UNIVERSITY OF NAPLES FEDERICO II

DEPARTMENT OF PHARMACY

PhD in Pharmaceutical Sciences — XXXIV cycle

Targeting of noncanonical nucleic acid structures for therapeutic intervention and biological investigation

PhD Thesis 2022

Anna Di Porzio

Tutor:
Prof. ANTONIO RANDAZZO

PhD Coordinator:
Prof. MARIA VALERIA D'AURIA

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Antonio Randazzo. He has been an exceptional motivator, providing me with valuable guidance and support for the entire duration of my PhD, encouraging me anytime, pushing me to sharpen my thinking and bringing my work to a higher level. His impressive determination and tenacity have greatly inspired me, making him the best supervisor I could have had. I truly hope he is proud of me.

I am also very thankful to Dr. Nunzia Iaccarino for her lovely assistance at every single stage of my PhD. She has given me about all the time she could, providing me with insightful comments and suggestions. She was there whenever I needed it, with her kindness and radiance. I will always be grateful to her for "driving" me to the end of this amazing journey.

Huge thanks go to Prof. Bruno Pagano and Prof. Jussara Amato for their unwavering scientific support. I value their dedication to work and attention to details, which have been really stimulating for me.

Warm thanks go to my colleagues Federica and Simona, for being incredibly nice and kind to me all the time, and to my friend Tommaso, for being my long-distance supporter during the pandemic, for sharing thoughts and feelings with me, and for his careful listening.

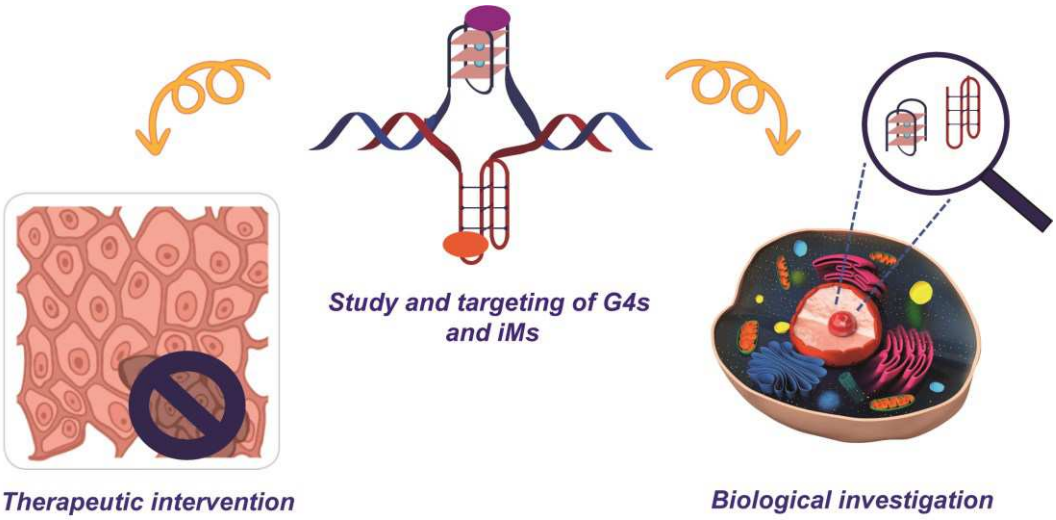
Furthermore, I would like to express my special appreciation and thanks to Dr. Marco Di Antonio, who has been my supervisor during the ten months I spent at Imperial College London. He has been an amazing co-tutor, offering me priceless personal and professional support. His enthusiasm and love for science have been contagious. I would choose him a thousand times as co-supervisor!

My heartfelt thanks also go to Denise Liano. She has been the first person I talked to as I joined Di Antonio's group. Her smile and sweetness immediately made me feel at home. We have been emotionally close to each other, despite the social distancing. I can definitely say that I have found a great friend other than a brilliant colleague!

Finally, and most importantly, I am indebted to my family for all the support they have shown me, since the very first day I chose such a career. I have always had my family to count on when times have been rough, and that was priceless. My hard-working parents have sacrificed their lives for me and my brother and provided unconditional love and care. I love them so much, and I would not have made it this far without them. I thank my beloved grandmothers, Anna, I wish she had lived for a few more years to let me see the pride in her eyes, and Nietta, who has never stopped believing in me. Last, but not least, I truly thank Salvatore, my boyfriend, for sticking by my side both during my good and bad times. He has been a true and extraordinary supporter during the past year, as well as the sweetest surprise life could have made to me.

Anna Di Porzio
Naples, January 2022

ABSTRACT



G-quadruplex (G4) structures fall among the most extensively studied noncanonical DNA/RNA conformations, with the current knowledge suggesting that they might play an integral part in many biological processes, including transcriptional regulation, telomeric maintenance, malignant transformation, and cancer development. In the last decade, mounting attention has been also focused on i-motif (iM) structures. Even though the structural aspects of iM formation, the factors leading to its stabilization, and its putative biological roles have been recently reviewed and discussed by the scientific community, there is still much to know about such noncanonical nucleic acid conformation.

This PhD project aimed, on the one hand, at providing a more in-depth knowledge about the i-motif structures and, on the other hand, at identifying and developing new molecular tools able to target alternative nucleic acid secondary structures for either therapeutic intervention or biological investigation.

Particularly, the thesis has been subdivided as follow:

Chapter 1 provides a general description of nucleic acids and noncanonical DNA/RNA conformations, stressing on their structural features and biological roles.

Chapter 2 offers an overview of the main methodologies employed in the herein presented studies.

Chapter 3 includes two detailed investigations on iM structures. The first one is a systematic study to simultaneously analyze the effect of pH, cation type, and cation concentration (and their possible interactions) on the formation of an i-motif structure *in vitro*. The second one uses multivariate data analysis to bring out valuable structural information on i-motif DNA from circular dichroism and thermal difference spectra.

Chapter 4 describes new G4-targeting compounds that we identified or rationally designed and synthesized as potential anti-cancer agents.

In *Chapter 5*, new molecules are reported that might be used as tools to clarify the biological roles of G4 and iM structures and their intriguing relationship.

The last section of this PhD thesis collects the general conclusions.

Finally, the *Appendix* reports the published scientific articles as cited throughout the thesis. Paper IV is not included since the Manuscript is still in preparation and has not been submitted at the present time.

LIST OF PUBLICATIONS

Paper I

"Assessing the influence of pH and cationic strength on i-motif DNA structure". Iaccarino N., Di Porzio A., Amato J., Pagano B., Brancaccio D., Novellino E., Leardi R., Randazzo A. *Analytical & Bioanalytical Chemistry*, 2019, *411*(28), 7473–7479.

Paper II

"Effects of sequence and base composition on the CD and TDS profiles of i-DNA". Iaccarino N., Cheng M., Qiu D., Pagano B., Amato J., Di Porzio A., Zhou J., Randazzo A., Mergny JL. *Angewandte Chemie,* 2021, *60*(18), 10295–10303.

Paper III

"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.

Paper IV

"Design, synthesis, biophysical and biological characterization of carbonic anhydrase inhibitors-Berberine hybrids as new multi-target anti-cancer drugs". *Manuscript in preparation*.

Paper V

"A short peptide that preferentially binds c-MYC G4 DNA". Minard A., Morgan D., Raguseo F., Di Porzio A., Liano D., Jamieson A., Di Antonio M. *Chemical Communications*, 2020, *56*(63), 8940–8943.

Paper VI

"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.

Other publications:

1. "Impact of phytosterols on liver and distal colon metabolome in experimental murine colitis model: an explorative study". Iaccarino N., Amato J., Pagano B., Di Porzio A., Micucci M., Bolelli L., Aldini R., Novellino E., Budriesi R., Randazzo A. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 2019, *34*(1), 1041–1050.
2. "Screening of G-quadruplex stabilizing ligands by nano differential scanning fluorimetry". Pagano B., Iaccarino N., Di Porzio A., Randazzo A., Amato J. *Analyst*, 2019, *144*(22), 6512–6516. Awarded with Front Cover of the Journal.
3. "Common G-quadruplex binding agents found to interact with i-motif-forming DNA: unexpected multi-target-directed compounds". Pagano A., Iaccarino N., Abdelhamid M.A.S., Brancaccio D., Garzarella E.U., Di Porzio A., Novellino E., Waller Z.A.E., Pagano B., Amato J., Randazzo A. *Frontiers in Chemistry*, 2018, *6*, 281.

LIST OF ABBREVIATIONS

A	Adenine
ALT	Alternative lengthening of telomeres
ANOVA	One-way analysis of variance
ATP	Adenosine 5'-triphosphate
Bcl-2	B-cell lymphoma 2
C	Cytosine
CA	Carbonic anhydrase
CAI	Carbonic anhydrase inhibitor
CD	Circular dichroism
ChIP-seq	Chromatin immunoprecipitation followed by sequencing
c-Myc	Cellular Myc
CuAAC	Copper(I)-catalyzed alkyne-azide cycloaddition
DAPI	4′,6-Diamidino-2-phenylindole
DBU	1,8-Diazabicycloundec-7-ene
DDR	DNA damage response
dG4s	DNA G-quadruplex
DMF	N, N′-dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DOE	Design of experiments
DPPA	Diphenylphosphoryl azide

EB	Ethidium bromide
FAM	Fluorescein amidite
FFD	Full factorial design
FID	Fluorescent intercalator displacement
FISH	Fluorescence in situ hybridization
FP	Fluorescence polarization
FRET	Förster resonance energy transfer
FT-IR	Fourier transform infrared
G	Guanine
G4	G-quadruplex
HIF-1/2	Hypoxia-inducible factor 1 and 2
HPLC	High-performance liquid chromatography
IF	Immunofluorescence
iM	i-Motif
L-CPL	Left-handed circularly polarized light
LncRNA	Long non-coding RNA
MLR	Multiple linear regression
MS	Mass spectrometry
MTT	3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
OVAT	One variable at a time
OXPHOS	Oxidative phosphorylation
PC	Principal component
PCA	Principal component analysis

PEM	Photoelastic modulator
PQS	Putative G-quadruplex-forming sequence
R-CPL	Right-handed circularly polarized light
rG4	RNA G-quadruplex
RNA	Ribonucleic acid
RT-qPCR	Real-time quantitative polymerase chain reaction
SD	Standard deviation
SEM	Standard error of the mean
SPPS	Solid-phase peptide synthesis
STD	Saturation transfer difference
T	Thymine
TCA	Tricarboxylic acid
TDS	Thermal difference spectra
TERRA	Telomeric repeat-containing RNA
THF	Tetrahydrofuran
TO	Thiazole orange
TSS	Transcription start site
U	Uracil
UTRs	Untranslated regions
UV	Ultraviolet
ZBG	Zinc-binding group

TABLE OF CONTENTS

Chapter 1
INTRODUCTION

1.1 **General description of nucleic acids**

Nucleic acids exist in two naturally occurring varieties: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The former represents the main information-carrying molecule within the cell, providing instructions for the proper functioning of the entire organism, whereas the latter plays critical roles in protein synthesis (messenger RNA, transfer RNA, ribosomal RNA) or gene expression regulation (micro RNAs). From a chemical point of view, nucleic acids can be imagined as necklaces strung with beads (the nucleotides) each of which is, in turn, made up of three components: a nitrogenous base, a five-carbon sugar, and at least one phosphate group.

Nitrogenous bases

The nitrogenous bases of nucleotides contain either two fused carbon-nitrogen rings, in the case of adenine (A) and guanine (G), or a single six-membered nitrogen-containing ring, in the case of cytosine (C), thymine (T) and uracil (U). A and G are classified as purines, while C, T and U are typically referred to as pyrimidines. Each base is characterized by a unique chemical structure, with its own set of functional groups (Figure 1.1). DNA contains A, T, C, and G, while RNA includes A, U, C, and G (that is, T is replaced by U).

Sugars

In addition to bearing slightly different nitrogenous bases, DNA and RNA also differ in the type of five-carbon sugar they contain: deoxyribose for the DNA and ribose for the RNA. The carbon atoms of a nucleotide's sugar moiety are numbered as 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 (or groups) (Figure 1.2).

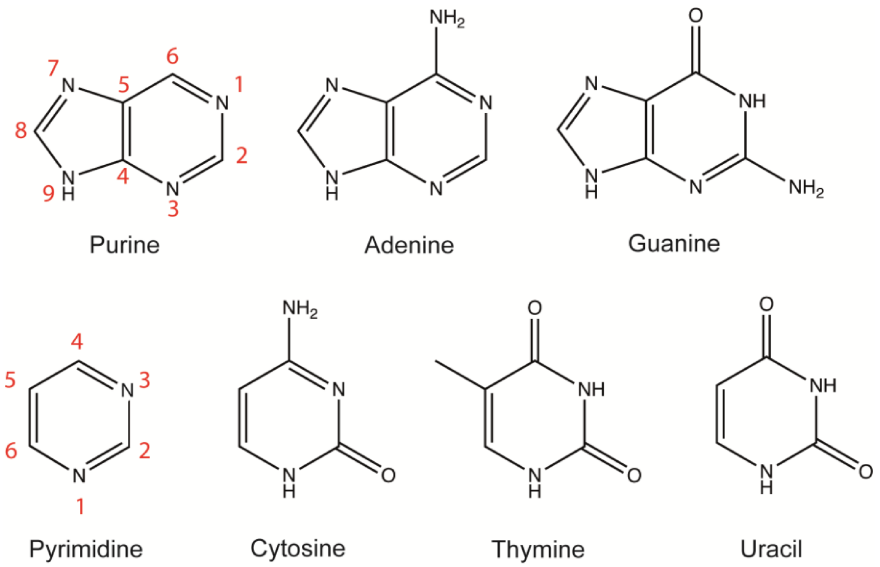


Figure 1.1. Structure of purine and pyrimidine bases.

Phosphate group(s)

A nucleotide which is still about to be added to the end of a polynucleotide chain, will bear three phosphate groups. Two out of three phosphate groups will be lost, instead, when the nucleotide joins a growing DNA or RNA chain.

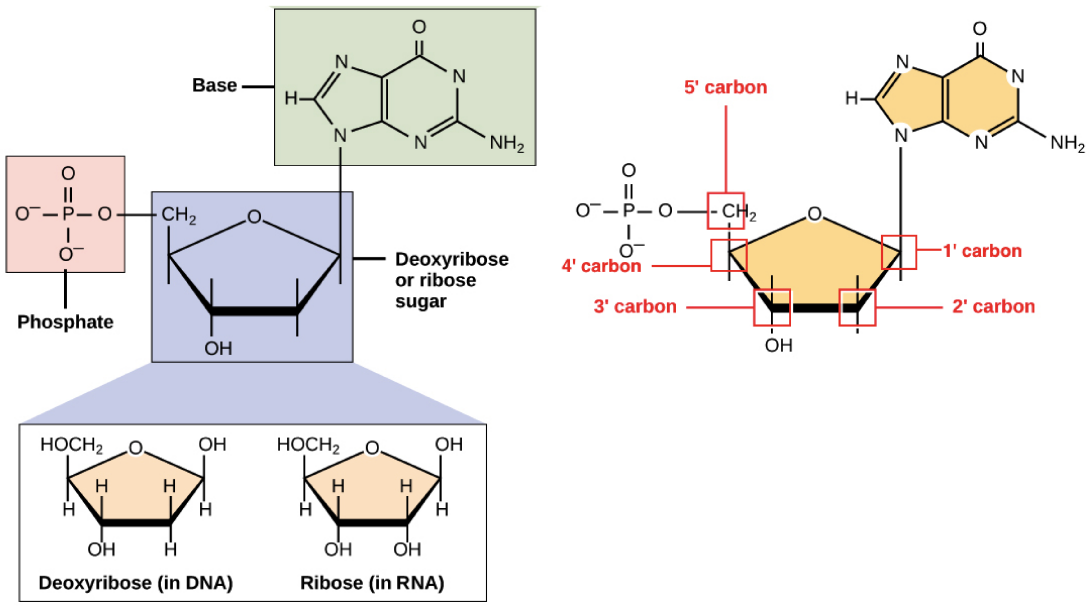


Figure 1.2. Nucleotide composition.

An important feature of the three-dimensional structure of nucleotides is the relative orientation of the base and sugar moieties. In the so-called “anti” conformation, the bulk of the base and the 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 to the N1-C2 bond of pyrimidines. By contrast, the “syn” conformation occurs when the C1′-O4′ bond of deoxyribose or ribose is *cis* to the N9-C4 purine bond or to the N1-C2 bond of pyrimidines (Figure 1.3).

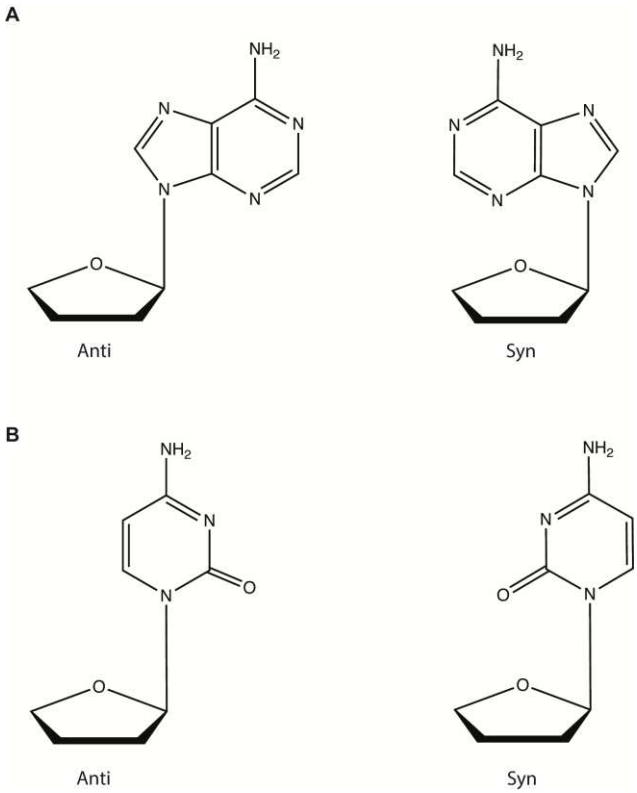


Figure 1.3. Anti and syn conformations of a nucleotide containing (**A**) a purine or (**B**) a pyrimidine base.

In a DNA/RNA strand, the 3′-hydroxyl group of a nucleotide and the 5′-phosphate group of its neighbor are covalently connected to each other through a phosphodiester bond. This gives the chain a chemical polarity, meaning that its two ends are chemically different: the 5′-end carries a free phosphate group, while the 3′-end bears an unlinked -OH group (Figure 1.4). RNA is usually single-stranded, whereas a DNA molecule, according to the James Watson and Francis Crick’s model (WATSON & CRICK, 1953), consists of two antiparallel strands (running in opposite directions) held together in a right-handed helix by hydrogen bonding between the bases. Particularly, A always pairs with T (by making two hydrogen bonds), and

G with C, establishing three hydrogen bonds (Figure 1.5), so that the sequence of one strand defines the composition of the other. Such complementary base-pairing allows the bases to be packed in the energetically most favorable arrangement within the double helix, while the hydrophilic sugar-phosphate backbone lays on the outside of the helix itself. Also, the reversible nature of the inter-strand bonds makes the information stored in the DNA readily accessible for replication or gene expression.

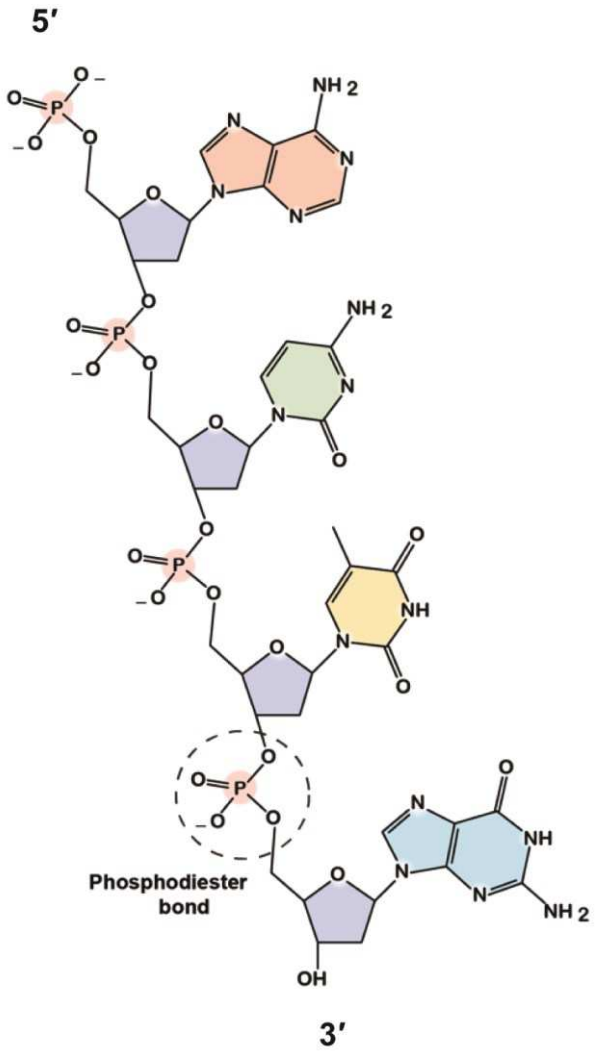


Figure 1.4. Schematic representation of the phosphodiester bond between nucleotides in a DNA strand.

The one described by Watson and Crick in 1953 is the most common DNA conformation in living cells, namely B-DNA, which is also characterized by a wide major groove and a narrow minor groove that run along the entire length of the molecule, being binding sites for several proteins (Figure 1.5).

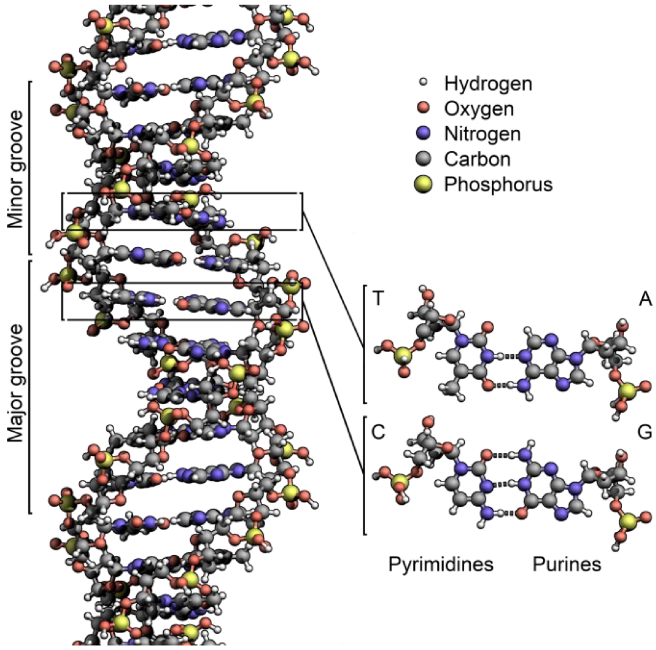


Figure 1.5. The three-dimensional double helix structure of DNA (B-DNA), correctly elucidated by James Watson and Francis Crick. Complementary bases are held together by hydrogen bonds.

1.2 **Noncanonical nucleic acid structures**

Under certain conditions and depending on the specific nucleotide sequence, both DNA and RNA can fold into a variety of noncanonical inter- or intramolecular structures, such as G-quadruplexes (G4s) and i-motifs (iMs) (Figure 1.6).

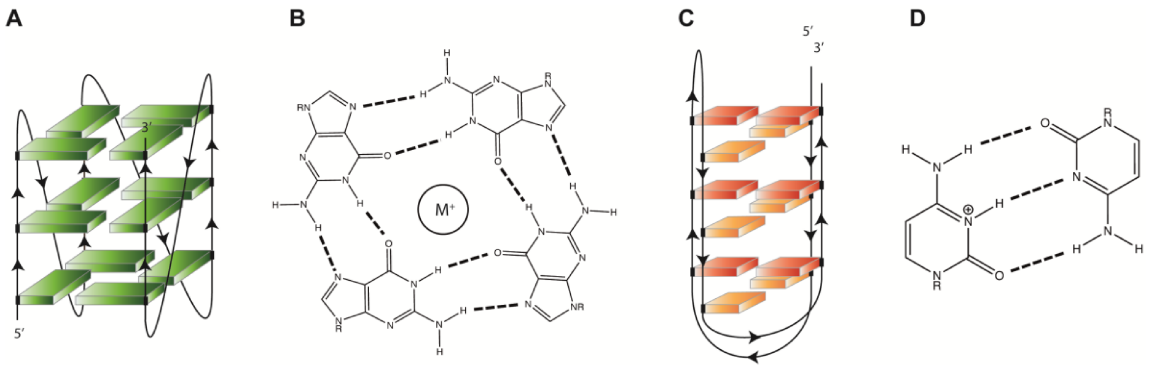


Figure 1.6. (**A**) An example of G4 structure. (**B**) The G-quartet, where M^+ is a monovalent cation. (**C**) An example of iM structure. (**D**) The cytosine-cytosine$^+$ base pair.

1.2.1 *G-quadruplexes*

G-quadruplex structures fall among the most intriguing and broadly studied alternative nucleic acid secondary structures (Spiegel et al., 2020). They are four-stranded helical complexes that can arise from self-stacking of two or more G-tetrads (Figure 1.6), cyclic arrays of four Hoogsteen hydrogen-bonded guanine residues (Huppert, 2010). The great majority of the oligonucleotides that fold into G-quadruplexes include four contiguous guanine tracts (each tract contributes one G per G-tetrad) linked by three runs of bases of undefined composition (the loops). The most thermodynamically stable G-quadruplexes generate from (at least) three stacked G-tetrads and loops containing from 1 to 7 bases, as demonstrated by detailed studies (Bugaut & Balasubramanian, 2008; Guédin et al., 2010; Hazel et al., 2004). Therefore, it has been proposed that motifs of the type $G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}$ (where N is any nucleotide), will fold into G-quadruplex structures (Todd, 2005). However, current experimental data suggests that G4s can also form in regions that do not follow such a stringent motif (Chambers et al., 2015; Marsico et al., 2019).

In a G-tetrad (also called G-quartet), the O6 atoms of the guanines face each other to form a central electron-rich channel where cations can place themselves, contributing to the thermal stabilization of a G4 structure more than intra-quartet hydrogen bonding or inter-quartet stacking (Gu et al., 1999). Particularly, K^+ ions (ionic radius 1.33 Å), being too large, are sandwiched between G-tetrads, whereas Na^+ ions (ionic radius 0.95 Å) are small enough to be coordinated within the plane of a G-quartet (Bhattacharyya et al., 2016; Williamson et al., 1989) (Figure 1.7). Although different monovalent cations can be encapsulated by G-quadruplexes ($K^+ > Na^+ > Li^+$), only sodium and potassium ions are physiologically relevant.

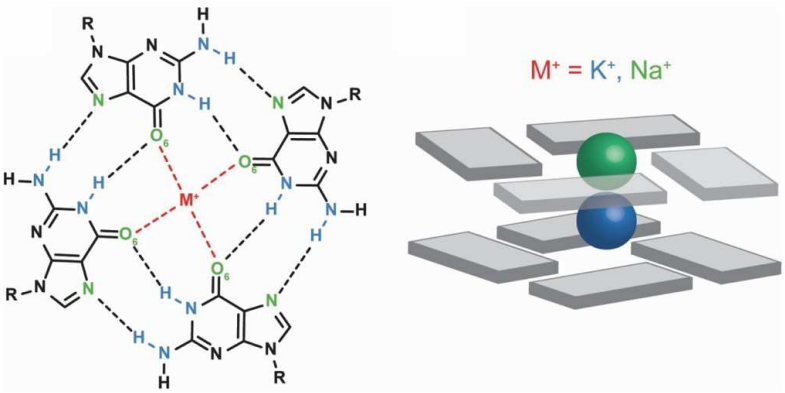


Figure 1.7. Coordination of cations within a G-quadruplex structure. Figure adapted from (Largy et al., 2016).

As already mentioned, the nucleotide sequences connecting the G-quartets form the so-called "loops". Noteworthy, loops can be of different geometries (Bugaut & Balasubramanian, 2008) (Figure 1.8):

- *Double chain reversal* (or *propeller*) *loops*, connecting two adjacent strands with the same orientation.
- *Lateral* (or e*dgewise*) *loops*, linking two adjacent strands with opposite orientations.
- *Diagonal loops*, connecting two opposing antiparallel strands across the G-quartet.

The length and composition of loops, along with strands orientation, can notably vary, making G4s extremely polymorphic structures and allowing them to adopt diverse topologies (Bugaut & Balasubramanian, 2008; Guédin et al., 2010) (Figure 1.8):

- *Parallel*, with the four strands sharing the same orientation.
- *Hybrid*, with three strands sharing an identical orientation and the fourth one oriented in the opposite direction.
- *Antiparallel*, in which two strands are oriented in one direction and the other two in the opposite one. Also, the strands sharing the same direction can be diagonally opposed or adjacent.

The previously discussed chemical differences between DNA and RNA also reflect on the properties of the corresponding G4 structures: the extra 2′-OH in RNA allows for additional hydrogen bonding and intramolecular interactions, thus making RNA G4s (rG4s) more stable than the DNA counterpart (Agarwal et al., 2012; Lightfoot et al., 2019; Malgowska et al., 2016). Also, as opposed to DNA G4s (dG4s) that can explore a broader conformational space, the 2′-OH moiety in rG4s is responsible for steric constraints that predominantly force them into the parallel topology (Banco & Ferré-D'Amaré, 2021; Tang & Shafer, 2006).
Finally, unlike the canonical double helix (which exhibits two grooves), all G-quadruplex structures display four grooves, whose properties and dimensions vary with respect to the topology and nature of loops: grooves in G4s with only lateral or diagonal loops are structurally simple. By contrast, if propeller loops are incorporated into the G4 structure, grooves show more complex structural features (Burge et al., 2006).

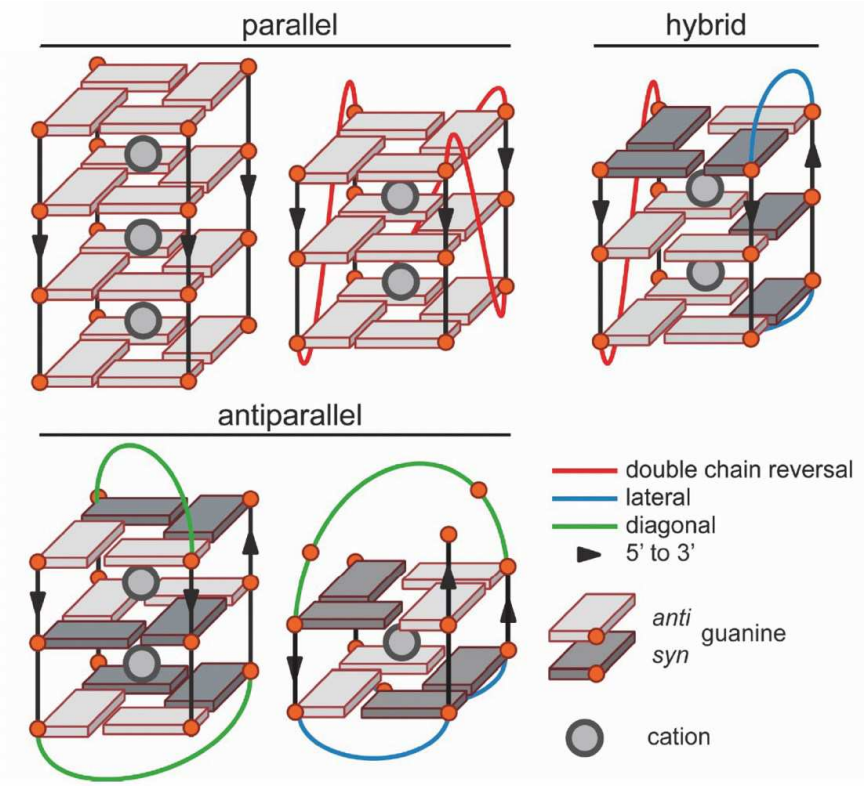


Figure 1.8. Schematic representation of different G4 topologies (parallel, hybrid, and antiparallel), differing one from the other in terms of relative orientation of the strands, number of tetrads, and geometry of the loops. Figure adapted from (Largy et al., 2016).

Beyond forming promptly *in vitro*, G-quadruplexes have been visualized in the genomic DNA of human cells by means of a structure-specific antibody (BG4)(Biffi et al., 2013) and mounting experimental data supports their relevance in the cellular context (Bochman et al., 2012; Varshney et al., 2020). Interestingly, it has been shown that DNA sequences able to fold into G-quadruplexes mainly exist at functional genomic loci, such as telomeres or gene promoters (Marsico et al., 2019). The specific location of dG4 motifs, along with the fact that they are highly conserved between different species (Capra et al., 2010; Marsico et al., 2019), suggest a selection pressure to retain G-quadruplex structures, which have been therefore speculated to govern fundamental biological processes such as telomere homeostasis or transcription (Kim, 2019; Rhodes & Lipps, 2015; Spiegel et al., 2020). Telomeres are repeats of the d(TTAGGG) sequence, located at the ends of each eukaryotic chromosome, and implicated in maintaining genome integrity (Figure 1.9). Being single-stranded DNA really delicate within the cellular environment, the short single-stranded region (around 50–200 bases) at the 3′-terminal of the human telomere is surrounded by a protective protein complex termed shelterin (De Lange, 2005). Such set of proteins also promotes T-loop formation (Griffith et al., 1999), ensuring

further protection to the 3′-overhang. Telomeres are critical for cell survival: in normal cells, they get shortened after each round of replication and, once the Hayflick limit is reached (Shay & Wright, 2000), cells cease to proliferate (senescence) or undergo apoptosis. Unlike what has been just said, around 85% of cancer cells gain immortality by overexpressing telomerase, a reverse-transcriptase enzyme that elongates telomeres (Greider & Blackburn, 1985), being an attractive target in designing anti-cancer agents. Particularly, since telomerase needs pairing with the single-stranded $d(TTAGGG)_n$ 3′-overhang to be functional, compounds able to stabilize G-quadruplexes formed within telomeres have been speculated to impair telomerase binding and interfere with its activity, leading to therapeutic advantages (Jean-Louis Mergny & Hélène, 1998).

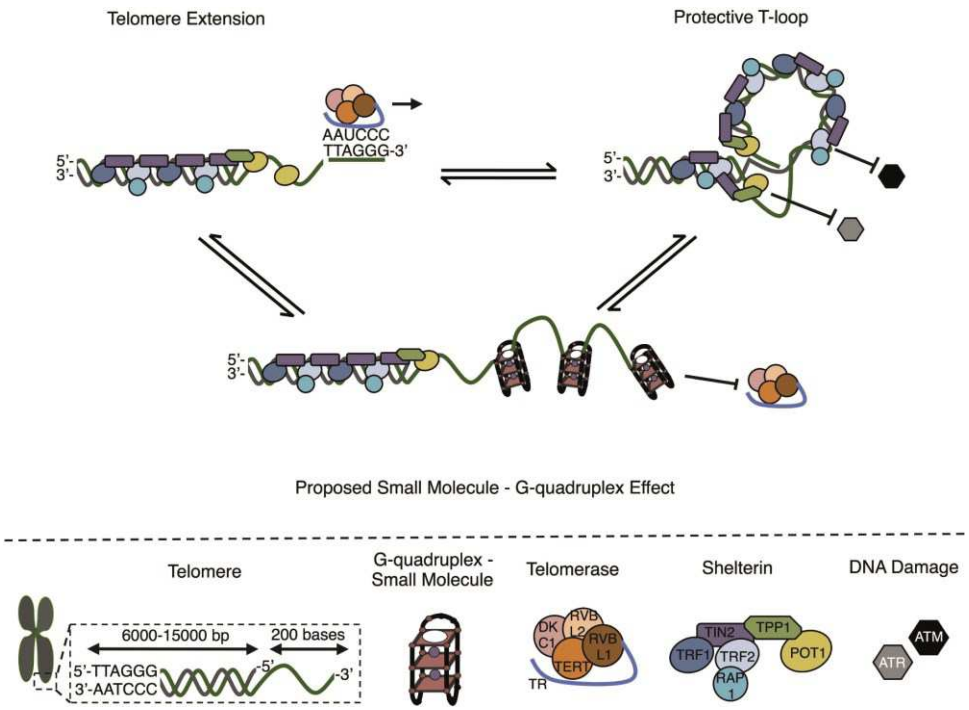


Figure 1.9. Representation of human telomeres, the proteins that bind to them (shelterin and telomerase), the protective T-loop they can adopt, and the potential role that G4 stabilizing small molecules can have on this equilibrium.

Besides telomeres, G4 motifs are also enriched at promoters, genomic elements that govern the first level of gene expression, i.e., transcription, containing binding sites for several regulatory proteins (transcription factors). Particularly, putative G-quadruplex-forming sequences (PQS) occur in the promoter region of oncogenes, normal genes that lead to the initiation and/or progression of cancer when aberrantly activated due to mutations or other genomic events. In

2002, Hurley and coworkers described the G-quadruplex structure in the regulatory region of the oncogene MYC, also showing MYC-downregulation as a direct consequence of Myc G4 stabilization by the TMPyP4 ligand (Siddiqui-Jain et al., 2002). The observation that G-quadruplexes could act as transcriptional roadblocks was further consolidated by later studies on the G4s forming at c-KIT, BCL-2, KRAS, and VEGF oncogene promoters (Balasubramanian et al., 2011), supporting the potential of G4s to be targeted by stabilizing ligands for anti-cancer therapeutic intervention. However, direct experimental evidence to demonstrate the role of endogenous G4s as antagonists to gene expression is still lacking. More recently, immunoprecipitation experiments (using BG4) followed by sequencing have revealed an almost inverted picture to the earlier studies. Indeed, BG4-based ChIP-seq performed in keratinocytes (Hänsel-Hertsch et al., 2016), human xenografts (Hänsel-Hertsch et al., 2020), and liposarcoma cells (Lago et al., 2021) has shown that G4s prevalently mark highly transcribed genes, potentially acting as transcriptional enhancers rather than repressors. Similar results have been obtained using a small G4-selective peptide probe in various mammalian cell lines (Zheng et al., 2020). The proposed ways by which G-quadruplexes might contribute to activate gene expression are through guanine oxidation (Fleming & Burrows, 2020) or transcription factor binding (Spiegel et al., 2021). In view of the above, further studies are still required to fully elucidate and decipher the intricate link between G4s in gene promoters and transcription.

In contrast to DNA G4s, which might affect nuclear events, RNA G4s may exert functions both in the nucleus and in the cytoplasm: rG4s at the 5′- and 3′-untranslated regions (UTRs) of messenger RNA have been shown to suppress translation (A. Arora & Suess, 2011; Kumari et al., 2007), whereas RNA G4s near splice junctions may affect alternative splicing (Marcel et al., 2011). rG4s might also influence RNA localization, as reported by Subramanian and coworkers (Subramanian et al., 2011). Despite the aforementioned evidence, Guo and Bartel have found that RNA G4s are globally unfolded in mammalian cells (J. U. Guo & Bartel, 2016). To reconcile such contradictory experimental findings, rG4s are now accepted to form transiently in living human cells, with only specific rG4 subsets existing at any given time as a function of exogenous and endogenous stimuli. It is worth noting that the unique chemical properties of RNA G4s could make it possible to develop small molecules able to selectively target rG4s over dG4s, just like the recently discovered carboxypyridostatin (Di Antonio et al., 2012). Taking into consideration the growing number of diseases that involve RNA G4s in their

pathogenesis (Cammas & Millevoi, 2017), selective targeting of rG4s may be a strategically important therapeutic approach for specific disorders.

1.2.2 *i-Motifs*

The C-rich complementary strand of a G4-forming sequence might fold into another four-stranded DNA conformation, the i-motif (Figure 1.6) or i-DNA, further supporting the huge structural versatility and dynamism of nucleic acids. iMs are composed of two intercalated duplexes held together by hemi-protonated cytosine-cytosine$^+$ (C-C$^+$) base pairs (Figure 1.6). Analogously to the guanine residues in a G-tetrad, base pairing of cytosines occurs via Hoogsteen hydrogen bonds. Since the formation of i-motif structures requires the protonation of cytosines, iMs are favored at pH values more acidic than the intracellular pH. Such a pH-dependency had raised skepticism within the scientific community about whether these structures persisted in living cells. Anyway, evidence for iMs formation *in vivo* has been recently provided by in-cell nuclear magnetic resonance (NMR) experiments (Dzatko et al., 2018), and by the generation and characterization of an antibody fragment (iMab) able to specifically detect the i-motifs in the nuclei of human cells (Zeraati et al., 2018). Furthermore, it has been shown that C-rich sequences can fold into an i-motif conformation also at neutral pH and under molecular crowding conditions (Rajendran et al., 2010; Zhou et al., 2010). The presence of positive charges in the core of the i-motif structures is key for their stability (Mir et al., 2017). However, several other factors can impact iM stability, such as the C-tract length (Fleming et al., 2017), the precise sequence and the resulting interactions of the nucleobases in the loops (Benabou et al., 2016; Dai et al., 2010; Fujii & Sugimoto, 2015), the ionic strength of the solution (Jean-Louis Mergny et al., 1995), the use of molecular crowding conditions (J. Cui et al., 2013).

As it happens for G4s, sequences capable of folding into iM structures are evolutionarily conserved across species, in both prokaryotes and eukaryotes, and frequently cluster within telomeres and the regulatory regions of numerous genes (Garavís et al., 2013), thus implying their involvement in a variety of processes such as telomere homeostasis and transcription. Despite iMs at telomeres might share similar biological significance with G-quadruplexes, only limited efforts have been made so far to develop compounds able to specifically stabilize telomeric i-motif DNA. Among them, the carboxylated single-wall carbon nanotubes described by Chen et al. in 2012 stand out as the first example of molecules able to interfere with telomere homeostasis in cancer cells, by stabilizing the iMs (Y. Chen et al., 2012).

The close proximity of iM elements to transcription start sites (TSS), along with the observation that iM formation is maximal at late G_1 phase (Zeraati et al., 2018), when transcription levels are high, suggest an association between iMs and transcription. There has been no direct evidence of i-motifs regulating transcription until 2014 (Kang et al., 2014; Kendrick et al., 2014) and, interestingly, data collected so far supports such cis-regulatory function of iMs, indicating their potential to act as activators or repressors of transcription, depending on the specific promoter region, the unique i-motif sequence, and the associated transcription factors. In a study on the Bcl-2 i-motif, Kendrick and coworkers utilized two compounds (the one able to stabilize the structure, the other one capable of preventing iM assembly) as molecular tools to modulate Bcl-2 i-motif formation within B-cell lymphoma and breast cancer cell lines. According to the results, BCL-2 transcription levels were reduced when destabilizing the iM and increased in the presence of a stable i-motif structure in the BCL-2 promoter (Kendrick et al., 2014). The role of the i-motif as a transcriptional activator was further supported by the discovery of the hnRNP LL transcription factor binding to the Bcl-2 i-motif and enhancing gene expression (Kang et al., 2014). Surprisingly, in contrast to the Bcl-2 iM, formation of the c-Myc i-motif has been found to correlate with lower transcriptional activity than when the structure is unfolded (Sutherland et al., 2016). Such model of transcriptional regulation has been further supported by the later use of small-molecule stabilizers (Shu et al., 2018).
Given the evidence that the i-motifs can act as transcriptional regulators or hinder the activity of telomerase, they might be ideal targets for shutting down oncogenic signaling or inhibiting tumor growth. Therefore, targeting i-motifs with small molecules may represent an important and feasible anti-tumor therapeutic strategy.

1.2.3 *Relationship between G-quadruplexes and i-motifs*

Previous *in vitro* investigations have reported that G4s and iMs formation are mutually exclusive (Y. Cui et al., 2016; Dhakal et al., 2012) due to the steric hindrance resulting from the proximity of the two potential tetraplex structures. Furthermore, it has been demonstrated that G-quadruplex and i-motif structures are highly dependent on the formation of each other: inducing and stabilizing the ones, simultaneously destabilize the others in human cells (King et al., 2020). Anyway, more in-depth studies and new molecular tools are needed to understand G4 and iM dynamics relative to each other and shed light on the mechanisms underlying their intricate relationship.

Chapter 2
GENERAL METHODOLOGIES

2.1 **Principal component analysis**

Chemometrics is defined as the application of mathematical and statistical methods to chemistry. It provides valuable tools to analyze complex data matrixes and bring out strong patterns from them. The most widely used chemometric technique is the principal component analysis (PCA), developed by Pearson and Hotelling (Hotelling, 1933; Pearson, 1901).

PCA is an unsupervised pattern-recognition method that allows to reduce the dimensionality of a dataset, consisting of an overwhelming number of interrelated variables, while preserving as much relevant information as possible. Therefore, the starting point is a large matrix of data with *N* rows (*observations* — analytical samples, biological samples from different individuals, batches from a batch process …) and *K* columns (*variables* — spectral or chromatographic variables, for example). Through PCA, such complex dataset is projected over a space defined by a smaller set of new orthogonal (uncorrelated) variables, called principal components (PCs), which are ordered according to the percentage of variance that they retain: most of the valuable information is typically explained by the first few PCs, whereas the following ones often contain only chance variation or noise. So, to put it simply, PCA captures the essence of the data in a few principal components, which convey the largest variance in the dataset.

Of note, the outcome of PCA consists of two plots:

- A scores plot, displaying the distribution of the samples along specific PCs. Objects that are close to each other share similar characteristics, whereas samples lying in opposite parts of the plot have different properties.
- A loadings plot for each calculated PC, indicating the variables responsible for the distribution of the samples along that specific PC in the scores plot. Again, close variables are strongly correlated.

Thus, in order to understand the reason for a particular grouping observed in the scores plot, it is necessary to look at the same direction in the loadings plot.

In the present PhD thesis, PCA has been employed to explore huge spectral datasets (Papers I and II), as well as a matrix containing a large amount of data from circular dichroism melting experiments (Paper VI).

2.2 **Circular dichroism**

Circular dichroism is a spectroscopic technique that uses circularly polarized light to provide information about the adopted secondary structure of a biomacromolecule (a protein, a nucleic acid, etc.), with very low sample amount required and high sensitivity.

Light is an electromagnetic radiation, a complex wave which is composed of an electric ($\vec{E}$) and a magnetic ($\vec{B}$) field vector, each one oscillating at right angles to the direction of propagation. Being the two components invariably perpendicular to each other, only one of them (i.e., the E-component) can be considered to simplify the wave description.

A light source usually consists of a collection of randomly oriented emitters, with all possible orientations of $\vec{E}$ (unpolarized light). When light passes through a polarizer, an optical filter that transmits light waves vibrating in a single plane while blocking the others, it becomes polarized (Figure 2.1).

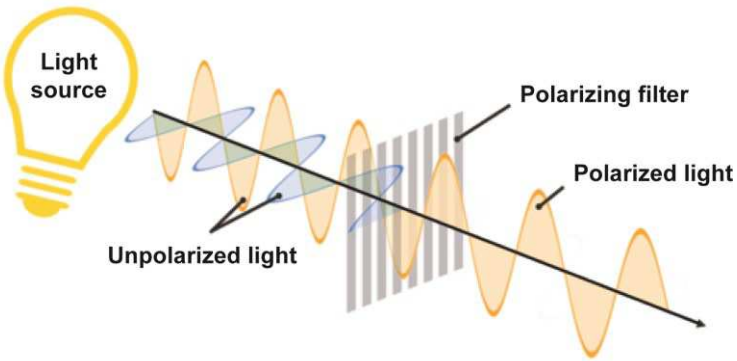


Figure 2.1. Schematic representation of unpolarized and polarized light.

As for linearly or plane-polarized light, the direction of the $\vec{E}$ vector is constant, while its magnitude varies. By contrast, in the case of circularly polarized light, the $\vec{E}$ vector rotates around the propagation axis maintaining a constant magnitude (Figure 2.2).

Furthermore, light can be circularly polarized in two directions: left and right. If the electric vector of the light coming towards the observer rotates counterclockwise, while looking at the

source, the light is said to be right-handed circularly polarized (R-CPL). If clockwise, then left-handed circularly polarized light (L-CPL).

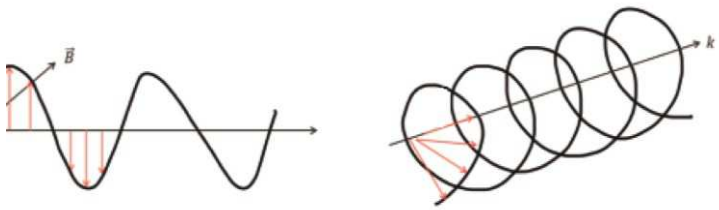


Figure 2.2. (Left) Linearly and (right) circularly polarized light.

Right- and left-handed circularly polarized lights will be absorbed to different extents when passing through a solution containing asymmetric chromophores or symmetric chromophores in asymmetric environments. Such difference is called circular dichroism (CD):

$$CD = A_L - A_R$$

where A_L is the absorbance of left-handed circularly polarized light and A_R is the absorbance of right-handed circularly polarized light.

According to the Lambert-Beer law:

$$A = \varepsilon \, l \, c$$

where ε is the wavelength-dependent molar extinction coefficient (M^{-1} cm^{-1}) of the species, l is the path length (cm), and c is the sample concentration (M).

During a CD measurement, l and c are constant. Therefore, the CD signal can be expressed as difference between molar extinction coefficients:

$$CD = \varepsilon_L - \varepsilon_R$$

CD experiments are performed on a spectrometer (also referred to as a spectropolarimeter) which is a highly specialized variation of an absorbance spectrophotometer (Figure 2.3): linearly polarized light of defined wavelength, after being selected by a monochromator, passes through a photoelastic modulator (PEM), which consists of a silica block attached to a piezoelectric element, and is converted into circularly polarized light. Particularly, an alternating drive voltage is applied to the PEM, so that the polarization state of the transmitted light alternates between R-CPL and L-CPL. An optically active sample will absorb the two

components of light to different extents. The spectrometer will measure this difference, at multiple wavelengths, and give a CD spectrum, which is similar to an absorption spectrum, but it can show either positive or negative bands, being CD a difference between molar absorptivities.

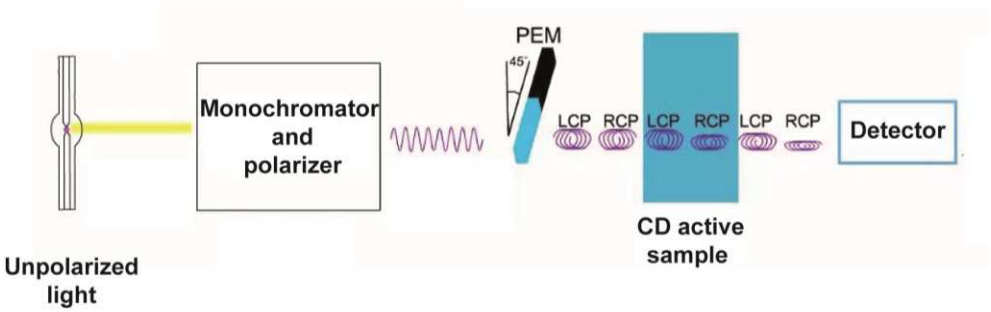


Figure 2.3. Schematic representation of a CD spectrometer.

Herein, CD spectroscopy has been widely employed, using standard protocols, to verify the secondary structure adopted by DNA sequences (Paper I) and to check the ability of the investigated compounds to affect their folding topology or thermal stability (Papers III, IV, and VI). Moreover, CD spectra have been acquired to validate the efficiency of chemical stapling in increasing the helicity of some peptides (Paper V).

2.3 **Fluorescence polarization**

Fluorescence is a two-stage physical phenomenon involving absorption of shorter-wavelength light by a chemical fluorophore (excitation), followed by the release of some of the absorbed energy as longer-wavelength light (emission).

Fluorescence polarization (FP) theory is based on the observation that when a small fluorescent-labeled molecule (often referred to as the probe or tracer) is excited by plane-polarized light, the emitted light will be largely depolarized because the molecule will tumble rapidly in solution during its fluorescence lifetime (time that elapses between excitation and emission), and a very low FP signal will be detected. By contrast, if the tracer is bound by another molecule and a large complex is formed, the rotation of the probe in solution will be slowed and the emitted light will remain polarized, giving a higher FP signal (Figure 2.4).

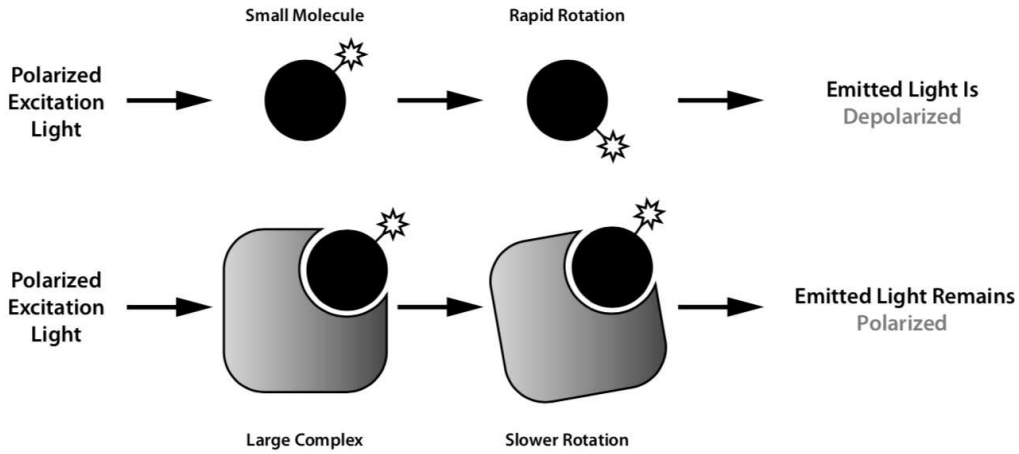


Figure 2.4. Principle behind the fluorescence polarization (FP) technique.

First described by Francis Perrin in 1926, FP is a versatile and truly homogeneous spectroscopic technique for studying molecular interactions directly in solution. It is adaptable to very low volumes (around 10–15 µL), it does not destroy the samples, and does not require the separation of bound and free species, which would be time-consuming and affect the reaction equilibrium, preventing an accurate quantitation of binding.

The mathematical equation behind an FP experiment is the following:

$$\frac{1}{FP} - \frac{1}{3} = \left(\frac{1}{FP_0} - \frac{1}{3}\right)\left(1 + \frac{RT}{\eta V}\,\tau\right)$$

where FP is the observed fluorescence polarization, FP_0 is the intrinsic polarization (polarization value in the absence of molecular rotation), R is the universal gas constant, T the absolute temperature, V the molecular volume (i.e., molecular size) of the fluorescent species, η the solution viscosity, and τ the fluorescence lifetime.

The above equation clearly illustrates that the observed FP mainly depends on V, if solution viscosity and temperature are held constant. Therefore, an increase in molecular volume due to receptor-ligand (Randall Bolger et al., 1998), DNA-protein (Lundblad, 1996), peptide-protein binding (Wu et al., 1997), or a decrease in V due to enzymatic degradation (R. Bolger & Checovich, 1994) can be followed by FP.

FP values are typically measured on an L-format, or single-channel, spectrofluorometer (Figure 2.5), which works as follows: excitation light at the proper wavelength, after being selected by a monochromator, passes through a polarizer that causes it to be plane-polarized. When the

polarized light illuminates the sample, the fluorophore is excited and emits light. The emitted light (at a specific wavelength), before reaching the detector, goes through the emission polarizer, which is rotated along parallel and orthogonal directions (with respect to the excitation plane), giving two intensity measurements (I|| and I⊥, respectively).

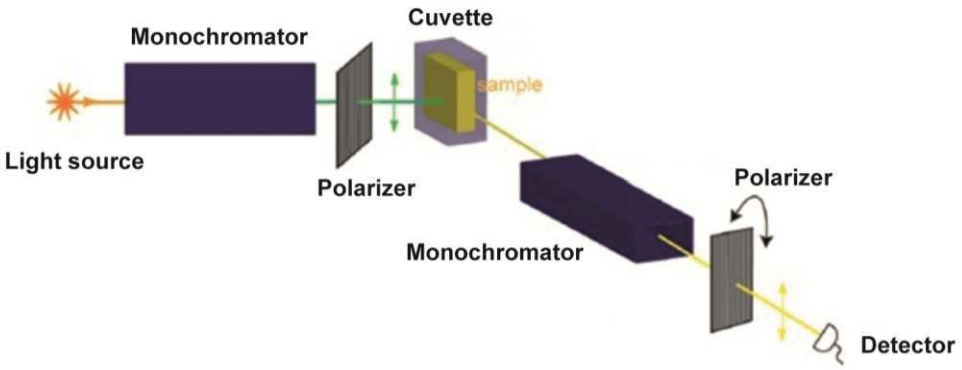


Figure 2.5. Schematic representation of a single-channel spectrofluorometer for FP.

Fluorescence polarization is then measured as the difference between the parallel and perpendicular emission intensities, normalized by the total fluorescence intensity of the emission beam:

$$FP = \frac{\mathrm{I}|| - \mathrm{I}\perp}{\mathrm{I}|| + \mathrm{I}\perp}$$

Herein, an FP assay (Figure 2.6) has been developed to measure binding of short peptides to G4 structures (Paper V).

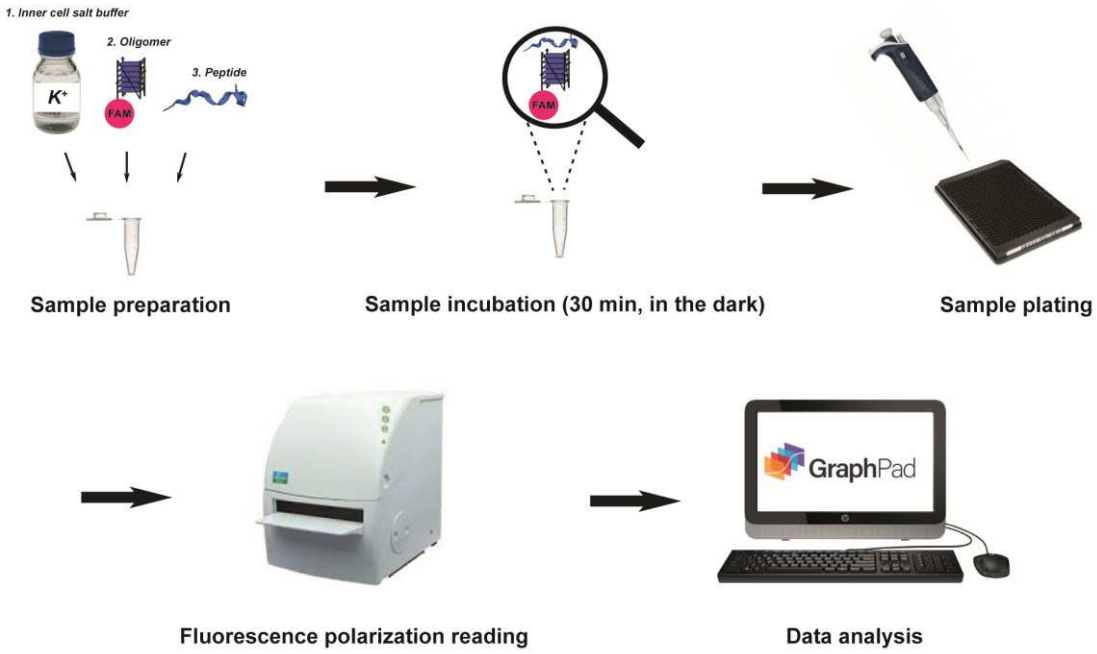


Figure 2.6. Experimental protocol of the FP assay reported in this PhD thesis.

2.4 **Trypan blue exclusion test**

Trypan blue is an anionic hydrophilic azo dye (~960 Da) which does not penetrate the cells unless their membrane is damaged. Therefore, it is used to quantify live cells by labeling dead cells exclusively: under light microscopy analysis, viable cells will have a clear cytoplasm, whereas dead cells will be stained blue.

In the experiments outlined here (Paper IV), the trypan blue exclusion test was performed by following the protocol depicted in Figure 2.7.

