations were performed in collaboration with Federica Moraca, Antonio Lupia, and Bruno Catalanotti from the Department of Pharmacy of University of Naples Federico II, Italy. UVRR experiments and corresponding data analysis were performed by Francesco D'Amico and Silvia di Fonzo from Sincrotrone Trieste. Reverse transcription inhibition assays were performed by Eleonora Vertecchi and Erica Salvati from the Institute of Molecular Biology and Pathology at the National Research Council in Rome, Italy. For further information, see APPENDIX.

4.5 UNVEILING THE INTERACTION BETWEEN DNA G-QUADRUPLEXES AND RG-RICH PEPTIDES (PAPER V)

4.5.1 INTRODUCTION

Protein-DNA interactions are crucial in cellular processes like transcriptional regulation, replication, and repair. Besides proteins that recognize specific DNA sequences (Garvie & Wolberger, 2001), some proteins bind to specific DNA structures, such as G4s (Meier-Stephenson, 2022). Under certain conditions, G-rich sequences have been observed to adopt G4 conformations as a result of specific stimuli, particular cell cycle phases, or in a cell type-specific manner. These conformations are often controlled by several proteins that bind and stabilize or resolve them (Linke et al., 2021; Mendoza et al., 2016).

Several studies suggest that human G4-binding proteins should be considered a specific group of nucleic acid-binding proteins since they differ from other proteins for the significant enrichment of glycine (G) and arginine (R) residues (Bartas et al., 2021). By comparing the amino acid sequences of 77 human G4-binding proteins available at the time of analysis, a common 20 amino acid-long RG-rich motif of sequence `RGRGRGRGGGSGGSGGRGRG`, named **NIQI** (**N**ovel **I**nteresting **Q**uadruplex **I**nteraction motif), has been identified (Brázda et al., 2018). Interestingly, the **NIQI** motif consists almost exclusively of R and G residues with a few serine (S) alternations. Structurally, **NIQI** is supposed to be inherently flexible to enable arginine and serine residues to interact, possibly by forming hydrogen bonds, with DNA in a wide range of G4 conformations. The interaction of RGG domains with G4s has sparked great interest, and the importance of elucidating RGG-mediated G4 recognition is underlined by the observation that this domain is indispensable for protein function (Z.-L. Huang et al., 2018; Masuzawa & Oyoshi, 2020;

Yan et al., 2021). Exploring the binding of **NIQI** to various G4s and its recognition is worth investigating, as it has never been thoroughly studied experimentally despite computational predictions.

Herein, we have characterized the interaction of the peptide with biologically relevant DNA G4s from oncogene promoters (*Bcl-2*, *c-Kit1*, *c-Kit2*, *c-Myc*, *Her2*, *HRAS1*, and *LWDLN1*) as well as telomeric sequence (*m-Tel*$_{24}$), characterized by parallel, antiparallel, or hybrid topologies by using a combination of biophysical techniques, including CD spectroscopy and MST. The interaction between **NIQI** and the most promising G4 targets has been further studied from both a thermodynamic and structural point of view through isothermal titration calorimetry (ITC) and NMR experiments. To further detail the interactions, key amino acids in the G4-binding peptide were determined by introducing point mutations. This study provides new insights into the interaction between RG-rich peptides and G4 structures, which can pave the way for the design and application of a novel class of peptide-based ligands.

4.5.2 RESULTS AND DISCUSSION

CD EXPERIMENTS. CD spectra in the wavelength range of 190-250 nm provide information on the presence of different types of secondary structures in peptides (Holzwarth & Doty, 1965). In agreement with the results obtained from peptide structure prediction using AlphaFold2 (McDonald et al., 2023) and PEP-FOLD3.5 (Rey et al., 2023) tools, the CD spectrum of **NIQI** showed very low ellipticity above 210 nm and a negative band near 200 nm, suggesting a random coil structure under experimental conditions (i.e., 5 mM KH_2PO_4/K_2HPO_4 aqueous buffer at pH 7.0, containing 20 mM KCl) (**Figure 4.41**, and **Table 4.17**). This agrees with the high glycine residue content of the peptide, which gives flexibility to the peptide structure.

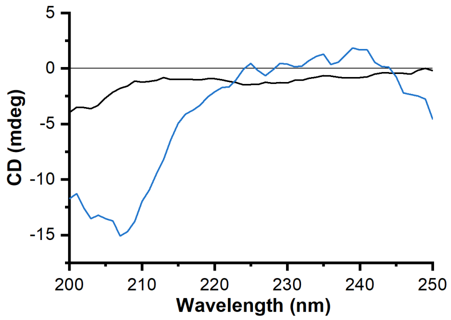


Figure 4.41 CD spectra of **NIQI** (6 µM) recorded at 20 °C in the absence (black line) and presence of *c-Kit2* G4 (blue line). To enable the detection of conformational changes in the peptide upon G4-binding, subtraction of the G4 signal was carried out.

Table 4.17 Secondary structure content estimation for **NIQI** derived from the CD spectra using BeStSel.

	Estimated secondary structure content (%)				
	Helix	**Antiparallel**	**Parallel**	**Turn**	**Others**
NIQI	0.0	37.2	0.0	17.1	45.7
NIQI + *c-Kit2 G4*	17.5	29.8	0.0	10.8	41.9

CD spectroscopy is also a valuable tool for analyzing the secondary structure of oligonucleotides, particularly in the range of 220-320 nm (Randazzo et al., 2013). The CD spectra of *c-Kit1*, *c-Kit2*, and *c-Myc* G4s exhibit the characteristic pattern associated with parallel-stranded G4s, whereas *HRAS1* and *LWDLN1* display the characteristic CD profile indicative of antiparallel G4 structures. Finally, *Bcl-2* and *m-Tel*$_{24}$ showed the typical profile associated with hybrid G4 structures.

Additional CD experiments were performed to verify the capability of **NIQI** to alter the native folding topology of these G4s. Interestingly, **NIQI** (3 mol equiv with respect to DNAs) influenced the CD spectra of the G4-forming sequences to different extents, with more pronounced effects observed in the case of the parallel G4s, causing a redshift of both the positive and negative bands located at 264 and 240 nm, respectively. In addition, a decrease in the intensity of the maximum and an increase in the depth of the minimum were observed in the case of *c-Kit1* and *c-Kit2* G4s, while a slight increase of the maximum was observed for *c-Myc* G4. As for *Bcl-2* hybrid G4, noticeable intensity variations of both the positive bands (at 265 and 290 nm) and the negative one (240 nm) occurred, which may indicate a structural rearrangement of the G4 upon peptide interaction. Conversely, only very small changes in the CD profile of *m-Tel24* hybrid G4 were observed upon peptide addition, as well as for the *HRAS1* and *LWDLN1* antiparallel G4 structures (**Figure 4.42**).

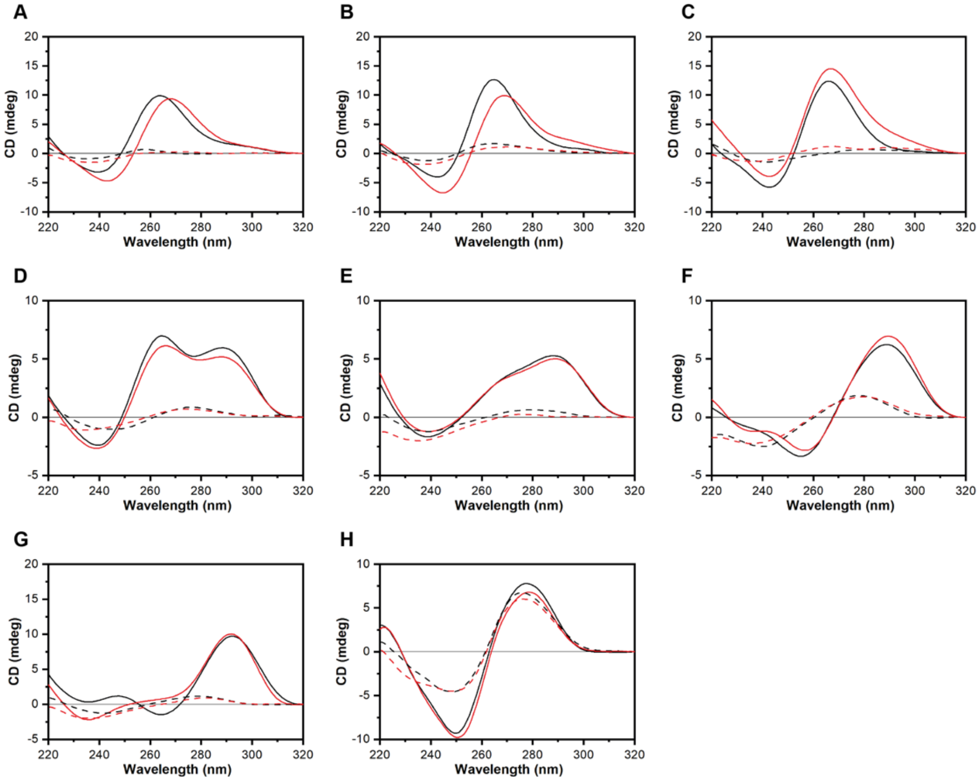


Figure 4.42 CD spectra of **A**) *c-Kit1*, **B**) *c-Kit2*, **C**) *c-Myc*, **D**) *Bcl-2*, **E**) *m-Tel24*, **F**) *HRAS1*, **G**) *LWDLN1* G4s, and **H**) *ds_{26}* duplex DNA in the absence and presence of **NIQI** (black and red lines, respectively) recorded at 20 and 100 °C (solid and dashed lines, respectively).

The stabilizing properties of **NIQI** were then evaluated by CD melting experiments measuring the peptide-induced changes in the melting temperature (ΔT_m) of the G4s (**Figure 4.43** and **Table 4.18**). Results of these experiments indicate that **NIQI** induces thermal stabilization most of the G4s, except for *HRAS1* and *LWDLN1* which adopt antiparallel conformations. However, **NIQI** does not exhibit inherent selectivity for a particular G4 topology. Indeed, the highest thermal stabilization effects were obtained for *c-Kit2* (parallel G4), followed by *Bcl-2* (hybrid G4), and c-*Myc* and *c-Kit1* G4s (both parallel G4s), as shown in **Table 4.18**. On the other hand, **NIQI** induced a slight destabilizing effect on *LWDLN1* G4, which might be attributed to a possible distortion of the DNA upon peptide interaction.

Additionally, with the aim of evaluating **NIQI** selectivity for G4 structures over double-stranded DNA, CD experiments were also performed by using a representative duplex model (ds_{26}). The CD spectrum of ds_{26} was not significantly altered after the addition of the peptide, indeed both the positive and negative bands (at around 280 nm and 250 nm, respectively) remained almost unchanged. Moreover, duplex thermal stability was not significantly affected by the peptide, indicating the preference of **NIQI** for G4 over duplex DNA.

Finally, to investigate the conformational changes of **NIQI** upon G4 binding, an analysis of the CD spectrum of the peptide in the presence of *c-Kit2*, that is the most stabilized G4, was performed in the 200-250 nm wavelength range. By subtracting the CD signal of the *c-Kit2* G4 from the spectrum of the G4/peptide mixture, it was possible to detect a conformational change of **NIQI**, with the appearance of an α-helical signal in the presence of the G4 (**Figure 4.41**, and **Table 4.17**). An increase in the helical content of **NIQI** from 0 to 17.5% was estimated in the presence of *c-Kit2* G4.

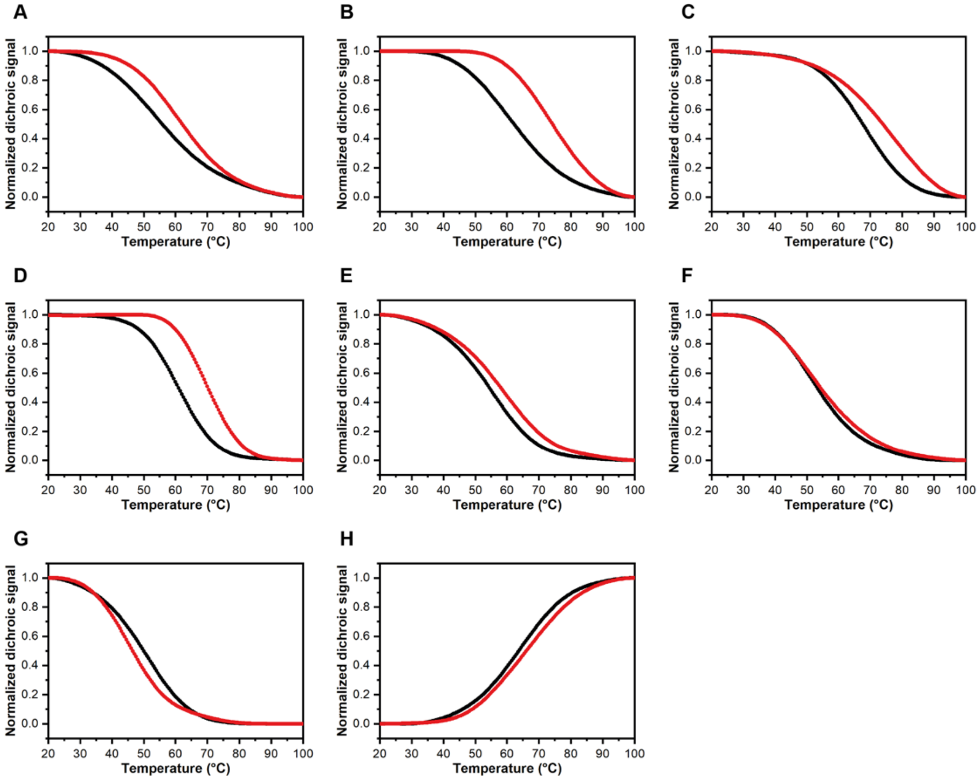


Figure 4.43 Normalized CD melting curves of **A**) *c-Kit1*, **B**) *c-Kit2*, **C**) *c-Myc*, **D**) *Bcl-2*, **E**) *m-Tel24*, **F**) *HRAS1*, **G**) *LWDLN1* G4s, and **H**) ds_{26} duplex DNA in the absence and presence of **NIQI** (black and red lines, respectively).

Table 4.18 NIQI-induced thermal stabilization of G4 and duplex DNAs measured by CD melting experiments.

DNA	Sequence (5'-3')	ΔT_m (°C)[a]
Parallel G4s		
c-*Kit1*	AGGGAGGGCGCTGGGAGGAGGG	6.3 (±0.2)
c-Kit2	CGGGCGGGCGCTAGGGAGGGT	12.6 (±0.2)
c-Myc	TGAGGGTGGGTAGGGTGGGTAA	7.9 (±0.4)
Antiparallel G4s		
HRAS1	TCGGGTTGCGGGCGCAGGGCACGGGCG	1.6 (±0.4)
LWDLN1	GGGTTTGGGTTTTGGGAGGG	-3.2 (±0.2)
Hybrid G4s		
Bcl-2	GGGCGCGGGAGGAATTGGGCGGG	10.5 (±0.2)
m-Tel$_{24}$	TTGGGTTAGGGTTAGGGTTAGGGA	3.7 (±0.4)
Duplex		
ds$_{26}$	CAATCGGATCGAATTCGATCCGATTG	2.0 (±0.4)

[a] ΔT_m represents the difference in melting temperature [$\Delta T_m = T_m$ (DNA + 3 peptide equiv) - T_m (DNA)]. The T_m values of DNA alone are: *c-Kit1* = 55.8 (±0.1) °C, *c-Kit2* = 62.1 (±0.1) °C, *c-Myc* = 68.0 (±0.2) °C, *HRAS1* = 53.1 (±0.2) °C, *LWDLN1* = 49.6 (±0.1) °C, *Bcl-2* = 60.5 (±0.1) °C, *m-Tel*$_{24}$ = 54.1 (±0.2) °C, *ds*$_{26}$ = 63.7 (±0.2) °C.

MST ASSAYS. To obtain quantitative data on the binding affinity of **NIQI** for the investigated G4 structures and to verify its selectivity for G4 over duplex DNA, MST was employed. For MST experiments, serial dilutions of DNA molecules were prepared and mixed with a constant concentration of FITC-labeled peptide, loaded into capillaries, and analyzed at 25 °C. The binding curves obtained from these experiments validated the binding of **NIQI** to all G4 structures, and notably, revealed a strong binding affinity of the peptide to certain G4s, with measured K_D values in the two-digit nanomolar range (**Figure 4.44**, and **Table 4.19**). In particular, **NIQI** exhibited excellent affinity for *c-Kit2*, followed by *Bcl-2*, *c-Myc*, and *c-Kit1* G4s, good affinity towards *m-Tel24* and *LWDLN1* G4s, and very low affinity for *HRAS1* G4. Also, **NIQI** displayed negligible binding to the *ds*$_{26}$ duplex, clearly indicating its selectivity for G4s over duplex DNA. These findings are in complete agreement with the results of CD melting experiments and identify *c-Kit2* and *Bcl-2* as the best G4 targets of this series for **NIQI**.

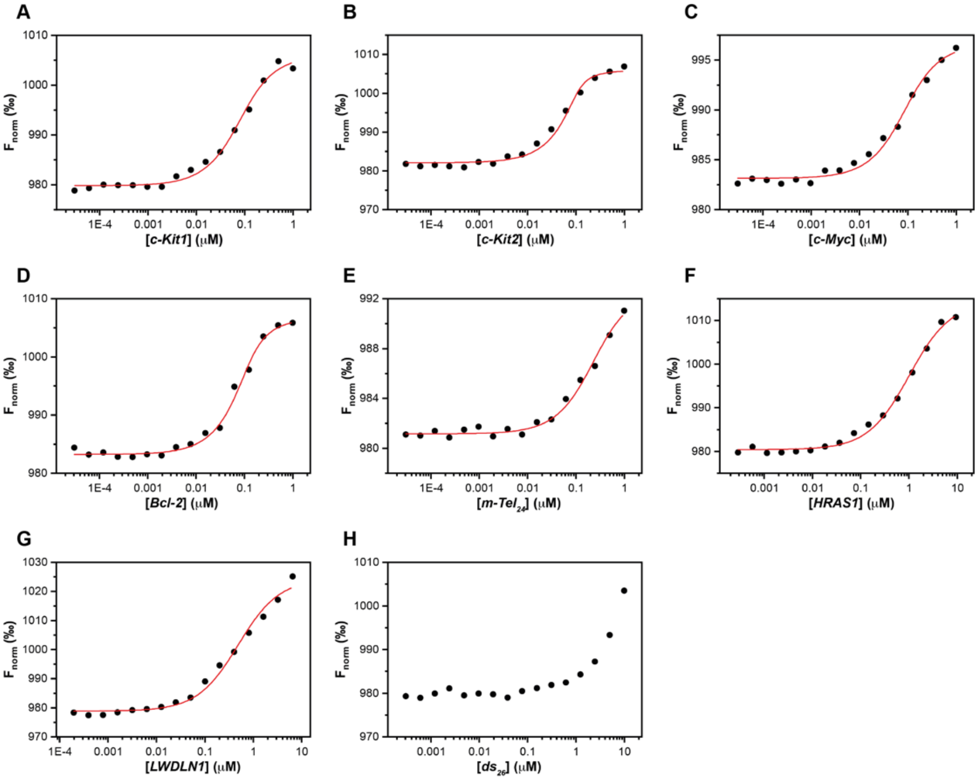


Figure 4.44 Dose-response curves of **A**) *c-Kit1*, **B**) *c-Kit2*, **C**) *c-Myc*, **D**) *Bcl-2*, **E**) *m-Tel24*, **F**) *HRAS1*, **G**) *LWDLN1* G4s, and **H**) *ds*$_{26}$ duplex DNA titrated against serial dilutions of fluorescein-conjugated **NIQI** (FITC-**NIQI**).

Table 4.19 Equilibrium K_D for the binding of **NIQI** to G4 and duplex DNAs obtained by MST experiments.

DNA	Sequence (5’-3’)	K_D (nM)
Parallel G4s		
c-*Kit1*	`AGGGAGGGCGCTGGGAGGAGGG`	63 (±11)
c-Kit2	`CGGGCGGGCGCTAGGGAGGGT`	10 (±6)
c-Myc	`TGAGGGTGGGTAGGGTGGGTAA`	60 (±10)
Antiparallel G4s		
HRAS1	`TCGGGTTGCGGGCGCAGGGCACGGGCG`	1,000 (±100)
LWDLN1	`GGGTTTGGGTTTTGGGAGGG`	460 (±60)
Hybrid G4s		
Bcl-2	`GGGCGCGGGAGGAATTGGGCGGG`	27 (±9)
m-Tel$_{24}$	`TTGGGTTAGGGTTAGGGTTAGGGA`	200 (±30)
Duplex		
ds$_{26}$	`CAATCGGATCGAATTCGATCCGATTG`	> 20,000

ISOTHERMAL TITRATION CALORIMETRY (ITC) EXPERIMENTS. ITC is a high accurate method to determine thermodynamic parameters of ligand binding to G4 DNA. It measures the enthalpic and entropic components of an interaction, revealing the thermodynamic driving forces in molecular recognition (Pagano et al., 2009). Therefore, to gain a deeper understanding of the binding of **NIQI** to *c-Kit2* and *Bcl-2* G4s, a thorough ITC calorimetric analysis was performed. Interestingly, both titrations result in binding isotherms revealing two-event binding processes, but with different features from each other.

Regarding *c-Kit2* G4, the best fit to the experimental data was obtained with a two-event model, where the first binding event was accompanied by a more exothermic enthalpy change and estimated to be >200-fold stronger than the second one (Le et al., 2013; Pirota et al., 2021). The overall stoichiometry for the binding of **NIQI** to *c-Kit2* was determined to be three peptides per G4 DNA, with one peptide involved in the first event and two peptides in the second one. The thermodynamic signatures of the two binding events indicate that the interaction of **NIQI** is enthalpically driven. Indeed, both binding events were characterized by unfavorable entropic contributions ($\Delta H° < 0$), indicating that the driving force behind the binding process is the formation of new interactions between the ligand and the DNA. On the other hand, the opposing entropic contribution suggests that the resulting complex is more rigid compared to the free molecules (**Figure 4.45**, and **Table 4.20**).

Regarding the thermodynamic characterization for the binding of **NIQI** to *Bcl-2* G4, it reveals again a two-event binding process but with features different from that observed for *c-Kit2* G4 (**Figure 4.45**). In this case, the binding isotherm has the typical trend of a system wherein the higher binding affinity site demonstrates a smaller exothermic enthalpy change than the lower binding affinity site, as confirmed by the $\Delta H°$ values (**Table 4.20**) (Le et al., 2013). The analysis reveals that the first event is about 40-fold stronger than the second one. The stoichiometry indicates that one G4 equiv can bind up to four equiv of **NIQI**. The thermodynamic signature of the two binding events suggests that the interaction is both enthalpically and entropically driven (**Table 4.20**). Indeed, the favorable enthalpic contribution ($\Delta H° < 0$), despite smaller than that observed for *c-Kit2*, acts synergistically with the entropic one ($T\Delta S° > 0$), which indicates that the driving force of the binding process is only partially represented by the formation of peptide-DNA interactions, but it may be also due to a relaxation of the G4 structure upon binding and/or to the displacement of water molecules. The hypothesis of peptide-induced changes in the structural features of *Bcl-2* G4 is also supported by the CD results. Overall, ITC data suggest that **NIQI** binds to G4 targets through distinct mechanisms driven by different thermodynamic contributions.

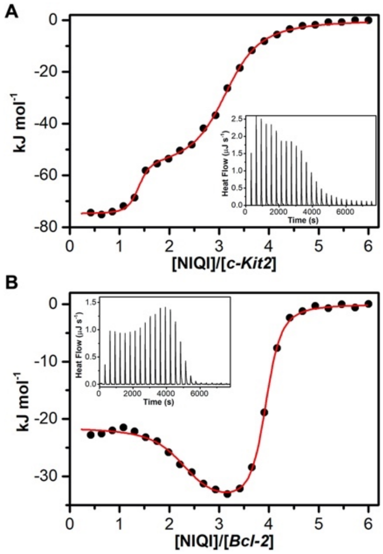


Figure 4.45 Raw ITC data (insets) and binding isotherms for titration of **A)** *c-Kit2* and **B)** *Bcl-2* G4s with **NIQI** obtained at 25 °C. The black dots represent the experimental data obtained by integrating the raw ITC data and subtracting the heat of peptide dilution into the buffer. The red lines represent the best fit obtained by the multiple binding sites model.

Table 4.20 Thermodynamic parameters for the interaction of **NIQI** with *c-Kit2* and *Bcl-2* G4s obtained by ITC at 25 °C.

DNA	No. of events	*n*	K_B (M^{-1})	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$T\Delta S°$ (kJ mol^{-1})
c-Kit2 G4	2	1	1.7×10^8	-47.0	-74.7	-27.7
		2	6.9×10^5	-33.3	-56.9	-23.6
Bcl-2 G4	2	2	1.1×10^8	-45.9	-14.8	31.1
		2	2.8×10^6	-36.8	-25.9	10.9

1D ^{1}H NMR ANALYSIS. To gain further information on the binding mode of **NIQI** to *c-Kit2* and *Bcl-2* G4s, 1D ^{1}H NMR titration experiments were performed. According to the literature, under the experimental conditions used, *c-Kit2* and *Bcl-2* sequences form a single G4 conformation characterized by 12 well-resolved imino proton peaks, corresponding to the 12 guanines involved in the three G-tetrad planes (J. Dai et al., 2006; Kuryavyi et al., 2010). Upon addition of increasing amounts of **NIQI**, a decrease in signal intensity was observed in both the imino and aromatic proton regions of the spectra of the two G4s, indicating a strong binding of the peptide to the tested G4 structures (**Figure 4.46**). The drop in signals turned out to be more intense in the case of *c-Kit2* for which the titration was virtually completed at the 3:1 peptide/G4 ratio, while for *Bcl-2* some signals are still visible in the spectrum at the 4:1 peptide/G4 ratio. Interestingly, the differences observed for the imino proton resonances of guanines clearly suggest a different peptide binding behavior to the stem of *c-Kit2* and *Bcl-2* G4s.

In the case of *c-Kit2 G4*, the imino proton signals most affected upon the addition of **NIQI** were those belonging to the external G-tetrads, especially those of the 5′ end G-tetrad, suggesting the formation of some stacking interactions. These are probably based on the large dipole moment associated with the amide group of the peptide and the prospect of dispersion involving amide π-electrons (James et al., 2009), allowing the peptide to arrange in a conformation that leads the arginine residues to approach and interact with specific phosphate groups of the G4 through electrostatic interactions, while the serine residues may form H-bonds with them. The formation of stacking interactions could explain the more favorable enthalpy changes observed by ITC. Conversely, in the case of *Bcl-2* G4, all guanines, including those of the central G-tetrad, displayed similar changes in their imino proton signals upon the addition of **NIQI**, suggesting a possible peptide interaction in the grooves of the G4. This kind of interaction could result in the displacement of a larger number of water molecules compared to stacking on G-tetrads, possibly explaining the favorable entropic contribution observed by ITC.

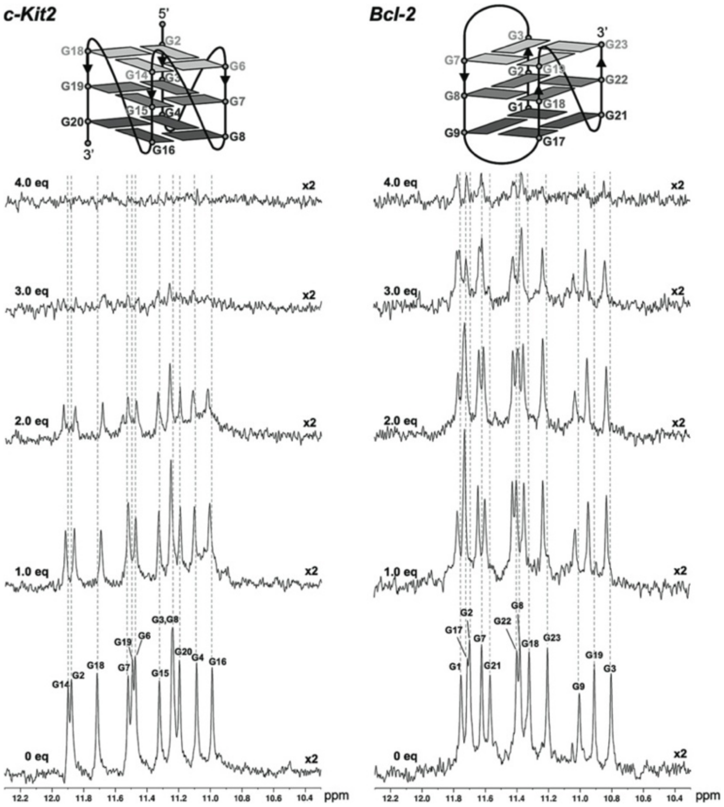


Figure 4.46 Imino proton region of NMR spectra of *c-Kit2* and *Bcl-2* G4s at 25 °C titrated with increasing mol equiv of **NIQI**.

MUTAGENESIS ANALYSIS. To further detail the interactions and determine the key amino acids involved in the binding of **NIQI** to *c-Kit2* and *Bcl-2* G4s, a mutagenesis analysis, consisting of replacing one at a time the arginine and serine residues of the **NIQI** sequence with the non-bulky and chemically inert alanine (A) residue, was performed. Glycine was not mutated since it has no sidechain and acts only as a flexible spacer.

CD melting experiments were performed and, in the case of *c-Kit2* G4, the results showed that: *i*) the arginine 3 and 7 have limited involvement in the interaction (induced ΔT_m of the mutants was comparable with the one induced by **NIQI**); *ii*) the arginine 1, 5, 17, and 19, as well as the serine 11 and 14, had greater involvement in the interaction with the G4 (induced ΔT_m of the mutants was at least 5 °C lower than the one induced by **NIQI**) (**Figure 4.47**, and **Table 4.21**). As for *Bcl-2*, a significant decrease in the peptide-induced G4 stabilization was observed upon substitution of arginine 3 and 7, as well as serine 14, suggesting a key role of these amino acids in the interaction. In contrast, negligible or small differences were observed when arginine 1 and 17 and serine 11, or arginine 5 and 19, respectively, were mutated (**Figure 4.47**, and **Table 4.21**).

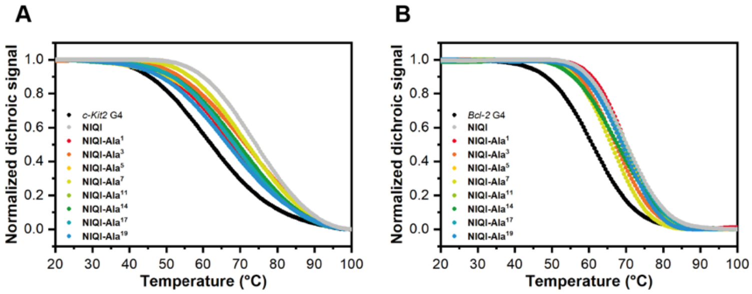


Figure 4.47 Normalized CD melting curves of **A**) *c-Kit2*, and **B**) *Bcl-2* G4s in the absence (black lines) and presence (color-coded lines) of 3 molar equiv of Ala-mutants **NIQI** peptides in comparison with **NIQI** (grey lines).

Table 4.21 **NIQI** Ala-mutants induced thermal stabilization of *c-Kit2* and *Bcl-2* G4s measured by CD melting experiments.

		ΔT_m (°C)[a]	
Peptide	**Sequence**	*c-Kit2 G4*	*Bcl-2 G4*
NIQI	RGRGRGRGGGSGGSGGRGRG	12.6 (±0.2)	10.5 (±0.2)
NIQI-Ala^1	**A**GRGRGRGGGSGGSGGRGRG	6.2 (±0.3)	9.5 (±0.2)
NIQI-Ala^3	RG**A**GRGRGGGSGGSGGRGRG	10.1 (±0.3)	6.1 (±0.3)
NIQI-Ala^5	RGRG**A**GRGGGSGGSGGRGRG	7.6 (±0.4)	8.3 (±0.4)
NIQI-Ala^7	RGRGRG**A**GGGSGGSGGRGRG	10.3 (±0.3)	5.1 (±0.2)
NIQI-Ala^{11}	RGRGRGRGGG**A**GGSGGRGRG	6.9 (±0.3)	9.5 (±0.2)
NIQI-Ala^{14}	RGRGRGRGGGSGG**A**GGRGRG	7.9 (±0.5)	7.0 (±0.2)
NIQI-Ala^{17}	RGRGRGRGGGSGGSGG**A**GRG	6.9 (±0.3)	9.2 (±0.4)
NIQI-Ala^{19}	RGRGRGRGGGSGGSGGRG**A**G	6.0 (±0.4)	8.4 (±0.4)

[a] ΔT_m represents the difference in melting temperature [$\Delta T_m = T_m$ (DNA + 3 peptide equiv) - T_m (DNA)]. The T_m values of DNA alone are: *c-Kit2* = 62.1 (±0.1) °C, and *Bcl-2* = 60.5 (±0.1) °C.

Overall, the alanine scanning analysis showed that the key amino acids of **NIQI** involved in the interaction with the two G4s are not identical. The involvement of fewer amino acids in the case of *Bcl-2* compared to *c-Kit2* also agrees with ITC data, which indicates a lower binding enthalpy for the interaction with the former than the latter G4.

4.5.3 CONCLUSIONS

The interaction between **NIQI**, a peptide containing an RGG motif shared by 77 human G4-binding proteins, and various biologically relevant G4 DNA structures with different topologies was investigated. **NIQI** induces some CD spectral changes and thermal stabilization in all parallel G4s and the hybrid *Bcl-2* G4. The peptide also exhibited good selectivity over duplex DNA. Among the investigated G4s, *c-Kit2* and *Bcl-2* showed the highest peptide-induced thermal stabilization values and gave the highest affinity binding constants determined by MST experiments, thus they were selected as the **NIQI** best targets. Results of ITC and NMR experiments clearly indicate a different binding mode of the peptide to the two G4s. As for *c-Kit2* G4, **NIQI** mainly affects the external G-tetrads, particularly the 5′ end G-tetrad, thus suggesting the formation of stacking interactions. **NIQI** itself, upon binding to *c-Kit2*, undergoes a transition from a random coil structure to a more helicized structure. On the other hand, in the case of *Bcl-2* G4, all G-tetrads are similarly affected by **NIQI**, suggesting a probable interaction of the peptide in the grooves of the G4 structure (**Figure 4.48**). Finally, by adopting the alanine scanning approach, the key and less relevant amino acids in the interaction with *c-Kit2* and *Bcl-2* G4s have been identified.

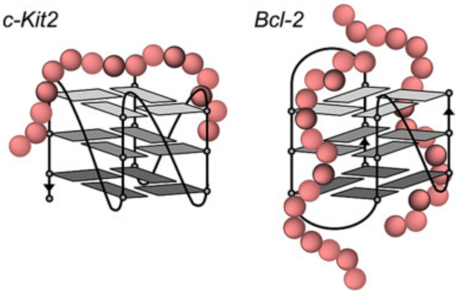


Figure 4.48 Schematic illustrations of possible binding modes for the strongest interactions of the **NIQI** peptide (in red) to *c-Kit2* and *Bcl-2* G4s.

The results obtained will undoubtedly help the development of other peptide-based G4 ligands with improved affinity and selectivity for G4 structures, as an alternative to small molecules. Indeed, due to their larger size and bulkiness compared to small molecules, we strongly believe that peptides could serve as more effective tools to interfere in G4-protein interactions and potentially exert significant effects in modulating oncogene expression.

4.5.4 EXPERIMENTAL SECTION

OLIGONUCLEOTIDE SYNTHESIS AND SAMPLE PREPARATION. DNA sequences listed in **Table 4.18** were synthesized on an ABI 394 DNA/RNA synthesizer (Applied Biosystem) at 5-μmol scale by using standard β-cyanoethyl phosphoramidite solid phase chemistry. DNA detachment from support and its deprotection were carried out by means of an aqueous solution of concentrated ammonia at 55 °C for 17 h. The filtrates and the washings were combined and concentrated under reduced pressure, solubilized in water, and then purified by high-performance liquid chromatography (HPLC) equipped with a Nucleo-gel SAX column (Macherey-Nagel, 1000–8/46) using a 30 min linear gradient going from 100% buffer A (20 mM KH_2PO_4/K_2HPO_4 aqueous solution (pH 7.0), containing 20 % (v/v) ACN) to 100% buffer B (1 M KCl,

KH_2PO_4/K_2HPO_4 aqueous solution (pH 7.0), containing 20 % (v/v) ACN). The purified fractions of the oligomers were then desalted by using C-18 cartridges (Sep-pak). The purity of the isolated oligomer was evaluated by NMR, and it turned out to be higher than 98 %. Next, oligonucleotides were lyophilized and then resuspended in 5 mM KH_2PO_4/K_2HPO_4 aqueous solution (pH 7.0), containing 20 mM KCl (or LiCl in the case of *c-Myc*) and their concentration was established by measuring the UV absorption at 90 °C considering the molar extinction coefficient values ε ($\lambda = 260$ nm) determined by the nearest neighbor model (Cantor et al., 1970). DNA samples were heated at 90 °C for 5 min, then gradually cooled to room temperature overnight, and finally stored at 4 °C for 24 h before use.

CD EXPERIMENTS. CD experiments were carried out on a Jasco J-815 spectropolarimeter equipped with a PTC- 423S/15 Peltier temperature controller. G4-forming oligonucleotides, as well as the duplex-forming DNA (ds26), were prepared at a concentration of 2 μM under conditions described above. CD spectra of DNA/peptide mixtures were obtained by adding 3 mol equiv of **NIQI** (stock solutions of peptide were 10 mM in H_2O) with respect to the oligonucleotide. Spectra of DNA molecules in the absence and presence of **NIQI** were recorded at 20 and 100 °C in the wavelength range of 220-320 nm using a scan rate of 100 nm min^{-1}, with a 1 s response time and 1 nm bandwidth. CD spectra of **NIQI** in the absence and presence of *c-Kit2* G4 were also recorded at 20 °C in the 200-250 nm wavelength range using a scan rate of 20 nm min^{-1}, with a 4 s response time and 1 nm bandwidth. All spectra were baseline subtracted and averaged over three scans. Subtraction of *c-Kit2* G4 spectrum from that of G4/peptide mixture was performed to have information about the spectrum of **NIQI** in the presence of G4. The percentage of secondary structures adopted by the peptide in the absence and presence of the G4 was estimated by using the BeStSel software (Micsonai et al., 2015).

CD melting experiments were carried out in the 20-100 °C temperature range at 1 °C min^{-1} heating rate by following CD signal changes at the wavelengths of the maximum CD intensity (264 nm for *c-Kit1*, *c-Kit2*, *c-Myc*, and *Bcl-2*; 290 nm for *m-Tel*$_{24}$, *HRAS1*, and *LWLDN1*; 252 nm for *ds*$_{26}$). CD melting experiments were performed in the absence and presence of the peptide (3 mol equiv) added to the folded DNA structures. The melting temperatures (T_m) were determined from a curve fit using Origin 7.0 software, and the ΔT_m values calculated as the difference in the T_m values of the DNA structures in the presence and absence of the peptide. All experiments were performed in triplicate, and the reported values are the average of the three measurements.

MST EXPERIMENTS. MST measurements were performed on a Monolith NT.115 (Nano-temper Technologies, Munich, Germany). Experiments were performed at 25 °C, using auto-tune LED power and medium MST power. The fluorescein isothiocyanate labelled peptide (FITC-**NIQI**) was prepared at 1 μM in H_2O, then diluted to 100 nM with the 5 mM KH_2PO_4/K_2HPO_4 buffer (pH 7.0), containing 20 mM KCl and supplemented with 0.1 % Tween and 5 % DMSO. Oligonucleotides were prepared at 1 mM concentration in 5 mM KH_2PO_4/K_2HPO_4 buffer (pH 7.0), containing 20 mM KCl, and annealed as described above. For the MST experiments, the final concentration of the labeled peptide was kept constant at 50 nM, while a serial dilution of the DNAs (1:2 from 2 or 40 mM) in the buffer supplemented with 0.1 % Tween and 5 % DMSO, was prepared and mixed with the peptide solution in a 1:1 volume ratio. All the samples were loaded into standard capillaries (Nano-Temper Technologies). MST data analysis was performed by employing the MO.Affinity Analysis software (v2.3) provided with the instrument.

ITC EXPERIMENTS. ITC measurements were performed at 25 °C using a nano-ITC Low Volume calorimeter (TA instruments, Lindon, UT, USA).

DNA and peptide solutions were all prepared with the same batch of buffer (5 mM KH_2PO_4/K_2HPO_4 buffer, pH 7.0, containing 20 mM KCl), to minimize the differences in buffer composition and pH. In each titration, volumes of 2 μL of peptide solution (200-400 μM) were injected, using a computer-controlled 50 μL syringe, into the calorimetric vessel (170 μL) containing the oligonucleotide (16-20 μM) with a spacing of 300 s between each injection to allow the system to reach the equilibrium. Heat produced by peptide dilution was evaluated in a control experiment by injecting the peptide solution into the buffer alone. The interaction heat for each injection was calculated after correction for the heat of peptide dilution. The corrected heat values were plotted as a function of the molar ratio, to give the corresponding binding isotherms, which were fitted with a multiple sites binding model by means of the NanoAnalyze software (TA instruments) supplied with the instrument, to give the binding enthalpy ($\Delta H°$), equilibrium binding constant (K_B), and stoichiometry of interaction (n). The Gibbs free-energy change and the entropy change were derived using the following relationships:

$$\Delta G° = -\mathrm{RT} \cdot \ln K_\mathrm{B}\,, \qquad T\Delta S° = \Delta H° - \Delta G°$$

All measurements were performed in triplicate.

NMR EXPERIMENTS. 1D ^1H NMR experiments were carried out on a Bruker AVANCE NEO NMR spectrometer, operating at 600 MHz (^1H Larmor frequency), equipped with a SampleJet autosampler and a 5 mm QCI H-P/C/N-D-5-Z CryoProbe, optimized for ^1H sensitivity. Experiments were carried out using 20 μM of *c-Kit2* or *Bcl-2* G4s in 5 mM KH_2PO_4/K_2HPO_4 buffer, pH 7.0, containing 20 mM KCl and 10 % D_2O, in the absence and presence of different amounts of **NIQI**. Spectra were recorded at 298 K, using 512 scans per spectrum with a recovery delay of 1.5 s, and suppressing water signal using the excitation sculpting with gradients (Hwang & Shaka, 1995). Spectra were phase-adjusted, baseline-corrected, and calibrated with respect to the water frequency. Processing and analysis were performed using the software package Bruker TOPSPIN 4.0.7.

Peptides synthesis was performed in collaboration with Nicola Grasso, Francesco Merlino and Paolo Grieco from the Department of Pharmacy of University of Naples Federico II, Italy. For further information, see APPENDIX.

CHAPTER 5
STUDY OF I-MOTIF STRUCTURES

Despite the importance of understanding the factors involved in stabilizing iM structures, there is still a lack of systematic characterization of favorable iM DNA folding conditions. In this CHAPTER, CD spectroscopy and DSC analysis were employed to characterize the thermodynamic stability of an iM-forming sequence in the presence of two different cations (K^+, and Na^+), at three different pH values (4.5, 5.0, and 5.5) (PAPER I).

5.1 ON THE THERMODYNAMICS OF FOLDING OF AN I-MOTIF IN SOLUTION UNDER FAVORABLE CONDITIONS (PAPER I)

5.1.1 INTRODUCTION

Various factors, such as an oligonucleotide and C-tract length, a loop structure, pH, salt concentration, and temperature, influence the formation of iM structures. A comprehensive understanding of iM behavior in different conditions is essential to develop nucleic acid nanodevices. Different iMs have been well-studied through various experimental techniques. However, there haven't been many systematic studies using differential scanning calorimetry (DSC) to analyze iMs under different solution conditions.

Here, a combination of CD spectroscopy and DSC have been employed to characterize the energetics of folding of a fragment of vertebrate telomere DNA, a 22-mer iM-forming sequence, `d(CCCTAA)`$_3$`CCCT`, under favorable conditions. At acidic pH, this C-rich sequence was shown to form an intramolecular structure that includes six C:C$^+$ base pairs (Phan et al., 2000). Its 21-mer truncated form, `d(CCCTAA)`$_3$`CCCT`, exists in two energetically close conformations, 5’E and 3’E, in equilibrium differing in the intercalation topology of the C:C$^+$ base pairs (Lieblein et al., 2012).

Transition pH (pH$_T$) indicates the pH level at which half of the oligodeoxynucleotide is folded. For C-rich telomeric oligonucleotides, pH$_T$ values have been observed between 5.9 and 6.4 across temperatures of 10-30 °C (McKim et al., 2016; Nguyen et al., 2017; S. Wu et al., 2013). The thermodynamics, including the evaluation of the energetic contributions of cytosine protonation, of this iM-forming sequence in solution at three different pH values below the pH$_T$ (4.5, 5.0, and 5.5) and in the presence of two different cations (Na$^+$ or K$^+$) has been investigated.

5.1.2 RESULTS AND DISCUSSION

To study the iM thermodynamic stability, the biophysical investigation through CD spectroscopy, DSC, and non-denaturing polyacrylamide gel electrophoresis (PAGE), has been conducted at pH of 4.5, 5.0, and 5.5, where the folded state was expected to be predominant. CD and DSC methods use different concentration scales. For successful DSC experiments, the DNA concentration needs to be at least twenty times higher than that of CD measurements. Thus, comparing T_m and enthalpy change values obtained by DSC with those obtained by CD can give valuable insight into the molecularity and mechanism of the iM folding/unfolding process. When the concentration of an oligonucleotide affects its T_m, it may form dimer or higher molecularities, whereas independent T_m suggests intramolecular structures. CD can be used to indirectly determine the enthalpy change by applying the van't Hoff equation ($\Delta_{vH}H°$), assuming a two-state process. In contrast, DSC directly measures the folding/unfolding enthalpy, which is obtained from the area of the DSC thermogram ($\Delta_{cal}H°$), without any model assumption (Pagano et al., 2013). The difference between $\Delta_{cal}H°$ and $\Delta_{vH}H°$ represents the deviation of the real process from a two-state transition.

CD EXPERIMENTS. CD spectra of the C-rich sequence showed the characteristic CD profile of an iM structure having a positive and a negative band at 288 and 257 nm, confirming the formation of iM at room temperature in both the Na^+- and K^+-containing buffers at the three pH values (Iaccarino et al., 2021). Denaturation and/or renaturation profiles in each condition were acquired both at a wavelength of maximum absorbance variation (λ = 288 nm, heating rate 0.5 °C min^{-1}) and recording whole CD spectra as a function of temperature using a step of 5 °C (**Figure 5.1**).

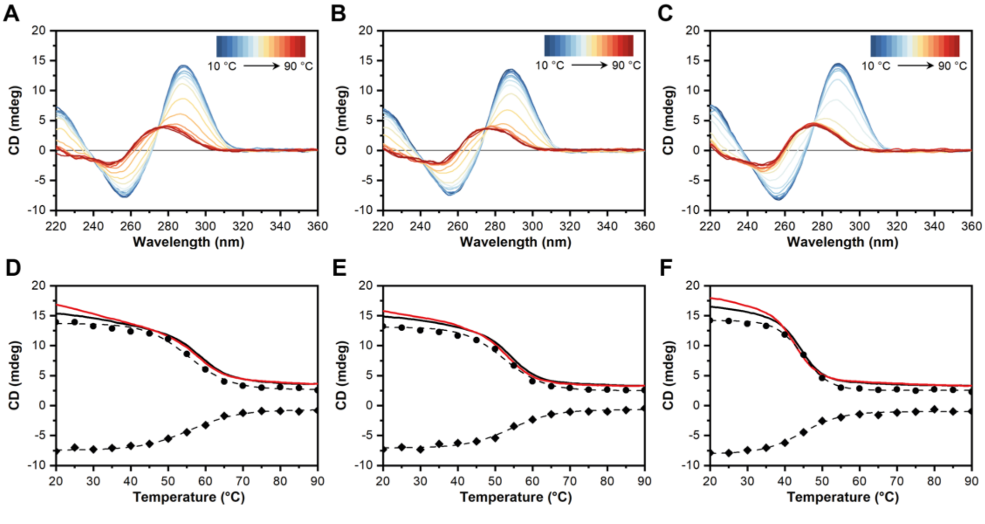


Figure 5.1 (*Top*) CD spectra for the iM-forming oligonucleotide as a function of temperature in 110 mM K^+-acetate buffer; (*Bottom*) CD melting and annealing curves (black and red solid lines, respectively) recorded at 288 nm with a scan rate of 0.5 °C min^{-1}, and CD melting curves at 288 and 257 nm (black dashed lines/circles or rhombus, respectively) derived from corresponding CD spectra at different temperatures (5 °C steps) at pH **A**,**D**) 4.5, **B**,**E**) 5.0, and **C**,**F**) 5.5.

Melting and annealing profiles were almost identical under the experimental conditions, indicating that the melting process is reversible. As expected, the pH increase resulted in a reduction of T_m, while the highest thermal stability was observed at a pH value near the cytosine's pK_a, which changes depending on temperature and ionic strength, ranging between 4.2 and 4.8 (Mergny et al., 1995). Conversely, $\Delta_{vH}H°$ appears to increase with increasing pH, even if the iM structure is less thermally stable due to lower protonation of cytosines (**Table 5.1**). Melting and annealing CD experiments showed no significant hysteresis for oligonucleotide heating/cooling processes or a different effect of scan rate on the T_m of the iM-forming sequence, suggesting that the folding-unfolding process is not kinetically controlled, and thermodynamic equilibrium is achieved during temperature changes.

Table 5.1 Thermodynamic parameters of iM determined by the van't Hoff analysis of CD melting curves.

Cation	**pH**	**T_m (°C)[a]**	**$\Delta_{vH}H°$ (kJ mol^{-1})[a]**
K^+	4.5	60	-188
	5.0	54	-220
	5.5	44	-274
Na^+	4.5	59	-150
	5.0	55	-203
	5.5	44	-262

[a] The error in T_m does not exceed 1 °C, while that in $\Delta_{vH}H°$ amounts to 10% of reported values.

PRINCIPAL COMPONENT ANALYSIS (PCA). The 3D melting curves were submitted to PCA, a decomposition method usually based on Singular Value Decomposition (SVD) (Gray & Chaires, 2011). The PCA decomposition resulted in two principal components that fully explain the variance in the data: *i*) PC1 (96.29% of explained variance), reflects the shape of the folded iM; and *ii*) PC2 (3.59% of explained variance) corresponds to the spectrum of the unfolded species. The PC3 and PC4 (respectively, 0.07% and 0.01% of explained variance) are essentially meaningless. Based on the loading plot with respect to the temperature, it appears that only two species are in balance in the solution. The folded iM structure is mostly present at temperatures below T_m, whereas the unfolded species predominates at higher temperatures. Thus, under these experimental conditions, the iM to single-strand transition is two-state and involves only two significant spectral species.

PAGE EXPERIMENTS. We performed native PAGE experiments at different pH (4.5, 5.0, and 5.5) levels to eliminate the possibility of conformational heterogeneity between monomer and dimer conformations. The electrophoretic mobility of the C-rich oligonucleotide was compared with T10 and T20 oligodeoxythymidylates. Results showed no formation of species with higher molecularity (**Figure 5.2**).

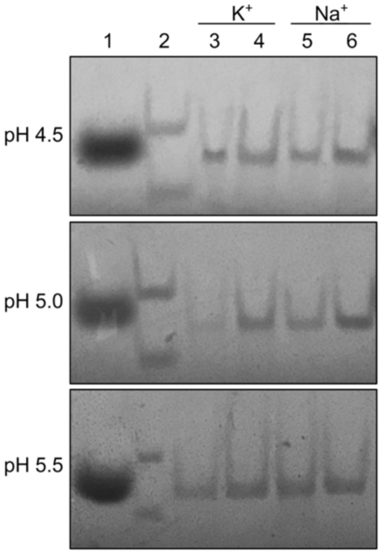


Figure 5.2 PAGE electrophoresis in K^+- or Na^+-acetate buffer at pH 4.5, 5.0, and 5.5, 4 °C. Lane 1: bromophenol blue (run marker). Lane 2: mixture of T10 and T20 oligodeoxynucleotides (size markers). Lanes 3–6: iM-forming oligonucleotide from CD (3,5) or DSC (4,6) samples.

DSC EXPERIMENTS. DSC melting experiments allowed model-independent thermodynamic characterization of the iM stability. iM transitions were confirmed reversible under these experimental conditions. Further, DSC and CD melting demonstrated almost symmetric curves with centered maxima at T_m values. The melting temperature was concentration-independent, indicating unimolecular folding. The integration of the DSC peaks gave $\Delta_{cal}H°$ values in good agreement with $\Delta_{vH}H°$ calculated from CD data (**Table 5.2**). Calorimetric and van't Hoff enthalpy results indicate a two-state equilibrium with no aggregation during the transition. Further, pH has no effect on intramolecular folding kinetics within the studied range.

Thermodynamic parameters for DNA (un)folding were calculated by considering a small heat capacity, (ΔC_p°), related to base dehydration and other contributions (Hadži & Lah, 2021). The folding ΔC_p°value for the iM was assumed as -2.6 kJ mol^{-1} K^{-1} (**Table 5.2**) (Assi et al., 2016).

Table 5.2 Thermodynamic parameters of iM determined by DSC.

Cation	**pH**	T_m **(°C)**[a]	$\Delta_{cal}H°$ **(kJ mol^{-1})**[a]	$\Delta_{cal}S°$ **(kJ mol^{-1} K^{-1})**[a]	$\Delta_{cal}G°_{25\,°C}$ **(kJ mol^{-1})**[a]	n_{CH+}
K^+	4.5	60.6	-158.3	-0.474	-11.8	-
	5.0	53.9	-224.1	-0.685	-16.4	3.1
	5.5	43.5	-261.9	-0.825	-13.9	5.0
Na^+	4.5	60.8	-145.2	-0.435	-10.4	-
	5.0	55.6	-195.5	-0.594	-14.4	2.4
	5.5	45.8	-273.4	-0.857	-16.0	6.1

[a] The error in T_m does not exceed 0.5 °C, while that in $\Delta_{cal}H°$ and $\Delta_{cal}S°$ amounts to 5%, and that in $\Delta_{cal}G°$ to 10% of reported values.

The iM folding is accompanied by a favorable Gibbs free energy change ($\Delta_{cal}G°$) resulting from the relatively large compensation of favorable enthalpy ($\Delta_{cal}H°$), corresponding to the formation of the hemiprotonated C:C$^+$

base pairs, and opposed entropy ($\Delta_{cal}S°$) contributions of the unfavorable ordering of oligonucleotide and the uptake of protons.

At different pH, the calculated $\Delta_{cal}H°$ and $\Delta_{cal}S°$ were very different due to their dependencies on oligonucleotide protonation. The negative entropy observed at high pH values is caused by an increase in proton uptake, while the enthalpy changes observed are affected by both pH and the buffer used, due to the protonation of the C:C$^+$ base pairs. However, to avoid the influence of enthalpy changes relative to the protonation of the buffer species ($\Delta_{buf}H°$), acetate buffer, whose $\Delta_{buf}H°$ is close to zero, has been employed. At pH = pK$_a$ of cytosine (around 4.5), half of them (six in this case) are protonated, no further protonation is required and, being the $\Delta_{buf}H°$ negligible, the measured $\Delta_{cal}H°$ corresponds to the enthalpy of iM formation ($\Delta_{iM}H°$). At pH > pKa, protonation of some of the cytosines is required, and the measured $\Delta_{cal}H°$ is the sum $\Delta_{iM}H°$ and the enthalpy of cytosine protonation ($\Delta_{CH+}H°$), which is:

$$\Delta_{\text{cal}}\text{H}^{\circ} = n_{\text{CH}^+} \cdot \Delta_{\text{CH}^+}\text{H}^{\circ} + \Delta_{\text{iM}}\text{H}^{\circ}$$

where n_{CH+} is the number of protonated cytosines. From the difference between the $\Delta_{cal}H°$ values at pH = pK$_a$ and those at pH 5.0 and 5.5, it is possible to determine n_{CH+}, assuming $\Delta_{CH+}H° = -20.9$ kJ mol^{-1} (Mergny et al., 1995). Finally, at 25 °C the enthalpy–entropy compensation results in a slightly more favorable $\Delta_{cal}G°$ value at pH 5.0 or 5.5 than at pH 4.5. The profile of the $\Delta_{cal}G°$ vs. temperature in K$^+$- or Na$^+$-containing buffer at the three investigated pH is quite similar (**Figure 5.3**). Despite slight differences, at temperatures below 40 °C, iM formation is more favorable at higher pH values (5.0 or 5.5) than at pH 4.5. Thus, T_m is not always a reliable indicator of thermodynamic stabilities, but for a proper comparison, the $\Delta_{cal}G°$ at a specific reference temperature should be considered the relevant parameter.

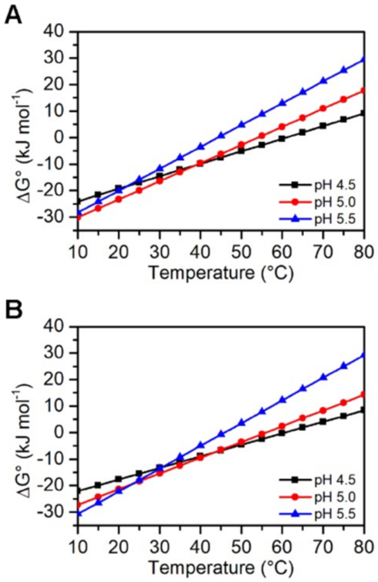


Figure 5.3 Temperature dependence of $\Delta_{cal}G°$ for iM formation in **A**) K^+- or **B**) Na^+-containing solutions at pH 4.5 (black line), 5.0 (red line), and 5.5 (blue line).