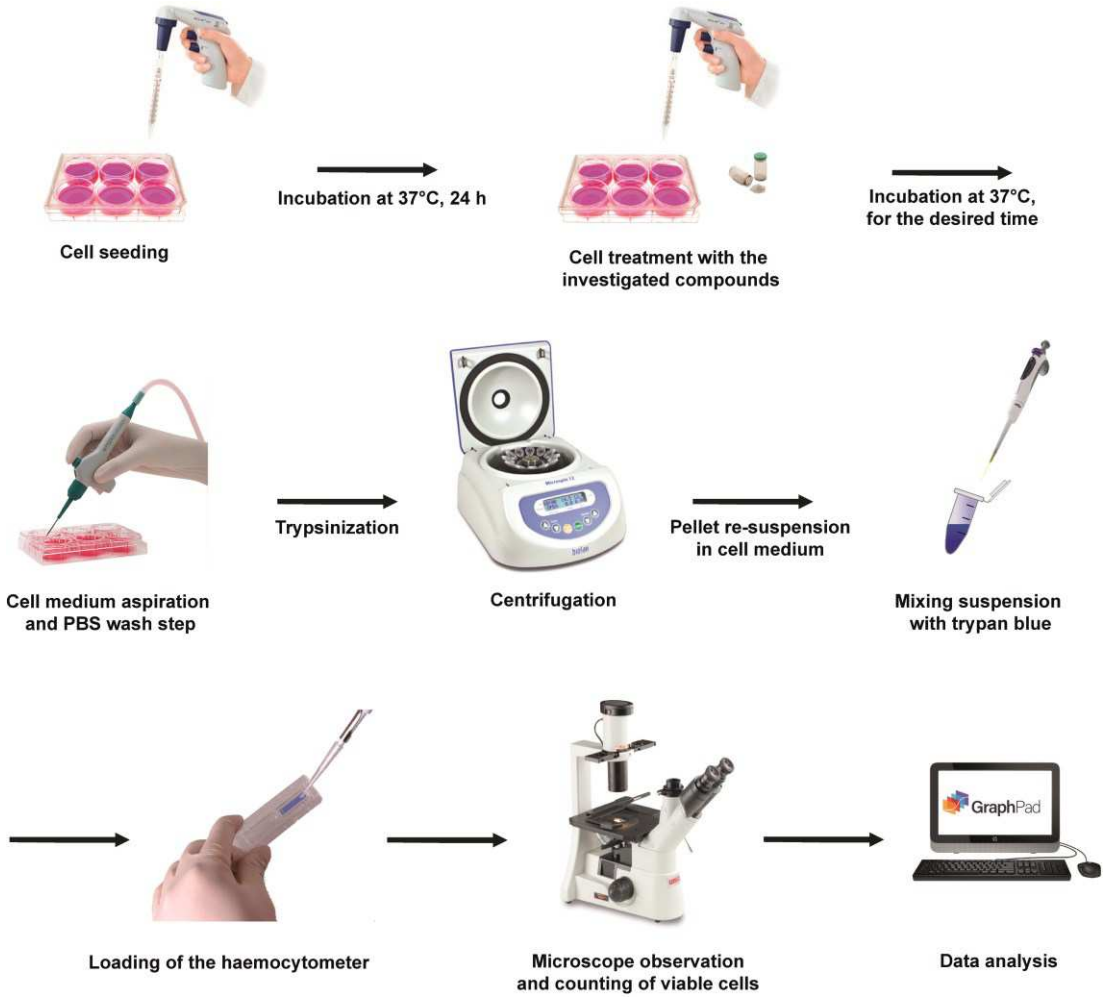


Figure 2.7. Experimental protocol of the trypan blue exclusion test reported in this PhD thesis.

2.5 **MTT assay**

The MTT assay is a standard colorimetric method to measure cellular metabolic activity as an indicator of cell viability and proliferation (Morgan, 1998). It is broadly employed to determine the *in vitro* cytotoxic or cytostatic activity of compounds on various cell lines. Such technique is based on the reduction of a soluble, yellow tetrazolium salt (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, or MTT) to insoluble purple formazan crystals, by active mitochondrial reductase enzymes in living cells (Figure 2.8). The absorbance (at 570 nm) of the colored formazan solution is then quantified using a spectrophotometer: the higher the absorbance of the solution, the higher the number of viable, metabolically active cells.

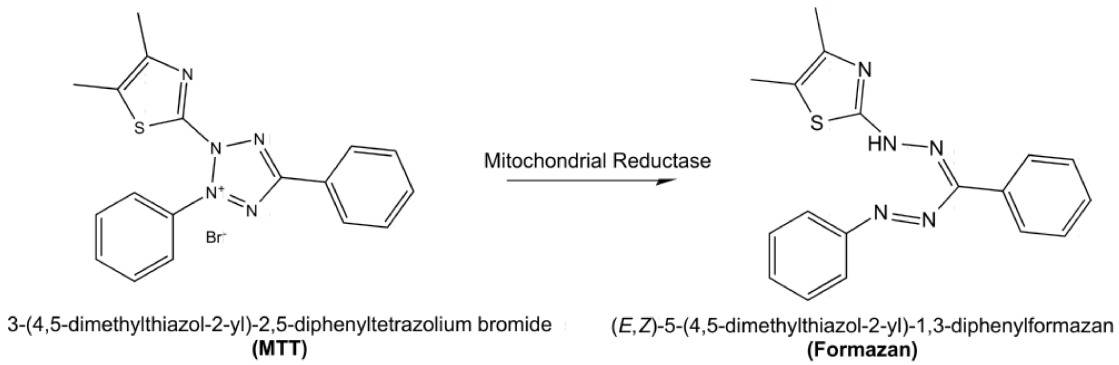


Figure 2.8. Metabolic reduction of MTT to formazan.

In the experiments described here (Paper IV), the MTT assay was carried out by following the protocol depicted in Figure 2.9.

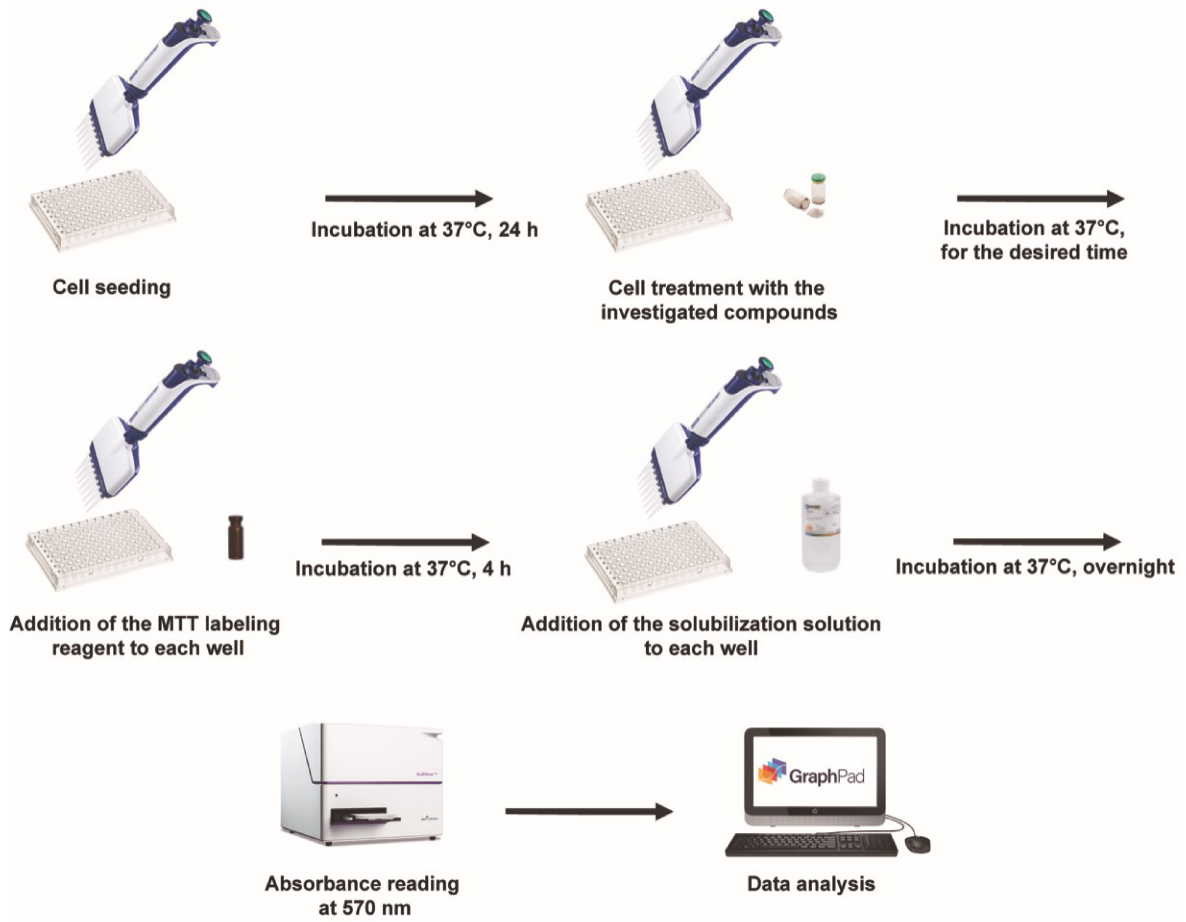


Figure 2.9. Experimental protocol of the MTT assay reported in this PhD thesis.

2.6 **Immunofluorescence microscopy**

Immunofluorescence (IF) is a powerful experimental method that exploits antibodies and microscopy to get insight into cellular processes, as well as to assess specific antigens (i.e., proteins) for their expression and localization.

Notably, there are two classes of immunofluorescence techniques, direct and indirect, which differ from each other in terms of number of antibodies used and fluorophore conjugation. Particularly, direct IF involves a single antibody (primary antibody) which is directly coupled to a fluorophore (also called fluorochrome), allowing for easy handling and quick visualization. By contrast, in indirect IF, a secondary fluorophore-coupled antibody, which specifically recognizes the primary antibody, is used to detect the structure of interest (Figure 2.10). Despite being more time-consuming (since it requires an additional step), indirect IF provides signal amplification, higher sensitivity and flexibility at a lower cost.

Fixation is an essential preliminary step, in IF staining, to immobilize target epitopes while retaining antigenicity, cellular and subcellular architecture. There is no universal fixative for every antigen: optimal fixatives and fixation methods may have to be selected empirically, based on the given antigen and sample type. Following fixation, cell membranes need to be properly permeabilized to give access to the antibody (or antibodies) applied successively. Of note, before any antibody treatment, blocking must be performed to prevent binding to off-targeted epitopes. Again, as no universal methodology exists, the best combination of blocking reagents and duration should be determined empirically. Finally, the sample is incubated with the primary (and, in case, secondary) antibody for antigen detection under a fluorescence microscope.

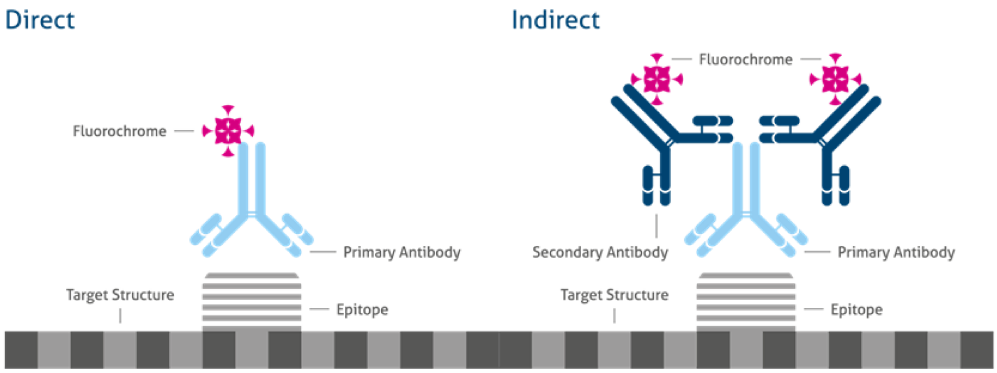


Figure 2.10. (Left) Direct and (right) indirect immunofluorescence techniques.

The main components of a fluorescence microscope overlap with those belonging to a traditional light microscope. However, there are two key features that sets the fluorescence microscope apart from the traditional microscope: the type of light source and the use of specialized filter elements (Figure 2.11). A detailed description of the most important parts of a fluorescence microscope is reported below:

- Light source (xenon or mercury arc lamp): it provides light, in a wide range of wavelengths, which is 10–100 times brighter than that from most incandescent lamps.
- Excitation filter: it filters out all wavelengths of the light source, except for the excitation range of the fluorophore under inspection.
- Dichroic mirror: it is placed at an angle of 45° between the excitation filter and the emission filter. Its function is to reflect the excitation signal towards the fluorophore and to transmit the emission signal towards the detector.
- Emission filter: it filters out the entire excitation range, only keeping the emission range of the fluorophore under investigation.
- Camera: it helps to record images of the specimen with high-resolution.

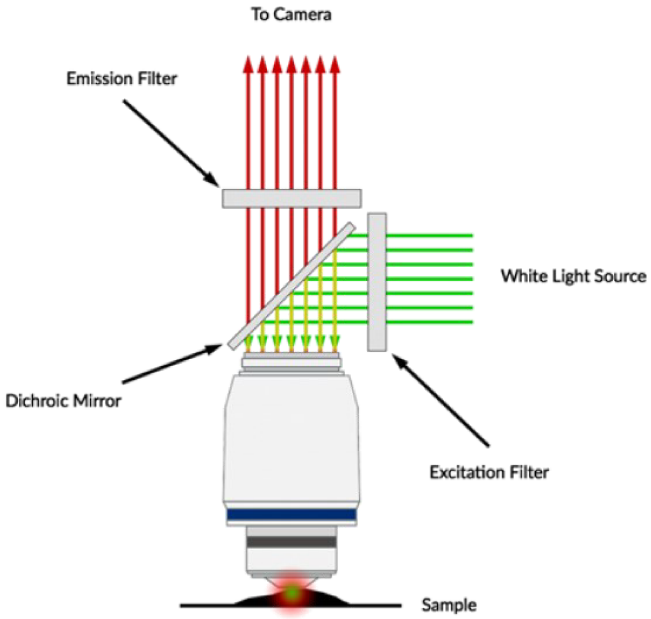


Figure 2.11. The main components of a fluorescence microscope.

In the studies presented here, indirect immunofluorescence was employed to evaluate the DNA damage response (Paper III) and visualize noncanonical DNA structures in living cells (Papers IV and VI).

Chapter 3
STUDY OF iM STRUCTURES

Despite the importance of understanding the role of the factors involved in the stabilization of iM structures, a systematic study that simultaneously explores the effect of pH, cation type, and cation concentration on iM formation *in vitro* still lacks. In this Chapter, a design of experiments to investigate the effect of those factors (and their interactions) on the formation of an i-motif structure is reported (Paper I).

Furthermore, the growing interest in the i-motifs as targets in anti-cancer therapy has increased the need for tools allowing a rapid and meaningful interpretation of the spectroscopic data of iM samples. In the present Chapter, a multivariate data analysis of the CD profiles and thermal difference spectra (TDS) of 180 DNA sequences is reported (Paper II), aiming at unveiling peculiar spectral regions that could be used as diagnostic features throughout the analysis of iM-forming sequences.

3.1 **Assessing the influence of pH and cationic strength on i-motif DNA structure** (Paper I)

3.1.1 *Introduction*

The design of experiments (DOE) is a chemometric tool to determine the relationship between factors affecting a system. Particularly, in such method, the factors of interest are systematically varied from the lowest to the highest value, and all possible factor combinations are executed in the same set of experiments, with the goal of screening the experimental space for optimum values (Leardi, 2009). This approach is opposed to the univariate one, also known as One-Variable-At-a-Time (OVAT) method, where all the parameters are kept constant while only one variable is optimized. The OVAT approach does not allow to investigate the potential interactions between the studied variables, whereas, in the DOE — using multiple linear regression (MLR) — it is possible to obtain a mathematical model by which the effect of the variables and their interactions can be estimated. In this work, with the aid of NMR and CD spectroscopies, we performed an entire set of experiments established by the DOE. Both the NMR and CD datasets were then examined using PCA, thus obtaining scores that were analyzed

by the MLR statistical tool to finally dissect the effect (and interactions) of pH, cation type, and cation concentration on the formation of an i-motif structure.

3.1.2 *Results and discussion*

We chose to explore the behavior of the iM structure formed by the C-rich human telomeric sequence (hTeloC, d[CCCT(AACCCT)$_3$]) at two pH values (quantitative variable), namely pH 5.0 and 6.6. The iM is completely structured at pH 5.0, whereas, at pH 6.6, the equilibrium is more shifted towards the unstructured form. We also wanted to investigate the effect of three different cations (qualitative variable), i.e., Na^+, K^+, and NH_4^+, on the formation of the hTeloC iM. Finally, the role of cation concentration (quantitative variable) was also considered, choosing 10 and 200 mM as values to be examined. The experimental plan is reported in Table 3.1. We decided to study all the 12 possible experimental conditions, this corresponding to a full factorial design (FFD)(Ebrahimi-Najafabadi et al., 2014). Therefore, 12 samples were prepared (in duplicate) and their ^{1}H-NMR and CD spectra were acquired.

Table 3.1. Experimental plan: each row corresponds to an experimental condition, whereas each column corresponds to a variable.

Experiment	X_1 (pH)	X_2 [cation] (mM)	X_3 (cation type)
1	5.0	10	Na^+
2	5.0	200	Na^+
3	6.6	10	Na^+
4	6.6	200	Na^+
5	5.0	10	K^+
6	5.0	200	K^+
7	6.6	10	K^+
8	6.6	200	K^+
9	5.0	10	NH_4^+
10	5.0	200	NH_4^+
11	6.6	10	NH_4^+
12	6.6	200	NH_4^+

NMR data. As shown in Figure 3.1, regardless of the cation type and concentration, the i-motif turned out to be very well-structured in all the samples prepared at pH 5.0 (showing the three typical imino signals of the base-paired cytosines in the range 15.80–15.40 ppm, and the aromatic signals at 8.44, 8.36, and 8.19 ppm (Phan et al., 2000). By contrast, the samples at pH 6.6, with all combinations of ion type and concentration, contained signals of both structured and unstructured DNA (the latter identified, among others, by the aromatic signals at 8.22 and

8.17 ppm), suggesting an equilibrium between those two species. Intriguingly, the NMR spectra of the samples containing 200 mM ion concentration, at pH 6.6, showed a smaller amount of structured DNA than those performed at 10 mM (pH 6.6), indicating a putative role of the ion concentration in stabilizing the iM structure.

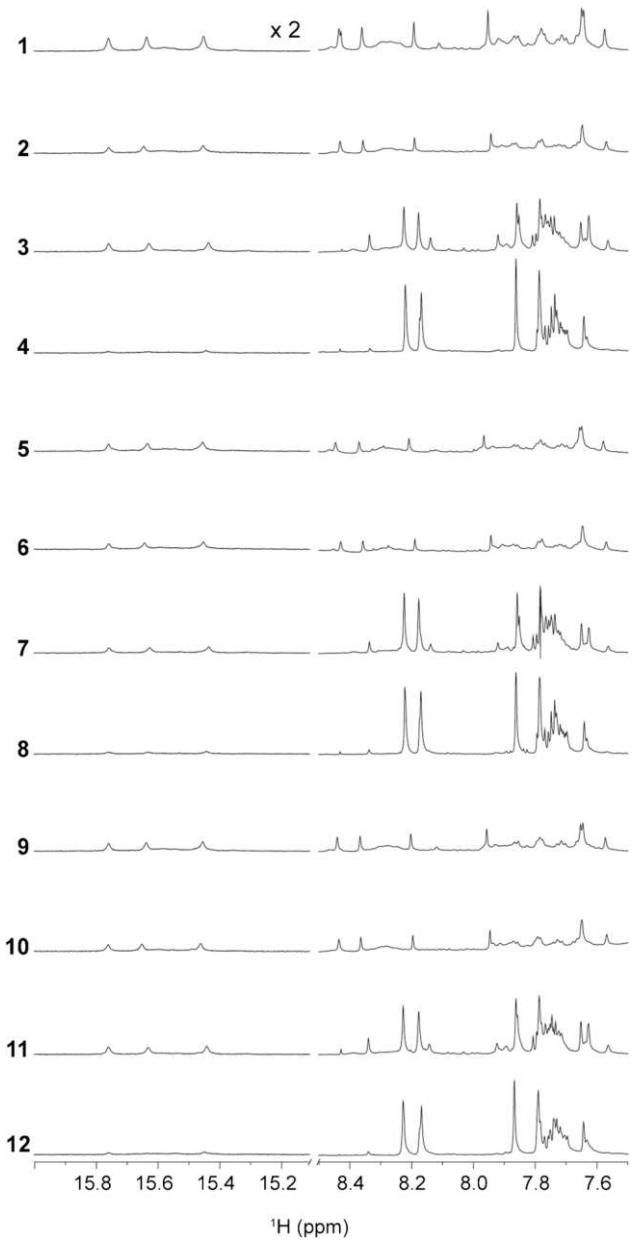


Figure 3.1. ^{1}H-NMR spectra of the 12 samples investigated in the full factorial design (FFD). Each number on the left-hand side refers to the corresponding experimental condition in Table 3.1. Figure from (Iaccarino et al., 2019).

In order to simultaneously take into account all the information contained in the NMR spectra, a PCA was performed and the resulting scores and loadings plots are reported in Figure 3.2.

The first principal component (PC1) clearly explained the effect of pH, while PC2 showed that the ion concentration affected iM formation only at pH 6.6.

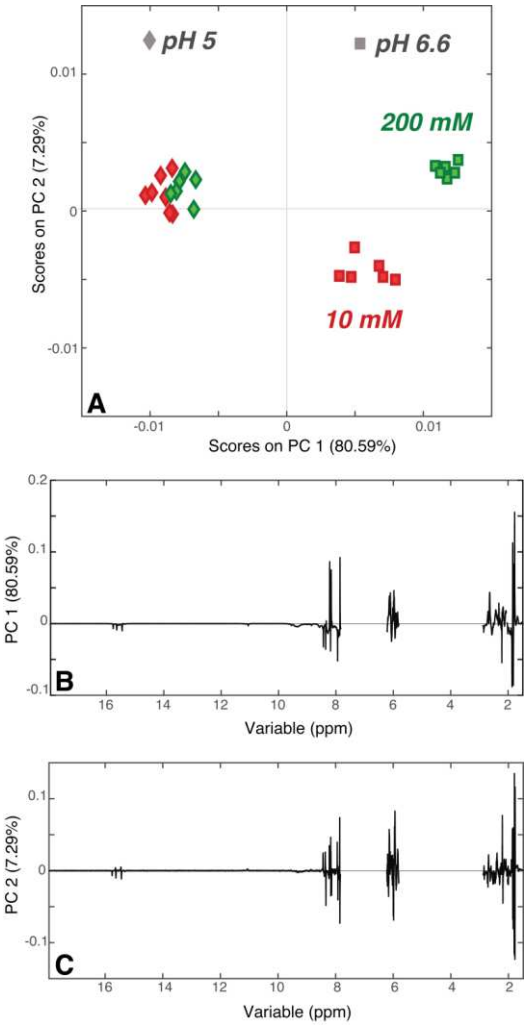


Figure 3.2. (**A**) PCA scores plot; (**B**) PC1 and (**C**) PC2 loadings plots of the FFD NMR data (including the replicates). Figure from (Iaccarino et al., 2019).

The NMR data was used to compute the following regression model:

$$Y = 0.0004 - 0.0081\,X_1^{(***)} - 0.0036\,X_2^{(***)} - 0.0006\,X_{3A} - 0.0008\,X_{3B} - 0.0030\,X_1X_2^{(***)}$$

where Y is the response value (iM formation), the first term is the constant term, X_1 is the linear term of pH, X_2 is the linear term of ion concentration, X_{3A} is the linear term of ion type at level K^+, X_{3B} is the linear term of ion type at level NH_4^+, and X_1X_2 corresponds to the interaction term between pH and ion concentration. It has to be noted that no linear term for Na^+ is present,

since Na^+ was the implicit level. Also, the statistical significance of the coefficients is reported (*p< 0.05, **p< 0.01, ***p< 0.001).

In the above model, the linear terms of pH and ion concentration, as well as the interaction term, had significant coefficients (meaning that those variables did affect iM formation), while the ion type had no relevant effect. Specifically, the fact that the coefficient of X_1X_2 was significant meant that it was not possible to interpret the effect of pH and ion concentration independently of each other. Indeed, we found that the highest percentage of iM formation was obtained at low pH values, with the ion concentration having basically no effect. Conversely, the ion concentration had a relevant effect on iM formation at higher pH values, with larger amounts of structure being obtained at low ion concentration.

CD data. CD spectra of the same iM samples were also acquired. As expected, the bands at around 288 and 260 nm (indicative of the formation of an iM structure and whose intensities are directly proportional to the amount of iM formed in solution) were significantly more intense in the CD spectra of the samples prepared at pH 5.0 (Figure 3.3). By contrast, the profile of the unstructured DNA (showing distinctive bands at 275 and 249 nm) was displayed at different extent in the other samples (Figure 3.3).

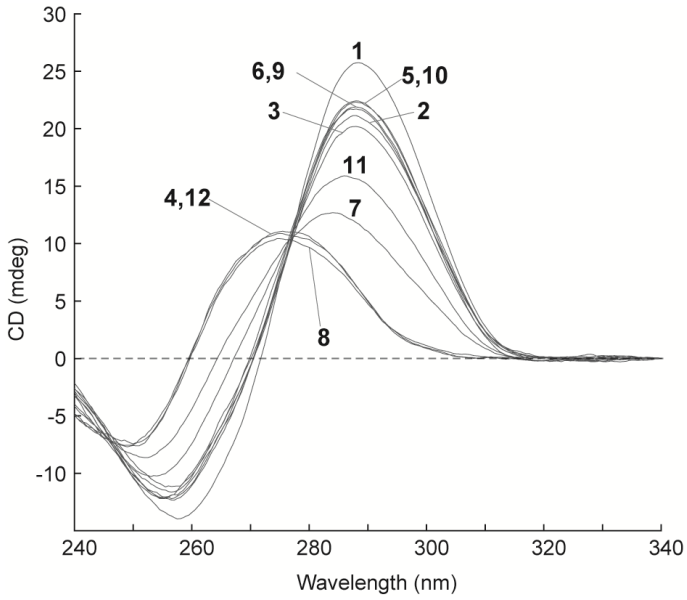


Figure 3.3. CD spectra of the 12 samples investigated in the FFD. Each number refers to the corresponding experimental condition in Table 3.1. Figure from (Iaccarino et al., 2019).

With the aim of taking into consideration all the acquired data points for each CD spectrum and retrieve potentially precious information, the CD dataset was then subjected to a principal component analysis. As shown in Figure 3.4A, the samples prepared at pH 5.0 had positive scores on PC1 (which explained 97.71% of the total variance and contained information about the effect of both pH and cation concentration) and placed themselves very close to each other, independently of the concentration of the ion, thus meaning that, at pH 5.0, the ion concentration had no effect. This perfectly agreed with what we observed with the NMR dataset. On the other hand, the samples prepared at pH 6.6 were spread over a wider area, ranging from those including 200 mM ion concentration (highly negative scores on PC1) to the ones at low ion concentration (10 mM) which turned out to be placed close to the samples at pH 5.0, thus confirming that the ion concentration had a pronounced effect only at higher pH values. Consistently with the NMR results, the PC1 loadings plot (Figure 3.4B) showed that the samples prepared under more acidic conditions (pH 5.0) contained the most structured i-motif (regardless of the ion concentration). By contrast, the effect of the ion concentration at pH 6.6 was considerable, with the experiments performed at higher ion concentration containing the least structured DNA and the experiments performed at lower ion concentration containing only partially structured i-motif.

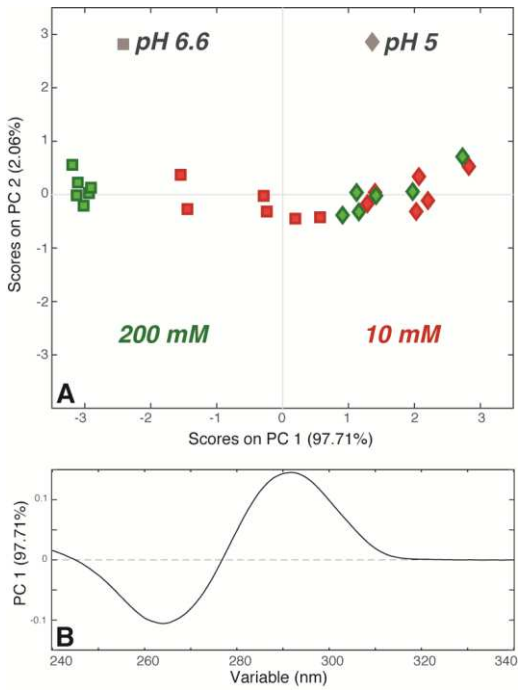


Figure 3.4. PCA (**A**) scores and (**B**) loadings plots of the CD data. Figure from (Iaccarino et al., 2019).

The CD data was also used to compute a regression model:

$$Y = 0.12 - 1.76\,X_1^{(***)} - 0.76\,X_2^{(***)} - 0.37\,X_{3A} - 0.00\,X_{3B} - 0.56\,X_1X_2^{(***)}$$

Once again, the interaction between pH and ion concentration proved to be highly significant, while the type of ion had no effect. The coefficients and the model resulting from the CD dataset were in perfect agreement with those arising from the NMR data, thus suggesting the uniqueness of the results obtained by two complementary spectroscopic techniques.

3.1.3 *Conclusions*

In the present investigation, we reported, for the first time, a systematic study on the role of pH, ion type and concentration on iM structure formation by combining DOE and multivariate data analysis. Our findings confirmed that the pH plays the main role in iM formation and revealed that, under experimental conditions close to neutral pH, the effect of cation concentration on structure formation is highly relevant (with a clear interaction between pH and ion concentration terms), while ion type has no significant effect.

3.2 **Effects of sequence and base composition on the CD and TDS profiles of i-DNA** (Paper II)

3.2.1 *Introduction*

The general formula of an intramolecular iM-forming sequence is $(C_nX_N)_3C_n$, where X can be either a C or non-C (T, A, G) deoxynucleotide; the presence of four cytosine tracts (C-tracts, C_n) allows for the generation of a C-stem, while the three spacer sequences (X_N), potentially forming loops, connect the four C-tracts (Figure 3.5). Herein, we employed multivariate data analysis to examine the CD spectral profiles and the UV thermal difference spectra of 180 iM-forming sequences with distinct C-tracts and spacer lengths (Table 3.2), reported in the companion paper of this investigation (Cheng et al., 2021), with the aim of detecting hidden but potentially informative bands in the spectra. Indeed, to the best of our knowledge, only the i-motif characteristic bands of the CD spectra (positive at 288 nm and negative at 264 nm) and TDS (positive at 240 nm and negative at 295 nm) have been considered so far, limiting the informative power of both CD and UV spectroscopies. Particularly, the CD spectra and TDS of the 180 sequences examined in this study were acquired at 13 different pH values (from 5.0 to 8.0, with 0.25 increments), but we primarily focused our attention on the spectra of the fully structured samples, at pH 5.0.

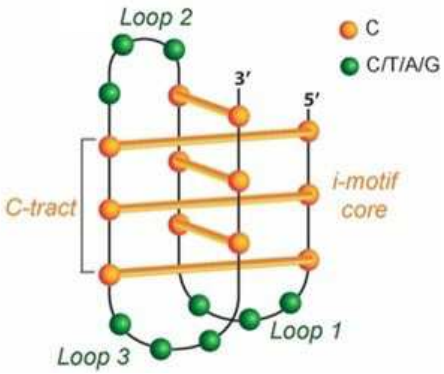


Figure 3.5. Schematic representation of an i-motif structure. Figure adapted from (Iaccarino et al., 2021).

3.2.2 *Results and discussion*

CD studies. The PCA computed on the 180 CD spectra acquired at pH 5.0 provided us with a series of valuable information. Firstly, samples with longer C-tracts were characterized by more intense CD bands at 288 nm and 264 nm, compared to those having shorter C-tracts. This was

in line with the fact that those bands are directly correlated to the number of C-C^+ base pairs in the iM structure (Jean-Louis Mergny et al., 1995; Školáková et al., 2019). In this regard, a very good positive correlation (Pearson's correlation coefficient r = 0.90) was found between the CD signal intensity at 288 nm and the number of Cs in the C-tracts. Secondly, the intensity of the CD signal at 275 nm turned out to be positively correlated with the total number of thymines in the sequence and negatively correlated with length of the central spacer (r = 0.65 and r = -0.61, respectively). Lastly, we found that, when the central spacer was shorter than 3 residues, the intensity ratio between the bands at 288 nm and 264 nm increased. The reason for this can be found in the structural requisites for an iM structure to form. Particularly, the central spacer is often responsible for the formation of a loop that requires at least 3 residues (Školáková et al., 2019). If the central spacer contains less than 3 residues, the i-motif structure will still be able to fold using some Cs of the adjacent C-tracts, thus including a lower number of C-C^+ pairs and some unpaired cytosines. Obviously, the CD spectrum of the sample will proportionally contain information about both paired Cs (a negative and a positive band at 264 and 288 nm, respectively) and unpaired Cs (a positive band centered at 275 nm) in the structure (Gray & Bollum, 1974). As a matter of fact, when both type of Cs contribute to define the general CD spectrum profile, the positive band of the unpaired Cs will be summed to the negative band at 264 nm, thus reducing its intensity, whereas the intensity of the positive band at 288 nm will further increased. Therefore, the higher the number of unpaired Cs, the higher the ratio between the intensity of the bands at 288 and 264 nm.

TDS studies. The TDS of a nucleic acid is obtained by subtracting the UV spectrum of the folded species, at low temperature, from that of the unfolded structure, at high temperature (J.-L. Mergny, 2005). The resulting profile has a unique shape and provides a distinctive signature for each type of nucleic acid structure. As for the CD spectra, the TDS (acquired at pH 5.0) of the 180 samples investigated in this study were subjected to PCA (after removing the spectral region between 220 and 240 nm, since affected by the buffer-related noise). According to the results, we found that the shorter the C-tract, the higher the band at about 295–310 nm. Furthermore, coherently with what we observed for the CD dataset, the information about the total number of Ts and length of the central spacer turned out to be hidden under the same wavelength (250–265 nm, in this case). Particularly, the band at around 250–265 nm proved to be positively correlated with the total number of thymines and negatively correlated with the length of the central spacer (if divided by the total number of Ts in the sequence).

3.2.3 *Conclusions*

In this work, we employed the power of multivariate data analysis to study the TDS and CD profiles of i-motif structures. Particularly, we were able to unveil peculiar informative bands that had not been reported before. Our results demonstrate that CD and TDS are much more informative for these structures than previously believed and that they can be used to retrieve interesting structural information on iM.

Table 3.2. List of the 180 DNA samples whose CD and TDS spectra were analyzed in this study.

C3 - tract	C4 - tract	C5 - tract	C6 - tract	Total spacer length
Group 1	*Group 16*	*Group 31*	*Group 46*	
T112-3	T112-4	T112-5	T112-6	
T121-3	T121-4	T121-5	T121-6	4
T211-3	T211-4	T211-5	T211-6	
Group 2	*Group 17*	*Group 32*	*Group 47*	
T113-3	T113-4	T113-5	T113-6	
T131-3	T131-4	T131-5	T131-6	5
T311-3	T311-4	T311-5	T311-6	
Group 3	*Group 18*	*Group 33*	*Group 48*	
T114-3	T114-4	T114-5	T114-6	
T141-3	T141-4	T141-5	T141-6	6
T411-3	T411-4	T411-5	T411-6	
Group 4	*Group 19*	*Group 34*	*Group 49*	
T115-3	T115-4	T115-5	T115-6	
T151-3	T151-4	T151-5	T151-6	7
T511-3	T511-4	T511-5	T511-6	
Group 5	*Group 20*	*Group 35*	*Group 50*	
T116-3	T116-4	T116-5	T116-6	
T161-3	T161-4	T161-5	T161-6	8
T611-3	T611-4	T611-5	T611-6	
Group 6	*Group 21*	*Group 36*	*Group 51*	
T221-3	T221-4	T221-5	T221-6	
T212-3	T212-4	T212-5	T212-6	5
T122-3	T122-4	T122-5	T122-6	
Group 7	*Group 22*	*Group 37*	*Group 52*	
T223-3	T223-4	T223-5	T223-6	
T232-3	T232-4	T232-5	T232-6	7
T322-3	T322-4	T322-5	T322-6	
Group 8	*Group 23*	*Group 38*	*Group 53*	
T224-3	T224-4	T224-5	T224-6	
T242-3	T242-4	T242-5	T242-6	8
T422-3	T422-4	T422-5	T422-6	
Group 9	*Group 24*	*Group 39*	*Group 54*	
T225-3	T225-4	T225-5	T225-6	
T252-3	T252-4	T252-5	T252-6	9
T522-3	T522-4	T522-5	T522-6	
Group 10	*Group 25*	*Group 40*	*Group 55*	
T226-3	T226-4	T226-5	T226-6	
T262-3	T262-4	T262-5	T262-6	10
T622-3	T622-4	T622-5	T622-6	
Group 11	*Group 26*	*Group 41*	*Group 56*	
T331-3	T331-4	T331-5	T331-6	
T313-3	T313-4	T313-5	T313-6	7
T133-3	T133-4	T133-5	T133-6	
Group 12	*Group 27*	*Group 42*	*Group 57*	
T332-3	T332-4	T332-5	T332-6	
T323-3	T323-4	T323-5	T323-6	8
T233-3	T233-4	T233-5	T233-6	
Group 13	*Group 28*	*Group 43*	*Group 58*	
T334-3	T334-4	T334-5	T334-6	
T343-3	T343-4	T343-5	T343-6	10
T433-3	T433-4	T433-5	T433-6	
Group 14	*Group 29*	*Group 44*	*Group 59*	
T335-3	T335-4	T335-5	T335-6	
T353-3	T353-4	T353-5	T353-6	11
T533-3	T533-4	T533-5	T533-6	
Group 15	*Group 30*	*Group 45*	*Group 60*	
T336-3	T336-4	T336-5	T336-6	
T363-3	T363-4	T363-5	T363-6	12
T633-3	T633-4	T633-5	T633-6	

The general formula of the iM-forming oligos under investigation is $C_{(3-6)}T_{(1-6)}C_{(3-6)}T_{(1-6)}C_{(3-6)}T_{(1-6)}C_{(3-6)}$. The samples are named as follows: the first 'T' letter means that all the spacers are composed of thymine bases only; the three consecutive numbers refer to lengths of the three spacers in the 5' to 3' direction; '-3, -4, -5 and -6' refer to the length of the C-tracts. For example, the T112-3 sequence is 5'-CCCTCCCTCCCTTCCC-3' (four repeats of 3 cytosines separated by one, one, and two thymines).

Chapter 4
TARGETING OF NONCANONICAL NUCLEIC ACID STRUCTURES FOR THERAPEUTIC INTERVENTION

The potential of noncanonical nucleic acid structures as promising therapeutic targets has been thoroughly discussed in Chapter 1. Herein, two different studies are reported in which we identified (Paper III) or rationally designed and synthesized (Paper IV) new G4-targeting molecules as potential anti-cancer agents. Of note, the latter investigation resulted from a fruitful collaboration with Prof. Claudiu Trandafir Supuran's research group in Florence (Italy).

4.1 **Targeting of telomeric repeat-containing RNA G-quadruplexes: from screening to biophysical and biological characterization of a new hit compound** (Paper III)

4.1.1 *Introduction*

The 3′ single-stranded overhang of telomeres is preceded by a double-stranded DNA region that is composed of the d(TTAGGG) repeated sequences on one strand and the complementary d(CCCTAA) repeats on the other one (Kwapisz & Morillon, 2020). The telomeric C-rich strand is transcribed into a long non-coding RNA (lncRNA) called TERRA (telomeric repeat-containing RNA), which includes the r(UUAGGG) G-rich motif (Azzalin et al., 2007), thus being able to fold into G4 structures (Figure 4.1).

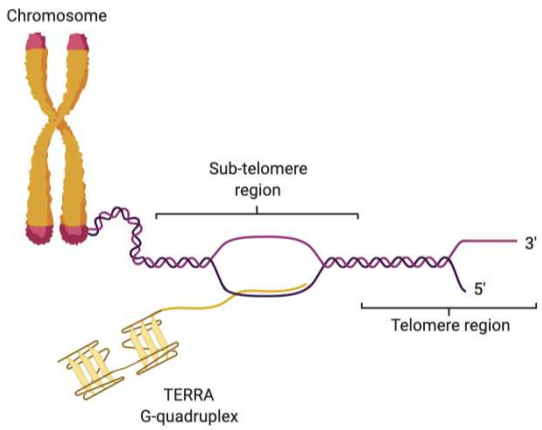


Figure 4.1. Schematic representation of telomeres and TERRA RNA.

TERRA molecules play critical roles in the telomere biology of both telomerase-positive and telomerase-negative cancer cells: in the former case, the TERRA G4 might bind to and sequester telomerase, preventing its access to the telomere, whereas in telomerase-negative cancer cells**,** which activate a mechanism of alternative lengthening of telomeres (ALT), TERRA could physically interact with the telomeric chromatin by forming RNA-DNA hybrids that are required for telomere homeostasis (Silva et al., 2021). As a consequence, the design of small molecules targeting the *TERRA G4s* is gaining increasing attention.

In the present study, a library of 103 drug-like putative *TERRA G4* binders was identified through a high-throughput *in silico* screening and tested for the ability to interact with and stabilize the *TERRA G4*. One out of 103 molecules proved to be a promising hit compound, and its ability to bind to and stabilize the *TERRA G4* was in-depth investigated both *in vitro* and in living cells.

4.1.2 *Results and discussion*

Virtual screening. Virtual-screening calculations were performed to identify drug-like molecules capable of binding to TERRA RNA, using the three-dimensional G4 structure formed by the r(UAGGGUUAGGGU) sequence as target (*TERRA G4*, Figure 4.2A). The process resulted in the identification of 103 compounds that were purchased and further studied.

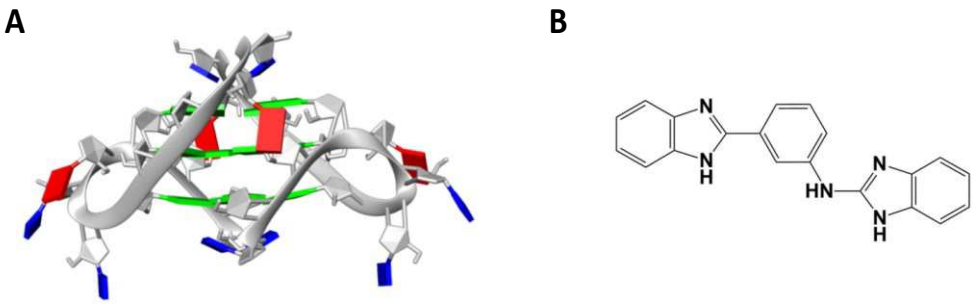


Figure 4.2. (**A**) Bimolecular G4 structure formed by TERRA RNA (guanine, green; adenine, red; uracil, blue). (**B**) Chemical structure of compound **BPBA**. Figure from (Marzano et al., 2021).

CD studies. In order to identify true hits, the 103 computationally selected small molecules were preliminarily screened for their ability to interact with the *TERRA G4* by CD spectroscopy. Although circular dichroism is not usually used for large-scale screening purposes, it is widely employed to reliably select nucleic acid-interacting compounds (Paramasivan et al., 2007; Santos et al., 2021). Indeed, CD experiments require unmodified oligonucleotides, so that

changes in the CD melting curves should only be ascribed to the direct interaction of the putative ligand with the nucleic acid structure. Firstly, CD spectra were recorded to examine the potential of the investigated compounds (10 molar equiv) to alter the dimeric propeller-type parallel conformation of the *TERRA G4* in K^+ buffer. No significant variations of the CD profile were detected upon addition of the compounds to the folded *TERRA G4* structure, suggesting that they did not modify its conformation. Then, the G4-stabilizing properties of the compounds were evaluated by CD melting experiments, measuring the ligand-induced changes in the apparent melting temperature ($\Delta T_{1/2}$) of the G4 structure. According to the results, only 1 of the 103 tested compounds (**BPBA**, Figure 4.2B) significantly increased the $T_{1/2}$ of the *TERRA G4* (Table 4.1). The remaining 102 compounds were discarded since unable to considerably stabilize such structure ($\Delta T_{1/2} < 3$ °C). Importantly, **BPBA** displayed selectivity for G4s over double-stranded DNAs (*Hairpin*$_{20}$ and *Hairpin*$_{27}$), as also highlighted by Förster resonance energy transfer (FRET) melting experiments (Table 4.1).

We next sought to explore the ability of **BPBA** to interact with other parallel DNA and RNA G-quadruplexes, and to discriminate between parallel and antiparallel G4 conformations. To this aim, we used G4-forming DNA sequences deriving from the c-KIT (*c-Kit2 G4*) and c-MYC (*c-Myc G4*) gene promoters, forming parallel G4 structures in K^+-containing buffer (Ambrus et al., 2005; Hsu et al., 2009); G4-forming RNA sequences deriving from the GSEC lncRNA (*GSEC G4*) and the 5′-UTR of BCL-2 mRNA (*Bcl-2 G4*) (S. Chen & Shen, 2020; Matsumura et al., 2017; Shahid et al., 2010), which fold into a parallel conformation; the 23-mer truncation of the human telomeric DNA sequence (*Tel*$_{23}$ *G4*), which rather folds into an antiparallel (3 + 1) hybrid G4 conformation in potassium-containing solution (Dai et al., 2008; Phan et al., 2007). The results pointed out **BPBA** ability to preferentially stabilize the parallel G4 conformations of *c-Kit2* and *c-Myc*, with only a negligible effect on the hybrid G4 structure of *Tel*$_{23}$ *G4*. Also, it could preferentially stabilize the *TERRA G4* over other RNA G4s.

Table 4.1. Analysis of **BPBA** interaction with the investigated oligonucleotides.

	Circular dichroism (CD) melting						
$\Delta T_{1/2}$ (°C) [1]	*TERRA G4*	*c-Kit2 G4*	*c-Myc G4*	*Tel*$_{23}$ *G4*	*GSEC G4*	*Bcl-2 G4*	*Hairpin*$_{20}$
	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* + *Hairpin*$_{27}$ (1:30)			*F-TERRA-T* + *Hairpin*$_{27}$ (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*		*Hairpin*$_{27}$	
	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)						

[1] $\Delta T_{1/2} = T_{1/2}$ (oligonucleotide + 10 ligand equiv) − $T_{1/2}$ (oligonucleotide). $T_{1/2}$ values in the absence of ligand are 74.3 (± 0.1) °C for the *TERRA G4*; 59.7 (± 0.1) °C for the *c-Kit2 G4*; 75.5 (± 0.1) °C for the *c-Myc G4*; 53.7 (± 0.1) °C for the *Tel*$_{23}$ *G4*; 78.8 (± 0.1) °C for the *GSEC G4*; 74.7 (± 0.2) °C for the *Bcl-2 G4*; 65.5 (± 0.2) °C for *Hrp*$_{20}$; 73.5 (± 0.1) °C for F-TERRA-T. [2] n.d. = not determinable.

Fluorescent intercalator displacement (FID) studies. To gain insight into the binding affinity of **BPBA** for the different RNA/DNA G4s, FID experiments were carried out. The assay is based on the competitive displacement of a light-up fluorescent probe (thiazole orange (TO), in this case) from the DNA, upon addition of increasing amounts of a candidate ligand (Largy et al., 2011; Monchaud et al., 2006) (Figure 4.3). TO is almost nonfluorescent when free in solution, whereas it is strongly fluorescent when bound to the DNA. Ligand-induced TO displacement decreases TO fluorescence, thus allowing to determine the relative ligand binding affinity for the structure under examination. According to the results (Table 4.1), **BPBA** showed higher affinity for the *TERRA G4* than for the *c-Kit2 G4* and *Hairpin*$_{27}$ (used as duplex model). In the case of the *c-Myc* and *Tel*$_{23}$ *G4s*, it was not possible to reach the 50% displacement of TO even after the addition of a large excess of the binder (20 molar equiv), clearly suggesting weaker ligand interactions.

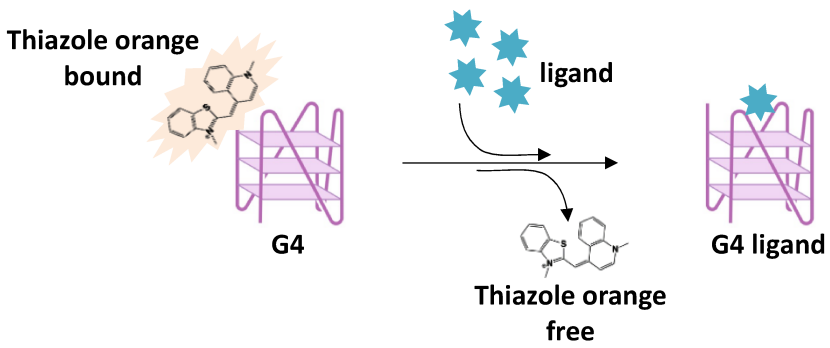


Figure 4.3. Thiazole orange as the fluorescent intercalator in our FID experiments.

Furthermore, with the aim of obtaining information about the binding mode of **BPBA** to the *TERRA G4*, an ethidium bromide (EB) displacement assay was performed using fluorescence spectroscopy. EB binds to duplex DNA through intercalation and to a DNA G4 through π-π stacking on the external G-tetrads (Q. Guo et al., 1992). In the absence of DNA, EB emits weak fluorescence at 595 nm, whereas its fluorescence emission is strongly enhanced upon association with G4s, due to the hydrophobic environment experienced by EB upon binding to the nucleic acid (Sengupta et al., 2013). Therefore, if a G4 ligand binds to a G-quadruplex structure via end-stacking mode, is should displace EB and decrease its fluorescence emission intensity (Figure 4.4).

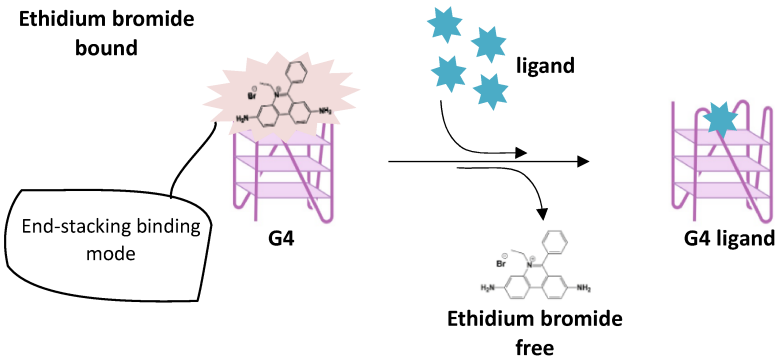


Figure 4.4. Ethidium bromide displacement assay.

Displacement titrations (performed by adding increasing amounts of **BPBA** to the *TERRA G4*/EB complex) showed a substantial decrease in the fluorescence emission intensity of EB, thus suggesting an end-stacking binding mode for this ligand to the *TERRA G4*. The concentration of **BPBA** required to give 50% decrease in EB fluorescence (DC_{50}) was also determined (Table 4.1), providing a value of 1.3 (± 0.4) μM which confirmed the strong interaction between **BPBA** and this G4 motif.

Biological studies. Afterwards, we used the crystal violet assay to evaluate the antiproliferative effect of **BPBA** on human cervix cancer cells (HeLa, characterized by high telomerase activity and low TERRA expression) and human osteosarcoma cells (U2OS, lacking telomerase activity and expressing high levels of TERRA) (R. Arora et al., 2014; Pompili et al., 2017). Interestingly, U2OS turned out to be significantly more sensitive to **BPBA**, with respect to HeLa cells (Figure 4.5). Moreover, in order to assess the ability of **BPBA** to bind to and stabilize TERRA also *in cellulo*, both HeLa and U2OS cell lines were treated (for 72 h) with 1, 5 or 10

μM **BPBA**, the RNA was then extracted and processed for TERRA analysis by real-time quantitative polymerase chain reaction (RT-qPCR). **BPBA** induced a stabilization of TERRA expression in U2OS that led to an accumulation of the RNA within the cells. 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.

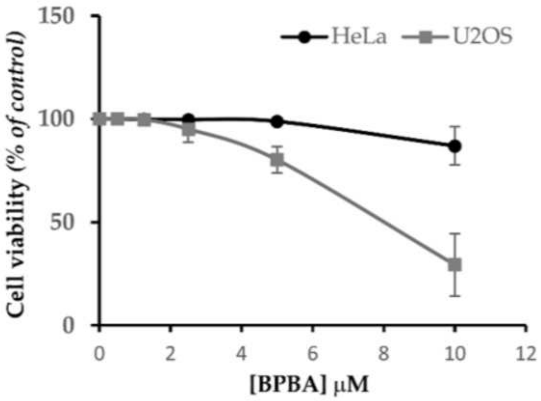


Figure 4.5. Antiproliferative effect of **BPBA** on low vs. high TERRA-expressing human cancer cells. Figure adapted from (Marzano et al., 2021).

Finally, as shown in Figure 4.6, we found that U2OS treatment with 10 μM **BPBA** induced a DNA damage response (DDR) at telomeres, potentially explaining the observed cytotoxic effect of the compound on human osteosarcoma cells. The DDR induction was negligible in HeLa cells.

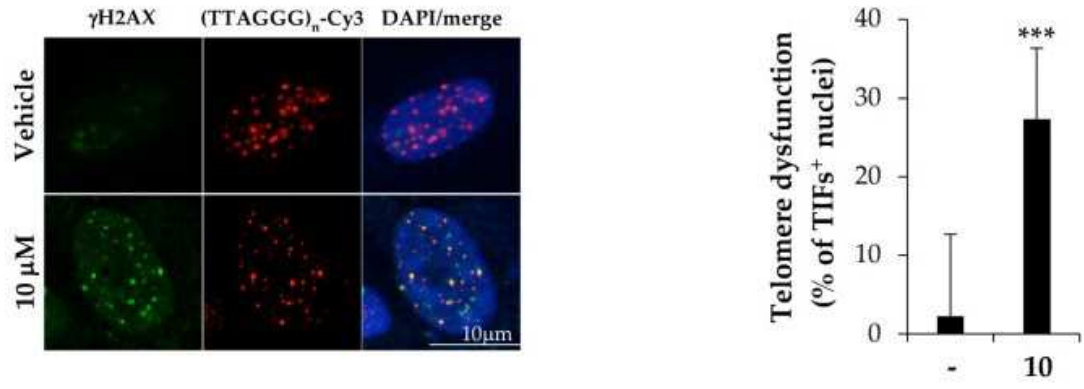


Figure 4.6. Immunofluorescence in combination with fluorescence in situ hybridization (FISH), using the telomere-specific $(TTAGGG)_n$-Cy3 PNA probe on U2OS. Figure adapted from (Marzano et al., 2021).

4.1.3 *Conclusions*

Herein, through the combination of virtual screening and several biophysical assays (CD, FRET, FID), we succeeded in the identification of a new *TERRA G4* binder (**BPBA**) with high selectivity for G4s over duplex DNA and enhanced binding affinity towards the *TERRA G4* vs. other G4-forming DNA/RNA sequences. The unique chemical structure of **BPBA**, two benzoimidazole units connected by an aniline residue, gives the molecule extensive planarity, probably also allowing for the optimal distribution of polar groups to interact with the *TERRA G4*. Biological studies demonstrated that **BPBA** can bind to and stabilize *in cellulo* TERRA lncRNAs, possibly by sequestering them from the physiological turnover cell machinery. Moreover, TERRA stabilization induced a DDR at telomeres, putatively by displacing TERRA from telomeric DNA. Overall, we speculate that targeting of TERRA with G4 ligands could represent an effective pharmacological strategy to hit ALT-positive tumors, which correlate with high TERRA expression and are generally associated with the worst prognosis.

4.2 **Design, synthesis, biophysical and biological characterization of carbonic anhydrase inhibitors-Berberine hybrids as new multi-target anti-cancer drugs** (Paper IV — Manuscript in preparation)

4.2.1 *Introduction*

Reprogramming of energetic metabolism has been recognized as one of the major hallmarks of cancer. In normal cells, glucose is oxidized to CO_2 and H_2O through the tricarboxylic acid (TCA) cycle, releasing 38 moles of adenosine 5'-triphosphate (ATP) per mole of glucose. By contrast, the uncontrolled division of cells within solid tumors creates poorly perfused regions — i.e., hypoxic conditions — in which oxidative phosphorylation (OXPHOS) is restricted and glucose is metabolized to lactate in an oxygen-independent process (glycolytic shift). Since glycolysis is less energy-efficient than oxidative phosphorylation, only 2 net molecules of ATP per 1 molecule of glucose are generated, cancer cells increase their glucose uptake (Gatenby & Gillies, 2004). Increased glucose intake and fermentation of glucose to lactate converge in the so-called "Warburg effect" (Liberti & Locasale, 2016) (Figure 4.7), which has been observed in several neoplasms, including breast, colon, cervical, and liver cancer (Jiao et al., 2018; Marshall et al., 1979; Park et al., 2013).

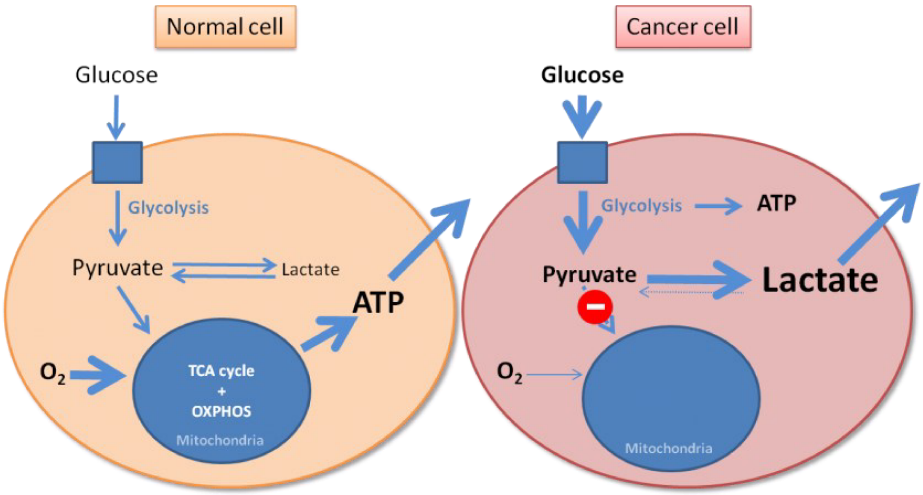


Figure 4.7. Schematic representation of the Warburg effect.

Furthermore, in hypoxic tumors, the massive production and rapid extrusion of lactic acid reduce the pH of the surrounding microenvironment (pH_e) in the range 6.7–7.1 (normal cells display a pH_e value around 7.4), with only a marginal decrease in the intracellular pH (> 7.2).

All the above metabolic changes and the selective adaptation of cancers cells to extracellular acidosis are orchestrated by transcription factors called hypoxia-inducible factor 1 and 2 (HIF-1/2), as demonstrated in the last decades. HIF-1/2 sense tumor oxygen levels and activate genes involved in metabolism, angiogenesis, pH regulation. Particularly, two of the proteins that have been identified to be upregulated in many hypoxic tumors are the carbonic anhydrase (CA) IX and XII isoforms (Wykoff et al., 2000, 2001). These transmembrane metalloenzymes, by catalyzing the reversible hydration of carbon dioxide to produce bicarbonate ions and H^+, have been shown to maintain the pH differential in tumor cells (Swietach et al., 2014), contributing to an aggressive metastatic phenotype (Martínez-Zaguilán et al., 1996; Stock & Schwab, 2009). Humans express a total of 15 CA isoforms, which differ in terms of cellular distribution, physiology, and function (Figure 4.8).

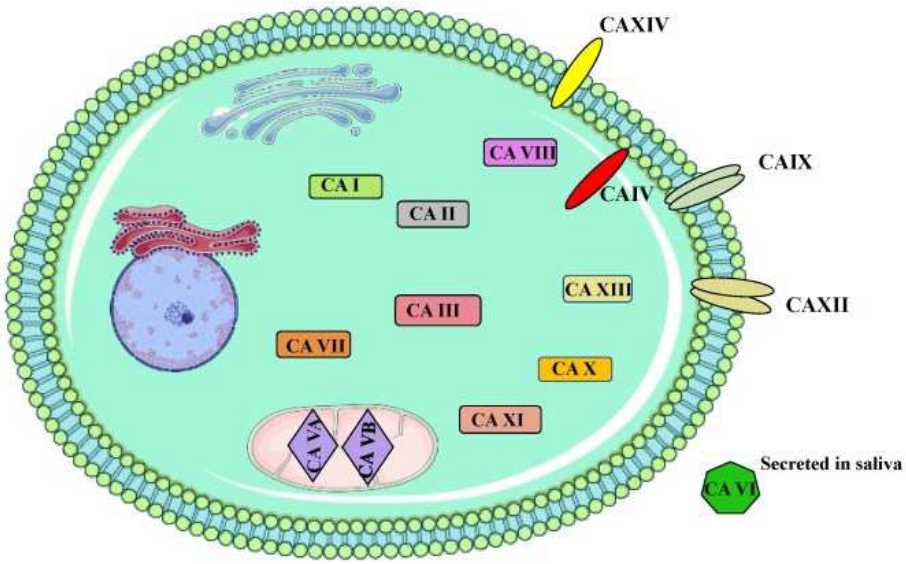


Figure 4.8. Subcellular localization of carbonic anhydrases. Figure from (Singh et al., 2018).

The CA active site is conical in shape and contains a zinc ion which is critical for the catalytic activity of the enzyme. The hypothesis that interfering with the functioning of such proteins may have anti-tumoral effects was first proposed by Pouysségur in 2006 (Pouysségur et al., 2006). Since then, the scientific community has made great efforts in developing CA inhibitors (CAIs) and inhibition of CA IX and XII is now a validated therapeutic approach to hamper growth, invasiveness, and metastasis of primary tumors.

Overall, CAIs are characterized by three structural components (Y. Mboge et al., 2015):

- A zinc-binding group (ZBG) which anchors the compound to the center of the CA active site.
- A linker which interacts with residues in the active site, further stabilizing the binding of the inhibitor.
- A tail moiety that promotes isoform specificity through interactions with isoform unique residues.

Sulfonamides and their thioesters are the most studied CAIs, with the nitrogen atom of the sulfonamide moiety directly coordinating the zinc ion in the catalytic site. These compounds are potent inhibitors, often exhibiting binding affinities in the micro- to nano-molar range (Supuran & Scozzafava, 2000). Coumarin-based derivatives have been recently identified as a new class of CAIs (Maresca *et al.*, 2009). They possess a unique inhibition mechanism, acting as "prodrug inhibitors". Particularly, the 2-hydroxy-cinnamic acid resulting from the hydrolysis (mediated by esterases) of their lactone ring, binds to the entrance of the enzyme active site, where a high variability of amino acid residues among the different mammalian CA isoforms is present, and where other CAIs were not found to bind earlier. This would explain why coumarins fall among the most isoform-selective CA inhibitors known to date.
In the present study, taking into account the roles of CA IX/XII and G-quadruplex structures in the context of cancer, we designed and synthesized molecular hybrids, containing both an inhibitor of CA IX/XII and a well-known G4 stabilizer, as potential multi-target anti-cancer agents. Indeed, the use of multi-target compounds is emerging as a valuable strategy to treat multi-factorial diseases like cancer, possibly providing a more predictive pharmacokinetic profile, reduced risk of drug interactions, and a higher treatment adherence compared to combined therapies. Particularly, we chose **Berberine** as G4 ligand to be incorporated in the structure of the hybrids: it is a natural product which has long been used in traditional Chinese medicine for its anti-cancer activity (Maiti & Kumar, 2010) and which binds to G4 DNA structures with a good selectivity over duplex DNA (Franceschin et al., 2006), thus being among the most attractive G4-targeting compounds. The ability of the synthesized hybrids to inhibit CA IX/XII isoforms and bind to and stabilize G4 structures *in vitro* was assessed, as well as their cytotoxic activity on CA IX-positive human cervix cancer cells.

4.2.2 *Results and discussion*

Synthesis. The Huisgen 1,3-dipolar cycloaddition between azides and alkynes represents a synthetic approach that has gained considerable interest in medicinal chemistry. Such click chemistry reaction is rapid, versatile, and allows to generate a stable linker (a 1,4- or 1,5-disubstituted 1,2,3-triazole, depending on whether it is catalyzed by Cu (I) or ruthenium-based catalysts, respectively) between two distinct moieties. For these reasons, it was employed to chemically connect the **Berberine** core with the CAI portion and to generate the hybrids (for a total of 21 compounds). Particularly, two series of hybrids were synthesized: the one containing a sulfonamide CAI portion (hereafter referred to as series PA), the other one a coumarin CAI portion (series AL) (Figure 4.9).

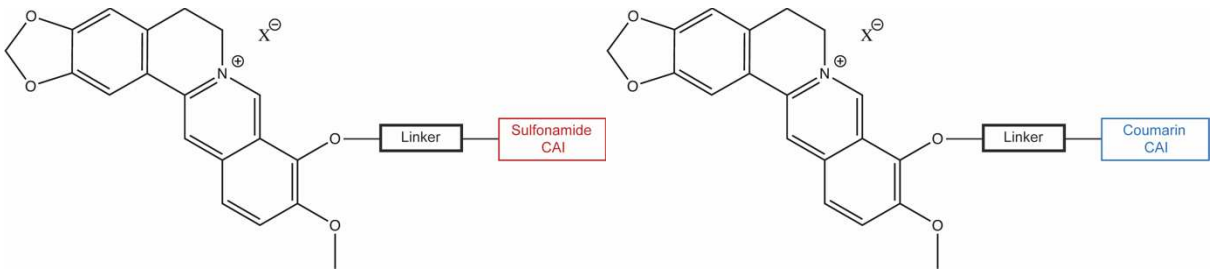


Figure 4.9. General structures of the molecular hybrids designed and synthesized in this study.

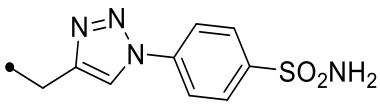
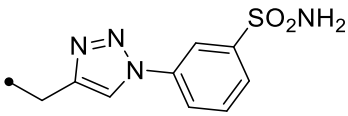
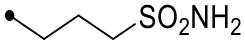
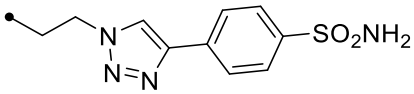
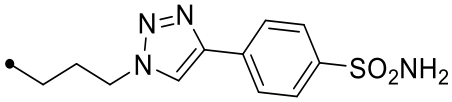
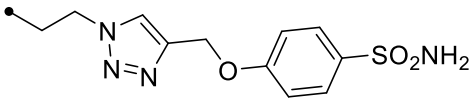
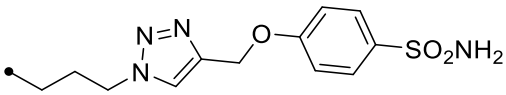
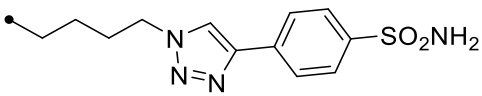
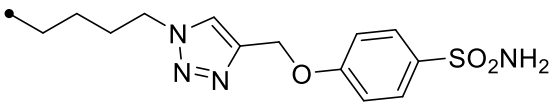
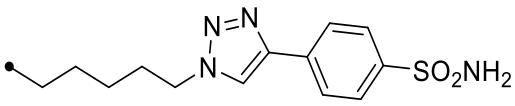
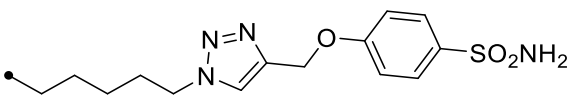
Evaluation of CA inhibition *in vitro*. The 21 synthesized hybrids were profiled by a stopped-flow CO_2 hydrase assay, originally reported by Khalifah (Khalifah, 1971), for their ability to inhibit the targeted CA IX/XII and the off-targeted CA I/II isoforms, *in vitro*.

As for compounds from series PA (Table 4.2), the lowest inhibition efficacy was found towards the off-targeted CA I isoform, with inhibition constants (K_i) ranging from 75.1 nM to 3670.2 nM. Both tumor-associated carbonic anhydrases (IX and XII) turned out to be effectively inhibited by most of the hybrids, with K_i values in the range 0.2–17.8 nM and 2.9–35.9 nM, respectively. Of note, the longer the linker between the **Berberine** scaffold and the aromatic CAI portion, the stronger the inhibition efficacy on CA IX. Finally, although the activity of the housekeeping CA II proved to be significantly blocked by the compounds, the sub-nanomolar K_i values measured towards CA IX delineated a quite selective profile, except for **PA-68** which turned out to be quite unspecific.

As for hybrids from series AL, their coumarin-like CAI portion clearly rendered them more selective CA IX/XII inhibitors than compounds from series PA (Table 4.3), in line with data reported in the literature.

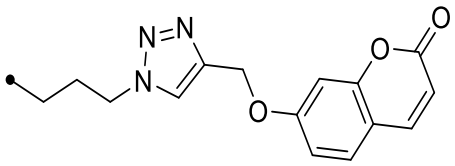
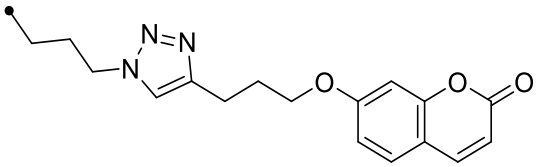
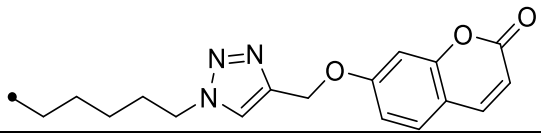
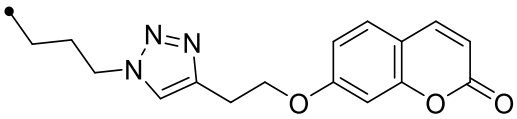
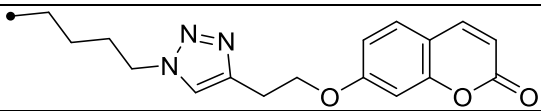
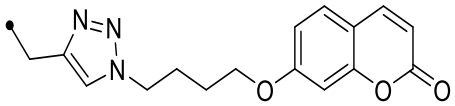
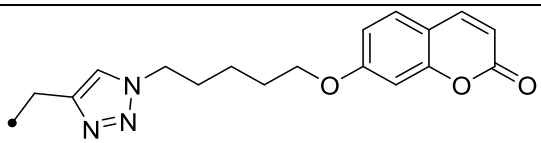
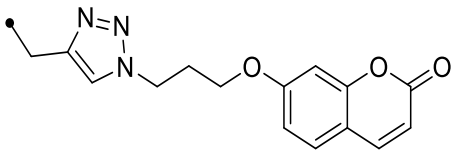
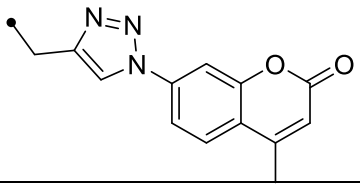
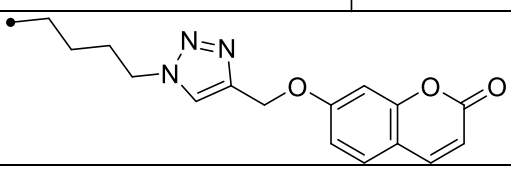
Overall, the synthesized hybrids showed a remarkable efficacy in inhibiting the tumor-associated carbonic anhydrases IX and XII *in vitro*, with a certain degree of selectivity over the housekeeping CA I and II isoforms.

Table 4.2. Inhibition properties of compounds from series PA, towards human CA I, II, IX, and XII isoforms.

Compound	Linker + CAI portion	K_i (nM)[a]			
		CA I	CA II	CA IX	CA XII
PA-59		829.1	22.9	2.7	6.1
PA-67		3670.2	90.1	3.9	24.8
PA-68		75.1	21.5	17.8	35.9
PA-77		580.5	59.4	1.4	7.3
PA-81		872.1	66.8	0.7	14
PA-84		940.2	6.0	0.7	4.4
PA-85		805	62.3	0.6	31.2
PA-86		592.4	55.1	0.5	25.5
PA-87		915.6	71.5	0.7	7.2
PA-88		573.6	15.8	0.2	18.3
PA-89		219.5	41.1	0.5	2.9

[a] Mean of three independent assays (errors were in the range of ± 5–10% of the reported values). Inhibitor-enzyme incubation time: 15 min for sulfonamides; 6 h for coumarins.

Table 4.3. Inhibition properties of compounds from series AL, towards human CA I, II, IX, and XII isoforms.

Compound	Linker + CAI portion	K_i (nM)[a]			
		CA I	CA II	CA IX	CA XII
AL-19		>10000	>10000	66.0	30.1
AL-36		>10000	>10000	58.9	28.4
AL-38		>10000	>10000	48.6	20.1
AL-39		>10000	>10000	34.5	18.9
AL-42		>10000	>10000	30.1	16.7
AL-47		>10000	>10000	25.5	4.2
AL-49		>10000	>10000	9.6	14.5
AL-60		>10000	>10000	35.1	10.2
AL-64		>10000	>10000	51.5	5.4
AL-96		>10000	>10000	17.3	8.4

[a] Mean of three independent assays (errors were in the range of ± 5–10% of the reported values). Inhibitor-enzyme incubation time: 15 min for sulfonamides; 6 h for coumarins.

Evaluation of compound-induced G4 stabilization and binding *in vitro*. The 21 synthesized hybrids were also screened for their ability to bind to and stabilize G4-forming DNA sequences, with respect to **Berberine** alone. Particularly, we employed the G-rich motifs from the promoter regions of c-MYC (*c-Myc G4*) and c-KIT (*c-Kit1 G4*) oncogenes, as well as the G-rich 23-mer truncation of the human telomeric DNA sequence (Tel_{23} *G4*). A 27-mer hairpin-duplex-forming sequence ($Hairpin_{27}$) was also used to evaluate the selectivity of the compounds for noncanonical DNAs over double-stranded DNA.

The overall topology adopted by each investigated DNA sequence, in K^+ buffer, was first verified by CD measurements. The *c-Myc* and *c-Kit1* G4-forming sequences adopted a parallel G4 conformation, displaying the distinctive positive band at around 260 nm and negative band at ~240 nm in their CD spectra. On the other hand, the Tel_{23} *G4* showed a CD spectrum having two positive bands at 290 and 270 nm and a weak negative band at 240 nm, consistent with a hybrid [3+1] G4 conformation. Finally, the $Hairpin_{27}$ showed a positive band at around 280 nm and a negative one at ~250 nm, confirming duplex formation. Additional CD spectra were acquired to explore the potential of the 21 synthesized hybrids (and **Berberine**) to alter the native folding topology of the investigated DNAs. No relevant variations of the DNA chiroptical signals were detected, suggesting an overall preservation of the DNA structure in the presence of the compounds. Afterwards, the G4-stabilizing properties of the molecules were assessed by CD melting experiments, measuring the $\Delta T_{1/2}$ values either of G4 or duplex structures (Table 4.4). According to the results, the hybrids turned out to be more effective than **Berberine** in stabilizing the G4s under investigation, with the strongest effect being on the *c-Kit1* and *c-Myc G4*, indicating that hybridization of **Berberine** with CAI portions could increase its G4-stabilizing properties, especially in the case of parallel G4 conformations. This might be ascribed to the hydrogen-bond acceptor and donor properties of the 1,2,3-triazole linker and/or to the aromatic character of both the linker and the CAI portion, which would guarantee further interactions with the G4 structures. Notably, $Hairpin_{27}$ thermal stability was not significantly affected by any of the compounds, except for **PA-68** ($\Delta T_{1/2}$ = 7.0 °C), possibly due to the lack of an aromatic moiety (in its CAI portion) which would favor a selective binding to the G-tetrads via π-π stacking interactions.

To get insight into the binding affinity of the compounds towards the examined G-quadruplexes and confirm their selectivity for G4s over duplex DNA, fluorescent intercalator displacement (FID) studies, using thiazole orange as a probe, were also performed. As shown in Table 4.5, all hybrids exhibited high selectivity for G4s over double-stranded DNA (not being able to

reach the 50% displacement of thiazole orange, on *Hairpin*$_{27}$), except for **PA-59** and **AL-64**, which, in the case of the duplex model, showed DC_{50} values of 4.7 and 4.9 μM, respectively. Of note, **PA-67**, **PA-77, PA-81**, and **PA-86 (**from the sulfonamide series), and **AL-38, AL-49**, and **AL-64** (belonging to the coumarin series), proved to be characterized by the strongest affinity to the *c-Myc G4*, with DC_{50} values in the range 0.9–1.5 μM.

Table 4.4. Compound-induced changes in the thermal stability of the investigated oligonucleotides, measured by circular dichroism melting experiments. $\Delta T_{1/2}$ values are the differences in the apparent melting temperatures of DNA in the presence and absence of compounds.

Compound	$\Delta T_{1/2}$ (°C) [1]			
	***Hairpin*$_{27}$**	***c-Kit1 G4***	***c-Myc G4***	***Tel*$_{23}$ *G4***
Berberine	0.6	15.5	4.7	10.2
PA-59	2.7	19.2	5.9	5.4
PA-67	−0.4	20.2	5.4	5.3
PA-68	7.0	18.7	5.3	6.6
PA-77	−0.3	27.2	9.3	9.8
PA-81	−3.0	20.9	5.3	5.9
PA-84	0.1	24.6	6.5	7.7
PA-85	−1.0	19.2	5.5	8.2
PA-86	−1.5	26.0	7.3	9.4
PA-87	0.8	19.0	6.9	7.8
PA-88	1.7	24.8	8.2	12.0
PA-89	0.9	22.2	7.5	10.3
AL-19	4.3	24.3	10.8	10.8
AL-36	1.5	22.2	7.2	12.9
AL-38	0.1	24.7	8.9	9.1
AL-39	3.2	23.0	7.3	7.7
AL-42	−2.4	20.9	7.6	7.3
AL-47	−1.3	20.2	7.5	7.3
AL-49	1.3	21.3	6.9	7.0
AL-60	1.5	20.7	7.8	10.4
AL-64	3.3	21.9	6.0	7.4
AL-96	1.3	22.2	5.7	9.2

[1] The error on the $\Delta T_{1/2}$ values is ± 1.0 °C. All experiments were performed in duplicate, and the values reported in the table are the average of two measurements.

Table 4.5. Compound concentrations required to give 50% thiazole orange displacement (DC_{50} values) from the DNA.

Compound	DC_{50} (μM)			
	Hairpin$_{27}$	***c-Kit1 G4***	***c-Myc G4***	***Tel$_{23}$ G4***
Berberine	n.d.[2]	n.d.[2]	n.d.[2]	n.d.[2]
PA-59	4.7	2.8	1.7	3.2
PA-67	n.d.[2]	3.8	1.4	4.0
PA-68	n.d.[2]	n.d.[2]	4.3	n.d.[2]
PA-77	n.d.[2]	n.d.[2]	0.9	3.4
PA-81	n.d.[2]	n.d.[2]	1.2	2.3
PA-84	n.d.[2]	n.d.[2]	2.9	n.d.[2]
PA-85	n.d.[2]	n.d.[2]	1.9	4.0
PA-86	n.d.[2]	n.d.[2]	1.0	3.0
PA-87	n.d.[2]	n.d.[2]	2.2	n.d.[2]
PA-88	n.d.[2]	n.d.[2]	2.2	n.d.[2]
PA-89	n.d.[2]	n.d.[2]	1.7	n.d.[2]
AL-19	n.d.[2]	n.d.[2]	1.8	n.d.[2]
AL-36	n.d.[2]	n.d.[2]	1.9	n.d.[2]
AL-38	n.d.[2]	n.d.[2]	1.5	n.d.[2]
AL-39	n.d.[2]	n.d.[2]	1.9	n.d.[2]
AL-42	n.d.[2]	n.d.[2]	2.3	n.d.[2]
AL-47	n.d.[2]	n.d.[2]	2.3	n.d.[2]
AL-49	n.d.[2]	n.d.[2]	1.5	n.d.[2]
AL-60	n.d.[2]	n.d.[2]	1.8	n.d.[2]
AL-64	4.9	2.4	1.2	2.4
AL-96	n.d.[2]	n.d.[2]	3.6	n.d.[2]

[2] n.d. = not determinable. The error on the DC_{50} values is ± 0.3 μM. All experiments were performed in duplicate, and the values reported in the table are the average of two measurements.

Evaluation of the effects on cell proliferation and viability. Combining our data on CA inhibition, G4 stabilization and binding affinity, compounds **PA-77**, **PA-86**, and **AL-38** were chosen as the most promising hybrids to be further investigated. Particularly, the effect of the selected molecules on the proliferation and viability of CA IX-positive cancer cells (HeLa) was determined. For this purpose, two standard methods were employed: the trypan blue exclusion test and the MTT assay.

Trypan blue exclusion test. Cells were treated with increasing concentrations of compounds (Table 4.6), including **Berberine** alone, **PA-80**, and **AL-15**, with the last two being the CAI portions of **PA-77**/**86** and **AL-38**, respectively (Figure 4.10). The experimental protocol described in Section 2.4 was then applied.

Table 4.6. Range of concentrations used for each compound.

Compound	Range of concentrations (μM)
PA-77	10, 25, 50, 100, 200
PA-86	2.5, 5, 10, 25, 50
AL-38	1, 10, 30, 40, 50
Berberine	50, 100, 150, 200, 250
PA-80	10, 25, 50, 100, 200
AL-15	1, 10, 30, 50, 100

As shown in Figures 4.11 and 4.12, the three CAI-**Berberine** hybrids (capable of both inhibiting the CA IX/XII isoforms and stabilizing G4 structures *in vitro*) produced a significant time- and concentration-dependent inhibition of cell growth, higher than that of **Berberine**, **PA-80** or **AL-15**, with IC_{50} values (compound concentrations required to give 50% of cell growth inhibition) in the range 18–62.9 μM and 4.5–20.9 μM after 24 h and 48 h of treatment, respectively. All calculated IC_{50} values are listed in Table 4.7.

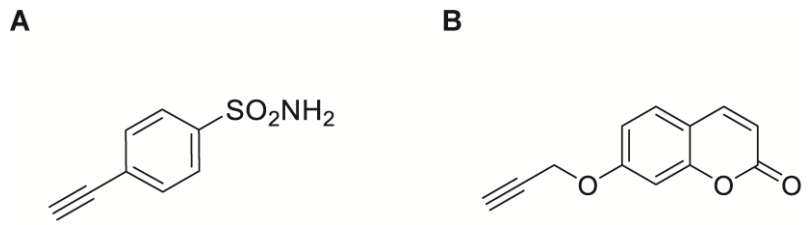


Figure 4.10. Chemical structures of (**A**) **PA-80** and (**B**) **AL-15**, as CAI portions of **PA-77**/**86** and **AL-38**, respectively.

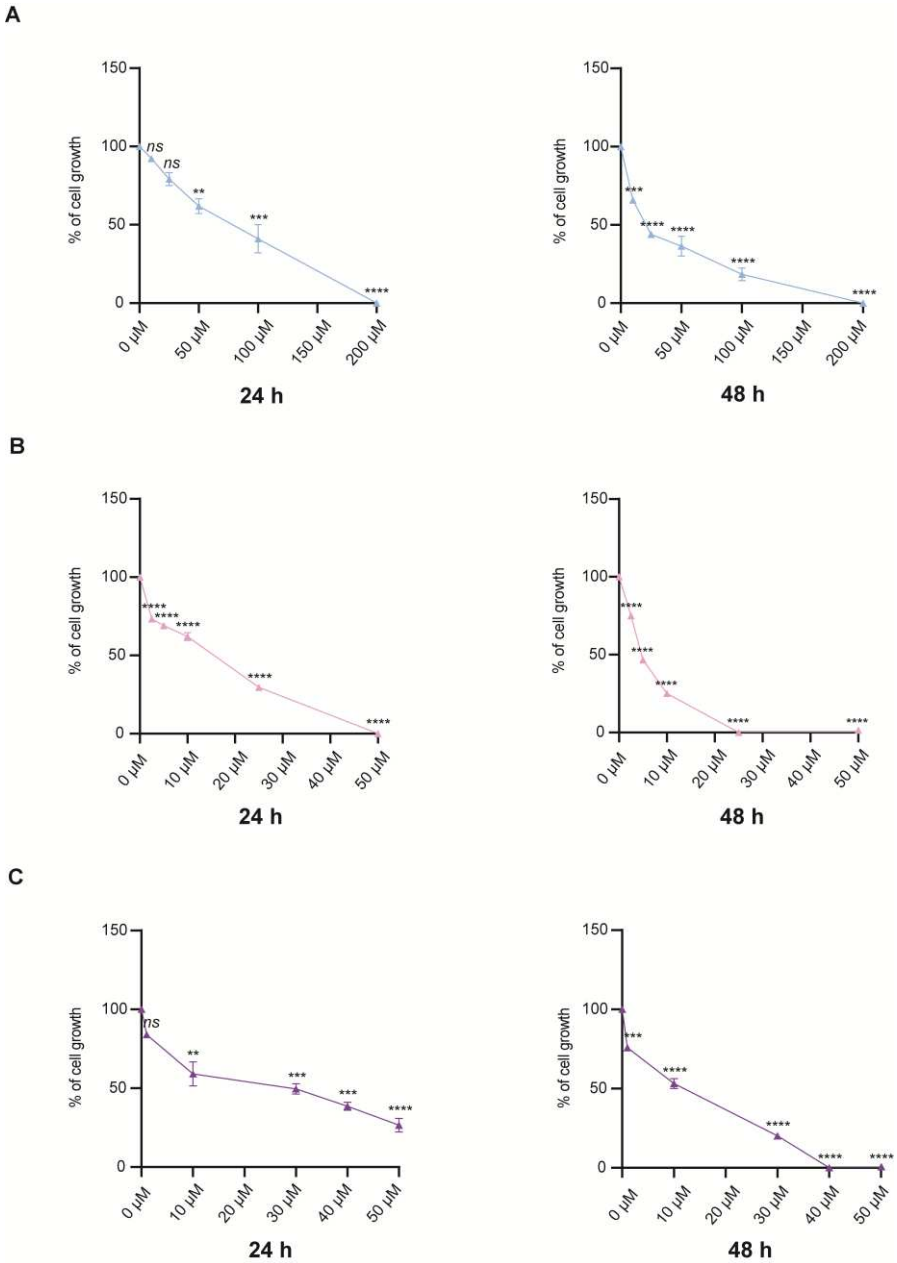


Figure 4.11. Growth curves of HeLa cells treated with (**A**) **PA-77**, (**B**) **PA-86**, and (**C**) **AL-38** for 24 h and 48 h. Each curve shows the mean ± SEM of three independent experiments. Differences in mean values between each treatment and the control group were determined by a one-way analysis of variance (ANOVA) on Prism 8.0.2 (ns = not significant, **p<0.01, ***p<0.001, ****p<0.0001).

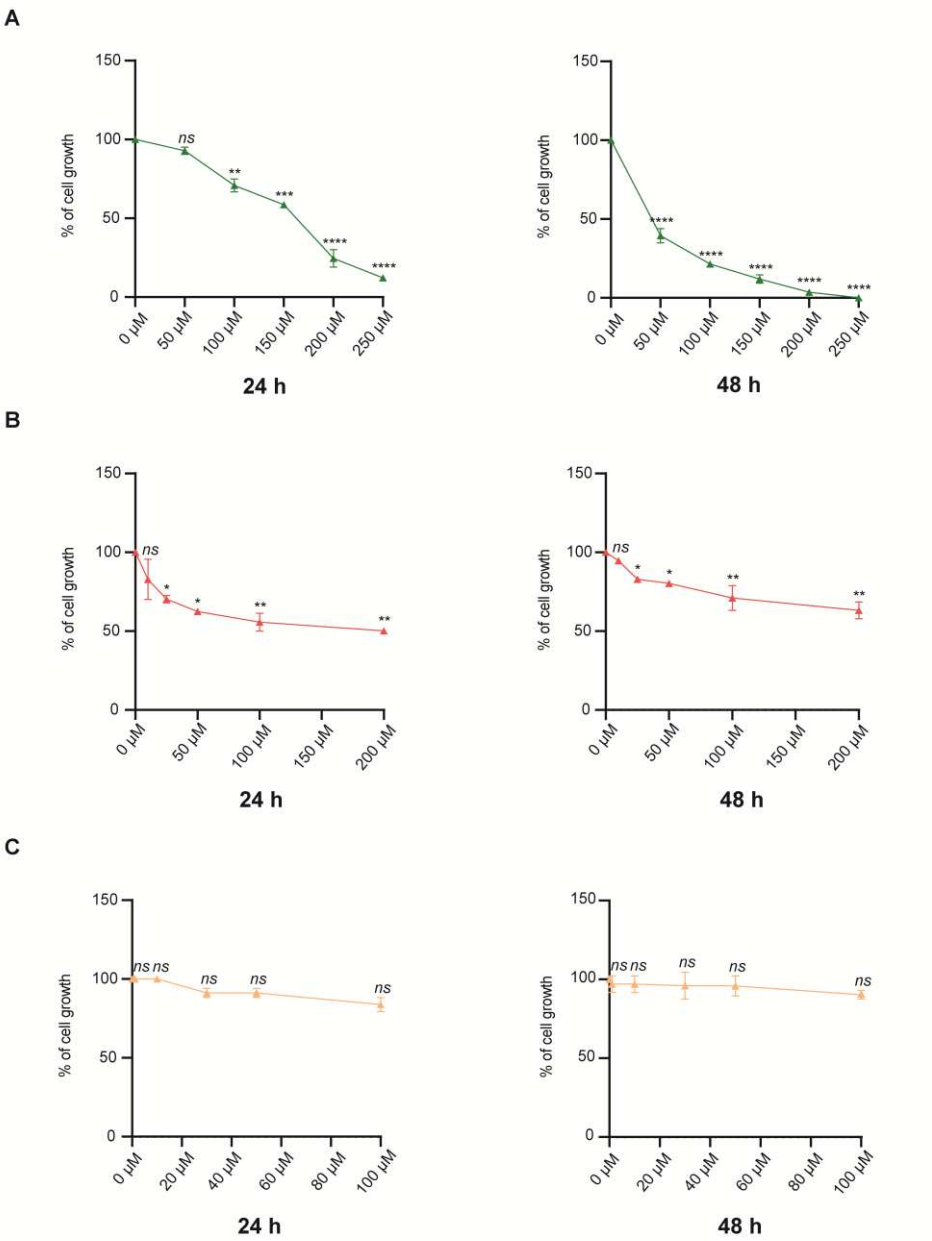


Figure 4.12. Growth curves of HeLa cells treated with (**A**) **Berberine**, (**B**) **PA-80**, and (**C**) **AL-15** for 24 h and 48 h. Each curve shows the mean ± SEM of three independent experiments. Differences in mean values between each treatment and the control group were determined by ANOVA on Prism 8.0.2 (ns = not significant, **p*<0.05, ***p*<0.01, ****p*<0.001, *****p*<0.0001).

Table 4.7. Compound concentrations required to give 50% of cell growth inhibition (IC_{50} values), with the trypan blue exclusion test.

Time	IC_{50} (µM)					
	PA-77	**PA-86**	**AL-38**	**Berberine**	**PA-80**	**AL-15**
24 h	62.9	21.5	18.0	132.0	> 200	> 100
48 h	20.9	9.5	4.5	< 50	> 200	> 100

MTT assay. HeLa cells were treated with increasing concentrations of the investigated compounds (Table 4.6) and processed by following the experimental protocol reported in Section 2.5. Results in Figures 4.13, 4.14 and in Table 4.8 show that the hybrids had a higher efficacy in reducing cellular viability than both **Berberine** and the CAI portions alone (in line with the results arising from the trypan blue exclusion method), except for **PA-77**, whose IC_{50} values fell outside the explored range of concentrations. Of note, the IC_{50} values obtained using the trypan blue dye were globally lower than those deriving from the MTT assay, due to the different sensitivity provided by the two methods.

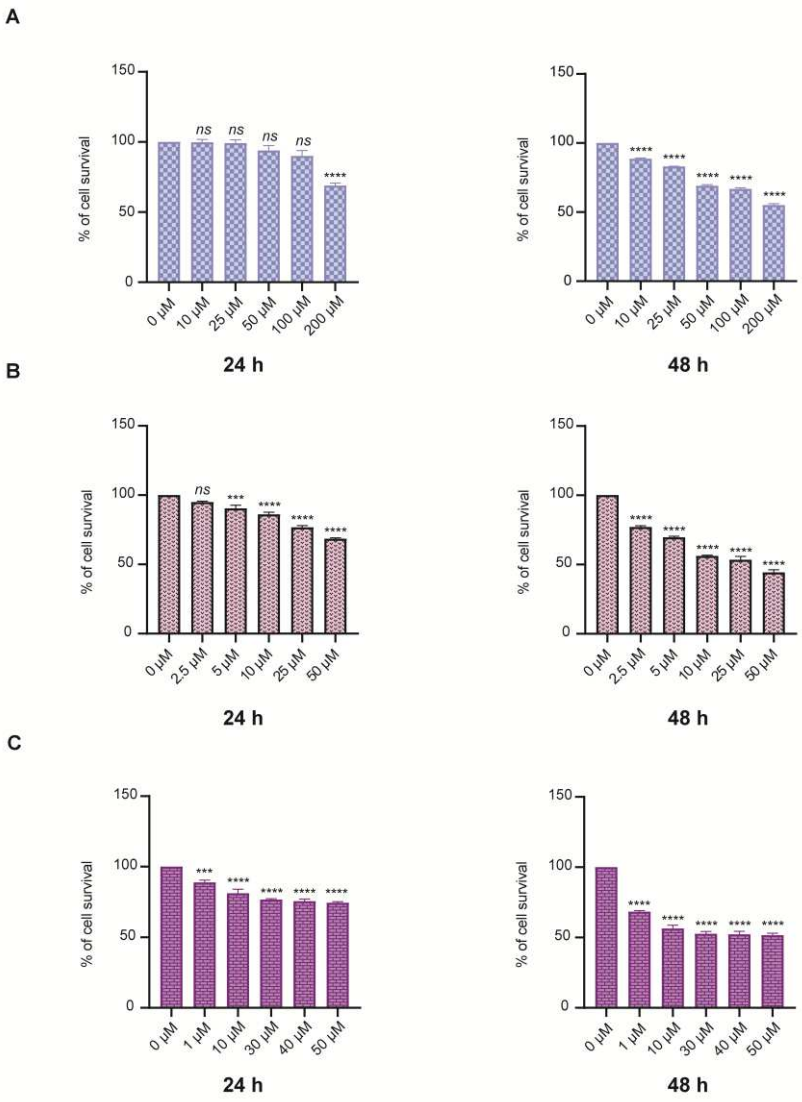


Figure 4.13. Percentage of survival of HeLa cells treated with (**A**) **PA-77**, (**B**) **PA-86**, and (**C**) **AL-38** for 24 h and 48 h. Histograms show the mean ± SEM of three independent experiments. Differences in mean values between each treatment and the control group were determined by ANOVA on Prism 8.0.2 (ns = not significant, ***p<0.001, ****p<0.0001).

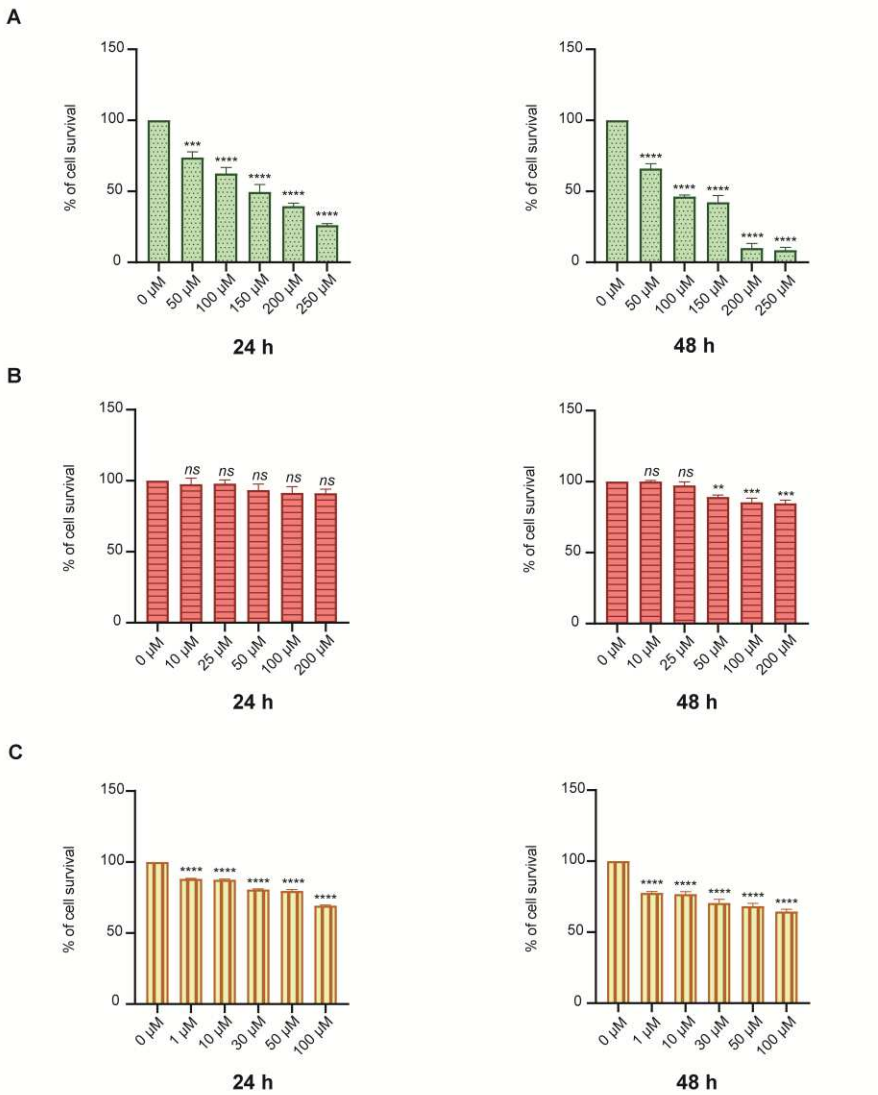


Figure 4.14. Percentage of survival of HeLa cells treated with (**A**) **Berberine**, (**B**) **PA-80**, and (**C**) **AL-15** for 24 h and 48 h. Histograms show the mean ± SEM of three independent experiments. Differences in mean values between each treatment and the control group were determined by ANOVA on Prism 8.0.2 (ns = not significant, **p<0.01, ***p<0.001, ****p<0.0001).

Table 4.8. Compound concentrations required to give 50% of cell viability reduction (IC_{50} values), with the MTT assay.

Time	IC_{50} (µM)					
	PA-77	**PA-86**	**AL-38**	**Berberine**	**PA-80**	**AL-15**
24 h	> 200	> 50	> 50	133.8	> 200	> 100
48 h	> 200	20.3	33.0	70.3	> 200	> 100

4.2.3 *Conclusions and ongoing experiments*

In the present study, we rationally designed and synthesized a library of molecular hybrids containing a coumarin/sulfonamide CAI portion and a **Berberine** core, bound together through a 1,2,3-triazole linker of varying length. The ability of the newly synthesized compounds to both inhibit the tumor-associated carbonic anhydrases IX/XII and bind to and stabilize G4 structures *in vitro* was then assessed. The most promising molecules (**PA-77**, **PA-86**, and **AL-38**) were subjected to a further biological characterization, displaying a good cytotoxic effect on CA IX-positive human cervix cancer cells. Additional experiments are still ongoing in our laboratory, including co-treatment of HeLa cells with equimolar concentrations of **Berberine** and **PA-80**/**AL-15**, to figure out whether the cytotoxicity of the hybrids is more pronounced than that resulting from the co-treatment with the single molecular entities. If that is the case, it would be potentially beneficial, as already said, in terms of pharmacokinetics or treatment adherence. Moreover, immunofluorescence experiments are being performed to assess the ability of the hybrids to affect the formation of G4 structures in cells. Overall, this study might pave the way towards the use of CAI-**Berberine** hybrids as new effective multi-target agents against hypoxic tumors.

Chapter 5
TARGETING OF NONCANONICAL NUCLEIC ACID STRUCTURES FOR BIOLOGICAL INVESTIGATION

As discussed in Chapter 1, various hypotheses have been postulated and proposed about the general biological roles of G-quadruplex structures. Nevertheless, selective tools able to unravel the specific functions of individual G4s still need to be developed. In this Chapter, a short peptide that preferentially binds (with nanomolar affinity) to the c-Myc G4 is reported (Paper V), paving the way towards the design of a Myc-selective probe to unveil the biological function of such peculiar G4. This study was carried out at Imperial College London (UK), under the supervision of Dr. Marco Di Antonio.

Furthermore, as already mentioned, the way G4 and iM structures reciprocally influence each other, and which biological implications are regulated by such a delicate balance are still open questions under intense investigation. Thus, the need for new experimental approaches able to disentangle this scientific dilemma is urgent. In the present Chapter, new compounds are reported (Paper VI) that might be used as tools to clarify the biological roles of G4 and iM structures and their relationship.

5.1 **A short peptide that preferentially binds c-MYC G-quadruplex DNA** (Paper V)

5.1.1 *Introduction*

Small molecule-based targeting approaches have been extensively used to explore G4 functions (Burger et al., 2005; Jain & Bhattacharya, 2011; Luedtke, 2009; Monchaud & Teulade-Fichou, 2008; Yaku et al., 2012), mainly relying on π-π stacking interactions of the molecular probes with the top (or bottom) end of G4 structures. This strategy is highly effective in achieving selectivity for G4s over duplex DNA, whereas it does not allow selectivity towards an intended G4 or a small set of G4s, due to the promiscuous presence of accessible π-stacking surfaces across different G4 structures. For the detailed investigation of the features and functions of individual G4s in the genome, probes showing specificity to certain G4s over others are required.

In the present work, taking inspiration from the recently reported crystal structure of the bovine DHX36 helicase bound to the G4 structure formed in the promoter region of the proto-oncogene c-MYC (M. C. Chen et al., 2018), we identified a short peptide to potentially probe, with selectivity, the c-Myc G4 structure in the human genome.

5.1.2 *Results and discussion*

We started our investigation by using PyMOL to analyze the crystal structure of the bovine DHX36 helicase bound to the c-Myc G4 (PDB ID: 5VHE) and identified a 22-amino acid sequence (Figure 5.1) as the minimal Myc-binding domain of the protein.

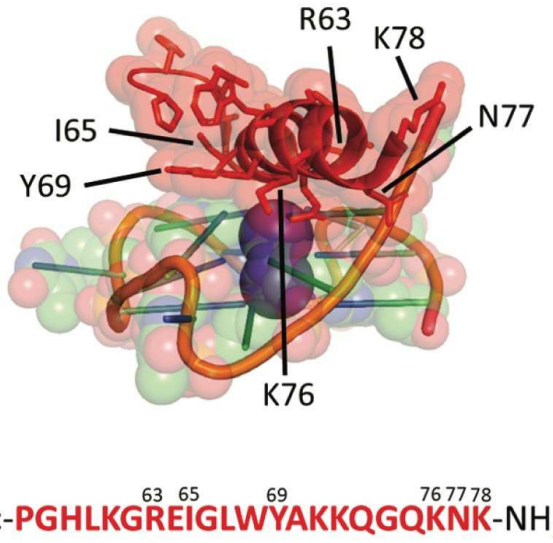


Figure 5.1. (Top) Minimal peptide binding domain of DHX36 bound to the Myc DNA G-quadruplex. (Bottom) The 22-amino acid sequence (**DM039**) used in this study. Figure from (Minard et al., 2020).

The 22-residue sequence (**DM039**) was then synthesized using previously established Fmoc/^{t}Bu solid-phase peptide synthesis (SPPS) protocols and acetylated at the N-terminal before being purified by high-performance liquid chromatography (HPLC) and characterized by mass spectrometry (MS). We next sought to evaluate the binding affinity of **DM039** towards a small panel of G4-forming sequences as well as single- and double-stranded DNA controls. Small-molecule ligands that bind to G4s are typically tested by melting experiments (FRET melting) for their ability to thermally stabilize G4 structures (De Cian et al., 2007). However, melting experiments are difficult to apply to peptides that are temperature sensitive. Also, even though stabilization properties are correlated, on some level, to the binding affinity of the ligands, they cannot be considered a direct affinity measurement. Thus, we decided to exploit

the higher molecular weight of the 22-amino acid peptide, compared to small molecules, to develop a fluorescence polarization (FP) assay to reliably measure G4 binding (Moerke, 2009). According to the FP results reported in Table 5.1, **DM039** showed a preferential binding (with an observed dissociation constant (K_D) of 112 nM [59–196 nM — 95% CI]) to the parallel c-Myc G4, over a small set of parallel and non-parallel G4s, single- and double-stranded DNA controls, but it was still able to bind to the Bcl-2 G4 (even if with lower affinity and a K_D value of 579 nM [447–749 nM — 95% CI]), possibly due to an unusually high accessibility of the G-tetrads in this mixed-type G4 structure.

Table 5.1. Results of the fluorescence polarization experiments.

Oligonucleotide	G4 topology	K_D (nM)
c-Myc	Parallel	112
ssDNA control	-	-
dsDNA control	-	-
c-Kit1	Parallel	-
c-Kit2	Parallel	-
HRAS	Antiparallel	-
hTelo	Hybrid	-
Bcl-2	Hybrid	579

The negligible binding to both single- and double-stranded control sequences suggested that G4 specific interactions were responsible for the observed binding of **DM039** to the Myc G4. In order to further test this hypothesis, we measured the peptide affinity to the targeted G4 when annealed in Li^+ buffer (Figure 5.2). The resulting K_D was equal to 1226 nM [491–3469 nM — 95% CI], proving a ~10-fold reduction in the affinity with respect to the binding measured in K^+ buffer and confirming that the binding to the G4 was dependent on the G4 folding status.

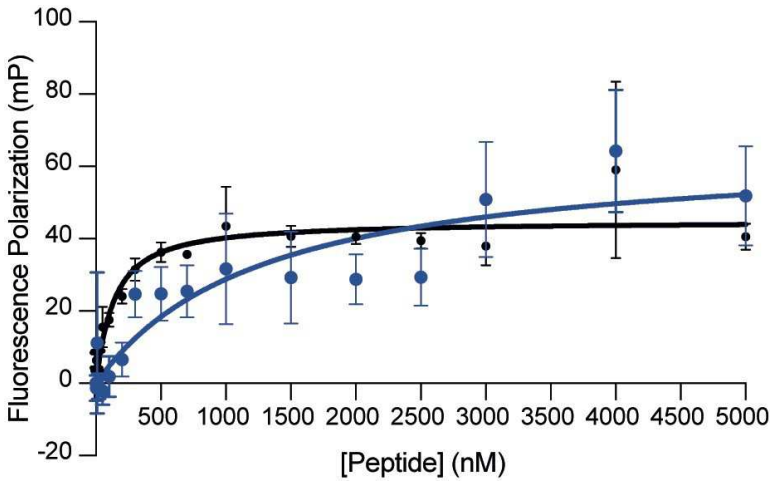


Figure 5.2. Results of the FP assay performed with **DM039** on the Myc G4, in Li^+ (blue) and K^+ (black) buffers. Each measurement was repeated three times. Figure from (Minard et al., 2020).

Strikingly, surface plasmon resonance experiments provided a measured K_D of 123 ± 23 nM, in excellent agreement with the value obtained by FP, confirming the high binding affinity of **DM039** to the G4 in the c-MYC gene promoter and proving the suitability of our FP assay to measure binding of short peptides to G4 structures.

Then, we moved one step forward and given that the sequence constituting **DM039** was folded into an α-helix in the crystal structure of the bovine DHX36 helicase bound to the Myc G4, we acquired CD spectra of **DM039** in the absence and presence of the Myc G4 to test whether the peptide also adopted a helical conformation when bound to the G4. The results revealed that **DM039** was mostly unstructured in solution, while an α-helical CD signal became detectable upon binding of the peptide to the Myc G4, displaying an increase in the calculated helical fraction (fH) from 0.073 (measured with **DM039** in solution) to 0.155 (observed upon G4 binding). Therefore, we wanted to investigate whether constraining the peptide into a helical conformation could improve its binding affinity to the targeted G4. To this aim, we designed and synthesized two hydrocarbon-stapled peptide analogues of **DM039**, **DM083** (*i*, *i*+4 stapling) and **DM102** (*i*, *i*+7 stapling) (Figure 5.3A and 5.3B). Once having the two cyclized peptides in our hands, we first validated, by CD, that stapling increased the helicity of the peptides, observing a fH of 0.18 and 0.22 for **DM083** and **DM102**, respectively. Crucially, the fH values of both **DM083** and **DM102** did not increase in the presence of the G4. The two stapled peptides were then assessed for binding to the Myc G4 by means of our FP assay (Figure 5.3C). Particularly, *i*, *i*+4 stapling in **DM083** reduced the affinity to the targeted G4 of ~5 fold (K_D = 548 nM [317–934 nM — 95% CI]), while the *i*, *i*+7 stapled peptide **DM102** still retained

substantial binding affinity to the Myc G4, with a reduction of around 3 fold compared to **DM039** (K_D = 332 nM [249–435 nM — 95% CI]). This was in agreement with a report revealing how lactam stapling of a similar peptide sequence overall retained G4 binding ability (Yaneva et al., 2020).

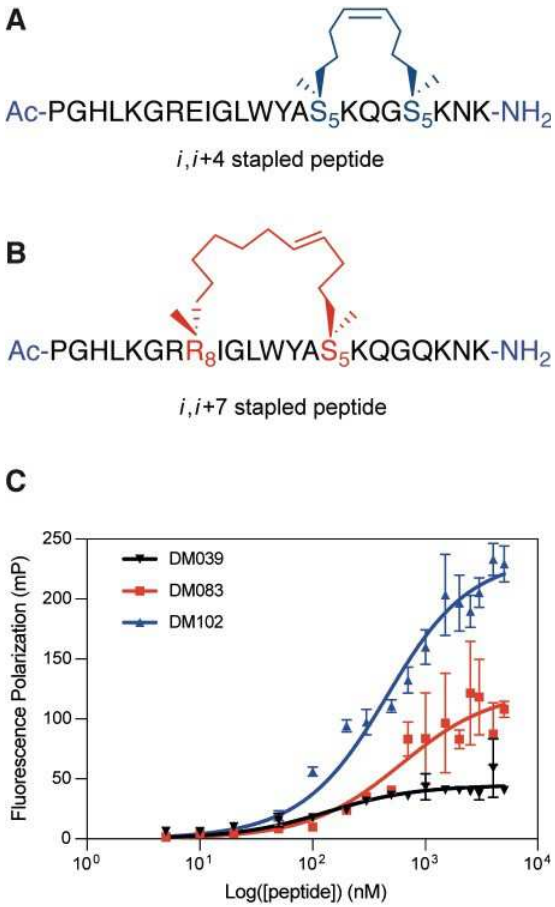


Figure 5.3. (**A**) *i*, *i*+4 hydrocarbon stapled peptide **DM083**, (**B**) *i*, *i*+7 hydrocarbon stapled peptide **DM102**. (**C**) Differential binding behavior against the Myc G4 observed for **DM039** and its two stapled versions (**DM083** and **DM102**) by means of FP measurements. Each measurement was repeated three times. Figure from (Minard et al., 2020).

Finally, with the aim of evaluating whether chemically locking of the peptide into a helical conformation resulted in retention of preferential selectivity for the Myc G4, we employed FP to assess **DM102** binding to the same panel of oligonucleotides used for **DM039** and, surprisingly, it turned out to be able to bind to all the DNAs tested, yielding observed K_D values of 574 nM [381–855 nM — 95% CI] for c-Kit1, 1081 nM [847–1383 nM — 95% CI] for hTelo, 831 nM [622–1112 nM — 95% CI] for Bcl-2, 1238 nM [827–1878 nM — 95% CI] for HRAS, 2628 nM [1648–4497 nM — 95% CI] for single-stranded DNA, and 2377 nM [1440–4222 nM

— 95% CI] for double-stranded DNA. This observation suggested that **DM039** might bind to the Myc G4 through an induced fit mechanism and a good degree of flexibility might be required for both high affinity interaction and selectivity, as chemical stapling of the sequence in **DM102** rendered the peptide a non-specific DNA binder.

5.1.3 *Conclusions*

Herein, we extracted a short peptide sequence (22 residues) from the crystal structure of the bovine DHX36 helicase bound to the c-Myc G4, synthesized it and developed a fluorescence polarization assay to measure its binding affinity to G4 structures. The peptide, **DM039**, displayed negligible binding to some parallel G4 structures (c-Kit1 and c-Kit2), binding to a mixed type G4 (Bcl-2), and a preferential binding (with nanomolar affinity) to the c-Myc G4. Furthermore, we observed that hydrocarbon stapling of the peptide reduced c-Myc binding affinity and abrogated both intra-G4 selectivity and selectivity over single- and double-stranded DNA, suggesting that a good degree of flexibility might be required for both high affinity and specificity to the targeted G4, which was not intuitive based on the helical conformation adopted by the peptide in the crystal structure.

Finally, we anticipate that further studies on **DM039** will be key to develop a Myc-selective probe and disentangle the biological role of such G4 over the ~700 000 mapped across the human genome. Our findings might pave the way towards the rational design of peptide-based molecular probes for selective targeting of individual G4s.

5.2 **Synthesis and characterization of bis-triazolyl-pyridine derivatives as noncanonical DNA-interacting compounds** (Paper VI)

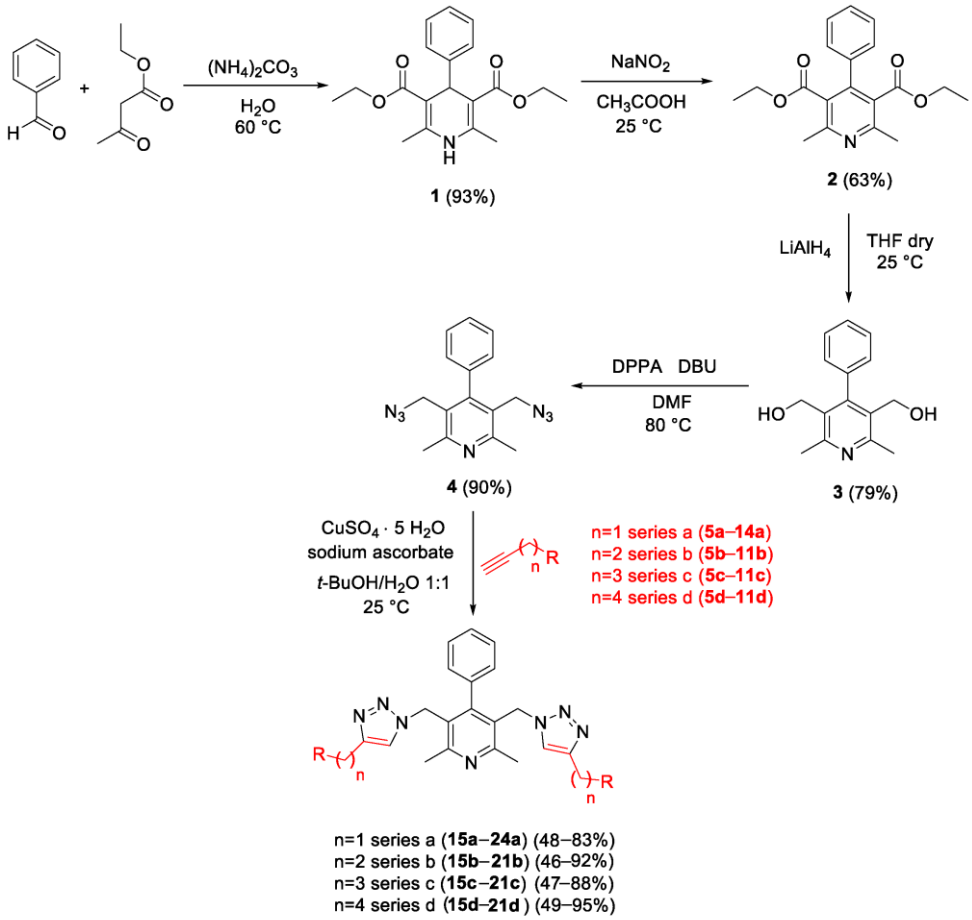
5.2.1 *Introduction*

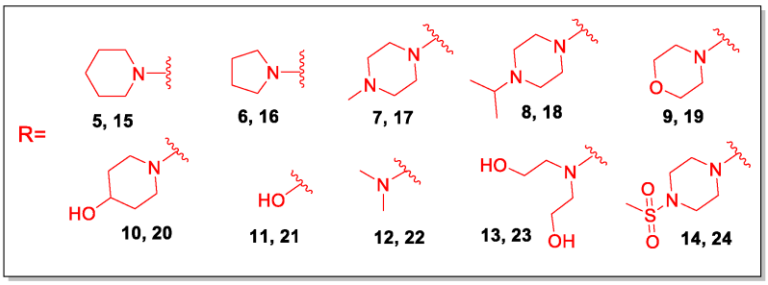
The scientific community has made great efforts in finding small molecules able to recognize G4 and iM structures to fully decipher their specific biological functions. To date, a large number of compounds that stabilize G4 structures and display considerable selectivity for G4s over single- and double-stranded DNA have been developed, whereas the i-motif specific ligand library still needs to be expanded (Shu et al., 2018).

In the present study, we combined the Hantzsch multicomponent reaction (A. Hantzsch, 1881; Arthur Hantzsch, 1882) and the Copper(I)-catalyzed Huisgen 1,3-dipolar alkyne-azide cycloaddition (CuAAC) (Rostovtsev et al., 2002) to generate a library of 31 potential G4 and/or iM ligands, containing a central pyridine scaffold (to make π-π stacking interactions with the nucleobases) and cationic side chains (to interact with the DNA negatively charged phosphate groups, stabilizing the interaction with the nucleic acid). The synthesized bis-triazolyl-pyridines were preliminarily screened for their ability to interact with G4 and/or iM DNAs through circular dichroism and clustered based on their capability to affect the thermal stability of the investigated DNA sequences by employing PCA. The ability of the best-selected compounds to interact with noncanonical DNAs was further investigated by using NMR and fluorescence spectroscopies. Finally, in order to translate the biophysical data into a more complex environment, the capability of the selected compounds to affect the formation of G4/iM structures in U2OS cells was assessed.

5.2.2 *Results and Discussion*

Synthesis. The library of 31 bis-triazolyl-pyridines (series a **15a**–**24a**, series b **15b**–**21b**, series c **15c**–**21c**, series d **15d**–**21d**) was prepared by following the synthetic route depicted in Scheme 1. The identity of the new compounds was established by their FT-IR, ^{1}H-NMR, ^{13}C-NMR, MS, and elemental analyses.





Scheme 1. Synthesis of the bis-triazolyl-pyridines **15a**–**24a**, **15b**–**21b**, **15c**–**21c**, and **15d**–**21d**.

CD studies. The 31 synthesized derivatives were preliminarily screened for their ability to interact with G4 and iM DNAs through CD spectroscopy (Pagano et al., 2012). Various G-rich sequences able to fold into parallel, antiparallel, and hybrid G4 structures and their iM-forming C-rich counterparts were employed: the G-/C-rich motifs from the promoter regions of BCL-2 (*Bcl-2 G4* and *Bcl-2 iM*) (Dai et al., 2006; Kendrick et al., 2009) and c-MYC (*c-Myc G4* and *c-Myc iM*) oncogenes (Ambrus et al., 2005; Dai et al., 2010), as well as the G-rich 23-mer (*Tel$_{23}$ G4)* and the C-rich 22-mer (*hTeloC iM*) truncations of the human telomeric DNA sequence (Dai et al., 2008; Pagano et al., 2015; Phan et al., 2000). Moreover, a 27-mer duplex-forming

sequence (*Hairpin*$_{27}$) was exploited to assess the selectivity of the compounds for noncanonical DNA structures over duplex DNA.

After having verified the overall topology adopted by each investigated DNA sequence in the proper buffer solution (potassium buffer, pH 7, for the G-rich and hairpin-forming sequences, and sodium buffer, pH 5, for the C-rich sequences), additional CD spectra were recorded to explore the ability of the 31 bis-triazolyl-pyridines to affect the native conformation of the DNAs themselves. Upon addition of an excess of compounds (10 molar equiv relative to the DNA), there were no significant variations in the DNA chiroptical signals, indicating an overall preservation of the DNA secondary structures in the presence of the compounds. Afterwards, CD melting experiments were performed and the $\Delta T_{1/2}$ values either of G4, iM, or duplex structures were measured to assess the DNA stabilizing/destabilizing properties of the synthesized derivatives (Table 5.2). The CD melting experiments resulted in a total of 217 $\Delta T_{1/2}$ values to be compared (31 compounds tested on 7 DNA sequences). In order to explore such a large dataset and to cluster the compounds based on their capability to affect the thermal stability of the investigated DNA sequences, the principal component analysis was employed. Particularly, given that most of the bis-triazolyl-pyridines destabilized the iM structures (negative $\Delta T_{1/2}$) while stabilizing the G4s, we decided to take into account the $\Delta T_{1/2}$ absolute values. Also, *Hairpin*$_{27}$ and *Tel*$_{23}$ *G4* were excluded from the analysis since their thermal stability was not significantly affected by any of the tested compounds ($|\Delta T_{1/2}| < 4$ °C). Figure 5.4 shows the scores and loadings plots generated by the PCA model.

Table 5.2. Compound-induced thermal stabilization or destabilization of the investigated oligonucleotides, measured by circular dichroism melting experiments.

Compound	$\Delta T_{1/2}$ (°C) [1]						
	Hairpin$_{27}$	***Bcl-2 G4***	***c-Myc G4***	***Tel$_{23}$ G4***	***Bcl-2 iM***	***c-Myc iM***	***hTeloC iM***
15a	0	5.0	7.0	1.1	−7.0	−18.4	−12.0
16a	−0.2	4.5	5.0	1.5	−11.0	−27.5	−25.0
17a	−1.0	4.0	5.0	0	−10.5	−27.0	−24.5
18a	0.2	4.5	7.5	3.5	−13.5	−24.5	−21.0
19a	1.7	0.5	2.0	1.0	−4.0	−5.5	−4.0
20a	3.9	0.5	1.0	0	−8.0	−10.4	−18.0
21a	2.9	0	0	0	0	−0.8	2.5
22a	−0.2	2.7	4.5	0.5	−11.0	−19.0	−25.5
23a	2.9	0	0.5	0.5	−6.0	−17.0	−12.5
24a	−0.7	1.5	2.0	−0.5	−2.0	−2.8	1.5
15b	−0.1	5.1	7.5	1.6	−6.0	−20.0	−26.5
16b	−0.1	4.5	8.0	1.6	−13.0	−19.4	−24.5
17b	−0.1	4.0	5.5	1.0	−10.0	−17.4	−23.5
18b	−0.1	5.5	8.5	1.5	−6.0	−11.4	−15.5
19b	0	3.0	2.0	1.2	−5.0	−4.0	−7.0
20b	2.2	5.9	4.0	0.1	−8.0	−6.0	−9.0
21b	1.9	1.0	1.0	2.0	0	−0.4	2.0
15c	−0.1	6.1	10.6	0	−11.0	−24.0	−29.0
16c	−0.1	5.6	9.5	0.5	−10.0	−17.4	−24.1
17c	−0.1	5.6	6.5	2.0	−12.0	−21.1	−27.4
18c	1.1	7.6	10.5	1.6	−10.5	−20.0	−25.5
19c	−1.1	5.1	4.2	−1.0	−3.0	−3.0	−10.0
20c	3.0	7.1	6.8	0	−9.0	−2.9	−9.0
21c	1.9	0	0	2.0	0	−0.9	1.5
15d	0	8.8	9.8	−1.0	−5.0	−4.9	−7.0
16d	−0.3	8.0	8.1	2.0	−5.0	−4.9	−7.0
17d	1.0	9.0	6.9	2.0	−8.0	−7.0	−8.9
18d	0	10.1	9.8	−1.0	−8.0	−8.0	−9.0
19d	0.9	4.2	2.8	−2.0	−5.0	−5.0	−7.0
20d	2.9	8.0	8.2	0	−8.0	−5.0	−9.0
21d	1.9	2.0	0.8	−0.8	−1.0	0	−1.0

[1] The error on $\Delta T_{1/2}$ values is ± 1.0 °C. All experiments were performed in duplicate, and the values reported in the table are the average of two measurements.