5.1.3 CONCLUSIONS

CD spectroscopy and DSC were used to study iM stability at various pH levels in a buffer with K^+/Na^+ cations. The iM folded and unfolded quickly, with reversible melting profiles. CD/DSC profiles showed identical T_m at different DNA concentrations, while native PAGE displayed only a single monomeric species. Our study revealed that thermodynamic and thermal stability are related but not identical properties of iM, cautioning against using thermal stability data for quantitative analysis of folding stability. This study is useful for understanding iM's biological functions and developing pH-sensitive molecular devices.

5.1.4 EXPERIMENTAL SECTION

OLIGONUCLEOTIDE SYNTHESIS AND SAMPLE PREPARATION. The 22-mer deoxyribonucleotide sequence was obtained by chemical synthesis at a 5 mmol scale on an ABI 394 DNA/RNA synthesizer (Applied Biosystem). After synthesis, DNA was detached from support and deprotected by treatment with an aqueous solution of ammonia 32%, at 55 °C for 12 h. The filtrates and washings were combined and concentrated under reduced pressure, solubilized in water, and then purified by high-performance liquid chromatography (HPLC) using an anionic exchange column (Nucleogel SAX, Macherey-Nagel, 1000-8/46) and eluting with a linear gradient from 100% 20 mM MH_2PO_4/M_2HPO_4 aqueous solution (pH 7.0) containing 20% CH_3CN to 100% 1.0 M MCl, 20 mM MH_2PO_4/M_2HPO_4 aqueous solution (pH 7.0) containing 20% CH_3CN % (where M stands for K^+ or Na^+ ion) in 30 min, and a flow rate of 1 mL min^{-1}. The purified fractions of the oligomers were then desalted by using Sep-pak cartridges (C-18). The purity of the isolated oligomer was evaluated by NMR and found to be higher than 98%. Samples for CD and DSC measurements were prepared at the appropriate oligonucleotide concentration in Na^+-or K^+-acetate buffer (110 mM CH_3COOM^+/CH_3COOH) at pH 4.5, 5.0 and 5.5. Before using, oligonucleotide solutions were heated at 90 °C for 5 min, then allowed to cool slowly to room temperature and stored at 4 °C overnight. Sample concentration was determined at 260 nm and 90 °C using the extinction coefficient of 196278 M^{-1} cm^{-1} calculated from the nearest neighbor model (Cantor et al., 1970).

CD EXPERIMENTS. CD experiments were carried out on a Jasco J-815 spectropolarimeter (JASCO Inc., Tokyo, Japan) equipped with a PTC-423S/15 Peltier temperature controller using a quartz cuvette with a path length of 0.1 cm. Oligonucleotide concentration was in the range of 18-20 μM for all CD samples. CD spectra were recorded using the following parameters: 220-360 nm spectral range, 1 nm data pitch, 2 nm bandwidth, 100 nm min^{-1} scan speed, and 4 s response. The buffer baseline was subtracted from each spectrum. CD melting and annealing curves were recorded in the 20–90 °C range by following the change of the CD signal at 288 nm, with a scan rate of 0.5 °C min^{-1}. Three-dimensional melting curves were obtained by recording whole CD spectra as a function of temperature using the same parameters reported above. CD spectra were collected using a temperature step of 5 °C. The CD melting curves at 288 nm were fitted by a two-state transition equation according to the van't Hoff analysis using Origin 7.0 software (OriginLab Corp., MA, USA) (Marky & Breslauer, 1987). The T_m and $\Delta_{\mathrm{vH}}H°$ values provide the best fit of the experimental melting data.

PCA. The CD spectra *vs*. temperature, acquired for each of the six investigated samples, were submitted to a singular value decomposition (SVD) based PCA to assess the number of significant spectral species involved in the equilibrium melting experiments (Wold et al., 1987). Each 3D melting consisted of a 141×17 matrix, the elements of which were ellipticity values arranged with 141 rows, corresponding to wavelengths from 220 to 360 nm, and 17 columns corresponding to temperatures from 10 to 90 °C. Each matrix was then imported in MATLAB R2015b (The MathWorks Inc., MA, USA) environment and submitted to PCA by employing the PLS Toolbox 8.6 (Eigenvector Research Inc., WA, USA). PCA allows the reduction of the data matrix dimensionality, providing a rapid visual representation of the major variance in the data. In fact, the original variables are transformed into a smaller set of new uncorrelated variables, called principal components (PCs), which are ordered according to their explained variance (PC1 explains the largest variance, PC2 contains the second largest variance, and so on). The PCs are visualized in two plots: the "*scores plot*" (where the predominant ellipticity values are reported for each PC) and the "*loadings plot*" (which reports how much of these components is present in the spectra as a function of the temperature). The choice of the correct number of PCs was made by considering both the spectral shapes observed in the scores plot and by evaluating the eigenvalues extracted from each model.

PAGE EXPERIMENTS. Non-denaturing 12% PAGE gels were prepared with 29:1 acrylamide/bis-acrylamide solution and TAE buffer (40 mM Tris base, 20 mM acetic acid, and 2 mM EDTA, adjusted to pH 5.5, 5.0, or 4.5). Gels were run on a Mini-PROTEAN electrophoresis system (Bio-Rad, CA, USA) at 4 °C, 4 mA, for 4–5 h. Appropriate DNA aliquots from CD or DSC samples were properly diluted to have a final amount of 0.1 mg mL^{-1} DNA in each lane (the total DNA amount in different lanes was kept the same). The buffer in DNA samples was 110 mM (Na^+ or K^+) acetate at the appropriate pH (5.5, 5.0, or 4.5). A solution of glycerol/TAE (2:1) was added to facilitate sample loading into the wells. The migration of C-rich DNA was compared with a mixture of two control oligodeoxythymidylates (T10 and T20). Bromophenol blue was used as a run marker. Bands in the gels were visualized by UV shadowing.

DSC EXPERIMENTS. DSC measurements were carried out on a Nano DSC (TA Instruments, USA). The experiments were performed using an oligonucleotide strand concentration of 320–500 μM. Samples were prepared in 110 mM Na^+- or K^+-acetate buffer at pH 4.5, 5.0 and 5.5. Scans were performed at 0.5 °C min^{-1} in the 20-90 °C temperature range. Reversibility was evaluated for each sample by rescanning it after cooling. A buffer-buffer scan was subtracted from the corresponding buffer-sample scan, and a polynomial baseline was drawn for each scan. Baseline-corrected thermograms were then normalized per mole of DNA to obtain the corresponding excess molar heat capacity,

$\langle \Delta C_\mathrm{p} \rangle$, as a function of temperature. The molar heat capacity change, $\Delta C_\mathrm{p}^\circ$, between the folded and unfolded states of DNA, was estimated from DSC thermograms by a second-order polynomial fitting and assumed to be temperature independent. $\Delta_\mathrm{cal}H^\circ$ were obtained by integrating the area under the $\langle \Delta C_\mathrm{p} \rangle$ *vs.* temperature curves. T_m values were determined as the temperatures (T) corresponding to the maximum of each thermogram peak. $\Delta_\mathrm{cal}S^\circ$ values were obtained by integrating the $\langle \Delta C_\mathrm{p} \rangle$/T versus T curves, and $\Delta_\mathrm{cal}G^\circ$ values were computed by:

$$\Delta_\mathrm{cal}G^\circ = \Delta_\mathrm{cal}H^\circ \left[1 - \left(\frac{T}{T_\mathrm{m}}\right)\right] + \Delta C_\mathrm{p}^\circ \left[T - T_\mathrm{m} - T \cdot \ln\frac{T}{T_\mathrm{m}}\right]$$

The reported thermodynamic parameters are the averages of three different heating experiments.

GENERAL CONCLUSIONS

Noncanonical DNA structures, such as G4s and iMs, play important biological roles, participating in essential cellular processes. They have garnered increased attention in recent years as potential targets for treating various diseases, including cancer. Thus, an in-depth investigation of their structures and interactions with potential binders, such as proteins, peptides, or small molecules, capable of either stabilizing or destabilizing them, stands as a pivotal focus in the field of therapeutic research.

Throughout my PhD studies, I successfully identified various novel molecules (proteins and small molecules) capable of binding nucleic acid G4 structures. This accomplishment was realized through the following different approaches: *i*) employing proteomic "*fishing-for-partners*" method to investigate DNA G4s interactomes in U2OS cancer cells (Paragraph 3.1, PAPER VI); *ii*) conducting virtual screening of a commercially available database to discover new small molecules able to bind telomeric RNA G4 (Paragraph 4.1, PAPER II); *iii*) undertaking rational design and synthesis of enhanced DNA G4 binders (Paragraph 4.3, PAPER III); *iv*) utilizing ligand-based drug repurposing to identify SARS-CoV-2 RNA G4 binders (Paragraph 4.4, PAPER IV). The discovery of at least seven novel proteins with the capability to bind G4-forming DNA has prompted biological and biophysical investigations of their interactions that will be conducted in the near future.

In our search for potential G4-targeting drugs, a new hit compound, BPBA, was found through the virtual screening of a commercially available database.

BPBA demonstrated selective binding to *TERRA* RNA G4 with relatively high affinity (K_D in low micromolar range), as well as to two other DNA G4-forming sequences from oncogene promoters, *c-Kit2*, and *c-Myc* (Marzano et al., 2021). Subsequently, an in-depth NMR structural investigation was undertaken to elucidate the structural details of the interaction between the BPBA and G4 structures (Paragraph 4.2, PAPER VII), in collaboration with the research group of Prof. Janez Plavec at the Slovenian NMR Center in Ljubljana (SLO), where I spent several months as a PhD visiting student.

The correlation between G4 affinity, selectivity, cytotoxicity, and immune gene stimulating activation was investigated for a series of rationally designed hydrazone derivatives. The results led to a proposed new rationale, based on low cell-killing potency and high immune stimulation to discover effective anticancer G4 ligands (Marzano et al., 2022).

My research group and I investigated not only G4 structures within the human genome but also within the SARS-CoV-2 viral genome. Utilizing a ligand-based drug repurposing strategy, three FDA-approved drugs (Netarsudil, Pranlukast, and Quercetin) were identified and characterized as effective SARS-CoV-2 G4 binders (Moraca et al., 2022).

Moreover, we recently provided valuable insights into the interaction between a peptide containing an RGG motif shared by at least 77 human G4RPs, NIQI, and various biologically relevant DNA G4s (Grasso et al., 2023).

Finally, our research has also contributed new insights into the energetic aspects of iM structures. Indeed, by means of a combination of spectroscopic and calorimetric techniques, we provided a full thermodynamic characterization of iM formation *in vitro* under favorable conditions, also revealing that thermodynamic and thermal stability are related but not identical properties of such DNA structures (Amato et al., 2021).

In conclusion, the research in the field of noncanonical nucleic acid structures is highly active and very intriguing. Despite the considerable advancements in recent years, there are still numerous questions that remain unanswered. Yet, it is precisely these unanswered queries that contribute to the beauty of scientific exploration!

REFERENCES

Abiri, A., Lavigne, M., Rezaei, M., Nikzad, S., Zare, P., Mergny, J. L., & Rahimi, H. R. (2021). Unlocking G-Quadruplexes as Antiviral Targets. *Pharmacological Reviews*, *73*(3), 897–923. https://doi.org/10.1124/PHARMREV.120.000230

Alba, J. J., Sadurní, A., & Gargallo, R. (2016). Nucleic Acid i-Motif Structures in Analytical Chemistry. *Critical Reviews in Analytical Chemistry*, *46*(5), 443–454. https://doi.org/10.1080/10408347.2016.1143347

Amato, J., D'Aria, F., Marzano, S., Iaccarino, N., Randazzo, A., Giancola, C., & Pagano, B. (2021). On the thermodynamics of folding of an i-motif DNA in solution under favorable conditions. *Physical Chemistry Chemical Physics*, *23*(28). https://doi.org/10.1039/d1cp01779a

Amato, J., Iaccarino, N., Pagano, B., Morigi, R., Locatelli, A., Leoni, A., Rambaldi, M., Zizza, P., Biroccio, A., Novellino, E., & Randazzo, A. (2014). Bis-indole derivatives with antitumor activity turn out to be specific ligands of human telomeric G-quadruplex. *Frontiers in Chemistry*, *2*(JUL), 104080. https://doi.org/10.3389/FCHEM.2014.00054/ABSTRACT

Amato, J., Madanayake, T. W., Iaccarino, N., Novellino, E., Randazzo, A., Hurley, L. H., & Pagano, B. (2018). HMGB1 binds to the KRAS promoter G-quadruplex: a new player in oncogene transcriptional regulation? *Chemical Communications*, *54*(68), 9442–9445. https://doi.org/10.1039/C8CC03614D

Amato, J., Mashima, T., Kamatari, Y. O., Kuwata, K., Novellino, E., Randazzo, A., Giancola, C., Katahira, M., & Pagano, B. (2020). Improved Anti-Prion Nucleic Acid Aptamers by Incorporation of Chemical Modifications. *Nucleic Acid Therapeutics*, *30*(6), 414–421. https://doi.org/10.1089/NAT.2020.0899

Amato, J., Morigi, R., Pagano, B., Pagano, A., Ohnmacht, S., De Magis, A., Tiang, Y. P., Capranico, G., Locatelli, A., Graziadio, A., Leoni, A., Rambaldi, M., Novellino, E., Neidle, S., & Randazzo, A. (2016). Toward the Development of Specific G-Quadruplex Binders: Synthesis, Biophysical, and Biological Studies of New Hydrazone Derivatives. *Journal of Medicinal Chemistry*, *59*(12), 5706–5720. https://doi.org/10.1021/ACS.JMEDCHEM.6B00129

Ambrus, A., Chen, D., Dai, J., Jones, R. A., & Yang, D. (2005). Solution Structure of the Biologically Relevant G-Quadruplex Element in the Human c-MYC Promoter. Implications for G-Quadruplex

Stabilization. *Biochemistry*, *44*(6), 2048–2058. https://doi.org/10.1021/bi048242p

Arora, R., Lee, Y., Wischnewski, H., Brun, C. M., Schwarz, T., & Azzalin, C. M. (2014). RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nature Communications*, *5*. https://doi.org/10.1038/NCOMMS6220

Assi, H. A., Harkness, R. W., Martin-Pintado, N., Wilds, C. J., Campos-Olivas, R., Mittermaier, A. K., González, C., & Damha, M. J. (2016). Stabilization of i-motif structures by 2'-β-fluorination of DNA. *Nucleic Acids Research*, *44*(11), 4998–5009. https://doi.org/10.1093/NAR/GKW402

Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E., & Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science (New York, N.Y.)*, *318*(5851), 798–801. https://doi.org/10.1126/SCIENCE.1147182

Baaske, P., Wienken, C. J., Reineck, P., Duhr, S., & Braun, D. (2010). Optical thermophoresis for quantifying the buffer dependence of aptamer binding. *Angewandte Chemie (International Ed. in English)*, *49*(12), 2238–2241. https://doi.org/10.1002/ANIE.200903998

Balasubramanian, S., Hurley, L. H., & Neidle, S. (2011). Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nature Reviews. Drug Discovery*, *10*(4), 261–275. https://doi.org/10.1038/NRD3428

Banco, M. T., & Ferré-D'Amaré, A. R. (2021). The emerging structural complexity of G-quadruplex RNAs. *RNA*, *27*(4), 390–402. https://doi.org/10.1261/RNA.078238.120

Bansal, A., Kaushik, S., & Kukreti, S. (2022). Non-canonical DNA structures: Diversity and disease association. *Frontiers in Genetics*, *13*, 959258. https://doi.org/10.3389/FGENE.2022.959258/BIBTEX

Barré, J., Sabatier, J. M., & Annweiler, C. (2020). Montelukast Drug May Improve COVID-19 Prognosis: A Review of Evidence. *Frontiers in Pharmacology*, *11*, 570004. https://doi.org/10.3389/FPHAR.2020.01344/BIBTEX

Bartas, M., Červeň, J., Guziurová, S., Slychko, K., & Pečinka, P. (2021). Amino Acid Composition in Various Types of Nucleic Acid-Binding Proteins. *International Journal of Molecular Sciences 2021, Vol. 22, Page 922*, *22*(2), 922. https://doi.org/10.3390/IJMS22020922

Bates, P. J., Reyes-Reyes, E. M., Malik, M. T., Murphy, E. M., O'Toole, M. G., & Trent, J. O. (2017). G-quadruplex oligonucleotide AS1411 as a cancer-targeting agent: Uses and mechanisms. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1861*(5), 1414–1428. https://doi.org/10.1016/J.BBAGEN.2016.12.015

Belmonte-Reche, E., Serrano-Chacón, I., Gonzalez, C., Gallo, J., & Bañobre-López, M. (2021). Potential G-quadruplexes and i-Motifs in the SARS-CoV-2. *PLOS ONE*, *16*(6), e0250654. https://doi.org/10.1371/JOURNAL.PONE.0250654

Bhattacharjee, S., Sengupta, P. K., & Bhowmik, S. (2017). Exploring the preferential interaction of quercetin with VEGF promoter G-quadruplex DNA and construction of a pH-dependent DNA-based logic gate. *RSC Advances*, *7*(59), 37230–37240. https://doi.org/10.1039/C7RA05930B

Biffi, G., Tannahill, D., McCafferty, J., & Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chemistry*, *5*(3), 182–186. https://doi.org/10.1038/NCHEM.1548

Bochman, M. L., Paeschke, K., & Zakian, V. A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. *Nature Reviews. Genetics*, *13*(11), 770. https://doi.org/10.1038/NRG3296

Brázda, V., Cerveň, J., Bartas, M., Mikysková, N., Coufal, J., & Pečinka, P. (2018). The Amino Acid Composition of Quadruplex Binding Proteins Reveals a Shared Motif and Predicts New Potential Quadruplex Interactors. *Molecules 2018, Vol. 23, Page 2341*, *23*(9), 2341. https://doi.org/10.3390/MOLECULES23092341

Brázda, V., Hároníková, L., Liao, J., & Fojta, M. (2014). DNA and RNA Quadruplex-Binding Proteins. *International Journal of Molecular Sciences*, *15*(10), 17493–17517. https://doi.org/10.3390/ijms151017493

Brazier, J. A., Shah, A., & Brown, G. D. (2012). I-Motif formation in gene promoters: unusually stable formation in sequences complementary to known G-quadruplexes. *Chemical Communications*, *48*(87), 10739–10741. https://doi.org/10.1039/C2CC30863K

Brenke, R., Kozakov, D., Chuang, G. Y., Beglov, D., Hall, D., Landon, M. R., Mattos, C., & Vajda, S. (2009). Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics*, *25*(5), 621–627. https://doi.org/10.1093/BIOINFORMATICS/BTP036

Brooks, T. A., Kendrick, S., & Hurley, L. (2010). Making sense of G-quadruplex and i-motif functions in oncogene promoters. *The FEBS Journal*, *277*(17), 3459–3469. https://doi.org/10.1111/J.1742-4658.2010.07759.X

Bryan, T. M. (2020). G-Quadruplexes at Telomeres: Friend or Foe? *Molecules 2020, Vol. 25, Page 3686*, *25*(16), 3686. https://doi.org/10.3390/MOLECULES25163686

Bryan, T. M., Englezou, A., Gupta, J., Bacchetti, S., & Reddel, R. R. (1995). Telomere elongation in immortal human cells without detectable telomerase activity. *The EMBO Journal*, *14*(17), 4240. https://doi.org/10.1002/J.1460-2075.1995.TB00098.X

Bugaut, A., & Balasubramanian, S. (2008). A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadraplexes. *Biochemistry*, *47*(2), 689–697. https://doi.org/10.1021/BI701873C/SUPPL_FILE/BI701873C-FILE002.PDF

Bugaut, A., & Balasubramanian, S. (2012). 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Research*, *40*(11), 4727–4741. https://doi.org/10.1093/NAR/GKS068

Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., & Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Research*, *34*(19), 5402. https://doi.org/10.1093/NAR/GKL655

Cantor, C. R., Warshaw, M. M., & Shapiro, H. (1970). Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, *9*(9), 1059–1077. https://doi.org/10.1002/BIP.1970.360090909

Capra, J. A., Paeschke, K., Singh, M., & Zakian, V. A. (2010). G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. *PLoS Computational Biology*, *6*(7), 9. https://doi.org/10.1371/JOURNAL.PCBI.1000861

Catasti, P., Chen, X., Deaven, L. L., Moyzis, R. K., Bradbury, E. M., & Gupta, G. (1997). Cystosine-rich strands of the insulin minisatellite adopt hairpins with intercalated cytosine+.cytosine pairs. *Journal of Molecular Biology*, *272*(3), 369–382. https://doi.org/10.1006/JMBI.1997.1248

Ceschi, S., Berselli, M., Cozzaglio, M., Giantin, M., Toppo, S., Spolaore, B., & Sissi, C. (2022). Vimentin binds to G-quadruplex repeats found at telomeres and gene promoters. *Nucleic Acids Research*, *50*(3), 1370–1381. https://doi.org/10.1093/nar/gkab1274

Chakraborty, P., & Grosse, F. (2011). Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair*, *10*(6), 654–665. https://doi.org/10.1016/J.DNAREP.2011.04.013

Chambers, V. S., Marsico, G., Boutell, J. M., Di Antonio, M., Smith, G. P., & Balasubramanian, S. (2015). High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nature Biotechnology*, *33*(8), 877–881. https://doi.org/10.1038/NBT.3295

Chen, Y., Qu, K., Zhao, C., Wu, L., Ren, J., Wang, J., & Qu, X. (2012). Insights into the biomedical effects of carboxylated single-wall carbon nanotubes on telomerase and telomeres. *Nature Communications 2012 3:1*, *3*(1), 1–13. https://doi.org/10.1038/ncomms2091

Chu, P. C., Yang, M. C., Kulp, S. K., Salunke, S. B., Himmel, L. E., Fang, C. S., Jadhav, A. M., Shan, Y. S., Lee, C. T., Lai, M. D., Shirley, L. A., Bekaii-Saab, T., & Chen, C. S. (2016). Regulation of oncogenic KRAS signaling via a novel KRAS-integrin-linked kinase-hnRNPA1 regulatory loop in human pancreatic cancer cells. *Oncogene*, *35*(30), 3897–3908. https://doi.org/10.1038/onc.2015.458

Ciliberto, G., & Cardone, L. (2020). Boosting the arsenal against COVID-19 through computational drug repurposing. *Drug Discovery Today*, *25*(6), 946. https://doi.org/10.1016/J.DRUDIS.2020.04.005

Clark, C. L., Cecil, P. K., Singh, D., & Gray, D. M. (1997). CD, absorption and thermodynamic analysis of repeating dinucleotide DNA, RNA and hybrid duplexes [d/r(AC)]12.[d/r(GT/U)]12 and the influence of phosphorothioate substitution. *Nucleic Acids Research*, *25*(20), 4098. https://doi.org/10.1093/NAR/25.20.4098

Cogoi, S., Paramasivam, M., Membrino, A., Yokoyama, K. K., & Xodo, L. E. (2010). The KRAS Promoter Responds to Myc-associated Zinc Finger and Poly(ADP-ribose) Polymerase 1 Proteins, Which Recognize a Critical Quadruplex-forming GA-element. *The Journal of Biological Chemistry*, *285*(29), 22003. https://doi.org/10.1074/JBC.M110.101923

Cogoi, S., Quadrifoglio, F., & Xodo, L. E. (2004). G-rich oligonucleotide inhibits the binding of a nuclear protein to the Ki-ras promoter and strongly reduces cell growth in human carcinoma pancreatic cells. *Biochemistry*, *43*(9), 2512–2523. https://doi.org/10.1021/BI035754F

Cogoi, S., & Xodo, L. E. (2006). G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Research*, *34*(9), 2536–2549. https://doi.org/10.1093/NAR/GKL286

Collie, G., Reszka, A. P., Haider, S. M., Gabelica, V., Parkinson, G. N., & Neidle, S. (2009). Selectivity in small molecule binding to human telomeric RNA and DNA quadruplexes. *Chemical Communications*, *48*, 7482–7484. https://doi.org/10.1039/B901889A

Collie, G. W., Haider, S. M., Neidle, S., & Parkinson, G. N. (2010). A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Research*, *38*(16), 5569–5580. https://doi.org/10.1093/NAR/GKQ259

Crowl, J. T., Gray, E. E., Pestal, K., Volkman, H. E., & Stetson, D. B. (2017). Intracellular Nucleic Acid Detection in Autoimmunity. *Annual Review of Immunology*, *35*, 313–336. https://doi.org/10.1146/ANNUREV-IMMUNOL-051116-052331

Cui, H., & Zhang, L. (2020). G-Quadruplexes Are Present in Human Coronaviruses Including SARS-CoV-2. *Frontiers in Microbiology*, *11*, 567317. https://doi.org/10.3389/FMICB.2020.567317/BIBTEX

Dai, J., Carver, M., Hurley, L. H., & Yang, D. (2011). Solution Structure of a 2:1 Quindoline–c-MYC G-Quadruplex: Insights into G-Quadruplex-Interactive Small Molecule Drug Design. *Journal of the American Chemical Society*, *133*(44), 17673–17680. https://doi.org/10.1021/ja205646q

Dai, J., Chen, D., Jones, R. A., Hurley, L. H., & Yang, D. (2006). NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Research*, *34*(18), 5133. https://doi.org/10.1093/NAR/GKL610

Dai, Y., Teng, X., Zhang, Q., Hou, H., & Li, J. (2023). Advances and challenges in identifying and characterizing G-quadruplex–protein interactions. *Trends in Biochemical Sciences*, *48*(10), 894–909. https://doi.org/10.1016/j.tibs.2023.06.007

De Armond, R., Wood, S., Sun, D., Hurley, L. H., & Ebbinghaus, S. W. (2005). Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry*, *44*(49), 16341–16350. https://doi.org/10.1021/BI051618U

De Cian, A., Guittat, L., Kaiser, M., Saccà, B., Amrane, S., Bourdoncle, A., Alberti, P., Teulade-Fichou, M. P., Lacroix, L., & Mergny, J. L. (2007). Fluorescence-based melting assays for studying quadruplex ligands. *Methods (San Diego, Calif.)*, *42*(2), 183–195. https://doi.org/10.1016/J.YMETH.2006.10.004

De La Faverie, A. R., Guédin, A., Bedrat, A., Yatsunyk, L. A., & Mergny, J. L. (2014). Thioflavin T as a fluorescence light-up probe for G4 formation. *Nucleic Acids Research*, *42*(8), e65. https://doi.org/10.1093/NAR/GKU111

De Lange, T. (2005). *Shelterin: the protein complex that shapes and safeguards human telomeres*. https://doi.org/10.1101/gad.1346005

De Magis, A., Manzo, S. G., Russo, M., Marinello, J., Morigi, R., Sordet, O., & Capranico, G. (2019). DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(3), 816–825. https://doi.org/10.1073/PNAS.1810409116/-/DCSUPPLEMENTAL

Debnath, M., Fatma, K., & Dash, J. (2019). Chemical Regulation of DNA i-Motifs for Nanobiotechnology and Therapeutics. *Angewandte Chemie International Edition*, *58*(10), 2942–2957. https://doi.org/10.1002/ANIE.201813288

Deng, Z., Wang, Z., Stong, N., Plasschaert, R., Moczan, A., Chen, H. S., Hu, S., Wikramasinghe, P., Davuluri, R. V., Bartolomei, M. S., Riethman, H., & Lieberman, P. M. (2012). A role for CTCF and cohesin in subtelomere chromatin organization, TERRA transcription, and telomere end protection. *EMBO Journal*, *31*(21), 4165–4178. https://doi.org/10.1038/EMBOJ.2012.266/SUPPL_FILE/EMBJ2012266.REVIEWER_COMMENTS.PDF

Derosa, G., Maffioli, P., D'Angelo, A., & Di Pierro, F. (2021). A role for quercetin in coronavirus disease 2019 (COVID-19). *Phytotherapy Research : PTR*, *35*(3), 1230–1236. https://doi.org/10.1002/PTR.6887

Di Antonio, M., Biffi, G., Mariani, A., Raiber, E. A., Rodriguez, R., & Balasubramanian, S. (2012). Selective RNA versus DNA G-quadruplex targeting by in situ click chemistry. *Angewandte Chemie (International Ed. in English)*, *51*(44), 11073–11078. https://doi.org/10.1002/ANIE.201206281

Di Fonzo, S., Amato, J., D'Aria, F., Caterino, M., D'Amico, F., Gessini, A., Brady, J. W., Cesàro, A., Pagano, B., & Giancola, C. (2020). Ligand binding to G-quadruplex DNA: new insights from ultraviolet resonance Raman spectroscopy. *Physical Chemistry Chemical Physics*, *22*(15), 8128–8140. https://doi.org/10.1039/D0CP01022G

Dong, Y., Yang, Z., & Liu, D. (2014). DNA nanotechnology based on i-motif structures. *Accounts of Chemical Research*, *47*(6), 1853–1860. https://doi.org/10.1021/AR500073A

Dzatko, S., Krafcikova, M., Hänsel-Hertsch, R., Fessl, T., Fiala, R., Loja, T., Krafcik, D., Mergny, J., Foldynova-Trantirkova, S., & Trantirek, L. (2018). Evaluation of the Stability of DNA i-Motifs in the Nuclei of Living Mammalian Cells. *Angewandte Chemie International Edition*, *57*(8), 2165–2169. https://doi.org/10.1002/anie.201712284

El Deeb, S., Al-Harrasi, A., Khan, A., Al-Broumi, M., Al-Thani, G., Alomairi, M., Elumalai, P., Sayed, R. A., & Ibrahim, A. E. (2022). Microscale thermophoresis as a powerful growing analytical technique for the investigation of biomolecular interaction and the determination of binding parameters. *Methods and Applications in Fluorescence*, *10*(4). https://doi.org/10.1088/2050-6120/AC82A6

Enokizono, Y., Konishi, Y., Nagata, K., Ouhashi, K., Uesugi, S., Ishikawa, F., & Katahira, M. (2005). Structure of hnRNP D complexed with single-stranded telomere DNA and unfolding of the quadruplex by heterogeneous nuclear ribonucleoprotein D. *The Journal of Biological Chemistry*, *280*(19), 18862–18870. https://doi.org/10.1074/JBC.M411822200

Erler, J. T., Cawthorne, C. J., Williams, K. J., Koritzinsky, M., Wouters, B. G., Wilson, C., Miller, C., Demonacos, C., Stratford, I. J., & Dive, C. (2004). Hypoxia-mediated down-regulation of Bid and Bax in tumors occurs via hypoxia-inducible factor 1-dependent and -independent mechanisms and contributes to drug resistance. *Molecular and Cellular Biology*, *24*(7), 2875–2889. https://doi.org/10.1128/MCB.24.7.2875-2889.2004

Ferino, A., Marquevielle, J., Choudhary, H., Cinque, G., Robert, C., Bourdoncle, A., Picco, R., Mergny, J.-L., Salgado, G. F., & Xodo, L. E. (2021). hnRNPA1/UP1 Unfolds KRAS G-Quadruplexes and Feeds a Regulatory Axis Controlling Gene Expression. *ACS Omega*, *6*(49), 34092–34106. https://doi.org/10.1021/acsomega.1c05538

Fernando, H., Reszka, A. P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A. R., Neidle, S., & Balasubramanian, S. (2006). A Conserved Quadruplex Motif Located in a Transcription Activation Site of the Human c-kit Oncogene. *Biochemistry*, *45*(25), 7854. https://doi.org/10.1021/BI0601510

Gaffarogullari, E. C., Krause, A., Balbo, J., Herten, D. P., & Jäschke, A. (2013). Microscale thermophoresis provides insights into mechanism and thermodynamics of ribozyme catalysis. *RNA Biology*, *10*(12), 1815. https://doi.org/10.4161/RNA.27101

Garvie, C. W., & Wolberger, C. (2001). Recognition of Specific DNA Sequences. *Molecular Cell*, *8*(5), 937–946. https://doi.org/10.1016/S1097-2765(01)00392-6

Gehring, K., Leroy, J. L., & Guéron, M. (1993). A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature 1993 363:6429*, *363*(6429), 561–565. https://doi.org/10.1038/363561a0

Ghosh, A., Pandey, S. P., Ansari, A. H., Sundar, J. S., Singh, P., Khan, Y., Ekka, M. K., Chakraborty, D., & Maiti, S. (2022). Alternative splicing modulation mediated by G-quadruplex structures in MALAT1 lncRNA. *Nucleic Acids Research*, *50*(1), 378–396. https://doi.org/10.1093/nar/gkab1066

Ghosh, A., Pandey, S. P., Joshi, D. C., Rana, P., Ansari, A. H., Sundar, J. S., Singh, P., Khan, Y., Ekka, M. K., Chakraborty, D., & Maiti, S. (2023). Identification of G-quadruplex structures in MALAT1 lncRNA that interact with nucleolin and nucleophosmin. *Nucleic Acids Research*. https://doi.org/10.1093/NAR/GKAD639

Giancola, C., & Pagano, B. (2012). Energetics of Ligand Binding to G-Quadruplexes. In J. B. Chaires & D. Graves (Eds.), *Quadruplex Nucleic Acid* (pp. 211–242). Springler. https://doi.org/10.1007/128_2012_347

Gomez, D., Aouali, N., Londoño-Vallejo, A., Lacroix, L., Mégnin-Chanet, F., Lemarteleur, T., Douarre, C., Shin-Ya, K., Mailliet, P., Trentesaux, C., Morjani, H., Mergny, J. L., & Riou, J. F. (2003). Resistance to the Short Term Antiproliferative Activity of the G-quadruplex Ligand 12459 Is Associated with Telomerase Overexpression and Telomere Capping Alteration. *Journal of Biological Chemistry*, *278*(50), 50554–50562. https://doi.org/10.1074/JBC.M308440200

Gomez, D., Wenner, T., Douarre, C., Macadré, J., Koebel, P., Giraud-Panis, M.-J., Kaplan, H., Kolkes, A., Shin-ya, K., & Riou, J.-F. (2006). The G-quadruplex Ligand Telomestatin Inhibits POT1 Binding to Telomeric Sequences In vitro and Induces GFP-POT1 Dissociation from Telomeres in Human Cells. *Cancer Res*, *66*(14), 6908–6920. https://doi.org/10.1158/0008-5472.CAN-06-1581

González, V., Guo, K., Hurley, L., & Sun, D. (2009). Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *The Journal of Biological Chemistry*, *284*(35), 23622–23635. https://doi.org/10.1074/JBC.M109.018028

Grasso, N., Graziano, R., Marzano, S., D'Aria, F., Merlino, F., Grieco, P., Randazzo, A., Pagano, B., & Amato, J. (2023). Unveiling the interaction between DNA G-quadruplexes and RG-rich peptides. *International Journal of Biological Macromolecules*, *253*. https://doi.org/10.1016/j.ijbiomac.2023.126749

Gray, R. D., & Chaires, J. B. (2011). Analysis of multidimensional G-quadruplex melting curves. *Current Protocols in Nucleic Acid Chemistry*, *Chapter 17*(SUPPL. 45). https://doi.org/10.1002/0471142700.NC1704S45

Guédin, A., Gros, J., Alberti, P., & Mergny, J. L. (2010). How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Research*, *38*(21), 7858–7868. https://doi.org/10.1093/NAR/GKQ639

Guéron, M., & Leroy, J. L. (2000). The i-motif in nucleic acids. *Current Opinion in Structural Biology*, *10*(3), 326–331. https://doi.org/10.1016/S0959-440X(00)00091-9

Gunaratnam, M., Merck, C. S., & Morjani, H. (2007). Mechanism of acridine-based telomerase inhibition and telomere shortening. *Biochemical Pharmacology*, *74*, 679–689. https://doi.org/10.1016/j.bcp.2007.06.011

Guo, Q., Marky, L. A., Kallenbach, N. R., & Lu, M. (1992). Interaction of the Dye Ethidium Bromide with DNA Containing Guanine Repeats. *Biochemistry*, *31*(9), 2451–2455. https://doi.org/10.1021/BI00124A002/ASSET/BI00124A002.FP.PNG_V03

Hadži, S., & Lah, J. (2021). Origin of heat capacity increment in DNA folding: The hydration effect. *Biochimica et Biophysica Acta. General Subjects*, *1865*(1). https://doi.org/10.1016/J.BBAGEN.2020.129774

Haider, S., Parkinson, G. N., & Neidle, S. (2008). Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophysical Journal*, *95*(1), 296–311. https://doi.org/10.1529/BIOPHYSJ.107.120501

Hazel, P., Huppert, J., Balasubramanian, S., & Neidle, S. (2004). Loop-length-dependent folding of G-quadruplexes. *Journal of the American Chemical Society*, *126*(50), 16405–16415. https://doi.org/10.1021/JA045154J/SUPPL_FILE/JA045154JSI20040927_115953.ZIP

Herdy, B., Mayer, C., Varshney, D., Marsico, G., Murat, P., Taylor, C., D'Santos, C., Tannahill, D., & Balasubramanian, S. (2018). Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Research*, *46*(21), 11592–11604. https://doi.org/10.1093/nar/gky861

Hirashima, K., & Seimiya, H. (2015). Telomeric repeat-containing RNA/G-quadruplex-forming sequences cause genome-wide alteration of gene expression in human cancer cells in vivo. *Nucleic Acids Research*, *43*(4), 2022–2032. https://doi.org/10.1093/NAR/GKV063

Hoffmann, A. C., Mori, R., Vallbohmer, D., Brabender, J., Klein, E., Drebber, U., Baldus, S. E., Cooc, J., Azuma, M., Metzger, R., Hoelscher, A. H., Danenberg, K. D., Prenzel, K. L., & Danenberg, P. V. (2008). High Expression of HIF1a Is a Predictor of Clinical Outcome in Patients with Pancreatic Ductal Adenocarcinomas and Correlated to PDGFA, VEGF, and bFGF. *Neoplasia (New York, N.Y.)*, *10*(7), 674. https://doi.org/10.1593/NEO.08292

Holzwarth, G., & Doty, P. (1965). The Ultraviolet Circular Dichroism of Polypeptides. *Journal of the American Chemical Society*, *87*(2), 218–228. https://doi.org/10.1021/JA01080A015/ASSET/JA01080A015.FP.PNG_V03

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/NPROT.2008.211

Huang, L., Yuen, T. T. T., Ye, Z., Liu, S., Zhang, G., Chu, H., & Yue, J. (2021). Berbamine inhibits SARS-CoV-2 infection by compromising TRPMLs-mediated endolysosomal trafficking of ACE2. *Signal Transduction and Targeted Therapy 2021 6:1*, *6*(1), 1–3. https://doi.org/10.1038/s41392-021-00584-6

Huang, Z.-L., Dai, J., Luo, W.-H., Wang, X.-G., Tan, J.-H., Chen, S.-B., & Huang, Z.-S. (2018). Identification of G-Quadruplex-Binding Protein from the Exploration of RGG Motif/G-Quadruplex Interactions. *Journal of the American Chemical Society*, *140*(51), 17945–17955. https://doi.org/10.1021/jacs.8b09329

Huppert, J. L. (2010). Structure, location and interactions of G-quadruplexes. *The FEBS Journal*, *277*(17), 3452–3458. https://doi.org/10.1111/J.1742-4658.2010.07758.X

Huppert, J. L., & Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucleic Acids Research*, *33*(9), 2908–2916. https://doi.org/10.1093/NAR/GKI609

Hwang, T. L., & Shaka, A. J. (1995). Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance, Series A*, *112*(2), 275–279. https://doi.org/10.1006/JMRA.1995.1047

Iaccarino, N., Cheng, M., Qiu, D., Pagano, B., Amato, J., Di Porzio, A., Zhou, J., Randazzo, A., & Mergny, J. L. (2021). Effects of Sequence and Base Composition on the CD and TDS Profiles of i-DNA. *Angewandte Chemie International Edition*, *60*(18), 10295–10303. https://doi.org/10.1002/ANIE.202016822

Ilc, T., Šket, P., Plavec, J., Webba Da Silva, M., Drevenšek-Olenik, I., & Spindler, L. (2013). Formation of G-wires: The role of G:C-base pairing and G-quartet stacking. *Journal of Physical Chemistry C*, *117*(44), 23208–23215. https://doi.org/10.1021/JP4019348/SUPPL_FILE/JP4019348_SI_001.PDF

James, W. H., Müller, C. W., Buchanan, E. G., Nix, M. G. D., Guo, L., Roskop, L., Gordon, M. S., Slipchenko, L. V., Gellman, S. H., & Zwier, T. S. (2009). Intramolecular amide stacking and its competition with hydrogen bonding in a small foldamer. *Journal of the American Chemical Society*, *131*(40), 14243–14245. https://doi.org/10.1021/JA9054965/SUPPL_FILE/JA9054965_SI_001.PDF

Jang, W. D., Jeon, S., Kim, S., & Lee, S. Y. (2021). Drugs repurposed for COVID-19 by virtual screening of 6,218 drugs and cell-based assay. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(30). https://doi.org/10.1073/PNAS.2024302118

Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P., & Duhr, S. (2011). Molecular interaction studies using microscale thermophoresis. *Assay and Drug Development Technologies*, *9*(4), 342–353. https://doi.org/10.1089/ADT.2011.0380

Ji, D., Juhas, M., Tsang, C. M., Kwok, C. K., Li, Y., & Zhang, Y. (2021). Discovery of G-quadruplex-forming sequences in SARS-CoV-2. *Briefings in Bioinformatics*, *22*(2), 1150–1160. https://doi.org/10.1093/BIB/BBAA114

Jia, Y., Ye, L., Ji, K., Toms, A.-M., Davies, M. L., Ruge, F., Ji, J., Hargest, R., & Jiang, W. G. (2014). Death associated protein 1 is correlated with the clinical outcome of patients with colorectal cancer and has a role in the regulation of cell death. *Oncology Reports*, *31*(1), 175–182. https://doi.org/10.3892/or.2013.2866

Joachimi, A., Benz, A., & Hartig, J. S. (2009). A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorganic & Medicinal Chemistry*, *17*(19), 6811–6815. https://doi.org/10.1016/J.BMC.2009.08.043

Kang, H.-J., Kendrick, S., Hecht, S. M., & Hurley, L. H. (2014). The Transcriptional Complex Between the *BCL2* i-Motif and hnRNP LL Is a Molecular Switch for Control of Gene Expression That Can Be Modulated by Small Molecules. *Journal of the American Chemical Society*, *136*(11), 4172–4185. https://doi.org/10.1021/ja4109352

Kendrick, S., Kang, H. J., Alam, M. P., Madathil, M. M., Agrawal, P., Gokhale, V., Yang, D., Hecht, S. M., & Hurley, L. H. (2014). The dynamic character of the BCL2 promoter i-motif provides a mechanism for modulation of gene expression by compounds that bind selectively to the alternative DNA hairpin structure. *Journal of the American Chemical Society*, *136*(11), 4161–4171. https://doi.org/10.1021/JA410934B

Kotar, A., Rigo, R., Sissi, C., & Plavec, J. (2019). Two-quartet kit* G-quadruplex is formed via double-stranded pre-folded structure. *Nucleic Acids Research*, *47*(5), 2641–2653. https://doi.org/10.1093/NAR/GKY1269

Krüger, A. C., Raarup, M. K., Nielsen, M. M., Kristensen, M., Besenbacher, F., Kjems, J., & Birkedal, V. (2010). Interaction of hnRNP A1 with telomere DNA G-quadruplex structures studied at the single molecule level. *European Biophysics Journal*, *39*(9), 1343–1350. https://doi.org/10.1007/s00249-010-0587-x

Kuryavyi, V., Phan, A. T., & Patel, D. J. (2010). Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter. *Nucleic Acids Research*, *38*(19), 6757–6773. https://doi.org/10.1093/nar/gkq558

Kwok, C. K., Marsico, G., & Balasubramanian, S. (2018). Detecting RNA G-Quadruplexes (rG4s) in the Transcriptome. *Cold Spring Harbor Perspectives in Biology*, *10*(7). https://doi.org/10.1101/CSHPERSPECT.A032284

Lago, S., Tosoni, E., Nadai, M., Palumbo, M., & Richter, S. N. (2017). The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. *Biochimica et Biophysica Acta. General Subjects*, *1861*(5 Pt B), 1371–1381. https://doi.org/10.1016/J.BBAGEN.2016.11.036

Lavezzo, E., Berselli, M., Frasson, I., Perrone, R., Palù, G., Brazzale, A. R., Richter, S. N., & Toppo, S. (2018). G-quadruplex forming sequences in the genome of all known human viruses: A comprehensive guide. *PLoS Computational Biology*, *14*(12). https://doi.org/10.1371/JOURNAL.PCBI.1006675

Le, V. H., Buscaglia, R., Chaires, J. B., & Lewis, E. A. (2013). Modeling complex equilibria in isothermal titration calorimetry experiments: Thermodynamic parameters estimation for a three-binding-site model. *Analytical Biochemistry*, *434*(2), 233–241. https://doi.org/10.1016/J.AB.2012.11.030

Lehmann, U., Celikkaya, G., Hasemeier, B., Länger, F., & Hans Kreipe, H. (2002). Promoter Hypermethylation of the Death-associated Protein Kinase Gene in Breast Cancer Is Associated with the Invasive Lobular Subtype1. *Cancer Research*, *62*(22), 6634–6638. https://aacrjournals.org/cancerres/article/62/22/6634/509327/Promoter-Hypermethylation-of-the-Death-associated

Leonetti, C., Amodei, S., D’Angelo, C., Rizzo, A., Benassi, B., Antonelli, A., Elli, R., Stevens, M. F. G., D’Incalci, M., Zupi, G., & Biroccio, A. (2004). Biological activity of the G-quadruplex ligand RHPS4 (3,11-difluoro-6,8,13-trimethyl-8H-quino[4,3,2-kl]acridinium methosulfate) is associated with telomere capping alteration. *Molecular Pharmacology*, *66*(5), 1138–1146. https://doi.org/10.1124/MOL.104.001537

Li, L., Williams, P., Ren, W., Wang, M. Y., Gao, Z., Miao, W., Huang, M., Song, J., & Wang, Y. (2020). YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nature Chemical Biology 2020 17:2*, *17*(2), 161–168. https://doi.org/10.1038/s41589-020-00695-1

Liao, Z. Z., Wang, Y. Di, Qi, X. Y., & Xiao, X. H. (2019). JAZF1, a relevant metabolic regulator in type 2 diabetes. *Diabetes/Metabolism Research and Reviews*, *35*(5), e3148. https://doi.org/10.1002/DMRR.3148

Lieblein, A. L., Buck, J., Schlepckow, K., Fürtig, B., & Schwalbe, H. (2012). Time-resolved NMR spectroscopic studies of DNA i-motif folding reveal kinetic partitioning. *Angewandte Chemie (International Ed. in English)*, *51*(1), 250–253. https://doi.org/10.1002/ANIE.201104938

Lieblein, A. L., Fürtig, B., & Schwalbe, H. (2013). Optimizing the Kinetics and Thermodynamics of DNA i-Motif Folding. *ChemBioChem*, *14*(10), 1226–1230. https://doi.org/10.1002/CBIC.201300284

Linke, R., Limmer, M., Juranek, S. A., Heine, A., & Paeschke, K. (2021). The Relevance of G-Quadruplexes for DNA Repair. *International Journal of Molecular Sciences 2021, Vol. 22, Page 12599*, *22*(22), 12599. https://doi.org/10.3390/IJMS222212599

Liu, K. W., Hu, B., & Cheng, S. Y. (2011). Platelet-derived growth factor signaling in human malignancies. *Chinese Journal of Cancer*, *30*(9), 581–584. https://doi.org/10.5732/CJC.011.10300

Ma, D., Xing, Z., Liu, B., Pedigo, N. G., Zimmer, S. G., Bai, Z., Postel, E. H., & Kaetzel, D. M. (2002). NM23-H1 and NM23-H2 repress transcriptional activities of nuclease-hypersensitive elements in the platelet-derived growth factor-A promoter. *The Journal of Biological Chemistry*, *277*(2), 1560–1567. https://doi.org/10.1074/JBC.M108359200

Mackenzie, D., Watters, A. K., To, J. T., Young, M. W., Muratori, J., Wilkoff, M. H., Abraham, R. G., Plummer, M. M., & Zhang, D. (2021). ALT Positivity in Human Cancers: Prevalence and Clinical Insights. *Cancers*, *13*(10). https://doi.org/10.3390/CANCERS13102384

Malgowska, M., Czajczynska, K., Gudanis, D., Tworak, A., & Gdaniec, Z. (2016). Overview of RNA G-quadruplex structures. *Acta Biochimica Polonica*, *63*(4), 609–621. https://doi.org/10.18388/ABP.2016_1335

Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., & Wüthrich, K. (1998). Recommendations for the presentation of NMR structures of proteins and nucleic acids. *Journal of Molecular Biology*, *280*(5), 933–952. https://doi.org/10.1006/JMBI.1998.1852

Marky, L. A., & Breslauer, K. J. (1987). Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers*, *26*(9), 1601–1620. https://doi.org/10.1002/BIP.360260911

Marsh, T. C., & Henderson, E. (1994). G-wires: self-assembly of a telomeric oligonucleotide, d(GGGGTTGGGG), into large superstructures. *Biochemistry*, *33*(35), 10718–10724. https://doi.org/10.1021/BI00201A020

Martadinata, H., & Phan, A. T. (2009). Structure of Propeller-Type Parallel-Stranded RNA G-Quadruplexes, Formed by Human Telomeric RNA Sequences in K+ Solution. *Journal of the American Chemical Society*, *131*(7), 2570–2578. https://doi.org/10.1021/ja806592z

Maruca, A., Ambrosio, F. A., Lupia, A., Romeo, I., Rocca, R., Moraca, F., Talarico, C., Bagetta, D., Catalano, R., Costa, G., Artese, A., & Alcaro, S. (2019). Computer-based techniques for lead identification and optimization i: Basics. *Physical Sciences Reviews*, *4*(6). https://doi.org/10.1515/PSR-2018-0113/ASSET/GRAPHIC/J_PSR-2018-0113_FIG_006.JPG

Marzano, S., Miglietta, G., Morigi, R., Marinello, J., Arleo, A., Procacci, M., Locatelli, A., Leoni, A., Pagano, B., Randazzo, A., Amato, J., & Capranico, G. (2022). Balancing Affinity, Selectivity, and Cytotoxicity of Hydrazone-Based G-Quadruplex Ligands for Activation of Interferon β Genes in Cancer Cells. *Journal of Medicinal Chemistry*, *65*(18). https://doi.org/10.1021/acs.jmedchem.2c00772

Marzano, S., Pagano, B., Iaccarino, N., Di Porzio, A., De Tito, S., Vertecchi, E., Salvati, E., Randazzo, A., & Amato, J. (2021). Targeting of telomeric repeat-containing RNA G-quadruplexes: From screening to biophysical and biological characterization of a new hit compound. *International Journal of Molecular Sciences*, *22*(19). https://doi.org/10.3390/ijms221910315

Masuzawa, T., & Oyoshi, T. (2020). Roles of the RGG Domain and RNA Recognition Motif of Nucleolin in G-Quadruplex Stabilization. *ACS Omega*, *5*(10), 5202–5208. https://doi.org/10.1021/ACSOMEGA.9B04221/ASSET/IMAGES/LARGE/AO9B04221_0003.JPEG

McDonald, E. F., Jones, T., Plate, L., Meiler, J., & Gulsevin, A. (2023). Benchmarking AlphaFold2 on peptide structure prediction. *Structure*, *31*(1), 111-119.e2. https://doi.org/10.1016/J.STR.2022.11.012

McKim, M., Buxton, A., Johnson, C., Metz, A., & Sheardy, R. D. (2016). Loop Sequence Context Influences the Formation and Stability of the i-Motif for DNA Oligomers of Sequence (CCCXXX)4, where X = A and/or T, under Slightly Acidic Conditions. *Journal of Physical Chemistry B*, *120*(31), 7652–7661. https://doi.org/10.1021/ACS.JPCB.6B04561/SUPPL_FILE/JP6B04561_SI_001.PDF

Meier-Stephenson, V. (2022). G4-quadruplex-binding proteins: review and insights into selectivity. *Biophysical Reviews*, *14*(3), 635–654. https://doi.org/10.1007/s12551-022-00952-8

Mendoza, O., Bourdoncle, A., Boulé, J. B., Brosh, R. M., & Mergny, J. L. (2016). G-quadruplexes and helicases. *Nucleic Acids Research*, *44*(5), 1989–2006. https://doi.org/10.1093/NAR/GKW079

Mergny, J. L., Lacroix, L., Hélène, C., Han, X., & Leroy, J. L. (1995). Intramolecular Folding of Pyrimidine Oligodeoxynucleotides into an i-DNA Motif. *Journal of the American Chemical Society*, *117*(35), 8887–8898. https://doi.org/10.1021/JA00140A001/ASSET/JA00140A001.FP.PNG_V03

Mergny, J. L., Lacroix, L., Teulade-Fichou, M. P., Hounsou, C., Guittat, L., Hoarau, M., Arimondo, P. B., Vigneron, J. P., Lehn, J. M., Riou, J. F., Garestier, T., & Hélène, C. (2001). Telomerase inhibitors based on quadruplex ligands selected by a fluorescence assay. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(6), 3062. https://doi.org/10.1073/PNAS.051620698

Micsonai, A., Wien, F., Kernya, L., Lee, Y. H., Goto, Y., Réfrégiers, M., & Kardos, J. (2015). Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(24), E3095–E3103. https://doi.org/10.1073/PNAS.1500851112/SUPPL_FILE/PNAS.201500851SI.PDF

Miglietta, G., Marinello, J., Russo, M., & Capranico, G. (2022). Ligands stimulating antitumour immunity as the next G-quadruplex challenge. *Molecular Cancer 2022 21:1*, *21*(1), 1–15. https://doi.org/10.1186/S12943-022-01649-Y

Miglietta, G., Russo, M., Duardo, R. C., & Capranico, G. (2021). G-quadruplex binders as cytostatic modulators of innate immune genes in cancer cells. *Nucleic Acids Research*, *49*(12), 6673–6686. https://doi.org/10.1093/NAR/GKAB500

Minenkova, O., Santapaola, D., Milazzo, F. M., Anastasi, A. M., Battistuzzi, G., Chiapparino, C., Rosi, A., Gritti, G., Borleri, G., Rambaldi, A., Dental, C., Viollet, C., Pagano, B., Salvini, L., Marra, E., Luberto, L., Rossi, A., Riccio, A., Merlo Pich, E., … De Santis, R. (2022). Human inhalable antibody fragments neutralizing SARS-CoV-2 variants for COVID-19 therapy. *Molecular Therapy*, *30*(5), 1979. https://doi.org/10.1016/J.YMTHE.2022.02.013

Mir, B., Serrano, I., Buitrago, D., Orozco, M., Escaja, N., & González, C. (2017). Prevalent Sequences in the Human Genome Can Form Mini i-Motif Structures at Physiological pH. *Journal of the American Chemical Society*, *139*(40), 13985–13988. https://doi.org/10.1021/JACS.7B07383/SUPPL_FILE/JA7B07383_SI_001.PDF

Monchaud, D., Allain, C., & Teulade-Fichou, M. P. (2006). Development of a fluorescent intercalator displacement assay (G4-FID) for establishing quadruplex-DNA affinity and selectivity of putative ligands. *Bioorganic & Medicinal Chemistry Letters*, *16*(18), 4842–4845. https://doi.org/10.1016/J.BMCL.2006.06.067

Montero, J. J., López De Silanes, I., Granã, O., & Blasco, M. A. (2016). Telomeric RNAs are essential to maintain telomeres. *Nature Communications 2016 7:1*, *7*(1), 1–13. https://doi.org/10.1038/ncomms12534

Moraca, F., Marzano, S., D'Amico, F., Lupia, A., Di Fonzo, S., Vertecchi, E., Salvati, E., Di Porzio, A., Catalanotti, B., Randazzo, A., Pagano, B.,

& Amato, J. (2022). Ligand-based drug repurposing strategy identified SARS-CoV-2 RNA G-quadruplex binders. *Chemical Communications*, *58*(85), 11913–11916. https://doi.org/10.1039/D2CC03135C

Mori, K., Lammich, S., Mackenzie, I. R. A., Forné, I., Zilow, S., Kretzschmar, H., Edbauer, D., Janssens, J., Kleinberger, G., Cruts, M., Herms, J., Neumann, M., Van Broeckhoven, C., Arzberger, T., & Haass, C. (2013). hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta Neuropathologica*, *125*(3), 413–423. https://doi.org/10.1007/S00401-013-1088-7

Moye, A. L., Porter, K. C., Cohen, S. B., Phan, T., Zyner, K. G., Sasaki, N., Lovrecz, G. O., Beck, J. L., & Bryan, T. M. (2015). Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nature Communications*, *6*. https://doi.org/10.1038/NCOMMS8643

Murat, P., Marsico, G., Herdy, B., Ghanbarian, A., Portella, G., & Balasubramanian, S. (2018). RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biology*, *19*(1), 1–24. https://doi.org/10.1186/S13059-018-1602-2/FIGURES/6

Nanni, P., Landuzzi, L., Nicoletti, G., De Giovanni, C., Giovarelli, M., Lalli, E., Facchini, A., & Lollini, P. L. (1992). Control of H-2 expression in transformed nonhaemopoietic cells by autocrine interferon. *British Journal of Cancer*, *66*(3), 479–482. https://doi.org/10.1038/BJC.1992.299

Ng, Y. L., Salim, C. K., & Chu, J. J. H. (2021). Drug repurposing for COVID-19: Approaches, challenges and promising candidates. *Pharmacology & Therapeutics*, *228*, 107930. https://doi.org/10.1016/J.PHARMTHERA.2021.107930

Nguyen, T., Fraire, C., & Sheardy, R. D. (2017). Linking pH, Temperature, and K+ Concentration for DNA i-Motif Formation. *The Journal of Physical Chemistry. B*, *121*(33), 7872–7877. https://doi.org/10.1021/ACS.JPCB.7B06317

Nieto, C. I., Cabildo, P., García, M. Á., Claramunt, R. M., Alkorta, I., & Elguero, J. (2014). An experimental and theoretical NMR study of NH-benzimidazoles in solution and in the solid state: proton transfer and tautomerism. *Beilstein Journal of Organic Chemistry*, *10*, 1620–1629. https://doi.org/10.3762/bjoc.10.168

Nonin-Lecomte, S., & Leroy, J. L. (2001). Structure of a C-rich strand fragment of the human centromeric satellite III: a pH-dependent intercalation topology. *Journal of Molecular Biology*, *309*(2), 491–506. https://doi.org/10.1006/JMBI.2001.4679

Nourian, A., Khalili, H., Ahmadinejad, Z., Kouchak, H. E., Jafari, S., Manshadi, S. A. D., Rasolinejad, M., & Kebriaeezadeh, A. (2020). Efficacy and safety of sofosbuvir/ ledipasvir in treatment of patients with COVID-19; A randomized clinical trial. *Acta Bio-Medica : Atenei Parmensis*, *91*(4), 1–14. https://doi.org/10.23750/ABM.V91I4.10877

Ohashi, H., Watashi, K., Saso, W., Shionoya, K., Iwanami, S., Hirokawa, T., Shirai, T., Kanaya, S., Ito, Y., Kim, K. S., Nomura, T., Suzuki, T., Nishioka, K., Ando, S., Ejima, K., Koizumi, Y., Tanaka, T., Aoki, S., Kuramochi, K., … Wakita, T. (2021). Potential anti-COVID-19 agents, cepharanthine and nelfinavir, and their usage for combination treatment. *IScience*, *24*(4), 102367. https://doi.org/10.1016/J.ISCI.2021.102367

Pagano, B., Margarucci, L., Zizza, P., Amato, J., Iaccarino, N., Cassiano, C., Salvati, E., Novellino, E., Biroccio, A., Casapullo, A., & Randazzo, A. (2015). Identification of novel interactors of human telomeric G-quadruplex DNA. *Chemical Communications*, *51*(14), 2964–2967. https://doi.org/10.1039/C4CC07231F

Pagano, B., Mattia, C. A., & Giancola, C. (2009). Applications of Isothermal Titration Calorimetry in Biophysical Studies of G-quadruplexes. *International Journal of Molecular Sciences 2009, Vol. 10, Pages 2935-2957*, *10*(7), 2935–2957. https://doi.org/10.3390/IJMS10072935

Pagano, B., Randazzo, A., Fotticchia, I., Novellino, E., Petraccone, L., & Giancola, C. (2013). Differential scanning calorimetry to investigate G-quadruplexes structural stability. *Methods (San Diego, Calif.)*, *64*(1), 43–51. https://doi.org/10.1016/J.YMETH.2013.02.018

Paramasivam, M., Membrino, A., Cogoi, S., Fukuda, H., Nakagama, H., & Xodo, L. E. (2009). Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription. *Nucleic Acids Research*, *37*(9), 2841–2853. https://doi.org/10.1093/nar/gkp138

Parkinson, G. N., Lee, M. P. H., & Neidle, S. (2002). Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, *417*(6891), 876–880. https://doi.org/10.1038/NATURE755

Patel, D. J., Phan, A. T., & Kuryavyi, V. (2007). Human telomere, oncogenic promoter and 5’-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Research*, *35*(22), 7429–7455. https://doi.org/10.1093/NAR/GKM711

Pavc, D., Sebastian, N., Spindler, L., Drevenšek-Olenik, I., Podboršek, G. K., Plavec, J., & Šket, P. (2022). Understanding self-assembly at molecular level enables controlled design of DNA G-wires of different properties. *Nature Communications*, *13*(1). https://doi.org/10.1038/S41467-022-28726-6

Perrone, R., Nadai, M., Frasson, I., Poe, J. A., Butovskaya, E., Smithgall, T. E., Palumbo, M., Paluì, G., & Richter, S. N. (2013). A Dynamic G-quadruplex region regulates the HIV-1 Long terminal repeat promoter. *Journal of Medicinal Chemistry*, *56*(16), 6521–6530. https://doi.org/10.1021/JM400914R/SUPPL_FILE/JM400914R_SI_002.PDF

Petr, M., Helma, R., Polášková, A., Krejči, A., Dvořáková, Z., Kejnovská, I., Navrátilová, L., Adámik, M., Vorlíčková, M., & Brázdová, M. (2016). Wild-type p53 binds to MYC promoter G-quadruplex. *Bioscience Reports*, *36*(5), 397. https://doi.org/10.1042/BSR20160232

Phan, A. T. (2010). Human telomeric G-quadruplex: structures of DNA and RNA sequences. *The FEBS Journal*, *277*(5), 1107–1117. https://doi.org/10.1111/J.1742-4658.2009.07464.X

Phan, A. T., Guéron, M., & Leroy, J. L. (2000). The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. *Journal of Molecular Biology*, *299*(1), 123–144. https://doi.org/10.1006/JMBI.2000.3613

Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S., & Patel, D. J. (2007). Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-kit Promoter. *Journal of the American Chemical Society*, *129*(14), 4386–4392. https://doi.org/10.1021/ja068739h

Phatak, P., Cookson, J. C., Dai, F., Smith, V., Gartenhaus, R. B., Stevens, M. F. G., & Burger, A. M. (2007). Telomere uncapping by the G-quadruplex ligand RHPS4 inhibits clonogenic tumour cell growth in vitro and in vivo consistent with a cancer stem cell targeting mechanism. *British Journal of Cancer 2007 96:8*, *96*(8), 1223–1233. https://doi.org/10.1038/sj.bjc.6603691

Pilger, D., Seymour, L. W., & Jackson, S. P. (2021). Interfaces between cellular responses to DNA damage and cancer immunotherapy. *Genes & Development*, *35*(9–10), 602–618. https://doi.org/10.1101/GAD.348314.121

Pirota, V., Platella, C., Musumeci, D., Benassi, A., Amato, J., Pagano, B., Colombo, G., Freccero, M., Doria, F., & Montesarchio, D. (2021). On the binding of naphthalene diimides to a human telomeric G-quadruplex multimer model. *International Journal of Biological Macromolecules*, *166*, 1320–1334. https://doi.org/10.1016/J.IJBIOMAC.2020.11.013

Pompili, L., Leonetti, C., Biroccio, A., & Salvati, E. (2017). Diagnosis and treatment of ALT tumors: Is Trabectedin a new therapeutic option? *Journal of Experimental and Clinical Cancer Research*, *36*(1). https://doi.org/10.1186/S13046-017-0657-3

Pouwels, K. B., Pritchard, E., Matthews, P. C., Stoesser, N., Eyre, D. W., Vihta, K. D., House, T., Hay, J., Bell, J. I., Newton, J. N., Farrar, J., Crook, D., Cook, D., Rourke, E., Studley, R., Peto, T. E. A., Diamond, I., & Walker, A. S. (2021). Effect of Delta variant on viral burden and vaccine effectiveness against new SARS-CoV-2 infections in the UK. *Nature Medicine*, *27*(12), 2127–2135. https://doi.org/10.1038/S41591-021-01548-7

Proctor, W. G., & Yu, F. C. (1950). The Dependence of a Nuclear Magnetic Resonance Frequency upon Chemical Compound. *Physical Review*, *77*(5), 717. https://doi.org/10.1103/PhysRev.77.717

Prokunina-Olsson, L., Fu, Y. P., Tang, W., Jacobs, K. B., Hayes, R. B., Kraft, P., Berndt, S. I., Wacholder, S., Yu, K., Hutchinson, A., Feigelson, H. S., Thun, M. J., Diver, W. R., Albanes, D., Virtamo, J., Weinstein, S., Schumacher, F. R., Cancel-Tassin, G., Cussenot, O., … Yeager, M. (2010). Refining the prostate cancer genetic association within the JAZF1 gene on chromosome 7p15.2. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *19*(5), 1349–1355. https://doi.org/10.1158/1055-9965.EPI-09-1181

Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., Doig, A., Guilliams, T., Latimer, J., McNamee, C., Norris, A., Sanseau, P., Cavalla, D., & Pirmohamed, M. (2019). Drug repurposing: progress, challenges and recommendations. *Nature Reviews. Drug Discovery*, *18*(1), 41–58. https://doi.org/10.1038/NRD.2018.168

Qin, Y., Rezler, E. M., Gokhale, V., Sun, D., & Hurley, L. H. (2007). Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Research*, *35*(22), 7698–7713. https://doi.org/10.1093/nar/gkm538

Raiber, E. A., Kranaster, R., Lam, E., Nikan, M., & Balasubramanian, S. (2012). A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Research*, *40*(4), 1499–1508. https://doi.org/10.1093/nar/gkr882

Rajendran, A., Nakano, S. I., & Sugimoto, N. (2010). Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH. *Chemical Communications*, *46*(8), 1299–1301. https://doi.org/10.1039/B922050J

Randazzo, A., Spada, G. P., & da Silva, M. W. (2013). Circular dichroism of quadruplex structures. *Topics in Current Chemistry*, *330*, 67–86. https://doi.org/10.1007/128_2012_331

Rankin, S., Reszka, A. P., Huppert, J., Zloh, M., Parkinson, G. N., Todd, A. K., Ladame, S., Balasubramanian, S., & Neidle, S. (2005). Putative DNA quadruplex formation within the human c-kit oncogene. *Journal of the American Chemical Society*, *127*(30), 10584–10589. https://doi.org/10.1021/JA050823U/ASSET/IMAGES/MEDIUM/JA050823UN00001.GIF

Redon, S., Reichenbach, P., & Lingner, J. (2010). The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Research*, *38*(17), 5797–5806. https://doi.org/10.1093/NAR/GKQ296

Rey, J., Murail, S., De Vries, S., Derreumaux, P., & Tuffery, P. (2023). PEP-FOLD4: a pH-dependent force field for peptide structure prediction in aqueous solution. *Nucleic Acids Research*, *51*(W1), W432–W437. https://doi.org/10.1093/NAR/GKAD376

Rhodes, D., & Lipps, H. J. (2015). G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Research*, *43*(18), 8627–8637. https://doi.org/10.1093/NAR/GKV862

Rocca, R., Moraca, F., Costa, G., Nadai, M., Scalabrin, M., Talarico, C., Distinto, S., Maccioni, E., Ortuso, F., Artese, A., Alcaro, S., & Richter, S. N. (2017). Identification of G-quadruplex DNA/RNA binders: Structure-based virtual screening and biophysical characterization. *Biochimica et Biophysica Acta. General Subjects*, *1861*(5 Pt B), 1329–1340. https://doi.org/10.1016/J.BBAGEN.2016.12.023

Rodriguez, R., Müller, S., Yeoman, J. A., Trentesaux, C., Riou, J. F., & Balasubramanian, S. (2008). A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *Journal of the American Chemical Society*, *130*(47), 15758–15759. https://doi.org/10.1021/JA805615W/SUPPL_FILE/JA805615W_SI_001.PDF

Roy, S., Ali, A., Kamra, M., Muniyappa, K., & Bhattacharya, S. (2020). Specific stabilization of promoter G-Quadruplex DNA by 2,6-disubstituted amidoanthracene-9,10-dione based dimeric distamycin analogues and their selective cancer cell cytotoxicity. *European Journal of Medicinal Chemistry*, *195*. https://doi.org/10.1016/J.EJMECH.2020.112202

Salvati, E., Leonetti, C., Rizzo, A., Scarsella, M., Mottolese, M., Galati, R., Sperduti, I., Stevens, M. F. G., D'Incalci, M., Blasco, M., Chiorino, G., Bauwens, S., Horard, B., Gilson, E., Stoppacciaro, A., Zupi, G., & Biroccio, A. (2007). Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *The Journal of Clinical Investigation*, *117*(11), 3236–3247. https://doi.org/10.1172/JCI32461

Sanchez-Martin, V. (2023). DNA G-Quadruplex-Binding Proteins: An Updated Overview. *DNA*, *3*(1), 1–12. https://doi.org/10.3390/dna3010001

Santos, T., Miranda, A., Campello, M. P. C., Paulo, A., Salgado, G., Cabrita, E. J., & Cruz, C. (2021). Recognition of nucleolin through interaction with RNA G-quadruplex. *Biochemical Pharmacology*, *189*. https://doi.org/10.1016/J.BCP.2020.114208

Santos, T., Salgado, G. F., Cabrita, E. J., & Cruz, C. (2021). G-Quadruplexes and Their Ligands: Biophysical Methods to Unravel G-Quadruplex/Ligand Interactions. *Pharmaceuticals 2021, Vol. 14, Page 769*, *14*(8), 769. https://doi.org/10.3390/PH14080769

Scognamiglio, P. L., Di Natale, C., Leone, M., Poletto, M., Vitagliano, L., Tell, G., & Marasco, D. (2014). G-quadruplex DNA recognition by nucleophosmin: new insights from protein dissection. *Biochimica et Biophysica Acta*, *1840*(6), 2050–2059. https://doi.org/10.1016/J.BBAGEN.2014.02.017

Seidel, S. A. I., Dijkman, P. M., Lea, W. A., van den Bogaart, G., Jerabek-Willemsen, M., Lazic, A., Joseph, J. S., Srinivasan, P., Baaske, P., Simeonov, A., Katritch, I., Melo, F. A., Ladbury, J. E., Schreiber, G., Watts, A., Braun, D., & Duhr, S. (2013). Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods (San Diego, Calif.)*, *59*(3), 301–315. https://doi.org/10.1016/J.YMETH.2012.12.005

Sengupta, B., Pahari, B., Blackmon, L., & Sengupta, P. K. (2013). Prospect of Bioflavonoid Fisetin as a Quadruplex DNA Ligand: A Biophysical Approach. *PLOS ONE*, *8*(6), e65383. https://doi.org/10.1371/JOURNAL.PONE.0065383

Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., Imamichi, T., & Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research*, *50*(W1), W216–W221. https://doi.org/10.1093/NAR/GKAC194

Shu, H., Zhang, R., Xiao, K., Yang, J., & Sun, X. (2022). G-Quadruplex-Binding Proteins: Promising Targets for Drug Design. *Biomolecules*, *12*(5), 648. https://doi.org/10.3390/biom12050648

Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., & Hurley, L. H. (2002). Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(18), 11593–11598. https://doi.org/10.1073/PNAS.182256799

Silva, B., Arora, R., Bione, S., & Azzalin, C. M. (2021). TERRA transcription destabilizes telomere integrity to initiate break-induced replication in human ALT cells. *Nature Communications 2021 12:1*, *12*(1), 1–12. https://doi.org/10.1038/s41467-021-24097-6

Simonsson, T., Pecinka, P., & Kubista, M. (1998). DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Research*, *26*(5), 1167–1172. https://doi.org/10.1093/NAR/26.5.1167

Smargiasso, N., Rosu, F., Hsia, W., Colson, P., Baker, E. S., Bowers, M. T., De Pauw, E., & Gabelica, V. (2008). G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. *Journal of the American Chemical Society*, *130*(31), 10208–10216. https://doi.org/10.1021/JA801535E

Sreeramulu, S., Richter, C., Berg, H., Wirtz Martin, M. A., Ceylan, B., Matzel, T., Adam, J., Altincekic, N., Azzaoui, K., Bains, J. K., Blommers, M. J. J., Ferner, J., Fürtig, B., Göbel, M., Grün, J. T., Hengesbach, M., Hohmann, K. F., Hymon, D., Knezic, B., … Schwalbe, H. (2021). Exploring the Druggability of Conserved RNA Regulatory Elements in the SARS-CoV-2 Genome. *Angewandte Chemie International Edition*, *60*(35), 19191–19200. https://doi.org/10.1002/ANIE.202103693

Sun, D., Guo, K., Rusche, J. J., & Hurley, L. H. (2005). 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 Research*, *33*(18), 6070–6080. https://doi.org/10.1093/NAR/GKI917

Sun, D., & Hurley, L. H. (2009). The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: Implications for drug targeting and control of gene expression. *Journal of Medicinal Chemistry*, *52*(9), 2863–2874. https://doi.org/10.1021/JM900055S/SUPPL_FILE/JM900055S_SI_001.PDF

Sun, D., Thompson, B., Cathers, B. E., Salazar, M., Kerwin, S. M., Trent, J. O., Jenkins, T. C., Neidle, S., & Hurley, L. H. (1997). Inhibition of human telomerase by a G-quadruplex-interactive compound. *Journal of Medicinal Chemistry*, *40*(14), 2113–2116. https://doi.org/10.1021/JM970199Z

Sung, Y., Park, S., Park, S. J., Jeong, J., Choi, M., Lee, J., Kwon, W., Jang, S., Lee, M.-H., Kim, D. J., Liu, K., Kim, S.-H., Lee, J.-H., Ha, Y.-S., Kwon, T. G., Lee, S., Dong, Z., Ryoo, Z. Y., Kim, M. O., … Ok Kim, M. (2017). Jazf1 promotes prostate cancer progression by activating JNK/Slug. *Oncotarget*, *9*(1), 755–765. https://doi.org/10.18632/ONCOTARGET.23146

Sutherland, C., Cui, Y., Mao, H., & Hurley, L. H. (2016). A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III1. *Journal of the American Chemical*

Society, *138*(42), 14138–14151. https://doi.org/10.1021/JACS.6B09196/SUPPL_FILE/JA6B09196_SI_001.PDF

Temime-Smaali, N., Guittat, L., Sidibe, A., Shin-Ya, K., Trentesaux, C., & Riou, J. F. (2009). The G-Quadruplex Ligand Telomestatin Impairs Binding of Topoisomerase IIIα to G-Quadruplex-Forming Oligonucleotides and Uncaps Telomeres in ALT Cells. *PLOS ONE*, *4*(9), e6919. https://doi.org/10.1371/JOURNAL.PONE.0006919

Thakur, R. K., Kumar, P., Halder, K., Verma, A., Kar, A., Parent, J. L., Basundra, R., Kumar, A., & Chowdhury, S. (2009). Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. *Nucleic Acids Research*, *37*(1), 172. https://doi.org/10.1093/NAR/GKN919

Todd, A. K., Johnston, M., & Neidle, S. (2005). Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Research*, *33*(9), 2901–2907. https://doi.org/10.1093/NAR/GKI553

Tong, X., Lan, W., Zhang, X., Wu, H., Liu, M., & Cao, C. (2011). Solution structure of all parallel G-quadruplex formed by the oncogene RET promoter sequence. *Nucleic Acids Research*, *39*(15), 6753–6763. https://doi.org/10.1093/NAR/GKR233

Tosoni, E., Frasson, I., Scalabrin, M., Perrone, R., Butovskaya, E., Nadai, M., Palù, G., Fabris, D., & Richter, S. N. (2015). Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Research*, *43*(18), 8884–8897. https://doi.org/10.1093/nar/gkv897

Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455–461. https://doi.org/10.1002/JCC.21334

Tsvetkov, V., Pozmogova, G., & Varizhuk, A. (2016). The systematic approach to describing conformational rearrangements in G-quadruplexes. *Journal of Biomolecular Structure & Dynamics*, *34*(4), 705. https://doi.org/10.1080/07391102.2015.1055303

Vo, T., Brownmiller, T., Hall, K., Jones, T. L., Choudhari, S., Grammatikakis, I., Ludwig, K. R., & Caplen, N. J. (2022). HNRNPH1 destabilizes the G-quadruplex structures formed by G-rich RNA sequences that regulate the alternative splicing of an oncogenic fusion transcript. *Nucleic Acids Research*, *50*(11), 6474–6496. https://doi.org/10.1093/nar/gkac409

Wang, Y., & Patel, D. J. (1993). Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure (London, England : 1993)*, *1*(4), 263–282. https://doi.org/10.1016/0969-2126(93)90015-9

Watson, J. D., & Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature 1953 171:4356*, *171*(4356), 737–738. https://doi.org/10.1038/171737a0

Webb, B. A., Chimenti, M., Jacobson, M. P., & Barber, D. L. (2011). Dysregulated pH: a perfect storm for cancer progression. *Nature Reviews. Cancer*, *11*(9), 671–677. https://doi.org/10.1038/NRC3110

Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D., & Duhr, S. (2010). Protein-binding assays in biological liquids using microscale thermophoresis. *Nature Communications 2010 1:1*, *1*(1), 1–7. https://doi.org/10.1038/ncomms1093

Williamson, J. R., Raghuraman, M. K., & Cech, T. R. (1989). Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell*, *59*(5), 871–880. https://doi.org/10.1016/0092-8674(89)90610-7

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*(1–3), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

Wright, E. P., Huppert, J. L., & Waller, Z. A. E. (2017). Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Research*, *45*(6), 2951–2959. https://doi.org/10.1093/NAR/GKX090

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., & Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265–269. https://doi.org/10.1038/S41586-020-2008-3

Wu, S., Wang, X., Ye, X., & Zhang, G. (2013). pH-Induced conformational change and dimerization of DNA chains investigated by analytical ultracentrifugation. *The Journal of Physical Chemistry. B*, *117*(39), 11541–11547. https://doi.org/10.1021/JP405561F

Xu, Y., Suzuki, Y., Ito, K., & Komiyama, M. (2010). Telomeric repeat-containing RNA structure in living cells. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(33), 14579–14584. https://doi.org/10.1073/PNAS.1001177107/SUPPL_FILE/PNAS.1001177107_SI.PDF

Yan, K. K. P., Obi, I., & Sabouri, N. (2021). The RGG domain in the C-terminus of the DEAD box helicases Dbp2 and Ded1 is necessary for G-quadruplex destabilization. *Nucleic Acids Research*, *49*(14), 8339–8354. https://doi.org/10.1093/nar/gkab620

You, J., Li, H., Lu, X. M., Li, W., Wang, P. Y., Dou, S. X., & Xi, X. G. (2017). Effects of monovalent cations on folding kinetics of G-quadruplexes. *Bioscience Reports*, *37*(4), 20170771. https://doi.org/10.1042/BSR20170771

Zakian, V. A. (1995). Telomeres: Beginning to Understand the End. *Science*, *270*(5242), 1601–1607. https://doi.org/10.1126/SCIENCE.270.5242.1601

Zeraati, M., Langley, D. B., Schofield, P., Moye, A. L., Rouet, R., Hughes, W. E., Bryan, T. M., Dinger, M. E., & Christ, D. (2018). I-motif DNA structures are formed in the nuclei of human cells. *Nature Chemistry 2018 10:6*, *10*(6), 631–637. https://doi.org/10.1038/s41557-018-0046-3

Zhang, D. H., Fujimoto, T., Saxena, S., Yu, H. Q., Miyoshi, D., & Sugimoto, N. (2010). Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. *Biochemistry*, *49*(21), 4554–4563. https://doi.org/10.1021/BI1002822

Zhang, Y., Ou, T.-M., Tan, J.-H., & Huang, Z. (2018). *Syntheses and Evaluation of New Acridone Derivatives for Selective Binders of Oncogene c-myc Promoter i-Motif in Gene Transcriptional Regulation*. https://doi.org/10.1039/C8CC00328A

Zhao, C., Qin, G., Niu, J., Wang, Z., Wang, C., Ren, J., & Qu, X. (2021). Targeting RNA G-Quadruplex in SARS-CoV-2: A Promising Therapeutic Target for COVID-19? *Angewandte Chemie (International Ed. in English)*, *60*(1), 432–438. https://doi.org/10.1002/ANIE.202011419

Zillner, K., Filarsky, M., Rachow, K., Weinberger, M., Längst, G., & Németh, A. (2013). Large-scale organization of ribosomal DNA chromatin is regulated by Tip5. *Nucleic Acids Research*, *41*(10), 5251–5262. https://doi.org/10.1093/NAR/GKT218

Zuffo, M., Guédin, A., Leriche, E. D., Doria, F., Pirota, V., Gabelica, V., Mergny, J. L., & Freccero, M. (2018). More is not always better: finding the right trade-off between affinity and selectivity of a G-quadruplex ligand. *Nucleic Acids Research*, *46*(19), e115–e115. https://doi.org/10.1093/NAR/GKY607

APPENDIX

Paper I

"On the thermodynamics of folding of an i-motif DNA in solution under favorable conditions". Amato J., D'Aria F., **Marzano S.**, Iaccarino N., Randazzo A., Giancola C., Pagano B. *Physical Chemistry Chemical Physics*, 2021, *23*(28), 15030–15037.

Paper II

"Targeting of telomeric repeat-containing RNA G-quadruplexes: from screening to biophysical and biological characterization of a new hit compound". **Marzano S.**, Pagano B., Iaccarino N., Di Porzio A., De Tito S., Vertecchi E., Salvati E., Randazzo A., Amato J. *International Journal of Molecular Sciences*, 2021, *22*(19), 10315–10335.

Paper III

"Balancing affinity, selectivity, and cytotoxicity of hydrazone-based G-quadruplex ligands for activation of interferon β genes in cancer cells". **Marzano S.**, Miglietta G., Morigi R., Marinello J., Arleo A., Procacci M., Locatelli A., Leoni A., Pagano B., Randazzo A., Amato J., Capranico G. *Journal of Medicinal Chemistry*, 2022, *65*(18), 12055–12067.

Paper IV

"Ligand-based drug repurposing strategy identified SARS-CoV-2 RNA G-quadruplex binders". Moraca F., **Marzano S.**, D'Amico F., Lupia A., Di Fonzo S., Vertecchi E., Salvati E., Di Porzio A., Catalanotti B., Randazzo A., Pagano B., Amato J. *Chemical Communications*, 2022, *58*(85), 11913–11916.

Paper V

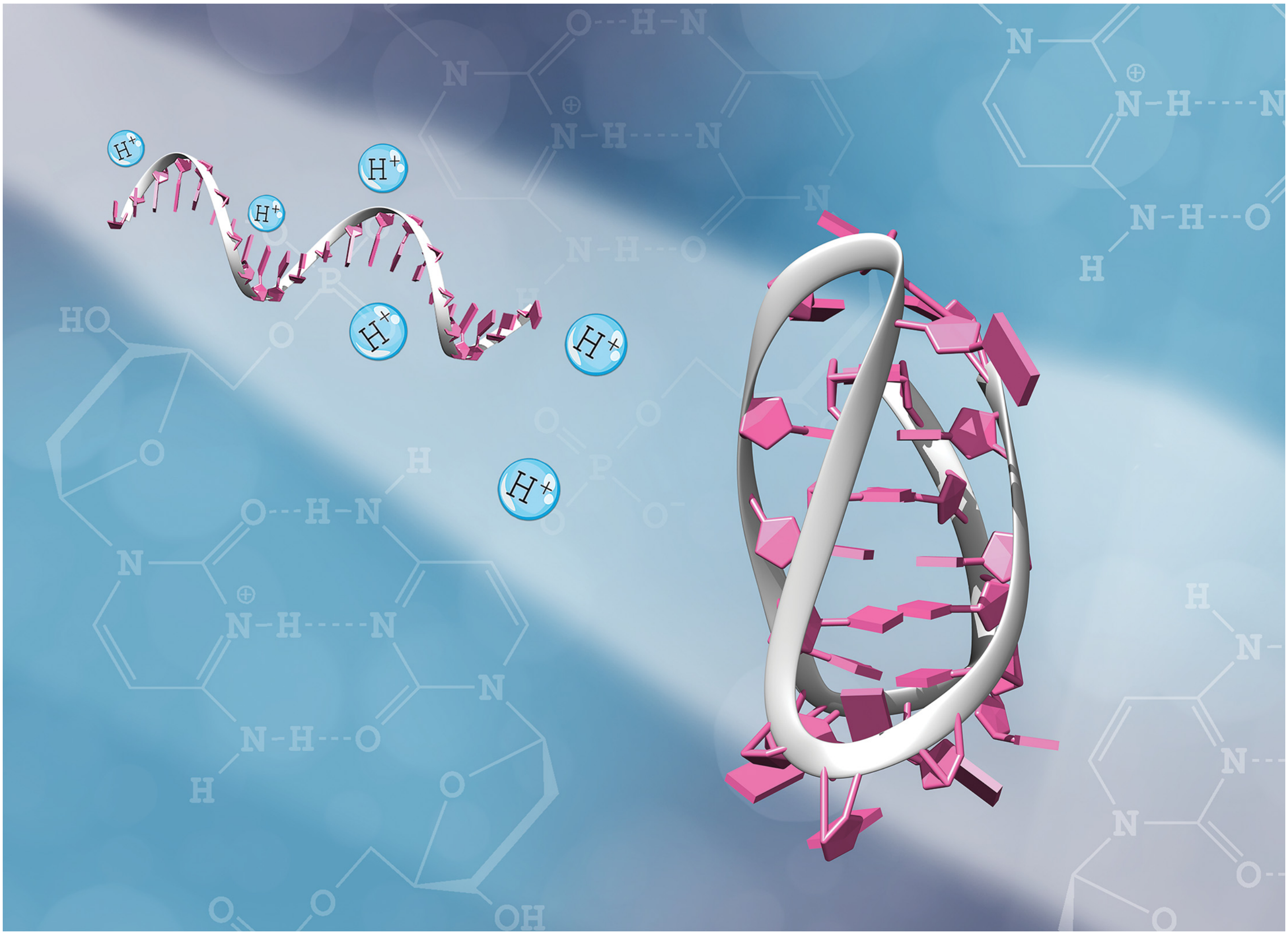
“Unveiling the interaction between DNA G-quadruplexes and RG-rich peptides”. Grasso N., Graziano R., **Marzano S.**, D'Aria F., Merlino F., Grieco P., Randazzo A., Pagano B., Amato J. *International Journal of Biological Macromolecules*, 2023, *253*(Part 3), 126749–126759.

PAPER I

On the thermodynamics of folding of an i-motif DNA in solution under favorable conditions

Amato J., D'Aria F., **Marzano S.**, Iaccarino N., Randazzo A., Giancola C., Pagano B.

Physical Chemistry Chemical Physics, 2021, Vol. 23(28)



Showcasing research from the group of Prof. Bruno Pagano at Department of Pharmacy, University of Naples Federico II, Italy

On the thermodynamics of folding of an i-motif DNA in solution under favorable conditions

This work investigates the energetics of an i-motif DNA structure in solution under mildly acidic conditions that promote its formation. Using spectroscopic and calorimetric techniques, a complete thermodynamic characterization of i-motif formation was performed, including the evaluation of the energetic contributions of cytosine protonation. This work provides new insights into the formation and stability of such noncanonical DNA structures that are useful for both those interested in their biological functions and those aiming to develop pH-sensitive molecular devices.

As featured in:

See Bruno Pagano *et al.*, *Phys. Chem. Chem. Phys.*, 2021, **23**, 15030.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/pccp

Registered charity number: 207890

PCCP

PAPER

Check for updates

Cite this: *Phys. Chem. Chem. Phys.*, 2021, **23**, 15030

Received 23rd April 2021,
Accepted 16th June 2021

DOI: 10.1039/d1cp01779a

rsc.li/pccp

On the thermodynamics of folding of an i-motif DNA in solution under favorable conditions†

Jussara Amato, ‡ Federica D'Aria, ‡ Simona Marzano, Nunzia Iaccarino, Antonio Randazzo, Concetta Giancola and Bruno Pagano *

Under slightly acidic conditions, cytosine-rich DNA sequences can form non-canonical secondary structures called i-motifs, which occur as four stretches of cytosine repeats form hemi-protonated $C{\cdot}C^+$ base pairs. The growing interest in the i-motif structures as important components in functional DNA-based nanotechnology or as potential targets of anticancer drugs, increases the need for a deep understanding of the energetics of their structural transitions. Here, a combination of spectroscopic and calorimetric techniques is used to unravel the thermodynamics of folding of an i-motif DNA under favorable conditions. The results give new insights into the energetic aspects of i-motifs and show that thermodynamic and thermal stability are related but not identical properties of such DNA structures.

Introduction

Nucleic acids have the ability to form a number of different hydrogen bonding patterns. This enables DNA to attain a variety of structural and conformational polymorphic forms.[1,2] Structurally, DNA can exist in single-stranded form or fold into secondary structures, which include the canonical double helix as well as non-canonical structures such as the G-quadruplex and i-motif species.[1,3] Non-canonical structures in turn encompass an ensemble of conformers depending on the sequence composition and environmental conditions.

In vivo, non-canonical DNA structures may form in important genomic regions and intervene in several biological processes, including the modulation of oncogenes expression, and therefore represent potential anticancer drug targets.[4–6] In addition, the nanoscale geometry and dynamic properties of non-canonical DNA, along with its inherent biocompatibility and biodegradability, have made such DNA motifs promising candidates for the development of novel functional nanostructures and nanodevices.[7,8]

Among the non-canonical DNA structures, growing interest is currently being paid to the i-motif (iM), a four stranded structure which can form in cytosine-rich sequences.[9] Indeed, despite the first iM structure was reported in 1993,[10] the presence of endogenous iM *in vivo* has only recently been confirmed.[11,12]

The iM consists of two parallel-stranded duplexes, stabilized by intercalated, hemi-protonated cytosine–cytosine ($C{\cdot}C^+$) base pairs (Fig. 1A).[9,13] The need for protonated cytosines in the formation of an iM results in a strong dependence of the structure's stability on pH,[14] which has also led to potential applications in the field of sensors and logic devices.[15] Indeed, one of the main limitations of nanostructures based on canonical Watson–Crick base pairing lies in the lack of a certain sensitivity to chemical stimuli. On the contrary, iM-based nanostructures would appear to be more useful due to their greater conformational flexibility in response to a chemical stimulus such as pH changes.[16]

The biological importance of iMs and their potential applications in nanotechnology as highly sensitive pH-responsive switches,[15] or, for example, as probes to monitor spatiotemporal pH changes in living cells or the dysregulated pH of a tumor microenvironment,[17,18] have prompted many research groups to devote their efforts to study this non-canonical fold and the factors that influence it. Intrinsic factors that affect iM behavior are the length of the oligonucleotide, the length and number of the C-tracts, and the nature of the loops (length and base composition).[19–24] Besides pH, extrinsic factors influencing iM structural transitions include temperature, and salt concentration and identity.[14,24–28]

In this frame, the knowledge of the thermodynamic and kinetic properties of iM structures under different experimental conditions, such as solution pH and cation type, is central to the construction as well as the function of nucleic acid nanodevices. Despite the formation of iMs using different DNA oligonucleotides has been extensively studied by a variety of experimental techniques, not many systematic studies have been carried out to date using differential scanning calorimetry (DSC) to perform a model-independent thermodynamic analysis of an iM *in vitro* under different solution conditions.

In this work, we used DSC and circular dichroism (CD) spectroscopy as complementary physicochemical methodologies

---

Department of Pharmacy, University of Naples Federico II, Naples, I-80131, Italy. E-mail: bruno.pagano@unina.it

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1cp01779a

‡ Co-first authors.

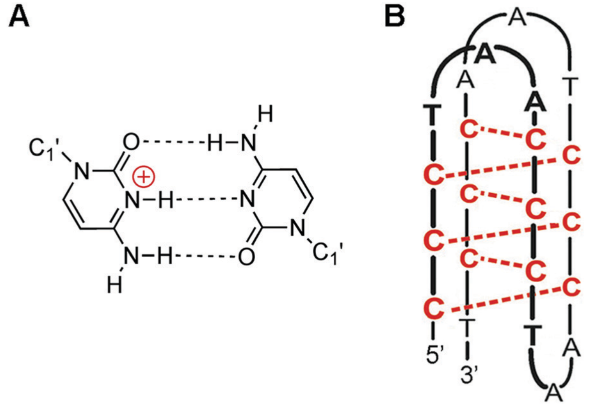


Fig. 1 (A) Hemi-protonated cytosine–cytosine (C·C$^+$) base pair, and (B) schematic representation of an intramolecular i-motif structure.

to characterize the energetics of a DNA iM formation. The iM-forming sequence we studied is the 22 nucleotide-long sequence d(CCCTAA)$_3$CCCT, which is a fragment of vertebrate telomere DNA. At acidic pH, this C-rich sequence was shown to form an intramolecular structure that includes six C·C$^+$ base pairs (Fig. 1B).[29] An NMR study on the shorter (21 nucleotides) telomeric sequence d(CCCTAA)$_3$CCC suggested also the existence of an equilibrium between two energetically close conformations differing in the intercalation topology of the C·C$^+$ base pairs.[30] In one, denoted 5′E, the outermost C·C$^+$ pair of the iM is formed by the cytidine at the 5′-end of the oligonucleotide. In the other, denoted 3′E, the outermost cytidine is that at the 3′-end.

Usually, for any iM-forming sequence, attention first focuses on a parameter called transition pH (pH$_T$), defined as the pH, at a given temperature, at which 50% of the oligodeoxynucleotide is in the folded state. As for the telomere-derived C-rich oligonucleotides, the pH$_T$ has been investigated under various experimental conditions and, at temperatures ranging from 10 to 30 °C, it was found in the range 5.9–6.4.[21,25,31]

Here our focus centers on the study of telomeric iM under favorable conditions that lead to its stable formation. Therefore, we have investigated the thermodynamics of the iM in solution at three different pH values below the pH$_T$ (4.5, 5.0, and 5.5) and in the presence of two different cations (Na$^+$ or K$^+$). For any application in which the iM would be the primary sensing element, a deep knowledge of energetics of iM structural transitions under various conditional constraints could be used to improve sensor accuracy and could be exploited to give highly specific information about environmental changes within narrow pH ranges. We report complete thermodynamic profiles of iM, including the evaluation of the energetic contributions of cytosine protonation.

Experimental

Oligonucleotide synthesis and sample preparation

The 22-mer deoxyribonucleotide sequence d(CCCTAA)$_3$CCCT was obtained by chemical synthesis at 5 μmol scale on an ABI 394 DNA/RNA synthesizer (Applied Biosystem). After synthesis, DNA was detached from support and deprotected by treatment with an aqueous solution of ammonia 32%, at 55 °C for 12 h. The filtrates and washings were combined and concentrated under reduced pressure, solubilized in water, and then purified by high-performance liquid chromatography (HPLC) using an anionic exchange column (Nucleogel SAX, Macherey-Nagel, 1000-8/46) and eluting with a linear gradient from 100% buffer A to 100% buffer B in 30 min, and a flow rate of 1 mL min^{-1}. Buffer A = 20 mM MH$_2$PO$_4$/M$_2$HPO$_4$ aqueous solution (pH 7.0) containing 20% CH$_3$CN (where M stands for K$^+$ or Na$^+$ ion). Buffer B = 1.0 M MCl, 20 mM MH$_2$PO$_4$/M$_2$HPO$_4$ aqueous solution (pH 7.0) containing 20% CH$_3$CN. The purified fractions of the oligomers were then desalted by using Sep-pak cartridges (C-18). The purity of the isolated oligomer was evaluated by NMR and found to be higher than 98%. Samples for CD and DSC measurements were prepared at the appropriate oligonucleotide concentration in Na$^+$- or K$^+$-containing acetate buffer (110 mM CH$_3$COO$^-$M$^+$/CH$_3$COOH) at pH 4.5, 5.0 and 5.5. Before using, oligonucleotide solutions were heated at 90 °C for 5 min, then allowed to cool slowly to room temperature and stored at 4 °C overnight. Sample concentration was determined at 260 nm and 90 °C using the extinction coefficient of 196 278 M^{-1} cm^{-1} calculated from the nearest neighbor model.[32]

Circular dichroism (CD) spectroscopy

CD experiments were carried out on a Jasco J-815 spectropolarimeter (JASCO Inc., Tokyo, Japan) equipped with a PTC-423S/15 Peltier temperature controller using a quartz cuvette with a path length of 0.1 cm. Oligonucleotide concentration was in the range 18–20 μM for all CD samples. CD spectra were recorded using the following parameters: 220–360 nm spectral range, 1 nm data pitch, 2 nm bandwidth, 100 nm min^{-1} scan speed, 4 s response. Buffer baseline was subtracted from each spectrum. CD melting and annealing curves were recorded in the 20–90 °C range by following the change of the CD signal at 288 nm, with a scan rate of 0.5 °C min^{-1}. Three-dimensional melting curves were obtained by recording whole CD spectra as a function of temperature using the same parameters reported above. CD spectra were collected using a temperature step of 5 °C. The CD melting curves at 288 nm were fitted by a two-state transition equation according to the van't Hoff analysis using Origin 7.0 software (OriginLab Corp., MA, USA).[33] The melting temperature (T_m) and enthalpy change ($\Delta_{vH}H°$) values provide the best fit of the experimental melting data.

Principal component analysis (PCA)

The CD spectra *versus* temperature (3D melting), acquired for each of the six investigated samples, were submitted to a singular value decomposition (SVD) based PCA[34] to assess the number of significant spectral species involved in the equilibrium melting experiments. Each 3D melting consisted of a 141 × 17 matrix, the elements of which were ellipticity values arranged with 141 rows, corresponding to wavelengths from 220 to 360 nm, and 17 columns corresponding to temperatures from 10 to 90 °C. Each matrix was then imported in MATLAB R2015b (The MathWorks Inc., MA, USA) environment and submitted to PCA by employing the PLS Toolbox 8.6 (Eigenvector Research Inc., WA, USA). PCA allows the reduction of the data matrix dimensionality, providing a rapid visual representation of the major variance in the data. In fact, the original variables are transformed into a smaller set of new uncorrelated variables, called principal components (PCs), which are ordered according to their explained variance (PC1 explains the largest variance, PC2 contains the second largest variance, and so on). The PCs are visualized in two plots: the "scores plot" (where the predominant ellipticity values are reported for each PC) and the "loadings plot" (which reports how much of these components is present in the spectra as a function of the temperature). The choice of the correct number of PCs was made by considering both the spectral shapes observed in the scores plot and by evaluating the eigenvalues extracted from each model. Table S1 (ESI†) reports the eigenvalues of the first four PCs generated from each of the six PCA models. In all cases, PC1 and PC2 have the highest eigenvalues while PC3 and PC4 are close to zero, thus indicating that only the first two PCs are relevant for explaining the data.

Non-denaturing polyacrylamide gel electrophoresis (PAGE)

Non-denaturing 12% PAGE gels were prepared with 29 : 1 acrylamide/bisacrylamide solution and TAE buffer (40 mM Tris base, 20 mM acetic acid, and 2 mM EDTA, adjusted to pH 5.5, 5.0, or 4.5).[35] Gels were run on a Mini-PROTEAN electrophoresis system (Bio-Rad, CA, USA) at 4 °C, 4 mA, for 4–5 h. Appropriate DNA aliquots from CD or DSC samples were properly diluted to have a final amount of 0.1 mg mL^{-1} DNA in each lane (the total DNA amount in different lanes was kept the same). The buffer in DNA samples was 110 mM (Na^+ or K^+) acetate at the appropriate pH (5.5, 5.0, or 4.5). A solution of glycerol/TAE (2 : 1) was added to facilitate sample loading into the wells. The migration of C-rich DNA was compared with a mixture of two control oligodeoxythymidylates (T_{10} and T_{20}). Bromophenol blue was used as run marker. Bands in the gels were visualized by UV shadowing.

Differential scanning calorimetry (DSC)

DSC measurements were carried out on a Nano DSC (TA Instruments, USA). The experiments were performed using an oligonucleotide strand concentration of 320–500 μM. Samples were prepared in 110 mM Na^+- or K^+-containing acetate buffer at pH 4.5, 5.0 and 5.5. Scans were performed at 0.5 °C min^{-1} in the 20–90 °C temperature range. Reversibility was evaluated for each sample by rescanning it after cooling. A buffer–buffer scan was subtracted from the corresponding buffer–sample scan, and a polynomial baseline was drawn for each scan. Baseline-corrected thermograms were then normalized per mole of DNA to obtain the corresponding excess molar heat capacity ($\langle\Delta C_p\rangle$) as a function of temperature. ΔC_p°, the molar heat capacity change between the folded and unfolded states of DNA, was estimated from DSC thermograms by a second-order polynomial fitting and assumed to be temperature independent. The model-independent transition enthalpies ($\Delta_{cal}H^\circ$) were obtained by integrating the area under the excess heat capacity *versus* temperature curves. T_m values were determined as the temperatures corresponding to the maximum of each thermogram peak. Entropy change ($\Delta_{cal}S^\circ$) values were obtained by integrating the $\langle\Delta C_p\rangle/T$ *versus T* curves (where T is the temperature in Kelvin), and the Gibbs free energy change ($\Delta_{cal}G^\circ$) values were computed by the equation $\Delta_{cal}G^\circ = \Delta_{cal}H^\circ[1-(T/T_m)] + \Delta C_p^\circ[T - T_m - T\ln(T/T_m)]$. The reported thermodynamic parameters are the averages of three different heating experiments.

Results and discussion

Presence of a folded oligonucleotide conformation under the solution conditions employed in this study and its stability as a function of pH and cation were examined by CD spectroscopy, DSC, and non-denaturing polyacrylamide gel electrophoresis (PAGE). Based on the above evaluation, we selected pH 4.5, 5.0, and 5.5, where the folded state was expected to be predominant, to study the iM thermodynamic stability. It's to be noted that CD and DSC methodologies work on different concentration scales. Indeed, DSC experiments require a DNA concentration at least twenty times higher than that of CD measurements. This means that the comparison of the melting temperature (T_m) and enthalpy change values obtained by DSC with those obtained by CD should provide information on the molecularity and mechanism of the iM folding/unfolding process. Indeed, intramolecular structures should be formed when the T_m is independent of oligonucleotide concentration, while dimer or higher molecularities are found when the T_m depends on oligonucleotide concentration. Moreover, being a spectroscopic method, CD allows to indirectly determine the enthalpy change by the van't Hoff equation assuming a two-states process, whereas DSC directly measures the folding/unfolding enthalpy (it is obtained from the area of the DSC thermogram) without any model assumption.[36] The difference between calorimetric and van't Hoff enthalpy changes is an indication of the deviation of the real process from a two-state transition.

CD spectra of the C-rich oligonucleotide confirmed the formation of iM at room temperature in both the Na^+- and K^+-containing buffers at the three pH values. Indeed, in all cases the characteristic CD profile of an iM structure having a positive and a negative band at 288 and 257 nm, respectively, is clearly distinguishable (Fig. 2A–C and Fig. S1, ESI†).[37]

The stability of the iM structure under the chosen experimental conditions was initially investigated by CD melting experiments. The CD denaturation/renaturation profiles of the oligonucleotide, recorded at the wavelength of maximum absorbance variation upon folding (λ = 288 nm), shows almost superimposable heating and cooling curves (Fig. 2D–F and Fig. S2, ESI†), thus indicating that the melting processes are reversible. The analysis of the melting curves yielded the T_m values and model-dependent van't Hoff enthalpies ($\Delta_{vH}H^\circ$) that are collected in Table 1. As expected, the thermal stability of iM decreased as pH increased. Indeed, regardless of the cation, the T_m values for the iM structure were found to be around 60, 55, and 44 °C at pH 4.5, 5.0 and 5.5, respectively. Therefore, the T_m was maximum at the pH value close to the pK_a of cytosine

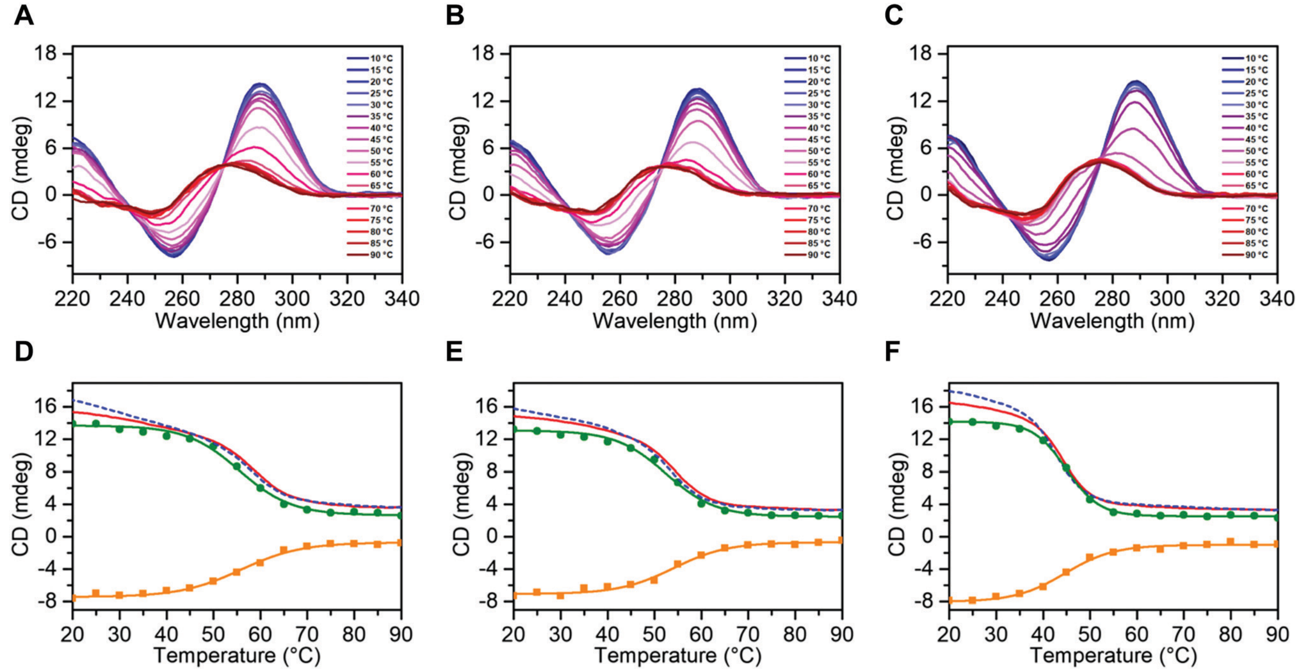


Fig. 2 CD spectra for the iM-forming oligonucleotide as a function of temperature in K^+-containing buffer at pH (A) 4.5, (B) 5.0, and (C) 5.5. (D–F) CD melting (red solid lines) and annealing (blue dashed lines) curves recorded at 288 nm with a scan rate of 0.5 °C min^{-1}, and CD melting curves at 288 (green lines/circles) and 257 (orange lines/squares) nm derived from the corresponding CD spectra at different temperatures (collected every 5 °C) for the iM in K^+-containing buffer at pH (D) 4.5, (E) 5.0, and (F) 5.5. All experiments were performed in acetate buffer.

Table 1 Thermodynamic parameters of $d(CCCTAA)_3CCCT$ iM determined by the van't Hoff analysis of CD melting curves

Cation	pH	T_m[a] (°C)	$\Delta_{vH}H^\circ$[a] (kJ mol^{-1})
K^+	4.5	60	−188
	5.0	54	−220
	5.5	44	−274
Na^+	4.5	59	−150
	5.0	55	−203
	5.5	44	−262

[a] The error in T_m does not exceed 1 °C, while that in $\Delta_{vH}H^\circ$ amounts to 10% of reported values.

(between 4.2 and 4.8, depending on temperature and ionic strength),[38] whereas an increase of 1 pH unit led to a decrease of the T_m of about 16 °C. The lower thermal stability of the iM structure can be ascribed to the lower extent of protonation of the cytosines at higher pH. Conversely, $\Delta_{vH}H^\circ$ appears to increase with increasing pH.

Three-dimensional melting curves for the iM were also obtained by recording whole CD spectra as a function of temperature using a step of 5 °C (Fig. 2 and Fig. S1, S2, ESI†). Thermodynamic parameters determined from the analysis of 3D melting curves are in perfect agreement with those obtained at a single wavelength at 0.5 °C min^{-1} heating rate. The T_m derived from the 3D melting curves at 288 and 257 nm (Fig. 2D–F and Fig. S2, ESI†) were identical within experimental error.

Overall, melting and annealing CD experiments showed no significant hysteresis for the oligonucleotide heating/cooling processes and no influence of the scan rate on the T_m of the iM-forming sequence, thus indicating that the folding–unfolding process for the iM is not kinetically controlled, *i.e.*, the thermodynamic equilibrium is achieved during the temperature changes.

In order to test the two-state transition assumption, the 3D melting curves were submitted to Principal Component Analysis (PCA), a decomposition method usually based on singular value decomposition (SVD). SVD has been extensively employed to decompose CD spectra of non-canonical DNA structures into their structural constituents, in particular it is a well-established method for the analysis of thermal denaturation processes and folding (pH- or salt-induced) kinetics.[39–41]

The PCA decomposition performed on the 3D melting data of each of the six samples, resulted in two principal components that fully explain (about 99%) the variance in the data. By way of example, we report the PCA model computed using the 3D melting curves of the iM-forming oligonucleotide in K^+-containing buffer at pH 4.5 (Fig. 3). The score plot (Fig. 3A) clearly shows that the first component (PC1), which explains the 96.29% of the total variance, reflects the shape of the folded iM, while the PC2 (3.59% of explained variance) corresponds to the spectrum of the unfolded species. The PC3 and PC4 (respectively, 0.07% and 0.01% of explained variance) scores are close to the zero, thus they are essentially meaningless. The amount by which each of these components is present in the spectra is shown in the loading plot (Fig. 3B) as a function of the temperature. Thus, we can infer that (i) only two species are in equilibrium in solution, and that (ii) the folded iM structure is mainly present at temperature values

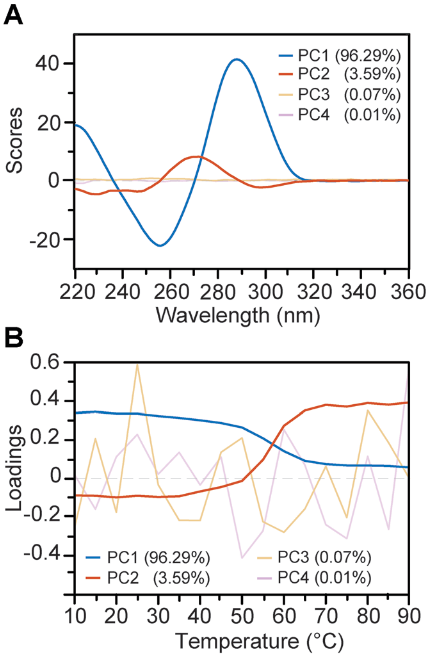


Fig. 3 (A) Score and (B) loadings plots of the four principal components (PC1–PC4) obtained from the PCA model computed using the CD spectra of the iM-forming oligonucleotide as a function of temperature in K^+-containing buffer at pH 4.5.

lower than T_m, whereas the unfolded species becomes predominant upon temperature increase. This suggests that, under these experimental conditions, the iM to single strand transition is two-state and involves only two significant spectral species. The remaining PCA models generated by the 3D melting data of iM-forming oligonucleotide in K^+-containing buffer at pH 5.0 and 5.5 are reported in Fig. S3 (ESI†), while the ones referred to the Na^+-containing buffer are showed in Fig. S4 (ESI†).

To further exclude the presence of a conformational heterogeneity involving equilibrium between monomer and dimer conformations, we performed native PAGE experiments at pH 4.5, 5.0 and 5.5 (Fig. 4 and Fig. S5, ESI†). The experiments were conducted starting from iM samples at two different DNA concentrations, that are those used for CD and DSC measurements (about 20 and 400 μM, respectively). The electrophoretic mobility of the C-rich oligonucleotide was compared with T_{10} and T_{20} oligodeoxythymidylates (lane 2). All the samples (lanes 3–6) migrated as a single band, thus excluding the formation of species with higher molecularity.

DSC melting experiments were carried out to obtain the complete model-independent thermodynamic characterization of the iM stability. The DSC profiles for the investigated DNA are shown in Fig. 5, and the corresponding thermodynamic parameters for iM formation are listed in Table 2.

Under these experimental conditions, the transitions of iM were confirmed to be reversible, as demonstrated by the recovery of the original DSC profile by repeating the thermal scan of each sample after cooling. The DSC curves showed an almost symmetric shape with maxima centered at T_m values very close to those obtained by CD melting. Since the DNA concentrations used in the DSC experiments were approximately 20-fold higher than those used in the CD melting experiments, the melting temperatures can be considered concentration-independent, which is consistent with a structure resulting from unimolecular folding. Moreover, the integration of the DSC peaks gave $\Delta_{cal}H^\circ$ values in good agreement with $\Delta_{vH}H^\circ$ calculated from CD data. The close correspondence between the calorimetric and van't Hoff enthalpies is once again consistent with a two-state equilibrium in which intermediate states are not significantly populated and indicates that no aggregation phenomenon occurs during the transition. Therefore, in the range investigated here, the pH seems to have no impact on the kinetics of intramolecular folding. On the other hand, it has been previously reported that at neutral pH the C-rich telomeric sequence does not melt in a reversible fashion, indicative of slow folding kinetics at pH close to 7.0.[42]

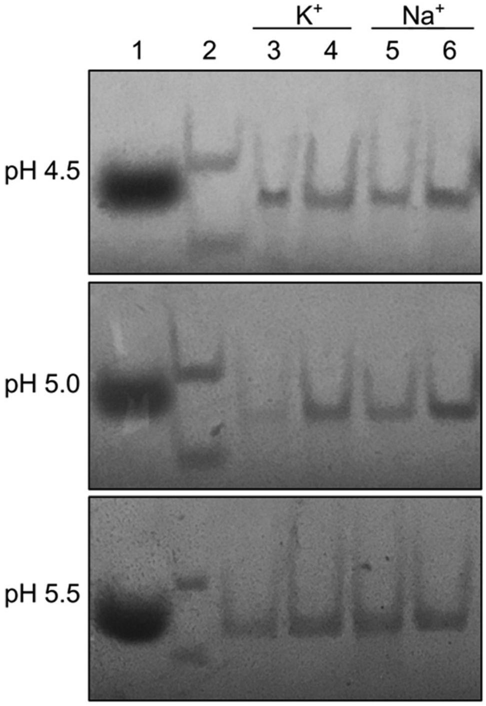


Fig. 4 Non-denaturing polyacrylamide gel electrophoresis in K^+- or Na^+-containing acetate buffer at pH 4.5, 5.0, and 5.5, 4 °C. Lane 1: bromophenol blue (run marker). Lane 2: mixture of T_{10} and T_{20} oligodeoxynucleotides (size markers). Lanes 3–6: d(CCCTAA)$_3$CCCT iM-forming oligonucleotide from CD (3,5) or DSC (4,6) samples. All DNA samples were diluted to the final desired strand concentration (0.1 mg mL^{-1}) before loading onto the gel.