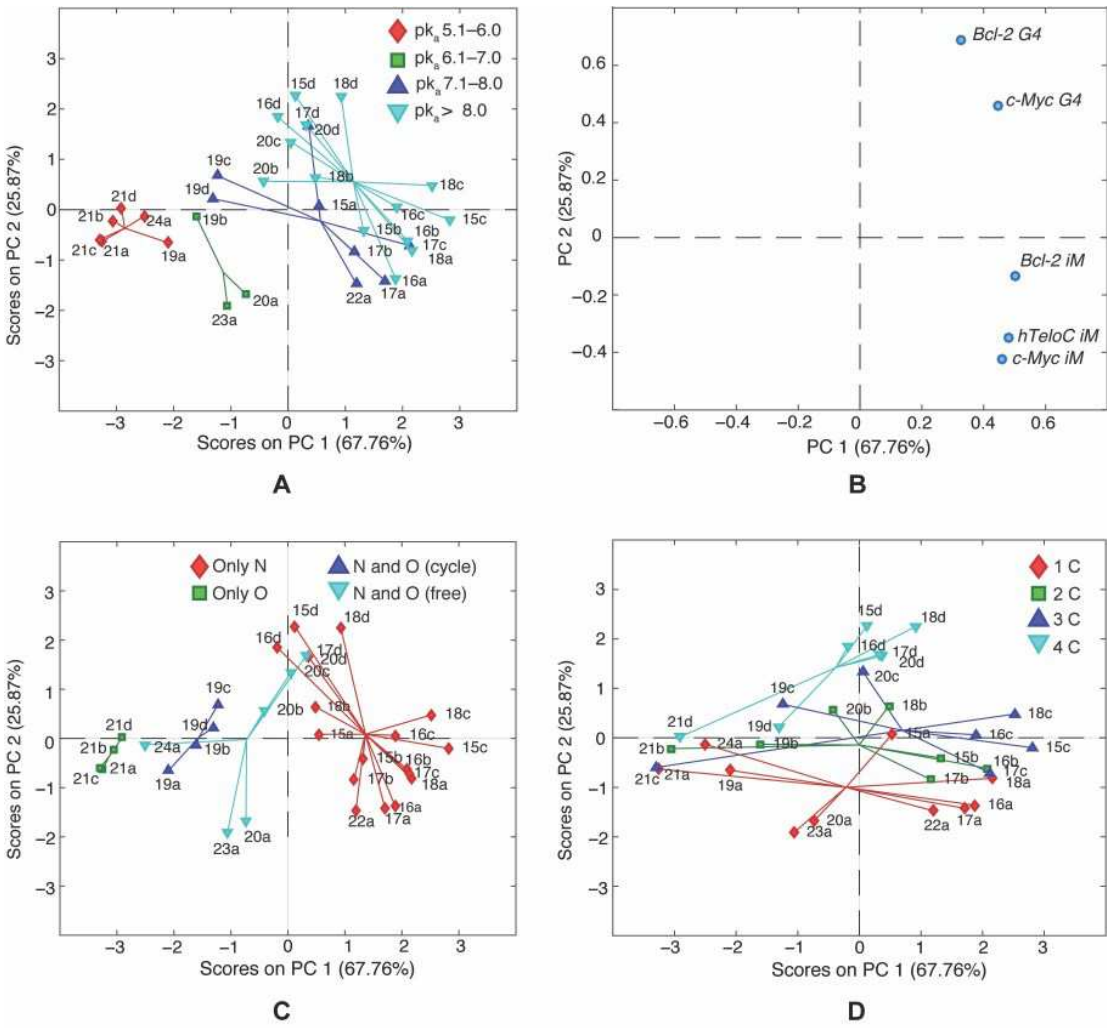


Figure 5.4. Output of the PCA model performed on the $\Delta T_{1/2}$ values dataset of the 31 compounds tested on *Bcl-2 G4*, *c-Myc G4*, *Bcl-2 iM*, *c-Myc iM,* and *hTeloC iM* DNAs. PC1/PC2 scores plot colored according to the (**A**) pK_a value of the conjugated acid of the strongest basic group in each molecule, (**C**) presence/absence of both nitrogen and oxygen atoms in the side chains, and (**D**) length of the side chain (from 1 up to 4 carbons); (**B**) PC1/PC2 loadings plot. Figure from (Di Porzio et al., 2021).

All the DNA sequences (variables) turned out to be clustered together on the very right side of the loadings plot (Figure 5.4B), meaning that the derivatives (samples) lying on the right side of the scores plot (Figure 5.4A) were characterized by higher $|\Delta T_{1/2}|$ values compared to the ones situated on the left side. Of note, by coloring each compound according to the pK_a value corresponding to the conjugate acid of the strongest basic group in the molecule (Figure 5.5 and Table 5.3 for details), we could observe a trend along PC1 (Figure 5.4A), going from the left (lower pK_a values) to the right (higher pK_a values), meaning that the compounds that least affected the thermal stability of the investigated DNA structures were the least likely to carry a

positive charge (under the experimental conditions), which is crucial for the interaction with the negatively charged DNA backbone (Amato et al., 2016, 2020).

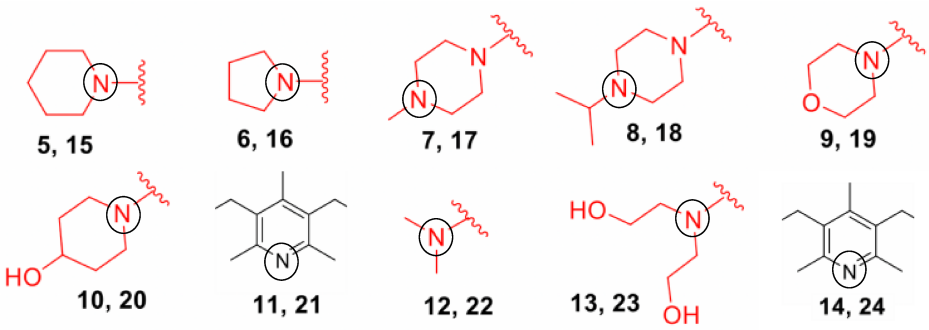


Figure 5.5. The strongest basic groups of the investigated compounds are circled in black. The pK_a values of their conjugated acids were used in the multivariate data analysis. Figure from (Di Porzio et al., 2021).

The samples were then colored according to the chemical characteristics of the side chain (R), with the aim of dissecting the structural features shared by the compounds able to interact with the DNA structures under investigation. For this purpose, the bis-triazolyl-pyridines were divided into four different classes where R (i) contained only nitrogen, (ii) only oxygen, (iii) both nitrogen and oxygen with the latter being closed in a cycle, and (iv) both nitrogen and oxygen with the latter being present as a free hydroxyl group. In doing so (Figure 5.4C), it became clear that both the lack of nitrogen and the presence of oxygen reduced the capability of the compounds to stabilize/destabilize the DNA structures. Indeed, the compounds that most affected the stability of the DNAs were those containing only nitrogen atoms in the R chain. Furthermore, by labeling the compounds in the scores plot according to the length of the R side chain, going from 1 carbon (series a) up to 4 carbons (series d), a clear separation along PC2 (Figure 5.4D) was detectable. Notably, the effect of the side chain length was observable only for the compounds that did affect the thermal stability of the investigated structures (right part of the plot) while it was not evident for the rest of the molecules (i.e., **21d**, for example). Particularly, compounds such as **20a** and **23a**, located in the lower part of the scores plot (where the iM-forming sequences are placed, in the corresponding area of the loadings plot) and characterized by a short carbon chain, altered the thermal stability of the iMs more significantly than that of G4 structures. On the contrary, compounds **15d**, **16d**, **17d**, **18d**, **20d**, and **20c**, located in the upper part of the plot (where the G4-forming sequences are placed, in the corresponding area of the loadings plot) and containing a longer carbon chain, preferentially

affected the thermal stability of the G4 over the iM structures. Interestingly, there were two compounds (**15c** and **18c**) that were located at the very right side of the PC1 axis, and thus characterized by high $\Delta T_{1/2}$ values, while being in the middle region of the PC2 axis, meaning that they were able to affect the thermal stability of both iMs and G4s. Specifically, they could stabilize the G4 structures while destabilizing the iM ones.

Table 5.3. pK_a values of the conjugated acid of the strongest base (see Figure 5.5) of the investigated compounds, with the corresponding pK_a interval.

Compound	pK_a value	pK_a interval
15a	7.5	7.1–8.0
16a	8.6	> 8.0
17a	7.7	7.1–8.0
18a	8.1	> 8.0
19a	5.8	5.1–6.0
20a	6.6	6.1–7.0
21a	5.7	5.1–6.0
22a	7.5	7.1–8.0
23a	6.1	6.1–7.0
24a	5.8	5.1–6.0
15b	9.0	> 8.0
16b	10.0	> 8.0
17b	7.9	7.1–8.0
18b	8.3	> 8.0
19b	6.8	6.1–7.0
20b	8.4	> 8.0
21b	5.7	5.1–6.0
15c	9.7	> 8.0
16c	10.5	> 8.0
17c	8.0	7.1–8.0
18c	8.4	> 8.0
19c	7.5	7.1–8.0
20c	8.8	> 8.0
21c	5.7	5.1–6.0
15d	10.0	> 8.0
16d	10.7	> 8.0
17d	8.0	7.1–8.0
18d	8.4	> 8.0
19d	7.7	7.1–8.0
20d	9.1	> 8.0
21d	5.8	5.1–6.0

NMR studies. To get information about the binding mode of **15c**, **18c**, **20a,** and **23a** to the *c-Myc G4* and the *hTeloC iM*, 1D ^{1}H-NMR experiments were performed. We focused on those

compounds since able to affect the thermal stability of both iM and G4 structures (**15c** and **18c**) or exclusively of the iMs (**20a** and **23a**). Compound **21d** was also employed as negative control. According to the literature and under the experimental conditions used in the present study, the *c-Myc G4* sequence forms a single G4 conformation characterized by 12 well-resolved imino proton peaks, corresponding to the 12 guanines involved in the three G-tetrad planes (Ambrus et al., 2005). For its part, the *hTeloC* sequence folds into an iM structure characterized by 3 well-resolved imino proton peaks that correspond to the 6 intercalated C-C^+ pairs (Phan et al., 2000). As expected, no relevant shifts of both the imino and aromatic signals of the *c-Myc G4* and the *hTeloC iM* were observed upon the addition of **21d**. By contrast, compounds **15c**, **18c**, **20a,** and **23a** were capable to affect the resonances of the imino protons of the *c-Myc G4,* as well as those of the aromatic protons of both the *c-Myc G4* and the *hTeloC iM*. Overall, our findings suggested a common end-stacking binding mode for **15c**, **18c**, **20a**, and **23a** to the 3′ G-tetrad of the *c-Myc G4* (combined with loop binding), and binding to the *hTeloC iM* loops, which is one of the most common mechanisms of interaction proposed for iM-targeting compounds (Masoud & Nagasawa, 2018).

Saturation transfer difference NMR studies. In order to confirm the binding of compounds **15c**, **18c**, **20a**, and **23a** to the *c-Myc G4* and *hTeloC iM*, a ligand-based NMR approach was used. Compound **21d** was also included. Particularly, saturation transfer difference NMR (STD NMR) experiments were performed in buffer solutions containing a large excess of compounds (250 μM) with respect to the DNA (20 μM) (Mayer & Meyer, 1999). STD NMR is based on the magnetization transfer, by the nuclear Overhauser effect (NOE), from the macromolecule to the ligand. STD NMR spectra are the result of the subtraction of a 1D ^{1}H spectrum where the macromolecule protons are selectively saturated (on-resonance spectrum) from a reference spectrum in which those protons are not saturated (off-resonance spectrum). The resulting STD spectrum will only display the signals of the ligand functional groups interacting with the macromolecule, if any. Herein, the STD spectra showed aromatic and aliphatic signals for **15c**, **18c**, **20a**, and **23a**, in the presence of both the *c-Myc G4* and *hTeloC iM*, confirming the binding of the compounds to the investigated noncanonical structures. Moreover, the results suggested that the aromatic moieties of the compounds could interact with the G4 and iM structures through π-π stacking with the guanines of the G-tetrads and the adenines of the loops, respectively, whereas their side chains could bind to the DNA backbone through electrostatic

interactions. Finally, the very low signal intensity of **21d** protons in the STD spectra corroborated its inability to interact with the DNA sequences under investigation.

Fluorescence studies. With the aim of obtaining quantitative data regarding the affinity of **15c**, **18c**, **20a**, and **23a** for the *c-Myc G4*, the *c-Myc iM*, and the *hTeloC iM*, fluorescence titration experiments were performed. As reported in Table 5.4, compound **18c**, which displayed the strongest affinity for the G4 structure of *c-Myc*, also exhibited the least affinity for the iM structure of both *c-Myc* and *hTeloC*. On the other hand, compounds **15c**, **20a** and **23a** showed higher affinity values to the *c-Myc iM* and the *hTeloC iM* rather than to the *c-Myc G4.* Furthermore, fluorescence titration experiments were carried out to investigate the interaction of **21d** with a G4 (*c-Myc G4*) and an iM (*c-Myc iM*) structure. The obtained curves were then fitted, giving binding constant (K_b) values of 1.1 (± 0.2) × 10^6 and 1.6 (± 0.2) × 10^6 M^{-1} for the *c-Myc G4* and the *c-Myc iM*, respectively, confirming the lower (although not null) affinity of **21d** for these targets. Finally, a fluorescence titration study was also performed by using **18c** and the [d(CT)$_{15}$] C-rich single-stranded oligonucleotide. The resulting K_b value (4.6 (± 0.4) × 10^6 M^{-1}) suggested that the compound-induced decrease in the melting temperature of the iMs could also be due to its preferred binding to the unfolded DNA, which would shift the folding equilibrium.

Table 5.4. Binding constant (K_b) values obtained by fitting the fluorescence titration curves to an independent and equivalent binding sites model (Giancola & Pagano, 2013).

Compound	K_b (×10^6 M^{-1}) [1]		
	c-Myc G4	***c-Myc iM***	***hTeloC iM***
20a	2.3	3.1	3.1
23a	2.2	3.2	3.1
15c	2.5	3.4	3.2
18c	3.1	2.8	2.6

[1] The error on the K_b values is ± 0.2 ×10^6 M^{-1}.

Immunofluorescence studies. Starting from the results of the biophysical assays, the selected molecules (**15c**, **18c**, **20a**, **23a**, and **21d** as a control) were tested for their capability to destabilize the iM and/or stabilize the G4 structures also in cells. To this aim, human osteosarcoma cells (U2OS), one of the models already used by Zeraati and colleagues to validate their anti-iM antibody (iMab) (Zeraati et al., 2018), were treated with 2 μM of each

compound, for 24 h, before proceeding with immunofluorescence (IF) microscopy. In accordance with the above results, all the investigated molecules, except for **21d**, succeeded in destabilizing the iMs, as demonstrated by the significant reduction in the iM structures following the treatments (Figures 5.6 and 5.8A). In this regard, the compounds differed from each other in terms of potency in destabilizing the iMs, with **15c** and **18c** that, inducing a reduction of about 70% and 60% in the iMs fluorescence signal, respectively, turned out to be the most effective compounds. Notably, the greater efficacy of **15c** and **18c** in destabilizing iMs well correlated with their capability to stabilize G4 structures (Figures 5.7 and 5.8B). The dual activity of **15c** and **18c**, already observed at the biophysical level, is in line with the literature data showing that, under certain conditions, the formation of G4s and iM structures can be reciprocally influenced (King et al., 2020). Altogether, our data confirmed the ability of the selected compounds to affect G4/iM structures also in the complexity of a cellular model.

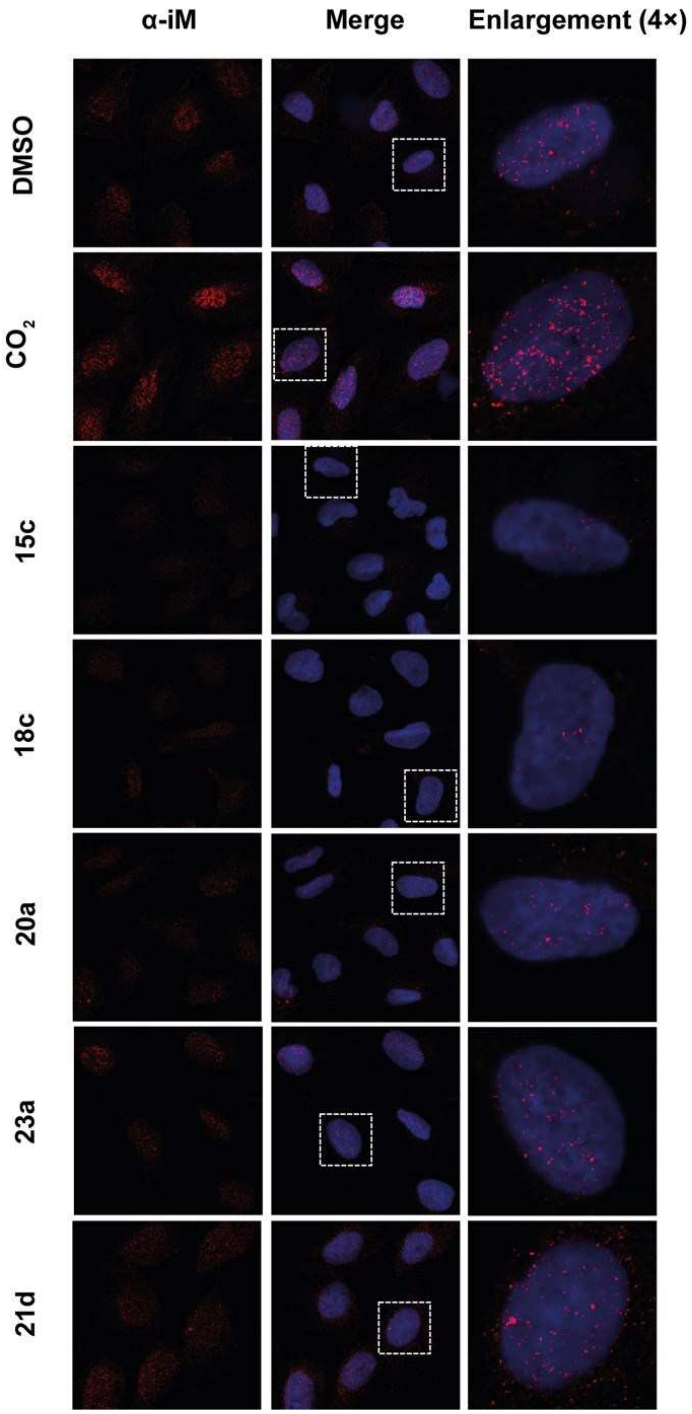


Figure 5.6. Biological evaluation of the iM-destabilizing activity of the selected compounds. Immunofluorescence analysis of iM structures in U2OS cells treated for 24 h with 2 μM of the selected compounds or an equivalent amount of DMSO (negative control). As a positive control, cells were maintained for 2.5 h in an atmosphere with 8% of CO_2. Representative images of confocal sections (63×) used to detect iM structures are shown. **Left panels**: iM structures (red) detected by anti-iM antibody (α-iM). **Middle panels**: merged images showing iM structures (red) and DAPI counterstained nuclei (blue). **Right panels**: 4× enlargements from the pictures in the middle panels. The experiment was performed in triplicate and at least 9 fields/experiments were evaluated for each condition. Figure from (Di Porzio et al., 2021).

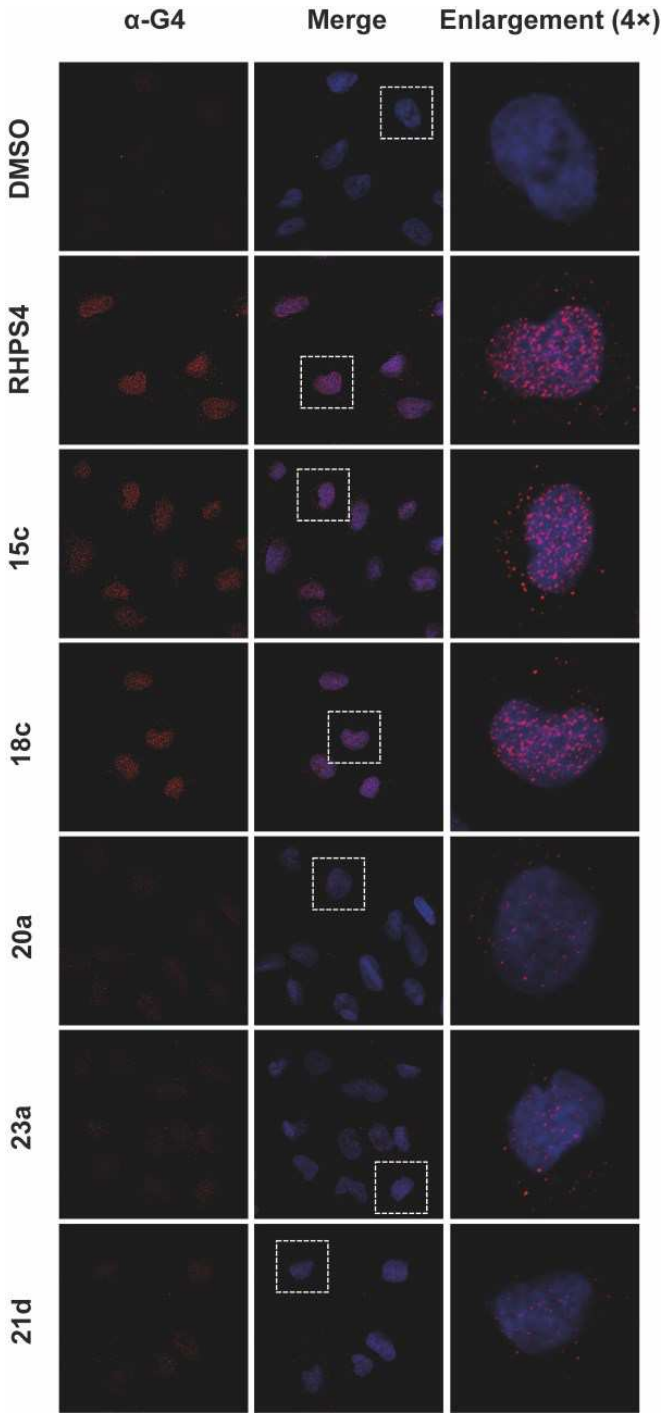


Figure 5.7. Biological evaluation of G4 stabilizing activity of the selected compounds. Immunofluorescence analysis of G4 structures in U2OS cells treated for 24 h with 2 μM of the selected compounds or an equivalent amount of DMSO (negative control). As a positive control, cells were treated for 24 h with 1 μM RHPS4. Representative images of confocal sections (63×) used for the detection of G4 structures are shown. **Left panels**: G4 structures (red) detected by anti-G4 antibody (α-G4). **Middle panels**: merged images showing G4 structures (red) and DAPI counterstained nuclei (blue). **Right panels**: 4× enlargements from the pictures in the middle panels. The experiment was performed in triplicate and at least 9 fields/experiments were evaluated for each condition. Figure from (Di Porzio et al., 2021).

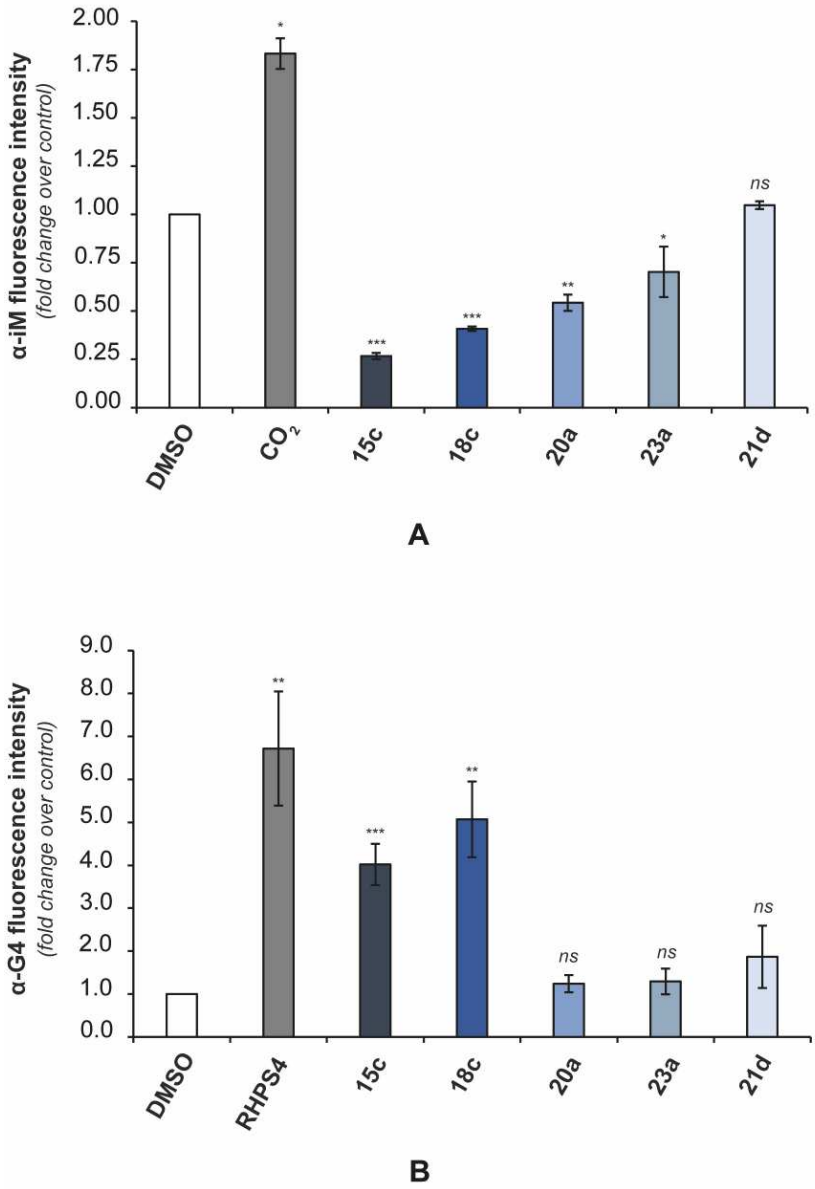


Figure 5.8. Quantitative analysis of fluorescence intensity of anti-iM (**A**) and anti-G4 (**B**) signals. For each panel, 25 cells/condition were analysed by ImageJ software and the results were expressed as fold change of fluorescence intensity over the negative control (DMSO). Results are referred to the images shown in Figures 5.6 and 5.7, respectively. Histograms show the mean ± SD of three independent experiments. Statistical significance was calculated using unpaired student *t*-tests on GraphPad Prism 6 (*p<0.05, **p<0.01, ***p<0.001). Figure from (Di Porzio et al., 2021).

5.2.3 *Conclusions*

The herein reported study brought us to the identification of compounds **15c** and **18c** as candidates capable of concomitantly stabilizing the G4 structures and destabilizing the iM ones, and compounds **20a** and **23a** that selectively affected the thermal stability of the iMs. NMR and fluorescence titration experiments gave insight into the interaction mode and affinity of the

selected compounds towards the noncanonical DNA structures under investigation. Intriguingly, **15c** and **18c** were able to significantly affect the formation of G4/iM structures also in the complexity of the cellular environment. To sum up, our investigation led to the identification of new compounds that could be utilized as molecular tools to elucidate the controversial biological roles of G4 and iM structures and the intricate way by which they influence each other.

GENERAL CONCLUSIONS

As first discovered in the early 1990s, stretches of cytosine-rich DNA are capable of folding into non-B conformations, called the i-motifs, demonstrating the great structural dynamism of nucleic acids. Particularly, the formation of iM structures is favored in mildly acidic solutions, where the protonation of Cs occurs. Such pH-dependency initially raised skepticism about whether the iMs also formed *in vivo*. In response to these doubts, it has been then shown that C-rich sequences can fold into an i-motif conformation also at neutral pH or under molecular crowding conditions, and the existence of these structures within the nuclei of living cells has been recently confirmed through in-cell NMR and immunofluorescence experiments. Sequences capable of folding into i-motif structures are evolutionarily conserved across species and frequently cluster within telomeres and gene promoters, thus suggesting their potential involvement in cellular processes such as telomere maintenance and transcription. On this point, increasing evidence supporting iMs ability to hinder telomerase activity or act as transcriptional regulators has been accumulated, making them ideal anti-cancer targets. In light of what has just been said, my research group and I thought that if we had found out the optimal experimental conditions (in terms of pH, cation type, and cation concentration in solution) to be used for studying iM-forming sequences *in vitro*, that would have had a remarkable impact on the field. To this aim, we exploited the unquestionable power of chemometric methods (i.e., DOE and PCA) and unveiled an unprecedented interaction between factors (pH and cation concentration) affecting iM formation: when working under experimental conditions close to neutral pH, the effect of cation concentration on iM structure formation is highly relevant, with greater amounts of structure being obtained at low cation concentration. Given the ever-growing interest in the iM structures as therapeutic targets, we also thought that the need for tools allowing a quick and valid interpretation of spectroscopic data acquired on iM samples was urgent and pressing. In this respect, chemometrics allowed us to identify CD and TDS bands which had never been reported before and which could provide researchers with precious information about the structure and chemical composition of an iM-forming sequence.

The complementary strand of a C-rich sequence might fold into another noncanonical secondary structure, the G-quadruplex. As for the iMs, G4-folding sequences are enriched in

functional genomic loci, including telomeres and gene promoters. G-rich motifs also occur in the lncRNA transcribed from the sub-telomere region (TERRA). Since the TERRA G4 structure has shown to play critical roles in telomere homeostasis, the design of small molecules to target it is receiving mounting attention. In this scenario, my research group and I sought to develop new TERRA G4 binders to potentially hit ALT-positive tumors, which correlate with high TERRA expression and are typically associated with the worst prognosis. By combining virtual screening and several biophysical/biological assays, we succeeded in the identification of a compound able to selectively bind to and stabilize the TERRA G4, both *in vitro* and in living cells.

Traditionally, drug discovery has been based on the rational design of highly selective chemical entities to hit a single biological target believed to be the dominant player in a certain pathology. Anyway, very few cancers have a single genetic mutation or factor that is entirely responsible for their development. Such genomic plasticity and resultant cellular heterogeneity preclude the efficacy of single-target therapy, which therefore is progressively fading in favor of multi-target approaches. The multi-target strategy, that acts on different targets associated with the multi-factorial disorder, comes in various forms, including the association of drugs or a single drug with multiple targets. Particularly, the combination, in a single chemical scaffold, of structural subunits that can be recognized by distinct bioreceptors would offer a more predictive pharmacokinetics, as well as a higher compliance, compared to the association of medicines. In this regard, taking into account the roles of carbonic anhydrases IX/XII and G-quadruplex structures in the context of a multi-factorial disease like cancer, we thought of designing and synthesizing carbonic anhydrase inhibitors-Berberine hybrids as potential multi-target anti-cancer agents. The newly synthesized molecules showed a remarkable efficacy in inhibiting the tumor-associated carbonic anhydrases IX and XII *in vitro*, with a certain degree of selectivity over the off-targeted CA I and II isoforms. Of note, the hybrids turned out to be more effective than Berberine in stabilizing (*in vitro*) the G4 structures under investigation, with the strongest effect being on the parallel *c-Kit1* and *c-Myc G4s*, indicating that hybridization of Berberine with CAI portions could increase its G4-stabilizing properties, especially in the case of parallel G4 conformations. The most promising molecules were then subjected to a further biological characterization, displaying a good cytotoxic effect on CA IX-positive human cervix cancer cells.

As already mentioned, the widespread localization of both G-quadruplexes and i-motifs in crucial areas of the human genome highlights breadth of processes that they might regulate. For

the detailed investigation of the features and roles of individual G4s/iMs, the identification and development of selective tools is needed. Particularly, we wanted to come up with a molecular entity allowing to selectively target and probe the therapeutically appealing G4 structure lying in the promoter region of the proto-oncogene c-MYC, which is overexpressed in more than 50% of human cancers. By taking inspiration from the crystal structure of the bovine DHX36 helicase bound to the *c-Myc G4*, we succeeded in identifying a short peptide able to preferentially bind, with nanomolar affinity, to the *c-Myc G4*, thus paving the way towards the development of a Myc-selective probe to uncover the biological significance of such G4 over the ~700 000 mapped across the human genome.

In general, small molecule-based targeting approaches fall among the most extensively used strategies to fully decipher the biological roles of noncanonical nucleic acid structures. More than a thousand compounds that bind to and stabilize G4 structures have been discovered so far, whereas there are comparatively very few iM-targeting molecules reported in the literature. In this context, we decided to combine a multicomponent reaction and a click chemistry functionalization to generate a set of 31 bis-triazolyl-pyridine derivatives as new potential G4/iM-interacting compounds. The pyridines were screened for their ability to interact with different G4 and/or iM DNAs and to affect their thermal stability, leading to the identification of new small molecules capable of either stabilizing the G4 structures while destabilizing the iM ones, or selectively affecting the thermal stability of the iMs, both *in vitro* and in living cells. Such compounds could be exploited as molecular tools to elucidate the biological functions of G4 and iM structures, as well as the basis underlying their intriguing interdependency.

To conclude, the noncanonical nucleic acids field is incredibly fervid and, although studies and research are moving at an incredible pace, there are still a lot of unaddressed questions … but that is exactly what makes science exciting and fascinating!

REFERENCES

Agarwal, T., Jayaraj, G., Prakash Pandey, S., Agarwala, P., & Maiti, S. (2012). RNA G-Quadruplexes: G-quadruplexes with "U" Turns. *Current Pharmaceutical Design*, *18*(14), 2102–2111. https://doi.org/10.2174/138161212799958468

Amato, J., Miglietta, G., Morigi, R., Iaccarino, N., Locatelli, A., Leoni, A., Novellino, E., Pagano, B., Capranico, G., & Randazzo, A. (2020). Monohydrazone Based G-Quadruplex Selective Ligands Induce DNA Damage and Genome Instability in Human Cancer Cells. *Journal of Medicinal Chemistry*, *63*(6), 3090–3103. https://doi.org/10.1021/acs.jmedchem.9b01866

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, A., & Suess, B. (2011). An RNA G-quadruplex in the 3' UTR of the proto-oncogene PIM1 represses translation. *RNA Biology*, *8*(5), 802–805. https://doi.org/10.4161/rna.8.5.16038

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*(1), 5220. https://doi.org/10.1038/ncomms6220

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*, *318*(5851), 798–801. https://doi.org/10.1126/science.1147182

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

Benabou, S., Garavís, M., Lyonnais, S., Eritja, R., González, C., & Gargallo, R. (2016). Understanding the effect of the nature of the nucleobase in the loops on the stability of the i-motif structure. *Physical Chemistry Chemical Physics*, *18*(11), 7997–8004. https://doi.org/10.1039/C5CP07428B

Bhattacharyya, D., Mirihana Arachchilage, G., & Basu, S. (2016). Metal Cations in G-

Quadruplex Folding and Stability. *Frontiers in Chemistry*, *4*, 38. https://doi.org/10.3389/fchem.2016.00038

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), 770–780. https://doi.org/10.1038/nrg3296

Bolger, R., & Checovich, W. (1994). A new protease activity assay using fluorescence polarization. *BioTechniques*, *17*(3), 585–589. http://www.ncbi.nlm.nih.gov/pubmed/7818914

Bolger, Randall, Wiese, T. E., Ervin, K., Nestich, S., & Checovich, W. (1998). Rapid screening of environmental chemicals for estrogen receptor binding capacity. *Environmental Health Perspectives*, *106*(9), 551–557. https://doi.org/10.1289/ehp.98106551

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-Quadruplexes. *Biochemistry*, *47*(2), 689–697. https://doi.org/10.1021/bi701873c

Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., & Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Research*, *34*(19), 5402–5415. https://doi.org/10.1093/nar/gkl655

Burger, A. M., Dai, F., Schultes, C. M., Reszka, A. P., Moore, M. J., Double, J. A., & Neidle, S. (2005). The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. *Cancer Research*, *65*(4), 1489–1496. https://doi.org/10.1158/0008-5472.CAN-04-2910

Cammas, A., & Millevoi, S. (2017). RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Research*, *45*(4), 1584–1595. https://doi.org/10.1093/nar/gkw1280

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), e1000861. https://doi.org/10.1371/journal.pcbi.1000861

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, M. C., Tippana, R., Demeshkina, N. A., Murat, P., Balasubramanian, S., Myong, S., & Ferré-D'Amaré, A. R. (2018). Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature*, *558*(7710), 465–469. https://doi.org/10.1038/s41586-018-0209-9

Chen, S., & Shen, X. (2020). Long noncoding RNAs: functions and mechanisms in colon cancer. *Molecular Cancer*, *19*(1), 167. https://doi.org/10.1186/s12943-020-01287-2

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*, *3*(1), 1074. https://doi.org/10.1038/ncomms2091

Cheng, M., Qiu, D., Tamon, L., Ištvánková, E., Víšková, P., Amrane, S., Guédin, A., Chen, J.,

Lacroix, L., Ju, H., Trantírek, L., Sahakyan, A. B., Zhou, J., & Mergny, J. (2021). Thermal and pH Stabilities of i-DNA: Confronting in vitro Experiments with Models and In-Cell NMR Data. *Angewandte Chemie International Edition*, *60*(18), 10286–10294. https://doi.org/10.1002/anie.202016801

Cui, J., Waltman, P., Le, V., & Lewis, E. (2013). The Effect of Molecular Crowding on the Stability of Human c-MYC Promoter Sequence I-Motif at Neutral pH. *Molecules*, *18*(10), 12751–12767. https://doi.org/10.3390/molecules181012751

Cui, Y., Kong, D., Ghimire, C., Xu, C., & Mao, H. (2016). Mutually Exclusive Formation of G-Quadruplex and i-Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA. *Biochemistry*, *55*(15), 2291–2299. https://doi.org/10.1021/acs.biochem.6b00016

Dai, J., Carver, M., & Yang, D. (2008). Polymorphism of human telomeric quadruplex structures. *Biochimie*, *90*(8), 1172–1183. https://doi.org/10.1016/j.biochi.2008.02.026

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), 5133–5144. https://doi.org/10.1093/nar/gkl610

Dai, J., Hatzakis, E., Hurley, L. H., & Yang, D. (2010). I-Motif Structures Formed in the Human c-MYC Promoter Are Highly Dynamic–Insights into Sequence Redundancy and I-Motif Stability. *PLoS ONE*, *5*(7), e11647. https://doi.org/10.1371/journal.pone.0011647

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*, *42*, 183–195. https://doi.org/10.1016/j.ymeth.2006.10.004

De Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & Development*, *19*(18), 2100–2110. https://doi.org/10.1101/gad.1346005

Dhakal, S., Yu, Z., Konik, R., Cui, Y., Koirala, D., & Mao, H. (2012). G-Quadruplex and i-Motif Are Mutually Exclusive in ILPR Double-Stranded DNA. *Biophysical Journal*, *102*(11), 2575–2584. https://doi.org/10.1016/j.bpj.2012.04.024

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 Edition*, *51*(44), 11073–11078. https://doi.org/10.1002/anie.201206281

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. (2021). Synthesis and Characterization of Bis-Triazolyl-Pyridine Derivatives as Noncanonical DNA-Interacting Compounds. *International Journal of Molecular Sciences*, *22*(21), 11959. https://doi.org/10.3390/ijms222111959

Dzatko, S., Krafcikova, M., Hänsel-Hertsch, R., Fessl, T., Fiala, R., Loja, T., Krafcik, D., Mergny, J.-L., 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

Ebrahimi-Najafabadi, H., Leardi, R., & Jalali-Heravi, M. (2014). Experimental Design in Analytical Chemistry—Part I: Theory. *Journal of AOAC INTERNATIONAL*, *97*(1), 3–11.

https://doi.org/10.5740/jaoacint.SGEEbrahimi1

Fleming, A. M., & Burrows, C. J. (2020). Interplay of Guanine Oxidation and G-Quadruplex Folding in Gene Promoters. *Journal of the American Chemical Society*, *142*(3), 1115–1136. https://doi.org/10.1021/jacs.9b11050

Fleming, A. M., Ding, Y., Rogers, R. A., Zhu, J., Zhu, J., Burton, A. D., Carlisle, C. B., & Burrows, C. J. (2017). 4 n –1 Is a "Sweet Spot" in DNA i-Motif Folding of 2′-Deoxycytidine Homopolymers. *Journal of the American Chemical Society*, *139*(13), 4682–4689. https://doi.org/10.1021/jacs.6b10117

Franceschin, M., Rossetti, L., D'Ambrosio, A., Schirripa, S., Bianco, A., Ortaggi, G., Savino, M., Schultes, C., & Neidle, S. (2006). Natural and synthetic G-quadruplex interactive berberine derivatives. *Bioorganic & Medicinal Chemistry Letters*, *16*(6), 1707–1711. https://doi.org/10.1016/j.bmcl.2005.12.001

Fujii, T., & Sugimoto, N. (2015). Loop nucleotides impact the stability of intrastrand i-motif structures at neutral pH. *Physical Chemistry Chemical Physics*, *17*(26), 16719–16722. https://doi.org/10.1039/C5CP02794B

Garavís, M., González, C., & Villasante, A. (2013). On the Origin of the Eukaryotic Chromosome: The Role of Noncanonical DNA Structures in Telomere Evolution. *Genome Biology and Evolution*, *5*(6), 1142–1150. https://doi.org/10.1093/gbe/evt079

Gatenby, R. A., & Gillies, R. J. (2004). Why do cancers have high aerobic glycolysis? *Nature Reviews Cancer*, *4*(11), 891–899. https://doi.org/10.1038/nrc1478

Giancola, C., & Pagano, B. (2013). Energetics of ligand binding to G-quadruplexes. *Topics in Current Chemistry*, *330*, 211–242. https://doi.org/10.1007/128-2012-347

Gray, D. M., & Bollum, F. J. (1974). A circular dichroism study of poly dG, poly dC, and poly dG:dC. *Biopolymers*, *13*(10), 2087–2102. https://doi.org/10.1002/bip.1974.360131011

Greider, C. W., & Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase activity in tetrahymena extracts. *Cell*, *43*(2), 405–413. https://doi.org/10.1016/0092-8674(85)90170-9

Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H., & de Lange, T. (1999). Mammalian Telomeres End in a Large Duplex Loop. *Cell*, *97*(4), 503–514. https://doi.org/10.1016/S0092-8674(00)80760-6

Gu, J., Leszczynski, J., & Bansal, M. (1999). A new insight into the structure and stability of Hoogsteen hydrogen-bonded G-tetrad: an ab initio SCF study. *Chemical Physics Letters*, *311*(3–4), 209–214. https://doi.org/10.1016/S0009-2614(99)00821-0

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

Guo, J. U., & Bartel, D. P. (2016). RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*, *353*(6306), aaf5371. https://doi.org/10.1126/science.aaf5371

Guo, Q., Lu, M., Marky, L. A., & Kallenbach, N. R. (1992). Interaction of the dye ethidium bromide with DNA containing guanine repeats. *Biochemistry*, *31*(9), 2451–2455. https://doi.org/10.1021/bi00124a002

Hänsel-Hertsch, R., Beraldi, D., Lensing, S. V., Marsico, G., Zyner, K., Parry, A., Di Antonio,

M., Pike, J., Kimura, H., Narita, M., Tannahill, D., & Balasubramanian, S. (2016). G-quadruplex structures mark human regulatory chromatin. *Nature Genetics*, *48*(10), 1267–1272. https://doi.org/10.1038/ng.3662

Hänsel-Hertsch, R., Simeone, A., Shea, A., Hui, W. W. I., Zyner, K. G., Marsico, G., Rueda, O. M., Bruna, A., Martin, A., Zhang, X., Adhikari, S., Tannahill, D., Caldas, C., & Balasubramanian, S. (2020). Landscape of G-quadruplex DNA structural regions in breast cancer. *Nature Genetics*, *52*(9), 878–883. https://doi.org/10.1038/s41588-020-0672-8

Hantzsch, A. (1881). Condensationsprodukte aus Aldehydammoniak und ketonartigen Verbindungen. *Berichte Der Deutschen Chemischen Gesellschaft*, *14*(2), 1637–1638. https://doi.org/10.1002/cber.18810140214

Hantzsch, Arthur. (1882). Ueber die Synthese pyridinartiger Verbindungen aus Acetessigäther und Aldehydammoniak. *Justus Liebig's Annalen Der Chemie*, *215*(1), 1–82. https://doi.org/10.1002/jlac.18822150102

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

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*(6), 417–441. https://doi.org/10.1037/h0071325

Hsu, S.-T. D., Varnai, P., Bugaut, A., Reszka, A. P., Neidle, S., & Balasubramanian, S. (2009). A G-Rich Sequence within the c-kit Oncogene Promoter Forms a Parallel G-Quadruplex Having Asymmetric G-Tetrad Dynamics. *Journal of the American Chemical Society*, *131*(37), 13399–13409. https://doi.org/10.1021/ja904007p

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

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

Iaccarino, N., Di Porzio, A., Amato, J., Pagano, B., Brancaccio, D., Novellino, E., Leardi, R., & Randazzo, A. (2019). Assessing the influence of pH and cationic strength on i-motif DNA structure. *Analytical and Bioanalytical Chemistry*, *411*(28), 7473–7479. https://doi.org/10.1007/s00216-019-02120-6

Jain, A. K., & Bhattacharya, S. (2011). Interaction of G-Quadruplexes with Nonintercalating Duplex-DNA Minor Groove Binding Ligands. *Bioconjugate Chemistry*, *22*(12), 2355–2368. https://doi.org/10.1021/bc200268a

Jiao, L., Zhang, H.-L., Li, D.-D., Yang, K.-L., Tang, J., Li, X., Ji, J., Yu, Y., Wu, R.-Y., Ravichandran, S., Liu, J.-J., Feng, G.-K., Chen, M.-S., Zeng, Y.-X., Deng, R., & Zhu, X.-F. (2018). Regulation of glycolytic metabolism by autophagy in liver cancer involves selective autophagic degradation of HK2 (hexokinase 2). *Autophagy*, *14*(4), 671–684. https://doi.org/10.1080/15548627.2017.1381804

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., Akiyama, Y., Hecht, S. M., & Hurley, L. H. (2009). The i-Motif in the bcl-2 P1 Promoter Forms an Unexpectedly Stable Structure with a Unique 8:5:7 Loop Folding Pattern. *Journal of the American Chemical Society*, *131*(48), 17667–17676. https://doi.org/10.1021/ja9076292

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

Khalifah, R. G. (1971). The Carbon Dioxide Hydration Activity of Carbonic Anhydrase. *Journal of Biological Chemistry*, *246*(8), 2561–2573. https://doi.org/10.1016/S0021-9258(18)62326-9

Kim, N. (2019). The Interplay between G-quadruplex and Transcription. *Current Medicinal Chemistry*, *26*(16), 2898–2917. https://doi.org/10.2174/0929867325666171229132619

King, J. J., Irving, K. L., Evans, C. W., Chikhale, R. V., Becker, R., Morris, C. J., Peña Martinez, C. D., Schofield, P., Christ, D., Hurley, L. H., Waller, Z. A. E., Iyer, K. S., & Smith, N. M. (2020). DNA G-Quadruplex and i-Motif Structure Formation Is Interdependent in Human Cells. *Journal of the American Chemical Society*, *142*(49), 20600–20604. https://doi.org/10.1021/jacs.0c11708

Kumari, S., Bugaut, A., Huppert, J. L., & Balasubramanian, S. (2007). An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. *Nature Chemical Biology*, *3*(4), 218–221. https://doi.org/10.1038/nchembio864

Kwapisz, M., & Morillon, A. (2020). Subtelomeric Transcription and its Regulation. *Journal of Molecular Biology*, *432*(15), 4199–4219. https://doi.org/10.1016/j.jmb.2020.01.026

Lago, S., Nadai, M., Cernilogar, F. M., Kazerani, M., Domíniguez Moreno, H., Schotta, G., & Richter, S. N. (2021). Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nature Communications*, *12*(1), 3885. https://doi.org/10.1038/s41467-021-24198-2

Largy, E., Hamon, F., & Teulade-Fichou, M.-P. (2011). Development of a high-throughput G4-FID assay for screening and evaluation of small molecules binding quadruplex nucleic acid structures. *Analytical and Bioanalytical Chemistry*, *400*(10), 3419–3427. https://doi.org/10.1007/s00216-011-5018-z

Largy, E., Mergny, J.-L., & Gabelica, V. (2016). Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. In A. Sigel, H. Sigel, & R. Sigel (Eds.), *The Alkali Metal Ions: Their Role for Life. Metal Ions in Life Sciences* (pp. 203–258). Springer. https://doi.org/10.1007/978-3-319-21756-7_7

Leardi, R. (2009). Experimental design in chemistry: A tutorial. *Analytica Chimica Acta*, *652*(1–2), 161–172. https://doi.org/10.1016/j.aca.2009.06.015

Liberti, M. V., & Locasale, J. W. (2016). The Warburg Effect: How Does it Benefit Cancer Cells? *Trends in Biochemical Sciences*, *41*(3), 211–218. https://doi.org/10.1016/j.tibs.2015.12.001

Lightfoot, H. L., Hagen, T., Tatum, N. J., & Hall, J. (2019). The diverse structural landscape of quadruplexes. *FEBS Letters*, *593*(16), 2083–2102. https://doi.org/10.1002/1873-3468.13547

Luedtke, N. W. (2009). Targeting G-Quadruplex DNA with Small Molecules. *CHIMIA International Journal for Chemistry*, *63*(3), 134–139. https://doi.org/10.2533/chimia.2009.134

Lundblad, J. R. (1996). Fluorescence polarization analysis of protein-DNA and protein-protein interactions. *Molecular Endocrinology*, *10*(6), 607–612. https://doi.org/10.1210/me.10.6.607

Maiti, M., & Kumar, G. S. (2010). Polymorphic Nucleic Acid Binding of Bioactive Isoquinoline Alkaloids and Their Role in Cancer. *Journal of Nucleic Acids*, *2010*, 1–23. https://doi.org/10.4061/2010/593408

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

Marcel, V., Tran, P. L. T., Sagne, C., Martel-Planche, G., Vaslin, L., Teulade-Fichou, M.-P., Hall, J., Mergny, J.-L., Hainaut, P., & Van Dyck, E. (2011). G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis*, *32*(3), 271–278. https://doi.org/10.1093/carcin/bgq253

Maresca, A., Temperini, C., Vu, H., Pham, N. B., Poulsen, S.-A., Scozzafava, A., Quinn, R. J., & Supuran, C. T. (2009). Non-Zinc Mediated Inhibition of Carbonic Anhydrases: Coumarins Are a New Class of Suicide Inhibitors. *Journal of the American Chemical Society*, *131*(8), 3057–3062. https://doi.org/10.1021/ja809683v

Marshall, M. J., Neal, F. E., & Goldberg, D. M. (1979). Isoenzymes of hexokinase, 6-phosphogluconate dehydrogenase, phosphoglucomutase and lactate dehydrogenase in uterine cancer. *British Journal of Cancer*, *40*(3), 380–390. https://doi.org/10.1038/bjc.1979.192

Marsico, G., Chambers, V. S., Sahakyan, A. B., McCauley, P., Boutell, J. M., Antonio, M. Di, & Balasubramanian, S. (2019). Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Research*, *47*(8), 3862–3874. https://doi.org/10.1093/nar/gkz179

Martínez-Zaguilán, R., Seftor, E. A., Seftor, R. E. B., Chu, Y.-W., Gillies, R. J., & Hendrix, M. J. C. (1996). Acidic pH enhances the invasive behavior of human melanoma cells. *Clinical & Experimental Metastasis*, *14*(2), 176–186. https://doi.org/10.1007/BF00121214

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), 1–20. https://doi.org/10.3390/ijms221910315

Masoud, S. S., & Nagasawa, K. (2018). i-Motif-Binding Ligands and Their Effects on the Structure and Biological Functions of i-Motif. *Chemical and Pharmaceutical Bulletin*, *66*(12), 1091–1103. https://doi.org/10.1248/cpb.c18-00720

Matsumura, K., Kawasaki, Y., Miyamoto, M., Kamoshida, Y., Nakamura, J., Negishi, L., Suda, S., & Akiyama, T. (2017). The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. *Oncogene*, *36*(9), 1191–1199. https://doi.org/10.1038/onc.2016.282

Mayer, M., & Meyer, B. (1999). Characterization of Ligand Binding by Saturation Transfer

Difference NMR Spectroscopy. *Angewandte Chemie International Edition*, *38*(12), 1784–1788. https://doi.org/10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q

Mergny, J.-L. (2005). Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Research*, *33*(16), e138. https://doi.org/10.1093/nar/gni134

Mergny, Jean-Louis, & Hélène, C. (1998). G-quadruplex DNA: A target for drug design. *Nature Medicine*, *4*(12), 1366–1367. https://doi.org/10.1038/3949

Mergny, Jean-Louis, Lacroix, L., Han, X., Leroy, J.-L., & Helene, C. (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

Minard, A., Morgan, D., Raguseo, F., Di Porzio, A., Liano, D., Jamieson, A. G., & Di Antonio, M. (2020). A short peptide that preferentially binds c-MYC G-quadruplex DNA. *Chemical Communications*, *56*(63), 8940–8943. https://doi.org/10.1039/d0cc02954h

Mir, B., Solés, X., González, C., & Escaja, N. (2017). The effect of the neutral cytidine protonated analogue pseudoisocytidine on the stability of i-motif structures. *Scientific Reports*, *7*(1), 2772. https://doi.org/10.1038/s41598-017-02723-y

Moerke, N. J. (2009). Fluorescence Polarization (FP) Assays for Monitoring Peptide-Protein or Nucleic Acid-Protein Binding. *Current Protocols in Chemical Biology*, *1*(1), 1–15. https://doi.org/10.1002/9780470559277.ch090102

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

Monchaud, D., & Teulade-Fichou, M.-P. (2008). A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.*, *6*(4), 627–636. https://doi.org/10.1039/B714772B

Morgan, D. M. (1998). Tetrazolium (MTT) assay for cellular viability and activity. *Methods in Molecular Biology (Clifton, N.J.)*, *79*, 179–183. https://doi.org/10.1385/0-89603-448-8:179

Pagano, B., Amato, J., Iaccarino, N., Cingolani, C., Zizza, P., Biroccio, A., Novellino, E., & Randazzo, A. (2015). Looking for efficient G-quadruplex ligands: Evidence for selective stabilizing properties and telomere damage by drug-like molecules. *ChemMedChem*, *10*(4), 640–649. https://doi.org/10.1002/cmdc.201402552

Pagano, B., Cosconati, S., Gabelica, V., Petraccone, L., De Tito, S., Marinelli, L., La Pietra, V., Saverio di Leva, F., Lauri, I., Trotta, R., Novellino, E., Giancola, C., & Randazzo, A. (2012). State-of-the-Art Methodologies for the Discovery and Characterization of DNA G-Quadruplex Binders. *Current Pharmaceutical Design*, *18*(14), 1880–1899. https://doi.org/10.2174/138161212799958332

Paramasivan, S., Rujan, I., & Bolton, P. H. (2007). Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding. *Methods*, *43*(4), 324–331. https://doi.org/10.1016/j.ymeth.2007.02.009

Park, S. I., Suh, D. S., Kim, S.-J., Choi, K. U., & Yoon, M.-S. (2013). Correlation between Biological Marker Expression and 18 F-Fluorodeoxyglucose Uptake in Cervical Cancer Measured by Positron Emission Tomography. *Onkologie*, *36*(4), 169–174.

https://doi.org/10.1159/000349944

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(1), 559–572. https://doi.org/10.1080/14786440109462720

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., Luu, K. N., & Patel, D. J. (2007). Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K + solution †. *Nucleic Acids Research*, *35*(19), 6517–6525. https://doi.org/10.1093/nar/gkm706

Pompili, L., Leonetti, C., Biroccio, A., & Salvati, E. (2017). Diagnosis and treatment of ALT tumors: is Trabectedin a new therapeutic option? *Journal of Experimental & Clinical Cancer Research*, *36*(1), 189. https://doi.org/10.1186/s13046-017-0657-3

Pouysségur, J., Dayan, F., & Mazure, N. M. (2006). Hypoxia signalling in cancer and approaches to enforce tumour regression. *Nature*, *441*(7092), 437–443. https://doi.org/10.1038/nature04871

Rajendran, A., Nakano, S., & 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. https://doi.org/10.1039/b922050j

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

Rostovtsev, V. V., Green, L. G., Fokin, V. V., & Sharpless, K. B. (2002). A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angewandte Chemie International Edition*, *41*(14), 2596–2599. https://doi.org/10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4

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*, *14*(8), 769. https://doi.org/10.3390/ph14080769

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

Shahid, R., Bugaut, A., & Balasubramanian, S. (2010). The BCL-2 5′ Untranslated Region Contains an RNA G-Quadruplex-Forming Motif That Modulates Protein Expression. *Biochemistry*, *49*(38), 8300–8306. https://doi.org/10.1021/bi100957h

Shay, J. W., & Wright, W. E. (2000). Hayflick, his limit, and cellular ageing. *Nature Reviews Molecular Cell Biology*, *1*(1), 72–76. https://doi.org/10.1038/35036093

Shu, B., Cao, J., Kuang, G., Qiu, J., Zhang, M., Zhang, Y., Wang, M., Li, X., Kang, S., Ou, T.-M., Tan, J.-H., Huang, Z.-S., & Li, D. (2018). Syntheses and evaluation of new acridone derivatives for selective binding of oncogene c- myc promoter i-motifs in gene transcriptional regulation. *Chemical Communications*, *54*(16), 2036–2039. https://doi.org/10.1039/C8CC00328A

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*, *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*, *12*(1), 3760. https://doi.org/10.1038/s41467-021-24097-6

Singh, S., Lomelino, C., Mboge, M., Frost, S., & McKenna, R. (2018). Cancer Drug Development of Carbonic Anhydrase Inhibitors beyond the Active Site. *Molecules*, *23*(5), 1045. https://doi.org/10.3390/molecules23051045

Školáková, P., Renčiuk, D., Palacký, J., Krafčík, D., Dvořáková, Z., Kejnovská, I., Bednářová, K., & Vorlíčková, M. (2019). Systematic investigation of sequence requirements for DNA i-motif formation. *Nucleic Acids Research*, *47*(5), 2177–2189. https://doi.org/10.1093/nar/gkz046

Spiegel, J., Adhikari, S., & Balasubramanian, S. (2020). The structure and function of DNA G-quadruplexes. *Trends in Chemistry*, *2*(2), 123–136. https://doi.org/10.1016/j.trechm.2019.07.002

Spiegel, J., Cuesta, S. M., Adhikari, S., Hänsel-Hertsch, R., Tannahill, D., & Balasubramanian, S. (2021). G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biology*, *22*(1), 117. https://doi.org/10.1186/s13059-021-02324-z

Stock, C., & Schwab, A. (2009). Protons make tumor cells move like clockwork. *Pflügers Archiv - European Journal of Physiology*, *458*(5), 981–992. https://doi.org/10.1007/s00424-009-0677-8

Subramanian, M., Rage, F., Tabet, R., Flatter, E., Mandel, J., & Moine, H. (2011). G–quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Reports*, *12*(7), 697–704. https://doi.org/10.1038/embor.2011.76

Supuran, C. T., & Scozzafava, A. (2000). Carbonic anhydrase inhibitors and their therapeutic potential. *Expert Opinion on Therapeutic Patents*, *10*(5), 575–600. https://doi.org/10.1517/13543776.10.5.575

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 III 1. *Journal of the American Chemical Society*, *138*(42), 14138–14151. https://doi.org/10.1021/jacs.6b09196

Swietach, P., Vaughan-Jones, R. D., Harris, A. L., & Hulikova, A. (2014). The chemistry, physiology and pathology of pH in cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1638), 20130099. https://doi.org/10.1098/rstb.2013.0099

Tang, C.-F., & Shafer, R. H. (2006). Engineering the Quadruplex Fold: Nucleoside Conformation Determines Both Folding Topology and Molecularity in Guanine Quadruplexes. *Journal of the American Chemical Society*, *128*(17), 5966–5973. https://doi.org/10.1021/ja0603958

Todd, A. K. (2005). Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Research*, *33*(9), 2901–2907. https://doi.org/10.1093/nar/gki553

Varshney, D., Spiegel, J., Zyner, K., Tannahill, D., & Balasubramanian, S. (2020). The regulation and functions of DNA and RNA G-quadruplexes. *Nature Reviews Molecular*

Cell Biology, *21*(8), 459–474. https://doi.org/10.1038/s41580-020-0236-x

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

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

Wu, P., Brasseur, M., & Schindler, U. (1997). A High-Throughput STAT Binding Assay Using Fluorescence Polarization. *Analytical Biochemistry*, *249*(1), 29–36. https://doi.org/10.1006/abio.1997.2158

Wykoff, C. C., Beasley, N. J. P., Watson, P. H., Turner, K. J., Pastorek, J., Sibtain, A., Wilson, G. D., Turley, H., Talks, K. L., Maxwell, P. H., Pugh, C. W., Ratcliffe, P. J., & Harris, A. L. (2000). Hypoxia-inducible expression of tumor-associated carbonic anhydrases. *Cancer Research*, *60*(24), 7075–7083. http://www.ncbi.nlm.nih.gov/pubmed/11156414

Wykoff, C. C., Beasley, N., Watson, P. H., Campo, L., Chia, S. K., English, R., Pastorek, J., Sly, W. S., Ratcliffe, P., & Harris, A. L. (2001). Expression of the Hypoxia-Inducible and Tumor-Associated Carbonic Anhydrases in Ductal Carcinoma in Situ of the Breast. *The American Journal of Pathology*, *158*(3), 1011–1019. https://doi.org/10.1016/S0002-9440(10)64048-5

Y. Mboge, M., McKenna, R., & C. Frost, S. (2015). Advances in Anti-Cancer Drug Development Targeting Carbonic Anhydrase IX and XII. *Topics in Anti-Cancer Research*, *5*, 3–42. https://doi.org/10.2174/9781681083339116050004

Yaku, H., Fujimoto, T., Murashima, T., Miyoshi, D., & Sugimoto, N. (2012). Phthalocyanines: a new class of G-quadruplex-ligands with many potential applications. *Chemical Communications*, *48*(50), 6203. https://doi.org/10.1039/c2cc31037f

Yaneva, M. Y., Cheong, V. V., Cheng, J. K., Lim, K. W., & Phan, A. T. (2020). Stapling a G-quadruplex specific peptide. *Biochemical and Biophysical Research Communications*, *531*(1), 62–66. https://doi.org/10.1016/j.bbrc.2020.02.144

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*, *10*(6), 631–637. https://doi.org/10.1038/s41557-018-0046-3

Zheng, K., Zhang, J., He, Y., Gong, J., Wen, C., Chen, J., Hao, Y., Zhao, Y., & Tan, Z. (2020). Detection of genomic G-quadruplexes in living cells using a small artificial protein. *Nucleic Acids Research*, *48*(20), 11706–11720. https://doi.org/10.1093/nar/gkaa841

Zhou, J., Wei, C., Jia, G., Wang, X., Feng, Z., & Li, C. (2010). Formation of i-motif structure at neutral and slightly alkaline pH. *Mol. BioSyst.*, *6*(3), 580–586. https://doi.org/10.1039/B919600E

APPENDIX

Paper I

"Assessing the influence of pH and cationic strength on i-motif DNA structure". Iaccarino N., Di Porzio A., Amato J., Pagano B., Brancaccio D., Novellino E., Leardi R., Randazzo A. *Analytical & Bioanalytical Chemistry*, 2019, *411*(28), 7473–7479.