The thermodynamic parameters were calculated taking into account the relatively small heat capacity changes associated with DNA (un)folding $\left(\Delta C_p^\circ\right)$. Indeed, base dehydration, among other compensating contributions, is now well accepted to be a source of negative heat capacity change in the folding of canonical and non-canonical DNA structures.[43–45] The folding ΔC_p° values for the iM were found in the range −2.0 to −3.7 kJ mol^{-1} K^{-1},[46] however, due to the high uncertainties associated with these values, we considered the average value of −2.6 kJ mol^{-1} K^{-1}.

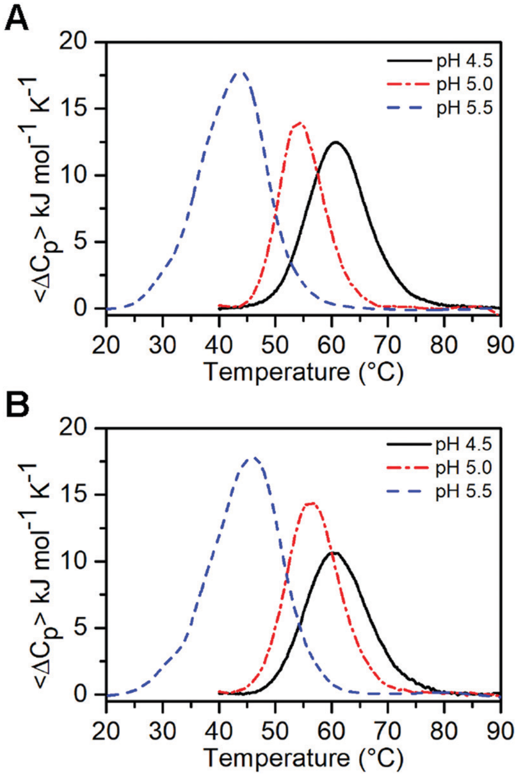


Fig. 5 DSC profiles for the iM-forming oligonucleotide at 0.5 °C min^{-1} heating rate in (A) K$^+$- or (B) Na$^+$-containing solutions at pH 4.5 (black solid line), 5.0 (red dash-dotted line), and 5.5 (blue dashed line). All experiments were performed in acetate buffer.

Table 2 Thermodynamic parameters of iM-forming oligonucleotide d(CCCTAA)$_3$CCCT determined by DSC

Cation	pH	T_m ^{a} (°C)	$\Delta_{cal}H^\circ$ ^{a} (kJ mol^{-1})	$\Delta_{cal}S^\circ$ ^{a} (kJ mol^{-1} K^{-1})	$\Delta_{cal}G^\circ_{25\,^\circ C}$ ^{a} (kJ mol^{-1})	n_{CH^+}
K$^+$	4.5	60.6	−158.3	−0.474	−11.8	—
	5.0	53.9	−224.1	−0.685	−16.4	3.1
	5.5	43.5	−261.9	−0.825	−13.9	5.0
Na$^+$	4.5	60.8	−145.2	−0.435	−10.4	—
	5.0	55.6	−195.5	−0.594	−14.4	2.4
	5.5	45.8	−273.4	−0.857	−16.0	6.1

[a] The error in T_m does not exceed 0.5 °C, while that in $\Delta_{cal}H^\circ$ and $\Delta_{cal}S^\circ$ amounts to 5%, and that in $\Delta_{cal}G^\circ$ to 10% of reported values.

Inspection of Table 2 shows that the iM folding is accompanied by a favorable Gibbs free energy change ($\Delta_{cal}G^\circ$) resulting from the relatively large compensation of favorable enthalpy ($\Delta_{cal}H^\circ$) and unfavorable entropy ($\Delta_{cal}S^\circ$) contributions. The favorable enthalpy terms correspond primarily to the formation of the hemi-protonated C·C$^+$ base pairs, while the unfavorable entropy terms mainly arise from contributions of the unfavorable ordering of oligonucleotide and the uptake of protons, and from the putative marginal uptake of counterions and/or water molecules.

It is interesting to compare the thermodynamic parameters of the iM-forming oligonucleotide at pH 4.5, 5.0, and 5.5. As a matter of fact, the calculated $\Delta_{cal}H^\circ$ and $\Delta_{cal}S^\circ$ were very different. At the lowest pH, the enthalpy of folding was less negative, and thus less favorable. This was in part compensated by a less negative, and thus more favorable, entropy of formation. To explain the enthalpy and entropy dependence on pH, we invoke differences in oligonucleotide protonation on changing the pH. Indeed, unfavorable entropy at increased pH is most probably due to higher proton uptake at those pH values. Moreover, protonation of the C·C$^+$ base pairs makes the observed enthalpy changes dependent on the pH as well as on the buffer used. In particular, the enthalpy changes relative to the protonation of the buffer species ($\Delta_{buf}H^\circ$) influences the thermodynamic parameters. To avoid this, we have used acetate buffer, whose $\Delta_{buf}H^\circ$ is close to zero. Therefore, the pH of this buffer is not temperature dependent, making it appropriate for melting experiments.

At pH = pK_a of cytosine (around 4.5), half of the cytosines on the oligonucleotide (six in this case) are protonated, which correspond to the optimum number for the formation of C·C$^+$ base pairs, and no further protonation is required (even though the distribution of protonated cytosines varies). In that case, being the $\Delta_{buf}H^\circ$ negligible, the measured $\Delta_{cal}H^\circ$ correspond to the enthalpy of iM formation ($\Delta_{iM}H^\circ$), which is $\Delta_{cal}H^\circ = \Delta_{iM}H^\circ$. On the other hand, at pH > pK_a, protonation of some of the cytosines is required. In that case, the measured $\Delta_{cal}H^\circ$ is the result of the sum of two contributions: $\Delta_{iM}H^\circ$ and the enthalpy of cytosine protonation ($\Delta_{CH^+}H^\circ$), which is $\Delta_{cal}H^\circ = n_{CH^+}\Delta_{CH^+}H^\circ + \Delta_{iM}H^\circ$ (where n_{CH^+} is the number of protonated cytosines).

From the difference between the $\Delta_{cal}H^\circ$ values at pH = pK_a and those at pH 5.0 and 5.5, it is possible to determine

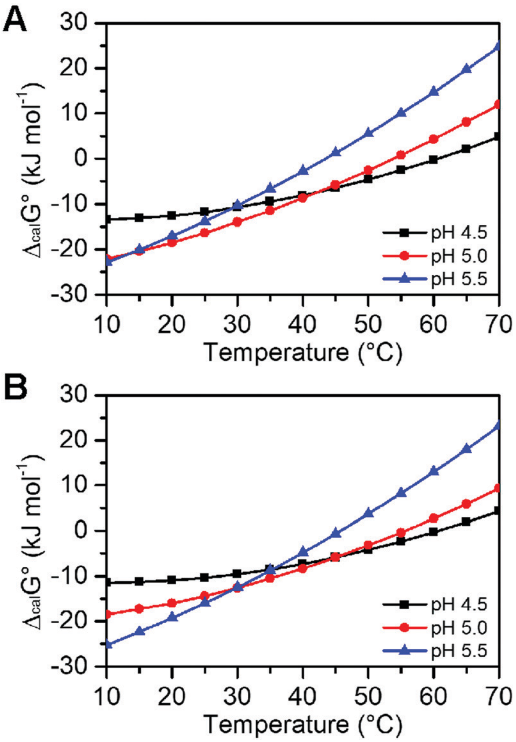


Fig. 6 Temperature dependence of the Gibbs energy change ($\Delta_{cal}G^\circ$) for iM formation in (A) K$^+$- or (B) Na$^+$-containing solutions at pH 4.5 (black line), 5.0 (red line), and 5.5 (blue lines).

$n_{CH^+}\Delta_{CH^+}H^\circ$. Mergny and coworkers calculated a value of -20.9 kJ mol^{-1} for $\Delta_{CH^+}H^\circ$.[38] Knowing this value and the other terms, we calculated the number of cytosines to be protonated for the formation of the intramolecular iM at pH 5.0 and 5.5 (n_{CH^+}, Table 2). This number was estimated to be 2–3 at pH 5.0 and 5–6 at pH 5.5.

A closer inspection of Table 2 shows that at 25 °C the enthalpy–entropy compensation results in a slightly more favorable $\Delta_{cal}G^\circ$ value at pH 5.0 or 5.5 than at pH 4.5. Thus, the temperature dependence of the Gibbs free energy change ($\Delta_{cal}G^\circ$) for iM formation was evaluated. The profile of the $\Delta_{cal}G^\circ$ *vs.* temperature in K^+- or Na^+-containing buffer at the three investigated pH are shown in Fig. 6. The overall trend in the $\Delta_{cal}G^\circ$ as a function of temperature in the presence of K^+ or Na^+ ions is similar. Interestingly, despite the small differences observed, at temperatures below 40 °C the $\Delta_{cal}G^\circ$ values for iM formation at pH 4.5 are less negative than at pH 5.0 or 5.5, thus indicating that, at such temperatures, the iM is favored at the higher pH values investigated here. This result shows that, despite thermal stabilities are often used as surrogates for thermodynamic stabilities in analyses of the iM stability, the correlation between T_m and $\Delta_{cal}G^\circ$ is not always obvious. Where stability comparisons are to be made, the relevant thermodynamic parameter is the Gibbs energy change at a chosen reference temperature.

Conclusions

We have investigated the energetics of an iM DNA structure in solution under mildly acidic conditions that promote its formation. We utilized CD spectroscopy and DSC to study the iM stability at multiple pH values in a buffer containing K^+ or Na^+ cations. Folding and unfolding processes of the iM followed relatively fast kinetics, as expected for intramolecular folding. In agreement with this observation, analysis of CD and DSC melting profiles revealed identical melting temperatures for samples over a range of DNA concentrations, and native PAGE showed only a single band corresponding to monomeric species. Since the melting profiles corresponded to equilibrium curves (*i.e.*, to reversible processes), we were able to provide a full thermodynamic characterization of iM formation *in vitro* under favorable conditions, thus broadening our understanding about the formation and thermodynamic stability of iM DNA structures. Our results also show that thermodynamic and thermal stability are related but not identical properties of an iM and underscore the risk in using thermal stability data in quantitative analysis of the iM folding stability. This study should be useful not only for those interested in the biological functions of this structure, but also for those seeking to exploit the distinct properties of iMs for the development of pH-sensitive molecular devices based on such structures.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

The research leading to these results has received funding from AIRC under IG 2020 – ID. 24590 project – to B. P.

References

1 J. Choi and T. Majima, Conformational changes of non-B DNA, *Chem. Soc. Rev.*, 2011, **40**, 5893–5909.

2 M. Gajarský, M. L. Živković, P. Stadlbauer, B. Pagano, R. Fiala, J. Amato, L. Tomáška, J. Šponer, J. Plavec and L. Trantírek, Structure of a Stable G-Hairpin, *J. Am. Chem. Soc.*, 2017, **139**, 3591–3594.

3 B. Pagano, C. A. Mattia, L. Cavallo, S. Uesugi, C. Giancola and F. Fraternali, Stability and Cations Coordination of DNA and RNA 14-Mer G-Quadruplexes: A Multiscale Computational Approach, *J. Phys. Chem. B*, 2008, **112**, 12115–12123.

4 H. Tateishi-Karimata and N. Sugimoto, Chemical biology of non-canonical structures of nucleic acids for therapeutic applications, *Chem. Commun.*, 2020, **56**, 2379–2390.

5 J. Amato, N. Iaccarino, A. Randazzo, E. Novellino and B. Pagano, Noncanonical DNA Secondary Structures as Drug Targets: the Prospect of the i-Motif, *ChemMedChem*, 2014, **9**, 2026–2030.

6 A. Pagano, N. Iaccarino, M. A. S. Abdelhamid, D. Brancaccio, E. U. Garzarella, A. Di Porzio, E. Novellino, Z. A. E. Waller, B. Pagano, J. Amato and A. Randazzo, Common G-Quadruplex Binding Agents Found to Interact With i-Motif-Forming DNA: Unexpected Multi-Target-Directed Compounds, *Front. Chem.*, 2018, **6**, 281.

7 C. Teller and I. Willner, *Curr. Opin. Biotechnol*, 2010, **21**, 376–391.

8 Y. Krishnan and F. C. Simmel, Nucleic Acid Based Molecular Devices, *Angew. Chem., Int. Ed.*, 2011, **50**, 3124–3156.

9 H. Abou Assi, M. Garavís, C. González and M. J. Damha, i-Motif DNA: structural features and significance to cell biology, *Nucleic Acids Res.*, 2018, **46**, 8038–8056.

10 K. Gehring, J.-L. Leroy and M. Guéron, A tetrameric DNA structure with protonated cytosine-cytosine base pairs, *Nature*, 1993, **363**, 561–565.

11 M. Zeraati, D. B. Langley, P. Schofield, A. L. Moye, R. Rouet, W. E. Hughes, T. M. Bryan, M. E. Dinger and D. Christ, I-motif DNA structures are formed in the nuclei of human cells, *Nat. Chem.*, 2018, **10**, 631–637.

12 W. Tang, K. Niu, G. Yu, Y. Jin, X. Zhang, Y. Peng, S. Chen, H. Deng, S. Li, J. Wang, Q. Song and Q. Feng, In vivo visualization of the i-motif DNA secondary structure in the Bombyx mori testis, *Epigenet. Chromatin*, 2020, **13**, 12.

13 S. Benabou, A. Aviñó, R. Eritja, C. González and R. Gargallo, Fundamental aspects of the nucleic acid i-motif structures, *RSC Adv.*, 2014, **4**, 26956–26980.

14 N. Iaccarino, A. Di Porzio, J. Amato, B. Pagano, D. Brancaccio, E. Novellino, R. Leardi and A. Randazzo, Assessing the influence of pH and cationic strength on i-motif DNA structure, *Anal. Bioanal. Chem.*, 2019, **411**, 7473–7479.

15 Y. Dong, Z. Yang and D. Liu, DNA nanotechnology based on i-motif structures, *Acc. Chem. Res.*, 2014, **47**, 1853–1860.

16 M. Debnath, K. Fatma and J. Dash, Chemical Regulation of DNA i-Motifs for Nanobiotechnology and Therapeutics, *Angew. Chem., Int. Ed.*, 2019, **58**, 2942–2957.

17 J. J. Alba, A. Sadurní and R. Gargallo, Nucleic Acid i-Motif Structures in Analytical Chemistry, *Crit. Rev. Anal. Chem.*, 2016, **46**, 443–454.

18 B. A. Webb, M. Chimenti, M. P. Jacobson and D. L. Barber, Dysregulated pH: a perfect storm for cancer progression, *Nat. Rev. Cancer*, 2011, **11**, 671–677.

19 S. P. Gurung, C. Schwarz, J. P. Hall, C. J. Cardin and J. A. Brazier, The importance of loop length on the stability of i-motif structures, *Chem. Commun.*, 2015, **51**, 5630–5632.

20 S. Benabou, M. Garavís, S. Lyonnais, R. Eritja, C. González and R. Gargallo, Understanding the effect of the nature of the nucleobase in the loops on the stability of the i-motif structure, *Phys. Chem. Chem. Phys.*, 2016, **18**, 7997–8004.

21 M. McKim, A. Buxton, C. Johnson, A. Metz and R. D. Sheardy, Loop Sequence Context Influences the Formation and Stability of the i-Motif for DNA Oligomers of Sequence (CCCXXX)4, where X = A and/or T, under Slightly Acidic Conditions, *J. Phys. Chem. B*, 2016, **120**, 7652–7661.

22 A. M. Fleming, Y. Ding, R. A. Rogers, J. Zhu, J. Zhu, A. D. Burton, C. B. Carlisle and C. J. Burrows, 4n-1 Is a "Sweet Spot" in DNA i-Motif Folding of 2′-Deoxycytidine Homopolymers, *J. Am. Chem. Soc.*, 2017, **139**, 4682–4689.

23 E. P. Wright, J. L. Huppert and Z. A. E. Waller, Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH, *Nucleic Acids Res.*, 2017, **45**, 2951–2959.

24 M. Cheng, D. Qiu, L. Tamon, E. Ištvánková, P. Víšková, S. Amrane, A. Guédin, J. Chen, L. Lacroix, H. Ju, L. Trantírek, A. B. Sahakyan, J. Zhou and J.-L. Mergny, Thermal and pH stabilities of i-DNA: confronting *in vitro* experiments with models and in-cell NMR data, *Angew. Chem., Int. Ed.*, 2021, **6**, 10286–10294.

25 T. Nguyen, C. Fraire and R. D. Sheardy, Linking pH, Temperature, and K+ Concentration for DNA i-Motif Formation, *J. Phys. Chem. B*, 2017, **121**, 7872–7877.

26 S. E. Kim, I.-B. Lee, C. Hyeon and S.-C. Hong, Destabilization of i-Motif by Submolar Concentrations of a Monovalent Cation, *J. Phys. Chem. B*, 2014, **118**, 4753–4760.

27 A. Garabedian, D. Butcher, J. L. Lippens, J. Miksovska, P. P. Chapagain, D. Fabris, M. E. Ridgeway, M. A. Park and F. Fernandez-Lima, Structures of the kinetically trapped i-motif DNA intermediates, *Phys. Chem. Chem. Phys.*, 2016, **18**, 26691–26702.

28 J. Somkuti and O. R. Molnár, and L. Smeller, Revealing unfolding steps and volume changes of human telomeric i-motif DNA, *Phys. Chem. Chem. Phys.*, 2020, **22**, 23816–23823.

29 A. T. Phan, M. Guéron and J.-L. Leroy, The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere, *J. Mol. Biol.*, 2000, **299**, 123–144.

30 A. L. Lieblein, J. Buck, K. Schlepckow, B. Fürtig and H. Schwalbe, Time-Resolved NMR Spectroscopic Studies of DNA i-Motif Folding Reveal Kinetic Partitioning, *Angew. Chem., Int. Ed.*, 2012, **51**, 250–253.

31 S. Wu, X. Wang, X. Ye and G. Zhang, pH-Induced Conformational Change and Dimerization of DNA Chains Investigated by Analytical Ultracentrifugation, *J. Phys. Chem. B*, 2013, **117**, 11541–11547.

32 C. R. Cantor, M. M. Warshaw and H. Shapiro, Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleolides, *Biopolymers*, 1970, **9**, 1059–1077.

33 L. A. Marky and K. J. Breslauer, Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves, *Biopolymers*, 1987, **26**, 1601–1620.

34 S. Wold, K. Esbensen and P. Geladi, Principal component analysis, *Chemom. Intell. Lab. Syst.*, 1987, **2**, 37–52.

35 W. Fu, L. Tang, G. Wei, L. Fang, J. Zeng, R. Zhan, X. Liu, H. Zuo, C. Z. Huang and C. Mao, Rational Design of pH-Responsive DNA Motifs with General Sequence Compatibility, *Angew. Chem., Int. Ed.*, 2019, **58**, 16405–16410.

36 B. Pagano, A. Randazzo, I. Fotticchia, E. Novellino, L. Petraccone and C. Giancola, Differential scanning calorimetry to investigate G-quadruplexes structural stability, *Methods*, 2013, **64**, 43–51.

37 N. Iaccarino, M. Cheng, D. Qiu, B. Pagano, J. Amato, A. Di Porzio, J. Zhou, A. Randazzo and J.-L. Mergny, Effects of Sequence and Base Composition on the CD and TDS Profiles of i-DNA, *Angew. Chem., Int. Ed.*, 2021, **60**, 10295–10303.

38 J.-L. Mergny, L. Lacroix, X. Han, J.-L. Leroy and C. Helene, Intramolecular Folding of Pyrimidine Oligodeoxynucleotides into an i-DNA Motif, *J. Am. Chem. Soc.*, 1995, **117**, 8887–8898.

39 R. D. Gray and J. B. Chaires, Analysis of Multidimensional G-Quadruplex Melting Curves, *Curr. Protoc. Nucleic Acid Chem.*, 2011, **45**, Unit17.4.

40 R. D. Gray, J. O. Trent, S. Arumugam and J. B. Chaires, Folding Landscape of a Parallel G-Quadruplex, *J. Phys. Chem. Lett.*, 2019, **10**, 1146–1151.

41 L. Petraccone, B. Pagano, V. Esposito, A. Randazzo, G. Piccialli, G. Barone, C. A. Mattia and C. Giancola, Thermodynamics and kinetics of PNA-DNA quadruplex-forming chimeras, *J. Am. Chem. Soc.*, 2005, **127**, 16215–16223.

42 J.-L. Leroy, M. Guéron, J.-L. Mergny and C. Hélène, Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif, *Nucleic Acids Res.*, 1994, **22**, 1600–1606.

43 S. Hadži and J. Lah, Origin of heat capacity increment in DNA folding: The hydration effect, *Biochim. Biophys. Acta, Gen. Subj.*, 2021, **1865**, 129774.

44 S. Hadži, M. Bončina and J. Lah, *Methods in Molecular Biology*, Humana Press Inc., 2019, vol. 2035, pp. 117–130.

45 S. Hadži, V. Kocman, D. Oblak, J. Plavec and J. Lah, Energetic Basis of AGCGA-Rich DNA Folding into a Tetrahelical Structure, *Angew. Chem., Int. Ed.*, 2019, **58**, 2387–2391.

46 H. A. Assi, R. W. Harkness, N. Martin-Pintado, C. J. Wilds, R. Campos-Olivas, A. K. Mittermaier, C. González and M. J. Damha, Stabilization of i-motif structures by 2′-β-fluorination of DNA, *Nucleic Acids Res.*, 2016, **44**, 4998–5009.

PAPER II

Targeting of telomeric repeat-containing RNA G-quadruplexes: from screening to biophysical and biological characterization of a new hit compound

Marzano S., Pagano B., Iaccarino N., Di Porzio A., De Tito S.,
Vertecchi E., Salvati E., Randazzo A., Amato J.

International Journal of Molecular Sciences, 2021, Vol. 22(19)

International Journal of
Molecular Sciences

MDPI

Article

Targeting of Telomeric Repeat-Containing RNA G-Quadruplexes: From Screening to Biophysical and Biological Characterization of a New Hit Compound

Simona Marzano [1], Bruno Pagano [1], Nunzia Iaccarino [1], Anna Di Porzio [1], Stefano De Tito [2,3], Eleonora Vertecchi [4], Erica Salvati [4], Antonio Randazzo [1] and Jussara Amato [1,*]

[1] Department of Pharmacy, University of Naples Federico II, Via D. Montesano 49, 80131 Naples, Italy; simona.marzano@unina.it (S.M.); bruno.pagano@unina.it (B.P.); nunzia.iaccarino@unina.it (N.I.); anna.diporzio@unina.it (A.D.P.); antonio.randazzo@unina.it (A.R.)
[2] Molecular Cell Biology of Autophagy, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK; stefano.de-tito@crick.ac.uk
[3] Institute of Experimental Endocrinology and Oncology, National Research Council, 80131 Naples, Italy
[4] c/o Department of Biology and Biotechnology "C. Darwin", Institute of Molecular Biology and Pathology, National Research Council, Sapienza University of Rome, Via degli Apuli 4, 00185 Rome, Italy; eleonora.vertecchi@uniroma1.it (E.V.); erica.salvati@cnr.it (E.S.)
* Correspondence: jussara.amato@unina.it; Tel.: +39-081678630

Citation: Marzano, S.; Pagano, B.; Iaccarino, N.; Di Porzio, A.; De Tito, S.; Vertecchi, E.; Salvati, E.; Randazzo, A.; Amato, J. Targeting of Telomeric Repeat-Containing RNA G-Quadruplexes: From Screening to Biophysical and Biological Characterization of a New Hit Compound. *Int. J. Mol. Sci.* **2021**, *22*, 10315. https://doi.org/10.3390/ijms221910315

Academic Editors: Manlio Palumbo and Claudia Sissi

Received: 16 August 2021
Accepted: 22 September 2021
Published: 24 September 2021

Abstract: DNA G-quadruplex (G4) structures, either within gene promoter sequences or at telomeres, have been extensively investigated as potential small-molecule therapeutic targets. However, although G4s forming at the telomeric DNA have been extensively investigated as anticancer targets, few studies focus on the telomeric repeat-containing RNA (TERRA), transcribed from telomeres, as potential pharmacological targets. Here, a virtual screening approach to identify a library of drug-like putative TERRA G4 binders, in tandem with circular dichroism melting assay to study their TERRA G4-stabilizing properties, led to the identification of a new hit compound. The affinity of this compound for TERRA RNA and some DNA G4s was analyzed through several biophysical techniques and its biological activity investigated in terms of antiproliferative effect, DNA damage response (DDR) activation, and TERRA RNA expression in high vs. low TERRA-expressing human cancer cells. The selected hit showed good affinity for TERRA G4 and no binding to double-stranded DNA. In addition, biological assays showed that this compound is endowed with a preferential cytotoxic effect on high TERRA-expressing cells, where it induces a DDR at telomeres, probably by displacing TERRA from telomeres. Our studies demonstrate that the identification of TERRA G4-targeting drugs with potential pharmacological effects is achievable, shedding light on new perspectives aimed at discovering new anticancer agents targeting these G4 structures.

Keywords: TERRA G-quadruplex; drug discovery; biophysical characterization; conformation-selective ligand; in vitro biological assays

1. Introduction

G-quadruplexes (G4s) are higher-order noncanonical nucleic acid structures formed by guanine-rich sequences [1]. These structures exhibit a common stem arrangement of stacked G-tetrads, where four guanine bases associate through Hoogsteen-type hydrogen bonding and their oxygen O6 atoms are arranged to coordinate metal cations, such as potassium or sodium, to give stability to the whole structure [1]. G4s can be intramolecular, i.e., formed by a single nucleic acid molecule, or intermolecular, i.e., formed by two or four strands [1]. Generally, potential intramolecular G4s have a consensus sequence of $G_{\geq 3}N_{1\text{–}7}G_{\geq 3}N_{1\text{–}7}G_{\geq 3}N_{1\text{–}7}G_{\geq 3}$, where N is any nucleotide, even if some non-consensus sequences were reported to fold into stable G4s [2,3]. Sequences between G-tracts are variable and, depending on their length and composition, determine different loop conformation and the overall G4 topology, which can be parallel, antiparallel, or hybrid [4,5].

Accumulating evidence shows that DNA and RNA G4 structures are formed in vivo [6–13], and play pivotal roles in regulating DNA transcription and replication, RNA translation, and the maintenance of genome integrity [14–18]. Indeed, G4-forming sequences are particularly enriched in proto-oncogene promoters [19], origins of replication [20], 5′- and 3′-untranslated regions (UTRs) of mRNA of a large number of genes [21], and at the ends of human chromosomes, the telomeres [22,23].

Telomeres are specialized nucleoprotein structures that protect chromosomal DNA from progressive degradation and ensure the integrity of linear chromosomes by preventing the natural ends from being recognized as DNA damage [22,23]. Human telomeres terminate with a 3′ single-stranded DNA overhang, which is composed of tandem repeats of the short guanine-rich sequence d(TTAGGG) synthesized by telomerase, a telomere-specific reverse transcriptase. In normal human cells, telomeres progressively shorten, leading to growth arrest upon telomere uncapping known as replicative ageing. Conversely, telomerase is overactive in numerous cancer subtypes, contributing to the ability of these cells to indefinitely proliferate due to the lack of chromosomal shortening [24]. Mounting evidence shows that G4s formation and stabilization at telomeres inhibits telomerase activity and induces DNA damage response, leading to chromosomal shortening and cell death [25,26]. Therefore, designing small molecules able to bind and stabilize telomeric G4s represents a potential avenue for developing novel selective anticancer agents [27–29].

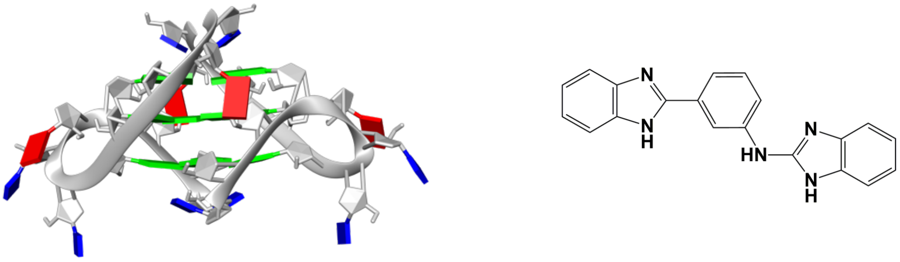
At telomeres, the 3′ single-stranded overhang is preceded by a double-stranded DNA region that is composed of the repeated sequences d(TTAGGG) on one strand and the complementary d(CCCTAA) repeats on the other [30]. The transcription of the telomeric C-rich strand in chromosomes produces telomeric repeat-containing RNA (TERRA), which has a canonical G-rich motif of sequence r(UUAGGG) [31–33]. As such, TERRA can fold into G4 structures. Besides regulating telomerase activity and protecting chromosome ends from telomere degradation, TERRA also takes part in heterochromatin formation and homologous recombination [34–36]. TERRA G4 structures are potentially more valuable therapeutic targets than their DNA counterparts, since telomere heterochromatin formation is required in all cancer cells, even in those that do not require telomerase to elongate their telomeres (ALT-positive tumors) [37]. These tumors display higher TERRA accumulation levels that appear to play a direct role in telomere elongation [38]. For this reason, the design of small molecules targeting TERRA G4s is attracting ever-increasing attention.

Structural analysis showed that TERRA RNA and its corresponding DNA sequences, which differ only for the presence of thymine instead of uracil bases and for deoxyribose sugars instead of ribose, can adopt different G4 topologies. For example, the 22-nucleotide-long telomeric DNA of sequence d[AGGG(TTAGGG)$_3$] adopts an antiparallel-stranded G4 conformation in Na^+-containing solution and a hybrid [3+1] (parallel/antiparallel-stranded) conformation in K^+-containing solution [1,39], while corresponding RNA sequence r[AGGG(UUAGGG)$_3$] folds into a parallel-stranded G4 structure in both Na^+ and K^+ solutions [40]. Direct evidence for the presence of parallel-stranded TERRA RNA G4s in living cells was also provided by Xu and coworkers by using a light-switching pyrene probe [41]. Further confirmation of their existence in vivo came from both optical imaging investigations performed with other small molecules [7,42,43] as well as with the BG4 antibody [6], and from sequencing-based methods [44,45].

Several studies suggested that RNA G4s are more compact and thermally stable than their DNA counterparts [46–48] due to the 2′-hydroxyl (2′-OH) group in the ribose sugar and the networks of water-mediated contacts within the grooves of RNA [49]. The presence of the 2′-OH group in the ribose also induces guanine bases to adopt the *anti* conformation in the G4 structure, which can thereby only be parallel [50,51]. In addition, the presence of the 2′-OH groups may interfere with the interaction of ligands with the loops of RNA G4 by reducing their depth and width [52], and/or affecting the π–π stacking surface of the external G-tetrads. Therefore, the presence of 2′-OH groups in RNA G4s could represent an important structural feature to be taken into account in the design of selective RNA G4-targeting ligands [6].

Di Antonio and coworkers showed that selective RNA vs. DNA G4 targeting can be achieved even by introducing small modifications into a generic G4 binder [53]. Indeed, pyridostatin is not able to discriminate between RNA and DNA G4s, while the carboxypyridostatin derivative has high preference for G4 RNA [53].

Herein, we identify new molecular scaffolds able to target *TERRA G4* by employing a strategy of high-throughput in silico screening of a large number of compounds from a commercially available database, followed by validation of the putative hits through a combination of experimental techniques. The selected 103 virtual screening-derived hits were experimentally evaluated for their binding properties by biophysical methodologies, identifying molecule N-[3-1H-1,3-benzodiazol-2-yl)phenyl]-1H-1,3-benzodiazol-2-amine (**BPBA**, Figure 1) as a promising hit compound able to bind and stabilize the G4 structure adopted by TERRA. The biophysical characterization of the binding profile of **BPBA** revealed that this ligand binds to parallel RNA and DNA G4 structures with a slightly higher affinity for the former, whereas it showed no affinity for double-stranded DNA.



(**a**) (**b**)

Figure 1. (**a**) Bimolecular G4 structure formed by TERRA RNA (guanine, green; adenine, red; uracil, blue). (**b**) Chemical structure of compound **BPBA**.

In addition, the investigation of the biological activity of **BPBA** showed that it is particularly effective in high TERRA-expressing human cancer cells by binding and displacing TERRA from telomeres.

2. Results and Discussion

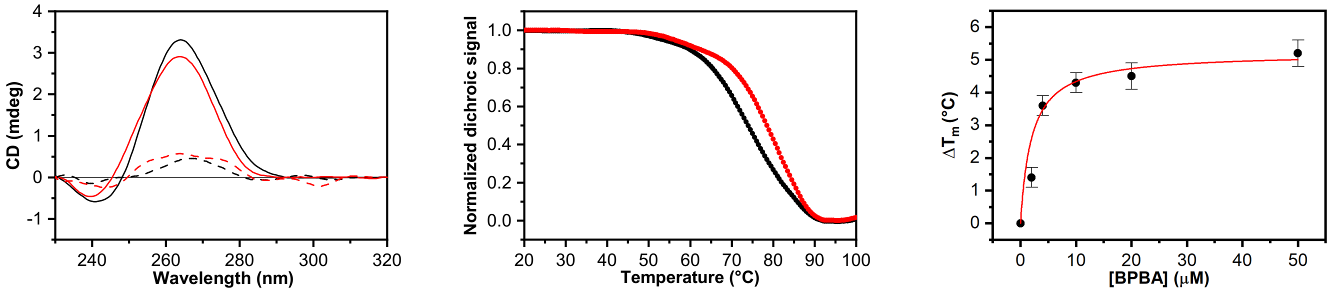
2.1. Compound Selection

Virtual-screening calculations were performed to identify drug-like molecules capable of binding to TERRA, using as target the three-dimensional G4 structures formed by the 12-nt r(UAGGGUUAGGGU) sequence (*TERRA G4*, Figure 1) [50,54]. We employed a receptor-based virtual-screening approach based on the identification of druggable RNA hot spots and molecular docking (see Section 3 for details). To improve the efficacy of the virtual screening, some precalculations to define a potential binding site on the RNA were performed. Since most true binding sites are cavities or regions that provide a large surface for favorable interactions, we looked for an accessible area on the target RNA structures that could form strong polar and nonpolar interactions with a putative ligand. These areas were detected by performing docking calculations with the AutoDock Vina [55] tool embedded in Mcule (https://mcule.com, accessed on 15 July 2017) using a set of small solvent molecules having different polarity [56]. Solvent molecules are used because, thanks to their small size, they can efficiently fit also in buried cavities. Potential hot spots were located where the top binding pose of at least three different molecules were overlapping.

Once RNA hot spot regions that could be targeted by putative ligands were identified, docking calculations, restricted to the identified binding sites, were carried out with AutoDock Vina by using a diverse set of 58870 commercially available compounds as a screening library. This virtual-screening process resulted in 103 drug-like compounds that were selected for further experimental investigations and purchased.

2.2. Circular Dichroism Screening

In order to identify true hits, the 103 computationally selected small molecules were screened for their ability to thermally stabilize *TERRA G4* by using circular dichroism (CD) melting assay. Although CD is not usually used for large-scale screening purposes, it is widely employed to select nucleic acid-interacting compounds with high reliability [57,58]. Indeed, CD experiments require unmodified oligonucleotides, so changes in CD melting curves should only be produced by the direct interaction of the putative ligand with the nucleic acid structure. First, CD spectra were recorded to examine the potential of the selected compounds to eventually alter the dimeric propeller-type parallel conformation of *TERRA G4* in K^+ buffer, whose CD spectrum in the absence of any compound resembles that reported in the literature, exhibiting a positive band at around 265 nm and a negative one at around 245 nm (Figure 2) [59]. RNA/ligand mixtures were obtained by adding putative ligands (10 molar equiv) to the folded *TERRA G4* structure. No significant variations of the CD profile were detected upon the addition of any compound (Figure S1), clearly suggesting they did not modify the parallel conformation adopted by *TERRA G4*. Then, the stabilizing properties of the compounds were evaluated by CD melting experiments measuring the ligand-induced change in the apparent melting temperature (ΔT_m) of the G4 structure. Melting experiments in the absence and presence of each compound (10 molar equiv) were recorded by following the CD signal at the wavelength of the maximal intensity (265 nm) of *TERRA G4* (Figure S2). These experiments showed that 1 out of 103 tested compounds, namely, **BPBA** significantly increased the T_m of *TERRA G4* (ΔT_m = 4.5 (±0.4) °C, Figure 2 and Table 1). The remaining 102 compounds were unable to significantly stabilize such structure ($\Delta T_m < 3$ °C) (Table S1), and thus were not further considered for the biophysical characterization of the interaction with *TERRA G4*.



(**a**) (**b**) (**c**)

Figure 2. (**a**) CD spectra of *TERRA G4* in the absence (black line) and presence (red line) of 10 molar equiv of **BPBA** recorded at 20 and 100 °C (solid and dashed lines, respectively); (**b**) CD melting profiles of *TERRA G4* in the absence and presence (black and red dots, respectively) of 10 molar equiv of **BPBA** recorded at 1 °C/min heating rate; (**c**) CD stabilization curve for *TERRA G4* with **BPBA**.

Additionally, the investigation of *TERRA G4*/**BPBA** interaction was extended by performing CD melting experiments by using a range of ligand concentrations (Figure S3). The thermal-shift curves of the ligand followed a dose–response pattern (Figure 2c and Table S2), suggesting that the interaction of **BPBA** with G4 is specific.

Table 1. Analysis of **BPBA** interaction with the investigated oligonucleotides.

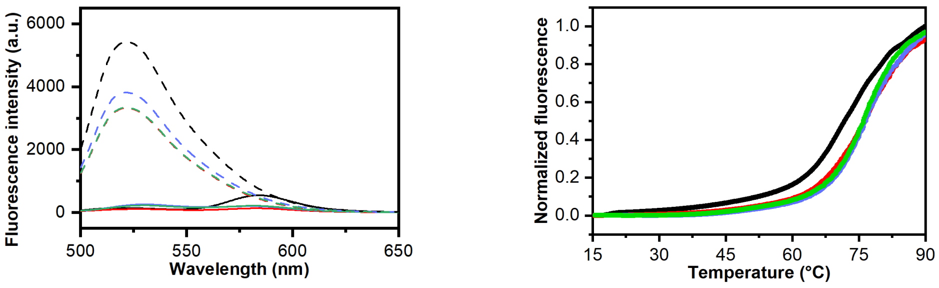
	Circular Dichroism (CD) Melting						
ΔT_m (°C) [1]	*TERRA G4*	*c-kit2 G4*	*c-myc G4*	*Tel*$_{23}$*G4*	*GSEC G4*	*Bcl-2 G4*	*Hrp*$_{20}$
	4.5 (±0.4)	18.7 (±0.3)	9.4 (±0.3)	0.7 (±0.2)	1.7 (±0.3)	2.9 (±0.3)	1.0 (±0.5)
	Förster resonance energy transfer (FRET) melting						
	F-TERRA-T	*F-TERRA-T* + *Hrp*$_{27}$ (1:30)	*F-TERRA-T* + *Hrp*$_{27}$ (1:100)				
	3.6 (±0.2)	3.7 (±0.2)	3.1 (±0.2)				
	Fluorescent intercalator displacement (FID)—thiazole orange						
DC$_{50}$ (μM)	*TERRA G4*	*c-kit2 G4*	*c-myc G4*	*Tel*$_{23}$*G4*	*Hrp*$_{27}$		
	2.4 (±0.4)	3.8 (±0.6)	n.d. [2]	n.d. [2]	4.1 (±0.5)		
	Fluorescent intercalator displacement (FID)—ethidium bromide						
	TERRA G4						
	1.3 (±0.4)						
	Microscale thermophoresis (MST)						
K_d (μM)	*TERRA G4*	*c-kit2 G4*	*c-myc G4*	*Tel*$_{23}$*G4*	*Hrp*$_{20}$		
	9.6 (±0.6)	23.5 (±0.6)	47.5 (±0.4)	n.d. [2]	n.d. [2]		

[1] $\Delta T_\mathrm{m} = T_{\mathrm{m}\,(\text{oligonucleotide+10 ligand equiv})} - T_{\mathrm{m}\,(\text{oligonucleotide})}$. T_m values in the absence of ligand are *TERRA G4* = 74.3 (±0.1) °C; *c-kit2 G4* = 59.7 (±0.1) °C; *c-myc G4* = 75.5 (±0.1) °C; *Tel*$_{23}$ *G4* = 53.7 (±0.1) °C; *GSEC G4* = 78.8 (±0.1) °C; *Bcl-2 G4* = 74.7 (±0.2) °C; *Hrp*$_{20}$ = 65.5 (±0.2) °C; *F-TERRA-T* = 73.5 (±0.1) °C. All experiments were performed in duplicate, and reported values are the average of two measurements. [2] n.d. = not determinable.

2.3. *BPBA Is a Selective G4 Binder That Exhibits Preference for Parallel G4 Conformations*

Once a G4 ligand is validated, its selectivity towards a certain nucleic acid structure must be assessed. Under physiological conditions, most DNA is in the B form. Since binding to B-DNA duplex can cause undesired toxicity effects, it is crucial to determine the selectivity of the ligand towards G4 vs. duplex structures before proceeding with a more in-depth characterization of its binding properties. Here, a 20-mer hairpin-forming sequence (*Hrp*$_{20}$) was chosen as a suitable duplex model. The CD spectrum of the hairpin in the presence of K^+ was characterized by a positive band at around 280 nm and a negative one at 250 nm, confirming duplex formation (Figure S4). These bands were not modified upon the addition of compound **BPBA** (10 molar equiv; Figure S4). Next, CD melting experiments of the hairpin were recorded both in the absence and presence of ligand following variations in CD signal intensity at 280 nm. No significant change of T_m was observed in this case (Figure S4 and Table 1), suggesting that **BPBA** selectively stabilizes the G4 over duplex DNA conformation.

The RNA G4 stabilizing properties of **BPBA** were further investigated by Förster resonance energy transfer (FRET) melting assay using a G4-forming telomeric RNA sequence dually labeled with donor FAM and acceptor TAMRA at the 5′ and 3′ ends, respectively (*F-TERRA-T*) [60]. FRET melting curves (Figure 3) confirmed that **BPBA** is able to stabilize the G4 structure formed by telomeric RNA (ΔT_m = 3.4 (±0.1) °C). Moreover, to further confirm the selectivity of **BPBA** for G4 over the duplex, a competition FRET melting experiment was carried out in the presence of a large excess of a duplex model, i.e., a 27-mer hairpin duplex-forming DNA (*Hrp*$_{27}$) [61–64]. The results of this experiment clearly showed that the stabilizing effect of **BPBA** on *TERRA G4* was not affected by the presence of the duplex competitor (Table 1), meaning that this compound preferably binds to G4s.



(**a**) (**b**)

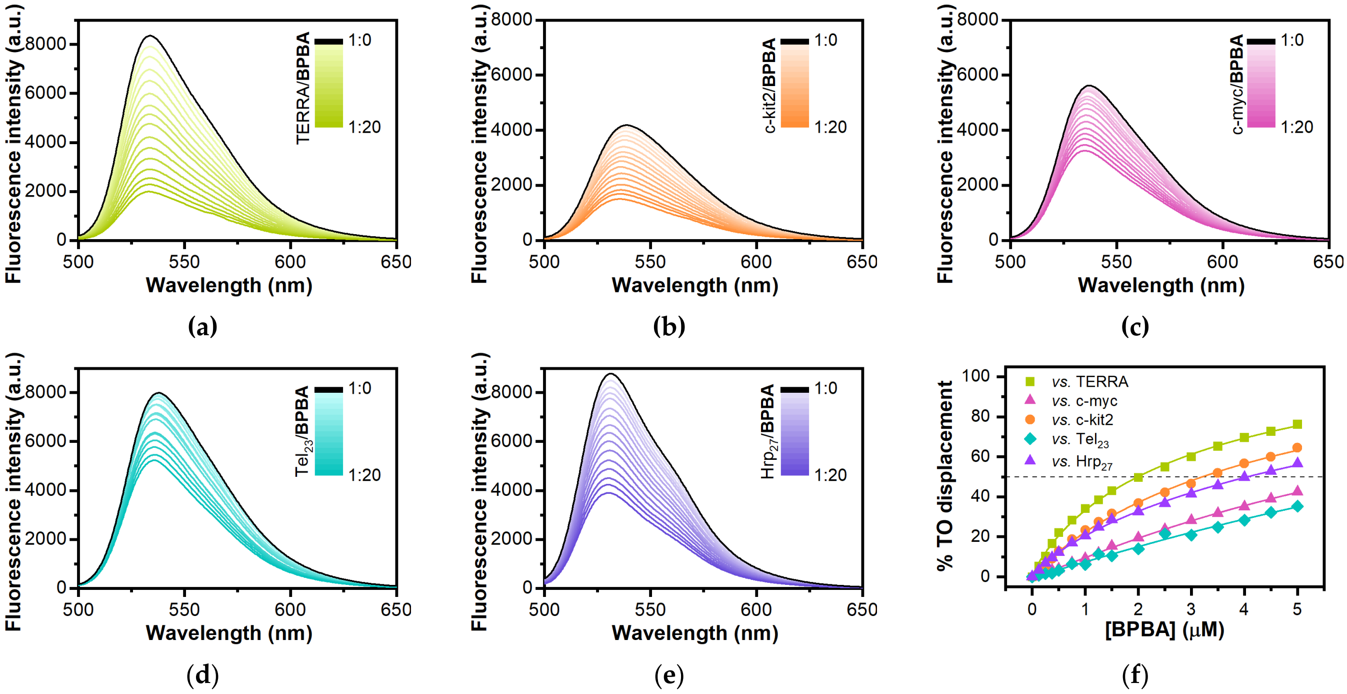
Figure 3. (**a**) Fluorescence emission spectra recorded at 15 °C (solid lines) and 90 °C (dashed lines); (**b**) FRET melting experiments for *F-TERRA-T* (0.2 μM) in the absence (black) and presence (red) of **BPBA** (2.0 μM). Experiments in the presence of **BPBA** were also performed by adding a large excess of Hrp_{27} duplex competitor (6.0 μM, blue; and 20.0 μM, green).

Since **BPBA** was selected for its ability to interact with *TERRA G4*, which adopts a parallel propeller-type conformation, we examined if **BPBA** also interacts with analogous parallel G4 structures formed by DNA and RNA G-rich sequences, and if it is potentially capable of discriminating between parallel and antiparallel G4 conformations. To this aim, we used three different G4-forming DNA sequences derived from the nuclease hypersensitive region of the *c-KIT* (*c-kit2 G4*) and *c-MYC* (*c-myc G4*) gene promoters, which form parallel propeller-type G4 structures in K^+-containing buffer, and from the human telomeric DNA sequence, particularly the 23-mer truncation (Tel_{23} *G4*), which rather folds in an antiparallel (3 + 1) hybrid G4 conformation in the same buffer conditions [65–68]. To obtain information about the ability of **BPBA** to interact with other parallel RNA G4 structures, two additional G4-forming RNA sequences derived from the GSEC lncRNA (*GSEC G4*) and the 5′-UTR of Bcl-2 mRNA (*Bcl-2 G4*) were also investigated [69–71]. The proper folding adopted by each of these G4-forming sequences was first confirmed by CD spectra. As expected, *c-kit2 G4*, *c-myc G4*, *GSEC G4*, and *Bcl-2 G4* displayed a positive band at around 265 nm and a negative one around 240 nm in the CD spectrum (Figures S5 and S6). These bands are characteristic of parallel-stranded G4 topologies [59]. On the other hand, Tel_{23} *G4* showed a positive band at 289 nm with a shoulder at ca. 268 nm and a weak negative band at around 240 nm (Figures S5 and S6), which are consistent with the presence of a (3+1) hybrid G4 as major conformation [59]. As already done for the other nucleic acid molecules, CD experiments were also performed to examine the potential of **BPBA** to alter the native folding topology of these G4s. No significant variations in CD signal were observed for any of these G4 structures (Figures S5 and S6), suggesting an overall preservation of their G4 architectures upon addition of the ligand (10 molar equiv). Hence, the stabilizing properties of **BPBA** on these G4s were evaluated by CD melting experiments (Figures S5 and S6). Results of these experiments clearly indicate the ligand ability to bind and stabilize *c-kit2 G4* and *c-myc G4* (ΔT_m = 18.7 (±1.0) and 9.4 (±0.5) °C, respectively), both having parallel G4 conformations with a negligible effect (ΔT_m = 0.7 (±0.2) °C) on the Tel_{23} *G4* hybrid structure (Table 1). No relevant ligand-induced thermal shift was observed in the case of both *GSEC* and *Bcl-2 G4s* ($\Delta T_m < 3$ °C), suggesting that **BPBA** could preferentially stabilize *TERRA G4* over other RNA G4s.

However, since the stabilization imparted by the ligand would naturally be more pronounced in intrinsically less stable oligonucleotides [72], the direct comparison of ΔT_m values cannot provide straightforward information on the binding affinity.

2.4. Analysis of Ligand Binding Affinity by Fluorescent Intercalator Displacement Assay

To gain insight into the binding affinity of **BPBA** for the different RNA/DNA G4s, fluorescent intercalator displacement (FID) experiments were carried out. This assay is based on the competitive displacement of a light-up fluorescent probe, in this case, thiazole orange (TO), from the DNA upon addition of increasing amounts of a candidate ligand [73–75]. TO is almost nonfluorescent when free in solution, while it is strongly fluorescent when bound to DNA [75]. Ligand-induced TO displacement decreases fluorescence, thus allowing for the determination of their relative binding affinity for the structure under examination. Here, TO displacement by **BPBA** was investigated for *TERRA G4*, *c-kit2 G4*, *c-myc G4*, *Tel*$_{23}$ *G4*, and *Hrp*$_{27}$. **BPBA** concentrations required to give 50% TO displacement (DC$_{50}$ values) were calculated from dose–response curves fitted to these data (Figure 4). As far as *TERRA G4* is concerned, the **BPBA** concentration at which 50% displacement was achieved was 2.4 (±0.4) µM, indicating good affinity for this G4 motif. On the other hand, a DC$_{50}$ value of 3.8 (±0.6) and 4.1 (±0.5) µM was obtained for the interaction of **BPBA** with *c-kit2 G4* and *Hrp*$_{27}$, respectively, suggesting a lower affinity than that for *TERRA G4*. In the case of *c-myc G4* and *Tel*$_{23}$ *G4*, it was not possible to reach a 50% displacement of TO even after addition of a large excess of the binder (20 molar equiv), clearly suggesting weaker ligand interactions.



(a) (b) (c)

(d) (e) (f)

Figure 4. TO displacement titrations for (**a**) *TERRA G4*, (**b**) *c-kit2 G4*, (**c**) *c-myc G4*, (**d**) *Tel*$_{23}$ *G4*, and (**e**) *Hrp*$_{27}$ upon addition of increasing amounts of **BPBA**; (**f**) Dose–response curves from FID experiments.

2.5. Study of Interaction between BPBA and TERRA G4 by Ethidium Bromide Displacement Assay

To obtain information on the binding mode of **BPBA** to *TERRA G4*, an ethidium bromide (EB) displacement assay was performed using fluorescence spectroscopy. EB binds to duplex DNA through intercalation, and to G4 DNA through π–π stacking on the external G-tetrads [76]. In the absence of DNA, EB emits weak fluorescence at 595 nm, while its fluorescence is strongly enhanced upon association with G4s as a consequence of the hydrophobic environment experienced by EB upon binding to the nucleic acid [77]. Therefore, the addition of a G4 ligand decreases EB fluorescence intensity if it binds to G4 via end-stacking mode. Displacement titrations performed by adding increasing amounts of **BPBA** to the *TERRA G4*/EB complex showed a substantial decrease in the fluorescence intensity of EB (Figure 5), thus suggesting an end-stacking binding mode for this ligand to *TERRA G4* [78,79]. In addition, the concentration of **BPBA** required to give the 50% decrease in EB fluorescence (DC$_{50}$ value) was calculated from dose–response curves obtained by plotting the percentage of EB displacement against ligand concentration. A DC$_{50}$ value of 1.3 (±0.4) µM was determined, confirming once again the strong interaction between **BPBA** and this G4 motif.

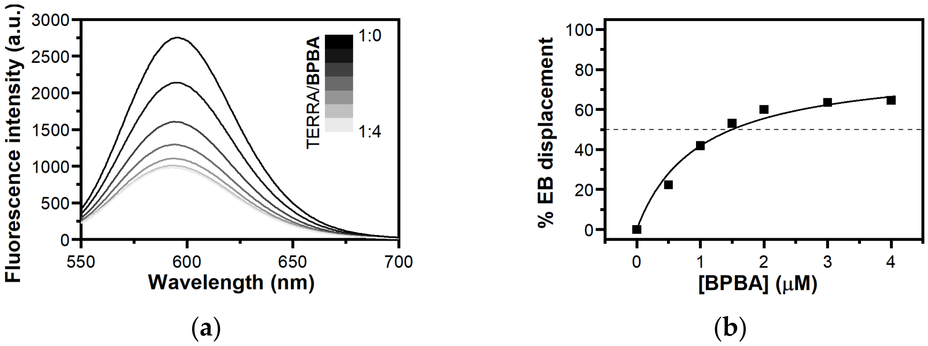


Figure 5. (**a**) Fluorescence spectra for ethidium bromide displacement from *TERRA G4* in the presence of increasing concentrations of **BPBA**; (**b**) Dose–response curves from FID experiments.

2.6. Determination of BPBA Affinity for RNA and DNA G4s

To obtain quantitative data on the binding affinity of **BPBA** for the investigated G4s, and to confirm ligand selectivity for G4 over the duplex, microscale thermophoresis (MST) experiments were carried out. MST is a rapid and easy methodology to measure the affinity of a small molecule for a nucleic acid target in solution [80–82]. This technique is based on thermophoresis, the directed motion of molecules in small temperature gradients. Thermophoresis is highly sensitive to all types of binding-induced changes of molecular properties, be it in size, charge, hydration shell, or conformation. Thus, if the ligand binding to the investigated molecule alters at least one of these parameters, it also changes the thermophoretic behavior of the target. This effect can be used to evaluate equilibrium dissociation constant K_d. To this purpose, serial dilutions of **BPBA** were prepared, mixed with a constant concentration of Cy5.5-labeled oligonucleotides (*TERRA G4*, *c-kit2 G4*, *c-myc G4*, *Tel*$_{23}$ *G4*, or *Hrp*$_{20}$), loaded into capillaries, and analyzed by MST. Results of MST binding curves (Figure 6 and Figure S7, Table 1) indicated that **BPBA** was able to bind to the parallel-stranded G4 structures, showing the lower K_d for *TERRA G4* (9.6 (±0.6) µM), followed by *c-kit2 G4* and *c-myc G4* (K_d = 23.5 (±0.6) µM and 47.5 (±0.4) µM, respectively) (Table 1). On the other hand, no significant change in the thermophoretic signal was observed for *Tel*$_{23}$ *G4* and *Hrp*$_{20}$ hairpin-duplex (Figure S7), clearly indicating the absence of a significant interaction in these cases.

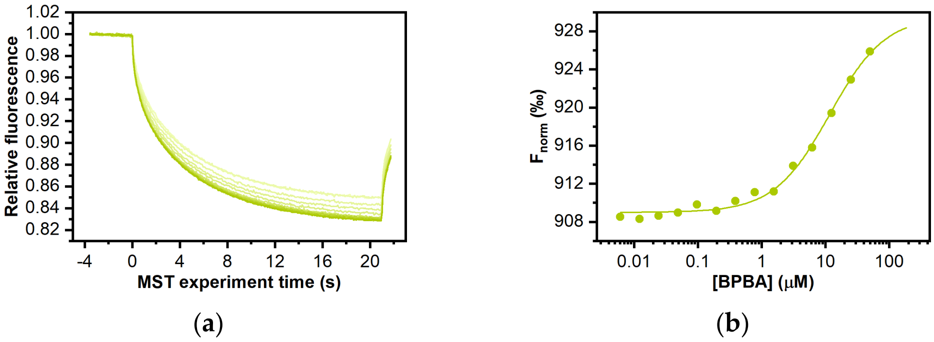
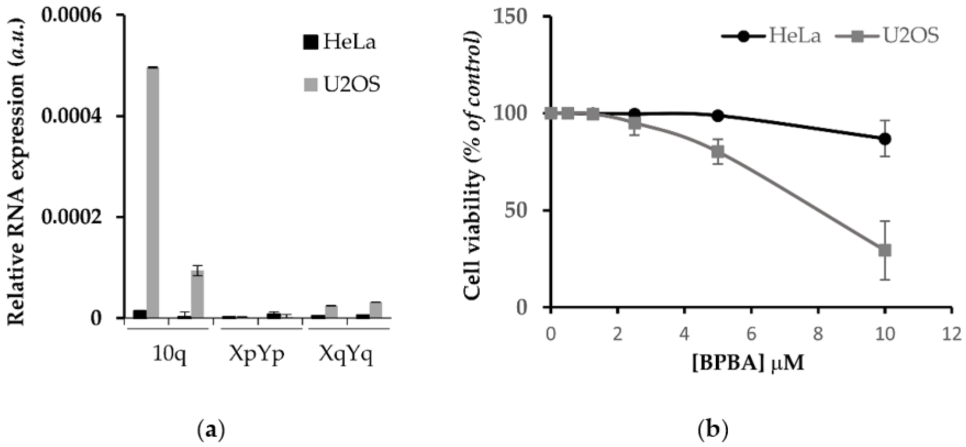


Figure 6. MST measurements on the interaction of compound **BPBA** with TERRA G4. (**a**) Time traces recorded by incubating increasing concentrations of **BPBA** with the labeled G4; (**b**) the corresponding binding curves.