Paper II

"Effects of sequence and base composition on the CD and TDS profiles of i-DNA". Iaccarino N., Cheng M., Qiu D., Pagano B., Amato J., Di Porzio A., Zhou J., Randazzo A., Mergny JL. *Angewandte Chemie,* 2021, *60*(18), 10295–10303.

Paper III

"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.

Paper V

"A short peptide that preferentially binds c-MYC G4 DNA". Minard A., Morgan D., Raguseo F., Di Porzio A., Liano D., Jamieson A., Di Antonio M. *Chemical Communications*, 2020, *56*(63), 8940–8943.

Paper VI

"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.

Paper I

Assessing the influence of pH and cationic strength on i-motif DNA structure

Iaccarino N., Di Porzio A., Amato J., Pagano B., Brancaccio D., Novellino E., Leardi R., Randazzo A.

Analytical & Bioanalytical Chemistry, 2019, Vol. 411 (28)

Analytical and Bioanalytical Chemistry (2019) 411:7473–7479
https://doi.org/10.1007/s00216-019-02120-6

RESEARCH PAPER

Assessing the influence of pH and cationic strength on i-motif DNA structure

Nunzia Iaccarino[1] · Anna Di Porzio[1] · Jussara Amato[1] · Bruno Pagano[1] · Diego Brancaccio[1] · Ettore Novellino[1] · Riccardo Leardi[2] · Antonio Randazzo[1]

Received: 2 August 2019 / Accepted: 2 September 2019 / Published online: 16 September 2019

Abstract

The i-motif is a biologically relevant non-canonical DNA structure formed by cytosine-rich sequences. Despite the importance of the factors affecting the formation/stability of such a structure, like pH, cation type and concentration, no systematic study that simultaneously analysed their effect on the i-motif in vitro has been carried out so far. Therefore, here we report a systematic study that aims to evaluate the effect of these factors, and their possible interaction, on the formation of an i-motif structure. Our results confirm that pH plays the main role in i-motif formation. However, we demonstrate that the effect of the cation concentration on the i-motif is strictly dependent on the pH, while no significant differences are observed among the investigated cation types.

Keywords Nucleic acids (DNA/RNA) · Design of experiments · Multivariate data analysis · i-Motif · NMR · Circular dichroism

Introduction

Non-canonical DNA secondary structures include hairpins, cruciforms, parallel-stranded duplexes, triplexes, G-quadruplexes and i-motifs (also known as C-quadruplexes) [1–6]. Interestingly, these DNA secondary structures, which exhibit sequence dependency, can be formed by sequences that are widely distributed throughout the human genome [7]. However, these structures are not randomly located within the genome, but are typically present in regions that play crucial roles in a number of biological functions in vivo, including replication, gene expression regulation and DNA recombination [8–11]. Among all the non-canonical DNA secondary structures, the i-motif (iM) has attracted a lot of attention in the last decade. iMs are formed by C-rich sequences and are composed of two intercalated duplexes, stabilized by hemi-protonated cytosine-cytosine$^+$ ($C{\cdot}C^+$) base pairs (Fig. 1) [12, 13]. Generally, studies on iMs have been limited based on the idea that, because they are stabilized by slightly acidic conditions, they are not physiologically relevant. Nevertheless, this assumption has been challenged with examples of iMs that can form at neutral pH and under molecular crowding conditions [14, 15]. Furthermore, evidences for iM formation in vivo have been recently provided by in cell NMR experiments [16], and by the discovery of an antibody that specifically binds iM in the nuclei of human cells [17]. Generally, to favour the formation of hemi-protonated $C{\cdot}C^+$ base pairs, iM structures are studied in vitro in acidic solutions and possibly in the presence of sodium or potassium ions, whose concentrations vary from 10 to 150 mM. Unfortunately, in spite of the importance of understanding the role of the factors involved in the stabilization of iM structures, there is no systematic study that simultaneously explores the effect of pH, cation type and cation concentration on the iM formation in vitro. Therefore, here we report a design of experiments (DoE), to study the effect of these factors and their interactions on the formation of an i-motif structure. DoE is a chemometric approach to determine the relationship between factors affecting a system [18]. Indeed, in such method, the factors of interest are varied systematically from their lowest to highest value, and all possible factor combinations are executed in the same set of experiments, with the goal to screen the experimental space for optimum values. This approach is opposed to the univariate

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00216-019-02120-6) contains supplementary material, which is available to authorized users.

✉ Antonio Randazzo
antonio.randazzo@unina.it

[1] Department of Pharmacy, University of Naples Federico II, Via D. Montesano 49, 80131 Naples, Italy

[2] Department of Pharmacy, University of Genova, Viale Cembrano 4, 16147 Genoa, Italy

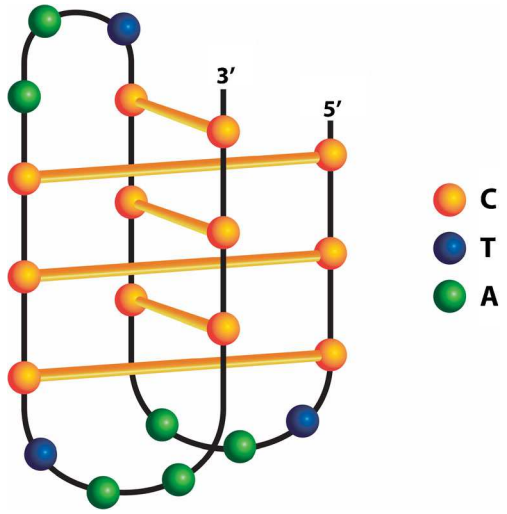


Fig. 1 Schematic representation of an i-motif structure

approach, also known as one-variable-at-time (OVAT), where all parameters are kept constant while only one variable is optimized. The OVAT approach requires a large number of experiments and does not allow investigation of the possible interactions between the studied variables. In the DoE approach, instead, the use of multiple linear regression (MLR) allows obtaining a mathematical model by which the effect of the variables and their interactions can be estimated. In this work, with the aid of spectroscopic techniques, such as NMR and circular dichroism (CD), we acquired an entire set of experiments established by the DoE. Both the NMR and CD datasets were then analysed using a multivariate approach, such as principal component analysis (PCA), to obtain the scores that were analysed by the MLR statistical tool.

Material and methods

Oligonucleotide synthesis and sample preparation

Oligonucleotide synthesis and purification were carried out following the same procedures described elsewhere [19]. The DNA sequences (5 μmol scale) were prepared by employing a standard ß-cyanoethylphosphoramidite solid phase chemistry on an ABI 394 DNA/RNA synthesizer (Applied Biosystem). DNA detachment from support and its deprotection were carried out by means of an aqueous solution of concentrated ammonia at 55 °C for 12 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). Different buffers were employed for the purification: buffer A, consisting of a 20 mM $KH_2PO_4/NaH_2PO_4/NH_4H_2PO_4$ (for $K^+/Na^+/NH_4^+$ sample purifications, respectively) aqueous solution (pH 7.0), containing 20% (v/v) CH_3CN, and buffer B, consisting of 1 M KCl/NaCl/NH_4Cl and 20 mM $KH_2PO_4/NaH_2PO_4/NH_4H_2PO_4$ (respectively for $K^+/Na^+/NH_4^+$ sample purifications) aqueous solution (pH 7.0) and containing 20% (v/v) CH_3CN. Buffers A and B were combined in a 30 min linear gradient going from 0 to 100% B with a flow rate of 1 mL/min. 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%. The sequence d[CCCT(AACCCT)$_3$] (hTeloC) was employed for the experiments. The oligonucleotide concentrations were established by measuring the UV absorption at 90 °C taking into account the molar extinction coefficient values ε (λ = 260 nm) determined by the nearest neighbour model [20].

For the full factorial design experiments (see below), hTeloC was dissolved in 10 and 200 mM potassium phosphate buffers, 10 and 200 mM sodium phosphate buffers and 10 and 200 mM ammonium phosphate buffers. Each buffer was used at two different pH values: 5.0 and 6.6. For the central composite design experiments, hteloC was dissolved in 2, 10 and 18 mM sodium phosphate buffers, at three pH values: 6.4, 6.6 and 6.8.

NMR experiments

For the full factorial design experiments, NMR spectra were acquired using a 700 MHz Varian Unity INOVA spectrometer. One-dimensional proton spectra were recorded at 10 °C, using pulsed-field gradient DPFGSE for water suppression, with 256 scans, spectral width 18,181.8 Hz, delay 3 s, receiver gain 20 and 32 k points. Twelve samples, referred to as *Block 1*, and the relative replicates, referred to as *Block 2*, were prepared at 280 μM strand concentration in 250 μL (H_2O/D_2O 9:1) of the relative buffer solution (see Electronic Supplementary Material (ESM) Table S2). NMR data were processed using the iNMR software (www.inmr.net) and then imported in Matlab (R2015b The Mathworks Inc.). A global alignment step was carried out, using icoshift algorithm [21]. The region between 7.82–6.22 ppm was discarded because of a signal characterizing the samples in ammonium buffer at pH 5.0; furthermore, the data interval between 5.79 ppm and 2.92 ppm was also discarded, because of the residual water signal; finally, the region below 1.48 ppm was discarded since it contains only noise. For the central composite design experiments, a Bruker 600 MHz spectrometer was employed. 1D ^{1}H-NMR spectra were recorded at 10 °C with a 'zgesgp' pulse-program (a gradient-based excitation sculpting using 180° water-selective pulses) with 256 scans, spectral width 15,625 Hz, delay 3 s, receiver gain 181 and 32 k points. Ten samples, referred to as *Set 1*, and the relative replicates, referred to as *Set 2*, were prepared at 250 μM strand concentration in 250 μL (H_2O/D_2O 9:1) sodium buffer solution at the corresponding pH and ion strength (see ESM Table S4). NMR data

were processed using the iNMR software (www.inmr.net) and then imported in Matlab (R2015b The Mathworks Inc.).

Circular dichroism spectroscopy

Circular dichroism (CD) experiments were recorded on a Jasco J-815 spectropolarimeter, equipped with a PTC-423S/15 Peltier temperature controller, following a previously reported protocol [19]. Each spectrum was recorded in the 220–360 nm wavelength range. The scan rate was set to 100 nm/min, with a 1 s response time, and 1 nm bandwidth. Spectra were analysed using Origin 7.0 software. CD experiments were performed using 28–30 μM oligonucleotide concentration. CD spectra were imported in Matlab. In order to consider only the variability related to the i-motif structure formation, the 240–340 nm wavelength interval was considered, thus discarding the regions above 340 nm and below 240 nm. Finally, each spectrum was normalized with respect to the relative DNA sample concentration.

Multivariate data analysis

Both the NMR and CD datasets were analysed using a multivariate approach, such as principal component analysis (PCA) [22]. Before data analysis, some pre-processing steps were performed. In order to minimize non-sample related variations, the final NMR matrix was normalized dividing each variable by the sum of all variables within a sample (1-norm). The variables were then mean centred. Mean centering was also performed on the CD dataset. PCA was then computed by using PLS Toolbox 8.6.1 in Matlab (R2015b The Mathworks Inc.) environment. The multiple linear regression (MLR) model for the NMR was computed by using as response the PC1/PC2 scores of the 24 experiments, while the MLR model for the CD was obtained by employing the PC1 scores of the relative PCA model. The MLR models were computed by using the R-based CAT toolbox freely available at "http://gruppochemiometria.it/index.php/software". The NMR dataset obtained from the central composite design was analysed analogously to the one from the full factorial design, using 1-norm and mean centering as pre-processing steps. Only the region between 14.6 and 15.8 ppm was considered in the PCA, being the most diagnostic for i-motif DNA structure formation. The PC1 score values of the 20 samples were submitted to multiple linear regression to obtain the mathematical model including linear, interaction and quadratic terms.

Results and discussion

In this study, we have used a truncation of the human telomeric C-rich sequence, namely d[CCCT(AACCCT)$_3$] (hTeloC), as sample model. This sequence is known to form a well-defined iM structure below pH 5.7 in solution containing 10 mM sodium ions [19], while it is basically unstructured at pH 7.0. So, we decided to explore the behaviour of iM between two pH values (quantitative variable), namely pH 5.0 and 6.6. At pH 5.0, the iM is completely structured, while at pH 6.6 the equilibrium is more shifted towards the unstructured form. Then, we wanted to observe the effect of three different cations (qualitative variable), namely Na^+, K^+ and NH_4^+, on the formation of the iM structure. Finally, the role of the concentration of these ions (quantitative variable) was also considered and it was evaluated between two values of concentration, namely at 10 and 200 mM. Therefore, since we had two variables at two levels and one variable at three levels, it has been decided to perform all the 12 (i.e. $2^2 \cdot 3^1$) possible experiments, this corresponding to a full factorial design (FFD) [23, 24].

This allowed us to estimate the coefficients of the following model:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_{3A}X_{3A} + b_{3B}X_{3B} + b_{12}X_1X_2$$

where X_1 = pH, X_2 = [ion] and X_3 = ion type. In this equation, b_0 is the constant term, b_1 is the coefficient of the linear term of pH, b_2 is the coefficient of the linear term of ion concentration, b_{3A} is the coefficient of the linear term of ion type at level K^+, b_{3B} is the coefficient of the linear term of ion type at level NH_4^+ and b_{12} is the coefficient of the interaction term between pH and concentration. It has to be noted that no coefficient for Na^+ is present, since Na^+ is the implicit level (the coefficients b_{3A} and b_{3B} must be interpreted as the difference between the results obtained when working with K^+ and NH_4^+, respectively, and the results obtained when working with Na^+). The experimental plan is reported in Table 1. The matrix has twelve rows, each row corresponding to an experimental condition, and three columns, each one corresponding to a variable. Thus, twelve experiments were performed and the

Table 1 Experimental plan

Experiment	X_1 (pH)	X_2 (cation) (mM)	X_3 (cation type)
1	5.0	10	Na^+
2	5.0	200	Na^+
3	6.6	10	Na^+
4	6.6	200	Na^+
5	5.0	10	K^+
6	5.0	200	K^+
7	6.6	10	K^+
8	6.6	200	K^+
9	5.0	10	NH_4^+
10	5.0	200	NH_4^+
11	6.6	10	NH_4^+
12	6.6	200	NH_4^+

resulting samples were analysed by NMR and CD spectroscopy. The experimental matrix, containing the coded values for each variable, is reported in the ESM (Table S1). In order to estimate the experimental variability and to have a better estimation of the coefficients of the model, the set of 12 experiments was performed twice (ESM Table S2). For the sake of simplicity, the figures reporting the NMR and CD spectra only refer to the first set of experiments. The 1D ^{1}H-NMR spectra of hTeloC in the different experimental conditions, defined by the experimental design, are reported in Fig. 2. Independently from the type of ion and concentration used, the i-motif in all the experiments

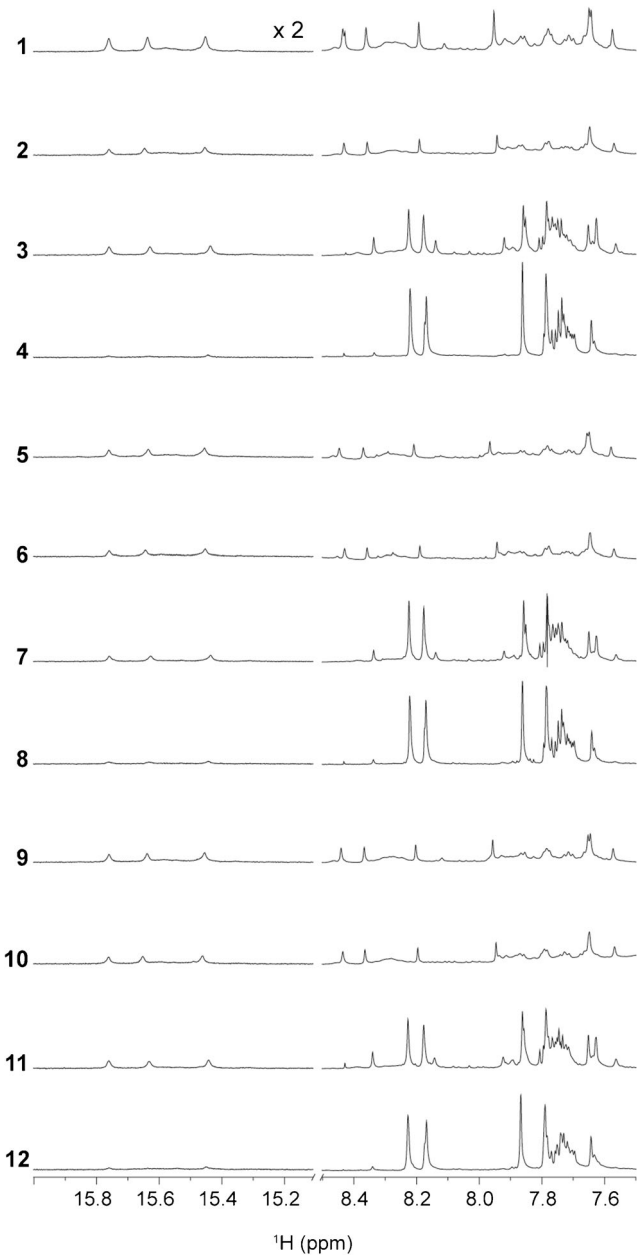


Fig. 2 ^{1}H-NMR spectra of the 12 samples investigated in the full factorial design

performed at pH 5.0 (1, 2, 5, 6, 9, 10) turned out to be very well structured, as witnessed by the presence, among others, of the imino signals of the base-paired cytosines at 15.76, 15.64 and 15.45 ppm, and by the aromatic signals at 8.44, 8.36 and 8.19 ppm [25]. On the other hand, all the experiments performed at pH 6.6 (3, 4, 7, 8, 11, 12) contained signals of both structured and unstructured DNA, this latter identified, among others, by the intense aromatic signals at 8.22 and 8.17 ppm. This suggests that, at pH 6.6, in all combinations of ion type and concentration, the oligodeoxynucleotide is in equilibrium between these two species. Interestingly, the spectra of the experiments performed at 200 mM ion concentration and pH 6.6 (4, 8, 12) show a smaller amount of structured DNA than those performed at 10 mM, indicating that the ion concentration may play a role in the stabilization of the iM structure. In order to take into account simultaneously all the information contained in the NMR spectra, a PCA has been performed. The score and loading plots of the PCA are reported in Fig. 3 (for sample identification in the score plot, see ESM Fig. S1). The first principal component (PC1) clearly explains the effect of pH, while PC2 shows that the ion concentration has an effect on the formation of the iM structure only at pH 6.6. In particular, having a careful inspection of the PC1 and PC2 loading plots, it is clear that the samples having negative scores on PC1, as expected, contain almost exclusively structured iM, while those having positive scores on PC2 contain a higher percentage of structured iM in comparison to those having negative scores on PC2, which contain almost exclusively unstructured DNA. We decided to use as response for the model the sum of the scores on PC1 and PC2, with changed sign, i.e. response = −(PC1 + PC2). As a result, the higher the response value the more structured is the sample. The following model has been obtained:

$$Y = 0.0004 - 0.0081\,X_1^{(***)} - 0.0036\,X_2^{(***)} - 0.0006\,X_{3A} - 0.0008\,X_{3B} - 0.0030\,X_1X_2^{(***)}$$

Since 24 experiments have been performed and six coefficients have been estimated, 18 degrees of freedom are available and therefore the significance of the coefficients can be reliably estimated (the level of significance is indicated according to the usual convention: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$). The plot of the coefficients is shown in Fig. S2 (ESM). The linear terms of pH (X_1) and ion concentration (X_2) and their interaction (X_1X_2) have significant coefficients, while the ion has no effect. The fact that the interaction term is significant means that it is not possible to interpret the effect of pH and ion concentration independently. In order to better understand this interaction, the 3D response surfaces on the pH–ion concentration plane are shown in Fig. S3 (see ESM). From these, it can be seen that at low pH values the highest

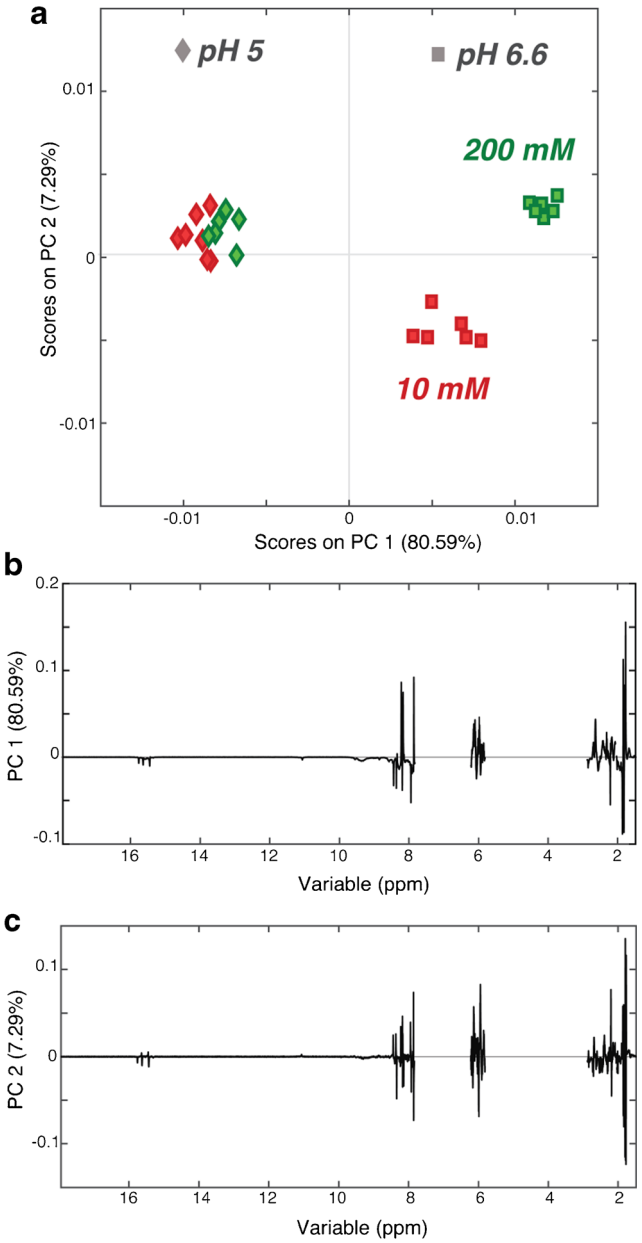


Fig. 3 PCA **a** score plot and loading plots of **b** PC1 and **c** PC2 of the FFD NMR data (including the replicates)

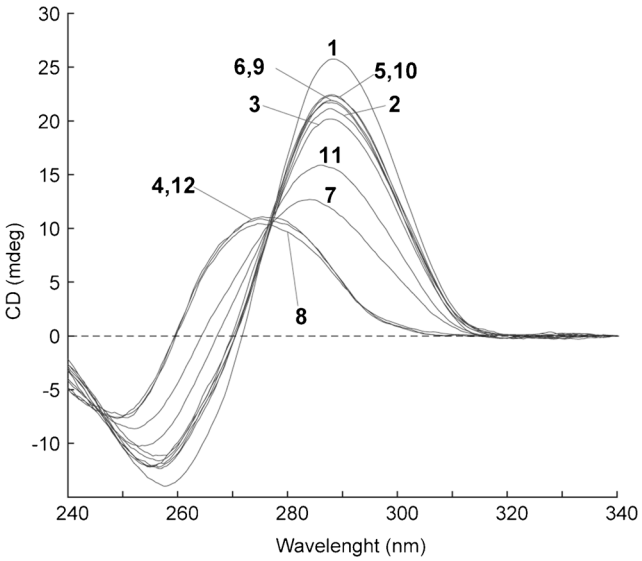


Fig. 4 CD spectra of the 12 samples investigated in the full factorial design

percentage of iM formation is obtained, and that at those conditions the ion concentration has basically no effect. At higher pH values, instead, the ion concentration has a relevant effect on the iM formation, with higher amounts obtained at lower ion concentration.

As mentioned above, CD spectra were also acquired on the same iM samples (Fig. 4). The maximum and minimum Cotton effects at 288 and 258 nm are indicative of the formation of the structure. Obviously, their intensities are directly proportional to the amount of iM formed in solution. On the other hand, the bands at 275 and 249 nm are indicative of the unstructured DNA. The bands at 288 and 258 nm are significantly more intense in the CD spectra of the samples acquired at pH 5.0, while the profile of the unstructured DNA is displayed at different extent in all other samples.

In order to take into consideration all the acquired points of each CD spectrum, a PCA has been computed and the relative score and loading plots are reported in Fig. 5 (for sample identification in the score plot, see ESM Fig. S4). The first principal component explains 97.7% of the total variance and contains information about the effect of both the pH and ion concentration. It can be seen that the experiments performed at pH 5.0 have positive scores on PC1 and are very close to each other, independently of the concentration of the ion, thus meaning that at pH 5.0 the ion concentration has no effect. This perfectly agrees with what we have observed from the NMR dataset. On the other hand, the experiments performed at pH 6.6 are distributed in a wider area, ranging from those performed at 200 mM ion concentration (highly negative scores on PC1) to the ones at low ion concentration (10 mM) that have intermediate scores on PC1 and are located close to the experiments performed at pH 5.0, this confirming that at higher pH the ion concentration has a relevant effect. Consistently with what observed for NMR data, from the analysis of the PC1 loading plot, the experiments performed at pH 5.0 contain the most structured i-motif (independently from the ion concentration), while at pH 6.6 there is a strong effect of the concentration, with the experiments performed at higher ion concentration containing the least structured DNA and the experiments performed at lower ion concentration containing only partially structured i-motif. The CD data were also used to compute a regression model, where the PC1 score values were used as response:

$$Y = 0.12 - 1.76\,X_1^{(***)} - 0.76\,X_2^{(***)} - 0.37\,X_{3A} - 0.00\,X_{3B} - 0.56\,X_1X_2^{(***)}$$

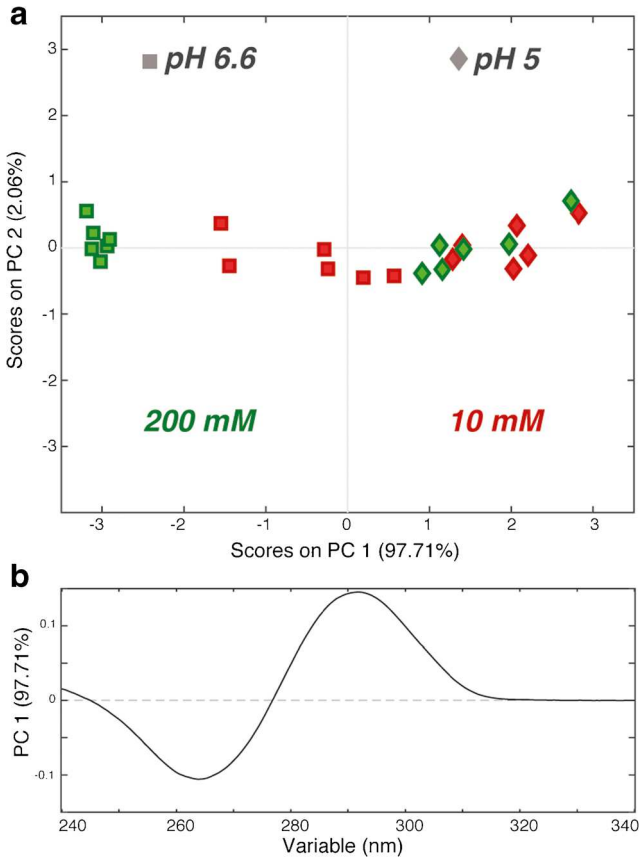


Fig. 5 PCA **a** score and **b** loading plots of the CD data

As previously discussed, the interaction between pH and ion concentration is highly significant, while the type of ion has no effect. Figures S5 and S6 (ESM) show the plot of the coefficients and the response surfaces, respectively. The coefficients (and therefore the model) are in perfect agreement with those computed on NMR data, thus suggesting the uniqueness of the results obtained by two complementary spectroscopic techniques.

Based on the above results, and with the aim of spanning the study to an environment as close as possible to the physiological one, further experiments were performed. In particular, by setting up a central composite design (CCD) [26], we decided to explore the i-motif behaviour varying the pH range from 6.4 to 6.8 and the cationic strength from 2 to 18 mM of sodium. Compared to the FFD, the CCD also allows getting information about a potential non-linear behaviour of the investigated variables (here pH and cationic strength). This is possible thanks to the additional experiments performed at the star and central points of the explored experimental space (Tables S3, ESM), allowing the calculation of the quadratic terms in the subsequent mathematical model:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_{12}X_1X_2 + b_{11}X_1^2 + b_{22}X_2^2$$

Thus, 10 one-dimensional NMR proton experiments including a replicate at the central point, which is represented by pH 6.6 and 10 mM Na^+, were performed in duplicate (Table S4, ESM). After spectra acquisition (Fig. S7, ESM), only the imino protons region was then submitted to the principal component analysis (Fig. S8, ESM). The obtained results (Fig. S9) clearly indicate that narrowing the pH and ionic strength range, the crucial factor driving the stability and formation of the i-motif structure is represented, once again, by the pH, while no significant variations depending on ionic strength are appreciable. It should be noted that the lack of significance of all the other coefficients could also be related to the intrinsic variability of the NMR response under these conditions (pH ≥ 6.6), especially in such close intervals.

In summary, we reported for the first time a systematic study on the role of pH, ion type and concentration on the iM structure formation by employing DoE and multivariate data analysis. Once again, we proved that the pH plays the main role on the iM formation. Interestingly, this study revealed that, close to neutral pH, the effect of cation concentration on structure formation is highly relevant (there is a clear interaction between the two variables), while ion type has no significant effect. In conclusion, experimental design and multivariate data analysis applied to the study of non-canonical DNA structures and of the factors affecting their stability may be useful to those researchers working on such complex structures.

Funding information This work was supported by a grant from Regione Campania-POR Campania FESR 2014/2020 (Project No. B61G18000470007) and by the Italian Association for Cancer Research (IG No. 18695 to AR and No. 16730 to BP).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Bacolla A, Wells RD. Non-B DNA conformations as determinants of mutagenesis and human disease. Mol Carcinog. 2009;48(4):273–85.
2. van de Sande J, Ramsing N, Germann M, Elhorst W, Kalisch BW, von Kitzing E, et al. Parallel stranded DNA. Science. 1988;241(4865):551–7.
3. Guéron M, Leroy J-L. The i-motif in nucleic acids. Curr Opin Struct Biol. 2000;10(3):326–31.
4. Gajarský M, Živković ML, Stadlbauer P, Pagano B, Fiala R, Amato J, et al. Structure of a stable G-hairpin. J Am Chem Soc. 2017;139(10):3591–4.
5. Cerofolini L, Amato J, Giachetti A, Limongelli V, Novellino E, Parrinello M, et al. G-triplex structure and formation propensity. Nucleic Acids Res. 2014;42(21):13393–404.
6. Neidle S. Quadruplex nucleic acids as targets for anticancer therapeutics. Nat Rev Chem. 2017;1:41.

7. Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP, Balasubramanian S. High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat Biotechnol. 2015;33(8):877–81.
8. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res. 2007;35(2):406–13.
9. Zhao J, Bacolla A, Wang G, Vasquez KM. Non-B DNA structure-induced genetic instability and evolution. Cell Mol Life Sci. 2010. https://doi.org/10.1007/s00018-009-0131-2.
10. Salvati E, Zizza P, Rizzo A, Iachettini S, Cingolani C, D'Angelo C, et al. Evidence for G-quadruplex in the promoter of vegfr-2 and its targeting to inhibit tumor angiogenesis. Nucleic Acids Res. 2014;42(5):2945–57.
11. Zizza P, Cingolani C, Artuso S, Salvati E, Rizzo A, D'Angelo C, et al. Intragenic G-quadruplex structure formed in the human CD133 and its biological and translational relevance. Nucleic Acids Res. 2016;44(4):1579–90.
12. Gehring K, Leroy JLJ-L, Gueron M, Guéron M. A tetrameric DNA structure with protonated cytosine-cytosine base pairs. Nature. 1993;363(6429):561–5.
13. Amato J, Iaccarino N, Randazzo A, Novellino E, Pagano B. Noncanonical DNA secondary structures as drug targets: the prospect of the i-motif. ChemMedChem. 2014;9(9):2026–30.
14. Zhou J, Wei C, Jia G, Wang X, Feng Z, Li C. Formation of i-motif structure at neutral and slightly alkaline pH. Mol BioSyst. 2010;6(3):580–6.
15. Rajendran A, Nakano S, Sugimoto N. Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH. Chem Commun. 2010;46(8):1299.
16. Dzatko S, Krafcikova M, Hänsel-Hertsch R, Fessl T, Fiala R, Loja T, et al. Evaluation of the stability of DNA i-motifs in the nuclei of living mammalian cells. Angew Chem Int Ed. 2018;57(8):2165–9.
17. Zeraati M, Langley DB, Schofield P, Moye AL, Rouet R, Hughes WE, et al. I-motif DNA structures are formed in the nuclei of human cells. Nat Chem. 2018;10(6):631–7.
18. Leardi R. Experimental design in chemistry: a tutorial. Anal Chim Acta. 2009;652(1–2):161–72.
19. Pagano A, Iaccarino N, Abdelhamid MAS, Brancaccio D, Garzarella EU, Di Porzio A, et al. Common G-quadruplex binding agents found to interact with i-motif-forming DNA: unexpected multi-target-directed compounds. Front Chem. 2018;6:281.
20. Cantor CR, Warshaw MM, Shapiro H. Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleolides. Biopolymers. 1970;9(9):1059–77.
21. Savorani F, Tomasi G, Engelsen SB. icoshift: a versatile tool for the rapid alignment of 1D NMR spectra. J Magn Reson. 2010;202(2):190–202.
22. Bro R, Smilde AK. Principal component analysis. Anal Methods. 2014;6(9):2812–31.
23. Box G, Hunter JS, Hunter W. Statistics Experimenters. 2nd ed. Wiley; 2009.
24. Ebrahimi-Najafabadi H, Leardi R, Jalali-Heravi M. Experimental design in analytical chemistry—Part I: theory. J AOAC Int. 2014;97(1):3–11.
25. Phan AT, Guéron M, Leroy JL. The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. J Mol Biol. 2000;299(1):123–44.
26. Eriksson L, Johansson E, Kettaneh-Wold N, Wikström C, Wold S. Design of experiments, principles and applications. 3rd ed. Umea: Umetrics Academy; 2009.

Paper II

Effects of sequence and base composition on the CD and TDS profiles of i-DNA

Iaccarino N., Cheng M., Qiu D., Pagano B., Amato J., Di Porzio A., Zhou J., Randazzo A., Mergny JL.

Angewandte Chemie, 2021, Vol. 60 (18)

How to cite: *Angew. Chem. Int. Ed.* **2021**, *60*, 10295 – 10303
International Edition: doi.org/10.1002/anie.202016822
German Edition: doi.org/10.1002/ange.202016822

DNA Motifs

Effects of Sequence and Base Composition on the CD and TDS Profiles of i-DNA

Nunzia Iaccarino[+], Mingpan Cheng[+], Dehui Qiu, Bruno Pagano, Jussara Amato, Anna Di Porzio, Jun Zhou, Antonio Randazzo, and Jean-Louis Mergny*

Abstract: *The i-motif DNA, also known as i-DNA, is a non-canonical DNA secondary structure formed by cytosine-rich sequences, consisting of two intercalated parallel-stranded duplexes held together by hemi-protonated cytosine–cytosine[+] (C:C[+]) base pairs. The growing interest in the i-DNA structure as a target in anticancer therapy increases the need for tools for a rapid and meaningful interpretation of the spectroscopic data of i-DNA samples. Herein, we analyzed the circular dichroism (CD) and thermal difference UV-absorbance spectra (TDS) of 255 DNA sequences by means of multivariate data analysis, aiming at unveiling peculiar spectral regions that could be used as diagnostic features during the analysis of i-DNA-forming sequences.*

Introduction

In the last decades, a variety of DNA secondary structures other than the canonical Watson–Crick duplex, have been documented. Such structural polymorphism depends on sequence, hydration, ions and/or ligands and superhelical stress; it occurs during biological processes such as replication and transcription, thus having an impact on genetic stability.[1] Non-canonical DNA structures include hairpins, cruciforms, triplexes, G-quadruplexes (G4s), and i-motifs (i-DNAs).[2–7] i-DNA was first observed in 1993 for the hexamer sequence d(TCCCCC) under acidic conditions.[8] It consists of two intercalated parallel-stranded duplexes held together by hemi-protonated cytosine–cytosine[+] (C:C[+]) base pairs (Figure 1 A).[8,9] i-DNA formation at physiological pH has been recently reported.[10,11] Moreover, the generation of an antibody able to detect i-DNA has proved its presence in the nucleus of human cells, arguing for regulatory roles in the genome, e.g., at proto-oncogene promoters and telomeres,[12] making this DNA structure a potential target for anticancer therapy. Putative i-DNA-forming sequences occur in C-rich strands complementary to G-rich regions that may form G4s. However, if the G4 counterpart folding conditions have already been extensively studied, i-DNA's optimal features for its formation near physiological pH in vitro are still under investigation. In fact, several studies have been conducted recently to better understand the influence of external conditions, such as presence of metals[13,14] or ligands,[15–17] molecular crowding,[11,18,19] cation type and ionic strength,[20] on i-DNA formation. However, i-DNA stability also depends on sequence composition. A typical formula of an intramolecular i-DNA-forming sequence is $(C_nX_N)_3C_n$, where X can be either a C or non-C (T, A, G); the presence of four C tracts (C_n) allows the generation of a C-stem, while the three spacers (X_N), connect the four cytosine tracts (C-tracts) and form the loops (Figure 1 B).

Much attention has been paid to the influence of the loops' length and composition as well as to the length of the C-tracts. In general, it was found that thymines confer a higher i-DNA stability compared to other non-C deoxynucleotides.[21,22] Very recently, the Vorlickova's group reported a systematic investigation of sequence requirements for i-

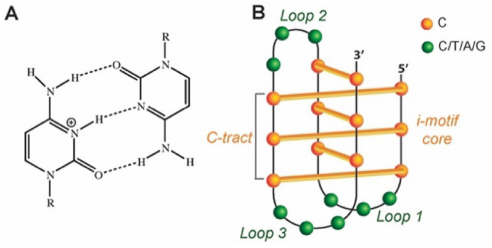


Figure 1. A) Hemi-protonated cytosine–cytosine[+] (C:C[+]) base pair. B) Schematic representation of an i-motif structure.

[*] Dr. N. Iaccarino,[+] Prof. B. Pagano, Prof. J. Amato, A. Di Porzio, Prof. A. Randazzo
Department of Pharmacy
University of Naples Federico II
Via D. Montesano 49, 80131 Naples (Italy)
E-mail: antonio.randazzo@unina.it

Dr. M. Cheng,[+] D. Qiu, Dr. J. Zhou, Dr. J.-L. Mergny
State Key Laboratory of Analytical Chemistry for Life Science
School of Chemistry & Chemical Engineering
Nanjing University
Nanjing 210023 (China)

Dr. M. Cheng,[+] Dr. J.-L. Mergny
ARNA Laboratory, Université de Bordeaux
Inserm U 1212, CNRS UMR5320, IECB
33607 Pessac (France)

Dr. J.-L. Mergny
Laboratoire d'Optique et Biosciences, Ecole Polytechnique
CNRS, INSERM, Institut Polytechnique de Paris
91128 Palaiseau (France)

[+] These authors contributed equally to this work.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202016822.

DNA formation. They found that the lower number of residues are present in the spacers, the more i-DNA is destabilized. This is due to the loss of $C:C^+$ base pairs as several Cs need to be incorporated into the loops to compensate for the short linkers.[23]

In the companion paper of this investigation, some of us have explored the simultaneous variation of both C-tract length and loop arrangements.[24] In particular, by analyzing a first set of 180 different DNA sequences (Table S1), the contribution of C-tracts and spacer length on i-DNA stability was explored. The general formula of the majority of i-DNA-forming oligos employed in that study is $C_{(3-6)}T_{(1-6)}C_{(3-6)}T_{(1-6)}C_{(3-6)}T_{(1-6)}C_{(3-6)}$. We investigated sequences with cytosine tracts of equal length made of 3 up to 6 Cs and three spacers made of 1 up to 6 thymines, in such a combination to have from 4 to 12 total thymines. The samples were named according to the following rationale: a "T" was used as prefix because the three spacers were composed of thymines only; three consecutive numbers were used to describe the lengths of the three spacers in the 5′ to 3′ direction; while the suffix referred to the length of the C-tracts ("-3", "-4", "-5" or "-6"′ for C_3, C_4, C_5, and C_6, respectively). Thus, for example, the sample T124-3 corresponded to the following DNA sequence: 5′-CCCTCCCTTCCCTTTTCCC-3′. An additional set of 75 i-DNA-forming sequences was added to evaluate the effects of different spacer lengths and compositions, terminal nucleobases, and non-equal C-tracts. The nomenclature of these additional sequences, listed in Table S2, uses the same rationale employed for the previous 180 samples. In particular, a subset of 40 samples was designed to generate sequences characterized by five Cs in each C-tract and differing for: (i) the length of central spacer, ranging from seven up to fifteen Ts (4 samples); (ii) the presence of adenines in the different positions of the spacers (24 samples); (iii) the length of the first and third spacers (12 samples). A second subset of 35 samples (all variants based on the original sequence "T252-5") included: (i) sequences having As, Ts or Gs as flanking nucleobases (9 samples); (ii) sequences with an increasing number of As or Gs in the spacers (20 samples); (iii) sequences with four non-equally sized C-tracts (6 samples). Thus, overall, a total of 255 samples was employed. In particular, ultraviolet (UV) and circular dichroism (CD) spectroscopies at different pH and temperature values were used to evaluate the thermal and pH stability of each i-DNA, as described in the companion paper.[24]

In the present work, we analyze the CD spectral profiles and UV thermal difference spectra (TDS) of these 255 samples by means of multivariate data analysis to detect hidden but potentially informative bands in the spectra of the i-DNA-forming sequences. Indeed, to our knowledge, only the well-known i-DNA characteristic bands of the CD (positive at 288 and negative at 264 nm) and TDS (positive at 240 and negative at 295 nm) spectra have been considered so far, limiting the informative power of the CD and UV-absorbance spectroscopies.

Results and Discussion

Considering the large number of spectra to be compared and the very high number of variables (i.e., the intensity at each sampled wavelength) they contain, a plain visual inspection of the TDS and CD spectra may not be sufficient to reveal hidden information in these spectra. However, this huge amount of data can be handled by multivariate data analysis. In particular, the Principal Component Analysis (PCA) is an unsupervised multivariate method that allows the reduction of the dimensionality of a data set, providing a visual representation of the major variance in the data.[25] Particularly, the original variables are transformed into a smaller set of new uncorrelated variables, called principal components (PCs), which are ordered according to the variance they explain (PC1 explains the greatest variance, PC2 contains the second greatest variance, and so on). The principal components are visualized in two plots, termed "scores plot" (where the samples appear close to each other when they are similar, and apart when they are dissimilar) and "loadings plot" (which highlights the variables responsible for the separation of the samples along each PC, in our case the spectral regions). Therefore, this method allows the clustering of the samples based on their similarities and the identification of the spectral regions of the spectra that are characteristic for each cluster. In the following two paragraphs the multivariate data analysis of the CD and TDS spectra is reported.

Circular Dichroism

We decided to first consider the 180 samples, listed in Table S1, characterized by cytosine tracts of equal length and thymine-based spacers. The characteristic CD profile of an i-DNA structure having a positive and a negative band, respectively, at 288 and 264 nm, is clearly distinguishable at acidic pH values (Figure S1). The intensity values of each data point of the 180 CD spectra acquired at pH 5.0 were used as variables in a PCA (Figure 2) that produced three meaningful principal components (PCs). Coloring the samples in the PC1/PC2 score plot (Figure 2 A) according to the numbers of Cs in the C-tract reveals the first information of this analysis. In fact, the samples turn out to be separated along PC1, where those having a higher number of Cs lie on the right side of the plot, and those having shorter C-tracts lie on the left part. The PC1 loading plot (Figure 2 B) reports the variables mainly responsible for the separation along the first principal component. Interestingly, this loading plot closely resembles the CD spectrum of an i-DNA structure. This result indicates that samples with longer C-tracts (that are in the positive region of the PC1) are characterized by more intense CD signals at 288 and 264 nm compared to those having shorter C-tracts. This is also evident from the superimposition of the 180 CD spectra colored according to C-tract length (Figure S2). This observation is not surprising since these bands are directly correlated to the number of $C:C^+$ base pairs in the i-DNA structure.[23,26] Indeed, calculating the Pearson's correlation coefficient r, a very good positive correlation is

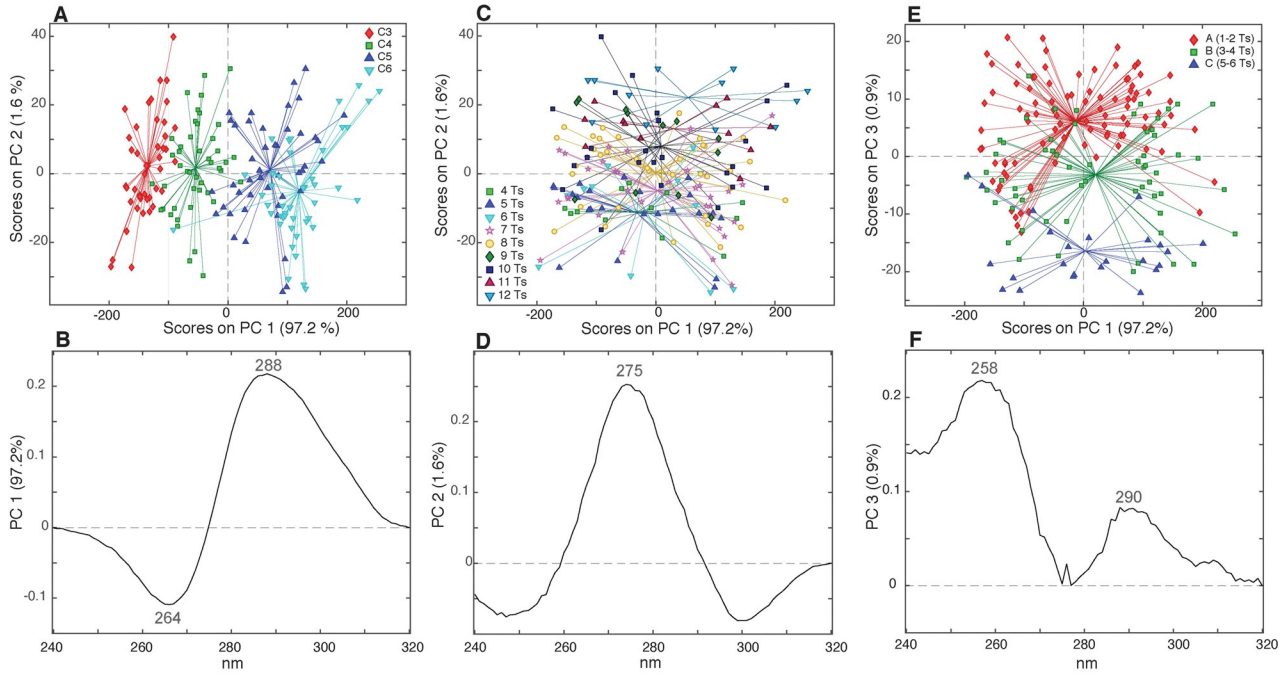


Figure 2. PC1/PC2 score plots of the PCA model calculated using the 180 CD spectra (acquired at pH 5.0) colored according to A) C-tract length and C) total number of Ts. E) PC1/PC3 score plot colored according to the length of the central spacer. Loading plots of B) PC1, D) PC2, and F) PC3.

found between the CD signal intensity at 288 nm and the number of Cs in the C-tract ($r = 0.90$). Interestingly, if the samples are colored according to the total number of Ts (Figure 2C), they turn out to be distributed along PC2. Particularly, the samples characterized by higher number of Ts are placed in the top of the score plot, while samples having a lower number of Ts lie in the bottom of the plot. The PC2 loading plot (Figure 2D) shows that the more Ts are present, the more intense is the signal at 275 nm. This can be explained considering that the CD spectrum of a single-stranded poly(dT) shows a positive band at 275 nm (Figure S3), thus the more Ts are in the sequence, the higher is the signal intensity at 275 nm, independently from the presence of a secondary structure. Indeed, also in this case, a good correlation coefficient between the signal intensity at 275 nm and the total number of Ts in the spacers is found ($r = 0.65$). Valuable information can also be retrieved by analyzing the PC1/PC3 score plot.

Indeed, an interesting trend along PC3 is observed when the samples are colored according to the length of the central spacer. This is particularly evident if the samples are grouped in three classes: "class A" for samples containing 1 or 2 Ts in the central spacer; "class B" for 3 or 4 Ts; and "class C" for 5 or 6 Ts (Figure 2E). Unfortunately, the PC3 loading plot (Figure 2F) is difficult to be interpreted, as it is characterized by two positive bands at 258 and 290 nm, apparently not related to any i-DNA bands or base composition of the DNA. The profile reported in the PC3 loading plot is ascribable to a combination of effects, maybe related, to different extents, to sequence composition and conformational features.

In order to verify this hypothesis, we computed another PCA on samples having the same nucleotide composition, but different residue sequences. By way of example, we show the results obtained by selecting the nine samples having twenty cytosines (five Cs in each C-tract) and seven Ts (differently distributed in the three spacers) namely T115-5/T151-5/T511-5, T223-5/T232-5/T322-5, T331-5/T313-5/T133-5. In order to improve the reliability of the multivariate analysis, we decided to increase the size of this data set by also including the spectra of the selected samples acquired at pH 5.25 and 5.50, after having carefully checked the irrelevance of the slight pH variation on CD spectra (Figure S4A). Thus, a total of 27 spectra (9 samples, whose CD spectra have been acquired at 3 different pH values) were used to compute a new PCA, and the resulting PC1/PC2 score and loading plot are showed in Figures 3A and B, respectively. The variance on PC1 was again explained by the bands at 288 nm and 264 nm (Figure S4B), however, in this case, the distribution of the samples is obviously not related to the content of Cs, since all

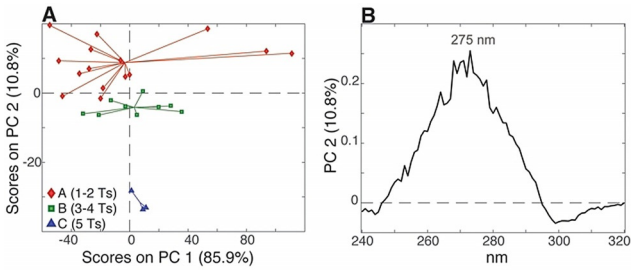


Figure 3. A) PC1/PC2 score plot of the PCA model calculated using the 9 samples, having a total of twenty Cs and seven Ts, acquired at pH 5.00, 5.25, and 5.50 (27 CD spectra in total), colored according to the central spacer length; B) relative PC2 loading plot.

the samples have the same base composition. Most probably, the variance observed in PC1 has to be ascribed simply to an intrinsic uncertainty of the DNA concentration and extinction coefficient values of the samples. On the other hand, the samples turn out to be distributed according to the length of the central spacer along PC2. In this case, the PC2 loading plot (Figure 3B) appears different from the one obtained in the previous analysis (Figure 2F). Indeed, the signal intensity at 275 nm seems to carry the information about the central spacer length; in particular, samples with shorter central spacer show more intense CD signal at 275 nm compared to those with longer central spacer. This result confirms that the PC3 loading plot of the analysis performed on all the samples (Figure 2F) is probably polluted by the information related to the total number of thymines in the sequences that also affects the signal around 275 nm. To shed light on this point, we decided to calculate the correlation coefficients between the CD signal intensity at 275 nm and the central spacer length ($r = -0.17$). The expected low correlation improves to a value of $r = -0.61$ when the CD signal intensity at 275 nm is divided by the total number of Ts of each sample. This confirms that the 275 nm wavelength hides both information. Furthermore, the negative sign of the correlation agrees with the PCA outcome, indicating that the longer is the central spacer the less intense is the band at 275 nm.

To better understand, in practice, how the chemical composition of the samples and the length of the central spacer affect the CD profile, it is useful to look at the superimposition of CD spectra of some samples having the same C-tract length and a different number of Ts and samples with the same number of Cs and Ts, but different length of the central spacer. By way of example, the comparison between the CD spectra of T112-4 and T336-4, which have the same number of Cs and different number of Ts, is reported in Figure 4A. The increased intensity around 275 nm, expected for the sample having a higher number of Ts (T336-4), turns into an overall shift of the spectrum towards shorter wavelengths, while the relative intensities of the two bands at 264 and 288 nm remain basically unchanged. In contrast, the comparison of CD spectra of samples that have the same number of cytosines and thymines, but different length of the central spacers (T116-5 and T161-5) (Figure 4B), shows that the decrement of the intensity of the signal around 275 nm for the sample having the longer central spacer (T161-5) turns

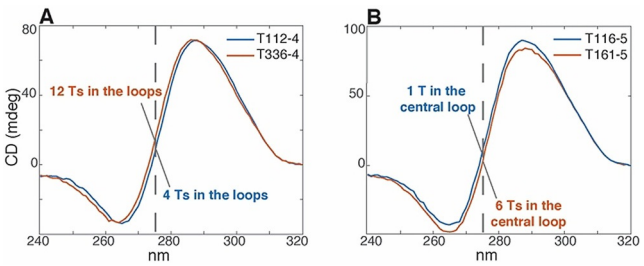


Figure 4. Superimposition of CD spectra of A) samples having the same C-tract length with different number of Ts (T112-4 vs. T336-4), B) samples having the same chemical composition but different central spacer lengths (T116-4 vs. T161-4). The dashed line is centered at 275 nm.

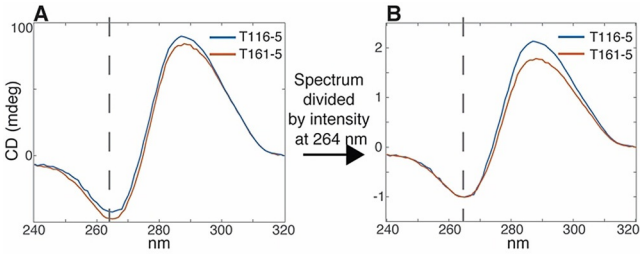


Figure 5. Superimposition of CD spectra of samples having the same chemical composition but with different central spacer length (T116-5 vs. T161-5) A) before and B) after dividing the spectra by their intensity value at 264 nm. The dashed line is centered at 264 nm.

into a general shift of the spectrum towards longer wavelengths.

Interestingly, in this case, a change in the relative intensities of the bands at 264 and 288 nm is also observable. In order to better visualize this, we decided to normalize each data point of the CD spectra by the intensity of the signal at 264 nm (Figure 5).

This basically provides normalized CD spectra whose bands' intensities are no more related to the number of Cs or Ts. Such normalization was applied to the 180 CD spectra and the new data set was submitted to a PCA (Figure 6).

As expected, the major variance (PC1) is no more related to the length of the C-tract, but it is related to the length of the central spacer (Figure 6A): indeed, samples having 5–6 nucleobases in the central spacer are mainly located in the left part of the plot, while samples having 1 or 2 Ts in their central spacer are placed on the right.

Interestingly, some samples having 1–2 Ts in the central spacer and three Cs in the C-tracts (T211-3, T112-3, T113-3, T311-3, T411-3, T114-3, T611-3, T116-3, T511-3, T115-3, T122-3, T121-3, T212-3) do not follow the general trend observed along PC1 (Figure 6A, inset). The reason of this apparent anomaly may be ascribed to the formation of bimolecular i-DNA structures, as proposed by Vorlickova's group.[23] Indeed, they observed that reducing the length both of the first and third, or the second and the third spacers, bimolecular i-DNAs are preferentially formed, and this only happens when the C-tract is made of three Cs.

As suggested by the PC1 loading plot (Figure 6B), samples having longer central spacer are characterized by a low intensity ratio between the bands at 288 and 264 nm. The reason of this can be explained taking into account the structural requisites for the formation of an i-DNA structure. In particular, the central spacer is often responsible for the formation of a loop that spans the major groove of the i-DNA (see companion paper[24]) and that this loop requires at least 3 residues.[23] If the central spacer contains a lower number of residues, the i-DNA structure can be formed in any case using some Cs of the adjacent C-tracts. In this case, a lower number of $C:C^+$ pairs are formed and the structure will contain some unpaired Cs. Obviously, the CD spectrum of the sample will proportionally contain information about both paired and unpaired Cs in the structure. As already mentioned, the $C:C^+$ base pairs in the stem provide a CD spectrum having a negative and a positive band at 264 and 288 nm, respectively,

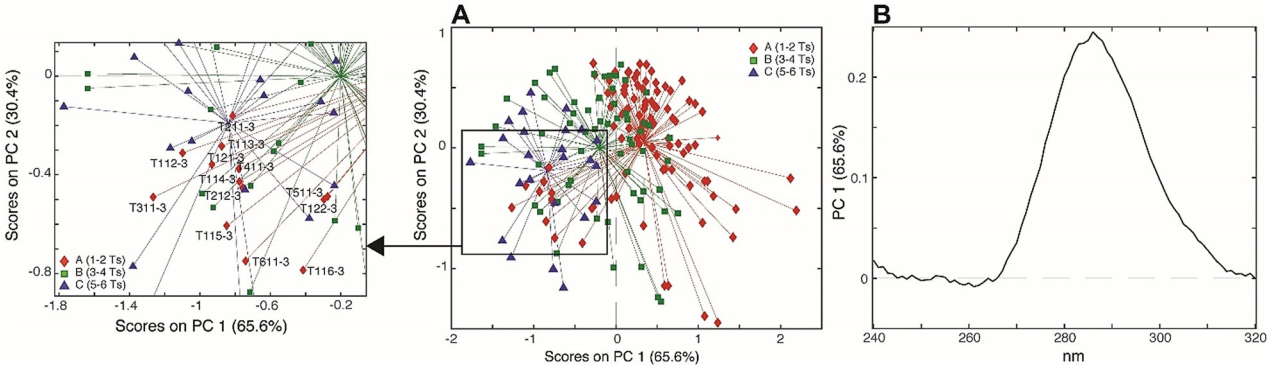


Figure 6. A) PC1/PC2 score plot and relative inset of the PCA model calculated using the 180 CD spectra (normalized by the signal intensity at 264 nm) colored according to the length of the central spacer; B) PC1 loading plot.

while cytosines not involved in the pairing are characterized by a CD spectrum having only a positive band centered around 275 nm.[27] As a matter of fact, when both type of Cs contribute to define the general appearance of the CD spectrum, the positive band of the unpaired Cs is summed to the negative band at 264 nm of the paired ones, reducing in this way the intensity of this latter band, while the intensity of the positive band at 288 nm is further increased. Therefore, overall, the relative intensities of the two bands changes proportionally to the number of unpaired Cs in the i-DNA structure. Hence, the higher is the number of unpaired Cs, the higher is the ratio between the intensity of the bands at 288 and 264 nm.