These results agree with those obtained with other techniques and confirm the preferential binding of **BPBA** to parallel over antiparallel-stranded G4 topologies, and its selectivity for the G4 over duplex conformation. Additionally, although **BPBA** showed propensity to bind both *TERRA G4* and the parallel G4s *c-myc G4* and *c-kit2 G4*, it showed higher binding affinity for TERRA.

2.7. Antiproliferative Effect of BPBA in Low vs. High TERRA-Expressing Human Cancer Cells

G4 binders have an established antiproliferative effect in cancer cells depending on their ability to induce DNA damage response (DDR) or to inhibit the expression of cellular oncogenes. Since **BPBA** showed a high affinity to *TERRA G4*, we assessed the antiproliferative effect of this compound in correlation with TERRA expression. To this aim, we employed human cervix cancer cells (HeLa) characterized by telomerase activity and low TERRA expression, as well as human osteosarcoma cells (U2OS) lacking telomerase activity and expressing high levels of TERRA [83,84]. In those cell lines, TERRA expression was measured by RT qPCR assay with primers against some of the most active TERRA promoters located at subtelomeres of chromosomes 10q, XqYq, and XpYp, showing a huge difference of expression between the two cell lines (Figure 7a). Then, the viability of cells exposed to **BPBA** concentration ranging from 50 nM to 10 µM was assessed by crystal violet assay. As shown in Figure 7b, U2OS cells were significantly more sensitive to **BPBA** (IC_{50} = 8.1 (±1.0) µM) with respect to HeLa (IC_{50} >> 10 µM).



(**a**) (**b**)

Figure 7. Differential effect of **BPBA** on viability of high vs. low TERRA expressing cells. (**a**) RT qPCR showing relative expression of specific TERRA RNA at different subtelomeric loci (10q, XpYp, XqYq) of HeLa, and U2OS cells; (**b**) HeLa and U2OS exposed to indicated doses of **BPBA** and, after 6 days, analyzed by crystal violet assay to determine the fraction of surviving cells. Percentages of surviving cells relative to untreated samples and mean of three independent experiments are shown. Error bars are SD.

2.8. BPBA Stabilizes TERRA Levels in U2OS Cells

To gain insight into the mechanism underlying the differential biological effect of **BPBA** on high vs. low TERRA-expressing cells, we assessed the ability of **BPBA** to bind and stabilize TERRA in cellulo by RT qPCR analysis of relative TERRA expression upon treatment. To this aim, both HeLa and U2OS cell lines were exposed to different concentrations of **BPBA**; after 72 h, RNA was extracted and processed for TERRA analysis. Relative TERRA expression reported in Figure 8 clearly shows that **BPBA** induced a stabilization of TERRA expression in U2OS that led to an accumulation of the RNA within the cell. This effect reached a saturation point at 5 µM, when presumably all TERRA molecules were bound by the ligand and sequestered from the degradation complexes binding that regulate the physiological turnover of the lncRNA. We did not observe the same effect in HeLa, depending on the low abundance of TERRA.

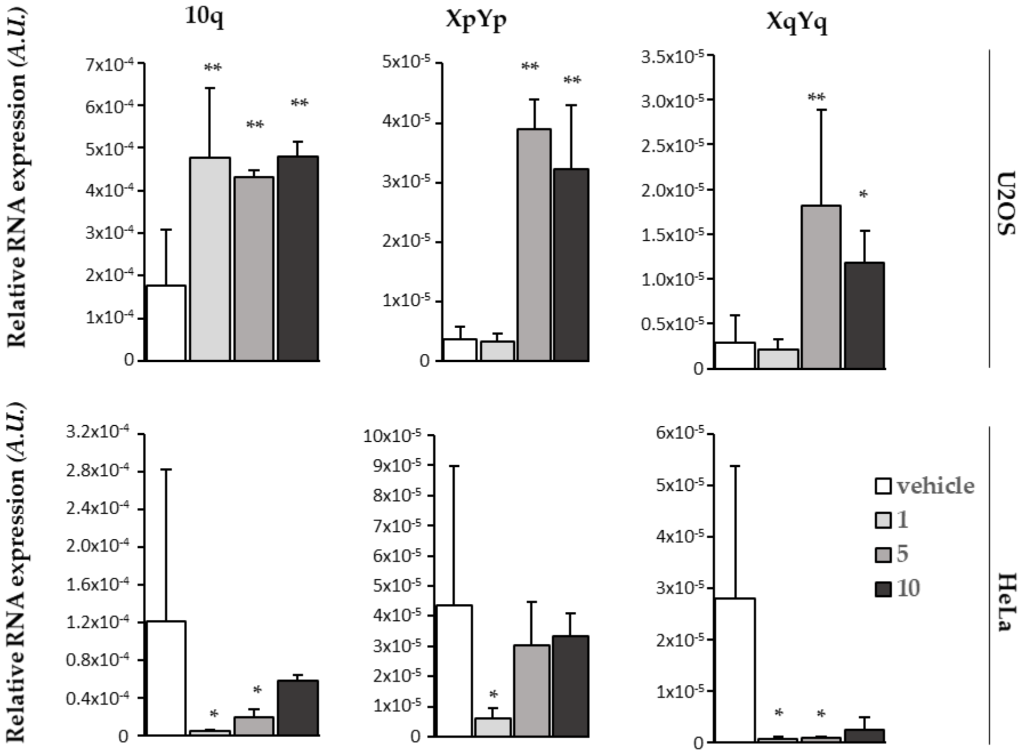
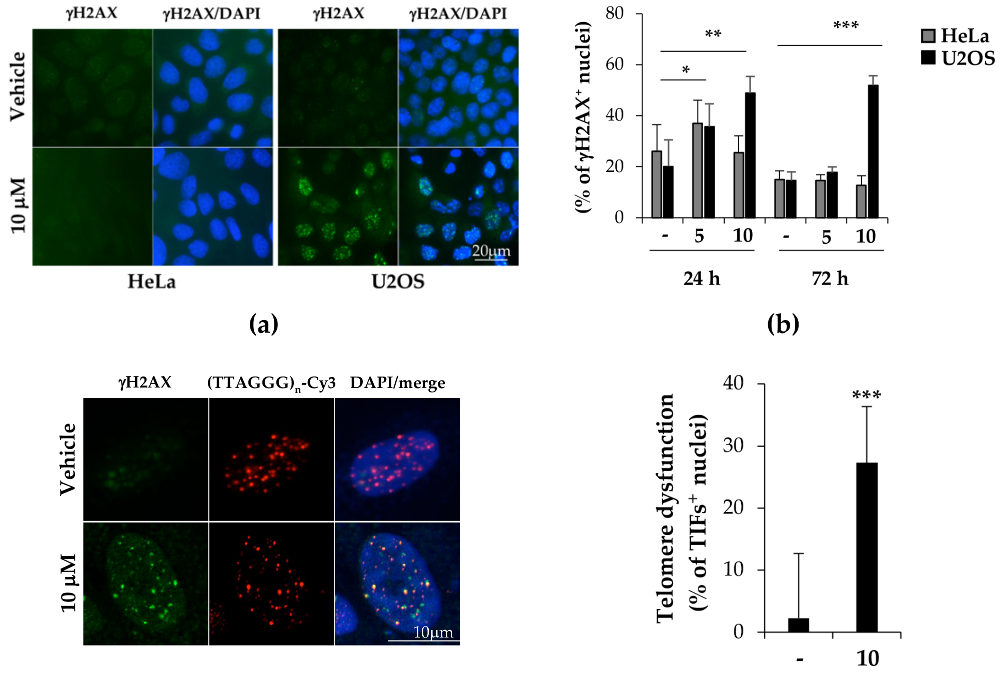


Figure 8. Effect of **BPBA** on TERRA expression. RT qPCR showing relative expression of specific TERRA RNA at different subtelomeric loci of HeLa and U2OS cells treated with the vehicle or indicated doses (1, 5, or 10 μM) of **BPBA** for 72 h. Mean of three independent experiments is shown. Error bars are SD; * = $p < 0.5$; ** = $p < 0.1$.

2.9. BPBA Induces Persistent DDR Activation in U2OS Cells

TERRA displacement from telomeres is supposed to induce telomere dysfunction [85]. To better understand the mechanisms underlying the cytotoxicity of **BPBA** in U2OS cells, we investigated the possibility that TERRA stabilization by **BPBA** could induce a DDR at telomeres (a marker for telomere dysfunction) leading to cell death. To this aim, both U2OS and HeLa cells were exposed to the compound for 24 or 72 h, and DDR activation was measured as the percentage of cells displaying γH2AX histone phosphorylation, a marker of DDR. As shown in Figure 9, HeLa treated with 5 or 10 μM **BPBA** showed a negligible DDR induction both in the first 24 h of treatment and in the following 72 h. Conversely, compared to the control, both 5 and 10 μM **BPBA** were able to induce DDR in a significant fraction of U2OS cells in the first 24 h (Figure 9b). However, while the 5 μM dose-induced DDR was rescued in the following 72 h, DDR was persistent at 72 h in 50% of the U2OS cells treated with 10 μM **BPBA**, in agreement with the calculated IC_{50} dose (Figure 7b). Lastly, to ascertain if the activated DDR coincided with telomeric loci, telomere-dysfunction-induced foci (TIF) activation was measured in U2OS cells treated with 10 μM **BPBA** for 72 h (Figure 9c). TIF positive cells (defined as cells displaying at least 4 of Cy3-conjugated telomere PNA probe/phosphorylated γH2AX histone colocalizing spots) were significantly increased in treated samples compared to in the control (Figure 9d).



(**a**) (**b**)

(c) (**d**)

Figure 9. TERRA binding by **BPBA** triggers persistent DDR and telomere dysfunction in U2OS cells. HeLa and U2OS cells treated with indicated doses of **BPBA**. At each endpoint, cells were fixed and processed for immunofluorescence with the antiphosphorylated γH2AX mAb primary antibody, followed by goat–antimouse 488 secondary antibody (**a**,**b**) alone or (**c**,**d**) in combination with FISH using the telomere-specific $(TTAGGG)_n$-Cy3 PNA probe. Fluorescence signals acquired with a Leica DMIRE deconvolution microscope (representative images at 72 h shown in (**a**)) or a Zeiss LMS confocal scanner (**b**) at 63× magnification. Percentage of cells displaying γH2AX signals or TIFs (>4 $(TTAGGG)_n$-Cy3/γH2AX colocalizing spots) was scored and reported in histograms ((**b**,**d**), respectively). Histograms report the mean of at least 6 different fields per sample ($n > 150$). Error bars are SD; * = $p < 0.5$; ** = $p < 0.1$; *** = $p < 0.05$.

3. Materials and Methods

3.1. Materials

CPG supports, (2′-OTBDMS)-RNA and DNA phosphoramidites, and all reagents for oligonucleotide synthesis were purchased from Link Technologies (Bellshill, UK). All other reagents and solvents were from Sigma-Aldrich (Merck KGaA, Darmstadt, Germany) and used without further purification. All buffers were produced from highly purified Milli-Q water and sterilized before use with diethylpirocarbonate (DEPC, from Merck KGaA, Darmstadt, Germany) and/or autocleavage. Putative ligands were purchased from Mcule (mcule.com Kft. Budapest, Hungary).

3.2. Virtual Screening

The 3D coordinates of the *TERRA G4* structures formed by the r(UAGGGUUAGGGU) sequence determined by NMR (PDB code: 2KBP) [54] and by X-ray diffraction (PDB code: 3IBK) [50] were downloaded from the Protein Data Bank website. PDB structures were prepared for docking using AutoDockTools by retaining nonstandard residues. Co-crystallized water molecules and counterions were removed from the X-ray structure.

A set of small organic solvent molecules, used as probes for hot spot mapping the binding surface of G4s [56], were gathered and prepared for docking in Mcule. Solvent molecules were docked to each target by using the Docking workflow step in Mcule (exhaustiveness: 8), which utilizes the Vina docking algorithm [55]. Binding sites were defined as cubes with 100 Å length in each direction to ensure that poses were evaluated on the surface of the whole RNA structures. Potential hot spots were located where the top binding pose of at least three different solvent molecules were overlapping. The centers of the hot spots were determined as the average of the X, Y, and Z coordinates of all atoms of the overlapping solvent molecules.

To analyze whether the proximity of the identified hot spots (hot spot regions) could be targeted by larger but still small molecules, a diverse set of 300 compounds were selected from the Mcule database containing 5.2 M compounds at the time of the selection. The selection included the following property filters: $150 \leq$ mass ≤ 300; $0 \leq \log P \leq 3$; H-bond acceptors ≤ 3; H-bond donors ≤ 3; rings ≥ 1; rotatable bonds ≤ 3; heavy atoms ≥ 15. Rapid elimination of swill (REOS) [86] filter was also applied to eliminate compounds containing unwanted motifs. Lastly, the 'diversity selection' workflow step was used to select the most diverse (dissimilar) molecules by eliminating the closest analogs, thus maximizing the coverage of the chemical space. The resulting 300 compounds were docked by the Docking Vina workflow step in Mcule to binding sites defined around the center of each of the previously identified hot spots (cubes, length in each direction: 22 Å). Docking scores of the compounds were analyzed for each hot spot region.

Then, a diverse set of commercially available compounds was prepared as a screening library. We started our selection from the Mcule database containing millions of purchasable compounds. The following property filters were applied: mass ≥ 200 Da; $\log P \geq 0$; rule-of-five violations = 0; rings ≥ 1; rotatable bonds ≤ 4; heavy atoms ≥ 15. Subsequently, the REOS filter was applied to eliminate compounds with toxic or non drug-like substructures [86]. Lastly, a diverse set of a maximum of 0.7 Tanimoto coefficient was created on the basis of the OpenBabel linear fingerprint. These filters resulted in a screening library of 58870 compounds. The most relevant protonation state of these compounds at pH 7.4 was generated by OpenBabel 2.3.1. The generated screening library was docked into each identified binding site of each RNA structure. Calculations were run on Mcule using the Docking Vina workflow step with default settings. All compounds were ranked on the basis of their docking score. The distribution of the docking scores for each virtual screening was analyzed to ensure that the docking method could distinguish between different ligands, i.e., the scores of the top hits significantly differ from those at the end of the ranked screening database. After this step, 103 compounds were selected and purchased for further analysis. Stock solutions of these compounds were prepared at 10 mM in DMSO. No solubility problems were encountered for the putative ligands at any of the concentrations used in the various experiments.

3.3. Oligonucleotide Synthesis and Sample Preparation

RNA/DNA sequences were synthesized on an ABI 394 DNA/RNA synthesizer (Applied Biosystem, Foster City, CA, USA) using standard β-cyanoethyl phosphoramidite solid phase chemistry at 1 μmol synthesis scale. Regarding RNA synthesis, 5-benzylthio-1-H-tetrazole (BTT) instead of 4,5-dicyanoimidazole (DCI) was used as activator reagent, and coupling steps were prolonged of 5 min. Another difference concerned the deprotection of bases and phosphates. A concentrated NH_4OH/EtOH (3:1, *v*/*v*) solution was used in the case of RNA, and the reaction was left at r.t. for 12 h. For DNA sequences, deprotection and detachment were performed by using a concentrated NH_4OH aqueous solution at 55 °C for 12 h. Both RNA and DNA sequences were purified by high-performance liquid chromatography (HPLC) on a Nucleogel SAX column (1000-8/46, Macherey-Nagel, GmbH & Co. KG, Dueren, Germany), as previously reported [87]. The fractions of the oligomers were collected and successively desalted by Sep-Pak cartridges (C-18). Lastly, 2′-TBDMS groups in RNA were removed by $Et_3N \cdot 3HF$/DMF (1:3, *v*/*v*) at r.t. for 12 h. The reaction was quenched with 0.1 M TEAA buffer (pH 7.0) and again desalted on a Sep-pak (C-18) cartridge. All oligonucleotides were proven to be >98% pure by NMR. The following oligonucleotides were synthesized: the 12-mer truncation of the human telomeric repeat-containing RNA sequence r(UAGGGUAAGGGU) (*TERRA G4*); the G4-forming sequence from GSEC long noncoding RNA r(GGGGUGGAGGAGGGGGAAGGGCGGGGG) (*GSEC G4*); the G-rich sequence of the 5′-UTR of Bcl-2 mRNA (GGGCCGUGGGGUGGGAGCUGGG) (*Bcl-2 G4*);

the *c-Kit2* sequence from the *c-Kit* oncogene promoter d(CGGGCGGGCGCTAGGGAGGGT) (*c-kit2 G4*); *c-Myc* promoter sequence d(TGAGGGTGGGTAGGGTGGGTAA) (*c-myc G4*); the 23-mer truncation of the human telomeric sequence d(TAGGGTTAGGGTTAGGGTTAGGG) (Tel_{23} *G4*); the 20-mer hairpin duplex-forming sequence d(CGAATTCGTTTTCGAATTCG) (Hrp_{20}); and the 27-mer hairpin duplex-forming sequence d(CGCGAATTCGCGTTTCGCGA ATTCGCG) (Hrp_{27}). Oligonucleotides were prepared in the appropriate buffer, and their concentration was measured by UV adsorption at 90 °C using the appropriate molar extinction coefficient values, ε (λ = 260 nm), calculated by the nearest-neighbor model [88].

3.4. Circular Dichroism (CD) Experiments

CD experiments were performed on a Jasco J-815 spectropolarimeter (Jasco, Easton, MD, USA) equipped with a PTC-423S/15 Peltier temperature controller. All spectra were recorded at 20 and 100 °C in the wavelength range of 230–320 nm, and averaged over three scans. A scan rate of 100 nm/min with a 1 s response time and 1 nm bandwidth were used. The buffer baseline was subtracted from each spectrum. Concentrations of 2 μM for G4s and 4 μM for Hrp_{20} were used. RNA G4s were prepared in 20 mM KH_2PO_4/K_2HPO_4 buffer (pH 7.0) containing 70 mM KCl, while a buffer solution consisting of 5 mM KH_2PO_4/K_2HPO_4 (pH 7.0) containing 20 mM KCl was used for all DNA samples. All oligonucleotide samples were annealed by heating at 90 °C for 5 min, followed by a slow cooling to room temperature overnight. CD spectra of oligonucleotide/ligand mixtures were obtained by adding 10 molar equiv of ligand (stock solutions of ligands were 10 mM in DMSO). CD melting experiments were carried out in the 20–100 °C temperature range at a 1 °C/min heating rate by following changes of the CD signal at the wavelengths of the maximal CD intensity (i.e., 264 nm for *TERRA G4*, *GSEC G4*, *Bcl-2 G4*, *c-kit2 G4*, and *c-myc G4*; 287 nm for Tel_{23} *G4*; 280 nm for Hrp_{20}). CD melting experiments were performed in the absence and presence of ligands (10 molar equiv) added to the folded nucleic acid structures. The apparent melting temperatures (T_m) were determined from a curve fit using Origin 7.0 software (OriginLab Corp., Northampton, MA, USA). ΔT_m values were determined as the difference in the T_m values of the nucleic acid structures in the presence and absence of ligands. All experiments were performed in triplicate, and the reported values are the average of the three measurements.

3.5. FRET Melting Experiments

Measurements were carried out on a FP-8300 spectrofluorometer (Jasco, Easton, MD, USA) equipped with a Peltier temperature controller system (Jasco PCT-818) using the dual-labeled G4-forming telomeric RNA sequence FAM-[r(GGGUAAGGGUAAGGGUAAGGG)]-TAMRA (*F-TERRA-T*), provided from Biomers (Ulm, Germany). The oligonucleotide was dissolved in water at 1 mM, diluted at 1 μM using 5 mM potassium phosphate buffer (pH 7.0) containing 20 mM KCl, and lastly annealed by heating to 90 °C for 5 min, followed by cooling to room temperature overnight and storage at 4 °C for 24 h before data acquisition. Experiments were performed in sealed quartz cuvettes with a path length of 1 cm by using 0.2 μM of prefolded *F-TERRA-T* G4 target, the ligand at 2 μM, and the Hrp_{27} duplex competitor at 0, 6, and 20 μM final concentrations [61–64]. In addition, a blank with no compound or competitor was also analyzed. Fluorescence spectra were acquired before (at 15 °C) and after (at 90 °C) melting assay. The dual-labeled oligonucleotide was excited at 492 nm, and emission spectra were recorded between 500 and 650 nm by using 100 nm/s scan speed. Excitation and emission slit widths were both set at 5 nm. FRET melting was performed by monitoring the emission of FAM at 522 nm (upon excitation at 492 nm), using a heating gradient of 1 °C/min over the range 15–90 °C. Emission of FAM was normalized between 0 and 1. Final analysis of the data was carried out using Origin 7.0 software (OriginLab Corp., Northampton, MA, USA).

3.6. Fluorescent Intercalator Displacement (FID) Assay with Thiazole Orange (TO)

A solution containing 0.25 μM of prefolded RNA (*TERRA G4*) or DNA G4 target and 0.5 μM of TO in 20 mM KH_2PO_4 (pH 7.0) and 70 mM KCl was prepared in a 1 cm path-length cell, and the corresponding fluorescence spectrum was acquired in the absence and presence of increasing concentrations of **BPBA** (10 mM stock solution in pure DMSO). Each ligand addition (from 0.5 to 20 molar equiv) was followed by a 3 min equilibration time before spectrum acquisition. The FID experiment was extended to a duplex DNA model (*Hrp*$_{27}$), in this case, three equivalents of TO (0.75 μM) were added to an oligonucleotide solution (0.25 μM). Measurements were run at 20 °C on a FP-8300 spectrofluorometer (Jasco, Easton, MD, USA) equipped with a Peltier cell holder (Jasco PCT-818), using an excitation wavelength of 485 nm and recording the emission in the 500–650 nm wavelength range. Both excitation and emission slits were set at 5 nm. The percentage of TO displacement was calculated as TO displacement (%) = $100 - [(F/F_0) \times 100]$, where F_0 is the fluorescence in the absence of ligand and F the fluorescence after each ligand addition. The percentage of displacement was then plotted as a function of the ligand concentration, and DC_{50} was calculated as the required concentration to displace 50% TO from each investigated DNA. Each titration was repeated at least in triplicate.

3.7. Fluorescent Intercalator Displacement (FID) Assay with Ethidium Bromide (EB)

A solution containing 10 μM of prefolded *TERRA G4* and 5 μM of EB in 20 mM potassium buffer (pH 7.0) containing 70 mM KCl was prepared in a 1 cm path-length cell. The fluorescence spectrum of the EB/*TERRA G4* complex in the absence of ligand was first recorded. Then, increasing concentrations of **BPBA** (10 mM stock solution in pure DMSO) were mixed to this EB/*TERRA G4* complex, and spectra were recorded 3 min after each ligand addition. Measurements were run at 20 °C on a FP-8300 spectrofluorometer (Jasco, Easton, MD, USA) equipped with a Peltier cell holder (Jasco PCT-818), using an excitation wavelength of 510 nm, and recording the emission in the 550–700 nm wavelength range. Both excitation and emission slits were set at 5 nm. Experiments were performed in duplicate.

3.8. MicroScale Thermophoresis (MST) Experiments

MST measurements were performed on a Monolith NT.115 (Nanotemper Technologies, Munich, Germany). The Cy5.5-fluorescently labelled oligonucleotides (from Biomers, Ulm, Germany) were prepared at 10–20 μM in 20 mM KH_2PO_4 buffer (pH 7.0) containing 70 mM KCl and annealed as described above. Nucleic acid samples were then diluted using the same phosphate buffer supplemented with 0.1% Tween. Ligand stock solution was 2 mM in pure DMSO. For the MST experiments, the concentration of the labelled G4s was kept constant at 50 nM, while a serial dilution of the ligand (1:2 from 0.4 mM) in the same buffer supplemented with 20% DMSO, was prepared and mixed with the oligonucleotide solution with a volume ratio of 1:1. All samples containing 10% DMSO as the final concentration were loaded into standard capillaries (NanoTemper Technologies, Munich, Germany). Measurements were performed at 20 °C using autotune LED power and medium MST power. MST data analysis was performed by employing the MO. Affinity Analysis software (v2.3) provided with the instrument.

3.9. Cells and Viability Assay (Crystal Violet)

HeLa and U2OS cells were purchased from ATCC and maintained according to the purchaser's instructions. Cells were seeded in 24 wells; after 24 h, cells were exposed to **BPBA** concentrations ranging from 1.25 to 10 μM for 6 days. Then, cells were washed twice in DPBS and fixed with 4% formaldehyde for 15 min at r.t. After washing with DPBS, 300 μL of crystal violet staining solution (Sigma-Aldrich, St. Louis, MO, USA) was added to each well and incubated for 30 min at r.t. Lastly, plates were rinsed twice with water, air-dried at r.t., and cell pellets were dissolved in 400 μL of acetic acid. The optical density of each well in triplicate was measured at 570 nm (OD_{570}) with a 96-well plate in an ELISA reader (Falcon, Corning, NY, USA). The average absorbance in each condition was used to calculate the survival expressed as percent of treated vs. untreated condition. IC_{50} (the necessary dose to reduce survival of 50%) was calculated by Calcusyn software (Biosoft, Cambridge, UK).

3.10. Immunofluorescence/FISH

HeLa and U2OS cells were seeded and treated with **BPBA**. At each endpoint, cells were fixed in 2% formaldehyde, permeabilized in 0.25% Triton X-100 in PBS for 5 min at r.t., and incubated with the mouse mAb anti-γH2AX (Millipore, Burlington, MA, USA) followed by the secondary Alexa 488 goat antimouse antibody. Lastly, nuclei were counterstained with DAPI (Sigma-Aldrich, St. Louis, MO, USA). For combined FISH experiments, after immunofluorescence, samples were refixed in 2% formaldehyde and dehydrated by ethanol series. Then, slices were hybridized with telomere-specific $(TTAGGG)_n$-Cy3 PNA probe (Panagene, Daejeon, South Korea) according to the manufacturer's instruction. Lastly, samples were counterstained with DAPI (Sigma-Aldrich, St. Louis, MO, USA). Fluorescence signals were recorded by using a Leica DMIRE2 microscope equipped with a Leica DFC 350FX camera and elaborated by a Leica FW4000 deconvolution software (Leica, Solms, Germany) at 63× magnification. TIFs images were acquired with a Zeiss LSM confocal laser scanner (Zeiss, Jena, Germany) at 63× magnification.

3.11. TERRA Real-Time qPCR

Real-time qPCR analysis of TERRA was performed as described [89]. Briefly, RNA was extracted from cells with an RNAeasy mini kit (Quiagen, Hilden, Germany) and accurately digested with the RNase-free DNase set (Quiagen, Hilden, Germany). Then, RNA quality was checked on FA gel electrophoresis and amplified in real-time PCR assay with subtelomere specific primers with a 7900HT Fast Real Time PCR System (Applied Biosystem, Waltham, MA, USA).

4. Conclusions

Targeting noncanonical nucleic acid structures such as G4s is an appealing opportunity for drug intervention in anticancer therapy. Indeed, these unusual arrangements, and in particular their structural conversions, appear to play roles in regulating some important disease-related biological processes. Low-molecular-weight compounds affecting nucleic acid conformational equilibria by preferentially binding to a given form could, therefore, represent a real chance for therapeutic applications. Besides regulating telomerase activity and protecting chromosome ends from telomere degradation, G4-forming TERRA RNA also takes part in heterochromatin formation and homologous recombination, thus representing a valuable therapeutic target. Herein, the application of a virtual screening approach in tandem with experimental screening via CD melting assay succeeded in the identification of a new hit compound (**BPBA**) as a binder of TERRA G4. The in vitro G4 binding properties of this compound were characterized by several biophysical assays (CD, FRET, FID, and MST). CD and FRET melting assays, as well as MST experiments, revealed that **BPBA** features high selectivity toward G4s, being its binding to duplex DNA negligible. Furthermore, TO-FID and MST experiments showed that **BPBA** has enhanced binding affinity towards TERRA G4 vs. other G4-forming DNA sequences present along the human genome. The examination of the molecular structure of **BPBA** compared to other screened compounds suggests that its preferable binding properties may be due to the presence of two benzoimidazole units connected by an aniline residue. This molecular arrangement, which is present only in this compound, gives the molecule extensive planarity, and probably also allows for the optimal distribution of polar groups for interaction with TERRA G4.

Biological characterization demonstrated that **BPBA** can bind and stabilize in cellulo TERRA lncRNAs, probably by sequestering them from the physiological turnover cell machinery. Moreover, TERRA stabilization induced a DDR, putatively by displacing

TERRA from telomeric DNA. Indeed, TERRA physically interacts with telomeric chromatin by forming DNA:RNA hybrids that are required for telomere homeostasis, especially in ALT cells such as U2OS, where TERRA downregulation causes the formation of TIFs [85]. In agreement with this, DDR activation prevalently occurs in U2OS cells, where **BPBA** also has the highest cytotoxic effect.

Overall, this study demonstrates that it is possible to identify TERRA G4 binders with potential pharmacological effects, thus paving the way for the search of new RNA-targeting drug candidates. A relevant percentage of human tumors (around 15%) possess ALT mechanisms for telomere elongation that correlate with high TERRA expression. These tumors are prevalently of mesenchymal origin. They are characterized by high genetic instability, and, in many histotypes, ALT positivity is associated with worse prognosis [90]. In this regard, TERRA G4 ligands could represent an effective pharmacological strategy to hit this class of tumors.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ijms221910315/s1.

Author Contributions: Conceptualization, B.P., A.R. and J.A.; methodology, B.P., E.S., A.R. and J.A.; formal analysis, S.M., N.I. and E.V.; investigation, S.M., A.D.P. and E.V.; resources, B.P., E.S., A.R. and J.A.; data curation, S.M., N.I., A.D.P. and S.D.T.; writing—original draft preparation, S.M., B.P., E.S. and J.A.; writing—review and editing, S.M., B.P., N.I., A.D.P., S.D.T., E.V., E.S., A.R. and J.A.; supervision, B.P., E.S., A.R. and J.A.; funding acquisition, B.P., A.R. and J.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by Italian Association for Cancer Research (AIRC, IG 2020 ID. 24590).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ALT	Alternative lengthening of telomeres
CD	Circular dichroism
DDR	DNA damage response
EB	Ethidium bromide
FID	Fluorescent intercalator displacement
FISH	Fluorescence in situ hybridization
FRET	Förster resonance energy transfer
G4	G-quadruplex
LncRNAs	Long noncoding RNAs
MST	Microscale thermophoresis
TERRA	Telomere repeat containing RNA
TIF	Telomere dysfunction-induced foci
TO	Thiazole orange

References

1. 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. [CrossRef]
2. Mukundan, V.T.; Phan, A.T. Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.* **2013**, *135*, 5017–5028. [CrossRef] [PubMed]
3. Lim, K.W.; Phan, A.T. Structural Basis of DNA Quadruplex-Duplex Junction Formation. *Angew. Chem.* **2013**, *125*, 8728–8731. [CrossRef]
4. Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S. Loop-Length-Dependent Folding of G-Quadruplexes. *J. Am. Chem. Soc.* **2004**, *126*, 16405–16415. [CrossRef]
5. Cheng, M.; Cheng, Y.; Hao, J.; Jia, G.; Zhou, J.; Mergny, J.-L.; Li, C. Loop permutation affects the topology and stability of G-quadruplexes. *Nucleic Acids Res.* **2018**, *46*, 9264–9275. [CrossRef] [PubMed]
6. Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* **2014**, *6*, 75–80. [CrossRef] [PubMed]
7. Chen, X.-C.; Chen, S.-B.; Dai, J.; Yuan, J.-H.; Ou, T.-M.; Huang, Z.-S.; Tan, J.-H. Tracking the Dynamic Folding and Unfolding of RNA G-Quadruplexes in Live Cells. *Angew. Chem. Int. Ed.* **2018**, *57*, 4702–4706. [CrossRef] [PubMed]
8. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **2013**, *5*, 182–186. [CrossRef] [PubMed]
9. Kouzine, F.; Wojtowicz, D.; Yamane, A.; Casellas, R.; Przytycka, T.M.; Levens, D.L. In Vivo Chemical Probing for G-Quadruplex Formation. *Methods Mol. Biol.* **2019**, *2035*, 369–382. [CrossRef] [PubMed]
10. Chen, S.-B.; Hu, M.-H.; Liu, G.-C.; Wang, J.; Ou, T.-M.; Gu, L.-Q.; Huang, Z.-S.; Tan, J.-H. Visualization of NRAS RNA G-Quadruplex Structures in Cells with an Engineered Fluorogenic Hybridization Probe. *J. Am. Chem. Soc.* **2016**, *138*, 10382–10385. [CrossRef] [PubMed]
11. Kwok, C.K.; Marsico, G.; Sahakyan, A.B.; Chambers, V.S.; Balasubramanian, S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* **2016**, *13*, 841–844. [CrossRef]
12. De Magis, A.; Götz, S.; Hajikazemi, M.; Fekete-Szücs, E.; Caterino, M.; Juranek, S.; Paeschke, K. Zuo1 supports G4 structure formation and directs repair toward nucleotide excision repair. *Nat. Commun.* **2020**, *11*, 3907. [CrossRef] [PubMed]
13. Summers, P.A.; Lewis, B.W.; Gonzalez-Garcia, J.; Porreca, R.M.; Lim, A.H.M.; Cadinu, P.; Martin-Pintado, N.; Mann, D.J.; Edel, J.B.; Vannier, J.B.; et al. Visualising G-quadruplex DNA dynamics in live cells by fluorescence lifetime imaging microscopy. *Nat. Commun.* **2021**, *12*, 162. [CrossRef]
14. Johnson, F.B. Fundamentals of G-quadruplex biology. *Annu. Rep. Med. Chem.* **2020**, *54*, 3–44. [CrossRef] [PubMed]
15. Masai, H.; Tanaka, T. G-quadruplex DNA and RNA: Their roles in regulation of DNA replication and other biological functions. *Biochem. Biophys. Res. Commun.* **2020**, *531*, 25–38. [CrossRef]
16. Rhodes, D.; Lipps, H.J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **2015**, *43*, 8627–8637. [CrossRef] [PubMed]
17. Malgowska, M.; Czajczynska, K.; Gudanis, D.; Tworak, A.; Gdaniec, Z. Overview of the RNA G-quadruplex structures. *Acta Biochim. Pol.* **2016**, *63*, 609–621. [CrossRef]
18. Zizza, P.; Cingolani, C.; Artuso, S.; Salvati, E.; Rizzo, A.; D'Angelo, C.; Porru, M.; Pagano, B.; Amato, J.; Randazzo, A.; et al. Intragenic G-quadruplex structure formed in the human CD133 and its biological and translational relevance. *Nucleic Acids Res.* **2016**, *44*, 1579–1590. [CrossRef] [PubMed]
19. Balasubramanian, S.; Hurley, L.H.; Neidle, S. Targeting G-quadruplexes in gene promoters: A novel anticancer strategy? *Nat. Rev. Drug Discov.* **2011**, *10*, 261–275. [CrossRef] [PubMed]
20. Besnard, E.; Babled, A.; Lapasset, L.; Milhavet, O.; Parrinello, H.; Dantec, C.; Marin, J.-M.; Lemaitre, J.-M. Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* **2012**, *19*, 837–844. [CrossRef]
21. Bugaut, A.; Balasubramanian, S. 5′-UTR RNA G-quadruplexes: Translation regulation and targeting. *Nucleic Acids Res.* **2012**, *40*, 4727–4741. [CrossRef]
22. 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*, 7429–7455. [CrossRef]
23. Bryan, T.M. G-Quadruplexes at Telomeres: Friend or Foe? *Molecules* **2020**, *25*, 3686. [CrossRef]
24. Moye, A.L.; Porter, K.C.; Cohen, S.B.; Phan, T.; Zyner, K.G.; Sasaki, N.; Lovrecz, G.O.; Beck, J.L.; Bryan, T.M. Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* **2015**, *6*, 7643. [CrossRef] [PubMed]
25. Riou, J.F.; Guittat, L.; Mailliet, P.; Laoui, A.; Renou, E.; Petitgenet, O.; Megnin-Chanet, F.; Helene, C.; Mergny, J.L. Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2672–2677. [CrossRef] [PubMed]
26. Tan, Z.; Tang, J.; Kan, Z.-Y.; Hao, Y.-H. Telomere G-Quadruplex as a Potential Target to Accelerate Telomere Shortening by Expanding the Incomplete End-Replication of Telomere DNA. *Curr. Top. Med. Chem.* **2015**, *15*, 1940–1946. [CrossRef] [PubMed]
27. Alessandrini, I.; Recagni, M.; Zaffaroni, N.; Folini, M. On the Road to Fight Cancer: The Potential of G-Quadruplex Ligands as Novel Therapeutic Agents. *Int. J. Mol. Sci.* **2021**, *22*, 5947. [CrossRef]

28. Sabharwal, N.; Chen, J.; Lee, J.; Gangemi, C.; D'Urso, A.; Yatsunyk, L. Interactions Between Spermine-Derivatized Tentacle Porphyrins and The Human Telomeric DNA G-Quadruplex. *Int. J. Mol. Sci.* **2018**, *19*, 3686. [CrossRef]
29. Pagano, B.; Amato, J.; Iaccarino, N.; Cingolani, C.; Zizza, P.; Biroccio, A.; Novellino, E.; Randazzo, A. Looking for Efficient G-Quadruplex Ligands: Evidence for Selective Stabilizing Properties and Telomere Damage by Drug-Like Molecules. *ChemMedChem* **2015**, *10*, 640–649. [CrossRef]
30. Kwapisz, M.; Morillon, A. Subtelomeric Transcription and its Regulation. *J. Mol. Biol.* **2020**, *432*, 4199–4219. [CrossRef]
31. Azzalin, C.M.; Reichenbach, P.; Khoriauli, L.; Giulotto, E.; Lingner, J. Telomeric Repeat Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends. *Science* **2007**, *318*, 798–801. [CrossRef] [PubMed]
32. Schoeftner, S.; Blasco, M.A. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat. Cell Biol.* **2008**, *10*, 228–236. [CrossRef] [PubMed]
33. Hirashima, K.; Seimiya, H. Telomeric repeat-containing RNA/G-quadruplex-forming sequences cause genome-wide alteration of gene expression in human cancer cells in vivo. *Nucleic Acids Res.* **2015**, *43*, 2022–2032. [CrossRef] [PubMed]
34. Montero, J.J.; López de Silanes, I.; Graña, O.; Blasco, M.A. Telomeric RNAs are essential to maintain telomeres. *Nat. Commun.* **2016**, *7*, 12534. [CrossRef] [PubMed]
35. Redon, S.; Reichenbach, P.; Lingner, J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res.* **2010**, *38*, 5797–5806. [CrossRef]
36. Cusanelli, E.; Chartrand, P. Telomeric repeat-containing RNA TERRA: A noncoding RNA connecting telomere biology to genome integrity. *Front. Genet.* **2015**, *6*, 143. [CrossRef]
37. Bryan, T.M.; Englezou, A.; Gupta, J.; Bacchetti, S.; Reddel, R.R. Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J.* **1995**, *14*, 4240–4248. [CrossRef]
38. Silva, B.; Arora, R.; Bione, S.; Azzalin, C.M. TERRA transcription destabilizes telomere integrity to initiate break-induced replication in human ALT cells. *Nat. Commun.* **2021**, *12*, 3760. [CrossRef]
39. Largy, E.; Mergny, J.-L.; Gabelica, V. Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. In *The Alkali Metal Ions: Their Role for Life*; Springer: Cham, Switzerland, 2016; pp. 203–258. [CrossRef]
40. Martadinata, H.; Phan, A.T. Structure of human telomeric RNA (TERRA): Stacking of two G-quadruplex blocks in K+ solution. *Biochemistry* **2013**, *52*, 2176–2183. [CrossRef]
41. Xu, Y.; Suzuki, Y.; Ito, K.; Komiyama, M. Telomeric repeat-containing RNA structure in living cells. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14579–14584. [CrossRef]
42. Laguerre, A.; Hukezalie, K.; Winckler, P.; Katranji, F.; Chanteloup, G.; Pirrotta, M.; Perrier-Cornet, J.-M.; Wong, J.M.Y.; Monchaud, D. Visualization of RNA-Quadruplexes in Live Cells. *J. Am. Chem. Soc.* **2015**, *137*, 8521–8525. [CrossRef]
43. Yang, S.Y.; Lejault, P.; Chevrier, S.; Boidot, R.; Robertson, A.G.; Wong, J.M.Y.; Monchaud, D. Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.* **2018**, *9*, 4730. [CrossRef]
44. Kwok, C.K.; Marsico, G.; Balasubramanian, S. Detecting RNA G-Quadruplexes (rG4s) in the Transcriptome. *Cold Spring Harb. Perspect. Biol.* **2018**, *10*, a032284. [CrossRef]
45. Lyu, K.; Chow, E.Y.-C.; Mou, X.; Chan, T.-F.; Kwok, C.K. RNA G-quadruplexes (rG4s): Genomics and biological functions. *Nucleic Acids Res.* **2021**, *49*, 5426–5450. [CrossRef]
46. Joachimi, A.; Benz, A.; Hartig, J.S. A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorg. Med. Chem.* **2009**, *17*, 6811–6815. [CrossRef]
47. Zhang, A.Y.Q.; Bugaut, A.; Balasubramanian, S. A Sequence-Independent Analysis of the Loop Length Dependence of Intramolecular RNA G-Quadruplex Stability and Topology. *Biochemistry* **2011**, *50*, 7251–7258. [CrossRef]
48. Bugaut, A.; Balasubramanian, S. A Sequence-Independent Study of the Influence of Short Loop Lengths on the Stability and Topology of Intramolecular DNA G-Quadruplexes. *Biochemistry* **2008**, *47*, 689–697. [CrossRef] [PubMed]
49. Zhang, D.-H.; Fujimoto, T.; Saxena, S.; Yu, H.-Q.; Miyoshi, D.; Sugimoto, N. Monomorphic RNA G-Quadruplex and Polymorphic DNA G-Quadruplex Structures Responding to Cellular Environmental Factors. *Biochemistry* **2010**, *49*, 4554–4563. [CrossRef] [PubMed]
50. Collie, G.W.; Haider, S.M.; Neidle, S.; Parkinson, G.N. A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.* **2010**, *38*, 5569–5580. [CrossRef] [PubMed]
51. Pagano, B.; Mattia, C.A.; Cavallo, L.; Uesugi, S.; Giancola, C.; Fraternali, F. Stability and Cations Coordination of DNA and RNA 14-Mer G-Quadruplexes: A Multiscale Computational Approach. *J. Phys. Chem. B* **2008**, *112*, 12115–12123. [CrossRef]
52. Collie, G.; Reszka, A.P.; Haider, S.M.; Gabelica, V.; Parkinson, G.N.; Neidle, S. Selectivity in small molecule binding to human telomeric RNA and DNA quadruplexes. *Chem. Commun.* **2009**, *48*, 7482–7484. [CrossRef] [PubMed]
53. Di Antonio, M.; Biffi, G.; Mariani, A.; Raiber, E.-A.; Rodriguez, R.; Balasubramanian, S. Selective RNA Versus DNA G-Quadruplex Targeting by In Situ Click Chemistry. *Angew. Chem. Int. Ed.* **2012**, *51*, 11073–11078. [CrossRef] [PubMed]
54. Martadinata, H.; Phan, A.T. Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K+ solution. *J. Am. Chem. Soc.* **2009**, *131*, 2570–2579. [CrossRef] [PubMed]
55. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2009**, *31*, 455–461. [CrossRef]
56. Brenke, R.; Kozakov, D.; Chuang, G.-Y.; Beglov, D.; Hall, D.; Landon, M.R.; Mattos, C.; Vajda, S. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25*, 621–627. [CrossRef]

57. 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. [CrossRef]
58. Santos, T.; Salgado, G.F.; Cabrita, E.J.; Cruz, C. G-Quadruplexes and Their Ligands: Biophysical Methods to Unravel G-Quadruplex/Ligand Interactions. *Pharmaceuticals* **2021**, *14*, 769. [CrossRef]
59. 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.* **2011**, *50*, 10645–10648. [CrossRef]
60. Giancola, C.; Pagano, B. Energetics of ligand binding to G-quadruplexes. *Top. Curr. Chem.* **2013**, *330*, 211–242. [CrossRef]
61. Pomeislová, A.; Vrzal, L.; Kozák, J.; Dobiaš, J.; Hubálek, M.; Dvořáková, H.; Reyes-Gutiérrez, P.E.; Teplý, F.; Veverka, V. Kinetic Target-Guided Synthesis of Small-Molecule G-Quadruplex Stabilizers. *Chem. Open* **2020**, *9*, 1236–1250. [CrossRef]
62. 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*, 183–195. [CrossRef]
63. Duskova, K.; Lamarche, J.; Amor, S.; Caron, C.; Queyriaux, N.; Gaschard, M.; Penouilh, M.-J.; de Robillard, G.; Delmas, D.; Devillers, C.H.; et al. Identification of Three-Way DNA Junction Ligands through Screening of Chemical Libraries and Validation by Complementary in Vitro Assays. *J. Med. Chem.* **2019**, *62*, 4456–4466. [CrossRef]
64. Luo, Y.; Granzhan, A.; Verga, D.; Mergny, J. FRET-MC: A fluorescence melting competition assay for studying G4 structures in vitro. *Biopolymers* **2021**, *112*, e23415. [CrossRef]
65. Hsu, S.-T.D.; Varnai, P.; Bugaut, A.; Reszka, A.P.; Neidle, S.; Balasubramanian, S. A G-Rich Sequence within the c-kit Oncogene Promoter Forms a Parallel G-Quadruplex Having Asymmetric G-Tetrad Dynamics. *J. Am. Chem. Soc.* **2009**, *131*, 13399–13409. [CrossRef]
66. 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 + solution †. *Nucleic Acids Res.* **2007**, *35*, 6517–6525. [CrossRef] [PubMed]
67. 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*, 2048–2058. [CrossRef]
68. Dai, J.; Carver, M.; Yang, D. Polymorphism of human telomeric quadruplex structures. *Biochimie* **2008**, *90*, 1172–1183. [CrossRef] [PubMed]
69. Chen, S.; Shen, X. Long noncoding RNAs: Functions and mechanisms in colon cancer. *Mol. Cancer* **2020**, *19*, 167. [CrossRef]
70. Matsumura, K.; Kawasaki, Y.; Miyamoto, M.; Kamoshida, Y.; Nakamura, J.; Negishi, L.; Suda, S.; Akiyama, T. The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. *Oncogene* **2017**, *36*, 1191–1199. [CrossRef]
71. Shahid, R.; Bugaut, A.; Balasubramanian, S. The BCL-2 5′ untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry* **2010**, *49*, 8300–8306. [CrossRef] [PubMed]
72. Rocca, R.; Moraca, F.; Costa, G.; Nadai, M.; Scalabrin, M.; Talarico, C.; Distinto, S.; Maccioni, E.; Ortuso, F.; Artese, A.; et al. Identification of G-quadruplex DNA/RNA binders: Structure-based virtual screening and biophysical characterization. *Biochim. Biophys. Acta Gen. Subj.* **2017**, *1861*, 1329–1340. [CrossRef] [PubMed]
73. Monchaud, D.; Allain, C.; Teulade-Fichou, M.-P. Development of a fluorescent intercalator displacement assay (G4-FID) for establishing quadruplex-DNA affinity and selectivity of putative ligands. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4842–4845. [CrossRef]
74. 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. [CrossRef]
75. Suss, O.; Motiei, L.; Margulies, D. Broad Applications of Thiazole Orange in Fluorescent Sensing of Biomolecules and Ions. *Molecules* **2021**, *26*, 2828. [CrossRef] [PubMed]
76. Guo, Q.; Lu, M.; Marky, L.A.; Kallenbach, N.R. Interaction of the dye ethidium bromide with DNA containing guanine repeats. *Biochemistry* **1992**, *31*, 2451–2455. [CrossRef]
77. Sengupta, B.; Pahari, B.; Blackmon, L.; Sengupta, P.K. Prospect of Bioflavonoid Fisetin as a Quadruplex DNA Ligand: A Biophysical Approach. *PLoS ONE* **2013**, *8*, e65383. [CrossRef] [PubMed]
78. Roy, S.; Ali, A.; Kamra, M.; Muniyappa, K.; Bhattacharya, S. Specific stabilization of promoter G-Quadruplex DNA by 2,6-disubstituted amidoanthracene-9,10-dione based dimeric distamycin analogues and their selective cancer cell cytotoxicity. *Eur. J. Med. Chem.* **2020**, *195*, 112202. [CrossRef]
79. Jana, J.; Mondal, S.; Bhattacharjee, P.; Sengupta, P.; Roychowdhury, T.; Saha, P.; Kundu, P.; Chatterjee, S. Chelerythrine down regulates expression of VEGFA, BCL2 and KRAS by arresting G-Quadruplex structures at their promoter regions. *Sci. Rep.* **2017**, *7*, 40706. [CrossRef]
80. Entzian, C.; Schubert, T. Studying small molecule–aptamer interactions using MicroScale Thermophoresis (MST). *Methods* **2016**, *97*, 27–34. [CrossRef]
81. Amato, J.; Platella, C.; Iachettini, S.; Zizza, P.; Musumeci, D.; Cosconati, S.; Pagano, A.; Novellino, E.; Biroccio, A.; Randazzo, A.; et al. Tailoring a lead-like compound targeting multiple G-quadruplex structures. *Eur. J. Med. Chem.* **2019**, *163*, 295–306. [CrossRef]
82. Catalano, R.; Moraca, F.; Amato, J.; Cristofari, C.; Rigo, R.; Via, L.D.; Rocca, R.; Lupia, A.; Maruca, A.; Costa, G.; et al. Targeting multiple G-quadruplex–forming DNA sequences: Design, biophysical and biological evaluations of indolo-naphthyridine scaffold derivatives. *Eur. J. Med. Chem.* **2019**, *182*, 111627. [CrossRef]

83. Arora, R.; Lee, Y.; Wischnewski, H.; Brun, C.M.; Schwarz, T.; Azzalin, C.M. RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat. Commun.* **2014**, *5*, 5220. [CrossRef]
84. Pompili, L.; Leonetti, C.; Biroccio, A.; Salvati, E. Diagnosis and treatment of ALT tumors: Is Trabectedin a new therapeutic option? *J. Exp. Clin. Cancer Res.* **2017**, *36*, 189. [CrossRef]
85. Deng, Z.; Wang, Z.; Stong, N.; Plasschaert, R.; Moczan, A.; Chen, H.-S.; Hu, S.; Wikramasinghe, P.; Davuluri, R.V.; Bartolomei, M.S.; et al. A role for CTCF and cohesin in subtelomere chromatin organization, TERRA transcription, and telomere end protection. *EMBO J.* **2012**, *31*, 4165–4178. [CrossRef] [PubMed]
86. Walters, W.P.; Stahl, M.T.; Murcko, M.A. Virtual screening—An overview. *Drug Discov. Today* **1998**, *3*, 160–178. [CrossRef]
87. Amato, J.; Iaccarino, N.; Pagano, B.; Morigi, R.; Locatelli, A.; Leoni, A.; Rambaldi, M.; Zizza, P.; Biroccio, A.; Novellino, E.; et al. Bis-indole derivatives with antitumor activity turn out to be specific ligands of human telomeric G-quadruplex. *Front. Chem.* **2014**, *2*, 54. [CrossRef]
88. Cantor, C.R.; Warshaw, M.M.; Shapiro, H. Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleolides. *Biopolymers* **1970**, *9*, 1059–1077. [CrossRef]
89. Feretzaki, M.; Lingner, J. A practical qPCR approach to detect TERRA, the elusive telomeric repeat-containing RNA. *Methods* **2017**, *114*, 39–45. [CrossRef] [PubMed]
90. MacKenzie, D.; Watters, A.K.; To, J.T.; Young, M.W.; Muratori, J.; Wilkoff, M.H.; Abraham, R.G.; Plummer, M.M.; Zhang, D. ALT Positivity in Human Cancers: Prevalence and Clinical Insights. *Cancers* **2021**, *13*, 2384. [CrossRef]

PAPER III

Balancing affinity, selectivity, and cytotoxicity of hydrazone-based G-quadruplex ligands for activation of interferon *β* genes in cancer cells

Marzano S., Miglietta G., Morigi R., Marinello J., Arleo A., Procacci M., Locatelli A., Leoni A., Pagano B., Randazzo A., Amato J., Capranico G.

Journal of Medicinal Chemistry, 2022, Vol. 65(18)

Journal of Medicinal Chemistry

pubs.acs.org/jmc

Article

Balancing Affinity, Selectivity, and Cytotoxicity of Hydrazone-Based G-Quadruplex Ligands for Activation of Interferon *β* Genes in Cancer Cells

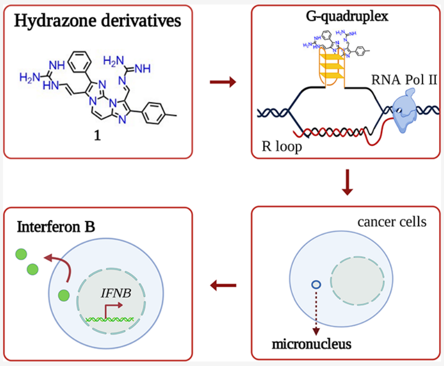
Simona Marzano, Giulia Miglietta, Rita Morigi,* Jessica Marinello, Andrea Arleo, Monica Procacci, Alessandra Locatelli, Alberto Leoni, Bruno Pagano, Antonio Randazzo, Jussara Amato,* and Giovanni Capranico*

Cite This: *J. Med. Chem.* 2022, 65, 12055−12067

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: G-quadruplex (G4) ligands are investigated to discover new anticancer drugs with increased cell-killing potency. These ligands can induce genome instability and activate innate immune genes at non-cytotoxic doses, opening the discovery of cytostatic immune-stimulating ligands. However, the interplay of G4 affinity/selectivity with cytotoxicity and immune gene activation is not well-understood. We investigated a series of closely related hydrazone derivatives to define the molecular bases of immune-stimulation activity. Although they are closely related to each other, such derivatives differ in G4 affinity, cytotoxicity, genome instability, and immune gene activation. Our findings show that G4 affinity of ligands is a critical feature for immune gene activation, whereas a high cytotoxic potency interferes with it. The balance of G4 stabilization *versus* cytotoxicity can determine the level of immune gene activation in cancer cells. Thus, we propose a new rationale based on low cell-killing potency and high immune stimulation to discover effective anticancer G4 ligands.

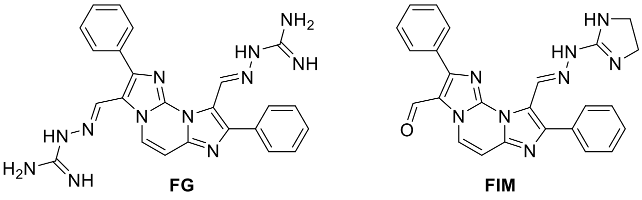


INTRODUCTION

G-quadruplex (G4) ligands are actively investigated to discover new effective anticancer drugs as G4s, non-canonical DNA structures, are considered promising targets.[1−4] Despite the large number of specific ligands developed, none has however shown efficacy in cancer patients and very few have reached early phases of clinical trials.[5,6] In line with the standard drug discovery rationale, several laboratories have previously searched for G4 binders with a high cell-killing potency.[1−5] Interestingly, we have recently demonstrated that the G4 binders pyridostatin (**PDS**) and PhenDC3 can effectively elicit an innate immune gene response [activation of interferon *β* (IFN-B) gene and IFN-B-dependent pathways] in human cancer cells, mediated by micronuclei accumulation at non-cytotoxic concentrations.[7] As recent advances clearly point to the potential of harnessing innate immunity for cancer immunotherapy,[8−12] non-cytotoxic immune-modulators may optimize immunotherapy in unresponsive cancers while having a marginal toxicity against proliferating normal cells. Thus, our recent findings[7] indicate that G4 ligands may be exploited as cytostatic immune-stimulating agents for anticancer immune-therapeutic combinations.[5] In particular, G4 binders can increase micronuclei,[7,13,14] which can be a source of cytoplasmic DNA that is able to induce the cGAS−STING pathway and activate innate immune genes.[7,15,16] However, the relationships among G4 affinity/selectivity, cell-killing potency, and genome instability determining a high level of immune gene activation by G4 ligands remains to be established.