In order to verify the robustness of these findings, we calculated the ratio between the intensity of the positive band at 288 nm and the intensity of the negative band at 264 nm for all the CD spectra (not normalized) employed in the study. A graphical representation of the calculated values is reported in the bar graph in Figure 7. This bar graph confirms that the higher is the central spacer length, the lower is the intensity ratio regardless of the number of Cs in the stem. The exception to this general "rule" is represented by sequences having very short spacers (T112, T121, T211) and samples having a spacer combination where two spacers are longer than the third one (T221, T212, T122, T331, T313, T133, T332, T323, T332).

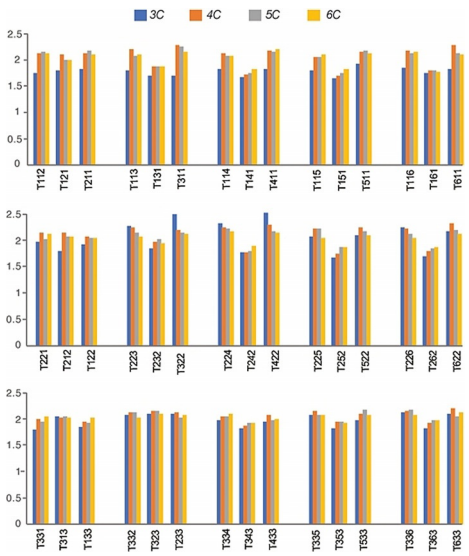


Figure 7. Graphic representation of the maximum (288 nm)/minimum (264 nm) ratio for the 180 CD spectra. Blue, orange, gray, and yellow represent C-tracts characterized by 3, 4, 5, and 6 Cs, respectively.

In order to evaluate if the findings reported for the first 180 sequences, characterized by equally sized cytosine tracts and thymine-based spacers, were still valid for a more heterogeneous set of i-DNA-forming sequences, we decided to perform multivariate data analyses of the CD spectra of 75 additional sequences divided in two subsets of 40 and 35 samples (Table S2), respectively.

The PCA computed on the subset made of 40 samples (characterized by five Cs in each C-tract and by different lengths and nucleotide composition of the spacers) generated a score plot where the samples are distributed along PC1 according to the length of the central spacer. In particular, the samples having a very short central spacer are characterized by lower intensities i-DNA bands around 264 and 288 nm, compared to the samples having a longer spacer, suggesting that a short central spacer is detrimental to the formation of i-DNA structures (Figures S5 and S6).

However, it should be noted that one sample ("T1151-5"), characterized by fifteen thymines in the central spacer, is clustered together with the samples having a single nucleotide in this central position. This observation suggests that a particularly long central spacer is also detrimental for the formation of i-DNA. This is in agreement with the pH and thermal stabilities reported in the companion paper.[24]

Then, the CD spectra of these 40 samples were normalized by the signal intensity at 264 nm (Figure S7) and the resulting spectra were submitted to PCA. As observed for the previously normalized 180 CD spectra, the samples having a longer central spacer are characterized by a lower ratio between the intensity of the bands at 288 and 264 nm. Interestingly, the T1151-5 sample is now clustered along those having a long central spacer, indicating that in this case, as expected, formation of the central loop involves this very long spacer and does not require the unpairing of $C:C^+$ base pairs (Figure S8). Thus, this additional data set allows to corroborate the fact that, independently from their composition, short and very long central spacers are detrimental for the

formation of i-DNA structures and that, when the central spacer is shorter than three residues, the ratio between the signal intensity at 288 nm and 264 nm increases.

The same analysis has also been performed on the subset made of 35 samples (Figure S9) listed in Table S2. In particular, these samples were analyzed in three subgroups to evaluate the impact on the i-DNA formation of three main sequence modifications: (i) addition of flanking bases, (ii) increasing number of purines in the spacers and (iii) non-equal C-tracts. In particular, the CD spectra of 9 samples having Ts, As, or Gs at the 5′- or 3′-ends (or both) were submitted to PCA. The PC1 loading plot indicates that the samples are distributed according to their propensities to form i-DNA. In particular, two samples (TT252-5 and TT252-5T), both characterized by a thymine at the 5′-end, turn out to be characterized by higher intensities of the i-DNA bands at 264 and 288 nm. We speculate that this is due to the formation of a T:T base pair between the thymine at the 5′-end and one of the thymines present in the central spacer (Figures S10A and S10B). The samples are also distributed along PC2 according to the Gs content (increasing number of Gs from the top to the bottom) and As content (increasing number of As from the bottom to the top) (Figures S10C and S10D). The PC2 loading plot indicates that the more Gs and, in turn, the less As, are in the sequence, the lower is the intensity of the band around 270–280 nm and the higher is the intensity of the band around 250–255 nm (Figure S10E). These spectral regions perfectly agree with the characteristic CD bands of a single-stranded poly(dG)[27] and a single-stranded poly-(dA).[28] Thus, the PC2 basically explains the contributions of the Gs and As to the i-DNA CD spectrum. Therefore, the analysis of this first subset of samples suggests that the addition of flanking residues to the i-DNA forming sequence may have effects on the CD spectrum both for the change in chemical composition of the samples and also for the different propensity of the sample to form an i-DNA structure, especially for the samples having an additional T at the 5′-end.

The CD spectra of twenty samples belonging to the second subgroup of samples, which contains an increasing number of As or Gs in the spacers, were also submitted to PCA. From the PC1/PC2 score plot (Figure S11A), we first observed the presence of two outliers (A252-5 and G252-5) that bothered the interpretation of the data (see discussion in the Supporting Information, Figure S11). Thus, a new PCA without these two samples was computed. The resulting PC1/PC2 score plot (Figure S12A) shows that samples are distributed according to the number of Gs present in the spacers along PC1. The PC1 loading plot (Figure S12B) reveals that the more Gs are in the spacers, the less intense are the bands around 250–265 nm and 275–300 nm. These bands are wider than those observed for the samples having Gs as flanking sequences (respectively, 250–255 nm and 270–280 nm) and include the wavelength typical of the i-DNA structure (264 and 288 nm). Therefore, this observation could be explained in two ways: (i) the presence of Gs in the sequence may favor the formation of G:C base pairs at the expense of the number of C:C$^+$ pairs, so that the bands of i-DNA structure decrease in intensity; (ii) since the presence of Gs in the sequence naturally increases the positive band at around 255 nm and the negative band at 278 nm,[27] these bands are summed to the bands of the i-DNA structure generating a decrease of their intensities. Instead, we were not able observe a clear contribution of the As (Figure S13). Thus, the analysis of this second subset of samples suggests that the complete substitution of the thymines with purines changes the appearance of the i-DNA CD spectrum, suggesting the formation of additional DNA secondary structures in solution. Instead, when few thymines are substituted with guanines, the i-DNA general spectral profile is still observable even though its characteristic bands are less intense compared to the same sequence having only thymine-based spacers (Figure S14). Also, we found that peculiar bands (255 and 278 nm) are indicative of the content of Gs, in agreement with the observations made for the first subgroup of samples.

Finally, the CD spectra of the last subgroup (six samples) characterized by samples having non-equally sized C-tracts (designed to obtain an odd number of C:C$^+$ pairs—Table S2) was also analyzed by PCAs. Interestingly, all observations retrieved from the analysis of the previous 180 samples are perfectly applicable to this sample subset (Figure S15).

The entire data set of CD spectra acquired at pH 7.0 was also analyzed. As discussed in the Supporting Information (Figures S16–S18), the analysis confirmed the absence of i-DNA structure for the majority of the analyzed samples and revealed the contribution of Cs, Ts, As and Gs to the spectrum. Interestingly, the presence of flanking bases (in particular adenines) seems to induce the i-DNA structure at neutral pH.

Thermal Difference Spectra (TDS)

The same approach used for CD was employed to study the TDS of the 255 samples investigated in this study. A TDS is calculated by subtracting the UV spectrum of the folded structure, at low temperature, from that of the unfolded structure, at high temperature. The resulting profile is unique and can be used to obtain a specific signature for DNA secondary structures.[29] At pH 5.0, the TDS profiles are in

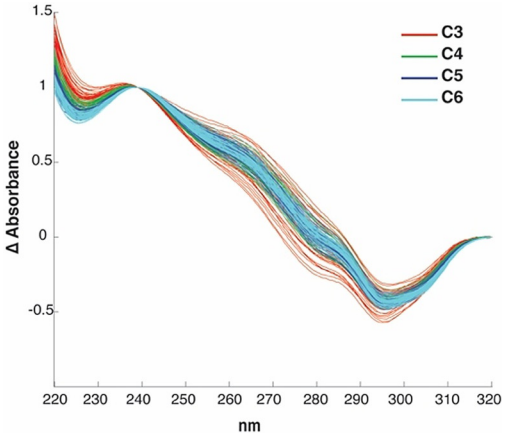


Figure 8. Thermal difference spectra (TDS) of the 180 samples acquired at pH 5.0. The different colors indicate distinct lengths of the C-tract.

perfect agreement with the typical i-DNA signature (Figure 8) with a positive peak at 240 nm (used to normalize the spectra) and a negative one at 295 nm.[29] As for CD spectra, we first analyzed the initial set of 180 sequences, characterized by equally sized cytosine tracts and thymine-based spacers. The PCA computed on the such profiles revealed that the band around 250–265 nm is more intense when there are more Ts in the sequence (Figures 9A and B) (for a more detailed discussion see the SI and Figures S19 and S20). Moreover, the lower is the C content the higher is the band around 295–310 nm (Figures 9C and D). Then, as done for the CD analysis, we decided to get rid of the variability related to the different number of Ts. Thus, we generated a data set including only sequences with seven and eight Ts in the spacers, accounting for a total of 72 samples and a new PCA was computed. The resulting PC1/PC2 score plot shows that the sequences with the longest central spacers (5 or 6 Ts) are separated from the rest of the samples along PC2 and that they are characterized by low values of ΔA around 250–265 nm (Figures 10A and 10B).

Thus, once again, as observed in the CD data set, the information about the total number of Ts and central spacer seems hidden under the same wavelength. This observation was also mathematically verified (see Supporting Information). Then, by coloring the samples in the PC1/PC3 score plot according to the C-tract length (Figure 10C), it is possible to observe the same trend along PC3 observed considering all the sample (Figure 9C), thus confirming that samples having longer central spacers are characterized by lower ΔA around 295–310 nm (Figure 10D). These results can be easily observed comparing the TDS profiles of the samples as showed in Figure 11.

As in the case for CD, the TDS profiles of the 75 additional sequences were analyzed through PCA. As discussed in Supporting Information (Figures S21–S23), all the observations retrieved from the analysis of the 180 samples

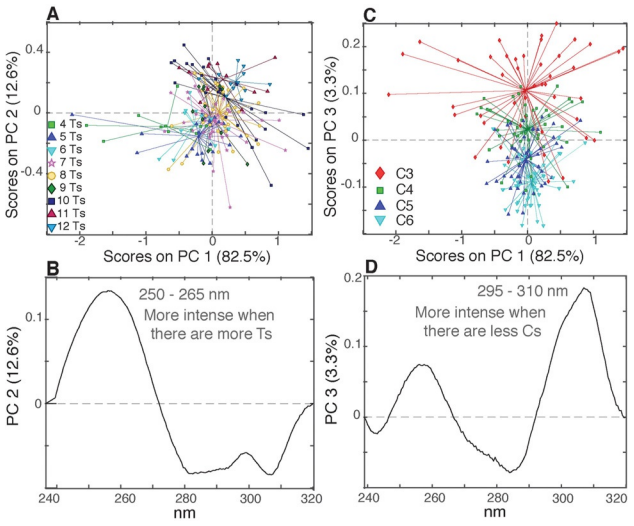


Figure 9. A) PC1/PC2 score plot of the PCA model calculated using the 180 TDS acquired at pH 5.0, colored according to the total number of Ts; B) PC2 loading plot. C) PC1/PC3 score plot colored according to the C-tract length; D) PC3 loading plot.

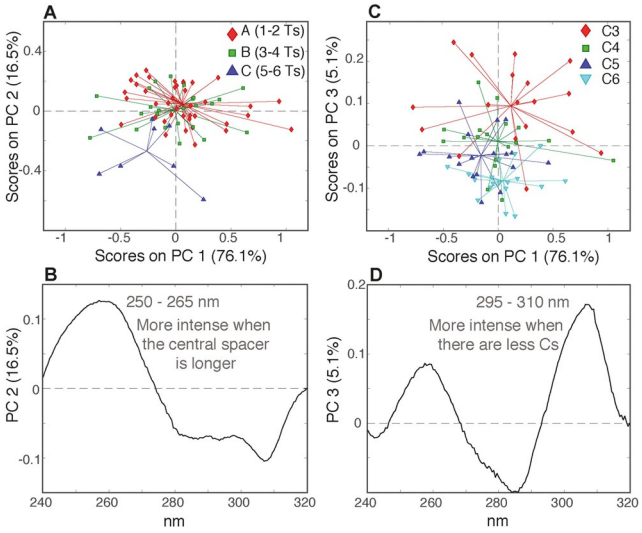


Figure 10. A) PC1/PC2 score plot of the PCA model calculated using the 72 TDS of samples having 7 and 8 Ts (acquired at pH 5.0), colored according to the number of Ts in the central spacer; B) PC2 loading plot. C) PC1/PC3 score plot colored according to the C-tract length; D) PC3 loading plot.

turned out to be perfectly applicable also to this additional subset of sequences. Unfortunately, no peculiar band could be ascribable to the presence of As or Gs. However, some samples (G252-5, TT252-5T, TT252-5, and 252-5_A6) turned out to have unprecedented TDS profiles that, for the time being, we are not able to explain.

Conclusion

In this work, multivariate data analysis was used to study the TDS and CD spectral profiles of an unprecedented large selection of i-DNA forming sequences (255 in total). This analysis reveals the impact of several kinds of sequence modifications on i-DNA formation and on the entirety of its CD spectral profile. In particular, our findings confirm that, the higher is the number of cytosines in a sequence, the higher its propensity to form i-DNA. This result is true both for sequences having an even and an odd number of C:C^+ base pairs. Moreover, our results corroborate the importance of the length of the central spacer for the i-DNA structure. Indeed, we observe that both a particularly short (i.e., one base) or long (i.e., fifteen bases) central spacers are detrimental for i-DNA. Interestingly, the presence of terminal bases (at 5′ and 3′ ends) is not detrimental for the formation of i-DNA structure. Instead, the presence of a T at the 5′ end seems to favor i-DNA formation. Moreover, the complete absence of thymines in the spacers (obtained by their substitution with purines) changes the appearance of the i-DNA CD spectrum, suggesting the formation of additional DNA secondary structures in solution. Instead, a partial substitution of some of the thymines with a few purines still allows i-DNA formation, even if to a lesser extent compared to the same sequence having only thymine-based spacers.

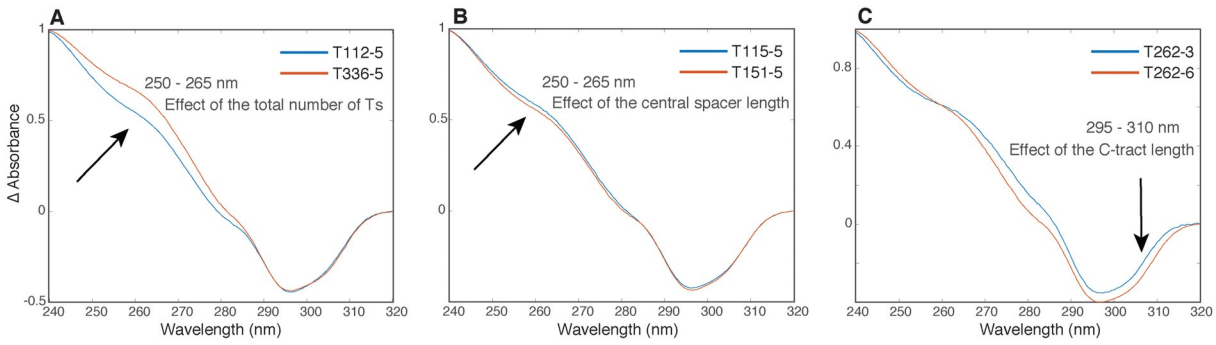


Figure 11. Superimposition of TDS of samples having A) the same C-tract length and different number of total Ts, B) the same chemical composition and different length of the central spacer, C) the same spacers compositions and different C-tract length.

Interestingly, a principal component analysis allows to detect other CD bands than those strictly related to the i-DNA (264 and 288 nm). In particular, we find peculiar informative bands that have never been reported before, related to the presence of thymines, guanines and adenines in the sequences. The intensity of the CD signal intensity at 275 nm is positively correlated with the total number of thymines, while the bands around 250–255 nm and 270–280 nm are indicative of the presence of adenines and guanines in the sequence. Moreover, the band around 275 nm is negatively correlated with the length of the central spacer. Moreover, the ratio of the intensities of the CD signals at 288 and 264 nm is negatively correlated with the length of the central spacer, when shorter than 3 residues.

The analysis of the CD spectra acquired at pH 7.0 confirmed the absence of i-DNA structure for the majority of the analyzed samples and revealed the contribution of Cs, Ts, As and Gs to spectrum. Interestingly, the presence of flanking bases (in particular adenines) seems to induce the i-DNA structure at neutral pH.

Furthermore, the multivariate analysis of the TDS data set reveals that the band around 250–265 nm is positively correlated with the total number of Ts, while it is negatively correlated with the length of the central spacer (if divided by the total number of Ts of the sequence). Moreover, the band around 295–310 nm deepens as the number of cytosines in the C-tracts increases.

Our results demonstrate that CD and TDS are much more informative for these structures than previously believed and that they can be used to retrieve interesting structural information on i-DNA.

Acknowledgements

This work was supported in part by the Italian Association for Cancer Research AIRC (IG 18695) and by Regione Campania-POR Campania FESR2014/2020 [B61G18000470007], National Natural Science Foundation of China (21977045), funding from Nanjing University (020514912216), and Fundamental Research Funds for the Central Universities (02051430210).

Conflict of interest

The authors declare no conflict of interest.

Keywords: circular dichroism · DNA · i-motif · multivariate data analysis · thermal difference spectroscopy

[1] G. Wang, K. M. Vasquez, *DNA Repair* **2014**, *19*, 143 – 151.
[2] A. Bacolla, R. D. Wells, *Mol. Carcinog.* **2009**, *48*, 273 – 285.
[3] J. van de Sande, N. Ramsing, M. Germann, W. Elhorst, B. Kalisch, E. von Kitzing, R. Pon, R. Clegg, T. Jovin, *Science* **1988**, *241*, 551 – 557.
[4] M. Guéron, J.-L. Leroy, *Curr. Opin. Struct. Biol.* **2000**, *10*, 326 – 331.
[5] M. Gajarský, M. L. Živković, P. Stadlbauer, B. Pagano, R. Fiala, J. Amato, L. Tomáška, J. Šponer, J. Plavec, L. Trantírek, *J. Am. Chem. Soc.* **2017**, *139*, 3591 – 3594.
[6] L. Cerofolini, J. Amato, A. Giachetti, V. Limongelli, E. Novellino, M. Parrinello, M. Fragai, A. Randazzo, C. Luchinat, *Nucleic Acids Res.* **2014**, *42*, 13393 – 13404.
[7] a) S. Neidle, *Nat. Rev. Chem.* **2017**, *1*, 41; b) V. Sanchez-Martin, C. Lopez-Pujante, M. Soriano-Rodriguez, J. A. Garcia-Salcedo, *Int. J. Mol. Sci.* **2020**, *21*, 8900.
[8] K. Gehring, J. L. Leroy, M. Guéron, *Nature* **1993**, *363*, 561 – 565.
[9] J. Amato, N. Iaccarino, A. Randazzo, E. Novellino, B. Pagano, *ChemMedChem* **2014**, *9*, 2026 – 2030.
[10] J. Zhou, C. Wei, G. Jia, X. Wang, Z. Feng, C. Li, *Mol. BioSyst.* **2010**, *6*, 580 – 586.
[11] A. Rajendran, S. Nakano, N. Sugimoto, *Chem. Commun.* **2010**, *46*, 1299.
[12] M. Zeraati, D. B. Langley, P. Schofield, A. L. Moye, R. Rouet, W. E. Hughes, T. M. Bryan, M. E. Dinger, D. Christ, *Nat. Chem.* **2018**, *10*, 631 – 637.
[13] H. A. Day, C. Huguin, Z. A. E. Waller, *Chem. Commun.* **2013**, *49*, 7696.
[14] S. Saxena, S. Joshi, J. Shankaraswamy, S. Tyagi, S. Kukreti, *Biopolymers* **2017**, *107*, e23018.
[15] H. A. Day, P. Pavlou, Z. A. E. Waller, *Bioorg. Med. Chem.* **2014**, *22*, 4407 – 4418.
[16] S. Fernández, R. Eritja, A. Aviñó, J. Jaumot, R. Gargallo, *Int. J. Biol. Macromol.* **2011**, *49*, 729 – 736.
[17] 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, A. Randazzo, *Front. Chem.* **2018**, *6*, 281.
[18] Y. P. Bhavsar-Jog, E. Van Dornshuld, T. A. Brooks, G. S. Tschumper, R. M. Wadkins, *Biochemistry* **2014**, *53*, 1586 – 1594.
[19] J. Zhou, G. Jia, Z. Feng, C. Li, *Chin. J. Chem.* **2010**, *31*, 309 – 311.

[20] N. Iaccarino, A. Di Porzio, J. Amato, B. Pagano, D. Brancaccio, E. Novellino, R. Leardi, A. Randazzo, *Anal. Bioanal. Chem.* **2019**, *411*, 7473 – 7479.
[21] M. McKim, A. Buxton, C. Johnson, A. Metz, R. D. Sheardy, *J. Phys. Chem. B* **2016**, *120*, 7652 – 7661.
[22] A. M. Fleming, K. M. Stewart, G. M. Eyring, T. E. Ball, C. J. Burrows, *Org. Biomol. Chem.* **2018**, *16*, 4537 – 4546.
[23] P. Školáková, D. Renčiuk, J. Palacký, D. Krafčík, Z. Dvořáková, I. Kejnovská, K. Bednářová, M. Vorlíčková, *Nucleic Acids Res.* **2019**, *47*, 2177 – 2189.
[24] M. Cheng, D. Qiu, L. Tamon, E. Maturová, P. Víšková, S. Amrane, A. Guédin, J. Chen, L. Lacroix, H. Ju, L. Trantírek, A. B. Sahakyan, J. Zhou, J. L. Mergny, *Angew. Chem. Int. Ed.* **2021**, https://doi.org/10.1002/anie.202016801; *Angew. Chem.* **2021**, https://doi.org/10.1002/ange.202016801.
[25] a) H. Hotelling, *J. Educ. Psychol.* **1933**, *24*, 417 – 441; b) J. Jaumot, R. Eritja, S. Navea, R. Gargallo, *Anal. Chim. Acta* **2009**, *642*, 117 – 126.
[26] J. L. Mergny, L. Lacroix, C. Hélène, X. Han, J. L. Leroy, *J. Am. Chem. Soc.* **1995**, *117*, 8887 – 8898.
[27] D. M. Gray, F. J. Bollum, *Biopolymers* **1974**, *13*, 2087 – 2102.
[28] J. Greve, M. F. Maestre, A. Levin, *Biopolymers* **1977**, *16*, 1489 – 1504.
[29] J. L. Mergny, J. Li, L. Lacroix, S. Amrane, J. B. Chaires, *Nucleic Acids Res.* **2005**, *33*, e138.

Manuscript received: December 18, 2020
Accepted manuscript online: February 22, 2021
Version of record online: March 24, 2021

Paper III

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

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

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

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 $G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}$, 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.

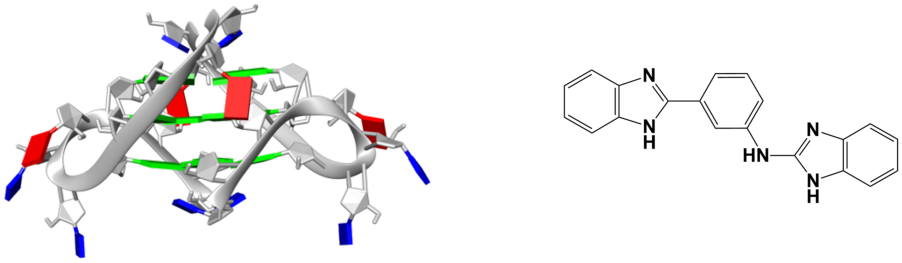
Structural analysis showed that TERRA RNA and its corresponding DNA sequences, which differ only for the presence of thymine instead of uracil bases and for deoxyribose sugars instead of ribose, can adopt different G4 topologies. For example, the 22-nucleotide-long telomeric DNA of sequence d[AGGG(TTAGGG)$_3$] 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 (ΔT_m < 3 °C) (Table S1), and thus were not further considered for the biophysical characterization of the interaction with *TERRA G4*.

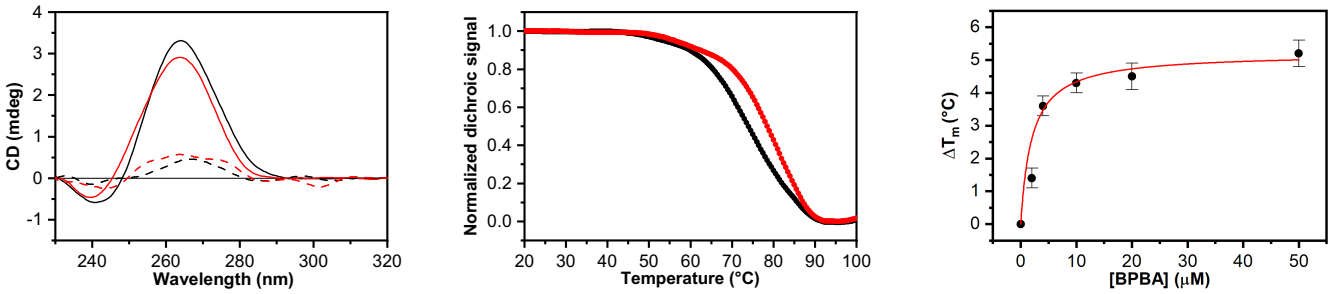


Figure 2. (**a**) CD spectra of *TERRA G4* in the absence (black line) and presence (red line) of 10 molar equiv of **BPBA** recorded at 20 and 100 °C (solid and dashed lines, respectively); (**b**) CD melting profiles of *TERRA G4* in the absence and presence (black and red dots, respectively) of 10 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.

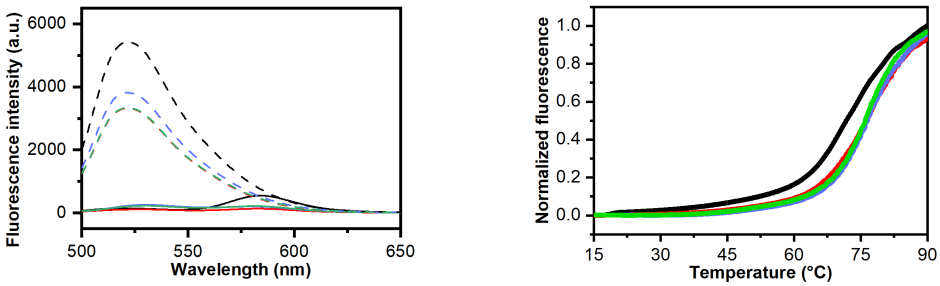
ΔT_m (°C) [1]	**Circular Dichroism (CD) Melting**						
	TERRA G4 4.5 (±0.4)	*c-kit2 G4* 18.7 (±0.3)	*c-myc G4* 9.4 (±0.3)	*Tel*$_{23}$*G4* 0.7 (±0.2)	*GSEC G4* 1.7 (±0.3)	*Bcl-2 G4* 2.9 (±0.3)	*Hrp*$_{20}$ 1.0 (±0.5)
	Förster resonance energy transfer (FRET) melting						
		F-TERRA-T 3.6 (±0.2)		*F-TERRA-T* + *Hrp*$_{27}$ (1:30) 3.7 (±0.2)		*F-TERRA-T* + *Hrp*$_{27}$ (1:100) 3.1 (±0.2)	
DC_{50} (μM)	**Fluorescent intercalator displacement (FID)—thiazole orange**						
	TERRA G4 2.4 (±0.4)	*c-kit2 G4* 3.8 (±0.6)		*c-myc G4* n.d. [2]		*Tel*$_{23}$*G4* n.d. [2]	*Hrp*$_{27}$ 4.1 (±0.5)
	Fluorescent intercalator displacement (FID)—ethidium bromide						
			TERRA G4 1.3 (±0.4)				
K_d (μM)	**Microscale thermophoresis (MST)**						
	TERRA G4 9.6 (±0.6)	*c-kit2 G4* 23.5 (±0.6)		*c-myc G4* 47.5 (±0.4)		*Tel*$_{23}$*G4* n.d. [2]	*Hrp*$_{20}$ n.d. [2]

[1] $\Delta T_m = T_{m\,(oligonucleotide+10\ ligand\ equiv)} - T_{m\,(oligonucleotide)}$. 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* (ΔT_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.

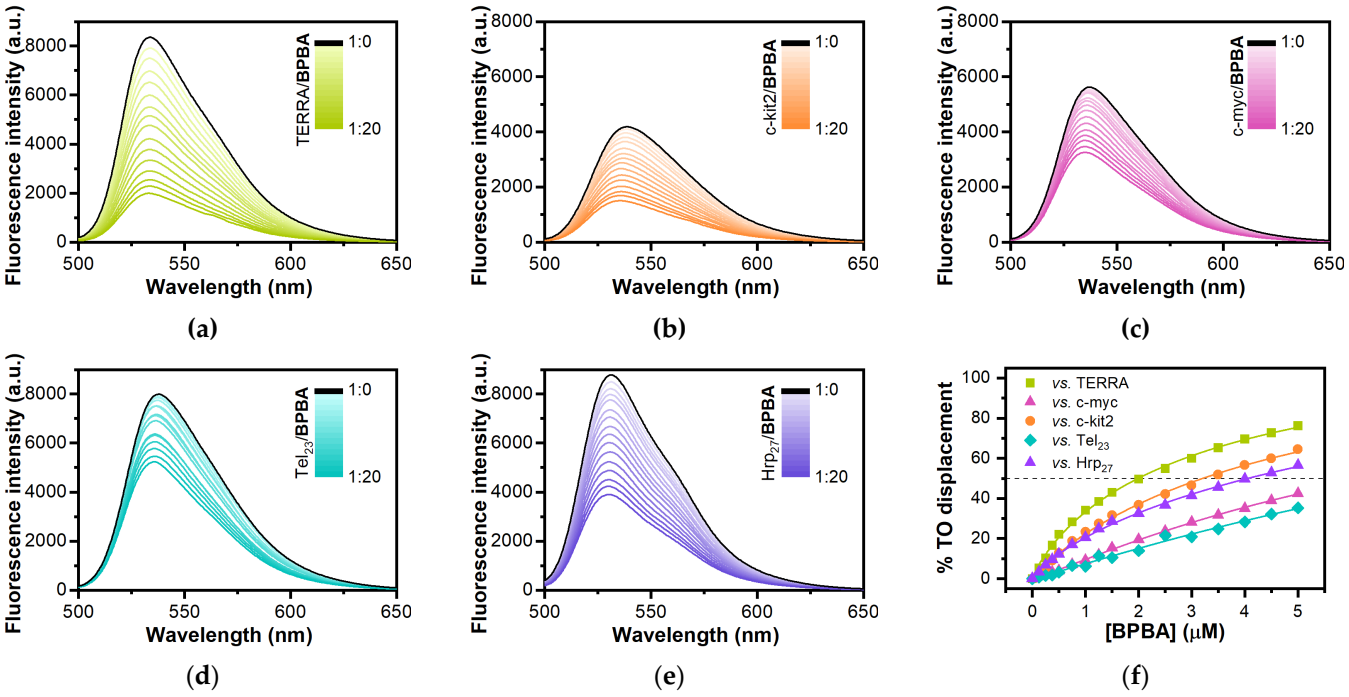
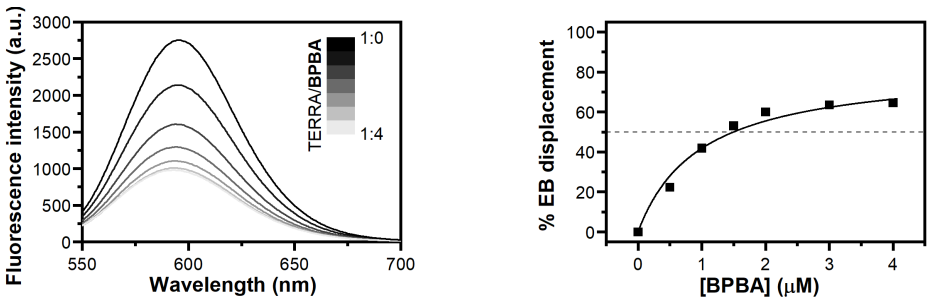


Figure 4. TO displacement titrations for (**a**) *TERRA G4*, (**b**) *c-kit2 G4*, (**c**) *c-myc G4*, (**d**) *Tel*$_{23}$ *G4*, and (**e**) *Hrp*$_{27}$ upon addition of increasing amounts of **BPBA**; (**f**) Dose–response curves from FID experiments.

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.

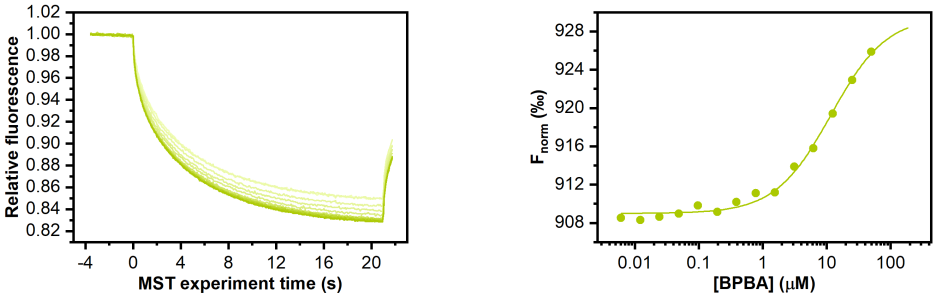


(**a**) (**b**)

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.



(**a**) (**b**)

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).

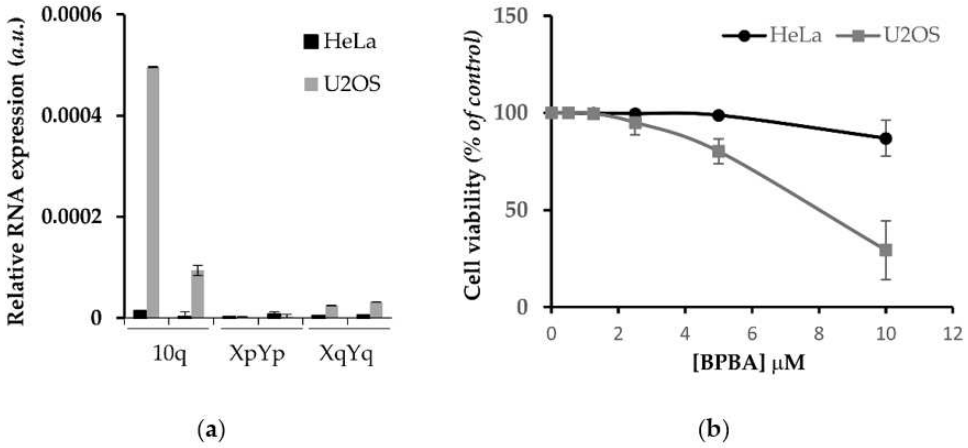


Figure 7. Differential effect of **BPBA** on viability of high vs. low TERRA expressing cells. (**a**) RT qPCR showing relative expression of specific TERRA RNA at different subtelomeric loci (10q, XpYp, XqYq) of HeLa, and U2OS cells; (**b**) HeLa and U2OS exposed to indicated doses of **BPBA** and, 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.

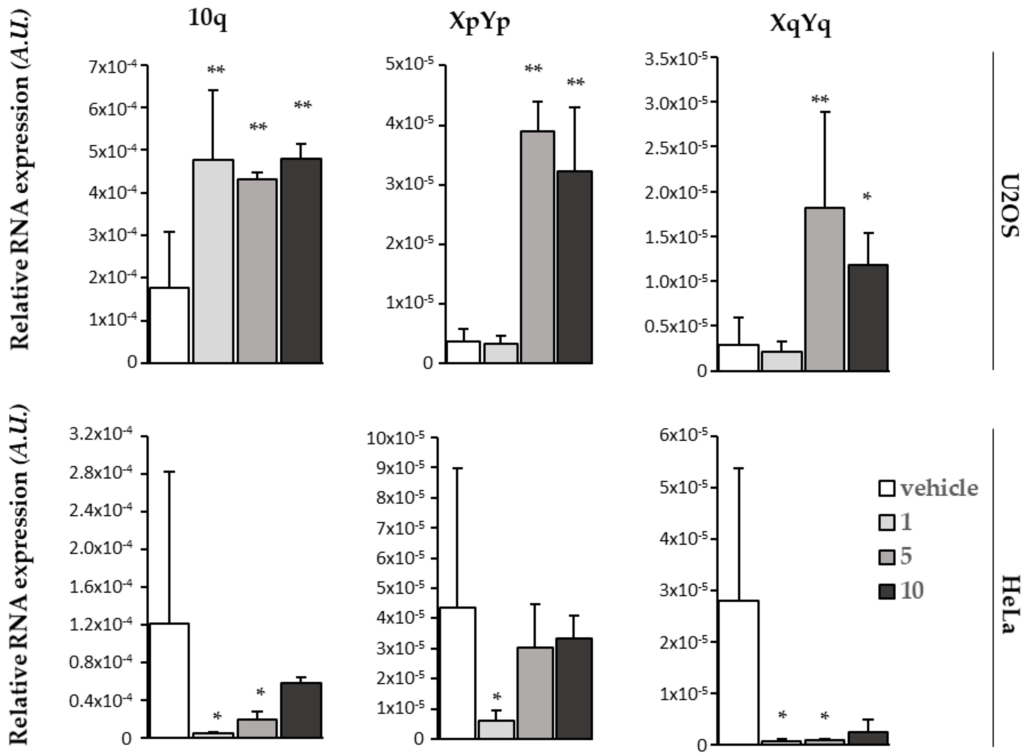


Figure 8. Effect of **BPBA** on TERRA expression. RT qPCR showing relative expression of specific TERRA RNA at different subtelomeric loci of HeLa and U2OS cells treated with the vehicle or indicated doses (1, 5, or 10 μM) of **BPBA** for 72 h. Mean of three independent experiments is shown. Error bars 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).

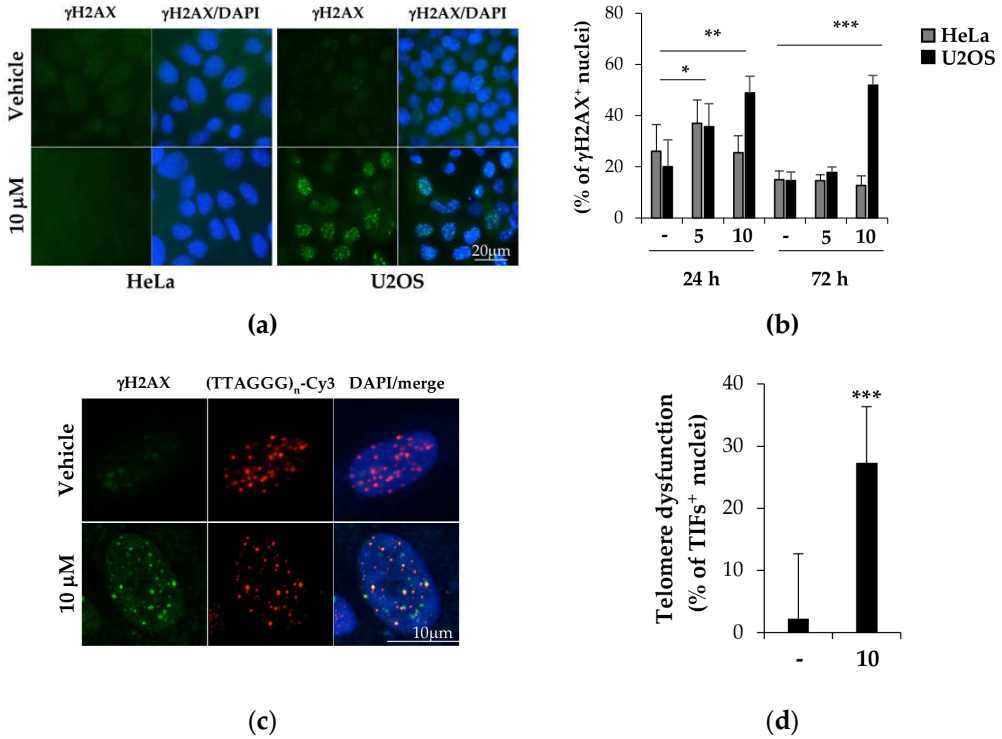


Figure 9. TERRA binding by **BPBA** triggers persistent DDR and telomere dysfunction in U2OS cells. HeLa and U2OS cells treated with indicated doses of **BPBA**. At each endpoint, cells were fixed and processed for immunofluorescence with the antiphosphorylated γH2AX mAb primary antibody, followed by 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; $0 \leq$ logP ≤ 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; logP ≥ 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 V

A short peptide that preferentially binds c-MYC G4 DNA

Minard A., Morgan D., Raguseo F., Di Porzio A., Liano D., Jamieson A., Di Antonio M.

Chemical Communications, 2020, Vol. 56 (63)

ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2020, **56**, 8940

Received 23rd April 2020,
Accepted 1st July 2020

DOI: 10.1039/d0cc02954h

rsc.li/chemcomm

A short peptide that preferentially binds c-MYC G-quadruplex DNA†

Aisling Minard, ‡[ab] Danielle Morgan,‡[c] Federica Raguseo,[a] Anna Di Porzio, [ad] Denise Liano, [ab] Andrew G. Jamieson *[c] and Marco Di Antonio *[ab]

G-quadruplexes (G4s) are non-canonical DNA secondary structures. The identification of selective tools to probe individual G4s over the ~700 000 found in the human genome is key to unravel the biological significance of specific G4s. We took inspiration from a crystal structure of the bovine DHX36 helicase bound to the G4 formed in the promoter region of the oncogene c-MYC to identify a short peptide that preferentially binds MYC G4 with nM affinity over a small panel of parallel and non-parallel G4s tested.

Besides the well-known double-helical structure, DNA can adopt several non-canonical structural arrangements under physiological conditions.[1] G-quadruplexes (G4s) have recently emerged as an interesting alternative DNA structure with respect to their potential for biological regulation. Sequences that form G4s are highly abundant, with more than 700 000 G4-structures experimentally detected across the human genome by sequencing experiments.[2] Because of their high genomic prevalence, DNA G4s have been speculated to play a role in several biological processes, including transcriptional regulation,[3] telomeric maintenance,[4] genomic instability,[5] cancer progression,[3] accelerating ageing[6] and neurodegenerative diseases.[7] To this end, several tools to map, visualise and stabilise G4s in biologically relevant contexts have been developed, in an attempt to unravel the specific biological roles played by G4-formation in cells.[8] Although genome-wide mapping experiments by means of Chromatin Immuno-Precipitation (ChIP) methods have facilitated the detection of G4s that are actually formed in cells,[9] the investigation of biological processes regulated by G4-stabilisation mainly relies on the use of small-molecule ligands that can be applied in living cells and monitored in real-time. To date, several ligands with high affinity and specificity for G4s over canonical double-stranded DNA have been reported.[10] Small-molecule targeting approaches mainly rely on π–π stacking interactions of the molecular probes with the top (or bottom) end of the G4-structure. Whilst this approach is highly effective to achieve G4 *vs.* duplex selectivity, it is less applicable to achieve selectivity towards a small subset of G4s over the ~700 000 available in the human genome. This is most likely due to the promiscuous presence of accessible π-stacking surfaces across different G4-structures, which hampers selective recognition of an individual G4 by means of π–π stacking interactions alone.

Here, we have taken inspiration from a recently reported crystal structure of the bovine DHX36 helicase bound to the G4-structure formed in the promoter region of the oncogene c-MYC[11] to identify a short-peptide sequence (**DM039**) that binds with nM affinity the MYC G4. Interestingly, **DM039** showed negligible binding to other G4-forming sequences of parallel, antiparallel and mixed type topologies, suggesting potential for selective targeting of MYC. Furthermore, we found that when constrained into a helical conformation by hydrocarbon stapling, the peptide displayed lower binding affinity to MYC and loss of selectivity towards double- and single-stranded DNA. This data suggested that the peptide requires a good degree of structural flexibility for high affinity and selective G4-binding, which is not intuitive based on the helical conformation adopted by the peptide in the crystal structure.[11]

We started our investigation analysing the crystal structure of the bovine DHX36 bound to MYC G4[11] (PDB ID: 5VHE) using PyMOL[12] (Fig. 1A) and identified the 22-amino acid sequence reported in Fig. 1B as the minimal MYC-binding domain of the protein in this particular context. This 22-amino acid sequence (Fig. 1B) is in agreement with that previously reported by NMR studies on the human version of the DHX36 helicase, which identified similar sequences (18, 23, 29 and 53 amino acids) as G4-binding motifs.[13] We then synthesised the 22-amino acids

[a] *Imperial College London, Chemistry Department, Molecular Science Research Hub, 80 Wood Lane, W12 0BZ, London, UK. E-mail: m.di-antonio@imperial.ac.uk*
[b] *Institute of Chemical Biology, Molecular Science Research Hub, 80 Wood Lane, W12 0BZ, London, UK*
[c] *School of Chemistry, University of Glasgow, G12 8QQ, Glasgow, UK. E-mail: andrew.jamieson.2@glasgow.ac.uk*
[d] *Department of Pharmacy, University of Naples Federico II, Via D. Montesano 49, 80131 Naples, Italy*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cc02954h

‡ These authors contributed equally to this work.

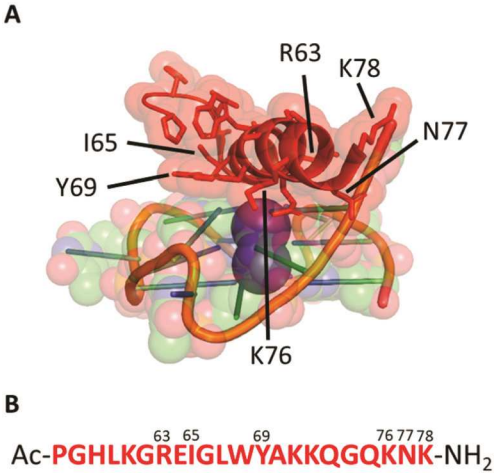


Fig. 1 (A) Minimal peptide binding domain of DHX36 bound to the MYC DNA G-quadruplex. (B) The 22-amino acid sequence (**DM039**) used in this study.

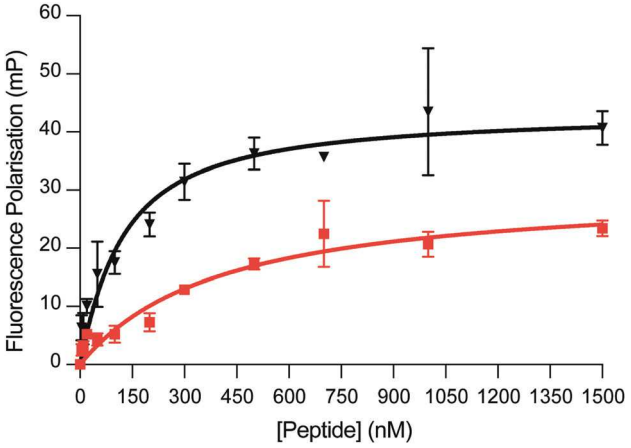


Fig. 2 FP binding traces obtained with DM039 in the presence of c-MYC (black) and BCL-2 (red). The FP experiments were performed in Inner cell buffer (see ESI†), at room temperature, by using a fixed concentration of oligomer (20 nM) and increasing concentrations of peptide (0–5000 nM. Here, only the range 0–1500 nM is shown) (see ESI†). Each measurement was repeated three times.

sequence (**DM039**) to investigate whether this particular peptide motif displayed an increased selectivity towards MYC G4 over other G4-structures. **DM039** was synthesised using previously established Fmoc/^{t}Bu solid-phase peptide synthesis (SPPS) protocols, employing an automated SPPS synthesiser, and acetylated at the N-terminal prior to HPLC purification and MS characterisation (see ESI†).

We next sought to evaluate the binding affinity of **DM039** towards a small panel of G4-forming sequences as well as a single-stranded and a double-stranded control sequence. Typically, small-molecule ligands that bind G4s are tested by melting experiments (FRET-melting) for their ability to stabilise G4-structures.[14] Even though stabilisation properties are correlated, on some level, to the binding affinity of the ligands, they cannot be considered a direct affinity measurement and are difficult to apply to peptides that are temperature sensitive. To overcome this limitation, we decided to exploit the higher molecular weight of the 22-amino acid peptide, compared to small-molecules, to develop a Fluorescence-Polarisation (FP) assay to reliably measure G4-binding. The FP measurements rely on the use of a fluorescein (FAM)-labelled oligonucleotide and the formation of a big complex (peptide–DNA) upon binding. The peptide–DNA complex formation causes a slower tumbling of the FAM-labelled oligonucleotide, which in turn generates a fluorescence polarised signal that can be detected as a function of the fraction of the oligonucleotide bound to the peptide.[15] We initially applied our FP-assay to evaluate the binding affinity of **DM039** to MYC G4. As displayed in Fig. 2, **DM039** binds with high affinity to MYC G4 with an observed K_d of 112 nM [59–196 nM – 95% CI], confirming the suitability of our FP-assay to measure the binding affinity of short peptides to G4s. We next tested whether the high binding affinity of **DM039** to MYC was specific to G4-folded oligonucleotides, testing its binding by FP to a FAM-labelled single-stranded mutant of MYC, no longer able to fold into a G4, and to a FAM-labelled double-stranded DNA sequence (see ESI†). **DM039** displayed negligible binding to both the MYC-mutant and double-stranded DNA controls, suggesting that G4-specific interactions are responsible for the observed binding to the MYC G4. To further test that the binding to MYC was dependent on the G4-folding status, we measured reduced binding affinity of **DM039** to the MYC G4 sequence when annealed in Li^+ buffer, which prevents stabilisation of the formed G4. The resulting K_d was 1226 nM [491–3469 nM – 95% CI], proving a ~10-fold reduction in binding affinity with respect to the binding measured in K^+ buffer (see ESI†).

Since previous reports on similar peptide sequences suggested selectivity for parallel G4s over antiparallel ones,[13] we decided to test the binding affinity of **DM039** to other parallel G4s. Specifically, we have tested FP-binding of **DM039** to FAM-labelled c-KIT1 and c-KIT2 G4s, which are known to fold into a parallel topology.[16] Surprisingly, **DM039** showed negligible binding to both c-KIT1 and c-KIT2 G4s (see ESI†), suggesting that this particular peptide sequence is not selective to all parallel conformations but can strongly discriminate against G4s of the same topology.

We further investigated whether **DM039** was unable to bind non-parallel G4s, as reported for similar peptide sequences extracted from human DHX36.[13] To this end, we performed the FP-assay against FAM-labelled HRAS G4 (antiparallel),[17] hTelo[18] and BCL-2[19] G4s (mixed type). **DM039** displayed negligible binding to either HRAS or hTelo G4s, confirming the inability of this peptide to bind some non-parallel G4-structures that were previously reported (see ESI†).[13] On the other hand, the DNA G4 found in the promoter region of the BCL-2 oncogene, which also folds into a mixed-type conformation,[19] was bound by **DM039** with an observed K_d of 579 nM [447–749 nM – 95% CI]. Although the binding affinity is lower to what observed for MYC G4 (~5 fold), this evidence suggested that **DM039** could interact with some non-parallel G4s, possibly due to an unusually high accessibility of the G-tetrads in this mixed-type structure.

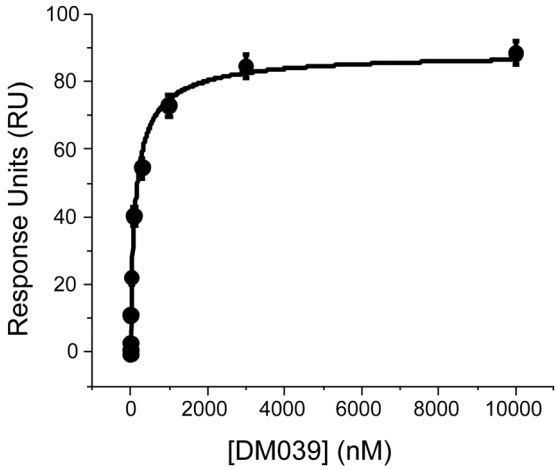


Fig. 3 SPR experiments to study the interaction of **DM039** and MYC G4 where the ligand concentration was varied from 0.1 nM to 10 μM.

We next sought to confirm the high binding affinity displayed by **DM039** to MYC G4 using Surface Plasmon Resonance (SPR) to measure the dissociation constant of **DM039** bound to MYC G4. Strikingly, SPR data confirmed high binding affinity of **DM039** to MYC G4, providing a measured K_d of 123 ± 23 nM (Fig. 3), which is in excellent agreement with the value obtained by FP.

Altogether, our biophysical study revealed that the **DM039** peptide preferentially binds MYC G4 with high selectivity over single-stranded and double-stranded DNA. Furthermore, **DM039** displayed a significant degree of intra-G4 selectivity that, in contrast with previous reports, is not simply dictated by G4-topology but rather dependent on specific G4-structural features. This suggests that peptide-based binding of G4s might be the key to achieve individual G4-targeting by means of molecular probes.

The 22-amino acid sequence that constitutes **DM039** is folded into a α-helix in the crystal structure reported of bovine DHX36 bound to MYC G4.[11] To test whether **DM039** also adopts a helical conformation when bound to MYC G4, we have performed CD analysis of **DM039** in solution and in the presence of MYC G4 (see ESI†). Whilst **DM039** appear to be unstructured in solution, an α-helical CD signal becomes detectable upon binding of the peptide to MYC G4, displaying an increase in the calculated helical fraction (fH) from 0.073 measured with **DM039** in solution to 0.155 observed upon MYC G4 binding (see ESI†). Therefore, we wanted to investigate whether conformationally constraining the wild type peptide using a hydrocarbon staple could pay the entropic penalty of folding and improve the binding affinity of **DM039** to MYC G4.

To test this hypothesis, we designed two stapled peptide analogues of the wild type peptide, **DM083** (*i*, *i* + 4 type staple) and **DM102** (*i*, *i* + 7 type staple) (Fig. 4A and B). On inspection of the helical interaction motif of bovine DHX36 bound to MYC G4 in the crystal structure and using the helix wheel, there are two ideal positions on the back face of the helix with which to incorporate a conformational constraint. The *i*, *i* + 4 stapled peptide **DM083** was designed to incorporate the hydrocarbon bridge across one loop of the helix between residues K71 and

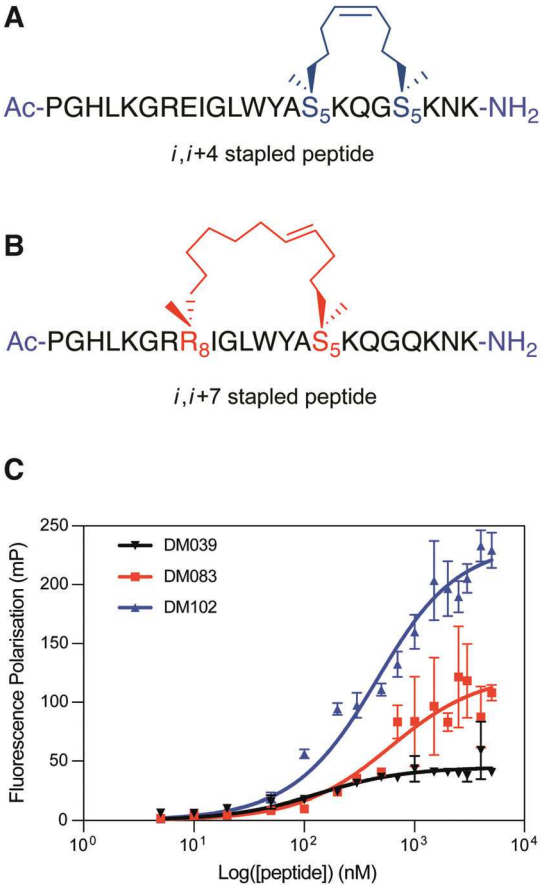


Fig. 4 (A) *i*, *i* + 4 hydrocarbon stapled peptide **DM083**, (B) *i*, *i* + 7 hydrocarbon stapled peptide **DM102**. (C) Differential binding behaviour against MYC G4 observed for **DM039** and its two stapled versions (**DM083** and **DM102**) by means of FP measurements. The FP experiments were performed in Inner cell buffer (see ESI†), at room temperature, by using a fixed concentration of oligomer (20 nM) and increasing concentrations of peptide (0–5000 nM) (see ESI†). Each measurement was repeated three times.

Q75 (Fig. 4A). A longer *i*, *i* + 7 hydrocarbon constrain was also designed that constrains across two loops of the helix, between residues E64 and K71 (Fig. 4B). Importantly, the residues mutated do not appear to take part in the binding event with MYC G4 or intramolecular stabilising interactions. The linear peptides were synthesised using standard Fmoc/^{t}Bu SPPS conditions using a microwave assisted peptide synthesiser. Commercially available Fmoc-S_5-OH and Fmoc-R_8-OH were used to incorporate the nonproteinogenic amino acids with the appropriate alkene functionality required for ring-closing metathesis, which was achieved using Grubbs 1st Generation Catalyst. The resulting cyclised peptides were cleaved from solid support and globally deprotected under standard acidic conditions (TFA/TIS/H_2O). The crude peptides were then purified by RP-HPLC and characterised using analytical RP-HPLC and mass spectrometry (ESI) (see ESI†).

We first validated that stapling increased the helicity of the peptides by CD, observing a fH of 0.18 and 0.22 for **DM083** and **DM102** respectively (see ESI†). Crucially, the fH values of both **DM083** and **DM102** did not increase in the presence of MYC G4

(see ESI†). The two stapled peptides were then assessed for binding to MYC G4 by means of our FP-assay. Both **DM083** and **DM102** displayed a significant reduction of binding affinity to MYC G4, with measured K_d of 548 nM [317–934 nM – 95% CI] and 332 nM [249–435 nM – 95% CI] respectively (Fig. 4C). This suggested that stabilising the helical peptide conformation reduces the affinity for the targeted G4-structure (see ESI†). Whilst *i*, *i* + 4 stapling in **DM083** reduces binding affinity to MYC G4 of ~5 fold, *i*, *i* + 7 stapled peptide **DM102** still retains substantial binding affinity to MYC G4 with a reduction of ~3 fold compared to **DM039**. This is in agreement with a recent report that revealed how lactam stapling of a similar peptide sequence overall retained G4-binding ability.[20] Nevertheless, we wanted to evaluate whether chemically locking of the peptide into a helical conformation, observed in the DHX36 crystal structure, resulted in retention of preferential selectivity for MYC G4, since the DHX36 helicase, unlike **DM039**, does not display intrinsic selectivity for a specific G4-structure or topology. We therefore tested the *i*, *i* + 7 stapled peptide **DM102** for binding to parallel and non-parallel DNA G4s as well as to single- and double-stranded DNA controls. Surprisingly, we observed that **DM102** binds with similar binding affinities all the oligonucleotides tested, yielding observed K_d values of 574 nM [381–855 nM – 95% CI] for c-KIT1, 1081 nM [847–1383 nM – 95% CI] for hTelo, 831 nM [622–1112 nM – 95% CI] for BCL-2, 1238 nM [827–1878 nM – 95% CI] for HRAS, 2628 nM [1648–4497 nM – 95% CI] for MYC-mutant and 2377 nM [1440–4222 nM – 95% CI] for double-stranded DNA. This observation suggested that **DM039** binds to MYC G4 through an induced fit mechanism and a good degree of flexibility is required of the extended peptide backbone for both high affinity interaction and selectivity, as chemical stapling of this sequence in **DM102** rendered the peptide a non-specific DNA binder.

In conclusion, we have developed a FP-assay to measure binding affinity of a short peptide (22 amino acids) extracted from a crystal structure of the bovine DHX36 helicase bound to MYC G4. Although selectivity towards parallel G4s was previously suggested for similar peptide sequences, we report that the particular peptide used in this work displays negligible binding to other parallel G4-structures (c-KIT1 and c-KIT2), binding to a mixed type G4 (BCL-2) and a preferential binding to MYC G4. Furthermore, we observed that chemical stapling of this peptide might preserve MYC binding but abrogates intra-G4 selectivity as well as selectivity over single and double-stranded DNA, suggesting that a rigid, extended peptide α-helical conformation is detrimental to the observed MYC selectivity. We anticipate that further investigation on **DM039** binding modes and non-canonical chemical stapling will be key to develop a MYC-selective probe and disentangle the biological role of this particular G4 over the ~700 000 present in the human genome. Our findings might pave the way towards the rational design of peptide-based molecular probes for selective targeting of individual G4s.

Conflicts of interest

There are no conflicts to declare.

References

1 R. Hänsel-Hertsch, M. Di Antonio and S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.*, 2017, **18**, 279–284.
2 V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith and S. Balasubramanian, *Nat. Biotechnol.*, 2015, **33**, 877–883.
3 S. Balasubramanian, L. H. Hurley and S. Neidle, *Nat. Rev. Drug Discovery*, 2011, **10**, 261–275.
4 D. Rhodes and H. J. Lipps, *Nucleic Acids Res.*, 2015, **43**, 8627–8637.
5 K. Paeschke, M. L. Bochman, P. D. Garcia, P. Cejka, K. L. Friedman, S. C. Kowalczykowski and V. A. Zakian, *Nature*, 2013, **497**, 458–462.
6 L. Froetscher, K. Marosi, D. M. Wilson, D. M. Eckley, S. Ghosh, M. P. Mattson, I. G. Goldberg, H. Kassahun, R. W. Maul, M. Gorospe, E. F. Fang, A. Tseng, S. K. Bharti, P. Bastian, M. Scheibye-Knudsen, V. A. Bohr, T. Iyama, S. De, M. Borch Jensen, R. M. Brosh, H. Nilsen and K. Scheibye-Alsing, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 12502–12507.
7 A. R. Haeusler, C. J. Donnelly, G. Periz, E. A. J. Simko, P. G. Shaw, M. S. Kim, N. J. Maragakis, J. C. Troncoso, A. Pandey, R. Sattler, J. D. Rothstein and J. Wang, *Nature*, 2014, **507**, 195–200.
8 F. Raguseo, S. Chowdhury, A. Minard and M. Di Antonio, *Chem. Commun.*, 2020, **56**, 1317–1324.
9 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 and S. Balasubramanian, *Nat. Genet.*, 2016, **48**, 1267–1272.
10 D. Monchaud and M. P. Teulade-Fichou, *Org. Biomol. Chem.*, 2008, **6**, 627–636.
11 M. C. Chen, R. Tippana, N. A. Demeshkina, P. Murat, S. Balasubramanian, S. Myong and A. R. Ferré-D'amaré, *Nature*, 2018, **558**, 465–469.
12 The PyMOL Molecular Graphics System, Version 2.1.1.
13 B. Heddi, V. V. Cheong, H. Martadinata and A. T. Phan, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 9609–9613.
14 A. De Cian, L. Guittat, M. Kaiser, B. Saccà, S. Amrane, A. Bourdoncle, P. Alberti, M. P. Teulade-Fichou, L. Lacroix and J. L. Mergny, *Methods*, 2007, **42**, 183–195.
15 N. J. Moerke, *Curr. Protoc. Chem. Biol.*, 2009, **1**, 1–15.
16 S. Rankin, A. P. Reszka, J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian and S. Neidle, *J. Am. Chem. Soc.*, 2005, **127**, 10584–10589.
17 A. Membrino, S. Cogoi, E. B. Pedersen and L. E. Xodo, *PLoS One*, 2011, **6**, e24421.
18 A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2006, **34**, 2723–2735.
19 J. Dai, D. Chen, R. A. Jones, L. H. Hurley and D. Yang, *Nucleic Acids Res.*, 2006, **34**, 5133–5144.
20 M. Y. Yanevaa, V. V. Cheong, J. K. Cheng, K. W. Lim and A. T. Phan, *Biochem. Biophys. Res. Commun.*, 2020, DOI: 10.1016/j.bbrc.2020.02.144.

Paper VI

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, Vol. 22 (21)

International Journal of
Molecular Sciences

MDPI

Article

Synthesis and Characterization of Bis-Triazolyl-Pyridine Derivatives as Noncanonical DNA-Interacting Compounds

Anna Di Porzio [1,†], Ubaldina Galli [2,†], Jussara Amato [1], Pasquale Zizza [3], Sara Iachettini [3], Nunzia Iaccarino [1], Simona Marzano [1], Federica Santoro [1], Diego Brancaccio [1], Alfonso Carotenuto [1], Stefano De Tito [4,5], Annamaria Biroccio [3], Bruno Pagano [1], Gian Cesare Tron [2,*] and Antonio Randazzo [1,*]

[1] Department of Pharmacy, University of Naples Federico II, Via D. Montesano 49, 80131 Naples, Italy; anna.diporzio@unina.it (A.D.P.); jussara.amato@unina.it (J.A.); nunzia.iaccarino@unina.it (N.I.); simona.marzano@unina.it (S.M.); federica.santoro@unina.it (F.S.); diego.brancaccio@unina.it (D.B.); alfonso.carotenuto@unina.it (A.C.); bruno.pagano@unina.it (B.P.)
[2] Department of Pharmaceutical Sciences, University of Piemonte Orientale, Largo Donegani 2/3, 28100 Novara, Italy; ubaldina.galli@uniupo.it
[3] Oncogenomic and Epigenetic Unit, IRCCS-Regina Elena National Cancer Institute, 00144 Rome, Italy; pasquale.zizza@ifo.gov.it (P.Z.); sara.iachettini@ifo.gov.it (S.I.); annamaria.biroccio@ifo.gov.it (A.B.)
[4] Molecular Cell Biology of Autophagy, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK; stefano.de-tito@crick.ac.uk
[5] Institute of Experimental Endocrinology and Oncology, National Research Council, 80131 Naples, Italy
* Correspondence: giancesare.tron@uniupo.it (G.C.T.); antonio.randazzo@unina.it (A.R.); Tel.: +39-0321375857 (G.C.T.); +39-081678514 (A.R.)
† These authors contributed equally to the work.

Abstract: Besides the well-known double-helical conformation, DNA is capable of folding into various noncanonical arrangements, such as G-quadruplexes (G4s) and i-motifs (iMs), whose occurrence in gene promoters, replication origins, and telomeres highlights the breadth of biological processes that they might regulate. Particularly, previous studies have reported that G4 and iM structures may play different roles in controlling gene transcription. Anyway, molecular tools able to simultaneously stabilize/destabilize those structures are still needed to shed light on what happens at the biological level. Herein, a multicomponent reaction and a click chemistry functionalization were combined to generate a set of 31 bis-triazolyl-pyridine derivatives which were initially screened by circular dichroism for their ability to interact with different G4 and/or iM DNAs and to affect the thermal stability of these structures. All the compounds were then clustered through multivariate data analysis, based on such capability. The most promising compounds were subjected to a further biophysical and biological characterization, leading to the identification of two molecules simultaneously able to stabilize G4s and destabilize iMs, both in vitro and in living cells.

Keywords: G-quadruplex; i-motif; bis-triazolyl-pyridine derivatives; click chemistry; biophysical study; immunofluorescence microscopy

check for updates

Citation: Di Porzio, A.; Galli, U.; Amato, J.; Zizza, P.; Iachettini, S.; Iaccarino, N.; Marzano, S.; Santoro, F.; Brancaccio, D.; Carotenuto, A.; et al. Synthesis and Characterization of Bis-Triazolyl-Pyridine Derivatives as Noncanonical DNA-Interacting Compounds. *Int. J. Mol. Sci.* **2021**, *22*, 11959. https://doi.org/10.3390/ijms222111959

Academic Editors: Manlio Palumbo and Claudia Sissi

Received: 28 September 2021
Accepted: 1 November 2021
Published: 4 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The canonical double helix is the most widely recognized genomic DNA structure. However, DNA is structurally dynamic and able to adopt a number of alternative secondary structures, such as cruciforms, triplexes, G-quadruplexes, and i-motifs [1–3]. G-quadruplex (G4) structures, four-stranded helical complexes that can arise from guanine-rich sequences, are among the most extensively studied noncanonical DNA secondary structures [4]. The core scaffold of a G4 is the G-tetrad (Figure 1), a cyclic array of four guanines connected by eight Hoogsteen hydrogen bonds. Two or more G-tetrads can self-associate into vertical stacks giving rise to G4s, which are extremely stable structures, although the stability of a G4 structure depends on many factors, including the length and sequence composition of the G-rich motif, the nature of the binding cations, and the folding topology [5–8]. Indeed,

G4s can adopt a wide variety of topologies, depending on the combinations of strand orientation and length and composition of the loops (the intervening sequences between the G-rich repeats) [9,10]. DNA sequences that can fold into G4 structures are broadly distributed across the human genome and mainly located at telomeres, gene promoters, and introns [11,12]. Due to such a high genomic prevalence, G4s have been speculated to play a role in several biological processes. Indeed, they may offer telomere protection and block replication fork progression [4], while in gene promoters, they may represent a physical obstacle in the way of the RNA polymerase machinery, thus affecting the expression of the downstream gene [13].

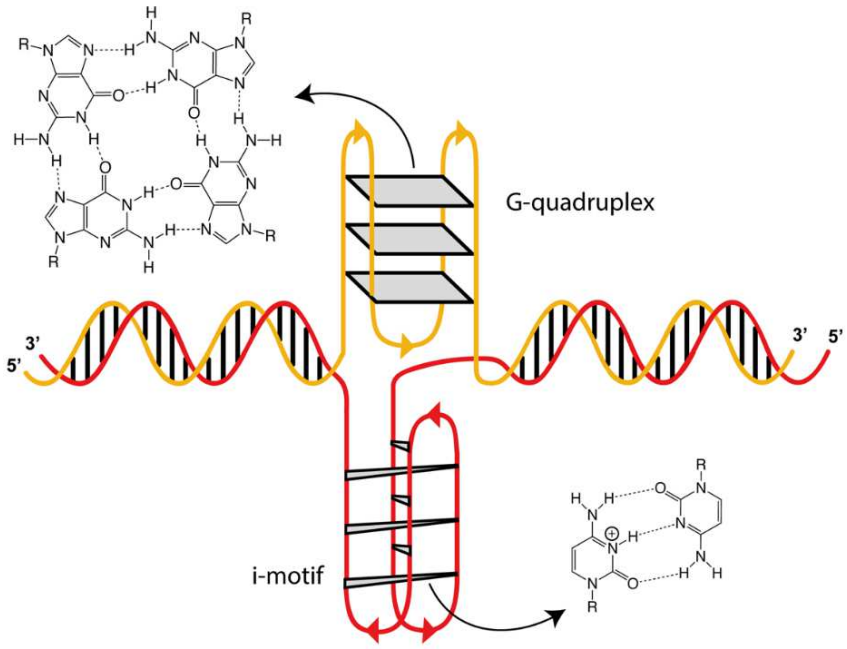


Figure 1. Noncanonical DNA secondary structures: G-quadruplex (G4) and i-motif (iM). **Top left**: a G-tetrad; **bottom right**: a hemi-protonated cytosine-cytosine (C-C$^+$) base pair.

The strand complementary to a G4-forming DNA motif is a C-rich sequence that can fold in another four-stranded structure called the i-motif (iM or i-DNA) [14]. This is composed of two head-to-tail, intercalated, parallel-stranded duplexes held together by hemi-protonated cytosine-cytosine (C-C$^+$) base pairs (Figure 1). Very recently, evidence for iM formation in cells has been provided by NMR experiments [15] and by the discovery of an antibody that specifically binds iM structures in the nuclei of human cells [16]. Analogously to G4s, the presence of iM structures may affect gene transcription and telomere biology [2]. Interestingly, the formation of iM structures is cell-cycle dependent, peaking at the late G1 phase [16], whereas G4 formation is maximal during the S phase [17]. This suggests that iMs and G4s might play different roles in regulating gene expression and transcription.

Many studies, focused on disentangling the biological roles of G4 and iM structures, demonstrate that such structures can be considered as molecular targets in cancer therapies [14,18,19], leading to significant therapeutic advantages in the treatment of cancer. For example, direct evidence of the role of DNA G4s in gene regulation has been provided, showing that oncogene promoter activity is generally repressed using G4s stabilizers, while destabilization of G4s generally leads to increased gene transcription [20]. On the other hand, constant efforts are still underway to fully decipher the specific biological functions of iM structures. In this contest, the scientific community has also made big efforts to find small molecules able to recognize G4 and iM structures. More than a thousand compounds that bind and stabilize G4 structures have been found so far. Many of them display considerable selectivity for G4s over single-stranded and double-stranded DNA, and some

of them are also able to discriminate among distinct G4 folding topologies. Compared to the well-documented examples of G4 ligands, the discovery of specific iM-targeting compounds is much more limited [21,22].

The design of small molecules able to interact with G4s is usually based on the following requisites: (i) the presence of a (hetero)-aromatic system that gives π-stacking interactions with planar G-tetrads and flanking bases, (ii) a V-shaped form to maximize the interaction with the G4, (iii) two or three cationic side chains able to participate in ionic interactions with the phosphate backbone of loops and grooves of G4 structures. On the other hand, loop binding and recognition of hemi-protonated C-C^+ base pairing seem to be the most favored binding modes of iM ligands [23,24].

To increase the number of putative noncanonical DNA-targeting compounds, medicinal chemists are in continuous search for the use of synthetic strategies able to give access to as many compounds as possible in a short period of time. In this context, over the years, the so-called click chemistry reaction, namely the Cu(I)-catalyzed Huisgen 1,3-dipolar cycloaddition between azides and alkynes (CuAAC) [25], has gained considerable interest in medicinal chemistry in general, and in the design of G4 binding ligands in particular [26]. The CuAAC exclusively generates a 1,4-disubstituted 1,2,3-triazole. This reaction is operationally simple, regioselective, modular, and bio-orthogonal, providing high yields of the products. Moreover, it can be performed in aqueous media under mild conditions and the azide and alkyne building blocks are either commercially available or easily synthesizable. In addition, the resulting 1,2,3-triazole system is stable under physiological conditions as it is resistant to hydrolysis and oxidation-reduction [27,28].

In 2006, Neidle's group described, for the first time, the use of click chemistry for the synthesis of G4-stabilizing ligands [29]. Afterward, this reaction has been employed to functionalize different aromatic/heteroaromatic moieties with appropriate side chain appendages to form novel molecular scaffolds that displayed efficient binding properties to G4s of different topologies [30–32].

Here, inspired by pyridostatin, a potent G4-binder based on an *N*,*N*′-bis(quinolinyl) pyridine-2,6-dicarboxamide scaffold [33–35] which has also been shown to interact with iM DNA (although without significant effects on the thermal stability of such structure) [36], a library of bis-triazolyl-pyridine derivatives has been generated. Then, the ability of such compounds of interacting with G4 and iM structures has been investigated both from a biophysical and biological point of view. In particular, we exploited the Hantzsch multicomponent reaction [37,38] to generate a central pyridine scaffold with two ester groups, which can act as potential arms ready to be functionalized. The resulting heterocyclic ring can work as a planar aromatic platform able to produce π-stacking interactions with the nucleobases, whereas the ester groups can be converted into azides. Then, a click chemistry reaction between azides and terminal alkynes containing a tertiary amine was carried out to install two cationic side chains able to interact with the negatively charged phosphate groups to stabilize the interaction with the nucleic acid.

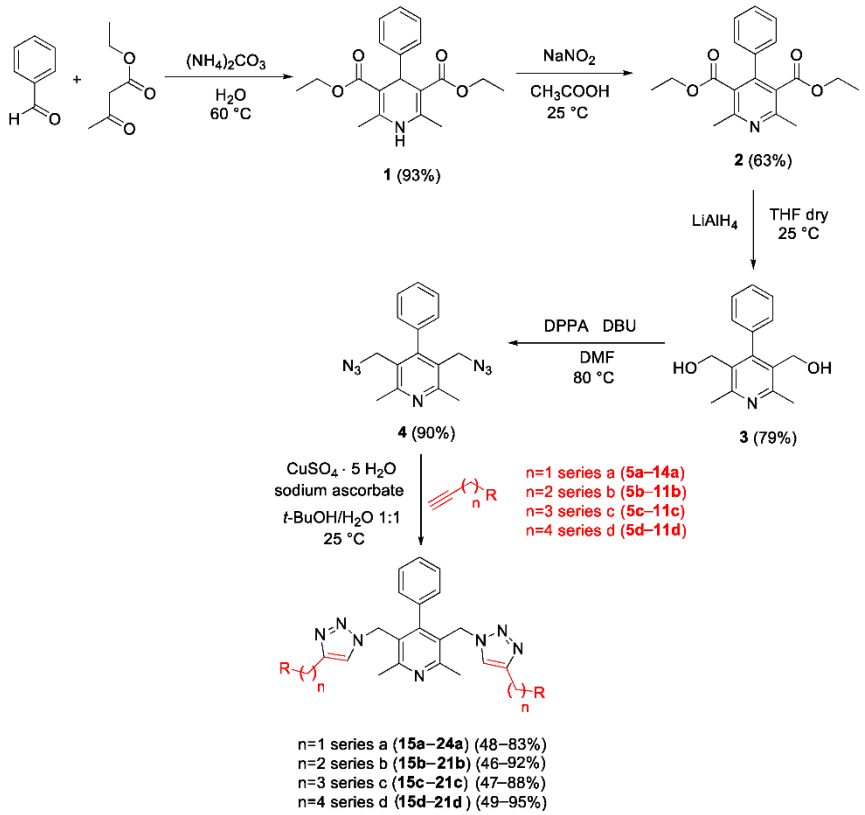
The synergistic effect of a multicomponent reaction and a click chemistry functionalization allowed for the generation of a library of 31 potential G4 and/or iM ligands thanks to the extensive modulation of the basic side chains, which is well-known to be crucial to impart the required binding selectivity.

The synthesized bis-triazolyl-pyridines were preliminarily screened for their ability to interact with G4 and/or iM DNAs through circular dichroism (CD) spectroscopy and clustered based on their capability to affect the thermal stability of the investigated DNA sequences by employing the Principal Component Analysis (PCA) technique. The ability of the best-selected compounds to interact with noncanonical DNAs was further investigated by using NMR and fluorescence spectroscopies. Finally, the proficiency of such compounds to modulate G4 and iM structures formation in living cells was assessed.

2. Results and Discussion

2.1. Chemistry

The synthetic route followed for the preparation of the putative noncanonical DNA-interacting compounds is depicted in Scheme 1. In brief, the pyridine scaffold was obtained reacting benzaldehyde, ethyl acetoacetate, and ammonium carbonate in water at 60 °C [39]. The dihydropyridine (**1**) obtained was then aromatized to give **2** via oxidation with sodium nitrite in glacial acetic acid at room temperature (RT). The two ester groups were then reduced to alcohol using $LiAlH_4$. The diol **3** was transformed into the diazide **4** using the DPPA, DBU protocol. The diazide **4** was reacted with twenty-seven terminal alkynes (see Supplementary Materials, General Procedures for the Synthesis of Alkynes and Schemes S1–S4), which contain at least one tertiary nitrogen atom, and with four commercial alkynes containing hydroxyl groups, by using the CuAAC protocol [25]. In particular, click reactions were carried out with four series of alkynes of different lengths: series a (**5a–14a**), series b (**5b–11b**), series c (**5c–11c**), series d (**5d–11d**) containing one, two, three or four methylene groups, respectively. A library of 31 bis-triazolyl-pyridines (series a **15a–24a**, series b **15b–21b**, series c **15c–21c**, series d **15d–21d**) were thus obtained and the compounds were tested for their ability to interact with G4 and iM DNA structures. The identity of the new compounds was established by their FT-IR, ^{1}H-NMR, ^{13}C-NMR, MS, and elemental analyses.



Scheme 1. Synthesis of bis-triazolyl-pyridines **15a–24a**, **15b–21b**, **15c–21c**, **15d–21d**.

2.2. Circular Dichroism Screening and Selection of the Most Promising Compounds

The 31 synthesized bis-triazolyl-pyridines were preliminarily screened for their ability to interact with G4 and iM DNAs through circular dichroism (CD) spectroscopy [40]. Various G-rich sequences that can form parallel, antiparallel, and hybrid G4 structures and their iM-forming C-rich counterparts were used in these experiments. In particular, we employed the G-/C-rich motifs from the promoter regions of BCL-2 (*Bcl-2 G4* and *Bcl-2 iM*) [41,42] and c-MYC (*c-Myc G4* and *c-Myc iM*) oncogenes [43,44], as well as the G-rich 23-mer (Tel_{23} *G4)* and the C-rich 22-mer (*hTeloC iM*) truncations of the human telomeric DNA sequence [45–47]. Moreover, a 27-mer hairpin-duplex-forming sequence (*Hairpin*) was also used to evaluate the selectivity of the compounds for noncanonical DNAs over duplex DNA. DNA samples were prepared in the appropriate buffers (see Materials and Methods section) at pH 7.0 for the G-rich and hairpin-duplex-forming sequences, and at pH 5.0 for the C-rich sequences since the formation of iM structures requires cytosine protonation to form the C-C^+ base pair.

The overall topology adopted by each investigated DNA sequence was first verified by CD measurements (Figures S1–S4). The *c-Myc G4*-forming sequence adopted a parallel G4 conformation, showing the characteristic positive band at around 260 nm and negative band at ~240 nm in its CD spectrum. The presence of an additional band at ~290 nm in the CD spectrum of *Bcl-2 G4* indicates the formation of a mixed parallel/antiparallel conformation. On the other hand, the Tel_{23} *G4* sequence showed a CD spectrum having two positive bands at 290 and 270 nm and a weak negative band at 240 nm, consistent with the formation of a hybrid [3+1] G4 folding topology. As for *Bcl-2 iM*, *c-Myc iM*, and *hTeloC iM*, all of them exhibited a CD spectrum with a positive band at 288 nm and a negative one at around 260 nm, which are characteristic values of iM formation [48], while the *Hairpin* showed a positive band at around 280 nm and a negative one at ~250 nm in the CD spectrum, confirming duplex formation (Figure S4).

Additional CD spectra were acquired to explore the potential of the 31 bis-triazolyl-pyridines to alter the native folding topology of the investigated DNAs. Upon addition of an excess of compounds (10 molar equiv relative to the DNA), no relevant variations of DNA chiroptical signals were observed, suggesting overall preservation of the DNA structure in the presence of the compounds (Figures S1–S4). Then, their DNA stabilizing/destabilizing properties were evaluated by CD-melting experiments measuring the compound-induced change in the apparent melting temperature ($\Delta T_{1/2}$) either of G4, iM, or duplex structures (Table 1 and Figures S5–S8).

Considering a large number of $\Delta T_{1/2}$ values to be compared (31 compounds tested on 7 DNA sequences give a total of 217 values of $\Delta T_{1/2}$), a multivariate data analysis tool, such as Principal Component Analysis (PCA) [49], was employed to explore the data set. In this way, similarities and differences in the behavior of the investigated compounds could be detected and correlated to their structural features. PCA is an unsupervised method that allows the reduction in the dimensionality of a data set, providing a visual representation of the major variance in the data. In particular, the original variables (5 DNA sequences) were transformed into 2 new uncorrelated variables termed Principal Component 1 (PC1), which explains 67.76% of the total variance, and Principal Component 2 (PC2) explaining 25.87% of the variance. PC1 and PC2 are visualized both in a scores plot (Figure 2a,c,d), where the compounds characterized by a similar behavior are clustered together and colored according to their structural features and in a loadings plot (Figure 2b) that indicates the DNA sequences responsible for the distribution of the compounds observed in the scores plot. Therefore, by employing this technique, we could cluster the 31 compounds based on their capability to affect the thermal stability of the investigated DNA sequences.

Table 1. Compound-induced thermal stabilization or destabilization of the investigated oligonucleotides, measured by circular dichroism melting experiments. $\Delta T_{1/2}$ values are the differences in the apparent melting temperatures of DNA in the presence and absence of compounds.

Compound	$\Delta T_{1/2}$ (°C) [1]						
	Hairpin	*Bcl-2 G4*	*c-Myc G4*	*Tel$_{23}$ G4*	*Bcl-2 iM*	*c-Myc iM*	*hTeloC iM*
15a	0	5.0	7.0	1.1	−7.0	−18.4	−12
16a	−0.2	4.5	5.0	1.5	−11.0	−27.5	−25.0
17a	−1.0	4.0	5.0	0	−10.5	−27.0	−24.5
18a	0.2	4.5	7.5	3.5	−13.5	−24.5	−21.0
19a	1.7	0.5	2.0	1.0	−4.0	−5.5	−4.0
20a	3.9	0.5	1.0	0	−8.0	−10.4	−18.0
21a	2.9	0	0	0	0	−0.8	2.5
22a	−0.2	2.7	4.5	0.5	−11.0	−19.0	−25.5
23a	2.9	0	0.5	0.5	−6.0	−17.0	−12.5
24a	−0.7	1.5	2.0	−0.5	−2.0	−2.8	1.5
15b	−0.1	5.1	7.5	1.6	−6.0	−20	−26.5
16b	−0.1	4.5	8.0	1.6	−13.0	−19.4	−24.5
17b	−0.1	4.0	5.5	1.0	−10	−17.4	−23.5
18b	−0.1	5.5	8.5	1.5	−6.0	−11.4	−15.5
19b	0	3.0	2.0	1.2	−5.0	−4.0	−7.0
20b	2.2	5.9	4.0	0.1	−8.0	−6.0	−9.0
21b	1.9	1.0	1.0	2.0	0	−0.4	2.0
15c	−0.1	6.1	10.6	0	−11.0	−24	−29
16c	−0.1	5.6	9.5	0.5	−10.0	−17.4	−24.1
17c	−0.1	5.6	6.5	2.0	−12.0	−21.1	−27.4
18c	1.1	7.6	10.5	1.6	−10.5	−20.0	−25.5
19c	−1.1	5.1	4.2	−1.0	−3.0	−3.0	−10.0
20c	3.0	7.1	6.8	0	−9.0	−2.9	−9.0
21c	1.9	0	0	2.0	0	−0.9	1.5
15d	0	8.8	9.8	−1.0	−5.0	−4.9	−7.0
16d	−0.3	8.0	8.1	2.0	−5.0	−4.9	−7.0
17d	1.0	9.0	6.9	2.0	−8.0	−7.0	−8.9
18d	0	10.1	9.8	−1.0	−8.0	−8.0	−9.0
19d	0.9	4.2	2.8	−2.0	−5.0	−5.0	−7.0
20d	2.9	8.0	8.2	0	−8.0	−5.0	−9.0
21d	1.9	2.0	0.8	−0.8	−1.0	0	−1.0

[1] The error on $\Delta T_{1/2}$ values is ±1.0 °C. All experiments were performed in duplicate, and the values reported are the average of two measurements.

Hairpin and *Tel$_{23}$ G4* were excluded from the analysis since their thermal stability was not significantly affected by any of the tested compounds ($|\Delta T_{1/2}| < 4$ °C). Therefore, the PCA was performed including two G4-forming sequences (*Bcl-2 G4* and *c-Myc G4*) and three iM-forming sequences (*hTeloC iM*, *c-Myc iM*, and *Bcl-2 iM*) for a total of 5 DNA sequences. Moreover, we decided to take into account the $\Delta T_{1/2}$ absolute values considering that most of the compounds destabilized the iM structures (negative $\Delta T_{1/2}$) while stabilizing the G4 structures. Figure 2 shows the scores and loadings plots generated by the PCA model. The fact that all the DNA sequences are clustered together on the very right side of the loadings plot (Figure 2b) suggests that the compounds that lie on the right side of the scores plot (Figure 2a) are characterized by higher $|\Delta T_{1/2}|$ values compared to the ones situated on the left side. Interestingly, by coloring each compound according to the pK_a value corresponding to the conjugate acid of the strongest basic group in the molecule (see Table S1 and Figure S9 for details), we could observe the presence of a trend along PC1, which goes from the left (lower pK_a values) to the right (higher pK_a values). This means that the compounds that least affect the thermal stability of the investigated DNA structures are the ones that are less prone to carry a positive charge (under the experimental conditions), which, in turn, is fundamental for the interaction with the negatively charged DNA backbone [50,51]. Then, in order to understand the common

structural features shared by the compounds able to interact with the DNA structures, the samples were colored according to the chemical features of the side chain (R). Thus, they were divided into different classes where R (i) contains only nitrogen; (ii) contains only oxygen; (iii) contains both nitrogen and oxygen with the latter being closed in a cycle; and (iv) contains both nitrogen and oxygen with the latter being present as a free hydroxyl group. Figure 2c shows the same scores plot reported in Figure 2a but colored according to these new classes and it is evident that both the lack of nitrogen and the presence of oxygen reduce the capability of the compounds to stabilize/destabilize the DNA structures. Indeed, the compounds that most affect the stability of the DNA structures are those containing only nitrogen atoms in the R chain.

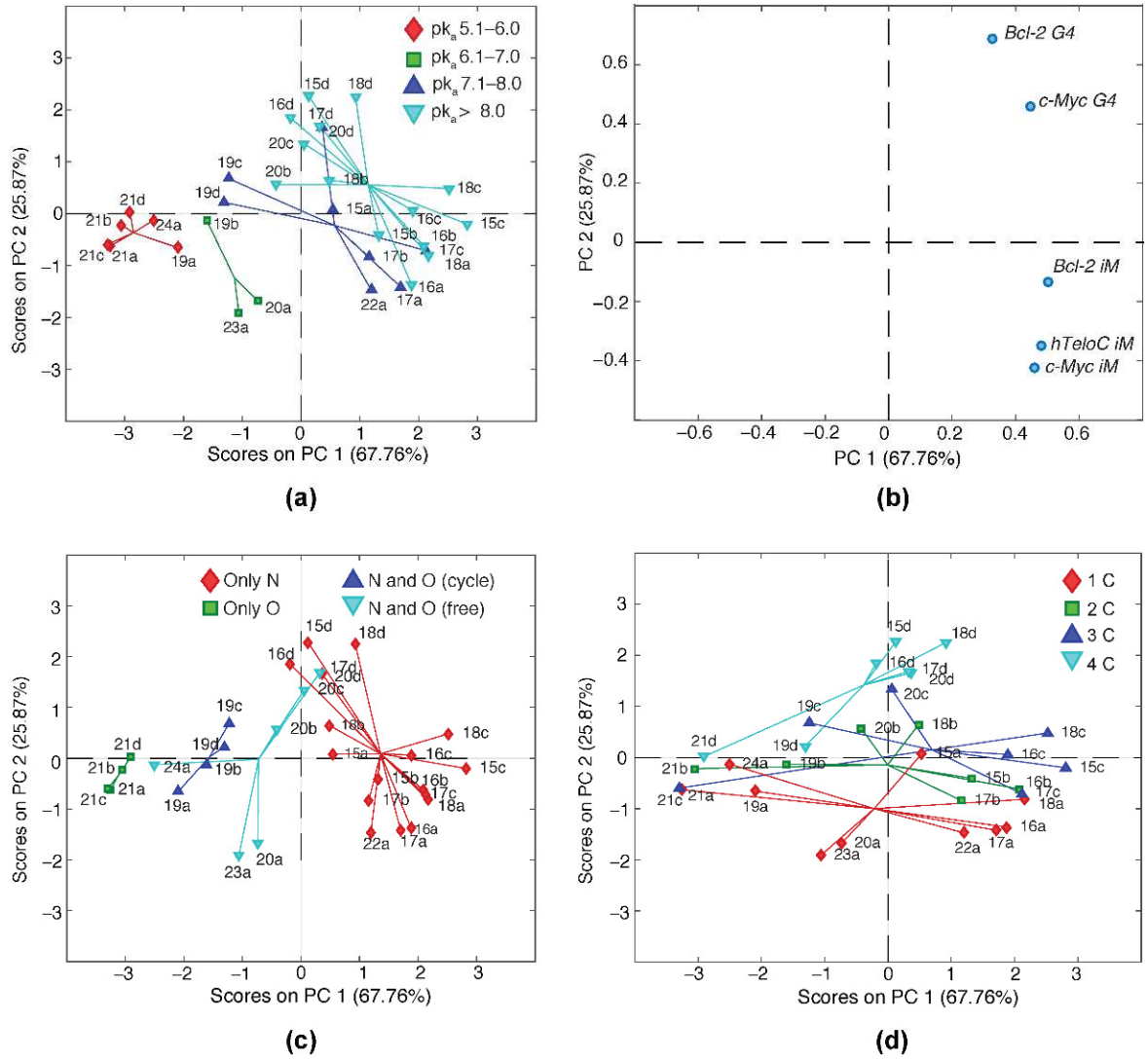


Figure 2. Output of the PCA model performed on the $\Delta T_{1/2}$ values dataset of the 31 compounds tested on *Bcl-2 G4*, *c-Myc G4*, *Bcl-2 iM*, *c-Myc iM*, and *hTeloC iM* DNAs. PC1/PC2 scores plot colored according to the (**a**) pK_a value corresponding to the conjugated acid of the strongest basic group in each molecule (**c**) presence/absence of both nitrogen and oxygen atoms in the side chains, and (**d**) length of the R side chain (from 1 up to 4 carbons); (**b**) PC1/PC2 loadings plot.

On the other hand, PC2 (28.87% of the total variance) shows the distribution of the compounds according to their effect on G4 rather than on the iM structures. In particular, **16a**, **17a**, **20a**, **22a**, and **23a**, located in the lower part of the scores plot where the iM-forming sequences are placed (in the corresponding area of the loadings plot), affect more significantly the thermal stability of the iMs rather than G4 structures. On the

contrary, compounds **15d**, **16d**, **17d**, **18d**, **20d,** and **20c**, located in the upper part of the plot, where the G4-forming sequences are placed (in the corresponding area of the loadings plot), preferentially affect the thermal stability of the G4 over the iM structures. In order to understand which structural feature could lead to such a selective effect on a DNA structure over another, we decided to label the compounds in the scores plot according to the length of the R side chain, which goes from 1 carbon (series a) up to 4 carbons (series d) (Figure 2d). Interestingly, we observed that the longer is the carbon chain, the higher is the ability to stabilize the G4 structures. The effect of the side chain length is observable only for the compounds that do affect the thermal stability of the investigated structures (right part of the plot) while it is not evident for the rest of the molecules. This confirms once again that the presence of positive charges is fundamental for the interaction with the DNA, while the length of the side chain could modulate compound selectivity. Interestingly, there are two compounds (**15c** and **18c**) that are located at the very right side of the PC1 axis, and thus characterized by high $\Delta T_{1/2}$ values, while being in the middle region of the PC2 axis, meaning that they are able to affect the thermal stability of both iM and G4 structures. In particular, they can stabilize the G4 structures while destabilizing the iM ones. This is a very interesting aspect from a biological point of view since previous studies have reported that G4 and iM structures may exert opposite roles in controlling gene transcription: while G4 ligand-induced stabilization generally blocks gene expression, stabilization of iMs seems to show transcription activating capabilities [36,52].

2.3. Nuclear Magnetic Resonance (NMR) Studies

To get information about the binding mode of **15c**, **18c**, **20a,** and **23a** to *c-Myc G4* and *hTeloC iM*, 1D ^{1}H-NMR experiments were performed. We focused on these compounds since able to affect the thermal stability of both iM and G4 structures (**15c** and **18c**) or exclusively of the iMs (**20a** and **23a**). In addition, **21d** was also employed to have a comparison with a compound unable to affect the stability of any of the investigated DNAs. Particularly, wanting to perform a comparative study on the binding mode and affinity of the most promising compounds towards the investigated G4/iMs and needing to use acidic conditions to guarantee iM folding, we performed NMR and fluorescence titration experiments under the same experimental conditions (10 mM potassium phosphate buffer, pH 5.0, for the studied G4 and 10 mM sodium phosphate buffer, pH 5.0, for the iMs).

The spectral regions of the imino and aromatic protons of *c-Myc G4* and *hTeloC iM* in the absence and presence of compounds (2 molar equiv) are shown in Figures 3 and 4, respectively. According to the literature, the *c-Myc G4* sequence forms, under the experimental conditions used, a single G4 conformation characterized by 12 well-resolved imino proton peaks, corresponding to the 12 guanines involved in the three G-tetrad planes [43]. On the other hand, *hTeloC iM* folds in an iM structure characterized by 3 well-resolved imino proton peaks that correspond to the 6 intercalated C-C^{+} pairs [47].

No relevant shift of both the imino and aromatic signals of *c-Myc G4* and *hTeloC iM* was observed upon the addition of **21d** (Figures 3 and 4, and Tables S2 and S3). Conversely, **15c**, **18c**, **20a,** and **23a** mainly affected the imino protons of *c-Myc G4*, as well as the aromatic protons of both *c-Myc G4* and *hTeloC iM* (Figures 3 and 4, and Tables S2 and S3, respectively). As for *c-Myc G4*, not negligible chemical shift changes were observed for the imino (H1) protons of G16 belonging to the 5′ G-tetrad (G7−G11−G16−G20), as well as for those of guanines involved in the 3′ G-tetrad (G9−G13−G18−G22). At the same time, the most perturbed aromatic protons turned out to be those of A15, located in the double nucleotide loop (T14A15) connecting G13 and G16, and of A25 which stacks over the 3′ G-tetrad, specifically, with residues G9 and G22. Overall, these results suggest a common end-stacking binding mode for **15c**, **18c**, **20a**, and **23a** to the 3′ G-tetrad of *c-Myc G4*, which is also combined with loop binding.

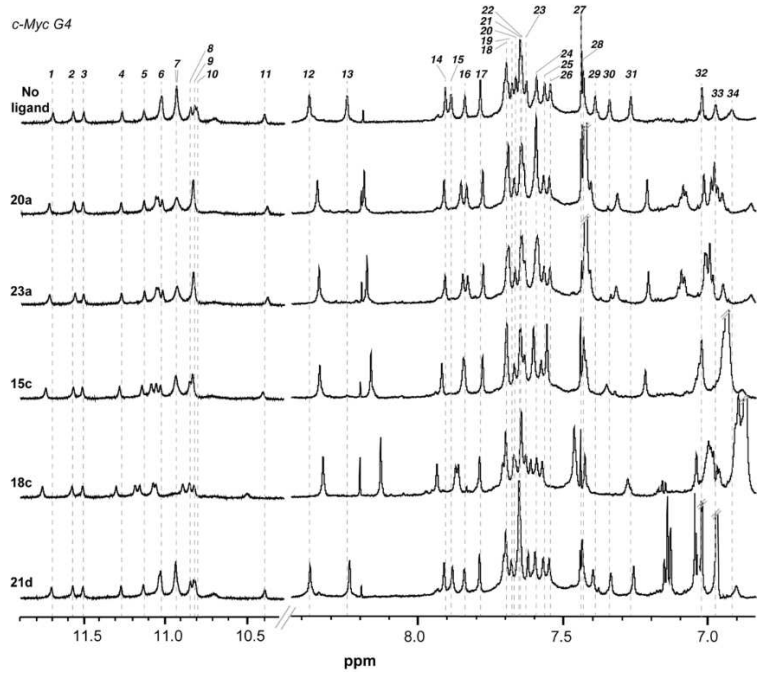


Figure 3. ^{1}H-NMR spectra of *c-Myc G4* in the absence and presence of compounds **20a**, **23a**, **15c**, **18c,** and **21d**. See Table S2 for signal assignment. The NMR sample was prepared at 100–150 μM final oligonucleotide concentration in 10 mM potassium phosphate buffer (pH 5.0).

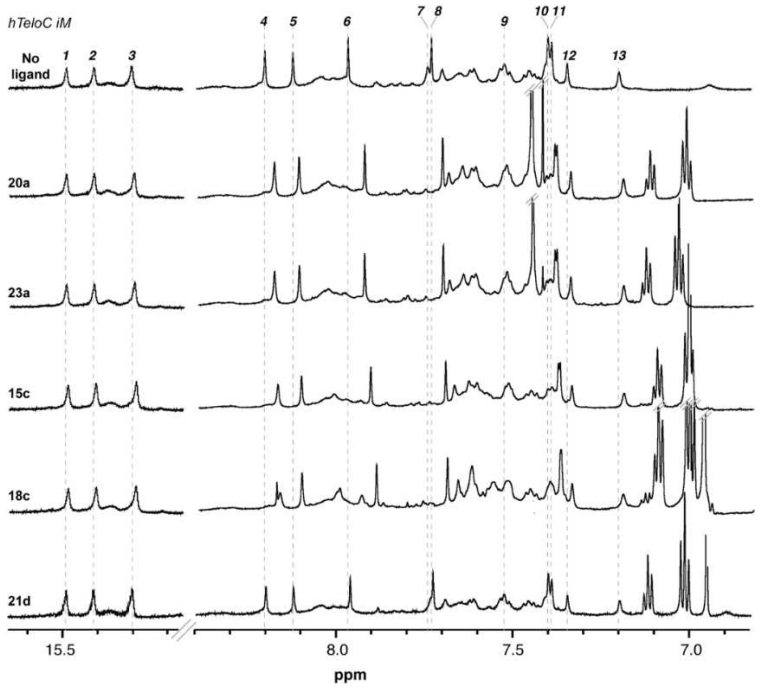


Figure 4. ^{1}H-NMR spectra of *hTeloC iM* in the absence and presence of compounds **20a**, **23a**, **15c**, **18c**, and **21d**. See Table S3 for signal assignment. The NMR sample was prepared at 100–150 μM final oligonucleotide concentration in 10 mM sodium phosphate buffer (pH 5.0).

Conversely, the imino proton signals of *hTeloC iM* were not affected by compounds **15c**, **18c**, **20a**, and **23a**. However, significant chemical shift changes of aromatic protons were observed especially for those residues located in the loops of *hTeloC iM*, and particularly for A5, A6, A11, A12, A17, and A18. These observations suggest interaction with iM loops,

which is one of the most common mechanisms of interaction proposed for iM-targeting compounds [24].

2.4. Saturation-Transfer Difference (STD) NMR Analysis

In order to confirm the binding of compounds **15c**, **18c**, **20a**, and **23a** to *c-Myc G4* and *hTeloC iM*, a ligand-based NMR approach was used. Particularly, Saturation Transfer Difference NMR (STD NMR) experiments were performed in buffer solutions containing a large excess of compounds (250 μM) with respect to the DNA (20 μM) [53]. STD NMR is based on the magnetization transfer by the Nuclear Overhauser Effect (NOE) from the macromolecule to the ligand. STD NMR spectra are the result of the subtraction of a 1D ^{1}H spectrum where the macromolecule protons are selectively saturated (on-resonance spectrum) from a reference spectrum in which the macromolecule is not saturated (off-resonance spectrum). The resulting difference STD spectrum will only show the ligand signals interacting with the macromolecule. The proton resonances for **15c**, **18c**, **20a**, and **23a** were observed in the STD spectra (Figures S10–S13) in presence of both *c-Myc G4* and *hTeloC iM*, demonstrating once again an interaction between the investigated compounds and the DNAs. In the STD spectra, the presence of both aromatic and aliphatic signals for **15c**, **18c**, **20a**, and **23a** suggests that the aromatic moieties of these compounds could interact with the G4 and iM through π-π stacking with the guanines of G-tetrads and the adenines of loops, respectively, while the side chains could interact with the backbone of the DNA molecules through electrostatic interactions. The very low signal intensity of the protons of **21d** in the STD spectra (Figure S14) shows that this compound does not interact with the DNA sequences under investigation.

2.5. Evaluation of Ligand Affinity by Fluorescence Experiments

To obtain quantitative data regarding the affinity of **15c**, **18c**, **20a**, and **23a** for *c-Myc G4*, *c-Myc iM*, and *hTeloC iM*, fluorescence titration experiments were performed. Fluorescence emission spectra of the compounds in the absence and presence of increasing amounts of oligonucleotides were recorded. The fluorescence intensity of the compounds decreased gradually with the addition of DNA until it reached saturation (Figure S15). The binding curves were obtained by plotting the fraction of bound compound (α), calculated following fluorescence changes at the emission maximum, as a function of DNA concentration (Figure 5). Binding constants (K_b) were obtained from curve fitting using an independent and equivalent binding sites model and reported in Table 2. The results of the interpolation analysis indicated that in those cases the data could not be fitted with a single binding site, but they were well fitted with two binding sites per DNA molecule. Compound **18c**, which displayed the strongest affinity for the G4 structure of *c-Myc*, also exhibited the least affinity for the iM structure of both *c-Myc* and *hTeloC*. On the other hand, compounds **15c**, **20a** and **23a** showed higher affinity values for *c-Myc iM* and *hTeloC iM* rather than to *c-Myc G4*.

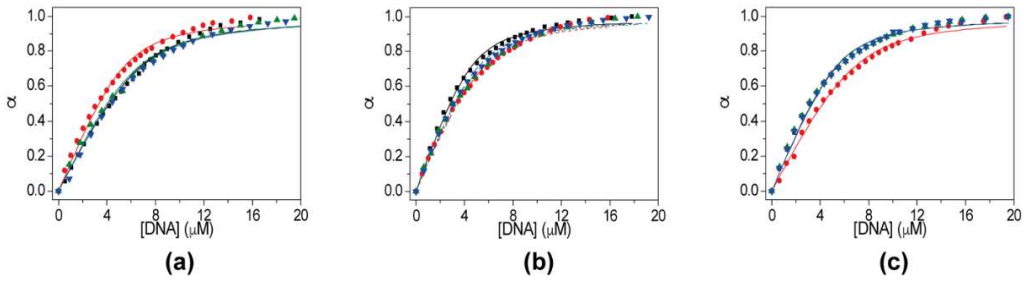


Figure 5. Fluorescence titration curves were obtained by plotting the fraction of bound compound (**20a** in green, **23a** in blue, **15c** in black, and **18c** in red) as a function of (**a**) *c-Myc G4*, (**b**) *c-Myc iM*, and (**c**) *hTeloC iM* concentration. Titrations were carried out in 10 mM potassium phosphate buffer (pH 5.0) for *c-Myc G4*, and 10 mM sodium phosphate buffer (pH 5.0) for *c-Myc iM* and *hTeloC iM*.

Table 2. Binding constant (K_b) values were obtained by fitting the fluorescence titration curves to an independent and equivalent binding sites model [54].

Compound	K_b ($\times 10^6$ M^{-1}) [1]		
	c-Myc G4	***c-Myc iM***	***hTeloC iM***
20a	2.3	3.1	3.1
23a	2.2	3.2	3.1
15c	2.5	3.4	3.2
18c	3.1	2.8	2.6

[1] The error on the K_b values is ± 0.2.

Furthermore, fluorescence titration experiments were carried out to investigate the interaction of **21d** with a G4 (*c-Myc G4*) and an iM (*c-Myc iM*) structure (Figure S16a,b). The obtained curves were then fitted giving K_b values of 1.1 (± 0.2) $\times$ 10^6 and 1.6 (± 0.2) $\times$ 10^6 M^{-1} for *c-Myc G4* and *c-Myc iM*, respectively, and stoichiometry of 2:1 in both cases. These results confirm the lower (although not null) affinity of this compound for these targets. Finally, the same experiment was performed for **18c** with a C-rich single-stranded oligonucleotide [$d(CT)_{15}$] (Figure S16c), giving a 1:1 stoichiometry and a K_b of 4.6 (± 0.4) $\times$ 10^6 M^{-1}, and thus suggesting that the decrease in melting temperature of the iMs could also be due to preferred binding to the unfolded DNA, which would shift the folding equilibrium.

2.6. Effect of the Compounds on G4/iM Structures in Living Cells

Starting from the results of the biophysical assays, the selected molecules (**15c**, **18c**, **20a**, **23a**, and **21d** as a control) were tested for their capability to destabilize the iM and/or stabilize the G4 structures also in the cells. To this aim, U2OS human osteosarcoma cells, one of the models already used by Zeraati and colleagues to validate their anti-iM antibody (iMab) [16], were treated for 24 h with 2 μM of each compound and the iM structures were evaluated by immunofluorescence (IF) microscopy. As a positive control, cells were maintained in 8% of CO_2 for 2.5 h, a condition that is known to favor the formation of intracellular iM structures [16]. In accordance with the previously reported results, all the investigated molecules, except for **21d**, succeeded in destabilizing the iMs as demonstrated by the significant reduction in the iM structures following the treatments (Figures 6 and S17a). IF experiments highlighted that the compounds differ from each other in terms of potency in destabilizing the iMs, with **15c** and **18c** that—inducing a reduction of about 70% and 60% in the iMs fluorescence signal, respectively—turned out to be the most effective compounds. Notably, the major efficacy of **15c** and **18c** in destabilizing iMs well correlates with their capability to stabilize G4 structures (Figures 7 and S17b). Indeed, the dual activity of **15c** and **18c**, already observed at the biophysical level, is in line with the literature data showing that, under certain conditions, the formation of G4s and iM structures can be reciprocally influenced [55]. Altogether, our data confirmed the ability of the selected compounds to affect G4/iM structures also in the complexity of a cellular model, demonstrating the biological translatability of the biophysical results.

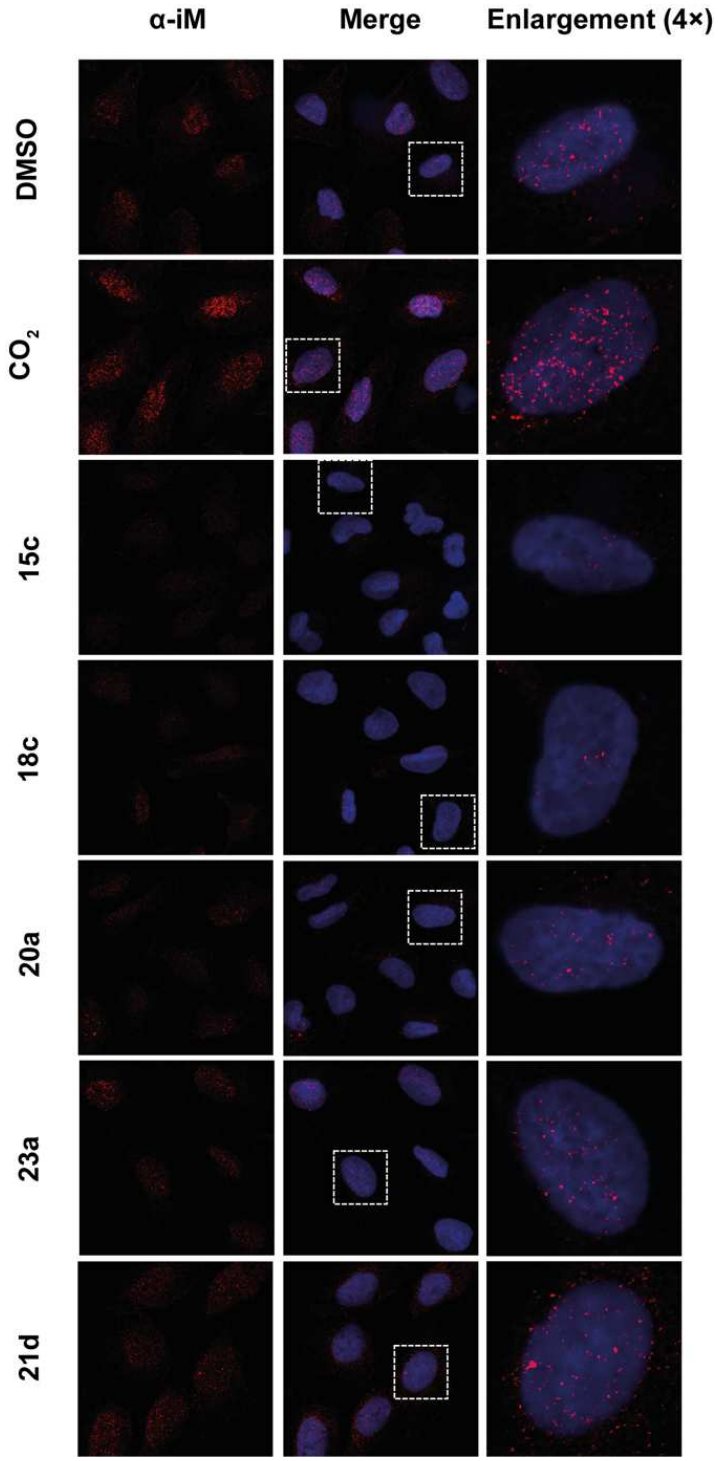


Figure 6. Biological evaluation of the iM-destabilizing activity of the selected compounds. Immunofluorescence analysis of iM structures in U2OS cells treated for 24 h with 2 μM of the selected compounds or an equivalent amount of DMSO (negative control). As a positive control, cells were maintained for 2.5 h in an atmosphere with 8% of CO_2. Representative images of confocal sections (63×) used to detect iM structures are shown. **Left panels**: iM structures (red) detected by anti-iM antibody (α-iM). **Middle panels**: merged images showing iM structures (red) and DAPI counterstained nuclei (blue). **Right panels**: 4× enlargements from the pictures in the middle panels. The experiment was performed in triplicate and at least 9 fields/experiments were evaluated for each condition. Scale bars are reported in the figures.

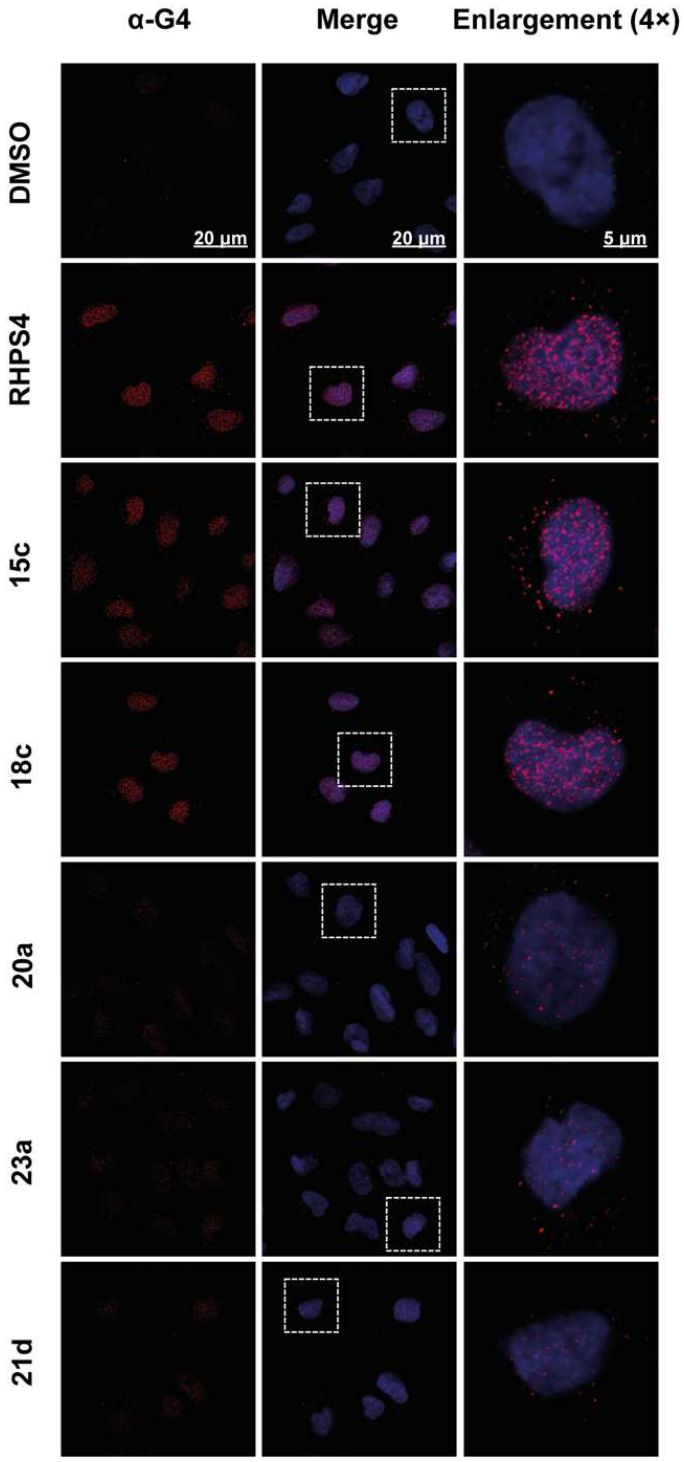


Figure 7. Biological evaluation of G4-stabilizing activity of the selected compounds. Immunofluorescence analysis of G4 structures in U2OS cells treated for 24 h with 2 µM of the selected compounds or an equivalent amount of DMSO (negative control). As a positive control, cells were treated for 24 h with 1 µM RHPS4. Representative images of confocal sections (63×) used for the detection of G4 structures are shown. **Left panels**: G4 structures (red) detected by anti-G4 antibody (α-G4). **Middle panels**: merged images showing G4 structures (red) and DAPI counterstained nuclei (blue). **Right panels**: 4× enlargements from the pictures in the middle panels. The experiment was performed in triplicate and at least 9 fields/experiments were evaluated for each condition. Scale bars are reported in the figures.

3. Materials and Methods

3.1. Materials

Commercially available reagents and solvents were purchased from Merck (Darmstadt, Germany) or Alfa Aesar (Karlsruhe, Germany) and used without any further purification. Immediately before using it, tetrahydrofuran (THF) was distilled from Na/benzophenone, under a slight positive atmosphere of dry nitrogen. Distillation from P_2O_5 was used to dry dichloromethane, which was then stored on activated molecular sieves (4 Å). When needed, the reactions were carried out in oven-dried glassware, under a positive pressure of dry nitrogen. The determination of the melting points (which are uncorrected) was done in open glass capillaries, with a Stuart scientific SMP3. All the compounds were characterized by employing FT-IR Nicolet Avatar (Thermo Fisher Scientific, Waltham, MA, USA), ^{1}H-NMR, ^{13}C-NMR (APT) (JEOL ECP 300 MHz, Tokyo, Japan), and mass spectrometry (Thermo Finnigan LCQ-deca XP-plus, Waltham, MA, USA), equipped with an ESI source and an ion trap detector. Chemical shifts (δ) are reported in parts per million (ppm). Flash column chromatography was performed on Kieselgel 60, 230–400 mesh ASTM silica gel (Merck, Kenilworth, NJ, USA). Thin-layer chromatography was carried out on 5 × 10 cm plates with a layer thickness of 0.25 mm silica gel 60 F_{254} (Merck, Kenilworth, NJ, USA). When necessary, they were developed with KMnO4 reagent. Elemental analysis (C, H, N) of the tested compounds was used to establish their purity which turned out to be ≥95%, being within ±0.4% of the calculated values. Alkynes **11a**, **11b**, **11c**, **11d**, **12a** were commercially available.

3.2. Chemistry

3.2.1. Synthesis of Diethyl 2,6-Dimethyl-4-Phenyl-1,4-Dihydropyridine-3,5-Dicarboxylate (**1**)

To a stirred solution of benzaldehyde (5.0 g, 47.17 mmol, 1 equiv) in water (50 mL), were added ethyl acetoacetate (13.21 mL, 103.77 mmol, 2.2 equiv) and ammonium carbonate (6.79 g, 70.7 mmol, 1.5 equiv). The mixture was heated at 60 °C for 4 h. After the reaction completion, cold water was added, and the product was extracted with dichloromethane (×2). The combined organic layers were washed with brine (×1), dried over anhydrous sodium sulfate, and concentrated *in vacuo*, to give 14.44 g of product as a pale-yellow solid, which was used in the next step without further purification. Yield 93%; mp 160–161 °C (lett. 156–157 °C) [56]; IR (KBr) 3341, 2981, 2934, 2902, 1687, 1650, 1488, 1372, 1211, 1091, 703 cm^{-1}; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 7.28–7.10 (m, 5-H), 5.87 (br s, 1-H), 4.98 (s, 1-H), 4.08–4.05 (m, 4-H), 2.30 (s, 6-H), 1.22–1.18 (m, 6-H) ppm; ^{13}C-NMR (75 MHz, $CDCl_3$) δ 167.8, 147.9, 144.1, 128.1, 127.9, 126.1, 104.1, 59.8, 39.7, 19.5, 14.3 ppm.

3.2.2. Synthesis of Diethyl 2,6-Dimethyl-4-Phenylpyridine-3,5-Dicarboxylate (**2**)

To a stirred and cooled (0 °C) solution of 1,4-dihydropyridine **1** (8.07 g, 24.6 mmol, 1 equiv) in glacial acetic acid (16.14 mL, 295.2 mmol, 12 equiv), $NaNO_2$ (8.48 g, 123 mmol, 5 equiv) was added in small portions. The resulting mixture was stirred for 2 h at 25 °C. After completion of the reaction, water was added, and the mixture was basified by the addition of sat. aq. $NaHCO_3$ solution dropwise. The product was extracted with dichloromethane (×2). The combined organic layers were dried over anhydrous sodium sulfate and concentrated *in vacuo*. Finally, the crude material was purified by column chromatography using PE/EtOAc 8:2 as eluant to give 5 g of a pale-yellow solid; yield 63%; mp 60–62 °C (lett. 63–64.5 °C) [57]; IR (KBr) 2980, 2933, 1728, 1555, 1289, 1228, 1098, 1041, 860, 704 cm^{-1}; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 7.28–7.15 (m, 5-H), 3.90 (qd, J = 7.0/2.7 Hz, 4-H), 2.52 (s, 6-H), 0.80 (td, J = 7.0/2.7 Hz, 6-H) ppm; ^{13}C-NMR (75 MHz, $CDCl_3$) δ 167.9, 155.4, 146.1, 136.6, 128.4, 128.14, 128.11, 126.9, 61.3, 22.9, 13.5 ppm.

3.2.3. Synthesis of (2,6-Dimethyl-4-Phenylpyridine-3,5-Diyl)Dimethanol (**3**)

Under a nitrogen atmosphere $LiAlH_4$ (2.71 g, 71.40 mmol, 3 equiv) was added portion-wise to a stirred solution of the corresponding diethylester **2** (7.75 g, 23.8 mmol, 1 equiv) in THF dry (155.0 mL) at 0 °C. The resulting mixture was stirred for 3 h at 25 °C. Then the

reaction was quenched by dropwise addition of NaOH 2 M solution under stirring and a white precipitate was formed. The suspension was filtered through a pad of Celite, and the filter cake was washed with MeOH. The filtrate obtained was concentrated *in vacuo* and the residue was purified by column chromatography using EtOAc/MeOH 9:1 as eluant to give 5 g of product as a white solid (lett. pale yellow oil) [58]; yield 79%; mp 164–166 °C; IR (KBr) 3300–2800 br, 1571, 1560, 1443, 1410, 1009, 711 cm^{-1}; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 7.45–7.44 (m, 3-H), 7.27–7.25 (m, 2-H), 4.30 (s, 4-H), 2.65 (s, 6-H) ppm; ^{13}C-NMR (75 MHz, $CDCl_3$) δ 158.2, 153.2, 138.3, 131.0, 130.2, 129.1, 128.9, 59.5, 21.6 ppm.