Here, to answer this question, we have focused on a highly homogenous series of new compounds able to selectively target G4s. In 2010, some of us identified **FG** (Chart 1), a bis-guanylhydrazone derivative of diimidazo[1,2-*a*:1,2-*c*]-pyrimidine, as a potent and selective G4 stabilizer.[17] Then, we identified highly selective analogues with a preference for parallel G4 topology and ability to stabilize G4s in living cancer

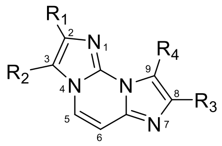
Chart 1. Chemical Structures of the Lead Compounds FG and FIM (1 and 3 by Amato *et al.*,[18] respectively)



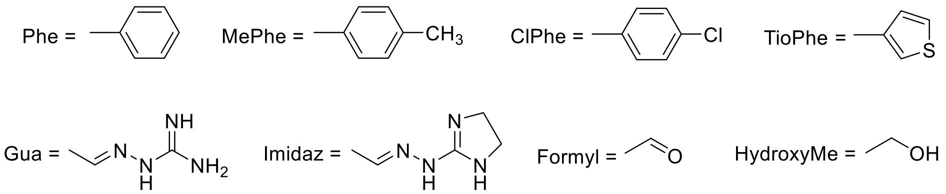
FG **FIM**

Received: May 16, 2022
Published: September 8, 2022

Chart 2. Chemical Structures of New FG and FIM Derivatives Synthesized in This Study



1-20



Comp	R_1	R_2	R_3	R_4
1	Phe	Gua	MePhe	Gua
2	Phe	Imidaz	MePhe	Imidaz
3	MePhe	Gua	Phe	Gua
4	MePhe	Imidaz	Phe	Imidaz
5	MePhe	Gua	MePhe	Gua
6	MePhe	Imidaz	MePhe	Imidaz
7	Phe	Gua	ClPhe	Gua
8	Phe	Imidaz	ClPhe	Imidaz
9	ClPhe	Gua	Phe	Gua
10	ClPhe	Imidaz	Phe	Imidaz
11	ClPhe	Gua	ClPhe	Gua
12	ClPhe	Imidaz	ClPhe	Imidaz
13	TioPhe	Imidaz	TioPhe	Imidaz
14	MePhe	Formyl	MePhe	Imidaz
15	Phe	Formyl	Phe	Gua
16	TioPhe	Formyl	TioPhe	Gua
17	MePhe	Formyl	Phe	Gua
18	MePhe	Formyl	MePhe	Gua
19	Phe	HydroxyMe	Phe	Imidaz
20	Phe	HydroxyMe	Phe	Gua

cells,[18] including **FG** and **FIM** (compounds **1** and **3** by Amato *et al.*,[18] respectively, Chart 1). The results established the diimidazo[1,2-*a*:1,2-*c*]pyrimidine core as a scaffold of selective G4 ligands and showed that both the iminoguanidine (Gua) and hydrazinoimidazoline (Imidaz) nitrogen chains are effective in achieving G4 binding properties. Then, **FG** was shown to induce DNA damage and micronuclei in human osteosarcoma U2OS cells in an R loop-dependent manner.[13]

As these agents are specific and effective G4 binders, we have now synthesized new close derivatives of **FG** and **FIM** to investigate the structural features eliciting a high immune gene activation relative to the cell-killing potency. The findings show that a proper balance between G4 affinity/selectivity and cytotoxicity is critical for immune gene activation in cancer cells.

■ RESULTS

Design of New Hydrazone-Based Compounds. In order to improve affinity and selectivity toward G4 structures and finely tune the biological effects of close **FG** and **FIM** analogues, we designed and synthesized a new series of molecules having different electron distribution and similar steric hindrance. For this purpose, the diimidazo[1,2-*a*:1,2-*c*]pyrimidine core was maintained unaltered and a chlorine or a methyl group was inserted at the para position of one or both the pending phenyl rings. In fact, chlorine and methyl have almost the same steric hindrance but opposite inductive effects, methyl being an electron donor group while chlorine has an electron withdrawing inductive effect. Both Gua and Imidaz moieties were considered as positively charged chains, either to obtain **FG** analogues (compounds **1**−**12**, Chart 2) or **FIM**

Table 1. Compound-Induced Thermal Stabilization of G4 and haipin structures Measured by CD Melting Experiments

	$\Delta T_{1/2}$ (°C)[a]					
Comp	*tel*$_{26}$	*c-kit1*	*c-kit2*	*c-myc*	*hairpin*	no. of positive charges
FG[b]	−4.5	>15.0	>20	>20	0.1[c]	2
1	4.8 (±0.2)	18.2 (±0.2)	>30[d]	16.9 (±0.2)	1.3 (±0.3)	2
2	5.3 (±0.2)	15.2 (±0.2)	>30[d]	9.1 (±0.3)	1.1 (±0.4)	2
3	6.3 (±0.4)	25.2 (±0.5)	>30[d]	16.9 (±0.4)	3.2 (±0.3)	2
4	6.2 (±0.2)	21.4 (±0.4)	>30[d]	13.1 (±0.6)	2.2 (±0.2)	2
5	5.5 (±0.2)	18.5 (±0.3)	>30[d]	15.6 (±0.2)	4.1 (±0.3)	2
6	4.5 (±0.3)	17.1 (±0.2)	26.5 (±0.4)	8.5 (±0.2)	2.4 (±0.2)	2
7	5.0 (±0.2)	21.4 (±0.4)	>30[d]	17.6 (±0.4)	2.3 (±0.2)	2
8	6.3 (±0.2)	15.5 (±0.3)	22.0 (±0.3)	7.4 (±0.3)	1.2 (±0.4)	2
9	8.3 (±0.2)	20.1 (±0.3)	>30[d]	14.3 (±0.2)	2.6 (±0.2)	2
10	6.6 (±0.2)	17.7 (±0.3)	19.2 (±0.3)	10.6 (±0.3)	1.2 (±0.3)	2
11	4.1 (±0.4)	11.1 (±0.2)	>30[d]	11.4 (±0.2)	2.9 (±0.2)	2
12	3.3 (±0.2)	9.6 (±0.2)	13.7 (±0.3)	5.5 (±0.2)	3.5 (±0.2)	2
13	3.8 (±0.3)	10.3 (±0.2)	14.9 (±0.3)	4.7 (±0.2)	2.0 (±0.2)	2
FIM[b]	−3.0	2.7	9.5	>20	−0.8[c]	1
14	1.3 (±0.2)	3.8 (±0.2)	6.8 (±0.2)	1.5 (±0.2)	0.0 (±0.3)	1
15	4.3 (±0.2)	11.2 (±0.2)	24.2 (±0.4)	5.8 (±0.3)	1.7 (±0.2)	1
16	4.8 (±0.3)	16.1 (±0.2)	25.2 (±0.3)	9.1 (±0.3)	2.7 (±0.2)	1
17	3.8 (±0.2)	6.6 (±0.3)	20.8 (±0.5)	3.1 (±0.2)	1.2 (±0.2)	1
18	2.1 (±0.2)	6.1 (±0.2)	16.2 (±0.2)	4.6 (±0.2)	2.3 (±0.3)	1
19	2.0 (±0.2)	7.1 (±0.2)	14.7 (±0.3)	1.7 (±0.2)	1.3 (±0.2)	1
20	1.8 (±0.3)	9.7 (±0.2)	18.6 (±0.4)	2.8 (±0.2)	2.3 (±0.3)	1

[a] $\Delta T_{1/2}$ represents the difference in melting temperature [$\Delta T_{1/2} = T_{1/2}$ (DNA + 2 ligand equiv) − $T_{1/2}$ (DNA)]. The $T_{1/2}$ values of DNA alone are: *c-kit1* = 54.0 ± 0.5 °C, *c-kit2* = 61.5 ± 0.5 °C, *c-myc* = 72.0 ± 0.5 °C, *tel*$_{26}$ = 47.9 ± 0.5 °C, and *hairpin* = 75.4 ± 0.2 °C. All experiments were performed in duplicate, and the reported values are the average of two measurements. [b] Data from ref 18. [c] A self-complementary 12-mer duplex-forming sequence was used as a duplex model. [d] ΔT1/2 could not be accurately determined as the compound increases hugely the thermal stability of c-kit2.

analogues (compounds **14**−**18**, Chart 2). In addition, since an **FG** analogue bearing thiophenes instead of phenyl groups proved to be a good G4 binder,[18] we also considered this kind of modification along with the replacement of the Gua chains with the Imidaz ones (compound **13**, Chart 2). Finally, the formyl group of **FIM** was replaced with a primary alcohol group, which is able to either accept or donate hydrogen bonds (compounds **19** and **20**, Chart 2). The complete synthesis of the derivatives is described in Supporting Information (Scheme S1 and Table S1).

Circular Dichroism Experiments. The stabilizing effects of compounds **1**−**20** on G4 structures formed by the G-rich DNA sequences from the nuclease hypersensitive region of the c-KIT (*c-kit1* and *c-kit2*) and c-MYC (*c-myc*) gene promoters as well as from the human telomeric sequence (*tel*$_{26}$) were analyzed by circular dichroism (CD) melting experiments. These DNA sequences were chosen for their ability to adopt different G4 topologies, characterized by parallel (*c-kit1*, *c-kit2*, and *c-myc*) or hybrid (*tel*$_{26}$) arrangements.[19−21] Consistently, CD spectra of *c-kit1*, *c-kit2*, and *c-myc* displayed a positive band at 264 nm and a negative one around 240 nm (Figure S1), which are characteristics of parallel-stranded G4 topologies.[22] On the other hand, *tel*$_{26}$ showed a positive band at 290 with a shoulder at *ca.* 268 nm and a weak negative band at around 240 nm (Figure S1), confirming the presence of a hybrid structure as the main conformation. A 27 residue-long hairpin-forming oligonucleotide (*hairpin*) was also used to evaluate the selectivity of the new analogues for G4s over a duplex. CD spectra of *hairpin* showed a positive band at around 280 nm and a negative one at ∼250 nm, confirming the formation of a duplex (Figure S1). Additional CD spectra were recorded to examine the potential of compounds **1**−**20** to modify the native folding topology of these G4s. DNA/ligand mixtures were prepared by adding each ligand (2 molar equiv) to folded G4 or hairpin structures. No significant variations in the CD signal were observed for any of the analyzed DNA structures (Figures S2−S6), suggesting no G4 topology changes upon addition of compounds. Then, their ability to bind and stabilize the DNA structures was evaluated by CD melting experiments measuring the compound-induced change in the apparent melting temperature ($\Delta T_{1/2}$) of G4 and duplex structures. CD melting curves of DNA with and without each ligand were obtained by following the variations of the intensity of the CD signals at 264, 290, and 252 nm for parallel G4s, hybrid G4, and duplex, respectively (Figures S7−S11). The results show that all compounds are good G4 stabilizers (with one exception, **14**) showing a higher preference for parallel than hybrid G4s (Table 1). In addition, as expected for ligands having the same core but different numbers of positive charges, the greater the charge number, the stronger the stabilizing effect on G4s (Table 1). However, compound interactions with dsDNA (*hairpin*) were also slightly increased by positive charges (Table 1). Thus, to assess the selectivity for G4 structures of this series of compounds, we selected the analogues showing a strong stabilizing effect on at least two G4s and a negligible effect ($\Delta T_{1/2}$ < 2.0 °C) on the *hairpin*, that is, compounds **1**, **2**, **8**, and **10** among those with two positively charged side chains and **15**, **19**, and **20** among those with a positive charge only.

FRET Melting Experiments. The Förster resonance energy transfer (FRET) methodology[23] was used to further evaluate G4-stabilizing properties and G4 *versus* duplex selectivity of **1**, **2**, **8**, **10**, **15**, **19**, and **20**. In this assay, the G4-forming *c-kit1* oligonucleotide labeled with FAM (F) and TAMRA (T) at the 5′ and 3′ ends, respectively, was employed (*F-c-kit1-T*) since, among the G4s more stabilized by these ligands, *c-kit1* is the one that has the lowest $T_{1/2}$ value, thus allowing to better estimate the stabilizing properties of different ligands and evaluate their ability to discriminate between G4 and duplex structures. Indeed, as the target G4 was the only labeled molecule, it was possible to evaluate the ligand selectivity by adding a large excess of the *hairpin* oligonucleotide (unlabeled competitor). Therefore, the ability of the investigated compounds to selectively stabilize the G4 was evaluated by measuring the effect of the presence of various concentrations of the competitor on the $\Delta T_{1/2}$ of the G4 in the presence of 2 molar equiv of each ligand. Results of these experiments (Figure S12 and Table 2) confirm that the selected compounds are efficient G4 stabilizers. However, in the case of compounds **1**, **2**, **8**, and **10** (carrying two positively charged side chains), G4/ligand interaction turned out to be somewhat challenged by the *hairpin* sequence being added in excess. This does not happen for **15**, **19,** and **20**, meaning that these compounds are more selective for G4 than the former.

Table 2. G4 Selectivity of the Selected Compounds[a]

Comp	$\Delta T_{1/2}$ (°C)[b]		
	F-c-kit1-T	*F-c-kit1-T + hairpin* (1:15)	*F-ckit1-T + hairpin* (1:50)
1	25.4 (±0.5)	22.2 (±0.5)	19.2 (±0.5)
2	22.4 (±0.5)	21.4 (±0.5)	20.2 (±0.5)
8	24.6 (±0.5)	21.4 (±0.5)	19.6 (±0.5)
10	26.6 (±0.5)	24.8 (±0.5)	23.8 (±0.5)
15	22.0 (±1.0)	21.6 (±1.0)	22.2 (±1.0)
19	8.0 (±0.4)	7.8 (±0.4)	8.4 (±0.4)
20	10.0 (±0.4)	11.1 (±0.5)	12.1 (±0.5)

[a] G4/dsDNA competition determined by ligand-induced thermal stabilization of *F-c-kit1-T* G4 measured by FRET. $\Delta T_{1/2}$ values are the differences between the $T_{1/2}$ of *F-c-kit1-T* in the presence (2 molar equiv) and absence of the ligands, without or with large excess of unlabeled *hairpin* (15 and 50 molar equiv with respect to G4). [b] The $T_{1/2}$ of *F-ckit1-T* is 57.4 (±0.2) °C. All experiments were performed at least in duplicate, and the reported values are the average of the measurements. The differences between results of CD and FRET melting experiments could be explained with different DNA sequences and/or experimental conditions.

Fluorescence Intercalator Displacement Assay. To gain insight into the affinity of the selected compounds for G4s, fluorescence intercalator displacement (G4-FID) experiments were performed by using the light-up fluorescent probe thiazole orange (TO), which binds to the DNA structure of interest.[24] The competitive displacement of TO from DNA by candidate ligands was monitored, thus enabling the determination of their relative binding affinity to the structures under examination, namely, *c-kit1*, *c-kit2*, and *c-myc* G4s, which were selected as they turned out to be those most stabilized by the ligands. Dose−response curves were obtained by plotting the percentage of TO displacement versus the concentration of each compound (Figure S13), and the concentrations at which 50% displacement was achieved (DC_{50}) were calculated. The lower the DC_{50} value, the higher should be the affinity of the compound for the DNA structure. Results of G4-FID assay (Table 3) indicate a good TO displacement ability for compounds **1**, **2**, **8**, and **10**. These ligands exhibited almost similar results for the investigated parallel G4s, suggesting that ligand/G4 interaction is not sequence-specific. On the other hand, compounds **15**, **19**, and **20** were not able to reach 50% displacement in any case, suggesting that **FIM** derivatives have a lower affinity for G4s than **FG** ones. Therefore, the TO-displacing ability seems to be in direct correlation with the number of positive charges on the ligands: the highly cationic molecules are the most efficient TO displacers. As for **15**, the apparent discrepancy between the results of G4-FID assay and melting experiments (CD and FRET) could also be explained considering that this ligand may bind to G4s without strictly competing with the TO.[25]

Table 3. Ligand DC_{50} Values for *c-kit1*, *c-kit2*, and *c-myc* G4s Determined with G4-FID Assay

Comp	DC_{50} (μM)[a]		
	c-kit1	*c-kit2*	*c-myc*
1	1.4	1.4	1.6
2	1.3	1.2	1.3
8	2.7	1.7	1.8
10	1.2	1.4	1.3

[a] The error in DC_{50} values is ±5%.

Microscale Thermophoresis Assay. Quantitative data on the binding affinity of **1**, **2**, **8**, **10**, **15**, **19**, and **20** for the selected G4s were obtained by microscale thermophoresis (MST), which is a powerful method for the quantitative analysis of the interactions between small molecules and nucleic acids in solution.[26] To perform MST experiments, one of the binding partners must be fluorescent (either intrinsically fluorescent or conjugated to a given fluorophore). Therefore, serial dilutions of ligands were prepared, mixed with a constant concentration of Cy5.5-labeled G4s (*c-kit1*, *c-kit2*, and *c-myc*), and analyzed by MST. Results of the binding curves showed that the compounds bind to G4s with different affinity (Table 4 and Figures S14−S16). In particular, compounds showed higher affinity values for *c-kit1* and *c-kit2* than for *c-myc*, and a slight preference for *c-kit1* over *c-kit2*, except for **19** and **20**. Noteworthy, compound **1** turned out to be the strongest G4 binder of the series, showing K_d values in the nanomolar range for the interaction with *c-kit1* and *c-kit2* [K_d = 0.03 and 0.04 μM, respectively], while **19** and **20** turned out to be the worst of the series in terms of affinity for G4s.

Table 4. Equilibrium Dissociation Constants for the Binding of the Ligands to *c-kit1*, *c-kit2*, and *c-myc* G4s Obtained by MST Experiments[a]

Comp	K_d (μM)		
	c-kit1	*c-kit2*	*c-myc*
1	0.03 ± 0.01	0.04 ± 0.01	1.5 ± 0.4
2	0.12 ± 0.02	0.13 ± 0.02	n.d.
8	0.38 ± 0.09	0.49 ± 0.09	3.0 ± 1.0
10	0.07 ± 0.02	0.50 ± 0.04	1.2 ± 0.4
15	0.30 ± 0.04	0.37 ± 0.03	7.0 ± 1.0
19	8.0 ± 2.0	2.5 ± 0.3	37 ± 1
20	2.0 ± 0.2	1.3 ± 0.2	8.0 ± 3.1

[a] Dissociation constant values were obtained with MST experiments. Comp, compound. n.d., not determined.

Cytotoxicity of Selected Hydrazone Derivatives. Next, we determined the cytotoxic potencies of compounds with two (**FG**, **1**, **2**, and **8**) or one (**FIM**, **15**, **19**, and **20**) positively charged chain in human osteosarcoma U2OS and murine fibrosarcoma MNMCA1 cells following 24 h of treatments by using **PDS** as a reference compound (Table 5). We selected these two lines as the former has been used in several G4 studies, included ours,[13,14,18] and the latter murine line is known to produce high levels of IFN-B.[27] The results show that **FIM** and **15**, both bearing an aldehyde moiety, are the most cytotoxic compounds among those analyzed (Table 5). In particular, they exhibited IC_{50} values around 6-fold and 5−18-fold higher than those of **19** and **20**, respectively, indicating that an aldehyde moiety confers a greater cytotoxicity than an alcohol group. Among the **FG** analogues, **8** is more cytotoxic than **1** and **2**. The compounds have similar IC_{50} in both the two lines; however, imino-guanidine chains confer around twofold higher cytotoxic activity than 2-hydrazino-2-imidazoline chains in humans but not in murine cells (compare **1** *vs* **2**, and **20** *vs* **19**, Table 5). To better define the interplay among G4 affinity/selectivity, G4 stabilization in cells, induction of genome instability and activation of IFN-B, **FG** and **FIM** analogues were discussed separately. The tested analogues show similar cytotoxic potency in normal human MRC5 fibroblasts (Table S2) as expected for a cell assay which measures cell-killing effects against proliferative cells.

Table 5. Cytotoxic Potency of Selected Hydrazone Derivatives[a]

Comp	human U2OS	murine MNMCA1
FG	15.9 ± 1.2[b]	n.d.
1	46.8 ± 12.7	38.5 ± 5.8
2	108.3 ± 35.4	28.6 ± 8.4
8	20.2 ± 1.0	23.5 ± 5.1
FIM	4.0 ± 0.33	2.3 ± 0.92
15	2.6 ± 0.87	1.9 ± 0.33
19	24.5 ± 1.2	12.1 ± 2.5
20	14.3 ± 9.3	35.7 ± 0.67
PDS	>50[b]	27.0 ± 16.0

[a] Data are IC_{50} (μM, concentration inhibiting 50% of cell growth) of each compound in human osteosarcoma U2OS cells and murine fibrosarcoma MNMCA1 cells. Treatments were for 24 h in exponentially growing cells. Cell survival was evaluated with the MTT test after 48 h of cell recovery in drug-free medium. IC_{50} values are means ± SEM of two independent experiments performed in triplicate. [b] See ref 13. n.d., not determined.

G4 Stabilization and DNA Damage by Hydrazone Derivatives with Two Side Chains. Next, we have evaluated cellular effects of closely related analogues starting with compounds **1**, **2**, and **8**, which have two positively charged side chains (Figure 1). We used an immunofluorescence (IF) assay to determine their ability to stabilize G4 structures in U2OS cells using the BG4 antibody, which specifically binds to G4 structures,[13,28] and **PDS** as the positive control.[13] The results (Figure 1A) show that **1** and **2** can stabilize G4 structures (around 2.4-fold increase) in living cells whereas compound **8** was ineffective (0.77-fold change). Analogues **1** and **2** showed a G4 stabilization *in vivo* similar to that of **PDS** (Figure 1A) and much higher than that of **8**, suggesting that the latter has other or additional cytotoxic mechanisms. Overall, these results agree with a higher ligand-induced G4 thermal stabilization observed for analogues **1** and **2** than that for **8**, particularly for *c-kit2* and *c-myc* G4s (Tables 1 and 3). The complete lack of G4 stabilization with **8** in nuclei suggests that G4 binding in living cells may be affected by interactions with other cellular components.

Next, we determined DNA damage induced by **1**, **2**, and **8** by evaluating the levels of S139-phosphorylated histone H2AX (γH2AX) (Figure 1B). We treated U2OS cancer cells with compounds for 24 h, at equal cytotoxic concentrations. **PDS** (10 μM) was used a reference compound.[13] The results show that **2** and **8** increased γH2AX foci levels (1.45- and 1.37-fold increase, respectively) whereas **1** did not. Thus, as **8** can induce DNA damage (Figure 1B) even without stabilizing G4s in cell (Figure 1A), while compound **2** induces G4 stabilization, it is reasonable to speculate that the cytotoxicity mechanism is likely different between the two compounds. On the other hand, G4 stabilization may lead to different levels of DNA damage, likely depending on *in vivo* G4 targeting. In contrast to **2**, compound **1** stabilizes G4 structures in cells (Figure 1A) but does not promote DNA damage (Figure 1B).

Micronuclei and IFN-B Activation by Hydrazone Derivatives with Two Side Chains. As non-cytotoxic doses of G4 binders, **PDS** and PhenDC3, can activate IFN-B-dependent pathways through micronuclei induction in human cancer cells,[7] we next ask if the new analogues can also affect immune gene expression in cancer cells. First, we determined the induction of micronuclei. **1**, **2**, and **8** at similar cytotoxic concentrations (15 μM, corresponding to 0.4−0.65 of their IC_{50}s) can induce almost the same micronuclei levels in MNMCA1 cells but less than those of **PDS** (Figure 1C). Then, we measured the amount of IFN-B secreted by murine MNMCA1 cells into the medium with ELISA assay. In agreement with experimental conditions reported for **PDS** previously,[7] murine cells were treated for 24 h with two concentrations (15 and 30 μM) of compounds and then allowed to recover for 2 days in fresh medium (Figure 1D). The results show that **PDS** induced higher IFN-B levels than the tested analogues, consistently with higher numbers of **PDS**-stimulated micronuclei. Among the studied analogues, **1** was more effective in the induction of IFN-B than **2**, whereas **8** was completely ineffective (Figure 1D). In particular, **1** induced IFN-B production at higher levels at non-cytotoxic concentrations (15 μM) and **8** was ineffective even at concentrations higher (30 μM) than the IC_{50} (Figure 1D and Table 5). Thus, compound **1**, which induces *in vivo* G4 stabilization without promoting DNA damage, can activate IFN-B gene expression at non-cytotoxic concentrations.

G4 Stabilization and DNA Damage by Hydrazone Derivatives with One Side Chain. Next, we have evaluated cellular effects of closely related analogues with one positively charged side chain, **FIM**, **15**, **19**, and **20**. Although these analogues showed a markedly decreased affinity for the tested G4 structures with respect to two positively charged analogues (Tables 1−3), their cytotoxic potencies are equal or higher than those of the latter (Table 5). Therefore, we asked whether **FIM** analogues could trigger G4 stabilization and DNA damage similar to **FG** analogues. The results show that they are all good G4 stabilizers in living cells (Figure 2A). **15** and **20** induced a somewhat higher stabilization (2.32−2.45-fold change) than **FIM** and **19** (1.55−1.91-fold change), indicating that the Gua moiety favors G4−ligand interactions better than the 2-hydrazino-2-imidazoline group (see also Table 1) probably due to its higher flexibility. Then, we investigated the ability of these analogues to induce DNA damage under the experimental conditions described above for **FG** analogues. The results show that these analogues increased γH2AX levels

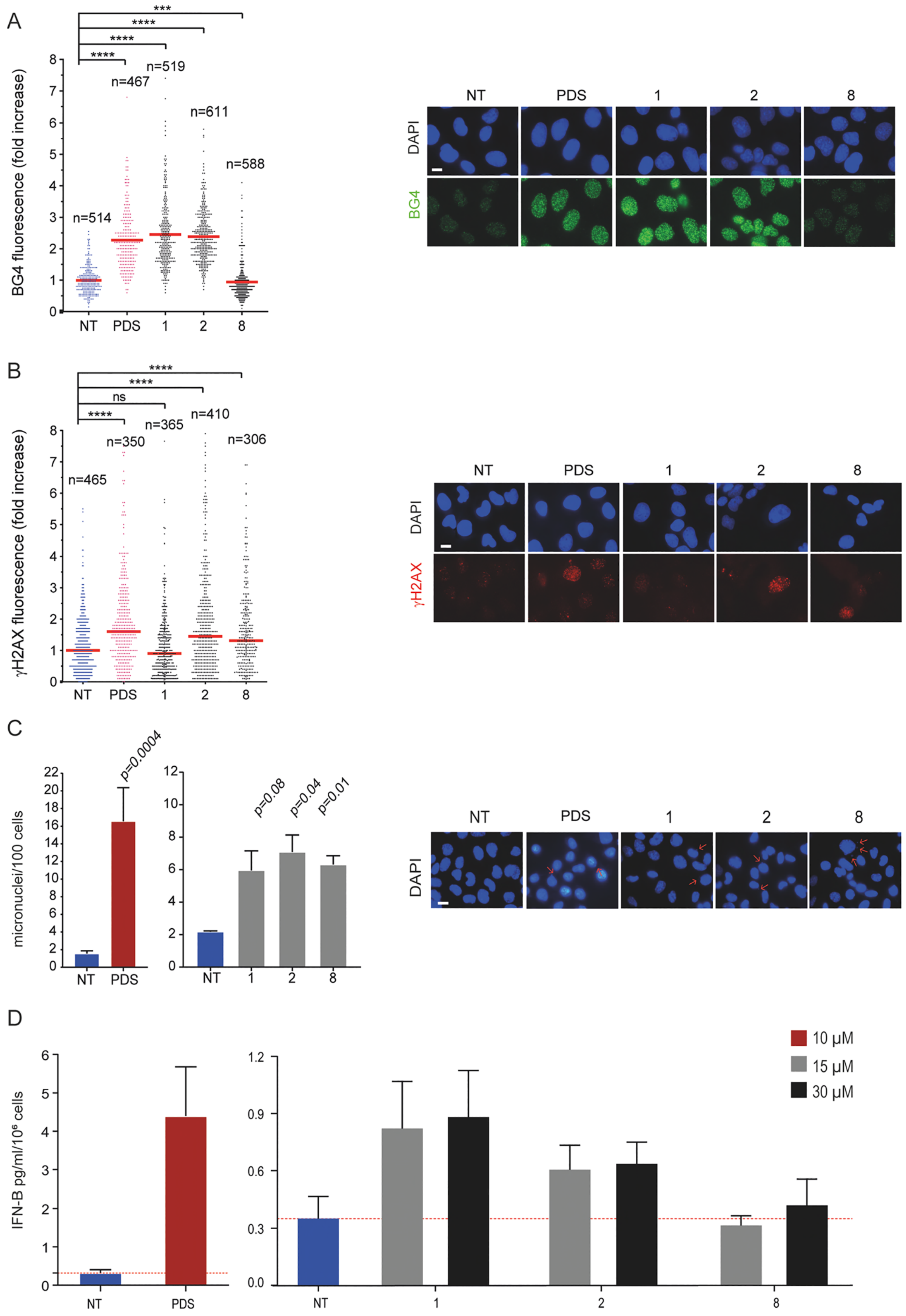


Figure 1. G4 stabilization, DNA damage, and IFN-B stimulation induced by **FG** derivatives. (A) Quantification of fluorescence signals of BG4 foci in U2OS cells being treated for 10 min with **PDS** or **FG** derivatives (compounds **1**, **2,** and **8**) at 10 μM concentration. The graph shows the fold increase reported as the mean ± SEM of three biological replicates, and the IF representative images are reported (left). (B) Quantification of

Figure 1. continued

fluorescence signals of γH2AX in U2OS cells being treated with **PDS** (10 μM) and **FG** derivatives at IC_{50} concentrations (46, 100, and 20 μM for **1**, **2**, and **8**, respectively) for 24 h of treatment. The graph shows the fold increase reported as the median ± SEM of two biological replicates and the IF representative images are reported (left). (C) Micronuclei quantification by DAPI staining in MNMCA1 cells treated (15 μM) after 24 h of treatment followed by 24 h of drug-free recovery. **PDS** (10 μM)-treated cells are also shown. The graph shows the mean ± SEM of two biological replicates, and the IF representative images are reported (left). Above the bar chart, the *p*-value are reported. The scale bar is 10 μm. (D) Quantification of IFN-B produced by MNMCA1 cells treated with **FG** derivatives at different concentrations (15 and 30 μM). **PDS** (10 μM)-treated cells are also shown. The IFN-B detection was performed with ELISA assay after 24 h of compounds treatment followed by 48 h of recovery. The bar chart reports the mean ± SEM of three biological replicates. Significance in all the graph was calculated by Mann−Witney test (*$p < 0.05$, **$p > 0.01$, ***$p > 0.001$, and ****$p < 0.0001$).

at similar levels in cancer cells, even though **15** was somewhat less effective (Figure 2B). As **FIM** and **15** were more cytotoxic than **19** and **20**, DNA damage features of the former are likely more lethal than those of the latter.

Micronuclei and IFN-B Activation by Hydrazone Derivatives with One Side Chain. Similar to **FG** analogues (Figure 1), we then tested the **FIM** analogues for the induction of micronuclei and the activation of IFN-B genes in murine MNMCA1 cells by using sub-cytotoxic concentrations (Figure 2C,D). The results show that **FIM**, **15**, **19**, and **20** induced a 2.5−4.0-fold increase in micronuclei levels in comparison to untreated cells with little difference among them (Figure 2C). Overall, the **FIM** analogues did not affect significantly IFN-B expression, showing a low, if any, with a maximum of less than twofold change for **FIM** (Figure 2D). No difference was observed between derivatives bearing Gua or 2-hydrazino-2-imidazoline groups as chains. Overall, the results indicate that analogues with one positively charged side chain were less effective in activating the IFN-B gene expression than the two positively charged analogues (Figure 1). The effect on gene expression was thus correlated with G4 affinity of the studied analogues.

DISCUSSION

Hydrazone-based compounds, including **FG** and **FIM** (Chart 1), are known to have a high selectivity for G4 structures relative to duplex DNA and to induce DNA damage and genome instability.[14,17,18] Here, we provide evidence that these agents can activate IFN-B gene expression in cancer cells at non-cytotoxic doses, therefore pointing to the exploitation of hydrazone-based G4 ligands as immunomodulating agents. In particular, **1** can be considered as a core structure for further analyses aiming at establishing a hit ligand with immune-stimulating anticancer activity.

In vivo G4 selectivity of structurally different G4 ligands is substantially unknown as the number and types of G4 structures in a living cell can be very high.[3−5] In addition, a ligand can have more molecular interactions affecting its biological outcome, in particular the cell-killing potency. Thus, our present investigation has been focused on very closely related analogues to minimize putative variations of unpredictable molecular interactions. In particular, a new series of **FG** and **FIM** having different electron distribution and similar steric hindrance were designed and synthesized. For this purpose, the diimidazo[1,2-*a*:1,2-*c*]pyrimidine core was maintained unaltered and a chlorine or a methyl group was inserted at the para-position of one or both the pending phenyl rings. In fact, chlorine and methyl have almost the same steric hindrance but opposite inductive effects, methyl being an electron donor group while chlorine has an electron-withdrawing inductive effect. Both Gua and Imidaz moieties were considered as positively charged chains, either to obtain **FG** analogues (compounds **1**−**12**, Chart 2) or **FIM** analogues (compounds **14**−**18**). In addition, since an **FG** analogue bearing thiophenes instead of the phenyl groups proved to be a good G4 binder,[14] we also considered this kind of modification along with the replacement of the Gua chains with Imidaz ones (compound **13**). Finally, the formyl group of **FIM** was replaced with a primary alcohol group, which is able to either accept or donate hydrogen bonds (compounds **19** and **20**).

The G4 binding properties of **1**−**20** in terms of either G4 stabilization, affinity, and selectivity over the duplex structure were measured by means of several biophysical techniques, including CD, G4-FID, MST, and competition FRET-melting. We used the results of CD melting experiments to select the best binders from **FG** and **FIM** series. In particular, the ligands were chosen on the basis of their selectivity for G4 over the duplex, that is, those compounds showing the most negligible effects on the hairpin−duplex model (compounds **1**, **2**, **8**, and **10** belonging to the **FG** series; **15** and **17** belonging to the **FIM** series and exhibiting the formyl group; and **19** and **20** in which the formyl group of **FIM** is reduced to the corresponding hydroxyl group). Next, among these compounds, we selected, within each series, those that showed the greatest stabilizing effects on at least two G4s and differed in the presence of Gua or Imidaz pendant groups, that is, compounds **1** and **2** as **FG** analogues (with Gua and Imidaz substituents, respectively), compound **15** among the **FIM** derivatives (with a formyl group in R_2 and a Gua substituent in R_4), and compounds **19** and **20** among the hydroxyl group-containing **FIM** derivatives (carrying a Gua and an Imidaz pendant group in R_4, respectively). Since compounds **8** and **10** belonging to the **FG** series also showed good stabilizing properties on the investigated G4s, we decided to include them in further biophysical assays aimed at assessing the selectivity of the ligands (FRET) and their affinity for G4s (G4-FID assay).

The results of these studies confirmed that compounds **1**, **2**, **8**, **10**, and **15** are stronger G4 stabilizers than **19** and **20**, and revealed that compounds with one positively charged side chain (*i.e.*, **15**, **19**, and **20**) have less affinity for G4s but are more selective binders compared to those having two positively charged side chains (**1**, **2**, **8**, and **10**), with compounds **1**, **2**, and **10** being the most efficient TO displacers. Despite the high chemical similarity between **8** and **10** (they differ in the inversion of the Phe and ClPhe substituents in R_1 and R_3), only compound **10** performed similarly to **1** and **2**, while compound **8** showed a slightly lower G4 affinity. These results were also confirmed by MST experiments, which allowed to evaluate the affinity of the ligands for the G4s. Indeed, compound **1** turned out to be the

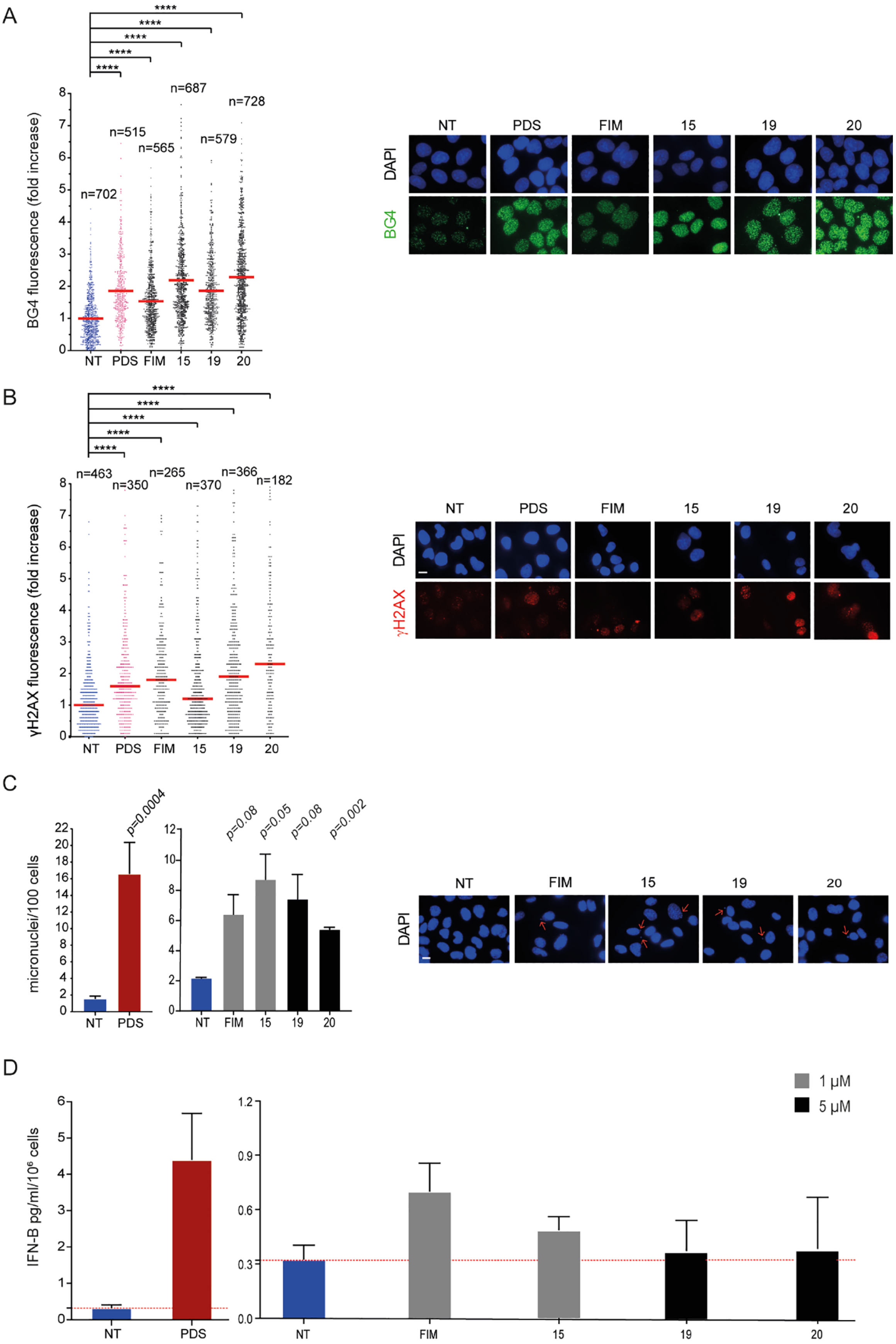


Figure 2. G4 stabilization, DNA damage, and IFN-B stimulation induced by **FIM** derivatives. (A) Quantification of fluorescence signals of BG4 foci in U2OS cells treated for 10 min with **PDS** or **FIM** derivatives at 10 μM concentration. Graphs show the fold increase reported as the mean ± SEM of three biological replicates. The images are representative of IF assays performed at reported concentrations (left). (B) Quantification of

Figure 2. continued

fluorescence signals of γH2AX in U2OS cells treated with **PDS** (10 μM) and **FIM** derivatives at IC_{50} concentrations (4, 2.5, 24, and 14 μM for **FIM**, **15**, **19**, and **20**, respectively). The graph shows the fold increase reported as the median ± SEM of two biological replicates, and the IF representative images are reported (left). (C) Micronuclei quantification by DAPI staining in MNMCA1 cells treated with 1 μM of compounds **FIM** and **15** and 5 μM for the analogues **19** and **20**. **PDS** (10 μM)-treated cells are also shown. Left, the graph shows the mean ± SEM of two biological replicates; right, representative cell images. The scale bar is 10 μm. Above the bar chart, the *p*-value are reported. (D) Quantification of IFN-B stimulated at the reported concentration has been detected after 24 h of treatment followed by 48 h of recovery. **PDS** (10 μM)-treated cells are also shown. The IFN-B protein levels were detected with ELISA assay. The bar chart reports the mean ± SEM of two biological replicates. Significance in all the graphs was calculated by Mann−Whitney test ($^{*}p < 0.05$, $^{**}p > 0.01$, $^{***}p > 0.001$, and $^{****}p < 0.0001$).

strongest G4 binder, followed by **2** and **10**, while **19** and **20** were the worst of the series.

Based on the whole set of biophysical data, the compounds were classified according to their affinity for G4s: strong binders in the case of **1**, **2**, and **10**; moderate binders for **8** and **15**; and modest binders for **19** and **20**.

Therefore, aimed at defining the interplay between G4 affinity, stabilization in cells, cytotoxicity, and immune-stimulation activity of these hydrazone-based compounds, derivatives **1**, **2**, **8**, **15**, **19**, and **20**, having different affinity for G4s, were selected for the biological investigations.

Interestingly, despite the minimal structural differences among the **FG** analogues **1**, **2** and **8**, they showed interesting differences in cytotoxic potency, in-cell G4 stabilization, and IFN-B gene activation. Compound **8** is more cytotoxic than **1** and **2** (Table 5); however, it minimally stabilizes G4 in nuclei (Figure 1A) and it does not trigger IFN-B production (Figure 1D). On the contrary, **1** shows a high G4 stabilization *in vivo* (Figure 1A) and the least cytotoxic potency in murine cells (Table 5), where it triggers a good activation of IFN-B genes (Figure 1D).

Conversely, **FIM** analogues bearing the chemically reactive aldehyde group (**FIM** and **15**) or the hydroxymethyl group (**19** and **20**) exhibit a greater cytotoxic potency than **FG** derivatives (Table 5), with compounds **FIM** and **15** being the most cytotoxic compounds of the series. In addition, they are able to stabilize G4 in cells (Figure 2A) but not able to trigger IFN-B activation. These data clearly show that a high cytotoxic ability interferes with the ability of a G4 ligand to activate the expression of IFN-B genes. Overall, **FG** analogues, characterized by two side chains, exhibit a markedly higher G4 affinity than that of **FIM** analogues (Tables 1 and 3), albeit with a reduction of G4 selectivity (Table 2). As **FIM** analogues overall do not activate IFN-B genes (Figure 2D), whereas the **FG** analogue **1** does (Figure 1D), we speculate that a high ligand affinity for G4 may be required for immune gene activation. **FG** and **FIM** analogues can stabilize G4s at similar levels in nuclear chromatin; however, we do not know whether G4 structures stabilized by each analogue are the same or not. Our data indicate that the specific pattern of stabilized G4s and, likely, the specific time and location may affect the molecular response to G4 ligand activity.

An important observation was that IFN-B activation was independent of the level of induced micronuclei (Figures 1 and 2, panels C and D), suggesting that cytosolic DNA from micronuclei was necessary but not sufficient for immune gene expression.[29,30] Even though the definition of the mechanism likely needs future investigations, however, the activation of other cytoplasmic signaling pathways may affect the recognition of micronuclei and activation of the STING pathway.[5,29,30] Autophagic processes are known to be activated by G4 binders[3,31,32] and can regulate the STING pathway through recycling micronuclei and DNA by forming autophagosomes.[33,34] Interestingly, autophagic gene pathways were not activated at high levels in MCF-7 cells treated with **PDS**, which can activate at very high levels the IFN-B gene and other genes stimulated by IFN-B.[7] Therefore, differences in autophagic pathway activation might explain differences in IFN-B production between analogue **1** and other studied derivatives.

CONCLUSIONS

Comparing very closely related G4 binders has allowed us to demonstrate that a proper balance between G4 affinity/selectivity and cytotoxicity is critical for immune gene activation, in particular a high G4 affinity and a relatively low cytotoxic potency are necessary for a G4 ligand to activate IFN-B genes in cancer cells (Figure 3). Thus, we propose a new rationale, based on low cell-killing potency and high G4 affinity, to discover effective anticancer G4 ligands with immune-stimulation activity.

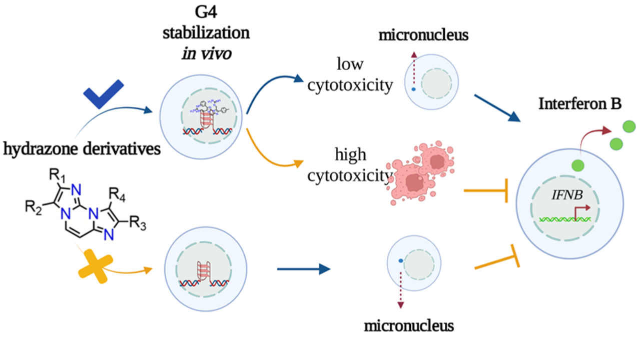


Figure 3. Schematic representation of cellular effects of hydrazone derivatives.

EXPERIMENTAL SECTION

Compound Synthesis and Materials. The synthesis and NMR spectra of **FG** and **FIM** analogues are reported in Supporting Information. All compounds are >95% pure by elemental analysis (see Supporting Information). Controlled pore glass supports, DNA phosphoramidites, all reagents for oligonucleotide synthesis and purification, and all other reagents and solvents were purchased from Merck KGaA (Darmstadt, Germany) and used without further purification. Dual-labeled FAM/TAMRA oligonucleotides and Cy5.5-labeled oligonucleotides were purchased from Biomers (Ulm, Germany).

Oligonucleotide Synthesis and Sample Preparation. The following deoxyribonucleotide sequences were used in this study: d(AGG GAG GGC GCT GGG AGG AGG G) (*c-kit1*), d(CGG GCG GGC GCT AGG GAG GGT) (*c-kit2*), d(TGA GGG TGG GTA GGG TGG GTA A) (*c-myc*), d(TTA GGG TTA GGG TTA GGG TTA GGG TT) (tel_{26}), and d(CGC GAA TTC GCG TTT CGC GAA TTC GCG) (*hairpin*). These oligonucleotides were chemically synthesized on the 1 μmol scale on an ABI 394 DNA/

RNA synthesizer (Applied Biosystems, CA, USA) by using the standard β-cyanoethyl phosphoramidite solid-phase chemistry, as described elsewhere.[35] After synthesis, oligonucleotides were detached from the support and deprotected by treating with an aqueous solution of concentrated ammonia at 55 °C, for 17 h. The filtrates and washings, after being combined and concentrated under reduced pressure, were solubilized in water and purified using a high-performance liquid chromatography system equipped with a Nucleogel SAX column (Macherey-Nagel, 1000-8/46), using a 30 min linear gradient from 100% buffer A to 100% buffer B at a flow rate of 1 mL/min, with buffer A consisting of a 20 mM KH_2PO_4/K_2HPO_4 aqueous solution (pH 7.0) and buffer B consisting of 1.0 M KCl and 20 mM KH_2PO_4/K_2HPO_4 aqueous solution (pH 7.0). Both buffer A and B also contained 20% (v/v) CH_3CN. The purified fractions were then desalted by means of C-18 cartridges (Sep-Pak). The purity of the isolated oligomers was checked by NMR and proved to be higher than 98%. All oligonucleotides were dissolved in a buffer solution consisting of 5 mM KH_2PO_4/K_2HPO_4 (pH 7.0) and 20 mM KCl (or LiCl in the case of *c-myc* because of its high thermal stability). The concentration of each oligonucleotide was verified by measuring the UV absorption at 90 °C, considering the appropriate molar extinction coefficient values ε (λ = 260 nm) calculated using the nearest-neighbor model.[36] Finally, to achieve the correct folding of the DNA sequences, oligonucleotide solutions were annealed by heating at 95 °C for 5 min followed by a slow cooling to room temperature and storage overnight at 4 °C.

CD Experiments. CD experiments were performed on a Jasco J-815 spectropolarimeter equipped with a PTC-423S/15 Peltier temperature controller. All the spectra were recorded at 20 and 100 °C in the wavelength range of 230−320 nm and averaged over three scans. A scan rate of 100 nm/min, with a 0.5 s response time and 1 nm bandwidth, was used. The buffer baseline was subtracted from each spectrum. For the CD experiments, 10 μM G4 and 15 μM duplex DNA in the absence or presence of 2 molar equiv of ligand were used. CD spectra were recorded 10 min after ligand addition. Ligand stock solutions were 10 mM in DMSO. CD melting experiments were carried out in the 20−100 °C temperature range at a 1 °C/min heating rate by following the changes in the CD signal at the wavelengths of the maximum CD intensity (263 nm) for *c-kit1*, *c-kit2*, *c-myc*, and (287 nm) tel_{26} G4s, or minimum CD intensity (252 nm) for the *hairpin*. CD melting experiments were recorded both in the absence and presence of compounds (2 molar equiv) added to the folded nucleic acid structures. The apparent melting temperatures ($T_{1/2}$) were determined from a curve fit using OriginPro 2021 software (OriginLab Corp., MA, USA). $\Delta T_{1/2}$ values were determined as the difference in the $T_{1/2}$ values of the nucleic acid structures in the presence and absence of the compounds. Normalization of melting curves between 0 and 1 was performed to better compare the results. In cases where the melting process was not completed even at 100 °C due to an exceptional ligand-induced G4 thermal stabilization, the relative melting curves were normalized by dividing only by the maximum.

FRET Melting Experiments. Measurements were carried out on a Jasco FP-8300 spectrofluorometer equipped with a Peltier temperature controller system (PCT-818) using a dual-labeled G4-forming sequence FAM-[d(CGG GCG GGC GCT AGG GAG GGT)]-TAMRA (*F-c-kit1-T*). The oligonucleotide was dissolved in water at 1 mM, diluted at 1 μM using 5 mM KH_2PO_4/K_2HPO_4 (pH 7.0) containing 20 mM KCl, and annealed by heating to 90 °C for 5 min, followed by slow cooling to room temperature overnight and storage at 4 °C for 24 h before data acquisition. Experiments were performed in sealed quartz cuvettes with a path length of 1 cm by using 0.2 μM prefolded *F-c-kit1-T* target, in the absence and presence of 2 molar equiv of the ligand and of the *duplex* competitor at 3 and 10 μM final concentrations. In addition, an experiment in the absence of compounds and competitors was also performed. Fluorescence spectra were acquired before and after melting assay (15 and 90 °C, respectively). The dual-labeled oligonucleotide was excited at 492 nm, and emission spectra were recorded between 500 and 650 nm by using a 100 nm/s scan speed. Excitation and emission slit widths were both set to 5 nm. FRET melting experiments were performed by monitoring the emission of FAM at 520 nm (upon excitation at 492 nm), using a heating gradient of 0.2 °C/min over the range 15−90 °C. Emission of FAM was normalized between 0 and 1. Final analysis of the data was carried out using OriginPro 2021 software.

Fluorescent Intercalator Displacement (G4-FID) Assay. A solution containing 0.25 μM G4 DNA (*c-kit1*, *c-kit2*, or *c-myc*) and 0.5 μM TO in 5 mM KH_2PO_4/K_2HPO_4 buffer (pH 7.0) containing 20 mM KCl (or LiCl in the case of *c-myc*) was prepared in a 1 cm-path length cell, and the corresponding fluorescence spectrum was acquired in the absence and presence of increasing concentrations of selected compounds (1 mM stock solution in DMSO). Each ligand addition (from 0.5 to 20 molar equiv) was followed by a 3 min equilibration time before spectrum acquisition. Measurements were run at 20 °C on a Jasco FP-8300 spectrofluorometer equipped with a Peltier cell holder (PCT-818), using an excitation wavelength of 485 nm and recording the emission in the 500−650 nm wavelength range. Both excitation and emission slits were set at 5 nm. Final analysis of the data was carried out using OriginPro 2021 software. The percentage of TO displacement was calculated as follows: TO displacement (%) = 100 − [$(F/F_0) \times 100$], where F_0 is the fluorescence in the absence of a ligand and F is the fluorescence after each ligand addition. The percentage of displacement was then plotted as a function of the ligand concentration, and DC_{50} was calculated as the required concentration to displace 50% TO. Each titration was performed in duplicate.

Microscale Thermophoresis. MST measurements were performed using a Monolith NT.115 instrument (NanoTemper Technologies). The Cy5.5 fluorescently labeled oligonucleotides (*c-kit1*, *c-kit2*, and *c-myc*) were prepared at 1 μM in 5 mM KH_2PO_4/K_2HPO_4 buffer (pH 7.0) containing 20 mM KCl and annealed as described above. DNA samples were then diluted using the same phosphate buffer supplemented with 0.1% Tween. For the MST experiments, the concentration of the labeled oligonucleotides was kept constant at 20 nM, while a serial dilution of the ligand (1:2 from 5.0, 40, 160, or 400 μM ligand stock solution) in the same buffer used for DNAs was prepared and mixed with the oligonucleotide solution with a volume ratio of 1:1. All the samples, containing 20% DMSO as the final concentration, were loaded into standard capillaries (NanoTemper Technologies). Measurements were performed and analyzed as previously reported.[37]

Cell Lines and Treatments. Human osteosarcoma U2OS and murine fibrosarcoma MNMCA1 cell lines were grown in monolayer cultures in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) (Gibco) and 1% L-glutamine (Gibco). Human fibroblast lung MRC5 cells were grown in a monolayer culture in DMEM, supplemented with Ham's F-10 nutrient mix (1:1), 10% FBS, and Pen/Strep 100 μg/mL. All cell lines were grown in a humidified incubator at 37 °C and 5% of CO_2. Cell line identity was routinely checked by genotyping (BMR Genomics). Compounds were dissolved in dimethyl sulfoxide (Sigma-Aldrich #472301) at 10 mM concentration, stored in aliquots at −20 °C, and diluted to final concentrations immediately prior to use.

MTT Cell Proliferation Assay. U2OS, MNMCA1, and MRC5 cells (3×10^4) were seeded in 24 wells. 24 h after seeding, cells were treated with increasing concentrations of compounds for 24 h. Then, compounds were removed, and the cells were grown in complete drug-free medium for 48 h. Then, thiazolyl blue tetrazolium bromide (MTT) (Merck #2128) solution (0.45 μg/mL) was added to each well and incubated for 1 h at 37 °C. After incubation, the medium was removed and 300 μL of dimethyl sulfoxide was added and incubated for 1 h at room temperature. Then, 100 μL of the solution was put in 96 wells, and absorbance at 540 nm was measured using a multiplate reader. The linear regression parameters were determined to calculate the IC_{50} (GraphPad Prism 4.0, Graph Pad Software Inc.).

IF Microscopy. U2OS cells (3.5×10^5) were seeded in 35 mm dish on coverslips. The BG4 fluorescence signal was determined after 10 min of treatment at the reported concentrations. The BG4 antibody was purified as described.[13] Briefly, BG4 was isolated from *Escherichia coli* extracts by using silica-based resin (Thermo #89964) precharged with Co^{2+} ions and eluted with 250 μM imidazole/PBS pH 8.0. The eluted antibody was concentrated in Pierce 30k MWCO tubes (Pierce #88529), and imidazole was finally removed by buffer exchange with intracellular cell salt buffer in Pierce 30k MWCO tubes. For BG4 staining, cells were pre-fixed with cell culture medium and fix solution (1:1) and then incubated with the fix solution composed of methanol and acetic acid (3:1) for 10 min at RT. The cells were permeabilized with 0.1% of Triton X-100 in PBS and blocked in 2% non-fat milk for 1 h at RT under gentle shaking. Next, cells were stained with 0.5 μg of BG4 for 2 h at room temperature. Next, cells were incubated with the anti-FLAG antibody (dilution 1/800) (Cell Signaling Technology #2368) for 1 h and then stained with the Alexa Fluor 488 anti-rabbit IgG (Life technologies #A11008). For S139-phosphorylated histone H2AX, γH2AX cells were treated with compounds at the reported concentrations for 24 h. Then, cells were fixed with 4% formaldehyde for 10 min, permeabilized with 0.5% Triton X-100 in PBS for 15 , and then incubated with 8% BSA in PBS for 30 min at RT. Next, cells were stained with anti-γH2AX antibodies (#05-636, Millipore) diluted to 1:500 and next incubated for 1 h with Alexa Fluor 594 anti-mouse IgG (#A11032, Life Technologies). For DNA staining, cells were incubated with 2 μg/μL DAPI for 20 min. The cover glasses were mounted with Mowiol 488. The slides were visualized at room temperature by using a fluorescence microscope (Eclipse TE 2000-S, Nikon) equipped with an AxioCam MRm (Zeiss) digital camera. The fluorescence signal was quantified by using ImageJ software and reported as a fold increase of the non-treated sample. Graphs were prepared with GraphPad Prism 8.

IFN-B ELISA Assay. MNMCA1 cells (8×10^5) were seeded in a 10 mm dish. IFN-B protein levels were measured in cell medium supernatants. Culture medium of untreated and treated MNMCA1 cells was collected after 24 h of treatment followed by 48 h of drug recovery. Supernatants were added with protease inhibitors (1 mg/mL pepstatin, leupeptin, and aprotinin, 2 mM DTT, and 0.5 mM PMSF) and then concentrated around 25-fold by using a Pierce Protein Concentrator PES, 3k MWCO, 5−20 mL (#88525, Thermo Fisher). IFN-B protein levels were quantified with a human IFN-B Quantikine ELISA kit (MIFNB0, R&D Systems) following manufacturer's instructions. IFN-B levels were normalized over the cell number.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jmedchem.2c00772.

Molecular formula strings (CSV)

Synthesis of compounds 1-20, NMR spectrometry characteristics, UHPLC system and conditions, and ^{1}H and ^{13}C spectra of compounds 1-20 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Rita Morigi − *Department of Pharmacy and Biotechnology, Alma Mater Studiorum—University of Bologna, 40126 Bologna, Italy;* orcid.org/0000-0003-4739-8733; Phone: +39 051 2099725; Email: rita.morigi@unibo.it

Jussara Amato − *Department of Pharmacy, University of Naples Federico II, 80131 Naples, Italy;* orcid.org/0000-0001-6096-3544; Phone: +39 081 678630; Email: jussara.amato@unina.it

Giovanni Capranico − *Department of Pharmacy and Biotechnology, Alma Mater Studiorum—University of Bologna, 40126 Bologna, Italy;* orcid.org/0000-0002-8708-6454; Phone: +39 051 2091209; Email: giovanni.capranico@unibo.it

Authors

Simona Marzano − *Department of Pharmacy, University of Naples Federico II, 80131 Naples, Italy*

Giulia Miglietta − *Department of Pharmacy and Biotechnology, Alma Mater Studiorum—University of Bologna, 40126 Bologna, Italy*

Jessica Marinello − *Department of Pharmacy and Biotechnology, Alma Mater Studiorum—University of Bologna, 40126 Bologna, Italy*

Andrea Arleo − *Department of Pharmacy and Biotechnology, Alma Mater Studiorum—University of Bologna, 40126 Bologna, Italy;* orcid.org/0000-0002-3262-7268

Monica Procacci − *Department of Pharmacy and Biotechnology, Alma Mater Studiorum—University of Bologna, 40126 Bologna, Italy*

Alessandra Locatelli − *Department of Pharmacy and Biotechnology, Alma Mater Studiorum—University of Bologna, 40126 Bologna, Italy*

Alberto Leoni − *Department of Pharmacy and Biotechnology, Alma Mater Studiorum—University of Bologna, 40126 Bologna, Italy*

Bruno Pagano − *Department of Pharmacy, University of Naples Federico II, 80131 Naples, Italy;* orcid.org/0000-0002-7716-9010

Antonio Randazzo − *Department of Pharmacy, University of Naples Federico II, 80131 Naples, Italy;* orcid.org/0000-0002-9192-7586

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jmedchem.2c00772

Author Contributions

S.M., G.M., and R.M. contributed equally. R.M., A. Locatelli, and A. Leoni synthesized derivatives. S.M., B.P., and J.A. performed oligonucleotide synthesis and biophysical experiments. G.M., A.A., J.M., and M.P. performed the biological experiments. A.R., J.A., and G.C. designed, planned, and supervised experimental work and interpreted the results with all authors. The manuscript was written by G.C., J.A., G.M., and R.M. and with contributions from other authors. All authors approved the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank all laboratory members for helpful discussion. The research leading to these results had received funding from AIRC (Associazione Italiana per la Ricerca sul Cancro), IG 2019—ID 23032 project—PI G.C., IG 2020—ID 24590 project—PI B.P and IG 2021—ID 26313 project-PI A.R. G.M. is a recipient of a FIRC-AIRC postdoc fellowship from Italy [23953].

■ ABBREVIATIONS

BSA, bovine serum albumin; CBR, condensed benzene ring; CD, circular dichroism; DAPI, 4′,6-diamidino-2-phenylindole; DMEM, Dulbecco's modified Eagle medium; DMF, dimethylformamide; DSB, double-strand break; FAM, fluorescein; FBS, fetal bovine serum; FID, fluorescence intercalator displacement; FRET, Förster resonance energy transfer; G4, G-quadruplex; γH2AX, S139-phosphorylated histone H2AX; IF, immunofluorescence; IFN-B, interferon β; MST, microscale thermophoresis; NOE, nuclear overhauser effect; PBS, phosphate-buffered saline; **PDS**, pyridostatin; PIPES, 1,4-piperazinediethanesulfonic acid; TAMRA, carboxytetramethylrhodamine; TO, thiazole orange

REFERENCES

(1) Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-Quadruplexes in Gene Promoters: A Novel Anticancer Strategy? *Nat. Rev. Drug Discovery* **2011**, *10*, 261−275.

(2) Neidle, S. Quadruplex Nucleic Acids as Novel Therapeutic Targets. *J. Med. Chem.* **2016**, *59*, 5987−6011.

(3) Carvalho, J.; Mergny, J.-L.; Salgado, G. F.; Queiroz, J. A.; Cruz, C. G-Quadruplex, Friend or Foe: The Role of the G-Quartet in Anticancer Strategies. *Trends Mol. Med.* **2020**, *26*, 848−861.

(4) Miglietta, G.; Russo, M.; Capranico, G. G-Quadruplex−R-Loop Interactions and the Mechanism of Anticancer G-Quadruplex Binders. *Nucleic Acids Res.* **2020**, *48*, 11942−11957.

(5) Miglietta, G.; Marinello, J.; Russo, M.; Capranico, G. Ligands Stimulating Antitumour Immunity as the next G-Quadruplex Challenge. *Mol. Cancer* **2022**, in press.

(6) Hilton, J.; Gelmon, K.; Bedard, P. L.; Tu, D.; Xu, H.; Tinker, A. V.; Goodwin, R.; Laurie, S. A.; Jonker, D.; Hansen, A. R.; Veitch, Z. W.; Renouf, D. J.; Hagerman, L.; Lui, H.; Chen, B.; Kellar, D.; Li, I.; Lee, S.-E.; Kono, T.; Cheng, B. Y. C.; Yap, D.; Lai, D.; Beatty, S.; Soong, J.; Pritchard, K. I.; Soria-Bretones, I.; Chen, E.; Feilotter, H.; Rushton, M.; Seymour, L.; Aparicio, S.; Cescon, D. W. Results of the Phase I CCTG IND.231 Trial of CX-5461 in Patients with Advanced Solid Tumors Enriched for DNA-Repair Deficiencies. *Nat. Commun.* **2022**, *13*, 3607.

(7) Miglietta, G.; Russo, M.; Duardo, R. C.; Capranico, G. G-Quadruplex Binders as Cytostatic Modulators of Innate Immune Genes in Cancer Cells. *Nucleic Acids Res.* **2021**, *49*, 6673−6686.

(8) Woo, S.-R.; Corrales, L.; Gajewski, T. F. Innate Immune Recognition of Cancer. *Annu. Rev. Immunol.* **2015**, *33*, 445.

(9) Poh, A. STINGing Antitumor Immunity into Action. *Cancer Discovery* **2018**, *8*, 259−260.

(10) Hartmann, G. Nucleic Acid Immunity. *Adv. Immunol.* **2017**, *133*, 121−169.

(11) van den Boorn, J. G.; Hartmann, G. Turning Tumors into Vaccines: Co-Opting the Innate Immune System. *Immunity* **2013**, *39*, 27−37.

(12) Fu, J.; Kanne, D. B.; Leong, M.; Glickman, L. H.; McWhirter, S. M.; Lemmens, E.; Mechette, K.; Leong, J. J.; Lauer, P.; Liu, W.; Sivick, K. E.; Zeng, Q.; Soares, K. C.; Zheng, L.; Portnoy, D. A.; Woodward, J. J.; Pardoll, D. M.; Dubensky, T. W.; Kim, Y. STING Agonist Formulated Cancer Vaccines Can Cure Established Tumors Resistant to PD-1 Blockade. *Sci. Transl. Med.* **2015**, *7*, 283ra52.

(13) De Magis, A.; Manzo, S. G.; Russo, M.; Marinello, J.; Morigi, R.; Sordet, O.; Capranico, G. DNA Damage and Genome Instability by G-Quadruplex Ligands Are Mediated by R Loops in Human Cancer Cells. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 816−825.

(14) Amato, J.; Miglietta, G.; Morigi, R.; Iaccarino, N.; Locatelli, A.; Leoni, A.; Novellino, E.; Pagano, B.; Capranico, G.; Randazzo, A. Monohydrazone Based G-Quadruplex Selective Ligands Induce DNA Damage and Genome Instability in Human Cancer Cells. *J. Med. Chem.* **2020**, *63*, 3090−3103.

(15) Harding, S. M.; Benci, J. L.; Irianto, J.; Discher, D. E.; Minn, A. J.; Greenberg, R. A. Mitotic Progression Following DNA Damage Enables Pattern Recognition within Micronuclei. *Nature* **2017**, *548*, 466−470.

(16) Mackenzie, K. J.; Carroll, P.; Martin, C.-A.; Murina, O.; Fluteau, A.; Simpson, D. J.; Olova, N.; Sutcliffe, H.; Rainger, J. K.; Leitch, A.; Osborn, R. T.; Wheeler, A. P.; Nowotny, M.; Gilbert, N.; Chandra, T.; Reijns, M. A. M.; Jackson, A. P. CGAS Surveillance of Micronuclei Links Genome Instability to Innate Immunity. *Nature* **2017**, *548*, 461−465.

(17) Sparapani, S.; Bellini, S.; Gunaratnam, M.; Haider, S. M.; Andreani, A.; Rambaldi, M.; Locatelli, A.; Morigi, R.; Granaiola, M.; Varoli, L.; Burnelli, S.; Leoni, A.; Neidle, S. Bis-Guanylhydrazone Diimidazo[1,2-a:1,2-c]Pyrimidine as a Novel and Specific G-Quadruplex Binding Motif. *Chem. Commun.* **2010**, *46*, 5680.

(18) Amato, J.; Morigi, R.; Pagano, B.; Pagano, A.; Ohnmacht, S.; De Magis, A.; Tiang, Y.-P. Y. P.; Capranico, G.; Locatelli, A.; Graziadio, A.; Leoni, A.; Rambaldi, M.; Novellino, E.; Neidle, S.; Randazzo, A. Toward the Development of Specific G-Quadruplex Binders: Synthesis, Biophysical, and Biological Studies of New Hydrazone Derivatives. *J. Med. Chem.* **2016**, *59*, 5706−5720.

(19) Hsu, S. T. D.; Varnai, P.; Bugaut, A.; Reszka, A. P.; Neidle, S.; Balasubramanian, S. A G-Rich Sequence within the c-Kit Oncogene Promoter Forms a Parallel G-Quadruplex Having Asymmetric G-Tetrad Dynamics. *J. Am. Chem. Soc.* **2009**, *131*, 13399−13409.

(20) 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*, 2048−2058.

(21) Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. Structure of the Hybrid-2 Type Intramolecular Human Telomeric G-Quadruplex in K+ Solution: Insights into Structure Polymorphism of the Human Telomeric Sequence. *Nucleic Acids Res.* **2007**, *35*, 4927−4940.

(22) Randazzo, A.; Spada, G. P.; da Silva, M. W. Circular Dichroism of Quadruplex Structures. *Top. Curr. Chem.* **2013**, *330*, 67−86.

(23) Decian, A.; Guittat, L.; Kaiser, M.; Sacca, B.; Amrane, S.; Bourdoncle, A.; Alberti, P.; Teuladefichou, M. P.; Lacroix, L.; Mergny, J. L. Fluorescence-Based Melting Assays for Studying Quadruplex Ligands. *Methods* **2007**, *42*, 183−195.

(24) Monchaud, D.; Allain, C.; Teulade-Fichou, M. P. Development of a Fluorescent Intercalator Displacement Assay (G4-FID) for Establishing Quadruplex-DNA Affinity and Selectivity of Putative Ligands. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4842−4845.

(25) Zuffo, M.; Guédin, A.; Leriche, E.-D.; Doria, F.; Pirota, V.; Gabelica, V.; Mergny, J.-L.; Freccero, M. More Is Not Always Better: Finding the Right Trade-off between Affinity and Selectivity of a G-Quadruplex Ligand. *Nucleic Acids Res.* **2018**, *46*, No. e115.

(26) Jerabek-Willemsen, M.; Wienken, C. J.; Braun, D.; Baaske, P.; Duhr, S. Molecular Interaction Studies Using Microscale Thermophoresis. *Assay Drug Dev. Technol.* **2011**, *9*, 342−353.

(27) Nanni, P.; Landuzzi, L.; Nicoletti, G.; De Giovanni, C.; Giovarelli, M.; Lalli, E.; Facchini, A.; Lollini, P. L. Control of H-2 Expression in Transformed Nonhaemopoietic Cells by Autocrine Interferon. *Br. J. Cancer* **1992**, *66*, 479−482.

(28) Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. Visualization and Selective Chemical Targeting of RNA G-Quadruplex Structures in the Cytoplasm of Human Cells. *Nat. Chem.* **2014**, *6*, 75−80.

(29) Crowl, J. T.; Gray, E. E.; Pestal, K.; Volkman, H. E.; Stetson, D. B. Intracellular Nucleic Acid Detection in Autoimmunity. *Annu. Rev. Immunol.* **2017**, *35*, 313−336.

(30) Pilger, D.; Seymour, L. W.; Jackson, S. P. Interfaces between Cellular Responses to DNA Damage and Cancer Immunotherapy. *Genes Dev.* **2021**, *35*, 602−618.

(31) Beauvarlet, J.; Bensadoun, P.; Darbo, E.; Labrunie, G.; Rousseau, B.; Richard, E.; Draskovic, I.; Londono-Vallejo, A.; Dupuy, J.-W.; Nath Das, R.; Guédin, A.; Robert, G.; Orange, F.; Croce, S.; Valesco, V.; Soubeyran, P.; Ryan, K. M.; Mergny, J.-L.; Djavaheri-Mergny, M. Modulation of the ATM/Autophagy Pathway by a G-Quadruplex Ligand Tips the Balance between Senescence and Apoptosis in Cancer Cells. *Nucleic Acids Res.* **2019**, *47*, 2739−2756.

(32) Zhou, W. J.; Deng, R.; Zhang, X. Y.; Feng, G. K.; Gu, L. Q.; Zhu, X. F. G-Quadruplex Ligand SYUIQ-5 Induces Autophagy by Telomere Damage and TRF2 Delocalization in Cancer Cells. *Mol. Cancer Ther.* **2009**, *8*, 3203−3213.

(33) Gui, X.; Yang, H.; Li, T.; Tan, X.; Shi, P.; Li, M.; Du, F.; Chen, Z. J. Autophagy Induction via STING Trafficking Is a Primordial Function of the CGAS Pathway. *Nature* **2019**, *567*, 262−266.

(34) Hopfner, K.-P.; Hornung, V. Molecular Mechanisms and Cellular Functions of CGAS−STING Signalling. *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 501−521.

(35) Amato, J.; Mashima, T.; Kamatari, Y. O.; Kuwata, K.; Novellino, E.; Randazzo, A.; Giancola, C.; Katahira, M.; Pagano, B. Improved Anti-Prion Nucleic Acid Aptamers by Incorporation of Chemical Modifications. *Nucleic Acid Ther.* **2020**, *30*, 414−421.

(36) Cantor, C. R.; Warshaw, M. M.; Shapiro, H. Oligonucleotide Interactions. III. Circular Dichroism Studies of the Conformation of Deoxyoligonucleolides. *Biopolymers* **1970**, *9*, 1059−1077.

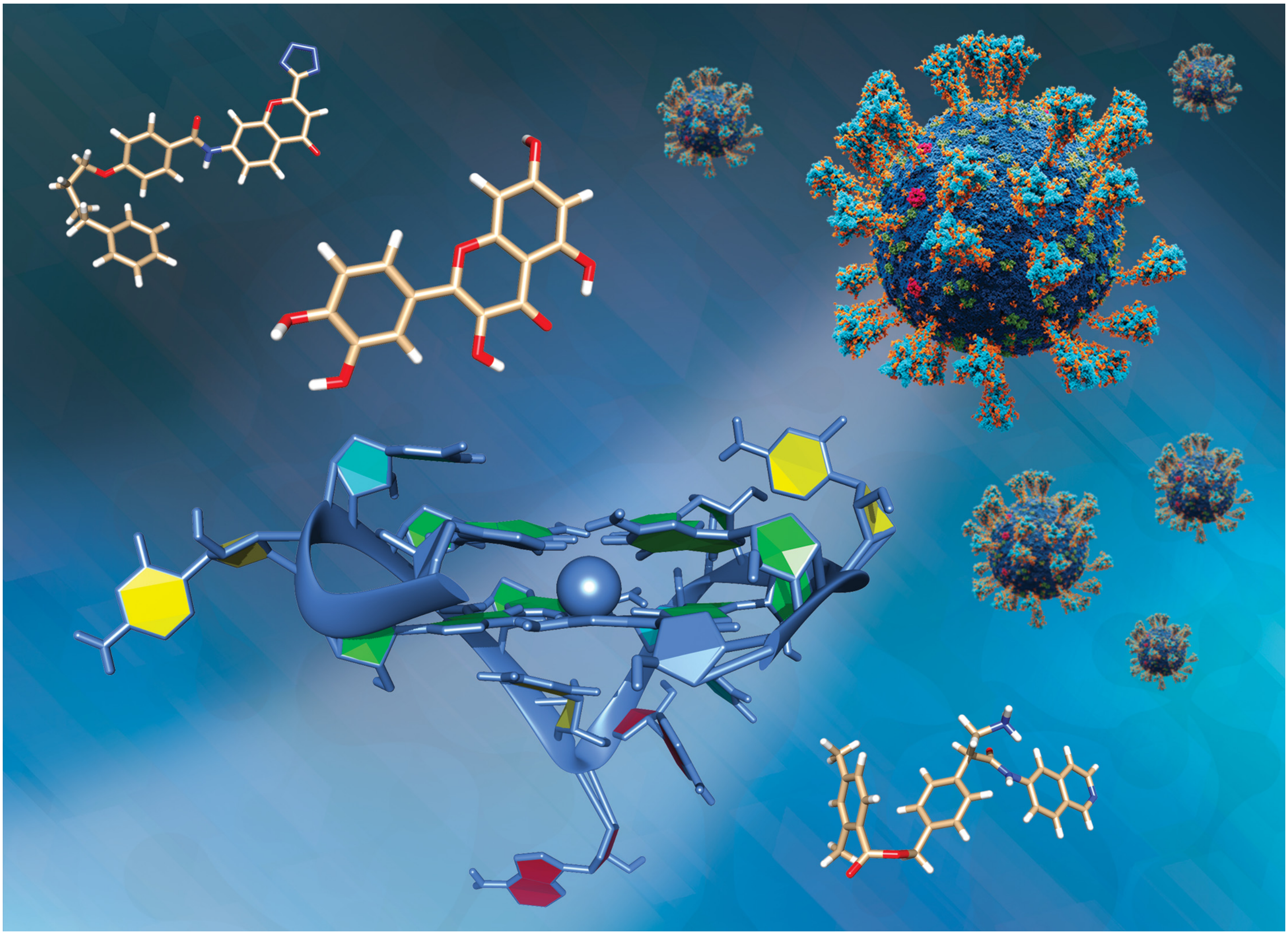
(37) Amato, J.; Madanayake, T. W.; Iaccarino, N.; Novellino, E.; Randazzo, A.; Hurley, L. H.; Pagano, B. HMGB1 Binds to the KRAS Promoter G-Quadruplex: A New Player in Oncogene Transcriptional Regulation? *Chem. Commun.* **2018**, *54*, 9442−9445.

PAPER IV

Ligand-based drug repurposing strategy identified SARS-CoV-2 RNA G-quadruplex binders

Moraca F., **Marzano S.**, D'Amico F., Lupia A., Di Fonzo S., Vertecchi E., Salvati E., Di Porzio A., Catalanotti B., Randazzo A., Pagano B., Amato J.

Chemical Communications, 2022,Vol. 58(85)

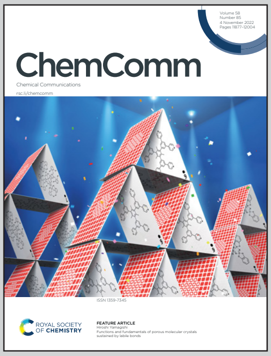


Showcasing research from the group of Prof. Jussara Amato at Department of Pharmacy, University of Naples Federico II, Italy

Ligand-based drug repurposing strategy identified SARS-CoV-2 RNA G-quadruplex binders

A ligand-based pharmacophore virtual screening strategy and multiple biophysical techniques were employed to identify some FDA-approved drugs that bind to SARS-CoV-2 G-quadruplex RNA. This work demonstrates that the methodological approach employed is effective in finding existing therapeutic molecules capable of targeting viral RNA secondary structures.

As featured in:



See Bruno Pagano,
Jussara Amato *et al.*,
Chem. Commun., 2022, **58**, 11913.

ROYAL SOCIETY
OF **CHEMISTRY**

rsc.li/chemcomm

ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2022, **58**, 11913

Received 2nd June 2022,
Accepted 27th September 2022

DOI: 10.1039/d2cc03135c

rsc.li/chemcomm

Ligand-based drug repurposing strategy identified SARS-CoV-2 RNA G-quadruplex binders†

Federica Moraca,[ab] Simona Marzano,[a] Francesco D'Amico,[c] Antonio Lupia,[ab] Silvia Di Fonzo,[c] Eleonora Vertecchi,[d] Erica Salvati,[d] Anna Di Porzio,[a] Bruno Catalanotti,[a] Antonio Randazzo,[a] Bruno Pagano*[a] and Jussara Amato*[a]

The single-stranded RNA genome of SARS-CoV-2 contains some G-quadruplex-forming G-rich elements which are putative drug targets. Here, we performed a ligand-based pharmacophore virtual screening of FDA approved drugs to find candidates targeting such RNA structures. Further *in silico* and *in vitro* assays identified three drugs as emerging SARS-CoV-2 RNA G-quadruplex binders.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the highly contagious virus responsible for the ongoing COVID-19 pandemic.[1] Although the SARS-CoV-2 vaccination campaign is showing its positive effect,[2,3] this pandemic remains a global emergency and the search for drugs to treat the viral infection is still an urgent task. However, traditional drug discovery is a long-cycle process with low success rates.[4]

In this context, repurposing of already approved drugs is currently one of the most attractive propositions because it involves the use of de-risked compounds, with potentially lower development costs and shorter development timelines.[4,5] A widely used approach to drug repurposing starts with virtual screening (VS) of existing drugs employing computational methods which are fast, low-cost screening processes. Many drugs are already being successfully repurposed to treat various diseases, including viral infections.[5] Thus, with the aim of boosting the arsenal against COVID-19, scientists also began to explore repurposed therapeutic molecules.[6–8]

So far, almost all new antiviral therapeutic strategies focus on targeting proteins.[8,9] However, the threat posed by SARS-CoV-2 infection requires exploring also plausible alternative approaches, such as targeting viral RNA and, in particular, its secondary structures.[10,11] Indeed, the folding of specific regions of the viral genomic RNA into certain secondary structures may hinder the viral genome expression and replication by acting as roadblocks for viral RNA transcription and/or as hallmarks for the attachment of RNA processing machinery.

Among these structures are the G-quadruplexes (G4s), four-stranded structures that can be formed by the folding on itself of single-stranded guanine-rich DNA or RNA sequences.[12a,b] A G4 structure is characterized by the stacking of two or more planar arrangements of four guanines (G-tetrads) stabilized by Hoogsteen hydrogen bonds and cation coordination. These structures may occur in sequences with at least four contiguous tracts of two or more guanines interspersed with sequences forming the so-called loops.

Critical roles for G4s have been described in several viruses,[13,14] including single-stranded RNA viruses, and some G4-targeting compounds have shown antiviral activity,[15] thus suggesting G4 specific compounds as potential antiviral agents. Recent reports have identified a number of putative G4-forming sequences in the genome of SARS-CoV-2, and some of them were demonstrated to form G4s *in vitro*.[11,16–18]

Here, by using ligand-based VS (LBVS) of FDA approved drugs and multiple biophysical techniques, we identified therapeutic molecules able to bind and stabilize SARS-CoV-2 G4-forming RNA and also provided a plausible mechanism of action of such molecules at the molecular level.

Molecules with similar structures tend to have similar properties and functions. Therefore, in LBVS, a pharmacophore model could be derived to define the structural features required to bind a target and exert biological activity.[19] Following this approach, drugs matching the structural and geometrical features of typical G4 binders should have analogous G4-binding properties. Thus, starting from known active RNA G4 ligands from the literature (*training set*, Table S1, ESI†), 3D pharmacophore models were generated and validated before performing the drug repurposing LBVS (for details on *training set* generation, see the Experimental section, ESI†). 3D ligand-based pharmacophores were built on the 2D Fingerprint clustered *training set*,

---

[a] *Department of Pharmacy, University of Naples Federico II, Via D. Montesano 49, 80131 Naples, Italy. E-mail: bruno.pagano@unina.it, jussara.amato@unina.it*
[b] *Net4Science srl, University "Magna Græcia" of Catanzaro, 88100 Catanzaro, Italy*
[c] *Elettra-Sincrotrone Trieste S. C. p. A., Science Park, 34149 Trieste, Italy*
[d] *Institute of Molecular Biology and Pathology, National Research Council, Via degli Apuli 4, 00185 Rome, Italy*

† Electronic supplementary information (ESI) available: Detailed Experimental section, Tables S1–S9, Fig. S1–S24 and Movie S1. See DOI: https://doi.org/10.1039/d2cc03135c

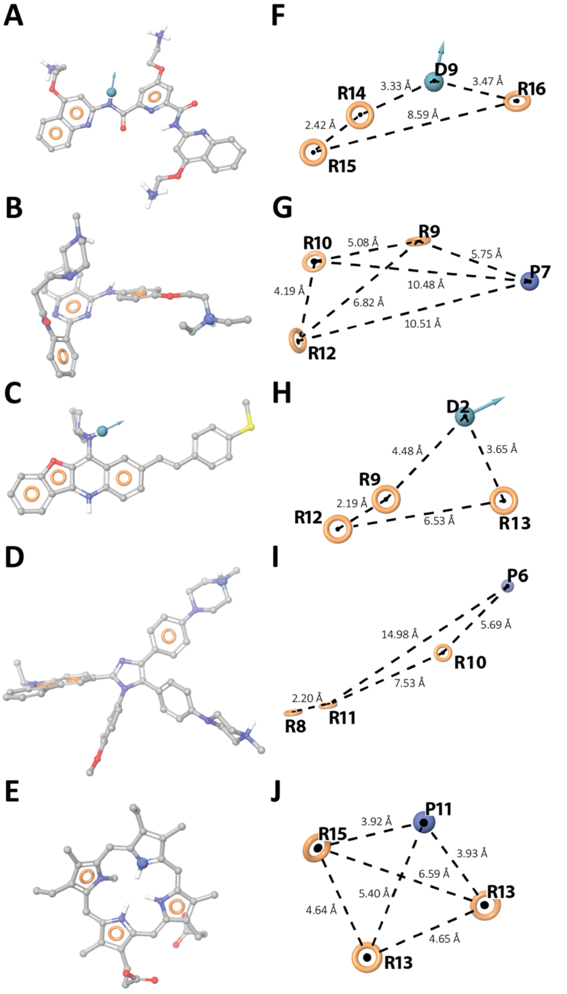


Fig. 1 Ligand-based 3D pharmacophore models. (A–E) *Hy1, Hy3, Hy5, Hy8, and Hy9* overlapping the reference ligands PDP, 19, 4a-10, IZCZ-3, and NMM-IX, respectively (Table S1, ESI†). (F–J) Pharmacophore hypotheses and distances between the pharmacophoric sites (dashed lines). Aromatic rings, H-bond donor, and positive ionizable groups are labelled as **R**, **D**, and **P**, respectively. The tolerance radius of each feature was set to 2 Å.

generating further 50 conformers to increase the chances that ligands matched the pharmacophores of the known G4 binders. A library of decoys was used as a benchmark *test set* to validate the predictive power of each hypothesis, leading to five best performing pharmacophore models (Fig. 1). All five models showed some typical features of G4 binders, namely three aromatic rings and a hydrogen bond donor or a positive ionisable group, which however are placed in different 3D spatial arrangements, as highlighted by their distances and by the different directionality of the vectors D (Fig. 1F–J), allowing us to retrieve different RNA G4 binders from the *training set* (Table S2, ESI†). Thus, the five pharmacophore models were employed to screen a library of ∼3000 FDA-approved drugs, considering the fitness score of the *training set* as a cut-off value to choose the best-matching drugs. The best hits, along with their original use, are listed in Table S3 (ESI†). Of note, 5 of the 15 putative G4 ligands (Table S4, ESI†) have already been evaluated as possible anti COVID-19 agents.

To investigate their G4-binding properties, we employed three biologically relevant G-rich sequences of SARS-CoV-2 (G4-1, G4-2, and G4-3, ESI†) found in the coding sequence regions of nucleocapsid protein, non-structural protein 10, and non-structural protein 3, as potential targets. Such sequences have already been shown to form stable G4s.[11,18] First, the ability of the drugs to bind those G4s was assessed by fluorescent intercalator displacement (G4-FID) assay, which relies on a light-up fluorescent probe (thiazole orange, TO) that binds to the G4 structures and can be competitively displaced by candidate ligands, thus enabling the determination of their relative affinity.[20] Pyridostatin (PDS), a well-known G4 binder, was used as a positive control. The results of G4-FID assay (Fig. S1 and Table S5, ESI†) showed that only Netarsudil and Quercetin (along with PDS) were able to effectively displace TO from all three G4 structures. Pranlukast turned out to be a very good TO competitor in the case of G4-1, while Ledipasvir and Osimertinib exhibited a certain degree of TO displacement only from G4-1 and G4-3, respectively. Based on DC_{50} values, the drugs were ranked as follows: (i) drugs with low to null affinity for G4s ($DC_{50} > 10.0$ μM), (ii) drugs with moderate affinity ($5.0 < DC_{50} < 10.0$ μM) (Ledipasvir for G4-1, Osimertinib and Quercetin for G4-3), and (iii) good G4 binders ($DC_{50} < 5.0$ μM) (Netarsudil, Pranlukast for G4-1, Quercetin for G4-1 and G4-2).

To study the G4-stabilizing effect of ligands showing at least moderate affinity for those targets, circular dichroism (CD) melting experiments were performed. First, the structures adopted by G4-1, G4-2 and G4-3 were verified by recording the relative CD spectra, which showed a maximum around 266 nm and a minimum at 240 nm (Fig. S2, ESI†) indicating the presence of parallel-stranded G4s. Melting and annealing profiles of G4s (recorded at 1.0 and 0.5 °C min^{-1}) were then collected by following the changes in CD signal at 266 nm (Fig. S3, ESI†). As for G4-1, melting and annealing curves were superimposable, showing that the unfolding and folding processes are at thermodynamic equilibrium and the T_m can be accurately determined. Conversely, hysteresis was observed for G4-2 and G4-3, indicating that they are affected by the kinetics of the process (which prevents reproducible measurements) and suggesting the presence of intermolecular G4 species or higher order structures through self-association between G4 units.[18] Therefore, in subsequent studies, we decided to focus only on G4-1. The selected drugs significantly increased the thermal stability of G4-1 (Fig. S4, ESI†). Interestingly, the best effects were found for drugs that showed higher affinity for G4-1 in the G4-FID assay (Netarsudil, Pranlukast, and Quercetin). Upon interacting with Ledipasvir, Osimertinib, and Quercetin, no significant variations in the CD spectrum of G4-1 were detected (Fig. S4, ESI†), implying that it kept its parallel G4 structure. Conversely, Netarsudil and Pranlukast seem to alter the native conformation of G4-1. Actually, these two drugs exhibit CD signal in the wavelength region of G4 (Fig. S5, ESI†). However, at the wavelength at which the melting experiments were recorded, the signal of both is close to zero, thus not significantly affecting the results. On the other hand, the experimental spectra of the two G4/drug mixtures differ sizeably from the corresponding spectra resulting from the arithmetic sum of the single spectra (Fig. S6, ESI†), highlighting once again their interaction.

To evaluate the affinity of Netarsudil, Pranlukast, and Quercetin for G4-1, fluorescence titration experiments were performed. Fluorescence emission spectra of drugs in the absence and presence of increasing amounts of G4 were recorded (Fig. S7, ESI†). On addition of RNA, fluorescence quenching and enhancement were observed for Netarsudil and Quercetin, respectively. Conversely, no relevant change in fluorescence intensity was observed for the intrinsically weakly fluorescent Pranlukast, not allowing us to quantify its affinity. Thus, binding isotherms were obtained by plotting fluorescence changes with G4 concentration (Fig. S7, ESI†), and the curves were fitted giving dissociation constants (K_d) of 0.8 (±0.2) and 13 (±5) μM for Netarsudil and Quercetin, respectively.

To get some information on the selectivity of the drugs, we evaluated their ability to stabilize the G4s derived from the 5′ untranslated region of *BCL-2* (BCL2-G4) and the long non-coding RNA *GSEC* (GSEC-G4) (Fig. S8 and S9, ESI†). CD melting data show that they do not significantly increase the thermal stability of GSEC-G4 ($\Delta T_m \leq 2$ °C). Conversely, while Netarsudil stabilizes BCL2-G4 marginally ($\Delta T_m = 2.6$ °C), Pranlukast and Quercetin show a moderate stabilizing capacity (ΔT_m of 5.3 and 5.4 °C, respectively), even if lower than that found for G4-1.

UV resonance Raman (UVRR) spectroscopy was then employed to get insights into the binding mode of drugs to G4-1.[21] UVRR spectra of G4-1, ligands, and corresponding complexes, all recorded at 266 nm and processed following standard procedures,[21] are shown in Fig. 2. In each panel of Fig. 2, the difference between the spectrum of the complex and that corresponding to the arithmetic sum of constituents (see ESI†) is shown to emphasize the spectral perturbations induced by the interaction. A change in the intensity and/or position of the bands (Fig. S10–S13, ESI†) indicates that an interaction is occurring and suggests the structural moieties involved. This analysis is not straightforward for Netarsudil, as most of its bands overlap those of RNA, precluding the possibility of obtaining clear information on the drug-binding regions of RNA. The only exceptions are the spectral variations at 1334, 1480, 1510, and 1579 cm^{-1}, which are associated to the adenine ring vibrations. Compared to the sum, the experimental spectrum of the RNA/Netarsudil complex shows an intensity increase of these bands, indicating that adenine residues of loop are involved in drug binding. Noteworthy, Netarsudil's strong spectral contribution allowed us to evaluate the parts of the drug involved in the interaction. The positive 1646 and negative 1657 cm^{-1} peaks clearly indicate a shift of the peak at 1652 cm^{-1}, which is attributed to stretching of the rA and rB rings of the drug (Table S6 and Fig. S11, ESI†). Similarly, the positive 1408 and negative 1417 cm^{-1} peaks reflect a downshift of the peak at 1414 cm^{-1}, which is mainly associated with the C–C/C–N stretching modes of the rA and rB rings. Conversely, no significant change in the peak at 1612 cm^{-1} was observed, corresponding to the stretching modes of rC and rD rings of the drug. These data suggest that rA and rB rings should be primarily involved in RNA binding.

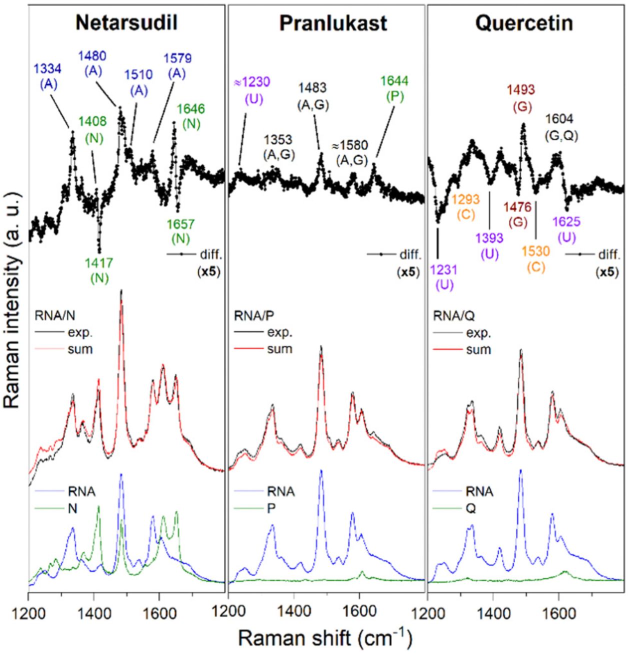


Fig. 2 UVRR spectra to study the binding of Netarsudil (N), Pranlukast (P), and Quercetin (Q) to G4-1 RNA. From bottom to top: RNA (blue); drug (green); RNA/drug complex (black); arithmetic sum of RNA and drug spectra (red); normalized difference between the spectra of the complex and the arithmetic sum (*diff.*). Spectra were normalized with the RNA spectrum intensity.

Pranlukast and Quercetin showed negligible UVRR spectral contributions compared to G4-1, so binding-induced spectral perturbations could be easily correlated with the band changes of nucleotides. Upon RNA/Quercetin interaction, a decrease in intensity of the difference spectrum at 1231, 1293, 1393, 1530 and 1625 cm^{-1}, related to bands of U and C residues (Fig. 2), was observed, suggesting their involvement in the interaction. Noteworthy, the presence of a minimum and a maximum at 1476 and 1493 cm^{-1}, respectively, due to a shift of the band at 1483 cm^{-1} which is mainly related to guanines,[21] indicate that these residues take part in the interaction. This is confirmed by the intensity increase and redshift of the band at 1606 cm^{-1} which could be due to both changes in the guanine normal modes and a downshift of the Quercetin peak (ring stretching vibration) from 1620 to 1604 cm^{-1} (Table S7, ESI†), overall suggesting that end-stacking is the favoured binding mode for this drug. As for Pranlukast, the difference spectrum shows changes associated to A, G, and U, although of lower intensity than those observed for Quercetin. Interestingly, a clear variation of the peak at 1642 cm^{-1} is observed, which corresponds mainly to C=O2 and C=O4 stretching vibrations of the drug in combination with N1–H and C26–H bending (Table S8, ESI†). Although it is a clear indication that Pranlukast binds to G4-1, this is not enough to obtain precise information on the molecular regions involved in the interaction.

To elucidate the binding poses of the three drugs to G4-1, docking and molecular dynamics simulations (MDs) were performed. The best docking poses showed that, as suggested by experimental data, Quercetin and Pranlukast preferentially bind to the 5′-end G-tetrad *via* stacking interactions with G11 and G14 and H-bonds with U4 and C13 (Fig. S14, ESI†). Conversely,

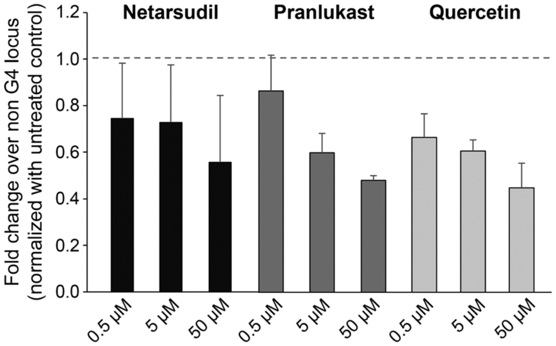


Fig. 3 Fold change in reverse transcription of the G4-1 region over a non-G4-forming control region from a synthetic SARS-CoV-2 RNA in the presence of the drugs. Untreated conditions were used as a control.

Netarsudil was found to interact with the loop on the 3′-end side, in agreement with UVRR results. To further verify the docking results and the geometrical stability of G4-1 upon binding, each complex was submitted to 500 ns of MDs. The most populated poses extracted after MDs are shown in Fig. S15 (ESI†). Root Mean Square Deviation (RMSD) analysis revealed that Quercetin adopts a stable binding pose (Fig. S16A, ESI†), keeping the H-bonds with U4 and C13 for ∼62% of MDs (Table S9, ESI†). As for Pranlukast, despite its chromen moiety is anchored by an H-bond with U4 for ∼82% of MDs (Table S9, ESI†), the flexibility of the phenylbutoxy group induces a slight change in the binding mode (Fig. S16B, ESI†). On the other hand, Netarsudil undergoes a deeper change of its binding mode during MDs (Fig. S16C and Movie S1, ESI†). This is mainly due to the CAAU loop, which fluctuates more than in the other complexes (Fig. S17, ESI†). The binding arrangements of Netarsudil are well highlighted by the analysis of solvent-accessible surface area (Fig. S18, ESI†), which decreases mainly for Netarsudil, A9 and U10 residues following the establishment of π–π stacking interactions between the isoquinoline and the phenyl moieties of the drug and the two RNA residues (Fig. S15C, ESI†). This explains the slightly higher RMSD values for G4-1 in the RNA/Netarsudil complex with respect to the G4-1 free state, contrarily to the Quercetin and Pranlukast effects (Fig. S19, ESI†). Nonetheless, the 2dRMSD matrices show the high tendency of Netarsudil to stabilize the G4-1 (Fig. S20, ESI†) as observed in the other experiments. Overall, these results fully agree with the band changes in the UVRR data induced by ligand binding. Changes in the G4-1 RNA geometry were also analysed by calculating some basic G4 parameters, *i.e.* the rotation (twist) angle between the two G-tetrads, their planarity, and gyration radii of the G-tetrads and whole G4 (Fig. S21, ESI†).[22] No significant changes in the position and planarity of G-tetrads and in G4 structural integrity were observed (Fig. S22–S24, ESI†), again indicating no major structural rearrangements in G4-1.

To evaluate if these drugs could interfere with biological functions such as viral RNA replication, we investigated whether their binding to G4 affected reverse transcription. A synthetic SARS-CoV-2 genomic RNA template was reverse transcribed using specific primers flanking the G4-1 or a non-G4-forming control region in the presence of growing concentrations of drugs. Then, G4-1 or the control region were used for quantitative RT-PCR to assay the levels of reverse transcription. As shown in Fig. 3, the drugs significantly reduced G4-1 reverse transcription compared to untreated samples, when normalized with the control region, thus showing that they may be able to interfere with viral replication.

In conclusion, three FDA approved drugs were identified as effective SARS-CoV-2 G4 binders through a ligand-based drug repurposing strategy. Our results lay the basis for further studies aiming to evaluate the antiviral activity of such drugs, while the methodological approach employed will certainly impact medicinal chemistry approaches for targeting of viral RNA G4s, even beyond SARS-CoV-2.

This work was supported by Italian Ministry of University and Research (FISR2020IP_04932 to JA). The authors thank CERIC for the support and facilities [20217014 to JA].

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang and J.-J. Zheng, *et al.*, *Nature*, 2020, **579**, 265–269.
2 E. J. Haas, F. J. Angulo, J. M. McLaughlin, E. Anis, S. R. Singer, F. Khan, N. Brooks, M. Smaja, G. Mircus, K. Pan, J. Southern, D. L. Swerdlow, L. Jodar, Y. Levy and S. Alroy-Preis, *Lancet*, 2021, **397**, 1819–1829.
3 K. B. Pouwels, E. Pritchard, P. C. Matthews, N. Stoesser, D. W. Eyre, K.-D. Vihta, T. House, J. Hay, J. I. Bell, J. N. Newton, J. Farrar, D. Crook and D. Cook, *et al.*, *Nat. Med.*, 2021, **27**, 2127–2135.
4 T. T. Ashburn and K. B. Thor, *Nat. Rev. Drug Discovery*, 2004, **3**, 673–683.
5 S. Pushpakom, F. Iorio, P. A. Eyers, K. J. Escott, S. Hopper, A. Wells, A. Doig, T. Guilliams, J. Latimer, C. McNamee, A. Norris, P. Sanseau, D. Cavalla and M. Pirmohamed, *Nat. Rev. Drug Discovery*, 2019, **18**, 41–58.
6 G. Ciliberto and L. Cardone, *Drug Discovery Today*, 2020, **25**, 946–948.
7 Y. L. Ng, C. K. Salim and J. J. H. Chu, *Pharmacol. Ther.*, 2021, **228**, 107930.
8 W. D. Jang, S. Jeon, S. Kim and S. Y. Lee, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**, e2024302118.
9 O. Minenkova, D. Santapaola, F. M. Milazzo, A. M. Anastasi, G. Battistuzzi, C. Chiapparino, A. Rosi, G. Gritti, G. Borleri, A. Rambaldi and C. Dental, *et al.*, *Mol. Ther.*, 2022, **30**, 1979–1993.
10 S. Sreeramulu, C. Richter, H. Berg, M. A. Wirtz Martin, B. Ceylan, T. Matzel, J. Adam, N. Altincekic, K. Azzaoui and J. K. Bains, *et al.*, *Angew. Chem., Int. Ed.*, 2021, **60**, 19191–19200.
11 C. Zhao, G. Qin, J. Niu, Z. Wang, C. Wang, J. Ren and X. Qu, *Angew. Chem., Int. Ed.*, 2021, **60**, 432–438.
12 (*a*) G. N. Parkinson, M. P. Lee and S. Neidle, *Nature*, 2002, **417**, 876–880; (*b*) S. Neidle and G. N. Parkinson, *Biochimie*, 2008, **90**, 1184–1196.
13 E. Lavezzo, M. Berselli, I. Frasson, R. Perrone, G. Palù, A. R. Brazzale, S. N. Richter and S. Toppo, *PLoS Comput. Biol.*, 2018, **14**, e1006675.
14 A. Abiri, M. Lavigne, M. Rezaei, S. Nikzad, P. Zare, J.-L. Mergny and H.-R. Rahimi, *Pharmacol. Rev.*, 2021, **73**, 897–923.
15 E. Ruggiero, I. Zanin, M. Terreri and S. N. Richter, *Int. J. Mol. Sci.*, 2021, **22**, 10984.
16 H. Cui and L. Zhang, *Front. Microbiol.*, 2020, **11**, 567317.
17 D. Ji, M. Juhas, C. M. Tsang, C. K. Kwok, Y. Li and Y. Zhang, *Briefings Bioinf.*, 2021, **22**, 1150–1160.
18 E. Belmonte-Reche, I. Serrano-Chacón, C. Gonzalez, J. Gallo and M. Bañobre-López, *PLoS One*, 2021, **16**, e0250654.
19 A. Maruca, F. A. Ambrosio, A. Lupia, I. Romeo, R. Rocca, F. Moraca, C. Talarico, D. Bagetta, R. Catalano, G. Costa, A. Artese and S. Alcaro, *Phys. Sci. Rev.*, 2019, **4**, 20180113.
20 D. Monchaud and M.-P. Teulade-Fichou, *Methods in Molecular Biology*, 2010, pp. 257–271.
21 S. Di Fonzo, J. Amato, F. D'Aria, M. Caterino, F. D'Amico, A. Gessini, J. W. Brady, A. Cesàro, B. Pagano and C. Giancola, *Phys. Chem. Chem. Phys.*, 2020, **22**, 8128–8140.
22 V. Tsvetkov, G. Pozmogova and A. Varizhuk, *J. Biomol. Struct. Dyn.*, 2016, **34**, 705–715.

PAPER V

Unveiling the interaction between DNA G-quadruplexes and RG-rich peptides

Grasso N., Graziano R., **Marzano S.**, D'Aria F., Merlino F., Grieco P., Randazzo A., Pagano B., Amato J.

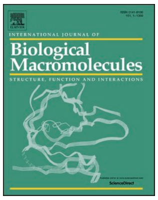
International Journal of Biological Macromolecules, 2023, Vol. 253(Part 3)

International Journal of Biological Macromolecules 253 (2023) 126749

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac



Unveiling the interaction between DNA G-quadruplexes and RG-rich peptides

Nicola Grasso [1], Raffaele Graziano [1], Simona Marzano, Federica D'Aria, Francesco Merlino, Paolo Grieco, Antonio Randazzo, Bruno Pagano *, Jussara Amato *

Department of Pharmacy, University of Naples Federico II, 80131 Naples, Italy

ARTICLE INFO

Keywords:
G-quadruplex
Peptide
RGG motif
DNA
Biophysics

ABSTRACT

G-quadruplexes are non-canonical DNA secondary structures formed within guanine-rich strands that play important roles in various biological processes, including gene regulation, telomere maintenance and DNA replication. The biological functions and formation of these DNA structures are strictly controlled by several proteins that bind and stabilize or resolve them. Many G-quadruplex-binding proteins feature an arginine and glycine-rich motif known as the RGG or RG-rich motif. Although this motif plays a crucial role in the recognition of such non-canonical structures, their interaction is still poorly understood. Here, we employed a combination of several biophysical techniques to provide valuable insights into the interaction between a peptide containing an RGG motif shared by numerous human G-quadruplex-binding proteins (NIQI) and various biologically relevant G-quadruplex DNA structures with different topologies. We also shed light on the key amino acids involved in the binding process. Our findings contribute to lay the basis for the development of a new class of peptide-based G-quadruplex ligands as an alternative to small molecules. These ligands may serve as valid tools for interfering in DNA-protein interactions, with potential therapeutic applications.

1. Introduction

Protein-DNA interactions play pivotal roles in almost all cellular processes, including regulation of transcription, replication, recombination, repair, chromatin organization, and more. Among the extensively investigated proteins are transcription factors, which bind to specific DNA sequences and modulate gene transcription by either activating or inhibiting it [1]. Besides proteins that recognize specific DNA sequences [2], some proteins bind to specific DNA structures, such as triplex, cruciform, or G-quadruplex (G4) structures [3–6]. Therefore, the formation of certain DNA structures is a crucial determinant for the effective binding of proteins involved in specific processes, such as tumorigenesis. An example is provided by the tumor suppressor protein p53: the imbalance of mutant p53 binding to target sites due to binding to other DNA structural motifs, such as G4s, appears to play a critical role in human tumor progression [7].

G4s are non-canonical nucleic acid secondary structures formed within DNA or RNA guanine (G)-rich strands, characterized by the presence of two or more G-tetrads, i.e. cyclic planar arrays of four guanines held together by Hoogsteen hydrogen bonds, which stack one each other by π–π interactions, and further stabilized by the presence of monovalent cations, usually K^+ or Na^+ (Fig. 1) [8,9]. G4s are characterized by high structural polymorphism, in fact they can adopt a wide variety of structures depending on the molecularity, the strand(s) direction, as well as the length and composition of the loops [10]. According to the molecularity, G4 structures can be distinguished as intramolecular or intermolecular, while considering the direction of the strands, they can be classified as parallel, antiparallel, or hybrid (Fig. 1). Interestingly, the topologies of G4s have an important influence on their stability, recognition by proteins or drugs, and in general on G4-related biological functions [10–13].

The genome-wide prevalence of G4s has also been demonstrated. Bioinformatics studies revealed that putative intramolecular G4-forming sequences are prevalent in the human genome [14,15] and are significantly enriched in oncogene promoters [16]. These findings were greatly supported by the experimental detection of G4s [17,18]. Even though the number of G4 structures found in human chromatin is substantially lower than predicted by computations, thus showing that not

---

* Corresponding authors.
E-mail addresses: bruno.pagano@unina.it (B. Pagano), jussara.amato@unina.it (J. Amato).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ijbiomac.2023.126749
Received 5 August 2023; Received in revised form 28 August 2023; Accepted 31 August 2023

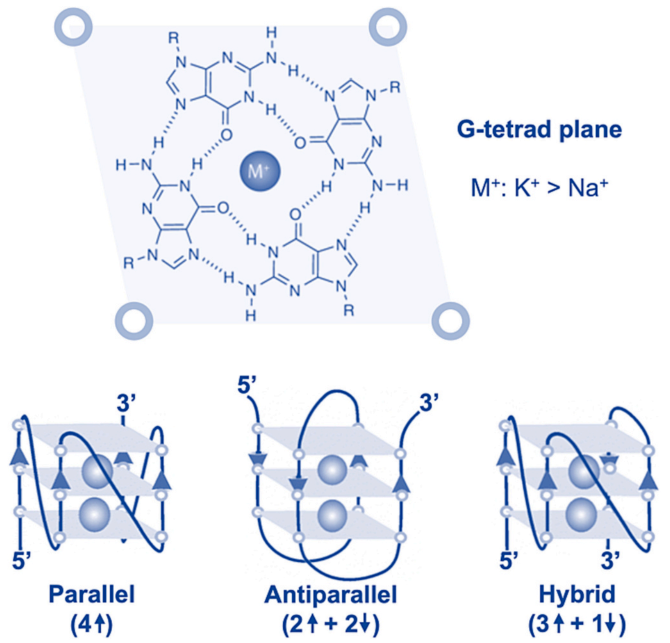


Fig. 1. (Top) Chemical structure of a G-tetrad showing the Hoogsteen hydrogen bonding between guanines and a central monovalent cation (M^+) that stabilizes the G4 structure, and (bottom) examples of representative topologies commonly found in G4 structures. The vertical arrows indicate the direction of the part of the strand involved in G-tetrad formation.

all sequences with G4-forming potential form these structures in a cellular context, they are particularly enriched in cancer-related genes and regions predisposed to amplification in cancer [18]. These data strongly support the potential of the G4 structural motif as a molecular target for therapeutic intervention.

G4s have been observed to adopt folded conformations under certain conditions, as a result of specific stimuli, during particular cell cycle phases, or in a cell type-specific manner [19]. Moreover, the formation of these DNA structures is strictly controlled by several proteins that bind and stabilize or resolve them [5,20,21].

Protein specificity for different G4 structures may arise from G4 strand orientation, loop features, groove dimensions, G4-duplex junctions, and/or a combination of these characteristics [5]. Several studies suggest that human G4-binding proteins should be considered a specific group of nucleic acid-binding proteins since they differ from other proteins for the significant enrichment of glycine (G) and arginine (R) residues [22,23]. Indeed, G4-binding proteins are particularly abundant in RGG domains/RG-rich motifs and exhibit depletion of LL-rich sequences.

By comparing the amino acid sequences of 77 human G4-binding proteins available at the time of analysis, Brazda et al. identified a common 20 amino acid-long RG-rich motif of sequence RGRGRGRGGGSGGSGGRGRG, named NIQI (Novel Interesting Quadruplex Interaction motif), [23]. Interestingly, the NIQI motif consists almost exclusively of R and G residues with a few serine (S) alternations. Both arginine and glycine are well known disorder-promoting amino acids, and they exhibit a relatively high level of flexibility (particularly glycine residues) [24]. From this point of view, NIQI is supposed to be inherently flexible to enable arginine and serine residues to interact, possibly by forming hydrogen bonds, with DNA in a wide range of G4 structures (parallel, antiparallel, or hybrid types).

The interaction of RGG domains with G4s has sparked great interest, and the importance of elucidating RGG-mediated G4 recognition is underlined by the observation that this domain is indispensable for protein function [25–27]. On the other hand, besides computational predictions, the interaction of NIQI with various G4s has never been thoroughly investigated experimentally.

Gaining a comprehensive understanding of the true nature and specificity of NIQI's recognition of different G4 structures is essential, not only to shed light on the G4-protein interaction processes but also for the design of peptide-based G4 ligands potentially capable of targeting DNA-protein interactions to control oncogene expression [28,29]. This strategy is particularly attractive, especially considering that the equilibria between different DNA secondary structures are regulated by different proteins, and that formation and resolution of G4s appear to serve as a regulatory mechanism in gene transcription within cancer cells [30–32].

Investigating the binding of NIQI to different G4s and how this recognition happens is definitely worth exploring. Herein, we have characterized the interaction of the peptide with biologically relevant DNA G4s with different topologies (parallel, antiparallel, or hybrid) by using a combination of biophysical techniques, including circular dichroism (CD) spectroscopy and microscale thermophoresis (MST). The interaction between NIQI and the most promising G4 targets has been further studied from both a thermodynamic and structural point of view through isothermal titration calorimetry (ITC) and nuclear magnetic resonance (NMR) experiments. To further detail the interactions, key amino acids in the G4-binding peptide were determined by introducing point mutations. Overall, this study aims at providing new insights into the interaction between RG-rich peptides and G4 structures, contributing to paving the way for the design and application of a novel class of peptide-based G4 ligands.

2. Materials and methods

2.1. Materials

N^{α}-Fmoc-protected amino acids [Fmoc-Gly, Fmoc-Ser(tBu), Fmoc-Arg(Pbf)], coupling reagents, [*N*,*N*,*N*′,*N*′-tetramethyl-O-(1H-benzotriazol-1-yl) uranium hexafluorophosphate (HBTU), 1-hydroxybenzotriazole (HOBt)], and the Rink amide resin (0.72 mmol/g of loading substitution) were purchased from GL Biochem Ltd. (Shanghai, China). *N*,*N*-diisopropylethylamine (DIPEA), piperidine, triethylamine (Et_3N), Fmoc-8-amino-3,6-dioxaoctanoic acid (Fmoc-O_2Oc-OH) and trifluoroacetic acid (TFA) were purchased from Iris-Biotech GmbH (Marktredwitz, Germany). Fluorescein isothiocyanate isomer I (FITC) was obtained from Merck KGaA (Darmstadt, Germany). Solvents used for peptide synthesis and purification [*N*,*N*-dimethylformamide (DMF), dichloromethane (DCM), diethyl ether (Et_2O), water, and acetonitrile (ACN)] were obtained from Merck KGaA and used without further purification. CPG supports, DNA phosphoramidites, and all reagents and solvents for oligonucleotide synthesis and purification were purchased from Merck KGaA and used without further purification. All buffers were produced from highly purified Milli-Q water and filtered before use.

2.2. Peptide synthesis

The peptide sequences of NIQI (RGRGRGRGGGSGGSGGRGRG) and its Ala derivatives were assembled by using the ultrasound-assisted solid-phase peptide synthesis (US-SPPS) protocol, integrated with the Fmoc/*t*Bu orthogonal protection strategy [33]. Peptide elongation was performed on the Fmoc Rink amide resin as solid support (0.1 mmol from 0.72 mmol/g as loading, 100–200 mesh as particle size) to obtain amidated C-terminus. The resin was placed into a 10 mL polypropylene tube, equipped with a filter, stopper, and top cap, and swollen in DMF on an automated shaker for 30 min at rt. First, the Fmoc group was removed from the Rink amide linker by treatment with 20 % piperidine in DMF solution and using ultrasonic irradiation (0.5 + 1 min). After each step, filtering and washings of the resin were executed (3 × 2 mL of DMF; 3 × 2 mL of DCM). Then, coupling reactions were performed by adding a solution of N^{α}-Fmoc-amino acid, HBTU, HOBt (3 equiv), and DIEA (6 equiv) in DMF to the resin, and exposing the resin to ultrasound irradiation for 10 min. Fmoc deprotection and coupling reactions were monitored by performing the colorimetric Kaiser test. Once completed the elongation of the sequences, the N-terminal Fmoc group was removed, and peptides were released from the resin and deprotected. The fluorescein labelling of NIQI peptide (FITC-O2Oc-RGRGRGRGGGSGGSGGRGRG, FITC-NIQI) was obtained as described elsewhere [34]. All peptides were cleaved from the support and deprotected by treating with a mixture of TFA/TIS/H_2O (95:2.5:2.5, *v/v/v*), at rt for 3 h. All peptides were purified by HPLC (Shimadzu Preparative Liquid Chromatography LC-8A) equipped with a preparative column (Phenomenex Kinetex C18 column, 5 μm, 100 Å, 150 × 21.2 mm) by using a linear gradient of ACN (0.1 % TFA) in water (0.1 % TFA) from 5 to 60 % over 30 min, with a flow rate of 10 mL/min, and UV detection at 210 nm. Final products were obtained by lyophilization of the appropriate fractions after removal of the ACN by rotary evaporation. Each peptide was ascertained to be of purity >95 % by analytical UHPLC (Shimadzu Nexera Liquid Chromatograph LC-30 AD) equipped with a Phenomenex Kinetex reverse-phase column (C18, 5 μm, 100 Å, 150 × 4.6 mm) with a flow rate of 1 mL/min by using a gradient of ACN (0.1 % TFA) in water (0.1 % TFA), from 0 to 60 % over 20 min, and the UV detection set at 210 nm (Table S1). Peptides were confirmed by high-resolution mass spectrometry (HRMS) (LTQ Orbitrap) (Table S1).

2.3. Oligonucleotide synthesis and sample preparation

DNA sequences listed in Table 1 were synthesized on an ABI 394 DNA/RNA synthesizer (Applied Biosystem) at 5-μmol scale by using standard β-cyanoethylphosphoramidite solid phase chemistry. DNA detachment from support and its deprotection were carried out by means of an aqueous solution of concentrated ammonia at 55 °C for 17 h. The filtrates and the washings were combined and concentrated under reduced pressure, solubilized in water, and then purified by high-performance liquid chromatography (HPLC) equipped with a Nucleogel SAX column (Macherey-Nagel, 1000–8/46) using a 30 min linear gradient going from 100 % buffer A to 100 % buffer B with a flow rate of 1 mL/min [buffer A: 20 mM KH_2PO_4/K_2HPO_4 aqueous solution (pH 7.0), containing 20 % (*v*/v) ACN; buffer B: 1 M KCl, KH_2PO_4/K_2HPO_4 aqueous solution (pH 7.0), containing 20 % (v/v) ACN]. The purified fractions of the oligomers were then desalted by using C-18 cartridges (Sep-pak). The purity of the isolated oligomer was evaluated by NMR, and it turned out to be higher than 98 %. Next, oligonucleotides were lyophilized and then resuspended in 5 mM KH_2PO_4/K_2HPO_4 aqueous solution (pH 7.0), containing 20 mM KCl (or LiCl in the case of *c-Myc*) and their concentration was established by measuring the UV absorption at 90 °C considering the molar extinction coefficient values ε (λ = 260 nm) determined by the nearest neighbor model [35]. DNA samples were heated at 90 °C for 5 min, then gradually cooled to room temperature overnight, and finally stored at 4 °C for 24 h before use.

Table 1
List of DNA sequences employed in the study, NIQI-induced thermal stabilization (**ΔT_m**) for G4 and duplex DNA structures measured by CD melting experiments, and equilibrium dissociation constants (K_d) for the binding of NIQI to DNAs obtained by MST experiments.

DNA	Sequence (5′ → 3′)	ΔT_m (°C)[a]	K_d (nM)
Parallel G4s			
c-Kit1	AGGGAGGGCGCTGGGAGGAGGG	6.3 (±0.2)	63 (±11)
c-Kit2	CGGGCGGGCGCTAGGGAGGGT	12.6 (±0.2)	10 (±6)
c-Myc	TGAGGGTGGGTAGGGTGGGTAA	7.9 (±0.4)	60 (±10)
Antiparallel G4s			
HRAS1	TCGGGTTGCGGGCGCAGGGCACGGGCG	1.6 (±0.4)	1000 (±100)
LWDLN1	GGGTTTGGGTTTTGGGAGGG	−3.2 (±0.2)	460 (±60)
Hybrid G4s			
Bcl-2	GGGCGCGGGAGGAATTGGGCGGG	10.5 (±0.2)	27 (±9)
m-Tel$_{24}$	TTGGGTTAGGGTTAGGGTTAGGGA	3.7 (±0.4)	200 (±30)
Duplex			
ds$_{26}$	CAATCGGATCGAATTCGATCCGATTG	2.0 (±0.4)	>20,000

[a] ΔT_m represents the difference in melting temperature [$\Delta T_m = T_m$ (DNA + 3 peptide equiv) - T_m (DNA)]. The T_m values of DNA alone are: *c-Kit1* = 55.8 ± 0.1 °C, *c-Kit2* = 62.1 ± 0.1 °C, *c-Myc* = 68.0 ± 0.2 °C, *HRAS1* = 53.1 ± 0.2 °C, *LWDLN1* = 49.6 ± 0.1 °C, *Bcl-2* = 60.5 ± 0.1 °C, *m-Tel$_{24}$* = 54.1 ± 0.2 °C, *ds$_{26}$* = 63.7 ± 0.2 °C. All experiments were performed in 5 mM KH_2PO_4/K_2HPO_4 buffer (pH 7.0) containing 20 mM KCl (or 20 mM LiCl in the case of *c-Myc*, because of its high thermal stability in KCl containing buffer). All experiments were performed in triplicate, and the reported values are the average of three measurements.

2.4. Circular dichroism (CD) spectroscopy

CD experiments were carried out on a Jasco J-815 spectropolarimeter equipped with a PTC- 423S/15 Peltier temperature controller. G4-forming oligonucleotides, as well as the duplex-forming DNA (*ds$_{26}$*), were prepared at a concentration of 2 μM in 5 mM KH_2PO_4/K_2HPO_4 (pH 7.0) containing 20 mM KCl, then annealed by heating at 90 °C for 5 min followed by slow cooling to room temperature overnight. CD spectra of DNA/peptide mixtures were obtained by adding 3 mol equiv of NIQI (stock solutions of peptide were 10 mM in H_2O) with respect to the oligonucleotide. Spectra of DNA molecules in the absence and presence of NIQI were recorded at 20 and 100 °C in the wavelength range of 220–320 nm using a scan rate of 100 nm/min, with a 1 s response time and 1 nm bandwidth. CD spectra of NIQI in the absence and presence of *c-Kit2* G4 were also recorded at 20 °C in the 200–250 nm wavelength range using a scan rate of 20 nm/min, with a 4 s response time and 1 nm bandwidth. All spectra were baseline subtracted and averaged over three scans. Subtraction of *c-Kit2* G4 spectrum from that of G4/peptide mixture was performed to have information about the spectrum of NIQI in the presence of G4. The percentage of secondary structures adopted by the peptide in the absence and presence of the G4 was estimated by using the BeStSel software [36]. CD melting experiments were carried out in the 20–100 °C temperature range at 1 °C/min heating rate by following CD signal changes at the wavelengths of the maximum CD intensity (264 nm for *c-Kit1*, *c-Kit2*, *c-Myc,* and *Bcl-2*; 290 nm for *m-Tel$_{24}$*, *HRAS1*, and *LWLDN1*; 252 nm for *ds$_{26}$*). CD melting experiments were performed in the absence and presence of the peptide (3 mol equiv) added to the folded DNA structures. The apparent melting temperatures (T_m) were determined from a curve fit using Origin 7.0 software, and the ΔT_m values calculated as the difference in the T_m values of the DNA structures in the presence and absence of the peptide. All experiments were performed in triplicate, and the reported values are the average of the three measurements.

2.5. Microscale thermophoresis (MST)

MST measurements were performed on a Monolith NT.115 (Nanotemper Technologies, Munich, Germany). Experiments were performed at 25 °C, using auto-tune LED power and medium MST power. The fluorescein isothiocyanate labelled peptide (FITC-NIQI) was prepared at 1 μM in H_2O, then diluted to 100 nM with the 5 mM KH_2PO_4/K_2HPO_4 buffer (pH 7.0), containing 20 mM KCl and supplemented with 0.1 % Tween and 5 % DMSO. Oligonucleotides were prepared at 1 mM concentration in 5 mM KH_2PO_4/K_2HPO_4 buffer (pH 7.0), containing 20 mM KCl, and annealed as described above. For the MST experiments, the final concentration of the labelled peptide was kept constant at 50 nM, while a serial dilution of the DNAs (1:2 from 2 or 40 mM) in the buffer supplemented with 0.1 % Tween and 5 % DMSO, was prepared and mixed with the peptide solution in a 1:1 volume ratio. All the samples were loaded into standard capillaries (NanoTemper Technologies). MST data analysis was performed by employing the MO.Affinity Analysis software (v2.3) provided with the instrument.

2.6. Isothermal titration calorimetry (ITC)

ITC measurements were performed at 25 °C using a nano-ITC Low Volume calorimeter (TA instruments, Lindon, UT, USA). DNA and peptide solutions were all prepared with the same batch of buffer (5 mM KH_2PO_4/K_2HPO_4 buffer, pH 7.0, containing 20 mM KCl), to minimize the differences in buffer composition and pH. In each titration, volumes of 2 μL of peptide solution (200–400 μM) were injected, using a computer-controlled 50 μL syringe, into the calorimetric vessel (170 μL) containing the oligonucleotide (16–20 μM) with a spacing of 300 s between each injection to allow the system to reach the equilibrium. Heat produced by peptide dilution was evaluated in a control experiment by injecting the peptide solution into the buffer alone. The interaction heat for each injection was calculated after correction for the heat of peptide dilution. The corrected heat values were plotted as a function of the molar ratio, to give the corresponding binding isotherms, which were fitted with a multiple-sites binding model by means of the NanoAnalyze software (TA instruments) supplied with the instrument, to give the binding enthalpy (ΔH°), equilibrium binding constant (K_a), and stoichiometry of interaction (n). The Gibbs free-energy change and the entropy change were derived using the following relationships: $\Delta G^\circ = -RT \ln K_a$ ($R = 8.314$ J mol^{-1} K^{-1}, $T = 298$ K) and $T\Delta S^\circ = \Delta H^\circ - \Delta G^\circ$. All measurements were performed in triplicate.

2.7. Nuclear magnetic resonance (NMR) spectroscopy

1D ^{1}H NMR experiments were carried out on a Bruker AVANCE NEO NMR spectrometer, operating at 600 MHz (^{1}H Larmor frequency), equipped with a SampleJet autosampler and a 5 mm QCI H-P/C/N-D-5-Z CryoProbe, optimized for ^{1}H sensitivity. Experiments were carried out using 20 μM of *c-Kit2* or *Bcl-2* G4s in 5 mM KH_2PO_4/K_2HPO_4 buffer, pH 7.0, containing 20 mM KCl and 10 % D_2O, in the absence and presence of different amounts of NIQI. Spectra were recorded at 298 K, using 512 scans per spectrum with a recovery delay of 1.5 s, and suppressing water signal using the excitation sculpting with gradients [37]. Spectra were phase-adjusted, baseline-corrected, and calibrated with respect to the water frequency. Processing and analysis were performed using the software package Bruker TOPSPIN 4.0.7.

3. Results and discussion

3.1. Research design

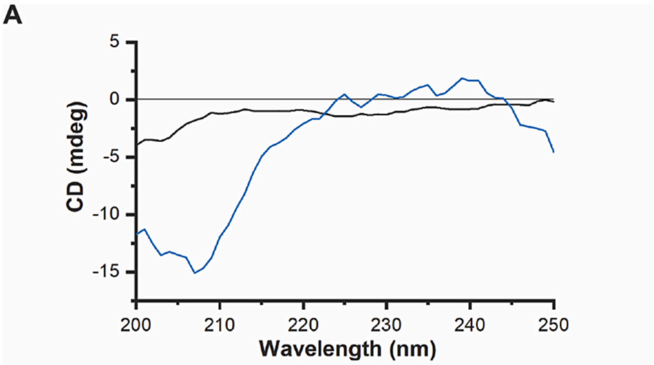
A series of G4-forming sequences found in human oncogene promoter regions (*Bcl-2*, *c-Kit1*, *c-Kit2*, *c-Myc*, *Her2*, *HRAS1*, and *LWDLN1*) as well as the telomeric sequence (*m-Tel$_{24}$*), which are capable of adopting different G4 topologies (parallel, antiparallel, and hybrid), were selected for this study (Table 1). These sequences were used to investigate the binding ability of the peptide NIQI (RGRGRGRGGGSGGSGGRGRG) and its capacity to discriminate between various G4 topologies. In addition, a 26-nucleotide duplex-forming sequence (*ds$_{26}$*) was also included to assess the selectivity of the peptide for G4 structures over duplex DNA.

3.2. Circular dichroism experiments

CD spectroscopy is a valuable technique for determining the secondary structures assumed by peptides and nucleic acids in solution. CD spectra in the wavelength range from 250 to around 190 nm provide information on the presence of different types of secondary structures in peptides. For instance, α-helical peptides exhibit two negative bands at 208 and 222 nm, along with a positive one at 193 nm, and peptides with well-defined antiparallel β-sheets show a negative band at 218 nm and a positive one at 195 nm. On the other hand, disordered peptides exhibit very low ellipticity above 210 nm and negative bands near 195 nm [38]. The CD spectrum of NIQI, recorded in 5 mM KH_2PO_4/K_2HPO_4 aqueous buffer containing 20 mM KCl (pH 7.0), shows that it is essentially in a random conformation under these experimental conditions (Fig. 2A) [39]. The presence of a high content of glycine residues, known to introduce flexibility in the peptide backbone [40], prevents the formation of an ordered structure for this peptide in an aqueous buffer. This observation is in agreement with the results obtained from peptide structure prediction using AlphaFold2 [41] and PEP-FOLD3.5 [42] tools, which suggest that NIQI tends to adopt a random coil structure (Fig. S1).

CD spectroscopy is also a valuable tool for analyzing the secondary structure of oligonucleotides, particularly in the range of 220–320 nm, which is diagnostic of the G4 motifs [43,44]. Specifically, a parallel G4 structure is identified by a positive band around 264 nm and a negative band at 240 nm, whereas an antiparallel G4 structure shows a positive band at 295 nm and a negative one at 260 nm. Hybrid (or [3 + 1]) structures instead display two positive bands at around 290 and 264 nm, and a negative band at 240 nm. The CD spectra of *c-Kit1*, *c-Kit2*, and *c-Myc* exhibit the characteristic pattern associated with parallel-stranded G4s, indicating that all three sequences adopt this conformation (Fig. S2). On the other hand, *HRAS1* and *LWDLN1* display the characteristic CD profile indicative of antiparallel G4 structures, while *Bcl-2* and *m-Tel$_{24}$* exhibit the typical profile associated with hybrid G4 structures.

Additional CD experiments were performed to verify the capability of NIQI to alter the native folding topology of these G4s. Thus, DNA/NIQI mixtures were prepared by adding the peptide (3 mol equiv) to the folded G4 structures in solution. The results of this analysis (Fig. S2) indicate that NIQI influenced the CD spectra of the G4-forming sequences to different extents, with more pronounced effects observed in the case of the parallel G4s. In particular, NIQI induced a redshift of both positive and negative bands at 264 and 240 nm, respectively, in the CD spectra of all parallel G4s, together with a decrease in the intensity of the maximum and an increase in the depth of the minimum in the case of *c-Kit1* and *c-Kit2*, while a slight increase of the maximum was observed for *c-Myc*. Despite no considerable variations, the spectra of DNA in the



B

Estimated secondary structure content (%)

	Helix	Antiparallel	Parallel	Turn	Others
NIQI	0.0	37.2	0.0	17.1	45.7
NIQI + *c-Kit2*	17.5	29.8	0.0	10.8	41.9

Fig. 2. (A) CD spectra of NIQI (6 μM) recorded at 20 °C in 5 mM KH_2PO_4/K_2HPO_4 buffer (pH 7.0) containing 20 mM KCl in the absence (black line) and presence of *c-Kit2* G4 (blue line). To enable the detection of conformational changes in the peptide upon G4-binding, subtraction of the G4 signal was carried out. (B) Secondary structure content estimation for NIQI derived from the CD spectra using BeStSel.

presence of the peptide differ to some extent from those in its absence, thus suggesting DNA-peptide interaction. Different effects after the addition of the peptide were also observed for the hybrid G4s *Bcl-2* and *m-Tel*$_{24}$. Noticeable intensity variations of both the positive bands (at 265 and 290 nm) and the negative one (240 nm) occurred in the case of *Bcl-2*. In particular, the concomitant decrease in the intensity of the two positive bands, more pronounced for the one at 265 nm compared to 290 nm, may indicate a structural rearrangement of the G4 upon peptide interaction. Conversely, only very small changes in the CD profile were observed for *m-Tel*$_{24}$ upon peptide addition, as well as for the *HRAS1* and *LWDLN1* antiparallel G4 structures. Altogether, these results may suggest a slightly different binding behavior of NIQI to the investigated G4 structures.

The stabilizing properties of NIQI were then evaluated by CD melting experiments measuring the peptide-induced changes in the melting temperature (ΔT_m) of the G4s. Results of these experiments, shown in Fig. S3 and summarized in Table 1, indicate that NIQI induces thermal stabilization in most of the G4s, except for *HRAS1* and *LWDLN1* (both belonging to the antiparallel G4 group). However, NIQI does not exhibit inherent selectivity for a particular G4 topology. Indeed, the highest thermal stabilization effects were observed for *c-Kit2* (parallel G4, $\Delta T_m = 12.6$ °C), followed by *Bcl-2* (hybrid G4, $\Delta T_m = 10.5$ °C), *c-Myc* and *c-Kit1* G4s (both parallel G4s, $\Delta T_m = 7.9$ and 6.3 °C, respectively). Conversely, the peptide displayed weak stabilization of *m-Tel*$_{24}$ (hybrid G4, $\Delta T_m = 3.7$ °C) and almost no effects on *HRAS1* (antiparallel G4, $\Delta T_m = 1.6$ °C). On the other hand, NIQI showed a slight destabilizing effect on *LWDLN1* G4 (antiparallel G4, $\Delta T_m = -3.2$ °C), which might be attributed to a possible distortion of this G4 upon peptide interaction.

Additionally, with the aim of evaluating NIQI selectivity for G4 structures over double-stranded DNA, CD experiments were also performed by using *ds*$_{26}$ as a representative duplex model. The CD spectrum of *ds*$_{26}$ typically shows a positive band at around 280 nm and a negative one at around 250 nm (Fig. S2). These bands were not significantly altered after the addition of the peptide. Moreover, duplex thermal stability was not significantly affected by the peptide ($\Delta T_m = 2.0$ °C), indicating the preference of NIQI for G4 over duplex DNA (Fig. S3 and Table 1).

Finally, to investigate the conformational changes of NIQI upon G4 binding, an analysis of the CD spectrum of the peptide in the presence of *c-Kit2* in the 200–250 nm wavelength range was performed. *c-Kit2* was selected as it was the most stabilized G4 among those investigated. By subtracting the CD signal of the *c-Kit2* G4 from the spectrum of the G4/peptide mixture, it was possible to detect any conformational changes in NIQI upon interaction with G4 (Fig. S4). Interestingly, while NIQI exhibited an unstructured conformation in solution, the CD spectrum revealed the appearance of an α-helical signal in the presence of the G4 (Fig. 2A), displaying an increase in the estimated helical content from 0 to 17.5 % (Fig. 2B).

3.3. Studies on the interaction of NIQI with G4 structures by MST assay

To obtain quantitative data on the binding affinity of NIQI for the investigated G4 structures and to verify its selectivity for G4 over duplex DNA, MST was employed. Basically, MST enables the monitoring of the thermophoretic behavior of the target molecule in small temperature gradients, which is sensitive to ligand-induced alterations in size, charge, and/or hydration shell of the target [45]. Consequently, changes in the thermophoretic behavior of the target can be used to determine the equilibrium dissociation constant, K_d. For MST experiments, serial dilutions of DNA molecules were prepared and mixed with a constant concentration of FITC-labelled peptide, loaded into capillaries, and analyzed at 25 °C. The binding curves obtained from these experiments validated the binding of NIQI to all G4 structures, and notably, revealed a strong binding affinity of the peptide to certain G4s, with measured K_d values in the two-digit nanomolar range (Table 1 and Fig. S5).

Indeed, NIQI exhibited the highest affinity for *c-Kit2* ($K_d = 10$ nM), followed by *Bcl-2* ($K_d = 27$ nM), *c-Myc* ($K_d = 60$ nM), and *c-Kit1* ($K_d = 63$ nM). NIQI also exhibited a good affinity towards *m-Tel*$_{24}$ and *LWDLN1* ($K_d = 200$ and 460 nM, respectively), while displaying significantly lower affinity for *HRAS1* G4 ($K_d = 1000$ nM). Interestingly, NIQI displayed negligible binding to the *ds*$_{26}$ duplex ($K_d > 20$ μM), clearly indicating its selectivity for G4s over duplex DNA. These findings are in complete agreement with the results of CD-melting experiments and identify *c-Kit2* and *Bcl-2* as the best G4 targets of this series for NIQI.

3.4. Thermodynamic data for the binding of NIQI to c-Kit2 and Bcl-2 G4s by ITC experiments

In order to gain a deeper understanding of the binding of NIQI to *c-Kit2* and *Bcl-2* G4s, a thorough calorimetric analysis was performed by using ITC. ITC is a highly accurate method for determining the thermodynamic parameters of ligand binding to G4 DNA [46,47]. It is the only technique capable of quantifying both the enthalpic and entropic components of an interaction, thereby elucidating the thermodynamic driving forces involved in molecular recognition.

The results of these experiments are presented in Fig. 3 and summarized in Table 2 (raw ITC data for control experiment performed by injecting the peptide into the buffer are shown in Fig. S6). Specifically, panels A and B of Fig. 3 show the binding isotherms for the interaction of NIQI with *c-Kit2* and *Bcl-2* G4s, respectively, while the insets represent the corresponding raw ITC data.

The thermodynamic characterization reveals some interesting

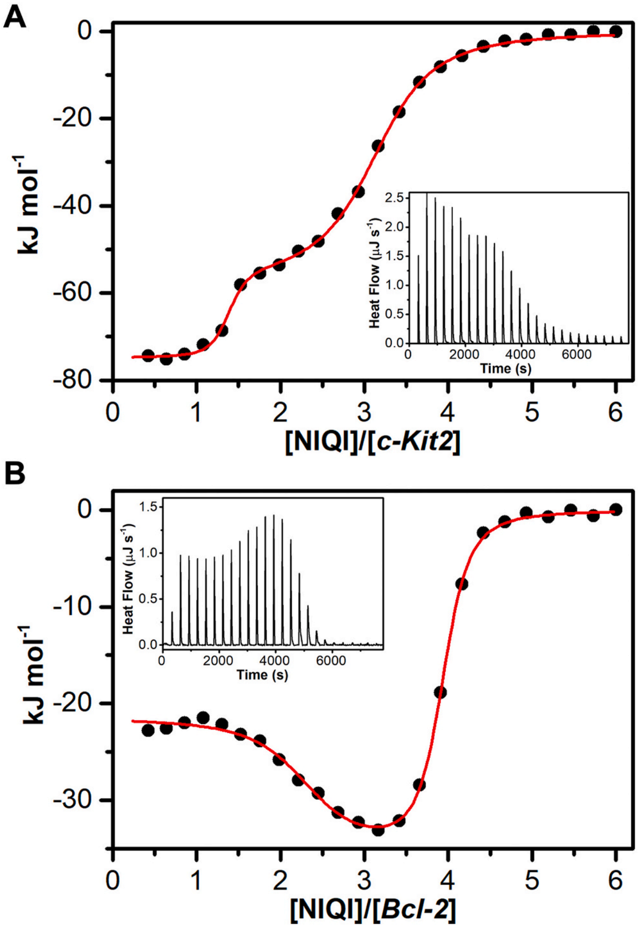


Fig. 3. Raw ITC data (insets) and binding isotherms for titration of (A) *c-Kit2* and (B) *Bcl-2* G4s with NIQI peptide obtained at 25 °C. The black dots represent the experimental data obtained by integrating the raw ITC data and subtracting the heat of peptide dilution into the buffer. The red lines represent the best fit obtained by multiple binding sites model.

Table 2
Thermodynamic parameters for the interaction of NIQI with *c-Kit2* and *Bcl-2* G4s obtained by ITC at 25 °C.

DNA	No. of events	n	K_a (M^{-1})	ΔG° (kJ mol^{-1})	ΔH° (kJ mol^{-1})	$T\Delta S^\circ$ (kJ mol^{-1})
c-*Kit*2	2	1	1.7×10^8	−47.0	−74.7	−27.7
		2	6.9×10^5	−33.3	−56.9	−23.6
Bcl-2	2	2	1.1×10^8	−45.9	−14.8	31.1
		2	2.8×10^6	−36.8	−25.9	10.9

features about the binding processes. Interestingly, both titrations result in binding isotherms revealing two-event binding processes, but with different features from each other.

Regarding *c-Kit2* G4, the binding isotherm exhibits a characteristic profile of a system wherein the binding site with higher affinity is accompanied by a more exothermic enthalpy change [48–50]. Indeed, the solid line (in red) drawn through the data points for *c-Kit2*, which represents the best fit to the data by a two-event model, allowed the estimation of two equilibrium constants, K_a, of 1.7×10^8 and 6.9×10^5 M^{-1} for the first and second event, respectively. Consequently, the first binding event was estimated to be >200-fold stronger than the second one. The overall stoichiometry for the binding of NIQI to *c-Kit2* was determined to be three peptides per G4 DNA, with one peptide involved in the first event and two peptides in the second one. The thermodynamic signatures of the two binding events (Table 2) indicate that the interaction of NIQI is enthalpically driven. Indeed, both binding events were characterized by unfavorable entropic contributions. The binding enthalpies (ΔH°) for the two events were determined to be −74.7 and −56.9 kJ mol^{-1} for the first and second event, respectively. These data clearly indicate that the driving force behind the binding process is the formation of new interactions between the ligand and the DNA. On the other hand, the opposing entropic contribution suggests that the resulting complex is more rigid compared to the free molecules.

Noteworthy, the thermodynamic characterization for the binding of NIQI to *Bcl-2* G4 reveals again a two-event binding process but with features different from that observed for *c-Kit2*. In this case, the binding isotherm has the typical trend of a system wherein the higher binding affinity site demonstrates a smaller exothermic enthalpy change than the lower binding affinity site [48], as confirmed by the ΔH° values (Table 2). The analysis reveals that the first event ($K_a = 1.1 \times 10^8$ M^{-1}) is about 40-fold stronger than the second one ($K_a = 2.8 \times 10^6$ M^{-1}). The stoichiometry indicates that one G4 equiv can bind up to four equiv of NIQI. The thermodynamic signature of the two binding events suggests that the interaction is both enthalpically and entropically driven. Indeed, the favorable enthalpic contribution (ΔH°), despite smaller than that observed for *c-Kit2*, acts synergistically with the entropic one ($T\Delta S^\circ$). Such a smaller enthalpic contribution clearly indicates that the driving force of the binding process is only partially represented by the formation of peptide-DNA interactions, but it may be also due to a relaxation of the G4 structure upon binding and/or to the displacement of water molecules, as suggested by the positive entropic contribution to the binding. The peptide-induced changes in the structural features of *Bcl-2* G4, as evidenced by the CD results (Fig. S2), also lend some support to this hypothesis.

Overall, ITC data indicate that the binding of NIQI to these G4 targets is driven by different thermodynamic contributions and suggest that the interactions could occur by distinct mechanisms, probably involving different regions of the G4 structures.

3.5. Studies on the interaction of NIQI with c-Kit2 and Bcl-2 G4s by NMR spectroscopy

To gain further information on the binding mode of NIQI to *c-Kit2* and *Bcl-2* G4s, ^{1}H NMR titration experiments were performed. According to the literature, under the experimental conditions used, *c-Kit2* and *Bcl-2* sequences form a single G4 conformation characterized by 12 well-resolved imino proton peaks, corresponding to the 12 guanines involved in the three G-tetrad planes (Fig. 4) [51,52]. The investigated oligonucleotides were titrated with NIQI up to a 4:1 peptide/G4 ratio. Upon addition of increasing amounts of NIQI, a decrease in signal intensity was observed in both the imino (Fig. 4) and aromatic (Fig. S7) proton regions of the spectra of the two G4s, indicating a strong binding of the peptide to the tested G4 structures. The drop in signals turned out to be more intense in the case of *c-Kit2* for which the titration was virtually completed at the 3:1 peptide/G4 ratio, while for *Bcl-2* some signals are still visible in the spectrum at the 4:1 peptide/G4 ratio. Interestingly, the differences observed for the imino proton resonances of guanines clearly suggest a different peptide binding behavior to the stem of *c-Kit2* and *Bcl-2* G4s.

In the case of *c-Kit2* G4, the imino proton signals most affected upon the addition of 1–2 mol equiv of NIQI were those belonging to the external G-tetrads, especially those of the 5′ end G-tetrad. This suggests the formation of some stacking interactions between these G-tetrads and the peptide, probably based on the large dipole moment associated with the amide group of the peptide and the prospect of dispersion involving amide π-electrons [53]. This in turn would allow the peptide to arrange in a conformation that leads the arginine residues to approach and interact with specific phosphate groups of the G4 through electrostatic interactions, while the serine residues may form H-bonds with them. Conversely, in the case of *Bcl-2* G4, all guanines, including those of the central G-tetrad, displayed similar changes in their imino proton signals upon the addition of NIQI, suggesting a possible peptide interaction in the grooves of the G4.

Overall, these results indicate a different binding mode of the peptide to the two G4s, in agreement with the differences observed by ITC. Indeed, the interaction of the peptide within the G4 grooves could result in the displacement of a larger number of water molecules compared to stacking on G-tetrads. This could explain the favorable entropic contribution observed in the case of NIQI binding to *Bcl-2*, while it appears unfavorable in the case of *c-Kit2*, which adopts a parallel topology where the grooves are occupied by loops. On the other hand, the formation of stacking interactions between NIQI and *c-Kit2* could explain the more favorable enthalpy changes observed by ITC in this case.

3.6. Alanine scanning to evaluate the role of individual amino acids in G4 binding

To further detail the interactions and determine the key amino acids involved in the binding of NIQI to *c-Kit2* and *Bcl-2* G4s, a mutagenesis analysis was performed. This analysis consisted in replacing one at a time the arginine and serine residues of the NIQI sequence with the non-bulky and chemically inert alanine residue. Glycine was not mutated since it has no sidechain and acts only as a flexible spacer. Therefore, a small library of eight peptides, each differing from NIQI by a single amino acid, was synthesized (Table 3) and their effects on *c-Kit2* and *Bcl-2* G4s were assessed by CD spectroscopy (Fig. S8–S11).

In particular, CD melting experiments allowed us to evaluate the individual contributions of specific arginine and serine residues to G4 binding and stabilization of NIQI. Interestingly, the results of these experiments (Figs. S9, S11, and Table 3) show that the arginine residues at positions 3 and 7 have limited involvement in the interaction with *c-Kit2* G4. Indeed, replacing them with alanine had minimal impact on the peptide's ability to stabilize this G4 (the difference in peptide-induced G4 stabilization for Ala^3 and Ala^7 derivatives compared to NIQI was

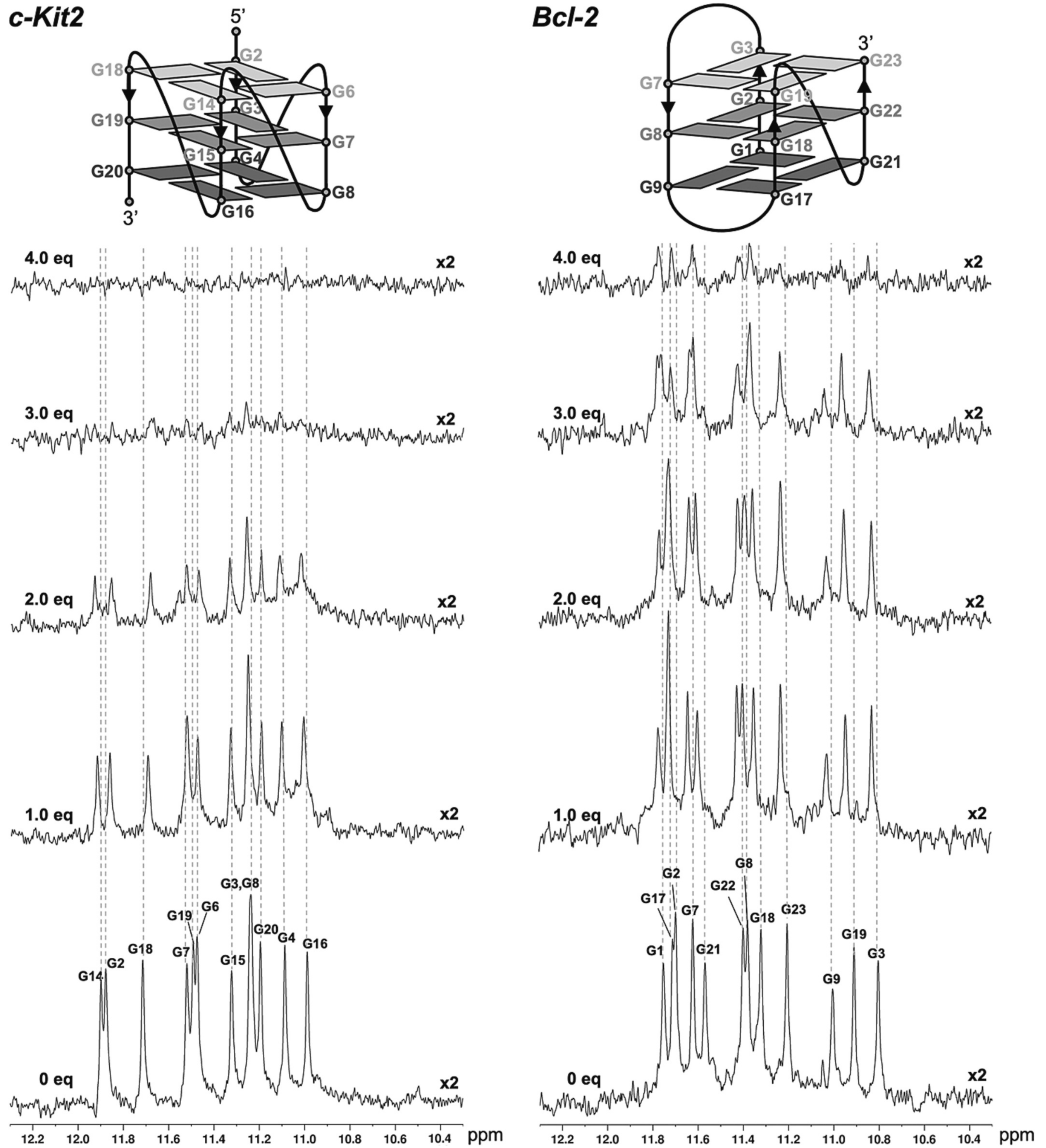


Fig. 4. Imino proton region of NMR spectra of *c-Kit2* and *Bcl-2* G4s at 25 °C titrated with NIQI peptide. The peptide equivalents are shown on the left of the spectra.

<2.5 °C). On the other hand, substituting the other arginine residues in the peptide sequence resulted in a significant decrease in the G4-stabilizing properties of the corresponding Ala-modified peptides (NIQI-Ala^1, NIQI-Ala^5, NIQI-Ala^{17}, and NIQI-Ala^{19}). These modified peptides induced a stabilization of *c-Kit2* G4 that was at least 5 °C lower than that induced by NIQI, indicating their greater involvement in the interaction with this G4. The same observation was made when substituting the serine residues (NIQI-Ala^{11} and NIQI-Ala^{14} peptides), suggesting an important role for them in the recognition and binding of the *c-Kit2* G4 structure.

As for *Bcl-2*, a significant decrease in the peptide-induced G4 stabilization was observed upon substitution of arginine 3 and 7 (Ala^3 and Ala^7 derivatives, respectively), as well as serine 14 (Ala^{14} derivative) (Fig. S11), suggesting a key role of these amino acids in the interaction with the *Bcl-2* G4. In contrast, negligible differences (around 1–1.5 °C) were observed when arginine 1 and 17 (Ala^1 and Ala^{17} derivatives,

Table 3
List of Ala derivatives of NIQI and their induced thermal stabilization (ΔT_m) for *c-Kit2* and *Bcl-2* G4 structures obtained by CD melting experiments.

Peptide	Sequence	ΔT_m (°C)[a]	
		c-Kit2	*Bcl-2*
NIQI	RGRGRGRGGGSGGSGGRGRG	12.6 (±0.2)	10.5 (±0.2)
NIQI-Ala^1	AGRGRGRGGGSGGSGGRGRG	6.2 (±0.3)	9.5 (±0.2)
NIQI-Ala^3	RGAGRGRGGGSGGSGGRGRG	10.1 (±0.3)	6.1 (±0.3)
NIQI-Ala^5	RGRGAGRGGGSGGSGGRGRG	7.6 (±0.4)	8.3 (±0.4)
NIQI-Ala^7	RGRGRGAGGGSGGSGGRGRG	10.3 (±0.3)	5.1 (±0.2)
NIQI-Ala^{11}	RGRGRGRGGGAGGSGGRGRG	6.9 (±0.3)	9.5 (±0.2)
NIQI-Ala^{14}	RGRGRGRGGGSGGAGGRGRG	7.9 (±0.5)	7.0 (±0.2)
NIQI-Ala^{17}	RGRGRGRGGGSGGSGGAGRG	6.9 (±0.3)	9.2 (±0.4)
NIQI-Ala^{19}	RGRGRGRGGGSGGSGGRGAG	6.0 (±0.4)	8.4 (±0.4)

[a] ΔT_m represents the difference in melting temperature [$\Delta T_m = T_m$ (DNA + 3 peptide equiv) - T_m (DNA)]. The T_m values of DNA alone are: *c-Kit2* = 62.1 ± 0.1 °C; *Bcl-2* = 60.5 ± 0.1 °C. All experiments were performed in 5 mM KH_2PO_4/K_2HPO_4 buffer (pH 7.0) containing 20 mM KCl. All experiments were performed in triplicate, and the reported values are the average of three measurements.

respectively) and serine 11 (Ala^{11}) were mutated, as well as only small variations (around 2–2.5 °C) were detected for the replacement of arginine 5 and 19 (Ala^5 and Ala^{19} derivatives). These results indicate either minimal to no involvement of these amino acids in the binding to *Bcl-2* G4 or that their substitution with alanine may be compensated by other amino acids present in the peptide sequence.

Overall, the alanine scanning analysis showed that the key amino acids of NIQI involved in the interaction with the two G4s are not identical. The involvement of fewer amino acids in the case of *Bcl-2* compared to *c-Kit2*, also agrees with ITC data, which indicate a lower binding enthalpy for the interaction with the former than the latter G4.

4. Conclusions

In this study, we analyzed the interaction between NIQI, a peptide containing an RGG motif shared by 77 human G4-binding proteins, and various biologically relevant G4 DNA structures with different topologies. CD experiments showed that NIQI induces some spectral changes in both parallel G4s and the hybrid G4 *Bcl-2*. These spectral variations are consistent with the peptide's ability to stabilize the G4 structures as well as with the affinity binding constants determined by MST experiments, which also indicated *c-Kit2* and *Bcl-2* as the best targets for this peptide.

Furthermore, the very weak interaction observed between NIQI and the duplex model ds_{26} confirmed that the binding of the RG-rich peptide to DNA is not only mediated by electrostatic interactions. It also suggested that there is a specific affinity for G4 structures, which is consistent with the fact that NIQI sequence originates from proteins that selectively bind to G4.

ITC and NMR experiments allowed us to elucidate the thermodynamics and gain insights into the binding mode of NIQI to the *c-Kit2* and *Bcl-2* G4 structures. The results of these experiments clearly indicate a different binding mode of the peptide to the two G4s. In particular, ^1H NMR spectra showed that NIQI affected the external G-tetrads of *c-Kit2* G4, particularly the 5′ end G-tetrad, thus suggesting the formation of stacking interactions, which would in turn allow the peptide to adopt a conformation enabling specific arginine residues to approach and interact with phosphate groups of the G4 through electrostatic interactions, while serine residues may form H-bonds with them. Remarkably, upon binding to *c-Kit2*, the peptide undergoes a transition from a completely random conformation to a structure containing 17.5 % helical content. On the other hand, in the case of *Bcl-2* G4, all G-tetrads were similarly affected by NIQI, suggesting a probable interaction of the peptide in the grooves of the G4 structure (Fig. 5). The NMR results are also corroborated by ITC data, which showed notable differences in the enthalpic and entropic contributions of NIQI binding to the two G4s, probably resulting from the distinct binding modes.

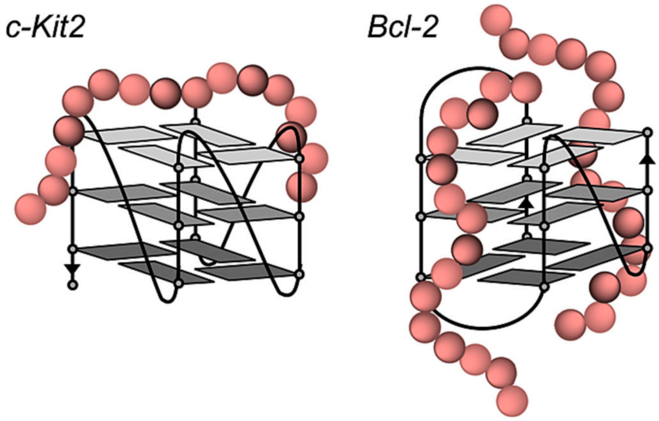


Fig. 5. Schematic representation of possible binding modes for the strongest interactions of the NIQI peptide (in red) to *c-Kit2* and *Bcl-2* G4s.

Finally, by adopting the alanine scanning approach, we systematically evaluated the impact of each arginine and serine residue on the binding to *c-Kit2* and *Bcl-2* G4s, thus identifying the key and less relevant amino acids in the interaction with these DNA molecules. The results obtained will undoubtedly help the development of other derivatives with the aim of further improving the peptide affinity and selectivity for G4 structures.

In conclusion, this study provides new insights into the interaction between peptides and G4 structures, thereby contributing to laying the groundwork for the design and use of a new class of peptide-based G4 ligands as an alternative to small molecules. Indeed, due to their larger size and bulkiness compared to small molecules, we strongly believe that peptides could serve as more effective tools to interfere in G4-protein interactions and potentially exert significant effects in modulating oncogene expression.

CRediT authorship contribution statement

Nicola Grasso: Methodology, Investigation, Formal analysis. **Raffaele Graziano:** Methodology, Investigation, Formal analysis. **Simona Marzano:** Investigation, Formal analysis. **Federica D'Aria:** Investigation, Formal analysis. **Francesco Merlino:** Validation, Formal analysis. **Paolo Grieco:** Resources, Supervision. **Antonio Randazzo:** Resources, Supervision. **Bruno Pagano:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Jussara Amato:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research leading to these results has received funding from AIRC (Italian Association for Cancer Research) under IG 2020 - ID. 24590 project - to B.P.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2023.126749.

References

[1] S.A. Lambert, A. Jolma, L.F. Campitelli, P.K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T.R. Hughes, M.T. Weirauch, The human transcription factors, Cell. 172 (2018) 650–665, https://doi.org/10.1016/j.cell.2018.01.029.

[2] C.W. Garvie, C. Wolberger, Recognition of specific DNA sequences, Mol. Cell 8 (2001) 937–946, https://doi.org/10.1016/S1097-2765(01)00392-6.

[3] F.A. Buske, J.S. Mattick, T.L. Bailey, Potential in vivo roles of nucleic acid triple-helices, RNA Biol. 8 (2011) 427–439, https://doi.org/10.4161/rna.8.3.14999.

[4] V. Brázda, R.C. Laister, E.B. Jagelská, C. Arrowsmith, Cruciform structures are a common DNA feature important for regulating biological processes, BMC Mol. Biol. 12 (2011) 33, https://doi.org/10.1186/1471-2199-12-33.

[5] V. Meier-Stephenson, G4-quadruplex-binding proteins: review and insights into selectivity, Biophys. Rev. 14 (2022) 635–654, https://doi.org/10.1007/s12551-022-00952-8.

[6] J. Amato, T.W. Madanayake, N. Iaccarino, E. Novellino, A. Randazzo, L.H. Hurley, B. Pagano, HMGB1 binds to the KRAS promoter G-quadruplex: a new player in oncogene transcriptional regulation? Chem. Commun. 54 (2018) 9442–9445, https://doi.org/10.1039/C8CC03614D.

[7] V. Brázda, M. Fojta, The rich world of p53 DNA binding targets: the role of DNA structure, Int. J. Mol. Sci. 20 (2019) 5605, https://doi.org/10.3390/ijms20225605.

[8] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, Quadruplex DNA: sequence, topology and structure, Nucleic Acids Res. 34 (2006) 5402–5415, https://doi.org/10.1093/nar/gkl655.

[9] D.J. Patel, A.T. Phan, V. Kuryavyi, Human telomere, oncogenic promoter and 5′-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics, Nucleic Acids Res. 35 (2007) 7429–7455, https://doi.org/10.1093/nar/gkm711.

[10] Y. Ma, K. Iida, K. Nagasawa, Topologies of G-quadruplex: biological functions and regulation by ligands, Biochem. Biophys. Res. Commun. 531 (2020) 3–17, https://doi.org/10.1016/j.bbrc.2019.12.103.

[11] J. Amato, L. Cerofolini, D. Brancaccio, S. Giuntini, N. Iaccarino, P. Zizza, S. Iachettini, A. Biroccio, E. Novellino, A. Rosato, M. Fragai, C. Luchinat, A. Randazzo, B. Pagano, Insights into telomeric G-quadruplex DNA recognition by HMGB1 protein, Nucleic Acids Res. 47 (2019) 9950–9966, https://doi.org/10.1093/nar/gkz727.

[12] S. Takahashi, J.A. Brazier, N. Sugimoto, Topological impact of noncanonical DNA structures on Klenow fragment of DNA polymerase, Proc. Natl. Acad. Sci. 114 (2017) 9605–9610, https://doi.org/10.1073/pnas.1704258114.

[13] Y. Xu, Chemistry in human telomere biology: structure, function and targeting of telomere DNA/RNA, Chem. Soc. Rev. 40 (2011) 2719–2740, https://doi.org/10.1039/c0cs00134a.

[14] A.K. Todd, M. Johnston, S. Neidle, Highly prevalent putative quadruplex sequence motifs in human DNA, Nucleic Acids Res. 33 (2005) 2901–2907, https://doi.org/10.1093/nar/gki553.

[15] J.L. Huppert, S. Balasubramanian, Prevalence of quadruplexes in the human genome, Nucleic Acids Res. 33 (2005) 2908–2916, https://doi.org/10.1093/nar/gki609.

[16] J.L. Huppert, S. Balasubramanian, G-quadruplexes in promoters throughout the human genome, Nucleic Acids Res. 35 (2007) 406–413, https://doi.org/10.1093/nar/gkl1057.

[17] V.S. Chambers, G. Marsico, J.M. Boutell, M. Di Antonio, G.P. Smith, S. Balasubramanian, High-throughput sequencing of DNA G-quadruplex structures in the human genome, Nat. Biotechnol. 33 (2015) 877–881, https://doi.org/10.1038/nbt.3295.

[18] R. Hänsel-Hertsch, D. Beraldi, S.V. Lensing, G. Marsico, K. Zyner, A. Parry, M. Di Antonio, J. Pike, H. Kimura, M. Narita, D. Tannahill, S. Balasubramanian, G-quadruplex structures mark human regulatory chromatin, Nat. Genet. 48 (2016) 1267–1272, https://doi.org/10.1038/ng.3662.

[19] D. Varshney, J. Spiegel, K. Zyner, D. Tannahill, S. Balasubramanian, The regulation and functions of DNA and RNA G-quadruplexes, Nat. Rev. Mol. Cell Biol. 21 (2020) 459–474, https://doi.org/10.1038/s41580-020-0236-x.

[20] O. Mendoza, A. Bourdoncle, J.-B. Boulé, R.M. Brosh, J.-L. Mergny, G-quadruplexes and helicases, Nucleic Acids Res. 44 (2016) 1989–2006, https://doi.org/10.1093/nar/gkw079.

[21] R. Linke, M. Limmer, S. Juranek, A. Heine, K. Paeschke, The relevance of G-quadruplexes for DNA repair, Int. J. Mol. Sci. 22 (2021) 12599, https://doi.org/10.3390/ijms222212599.

[22] M. Bartas, J. Červeň, S. Guziurová, K. Slychko, P. Pečinka, Amino acid composition in various types of nucleic acid-binding proteins, Int. J. Mol. Sci. 22 (2021) 922, https://doi.org/10.3390/ijms22020922.

[23] V. Brázda, J. Cerveň, M. Bartas, N. Mikysková, J. Coufal, P. Pečinka, J. Červeň, M. Bartas, N. Mikysková, J. Coufal, P. Pečinka, The amino acid composition of quadruplex binding proteins reveals a shared motif and predicts new potential quadruplex interactors, Molecules. 23 (2018) 2341, https://doi.org/10.3390/molecules23092341.

[24] R.M. Williams, Z. Obradovic, V. Mathura, W. Braun, E.C. Garner, J. Young, S. Takayama, C.J. Brown, A.K. Dunker, The protein non-folding problem: amino acid determinants of intrinsic order and disorder, in: Biocomputing, World Scientific, 2001, pp. 89–100, https://doi.org/10.1142/9789814447362_0010.

[25] Z.L. Huang, J. Dai, W.H. Luo, X.G. Wang, J.H. Tan, S. Bin Chen, Z.S. Huang, Identification of G-quadruplex-binding protein from the exploration of RGG motif/G-quadruplex interactions, J. Am. Chem. Soc. 140 (2018) 17945–17955, https://doi.org/10.1021/jacs.8b09329.

[26] K. Takahama, T. Oyoshi, Specific binding of modified RGG domain in TLS/FUS to G-quadruplex RNA: tyrosines in RGG domain recognize 2′-OH of the riboses of loops in G-quadruplex, J. Am. Chem. Soc. 135 (2013) 18016–18019, https://doi.org/10.1021/ja4086929.

[27] N. Vasilyev, A. Polonskaia, J.C. Darnell, R.B. Darnell, D.J. Patel, A. Serganov, Crystal structure reveals specific recognition of a G-quadruplex RNA by a β-turn in the RGG motif of FMRP, Proc. Natl. Acad. Sci. 112 (2015), https://doi.org/10.1073/pnas.1515737112.

[28] S. Pelliccia, J. Amato, D. Capasso, S. Di Gaetano, A. Massarotti, M. Piccolo, C. Irace, G.C.G.C. Tron, B. Pagano, A. Randazzo, E. Novellino, M. Giustiniano, Bio-inspired dual-selective BCL-2/c-MYC G-quadruplex binders: design, synthesis, and anticancer activity of drug-like imidazo[2,1-i]purine derivatives, J. Med. Chem. 63 (2020) 2035–2050, https://doi.org/10.1021/acs.jmedchem.9b00262.

[29] J. Amato, A. Pagano, D. Capasso, S. Di Gaetano, M. Giustiniano, E. Novellino, A. Randazzo, B. Pagano, Targeting the BCL2 gene promoter G-quadruplex with a new class of furopyridazinone-based molecules, ChemMedChem. 13 (2018) 406–410, https://doi.org/10.1002/cmdc.201700749.

[30] N. Kosiol, S. Juranek, P. Brossart, A. Heine, K. Paeschke, G-quadruplexes: a promising target for cancer therapy, Mol. Cancer 20 (2021) 40, https://doi.org/10.1186/s12943-021-01328-4.

[31] S. Neidle, Quadruplex nucleic acids as targets for anticancer therapeutics, Nat. Rev. Chem. 1 (2017) 0041, https://doi.org/10.1038/s41570-017-0041.

[32] D. Rhodes, H.J. Lipps, G-quadruplexes and their regulatory roles in biology, Nucleic Acids Res. 43 (2015) 8627–8637, https://doi.org/10.1093/nar/gkv862.

[33] F. Merlino, S. Tomassi, A.M. Yousif, A. Messere, L. Marinelli, P. Grieco, E. Novellino, S. Cosconati, S. Di Maro, Boosting Fmoc solid-phase peptide synthesis by ultrasonication, Org. Lett. 21 (2019) 6378–6382, https://doi.org/10.1021/acs.orglett.9b02283.

[34] H. Adihou, R. Gopalakrishnan, T. Förster, S.M. Guéret, R. Gasper, S. Geschwindner, C. Carrillo García, H. Karatas, A.V. Pobbati, M. Vazquez-Chantada, P. Davey, C. M. Wassvik, J.K.S. Pang, B.S. Soh, W. Hong, E. Chiarparin, D. Schade, A. T. Plowright, E. Valeur, M. Lemurell, T.N. Grossmann, H. Waldmann, A protein tertiary structure mimetic modulator of the Hippo signalling pathway, Nat. Commun. 11 (2020) 5425, https://doi.org/10.1038/s41467-020-19224-8.

[35] H. Cantor, R. C., M.M. Warshaw, C.R. Cantor Shapiro, M.M. Warshaw, H. Shapiro, Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleolides, Biopolymers. 9 (1970) 1059–1077, https://doi.org/10.1002/bip.1970.360090909.

[36] A. Micsonai, F. Wien, L. Kernya, Y.-H. Lee, Y. Goto, M. Réfrégiers, J. Kardos, Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) E3095–E3103, https://doi.org/10.1073/pnas.1500851112.

[37] T.L. Hwang, A.J. Shaka, Water Suppression That Works, Excitation sculpting using arbitrary wave-forms and pulsed-field gradients, J. Magn. Reson. Ser. A 112 (1995) 275–279, https://doi.org/10.1006/jmra.1995.1047.

[38] G. Holzwarth, P. Doty, The ultraviolet circular dichroism of polypeptides, J. Am. Chem. Soc. 87 (1965) 218–228, https://doi.org/10.1021/ja01080a015.

[39] F. Abedin, N. Kandel, S.A. Tatulian, Effects of Aβ-derived peptide fragments on fibrillogenesis of Aβ, Sci. Rep. 11 (2021) 19262, https://doi.org/10.1038/s41598-021-98644-y.

[40] J.S. Richardson, D.C. Richardson, Principles and patterns of protein conformation, in: Predict. Protein Struct. Princ. Protein Conform., Springer US, Boston, MA, 1989: pp. 1–98. doi:https://doi.org/10.1007/978-1-4613-1571-1_1.

[41] E.F. McDonald, T. Jones, L. Plate, J. Meiler, A. Gulsevin, Benchmarking AlphaFold2 on peptide structure prediction, Structure. 31 (2023) 111–119.e2, https://doi.org/10.1016/j.str.2022.11.012.

[42] J. Rey, S. Murail, S. de Vries, P. Derreumaux, P. Tuffery, PEP-FOLD4: a pH-dependent force field for peptide structure prediction in aqueous solution, Nucleic Acids Res. 51 (2023) W432–W437, https://doi.org/10.1093/nar/gkad376.

[43] R. del Villar-Guerra, J.O. Trent, J.B. Chaires, G-Quadruplex secondary structure obtained from circular dichroism spectroscopy, Angew. Chemie Int. Ed. 57 (2018) 7171–7175, https://doi.org/10.1002/anie.201709184.

[44] A. Randazzo, G.P. Spada, M. Webba da Silva, Circular dichroism of quadruplex structures, Top. Curr. Chem. 330 (2013) 67–86, https://doi.org/10.1007/128-2012-331.

[45] M. Jerabek-Willemsen, C.J. Wienken, D. Braun, P. Baaske, S. Duhr, Molecular interaction studies using microscale thermophoresis, Assay Drug Dev. Technol. 9 (2011) 342–353, https://doi.org/10.1089/adt.2011.0380.

[46] B. Pagano, C.A. Mattia, C. Giancola, Applications of isothermal titration calorimetry in biophysical studies of G-quadruplexes, Int. J. Mol. Sci. 10 (2009) 2935–2957, https://doi.org/10.3390/ijms10072935.

[47] M. Campanile, R. Oliva, P. Del Vecchio, R. Winter, L. Petraccone, The anticancer peptide LL-III binds with nanomolar affinity to human telomeric and cMyc G-quadruplexes, Chem. Commun. 59 (2023) 6179–6182, https://doi.org/10.1039/D3CC00737E.

[48] V.H. Le, R. Buscaglia, J.B. Chaires, E.A. Lewis, Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model, Anal. Biochem. 434 (2013) 233–241, https://doi.org/10.1016/j.ab.2012.11.030.

[49] V. Pirota, C. Platella, D. Musumeci, A. Benassi, J. Amato, B. Pagano, G. Colombo, M. Freccero, F. Doria, D. Montesarchio, On the binding of naphthalene diimides to a human telomeric G-quadruplex multimer model, Int. J. Biol. Macromol. 166 (2021) 1320–1334, https://doi.org/10.1016/j.ijbiomac.2020.11.013.

[50] A. Peduto, B. Pagano, C. Petronzi, A. Massa, V. Esposito, A. Virgilio, F. Paduano, F. Trapasso, F. Fiorito, S. Florio, C. Giancola, A. Galeone, R. Filosa, Design, synthesis, biophysical and biological studies of trisubstituted naphthalimides as G-quadruplex ligands, bioorganic, Med. Chem. 19 (2011) 6419–6429, https://doi.org/10.1016/j.bmc.2011.08.062.

[51] V. Kuryavyi, A.T. Phan, D.J. Patel, Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter, Nucleic Acids Res. 38 (2010) 6757–6773, https://doi.org/10.1093/nar/gkq558.

[52] J. Dai, D. Chen, R.A. Jones, L.H. Hurley, D. Yang, NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region, Nucleic Acids Res. 34 (2006) 5133–5144, https://doi.org/10.1093/nar/gkl610.

[53] W.H. James III, C.W. Müller, E.G. Buchanan, M.G.D. Nix, L. Guo, L. Roskop, M. S. Gordon, L.V. Slipchenko, S.H. Gellman, T.S. Zwier, Intramolecular amide stacking and its competition with hydrogen bonding in a small foldamer, J. Am. Chem. Soc. 131 (2009) 14243–14245, https://doi.org/10.1021/ja9054965.