3.2.4. Synthesis of 3,5-Bis(Azidomethyl)-2,6-Dimethyl-4-Phenylpyridine (**4**)

Under a nitrogen atmosphere, DPPA (diphenylphosphoryl azide) (6.64 mL, 30.86 mmol, 3 equiv) and DBU (1,8-Diazabicycloundec-7-ene) (4.60 mL, 30.86 mmol, 3 equiv) were subsequently added dropwise to a cooled (0 °C) and stirred solution of alcohol **3** (2.5 g, 10.28 mmol, 1 equiv) in 75.0 mL di *N,N′*-dimethylformamide dry. Once the additions were completed, the ice bath was removed and allowed to react at RT for 30 min, then it was heated to 80 °C for 1 h. After cooling to RT, the reaction was worked up by dilution with water and extraction with EtOAc (×2). The combined organic layers were washed with water (×3), dried over anhydrous sodium sulfate, and concentrated *in vacuo.* The residue was purified by column chromatography using PE/EtOAc 98:2 and PE/EtOAc 9:1 as eluants to give 2.74 g of product as a white solid; yield 90%, mp 74–76 °C; IR (KBr) 2959, 2932, 2104, 1554, 1488, 1458, 1260, 1193, 960, 770 cm^{-1}; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 7.49–7.46 (m, 3-H), 7.15–7.13 (m, 2-H), 4.07 (br s, 4-H), 2.64 (br s, 6-H) ppm; ^{13}C-NMR (75 MHz, $CDCl_3$) δ 157.8, 151.8, 135.8, 129.0, 128.8, 128.5, 125.0, 49.1, 22.6 ppm; ESI-MS m/z 294 $[M + H]^+$.

3.2.5. General Procedure for the Preparation of Bis-Triazolyl-Pyridines **15a**–**24a**, **15b**–**21b**, **15c**–**21c**, **15d**–**21d**

3,5-bis(azidomethyl)-2,6-dimethyl-4-phenylpyridine **4** (0.13 g, 0.44 mmol, 1 equiv) and the corresponding alkyne (0.968, 2.2 equiv) were suspended in a mixture of water/*tert*-butanol (1:1) (4 mL), sodium ascorbate of a freshly prepared 1 M solution in water (0.09 mL, 0.097 mmol, 0.22 equiv) was added, followed by the addition of copper (II) sulfate pentahydrate (0.002 g, 0.0097 mmol, 0.022 equiv). The resulting mixture was vigorously stirred at 25 °C for 3 h under nitrogen. Then the mixture was concentrated *in vacuo* and the crude was purified by column chromatography to give the desired products.

Series a: Bis-Triazolyl-Pyridines **15a**–**24a**

1,1′-((((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))bis(methylene))dipiperidine (**15a**)

The title compound was prepared from azide **4** and 1-(prop-2-yn-1-yl)piperidine (**5a**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 8:2 and EtOAc/MeOH 8:2 + 1% conc. NH_4OH as eluants to give a white solid: yield 51%; mp 218–220 °C dec; IR (KBr) 3062, 2933, 2851, 2778, 1563, 1465, 1438, 1318, 1299, 1111, 1051, 703 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.44–7.37 (m, 5-H), 7.01 (d, J = 7.0 Hz, 2-H), 5.34 (s, 4-H), 3.57 (s, 4-H), 2.59 (s, 6-H), 2.41 (br s, 8-H), 1.58 (br s, 8-H), 1.46(br s, 4-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.7, 154.4, 143.8, 136.8, 130.0, 129.9, 129.3, 126.2, 125.4, 54.8, 54.0, 49.7, 26.3, 24.8, 22.4 ppm; ESI-MS m/z 540 $[M + H]^+$. Anal. Calcd for $C_{31}H_{41}N_9$: C, 68.99; H, 7.66; N, 23.36; found: C, 69.14; H, 7.57; N, 23.49. The ^{1}H- and ^{13}C-NMR (APT) spectra of **15a** are shown in Figure S18.

2,6-dimethyl-4-phenyl-3,5-bis((4-(pyrrolidin-1-ylmethyl)-1H-1,2,3-triazol-1-yl)methyl) pyridine (**16a**)

The title compound was prepared from azide **4** and 1-(prop-2-yn-1-yl)pyrrolidine (**6a**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 8:2 and EtOAc/MeOH 7:3 + 1% conc. NH_4OH as eluants to give a white solid: yield 67%; mp 197–199 °C; yield 67%; IR (KBr) 3061, 2961, 2787, 1558, 1444, 1347, 1145, 1053, 702 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.50–7.37 (m, 5-H), 7.07 (d,

J = 7.3 Hz, 2-H), 5.35 (s, 4-H), 3.70 (s, 4-H), 2.58–2.54 (m, 14-H), 1.79 (br s, 8-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.6, 154.3, 145.2, 136.8, 130.0, 129.8, 129.3, 126.2, 124.8, 54.5, 50.6, 49.7, 24.2, 22.4 ppm; ESI-MS *m*/*z* 512 $[M + H]^{+\cdot}$. Anal. Calcd for: $C_{29}H_{37}N_9$: C, 68.07; H, 7.29; N, 24.64; found: C, 68.29; H, 7.42; N, 24.79. The ^{1}H- and ^{13}C-NMR (APT) spectra of **16a** are shown in Figure S19.

4,4′-((((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))bis(methylene))bis(1-methylpiperazine) (**17a**)

The title compound was prepared from azide **4** and 1-methyl-4-(prop-2-yn-1-yl)piperazine (**7a**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 3:7 and EtOAc/MeOH 3:7 + 1% conc. NH_4OH as eluants to give a white solid: yield 80%; mp 221–223 °C dec; IR (KBr) 3101, 3060, 2933, 2797, 2741, 1557, 1455, 1280, 1163, 1139, 1031, 822, 705 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.48–7.37 (m, 5-H), 7.06–7.04 (m, 2-H), 5.35 (s, 4-H), 3.60 (s, 4-H), 2.59 (s, 6-H), 2.50 (br s, 16-H), 2.28 (s, 6-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.6, 154.3, 144.1, 136.7, 130.0, 129.8, 129.3, 126.1, 125.2, 55.4, 53.2, 53.0, 49.6, 45.9, 22.4 ppm; ESI-MS *m*/*z* 570 $[M + H]^+$. Anal. Calcd for: $C_{31}H_{43}N_{11}$: C, 65.35; H, 7.61; N, 27.04; found: C, 65.12; H, 7.78; N, 27.19. The ^{1}H- and ^{13}C-NMR (APT) spectra of **17a** are shown in Figure S20.

4,4′-((((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))bis(methylene))bis(1-isopropylpiperazine) (**18a**)

The title compound was prepared from azide **4** and 1-isopropyl-4-(prop-2-yn-1-yl)piperazine (**8a**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 7:3 and EtOAc/MeOH 7:3 + 1% conc. NH_4OH as eluants to give an off-white solid: yield 67%; mp 214–216 °C dec; IR (KBr) 3101, 2964, 2879, 2802, 1566, 1447, 1316, 1177, 1133, 1048, 842, 705 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.47–7.40 (m, 5-H), 7.04 (d, *J* = 7.0 Hz, 2-H), 5.34 (s, 4-H), 3.60 (s, 4-H), 2.72–2.53 (m, 24-H), 1.08 (br d, 12-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.7, 154.4, 144.1, 136.8, 130.1, 129.9, 129.4, 126.2, 125.4, 56.0, 53.3, 49.7, 49.3, 22.4, 18.6 ppm; ESI-MS *m*/*z* 626 $[M + H]^+$. Anal. Calcd for: $C_{35}H_{51}N_{11}$: C, 67.17; H, 8.21; N, 24.62; found: C, 67.38; H, 8.35; N, 24.48. The ^{1}H- and ^{13}C-NMR (APT) spectra of **18a** are shown in Figure S21.

4,4′-((((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))bis(methylene))dimorpholine (**19a**)

The title compound was prepared from azide **4** and 4-(prop-2-yn-1-yl)morpholine (**9a**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 7:3 and EtOAc/MeOH 7:3 + 1% conc. NH_4OH as eluants to give a white solid: 78% yield; mp 239–241 °C dec.; IR (KBr) 3063, 2926, 2856, 2813, 1558, 1453, 1328, 1115, 1003, 863, 705 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.44–7.35 (m, 5-H), 7.00 (d, *J* = 6.7 Hz, 2-H), 5.34 (br s, 4-H), 3.66 (br s, 8-H), 3.59 (br s, 4-H), 2.59 (br s, 6-H), 2.43 (br s, 8-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.7, 154.4, 144.0, 136.7, 130.0, 129.8, 129.3, 126.2, 125.3, 67.6, 54.1, 53.8, 49.6, 22.3 ppm; ESI-MS *m*/*z* 544 $[M + H]^+$. Anal. Calcd for: $C_{29}H_{37}N_9O_2$: C, 64.07; H, 6.86; N, 23.19; found: C, 63.89; H, 7.03; N, 23.48. The ^{1}H- and ^{13}C-NMR (APT) spectra of **19a** are shown in Figure S22.

1,1′-((((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))bis(methylene))bis(piperidin-4-ol) (**20a**)

The title compound was prepared from azide **4** and 1-(prop-2-yn-1-yl)piperidin-4-ol (**10a**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 7:3 and EtOAc/MeOH 7:3 + 1% conc. NH_4OH as eluents to give a white solid: yield 67%; mp 90–92 °C; IR (KBr) 3400–2700 br, 3139, 2939, 2824, 1557, 1442, 1331, 1223, 1137, 1056, 707 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.45–7.35 (m, 5-H), 7.00 (d, *J* = 6.7 Hz, 2-H), 5.34 (s, 4-H), 3.60 (br s, 6-H), 2.75 (m, 4-H), 2.60 (s, 6-H), 2.20 (m, 4-H), 1.84 (m, 4-H), 1.54 (m, 4-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.7, 154.4, 144.2, 136.8, 130.0, 129.9, 129.3, 126.2, 125.3, 67.8, 53.3, 51.6, 49.7, 34.6, 22.4 ppm; ESI-MS *m*/*z* 572 $[M + H]^+$. Anal. Calcd for: $C_{31}H_{41}N_9O_2$: C, 65.13; H, 7.23; N, 22.05; found: C, 64.89; H, 7.36; N, 22.17. The ^{1}H- and ^{13}C-NMR (APT) spectra of **20a** are shown in Figure S23.

(((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))dimethanol (**21a**)

The title compound was prepared from azide **4** and propargyl alcohol (**11a**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 9:1 and EtOAc/MeOH 8:2 as eluants to give an amorphous yellow solid: yield 83%; IR (KBr) 3400–2800 br, 1557, 1445, 1218, 1142, 1037 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.45–7.39 (m, 5-H), 7.04 (d, J = 7.3 Hz, 2-H), 5.33 (s, 4-H), 4.59 (s, 4-H), 2.56 (s, 6-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.7, 154.4, 148.9, 136.7, 130.0, 129.9, 129.3, 126.2, 125.0, 56.3, 49.7, 22.3 ppm; ESI-MS m/z 406 $[M + H]^+$. Anal. Calcd for: $C_{21}H_{23}N_7O_2$: C, 62.21; H, 5.72; N, 24.18; found: 62.03; H, 5.89; N, 24.43. The ^{1}H- and ^{13}C-NMR (APT) spectra of **21a** are shown in Figure S24.

1,1′-(((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))bis(N,N-dimethylmethanamine) (**22a**)

The title compound was prepared from azide **4** and 1-dimethylamino-2-propyne (**12a**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 8:2 and EtOAc/MeOH 8:2 + 1% conc. NH_4OH as eluants to give a white solid: yield 48%; mp 184–186 °C; IR (KBr) 3110, 3069, 2941, 2814, 2763, 1561, 1456, 1142, 1052, 1036, 841, 704 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.51–7.05 (m, 5-H), 7.06 (d, J = 6.7 Hz, 2-H), 5.36 (s, 4-H), 3.56 (s, 4-H), 2.58 (s, 6-H), 2.22 (br s, 12-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.6, 154.3, 144.3, 136.7, 130.0, 129.8, 129.3, 126.1, 125.2, 54.1, 49.7, 44.8, 22.4 ppm; ESI-MS m/z 460 $[M + H]^+$. Anal. Calcd for: $C_{25}H_{33}N_9$: C, 65.33; H, 7.24; N, 27.43; found: C, 65.59; H, 7.39; N, 27.22. The ^{1}H- and ^{13}C-NMR (APT) spectra of **22a** are shown in Figure S25.

2,2′,2″,2‴-(((((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))bis(methylene))bis(azanetriyl))tetrakis(ethan-1-ol) (**23a**)

The title compound was prepared from azide **4** and 2,2′-(prop-2-yn-1-ylazanediyl)bis(ethan-1-ol) (**13a**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 7:3 and EtOAc/MeOH 7:3 + 1% conc. NH_4OH as eluants to give a yellow oil: yield 66%; IR (KBr) 3500–2800 br, 1576, 1443, 1336, 1219, 1141, 1044, 764, 708 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.48 (s, 2-H), 7.41–7.39 (m, 3-H), 7.01 (d, J = 6.7 Hz, 2-H), 5.33 (br s, 4-H), 3.78 (br s, 4-H), 3.61–3.58 (m, 8-H), 2.59 (br s, 14-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.7, 154.5, 145.5, 136.7, 130.1, 129.9, 129.3, 126.2, 125.0, 60.4, 57.1, 49.8, 49.7, 22.3 ppm; ESI-MS m/z 602 $[M + Na]^+$. Anal. Calcd for: $C_{29}H_{41}N_9O_4$: C, 60.09; H, 7.13; N, 21.75; found: C, 60.32; H, 6.95; N, 21.58. The ^{1}H- and ^{13}C-NMR (APT) spectra of **23a** are shown in Figure S26.

4,4′-((((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))bis(methylene))bis(1-(methylsulfonyl)piperazine) (**24a**)

The title compound was prepared from azide **4** and 1-(methylsulfonyl)-4-(prop-2-yn-1-yl)piperazine (**14a**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 7:3 and EtOAc/MeOH 7:3 + 1% conc. NH_4OH as eluants to give a light orange oil: yield 67%; mp 244–246 °C dec.; IR (KBr) 3067, 2925, 2850, 2822,1561, 1455, 1342, 1323, 1156, 1133, 960, 785 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.37–7.35 (m, 3-H), 6.99–6.94 (m, 4-H), 5.21 (br s, 4-H), 3.61 (br s, 4-H), 3.19 (br s, 8-H), 2.74 (br s, 6-H), 2.55 (br s, 14-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 158.8, 152.2, 143.4, 135.4, 129.2, 129.1, 128.1, 124.0, 122.5, 52.7, 52.0, 48.8, 45.7, 34.4, 22.9 ppm; ESI-MS m/z 698 $[M + H]^+$. Anal. Calcd for: $C_{31}H_{43}N_{11}O_4S_2$: C, 53.35; H, 6.21; N, 22.08; found: C, 53.54; H, 5.98; N, 22.23. The ^{1}H- and ^{13}C-NMR (APT) spectra of **24a** are shown in Figure S27.

Series b: Bis-Triazolyl-Pyridines **15b**–**21b**

1,1′-((((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))bis(ethane-2,1-diyl))dipiperidine (**15b**)

The title compound was prepared from azide **4** and 1-(but-3-yn-1-yl)piperidine (**5b**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 8:2 and EtOAc/MeOH 8:2 + 1% conc. NH_4OH as eluants to give a

white solid: yield 92%; mp 214–216 °C dec.; IR (KBr) 3110, 3058, 2933, 2804, 2769, 1560, 1438, 1143, 1055, 1030, 703 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.43–7.30 (m, 5-H), 6.99 (d, J = 6.7 Hz, 2-H), 5.30 (s, 4-H), 2.82 (t, J = 7.6 Hz, 4-H), 2.60–2.48 (m, 18-H), 1.60 (br s, 8-H), 1.48 (br s, 4-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.6, 154.3, 146.7, 136.8, 130.0, 129.8, 129.3, 126.3, 123.6, 59.6, 55.2, 26.5, 25.1, 23.5, 22.3 ppm; ESI-MS m/z 568 $[M + H]^+$. Anal. Calcd for: $C_{33}H_{45}N_9$: C, 69.81; H, 7.99; N, 22.20; found: C, 69.69; H, 8.21; N, 21.97. The ^{1}H- and ^{13}C-NMR (APT) spectra of **15b** are shown in Figure S28.

2,6-dimethyl-4-phenyl-3,5-bis((4-(2-(pyrrolidin-1-yl)ethyl)-1H-1,2,3-triazol-1-yl)methyl)pyridine (**16b**)

The title compound was prepared from azide **4** and 1-(but-3-yn-1-yl)pyrrolidine (**6b**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 7:3 and EtOAc/MeOH 7:3 + 1% conc. NH_4OH as eluants to give an off-white solid: yield 69%; mp 182–184 °C; IR (KBr) 3112, 3062, 2955, 2787, 1562, 1442, 1146, 1052, 1033, 841, 702 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.41–7.31 (m, 5-H), 7.01 (d, J = 6.7 Hz, 2-H), 5.30 (s, 4-H), 2.87–2.82 (m, 4-H), 2.76–2.71 (m, 4-H), 2.61–2.57 (m, 14-H), 1.82 (br s, 8-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.6, 154.4, 146.5, 136.8, 130.0, 129.8, 129.3, 126.3, 123.6, 56.7, 54.9, 25.6, 24.2, 22.3 ppm; ESI-MS m/z 540 $[M + H]^+$. Anal. Calcd for: $C_{31}H_{41}N_9$: C, 68.99; H, 7.66; N, 23.36; found: C, 69.25; H, 7.82; N, 23.17. The ^{1}H- and ^{13}C-NMR (APT) spectra of **16b** are shown in Figure S29.

4,4′-((((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))bis(ethane-2,1-diyl))bis(1-methylpiperazine) (**17b**)

The title compound was prepared from azide **4** and 1-(but-3-yn-1-yl)-4-methylpiperazine (**7b**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 3:7 and EtOAc/MeOH 3:7 + 1% conc. NH_4OH as eluants to give an off-white solid: yield 81%; mp 149–152 °C dec.; IR (KBr) 2937, 2810, 2690, 1556, 1455, 1284, 1164, 1146, 1055, 1009, 839 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.42–7.35 (m, 5-H), 7.02 (d, J = 6.4 Hz, 2-H), 5.31 (s, 4-H), 2.82 (t, J = 7.4 Hz, 4-H) 2.63–2.56 (m, 26-H), 2.30 (s, 6-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.5, 154.2, 146.6, 136.8, 130.0, 129.8, 129.3, 126.3, 123.7, 58.5, 55.6, 53.3, 49.5, 45.9, 23.7, 22.4 ppm; ESI-MS m/z 598 $[M + H]^+$. Anal. Calcd for: $C_{33}H_{47}N_{11}$: C, 66.30; H, 7.92; N, 25.77; found: C, 66.12; H, 8.16; N, 26.09. The ^{1}H- and ^{13}C-NMR (APT) spectra of **17b** are shown in Figure S30.

4,4′-((((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))bis(ethane-2,1-diyl))bis(1-isopropylpiperazine) (**18b**)

The title compound was prepared from azide **4** and 1-(but-3-yn-1-yl)-4-isopropylpiperazine (**8b**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 7:3 and EtOAc/MeOH 7:3 + 1% conc. NH_4OH as eluants to give a white solid: yield 82%; mp 224–226 °C dec.; IR (KBr) 3113, 3058, 2960, 2938, 2813, 1561, 1464, 1448, 1271, 1178, 1147, 1055, 1030, cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.41–7.34 (m, 5-H), 7.01 (d, J = 6.7 Hz, 2-H), 5.30 (s, 4-H), 2.82 (t, J = 6.7 Hz, 4-H) 2.69–2.56 (m, 28-H), 1.07 (br d, 12-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.5, 154.3, 146.6, 136.8, 130.0, 129.8, 129.3, 126.3, 123.6, 58.6, 55.9, 53.7, 49.5, 49.4, 23.6, 22.3, 18.7 ppm; ESI-MS m/z 654 $[M + H]^+$. Anal. Calcd for: $C_{37}H_{55}N_{11}$: C, 67.96; H, 8.48; N, 23.56; found: C, 67.84; H, 8.65; N, 23.31. The ^{1}H- and ^{13}C-NMR (APT) spectra of **18b** are shown in Figure S31.

4,4′-((1,1′-((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-4,1-diyl))bis(ethane-2,1-diyl))dimorpholine (**19b**)

The title compound was prepared from azide **4** and 4-(but-3-yn-1-yl)morpholine (**9b**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 8:2 as eluant to give a white solid: yield 46%; mp 200–204 °C dec.; IR (KBr) 3116, 3064, 2942, 2852, 2812, 1564, 1461, 1446, 1274, 1212, 1115, 1006, 869, 703 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.43–7.38 (m, 3-H), 7.29 (s, 2-H), 7.00 (d, J = 6.7 Hz, 2-H), 5.30 (s, 4-H), 3.66 (br s, 8-H), 2.82 (t, J = 7.4 Hz, 4-H), 2.61–2.57 (m, 12-H), 2.48 (br s, 6-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.6, 154.3, 146.6, 136.8, 130.0, 129.9, 129.3, 126.3, 123.7, 67.7, 59.1, 54.5, 49.6, 23.4, 22.4 ppm; ESI-MS m/z 572 $[M + H]^+$. Anal. Calcd for:

$C_{31}H_{41}N_9O_2$: C, 65.13; H, 7.23; N, 22.05; found: C, 65.29; H, 7.48; N, 21.86. The ^{1}H- and ^{13}C-NMR (APT) spectra of **19b** are shown in Figure S32.

1,1′-((1,1′-((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-4,1-diyl))bis(ethane-2,1-diyl))bis(piperidin-4-ol) (**20b**)

The title compound was prepared from azide **4** and 1-(but-3-yn-1-yl)piperidin-4-ol (**10b**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 7:3 + 1% conc. NH_4OH as eluant to give a white solid: yield 77%; mp 58–62 °C; IR (KBr) 3135, 2939, 2813, 1554, 1444, 1340, 1218, 1136, 1072, 762, 707, cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.41–7.38 (m, 3-H), 7.33 (s, 2-H), 7.00 (d, J = 6.7 Hz, 2-H), 5.31 (s, 4-H), 3.64 (br s, 2-H), 2.85–2.80 (m, 8-H), 2.63–2.57 (m, 10-H), 2.24 (br t, 4-H), 1.87 (br d, 4-H), 1.61–1.55 (m, 4-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.6, 154.3, 146.6, 136.8, 130.0, 129.9, 129.3, 126.3, 123.7, 68.0, 58.7, 52.0, 49.6, 34.7, 23.8, 22.4 ppm; ESI-MS m/z 600 $[M + H]^+$. Anal. Calcd for: $C_{33}H_{45}N_9O_2$: C, 66.08; H, 7.56; N, 21.02; found: C, 65.84; H, 7.73; N, 21.28. The ^{1}H- and ^{13}C-NMR (APT) spectra of **20b** are shown in Figure S33.

2,2′-(((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))bis(ethan-1-ol) (**21b**)

The title compound was prepared from azide **4** and 3-butyn-1-ol (**11b**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 9:1 and EtOAc/MeOH 8:2 eluants to give an amorphous off-white solid: yield 88%; IR (KBr) 2800–3200 br, 1556, 1444, 1218, 1131, 1053, 707cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.40–7.29 (m, 5-H), 7.01 (d, J = 7.0 Hz, 2-H), 5.30 (s, 4-H), 3.73; (td, J = 6.7/1.8 Hz, 4-H), 2.81(td, J = 6.7/1.8 Hz, 4-H), 2.57 (s, 6-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.6, 154.4, 146.0, 136.7, 130.0, 129.8, 129.3, 126.3, 123.9, 61.9, 49.5, 29.7, 22.3 ppm; ESI-MS m/z 434 $[M + H]^+$. Anal. Calcd for: $C_{23}H_{27}N_7O_2$: C, 63.72; H, 6.28; N, 22.62; found: C, 63.44; H, 6.42; N, 22.32. The ^{1}H- and ^{13}C-NMR (APT) spectra of **21b** are shown in Figure S34.

Series c: Bis-Triazolyl-Pyridines **15c**–**21c**

1,1′-((((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))bis(propane-3,1-diyl))dipiperidine (**15c**)

The title compound was prepared from azide **4** and 1-(pent-4-yn-1-yl)piperidine (**5c**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 8:2 and EtOAc/MeOH 8:2 + 1% conc. NH_4OH as eluants to give a white solid: yield 79%; mp 159–161 °C; IR (KBr) 3058, 2931, 2850, 2800, 2762, 1562, 1441, 1150, 1110, 1053, 1037, 702 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.43–7.29 (m, 5-H), 7.02 (d, J = 7.0 Hz, 2-H), 5.30 (s, 4-H), 2.62 (t, J = 7.0 Hz, 4-H), 2.56 (br s, 6-H), 2.43 (br s, 8-H), 2.35 (d, J = 7.2 Hz, 4-H), 1.81 (quint, J = 7.0 Hz, 4-H), 1.59 (br s, 8-H), 1.47 (br s, 4-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.5, 154.3, 148.4, 136.8, 130.0, 129.8, 129.3, 126.2, 123.2, 59.5, 55.4, 49.5, 27.1, 26.4, 25.1, 24.2, 22.3 ppm; ESI-MS m/z 596 $[M + H]^+$. Anal. Calcd for: $C_{35}H_{49}N_9$: C, 70.55; H, 8.29; N, 21.16; found: C, 70.32; H, 8.53; N, 21.35. The ^{1}H- and ^{13}C-NMR (APT) spectra of **15c** are shown in Figure S35.

2,6-dimethyl-4-phenyl-3,5-bis((4-(3-(pyrrolidin-1-yl)propyl)-1H-1,2,3-triazol-1-yl)methyl)pyridine (**16c**)

The title compound was prepared from azide **4** and 1-(pent-4-yn-1-yl)pyrrolidine (**6c**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 6:4 and EtOAc/MeOH 6:4 + 1% conc. NH_4OH as eluants to give a dark-beige solid: yield 47%; mp 130–133 °C; IR (KBr) 3107, 3056, 2941, 2775, 1557, 1444, 1144, 1052, 1029, 840, 702 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.44–7.31 (m, 5-H), 7.02 (d, J = 6.5 Hz, 2-H), 5.31 (s, 4-H), 2.68–2.53 (m, 22-H), 1.89–1.82 (m, 12-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.5, 154.2, 148.2, 136.8, 130.0, 129.8, 129.3, 123.3, 56.6, 54.9, 49.5, 29.1, 24.1, 22.3 ppm; ESI-MS m/z 568 $[M + H]^+$. Anal. Calcd for: $C_{33}H_{45}N_9$: C, 69.81; H, 7.99; N, 22.20; found: C, 70.03; H, 8.12; N, 22.47. The ^{1}H- and ^{13}C-NMR (APT) spectra of **16c** are shown in Figure S36.

4,4′-((((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1,4-diyl))bis(propane-3,1-diyl))bis(1-methylpiperazine) (**17c**)

The title compound was prepared from azide **4** and 1-methyl-4-(pent-4-yn-1-yl)pipera zine (**7c**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 3:7 and EtOAc/MeOH 3:7 + 1% conc. NH_4OH as eluants to give a white solid: yield 72%; mp 139–141 °C; IR (KBr) 3109, 3057, 2937, 2794, 1558, 1457, 1296, 1161, 1051, 814, 705 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.41–7.30 (m, 5-H), 7.02 (d, J = 6.4 Hz, 2-H), 5.30 (s, 4-H), 2.64; (t, J = 7.3 Hz, 4-H), 2.56 (br s, 6-H), 2.50 (br s, 8-H), 2.37 (t, J = 7.0 Hz, 4-H), 2.28 (br s, 6-H), 1.80 (quint, J = 7.3 Hz, 4-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.5, 154.2, 148.4, 136.8, 130.0, 129.8, 129.3, 126.2, 123.2, 58.6, 55.5, 53.6, 49.5, 45.9, 27.2, 24.0, 22.4 ppm; ESI-MS m/z 626 $[M + H]^+$. Anal. Calcd for: $C_{35}H_{51}N_{11}$: C, 67.17; H, 8.21; N, 24.62; found: C, 66.99; H, 8.37; N, 24.45. The ^{1}H- and ^{13}C-NMR (APT) spectra of **17c** are shown in Figure S37.

4,4′-((((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1, 4-diyl))bis(propane-3,1-diyl))bis(1-isopropylpiperazine) (**18c**)

The title compound was prepared from azide **4** and 1-isopropyl-4-(pent-4-yn-1-yl)pipe razine (**8c**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 7:3 and EtOAc/MeOH 7:3 + 1% conc. NH_4OH as eluants to give a white solid: yield 79%; mp 153–156 °C; IR (KBr) 3115, 3061, 2943, 2804, 2770, 1560, 1442, 1268, 1182, 1053, 984, 703 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.42–7.32 (m, 5-H), 7.06 (d, J = 6.1 Hz, 2-H), 5.32 (s, 4-H), 2.67–2.37 (m, 32-H), 1.81 (m, 4-H), 1.09–1.06 (m, 12-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.4, 154.0, 148.3, 136.8, 130.0, 129.8, 129.3, 126.2, 123.1, 58.6, 55.8, 53.9, 49.5, 49.4, 27.2, 24.0, 22.5, 18.8 ppm; ESI-MS m/z 682 $[M + H]^+$. Anal. Calcd for: $C_{39}H_{59}N_{11}$: C, 68.69; H, 8.72; N, 22.59; found: C, 68.94; H, 8.55; N, 22.35. The ^{1}H- and ^{13}C-NMR (APT) spectra of **18c** are shown in Figure S38.

4,4′-((1,1′-((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazo le-4,1-diyl))bis(propane-3,1-diyl))dimorpholine (**19c**)

The title compound was prepared from azide **4** and 4-(pent-4-yn-1-yl)morpholine (**9c**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 9:1 + 1% conc. NH_4OH as eluant to give a white solid: yield 52%; mp 164–166 °C; IR (KBr) 3114, 3061, 2943, 2849, 2809, 1710, 1562, 1453, 1293, 1137, 1117, 864, 706, cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.44–7.39 (m, 3-H), 7.29 (s, 2-H), 7.01 (d, J = 6.4 Hz, 2-H), 5.30 (s, 4-H), 3.69–3.66 (m, 8-H), 2.64 (t, J = 7.3, 4-H), 2.56 (br s, 6-H), 2.44 (br s, 8-H), 2.35 (t, J = 7.3, 4-H), 1.80 (quint, J = 7.7 Hz, 4-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.6, 154.4, 148.5, 136.9, 130.1, 129.9, 129.3, 126.4, 123.4, 67.6, 59.2, 54.8, 49.6, 27.0, 24.0, 22.4 ppm; ESI-MS m/z 600 $[M + H]^+$. Anal. Calcd for: $C_{33}H_{45}N_9O_2$: C, 66.08; H, 7.56; N, 21.02; found: C, 65.82; H, 7.76; N, 20.91. The ^{1}H- and ^{13}C-NMR (APT) spectra of **19c** are shown in Figure S39.

1,1′-((1,1′-((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazo le-4,1-diyl))bis(propane-3,1-diyl))bis(piperidin-4-ol) (**20c**)

The title compound was prepared from azide **4** and 1-(pent-4-yn-1-yl)piperidin-4-ol (**10c**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 8:2 + 1% conc. NH_4OH and EtOAc/MeOH 7:3 + 1% conc. NH_4OH as eluants to give a pale yellow solid: yield 57%; mp 78–81 °C; IR (KBr) 3300, 3134, 2940, 2812, 1672, 1561, 1444, 1340, 1217, 1135, 1062, 763, 707, cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.41–7.39 (m, 3-H), 7.28 (s, 2-H), 7.01 (d, J = 7.3 Hz, 2-H), 5.30 (s, 4-H), 3.63 (m, 2-H), 2.82–2.78 (m, 4-H), 2.63 (t, J = 7.4 Hz, 4-H), 2.56 (br s, 6-H), 2.36 (t, J = 7.1 Hz, 4-H), 2.19 -2.13 (m, 4-H), 1.93–1.78 (m, 8-H), 1.61–1.51 (m, 4-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.5, 154.3, 148.4, 136.8, 130.0, 129.8, 129.3, 126.3, 123.3, 68.1, 58.7, 52.2, 49.6, 34.7, 27.4, 24.1, 22.4 ppm; ESI-MS m/z 628 $[M + H]^+$. Anal. Calcd for: $C_{35}H_{49}N_9O_2$: C, 66.96; H, 7.87; N, 20.08; found: C, 66.83; H, 7.99; N, 20.27. The ^{1}H- and ^{13}C-NMR (APT) spectra of **20c** are shown in Figure S40.

3,3′-(((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-1, 4-diyl))bis(propan-1-ol) (**21c**)

The title compound was prepared from azide **4** and 4-pentyn-1-ol (**11c**) according to the general procedure. The crude was purified by column chromatography using

EtOAc/MeOH 9:1 and EtOAc/MeOH 8:2 as eluants to give an amorphous yellow solid: yield 88%; IR (KBr) 3400–3000 br, 2944, 2870, 1556, 1442, 1216, 1054, 707 cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.42–7.35 (m, 3-H), 7.26 (s, 2-H), 6.98 (d, J = 7.0 Hz, 2-H), 5.30 (s, 4-H), 3.54 (td, J = 6.4/2.1 Hz, 4-H), 2.68 (t, J = 6.4 Hz, 4-H), 2.57 (s, 6-H), 1.80 (quint, J = 6.4 Hz, 4-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.5, 154.3, 148.5, 136.7, 130.0, 129.8, 129.2, 126.3, 123.2, 61.9, 49.5, 33.2, 22.5, 22.3 ppm; ESI-MS m/z 462 $[M + H]^+$. Anal. Calcd for: $C_{25}H_{31}N_7O_2$: C, 65.06; H, 6.77; N, 21.24; found: C, 65.32; H, 6.93; N, 20.93. The ^{1}H- and ^{13}C-NMR (APT) spectra of **21c** are shown in Figure S41.

Series d: Bis-Triazolyl-Pyridines **15d**–**21d**

1,1′-((1,1′-((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-4,1-diyl))bis(butane-4,1-diyl))dipiperidine (**15d**)

The title compound was prepared from azide **4** and 1-(hex-5-yn-1-yl)piperidine (**5d**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 9:1 and EtOAc/MeOH 8:2 + 1% conc. NH_4OH as eluants to give a pale yellow solid: yield 65%; mp 173–176 °C dec.; IR (KBr) 3116, 3062, 2930, 2856, 2768, 1700, 1561, 1440, 1146, 1053, 702, cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.44–7.38 (m, 3-H), 7.28 (s, 2-H), 7.02 (d, J = 6.5 Hz, 2-H), 5.30 (s, 4-H), 2.66–2.55 (m, 10-H), 2.42–2.31 (m, 12-H), 1.60–1.48 (m, 20-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.6, 154.4, 148.7, 136.9, 130.0, 129.9, 129.4, 126.3, 123.3, 60.1, 55.5, 49.6, 28.6, 26.8, 26.4, 26.0, 25.2, 22.3 ppm; ESI-MS m/z 624 $[M + H]^+$. Anal. Calcd for: $C_{37}H_{53}N_9$: C, 71.23; H, 8.56; N, 20.21; found: C, 71.05; H, 8.79; N, 20.48. The ^{1}H- and ^{13}C-NMR (APT) spectra of **15d** are shown in Figure S42.

2,6-dimethyl-4-phenyl-3,5-bis((4-(4-(pyrrolidin-1-yl)butyl)-1H-1,2,3-triazol-1-yl)methyl)pyridine (**16d**)

The title compound was prepared from azide **4** and 1-(hex-5-yn-1-yl)pyrrolidine (**6d**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 8:2 and EtOAc/MeOH 7:3 + 1% conc. NH_4OH as eluants to give a pale yellow solid: yield 49%; mp 138–141 °C; IR (KBr) 3114, 3062, 2929, 2858, 1618, 1562, 1462, 1219, 1144, 1051, 700, 587, cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.44–7.39 (m, 3-H), 7.30 (s, 2-H), 7.03 (d, J = 7.1 Hz, 2-H), 5.30 (s, 4-H), 2.67–2.50 (m, 22-H), 1.81 (br s, 8-H), 1.63–1.55 (m, 8-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.6, 154.3, 148.7, 136.9, 130.0, 129.9, 129.4, 126.3, 123.3, 57.1, 55.0, 49.6, 28.9, 28.5, 26.0, 24.1, 22.3 ppm; ESI-MS m/z 596 $[M + H]^+$. Anal. Calcd for: $C_{35}H_{49}N_9$: C, 70.55; H, 8.29; N, 21.16; found: C, 70.43; H, 8.15; N, 21.41. The ^{1}H- and ^{13}C-NMR (APT) spectra of **16d** are shown in Figure S43.

4,4′-((1,1′-((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-4,1-diyl))bis(butane-4,1-diyl))bis(1-methylpiperazine) (**17d**)

The title compound was prepared from azide **4** and 1-(hex-5-yn-1-yl)-4-methylpiperazine (**7d**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 8:2 + 1% conc. NH_4OH as eluant to give a yellow solid: yield 76%; mp 146–149 °C dec; IR (KBr) 3115, 3063, 2934, 2860, 2793, 1618, 1561, 1459, 1446, 1280, 1164, 820, 700, cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.41–7.39 (m, 3-H), 7.28 (s, 2-H), 7.01 (d, J = 7.3 Hz, 2-H), 5.30 (s, 4-H), 2.64–2.32 (m, 36-H), 1.65–1.52 (m, 8-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.6, 154.4, 148.7, 136.9, 130.0, 129.8, 129.3, 126.3, 123.3, 59.0, 55.4, 53.5, 49.6, 45.8, 28.4, 26.8, 25.9, 22.3 ppm; ESI-MS m/z 654 $[M + H]^+$. Anal. Calcd for: $C_{37}H_{55}N_{11}$: C, 67.96; H, 8.48; N, 23.56; found: C, 67.65; H, 8.66; N, 23.75. The ^{1}H- and ^{13}C-NMR (APT) spectra of **17d** are shown in Figure S44.

4,4′-((1,1′-((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-4,1-diyl))bis(butane-4,1-diyl))bis(1-isopropylpiperazine) (**18d**)

The title compound was prepared from azide **4** and 1-(hex-5-yn-1-yl)-4-isopropylpiperazine (**8d**) according to the general procedure. The crude was purified by column chromatography using EtOAc and EtOAc/MeOH 9:1 + 1% conc. NH_4OH as eluants to give a white solid: yield 61%; mp189–193 °C; IR (KBr) 3116, 3061, 2962, 2931, 2811, 1561, 1467, 1446, 1269, 1183, 1146, 1054, 703, cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.42–7.39 (m, 3-H), 7.28 (s, 2-H), 7.02 (d, J = 6.8 Hz, 2-H), 5.30 (s, 4-H), 2.66–2.37 (m, 32-H), 1.62–1.51 (m, 8-H),

1.08 (d, J = 6.5 Hz, 12-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.6, 154.4, 148.7, 136.9, 130.1, 129.9, 129.4, 126.3, 123.3, 59.3, 55.9, 54.0, 49.6, 49.4, 28.5, 26.9, 26.0, 22.4, 18.7 ppm; ESI-MS m/z 711 $[M + H]^+$. Anal. Calcd for: $C_{41}H_{63}N_{11}$: C, 69.36; H, 8.94; N, 21.70; found: C, 69.55; H, 8.73; N, 21.89. The ^{1}H- and ^{13}C-NMR (APT) spectra of **18d** are shown in Figure S45.

4,4′-((1,1′-((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-4,1-diyl))bis(butane-4,1-diyl))dimorpholine (**19d**)

The title compound was prepared from azide **4** and 4-(hex-5-yn-1-yl)morpholine (**9d**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 9:1 and EtOAc/MeOH 9:1 + 1% conc. NH_4OH as eluants to give a white solid: yield 62%; mp 152–156 °C; IR (KBr) 3060, 2931, 2855, 2808, 1703, 1560, 1445, 1115, 1053, 1031, 865, 702, 589, cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.41–7.39 (m, 3-H), 7.29 (s, 2-H), 7.04 (d, J = 5.8 Hz, 2-H), 5.30 (s, 4-H), 3.67–3.66 (m, 8-H), 2.64 (t, J = 6.7 Hz, 4-H), 2.55 (br s, 6-H), 2.42–2.33 (m, 12-H), 1.66–1.61 (m, 4-H), 1.54–1.51 (m, 4-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.5, 154.2, 148.7, 136.9, 130.0, 129.8, 129.3, 126.3, 123.2, 67.6, 59.7, 54.7, 49.6, 28.3, 26.6, 26.0, 22.4 ppm; ESI-MS m/z 628 $[M + H]^+$. Anal. Calcd for: $C_{35}H_{49}N_9O_2$: C, 66.96; H, 7.87; N, 20.08; found: C, 67.25; H, 7.99; N, 20.25. The ^{1}H- and ^{13}C-NMR (APT) spectra of **19d** are shown in Figure S46.

1,1′-((1,1′-((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-4,1-diyl))bis(butane-4,1-diyl))bis(piperidin-4-ol) (**20d**)

The title compound was prepared from azide **4** and 1-(hex-5-yn-1-yl)piperidin-4-ol (**10d**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 9:1 + 1% conc. NH_4OH and EtOAc/MeOH 7:3 + 1% conc. NH_4OH as eluants to give a white solid: yield 87%; mp 159–162 °C; IR (KBr) 3406, 3058, 2940, 2765, 1557, 1444, 1362, 1130, 1062, 1031, 699, 584, cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.44–7.36 (m, 3-H), 7.28 (s, 2-H), 7.02 (d, J = 6.4 Hz, 2-H), 5.30 (s, 4-H), 3.64–3.62 (m, 2-H), 2.83–2.79 (m, 4-H), 2.64 (t, J = 7.4 Hz, 4-H), 2.56 (br s, 6-H), 2.37 (t, J = 7.4 Hz, 4-H), 2.20–2.14 (m, 4-H), 1.88–1.84 (m, 4-H), 1.61–1.51 (m, 12-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.6, 154.3, 148.7, 136.9, 130.0, 129.8, 129.3, 126.3, 123.2, 68.0, 59.2, 52.1, 49.6, 34.6, 28.5, 27.0, 25.9, 22.4 ppm; ESI-MS m/z 656 $[M + H]^+$. Anal. Calcd for: $C_{37}H_{53}N_9O_2$: C, 67.76; H, 8.15; N, 19.22; found: C, 67.89; H, 8.23; N, 18.97. The ^{1}H- and ^{13}C-NMR (APT) spectra of **20d** are shown in Figure S47.

4,4′-(1,1′-((2,6-dimethyl-4-phenylpyridine-3,5-diyl)bis(methylene))bis(1H-1,2,3-triazole-4,1-diyl))bis(butan-1-ol) (**21d**)

The title compound was prepared from azide **4** and 5-hexyn-1-ol (**11d**) according to the general procedure. The crude was purified by column chromatography using EtOAc/MeOH 9:1 as eluant to give a white solid: yield 95%; mp 105–108 °C; IR (KBr) 3275, 3109, 3055, 2923, 2853, 1554, 1442, 1215, 1146, 1076, 1048, 842, 703, 589, cm^{-1}; ^{1}H-NMR (300 MHz, CD_3OD) δ 7.40–7.38 (m, 3-H), 7.28 (s, 2-H), 7.01 (d, J = 6.4 Hz, 2-H), 5.30 (s, 4-H), 3.56 (t, J = 6.4 Hz, 4-H), 2.63 (t, J = 7.3 Hz, 4-H), 2.57 (s, 6-H), 1.66 (quint, J = 7.4 Hz, 4-H), 1.52 (quint, J = 7.4 Hz, 4-H) ppm; ^{13}C-NMR (75 MHz, CD_3OD) δ 159.5, 154.3, 148.8, 136.8, 130.0, 129.8, 129.3, 126.3, 123.2, 62.4, 49.6, 33.0, 26.8, 25.9, 22.4 ppm; ESI-MS m/z 512 $[M + Na]^+$. Anal. Calcd for: $C_{27}H_{35}N_7O_2$: C, 66.23; H, 7.21; N, 20.03; found: C, 66.42; H, 7.01; N, 20.36. The ^{1}H- and ^{13}C-NMR (APT) spectra of **21d** are shown in Figure S48.

3.3. Oligonucleotide Synthesis and Sample Preparation

The following deoxyribonucleotide sequences were selected for the experiments: the promoter sequences of d(GGG CGC GGG AGG AAT TGG GCG GG) (*Bcl-2 G4*), d(TGA GGG TGG GTA GGG TGG GTA A) (*c-Myc G4*), d(TAG GGT TAG GGT TAG GGT TAG GG) (*Tel$_{23}$ G4*), d(CAG CCC CGC TCC CGC CCC CTT CCT CCC GCG CCC GCC CCT) (*Bcl-2 iM*), d(TTC CCC ACC CTC CCT ACC CTA A) (*c-Myc iM*), d(CCC TAA CCC TAA CCC TAA CCC T) (*hTeloC iM*), d(CGC GAA TTC GCG TTT CGC GAA TTC GCG) (*Hairpin*). The oligonucleotides were chemically synthesized at a 5-µmol scale on an ABI 394 DNA/RNA synthesizer (Applied Biosystem, Foster City, CA, USA), by using the standard β-cyanoethylphosphoramidite solid-phase chemistry, as described elsewhere [59].

The subsequent DNA detachment from support and removal of the semi-permanent protection groups were achieved upon treatment with an aqueous solution of concentrated ammonia at 55 °C, for 12 h. After being combined and concentrated under reduced pressure, the filtrates and the washings were solubilized in water and then purified by high-performance liquid chromatography (HPLC) equipped with a Nucleogel SAX 1000-8/46 column (Macherey-Nagel, GmbH & Co. KG, Düren, Germany). Different buffers were used for the purification step: buffer A, consisting of a 20 mM MH_2PO_4/M_2HPO_4 aqueous solution (pH 7.0) (where M stands for K^+ or Na^+ ion), and buffer B, consisting of 1.0 M MCl, 20 mM MH_2PO_4/M_2HPO_4 aqueous solution (pH 7.0). Both buffer A and B also contained 20% (*v*/*v*) CH_3CN. A 30 min linear gradient going from 0 to 100% buffer B with a flow rate of 1 mL/min was employed. Desalting of the purified fractions was obtained by means of Sep-pak cartridges (C-18). The purity of the isolated oligomers was checked by NMR and proved to be higher than 98%. The oligonucleotide concentrations were established by measuring the UV absorption at 90 °C, taking into account the appropriate molar extinction coefficient values ε (λ = 260 nm) calculated by the nearest neighbor model [60]. The DNA samples were then dissolved in the proper buffer, as described below. Finally, to achieve the correct folding of the DNA sequences, the oligonucleotide solutions were heated in a water bath at 90 °C, for 5 min, and then left to cool down slowly, at RT, overnight.

3.4. Circular Dichroism (CD) Experiments

The samples for CD measurements were prepared by dissolving *Bcl-2 G4*, *c-Myc G4,* and *Tel_{23} G4* in 25 mM potassium phosphate buffer (pH 7.0), and *Bcl-2 iM*, *c-Myc iM*, and *hTeloC iM* in 10 mM sodium phosphate buffer (pH 5.0) [61], and the *Hairpin* in 10 mM sodium phosphate buffer (pH 7.0). Circular dichroism (CD) experiments (spectra and melting) were performed 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. Data was obtained from samples at 15–20 μM final oligonucleotide concentration, in the absence and presence of 10 molar equivalents of each compound (2.5, 5, or 10 mM in 100% DMSO). CD spectra were recorded at 5 and 90 °C for *Bcl-2 iM*, *c-Myc iM*, and *hTeloC iM*, and at 20 °C and 100 °C for *Bcl-2 G4*, *c-Myc G4*, *Tel_{23} G4*, and *Hairpin*. Each spectrum was acquired in a wavelength range of 220–360 nm, averaged over three scans, and subtracted from the buffer baseline. The scanning speed was set to 100 nm/min, with a 4 s response time, 1 nm data pitch, and 2 nm bandwidth. CD melting experiments were performed at 1 °C/min heating rate, in the 5–90 °C temperature range for *Bcl-2 iM*, *c-Myc iM*, and *hTeloC iM*, and in the 20–100 °C temperature range for *Bcl-2 G4*, *c-Myc G4*, *Tel_{23} G4,* and *Hairpin*. Changes in the CD signal were followed at the wavelengths of the maximum CD intensity: 264 nm for *Bcl-2 G4* and *c-Myc G4*, 287 nm for *Tel_{23} G4*, 288 nm for *Bcl-2 iM*, *c-Myc iM,* and *hTeloC iM*. As for *Hairpin*, CD melting curves were recorded by following the change in the CD signal at 251 nm, the wavelength of the minimum intensity value of the respective CD spectrum. All CD melting curves were normalized between 0 and 1. The apparent melting temperatures ($T_{1/2}$) were mathematically calculated by using the curve fitting function in Origin 7.0 software (OriginLab, Northampton, MA, USA). The $T_{1/2}$ values of the DNAs alone are: *Bcl-2 G4* = 59.1 (±0.9) °C; *c-Myc G4* = 78.6 (±0.4) °C; *Tel_{23} G4* = 55 (±0.7) °C; *Bcl-2 iM* = 62.8 (±0.9) °C; *c-Myc iM* = 73.6 (±0.8) °C; *hTeloC iM* = 51.5 (±0.2) °C; *Hairpin* = 71.1 (±0.9) °C. $\Delta T_{1/2}$ values correspond to the difference between the DNA melting temperatures with and without compounds.

3.5. Principal Component Analysis

The $\Delta T_{1/2}$ values obtained from the CD experiments were collected in a data matrix consisting of 31 rows (compounds) and 7 columns (DNA sequences) as reported in Table 1. Before PCA, data was converted to absolute values, then mean-centered and scaled to unit variance (autoscaling). Autoscaling employs the standard deviation as a scaling factor thus giving all the $\Delta T_{1/2}$ values the same chance to affect the model. Moreover, data relative to the *Hairpin* and *Tel_{23} G4* sequences were removed before the analysis because of the very

small variation of the $\Delta T_{1/2}$ values (<4 °C). PCA was then computed by using PLS Toolbox 8.6.1 (Eigenvector Research Inc., Wenatchee, WA, USA) in Matlab R2015b (The Mathworks Inc., Natick, MA, USA) environment. The pK_a values for each compound were calculated by using the Percepta Software version 14.3.0 (ACD/Labs, Toronto, ON, Canada).

3.6. Nuclear Magnetic Resonance Experiments

3.6.1. 1D ^{1}H-NMR Experiments

The samples for nuclear magnetic resonance (NMR) spectroscopy were prepared at 100–150 μM final oligonucleotide concentration, in 250 μL (H_2O/D_2O 9:1) buffer solution (10 mM sodium phosphate buffer (pH 5.0) for *hTeloC iM*, and 10 mM potassium phosphate buffer (pH 5.0) for *c-Myc G4*). DNA/compound mixtures were obtained by adding 2 equivalents of the compounds directly to the DNA solution inside the NMR tube [59], with a final DMSO concentration of 6%. NMR experiments were performed by employing a 700 MHz Bruker spectrometer. In particular, 1D ^{1}H-NMR spectra were recorded at 10 °C with a 'zgesgp' pulse-program (a gradient-based excitation sculpting using 180° water-selective pulses), including the following parameters: 128 scans, spectral width 17,241 Hz, delay 3 s, receiver gain 101, and 32 k points. All NMR spectra were calibrated by centering the water signal at 4.69 ppm (at 10 °C and pH 5.0). NMR data processing was done by using the vendor software TOPSPIN 4.0.7 (Bruker Biospin Gmbh, Rheinstetten, Germany).

3.6.2. STD NMR Experiments

In the 1D ^{1}H ligand-based STD NMR experiments, each compound (250 μM), previously solubilized in DMSO-d_6, was added to a DNA solution (20 μM, 600 μL) in 100% of deuterium oxide containing 10 mM phosphate buffer, pH 5.0. All the spectra were acquired at 298 K with a Bruker AVANCE NEO NMR spectrometer (Bruker Biospin Gmbh, Rheinstetten, Germany) operating at 700 MHz (^{1}H Larmor frequency), equipped with a 5 mm TCI 3 channels HCN cryo-probe head, optimized for ^{1}H sensitivity. The spectrometer was also equipped with a SampleCase (autosampler) for NMR screening. The spectra were processed with the vendor software TOPSPIN 4.0.7 (Bruker Biospin Gmbh, Rheinstetten, Germany). Saturation Transfer Difference spectra were acquired with 1024 scans, with on-resonance irradiation at 5.95 and 9.12 ppm for selective saturation of *c-Myc G4* and *hTeloC iM* resonances, respectively, and off-resonance irradiation at 40 ppm for reference spectra. A train of 40 Gaussian-shaped pulses of 50 ms (with 1 ms delay between pulses) was used, for a total saturation time of 4 s. The saturation width of the used radiofrequency pulses was 100 Hz. The STD spectra were obtained by internal subtraction of the saturated spectrum from the reference spectrum by phase cycling, with a spectral width of 20 ppm, relaxation delay 1.0 s, 16 k data points for acquisition, and 65 k for transformation.

3.7. Fluorescence Titrations

Fluorescence titration experiments were performed at 25 °C on an FP-8300 spectrofluorometer (Jasco, Easton, MD, USA) equipped with a Peltier temperature controller system (Jasco PCT-818, Jasco, Easton, MD, USA). A 1 cm path length, sealed quartz cuvette was used. Titrations were carried out by stepwise addition (5 μL) of *c-Myc G4*, *c-Myc iM*, *hTeloC iM*, or a single-stranded DNA [$d(CT)_{15}$] solution (150–200 μM) to a cell containing a fixed concentration of a compound solution (2.5–3.5 μM) in the appropriate buffer (10 mM potassium phosphate buffer (pH 5.0) for *c-Myc G4*, and 10 mM sodium phosphate buffer (pH 5.0) for *c-Myc iM*, *hTeloC iM*, and single-stranded DNA). The excitation wavelength was set at 275 nm, and emission spectra were recorded in the wavelength range of 285–600 nm. Both excitation and emission slit widths were set at 5 nm. After each DNA addition, the solution was stirred and allowed to equilibrate for 5 min before spectrum acquisition. The fraction of bound compound (α) at each point of the titration was calculated following the changes of fluorescence intensity at the maximum of intensity (307 nm). The compound concentration was corrected for dilution effects resulting from the change in volume due to DNA solution addition. It should be pointed out that the excitation of compounds at

275 nm led to a certain filter effect due to DNA absorption at this wavelength, which in part contributes to the observed quenching of compound fluorescence independent of binding. Titration curves were obtained by plotting α versus the DNA concentration. The equilibrium dissociation constant (K_b) and the stoichiometry of interaction were estimated by fitting the resulting curve to an independent and equivalent binding site model as previously described [62]. The experiments were repeated in duplicate, and the results are presented as the mean ± S.D.

3.8. Biological Experiments

3.8.1. Cell Culture and Treatments

Human osteosarcoma cells (U2OS) were purchased from American Type Culture Collection (ATCC) and were grown in Dulbecco's Modified Eagle Medium (DMEM, Euroclone, Milan, Italy) supplemented with 10% Fetal Bovine Serum (FBS), 2 mM L-glutamine, and antibiotics, at 37 °C, in a 5% CO_2-95% air atmosphere. As a positive control for the stabilization of iM structures, U2OS were cultured in the presence of 8% of CO_2 for 2.5 h, as reported elsewhere [16]. As for G-quadruplex structures stabilization, treatment with 1 μM RHPS4 for 24 h was used as a positive control [63]. All compounds were dissolved at a concentration of 10 mM in DMSO. Cells were incubated for 24 h with the compounds, at a final concentration of 2 μM.

3.8.2. Immunofluorescence (IF) Microscopy Experiments

For the IF experiments, cells were fixed in 4% formaldehyde in phosphate-buffered saline (PBS), at RT, for 10 min, permeabilized in 0.25% Triton X-100 in PBS 1×, at RT, for 5 min, and incubated with blocking solution (3% FBS in PBS 1×) at RT for 1 h. For immuno-labeling, cells were incubated at RT, for 2 h, with the flag-tagged antibody that recognizes iM structures. Successively, cells were washed three times with PBS 1× and incubated with the antibody anti-Flag (Sigma-Aldrich, St. Louis, MO, USA, #F7425) for 1 h. Finally, after three washes in PBS 1×, cells were incubated with the antibody anti-rabbit IgG (H + L), F(ab′)2 Fragment (Alexa Fluor 555 Conjugate) (Cell Signaling, Danvers, MA, USA, #4413S) for 1 h. For G4 structures staining, after the permeabilization, cells were incubated with the specific antibody (Anti-DNA/RNA G-quadruplex (BG4), Absolute Antibody, Oxford, UK, #Ab00174-1.1) at RT, for 2 h, washed three times with PBS 1×, and incubated with the antibody anti-mouse IgG (H + L), F(ab')2 Fragment (Alexa Fluor 555 Conjugate) (Cell Signaling Technology, Danvers, MA, USA, #4409S) for 1 h. For all the IF experiments, nuclei were stained with 4′,6-diamidino-2-phenylindole (DAPI, Sigma-Aldrich, St. Louis, MO, USA, #D9542), and fluorescence signals were recorded by using a Zeiss LSM 880 with airyscan (Zeiss, Jena, Germany), 63× magnification. IF experiments were quantified by Image J (version 1.53e, National Institutes of Health, NIH, Bethesda, MD, USA). 25 cells were screened for each condition and the results were expressed as fold change of fluorescence intensity (anti-G4 or anti-iM signal) over the negative control (DMSO). Histograms show the mean ± SD of three independent experiments. Statistical significance was calculated using unpaired student *t*-tests on Prism 6 (GraphPad, San Diego, CA, USA).

4. Conclusions

In the present study, we rationally designed a new library of potential G4/iM-targeting compounds according to the main structural features required to design effective noncanonical DNA-interacting compounds. To this aim, we took advantage of a novel synergistic approach by combining the Hantzsch multicomponent reaction and the CuAAC click chemistry reaction to generate a library of 31 bis-triazolyl-pyridine derivatives. Then, using multiple biophysical techniques, we dissected the profile of each compound in terms of effects on G4/iM structures. Of note, these experiments were corroborated by the application of the PCA method to select the most promising compounds. Finally, in order to translate the biophysical data into a more complex environment, we probed the capability of the selected compounds to affect the formation of G4/iM structures in U2OS cells.

Interestingly, our biophysical screening led to the identification of compounds **15c** and **18c** as candidates capable of concomitantly stabilizing the G4 structures and destabilizing the iM ones, and compounds **20a** and **23a** that selectively affected the thermal stability of the iMs, showing negligible effects on G4 structures. NMR and fluorescence titration experiments provided insights into the interaction mode and affinity of the selected compounds towards the noncanonical DNA structures under investigation. Noteworthy, the results of the biophysical study found confirmation also at the biological level. Indeed, analyses performed by confocal microscopy showed that the same compounds were able to destabilize iM structures in U2OS cells, the best-characterized cell model used so far to study iMs formation. Intriguingly, **15c** and **18c** were able to effectively destabilize iM structures by simultaneously stabilizing G4s, so resulting—also in the complexity of the cellular environment—the most effective compounds among those synthesized and tested.

How G4 and iM structures reciprocally influence each other, and which biological implications are regulated by such a delicate balance are still open questions under intense investigation. Thus, the need for new experimental approaches able to disentangle this scientific dilemma is urgent. In this scenario, our investigation led to the identification of new compounds that might be used as tools to shed light on the mechanisms underlying the controversial biological roles of G4 and iM structures and their intricated relationship.

Supplementary Materials: The Supplementary Materials can be found at https://www.mdpi.com/article/10.3390/ijms222111959/s1.

Author Contributions: Conceptualization, J.A., B.P., G.C.T. and A.R.; methodology, U.G., J.A., P.Z., N.I., D.B., A.B., B.P., G.C.T. and A.R.; formal analysis, A.D.P., U.G., P.Z., S.I., N.I., S.M., F.S. and D.B.; investigation, A.D.P., U.G., P.Z., S.I., S.M., F.S. and D.B.; resources, A.B., B.P., G.C.T. and A.R.; data curation, A.D.P., U.G., J.A., P.Z., N.I. and S.D.T.; writing—original draft preparation, A.D.P., U.G., J.A., P.Z., N.I., D.B., B.P., G.C.T. and A.R.; writing—review and editing, A.D.P., J.A., N.I., S.D.T., A.B., B.P., G.C.T. and A.R.; supervision, J.A., A.C., A.B., B.P., G.C.T. and A.R.; funding acquisition, B.P. and A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Italian Association for Cancer Research (AIRC, IG 21579, IG 24590, and IG 18695 to A.B., B.P. and A.R., respectively).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The antibody used to detect i-motif structures was kindly gifted by Mahdi Zeraati, Garvan Institute of Medical Research, Darlinghurst, Sydney, Australia.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

APT	Attached proton test
ATCC	American type culture collection
BCL-2	B-cell lymphoma 2
CD	Circular dichroism
c-MYC	Cellular MYC
CuAAC	Copper(I)-catalyzed alkyne-azide cycloaddition
DAPI	4′,6-diamidino-2-phenylindole
DBU	1,8-diazabicycloundec-7-ene
DMEM	Dulbecco's modified eagle medium
DMF	*N*,*N*′-dimethylformamide
DMSO	Dimethyl sulfoxide
DPPA	Diphenylphosphoryl azide
ESI	Electrospray ionization
FBS	Fetal bovine serum

FT-IR	Fourier transform infrared
G4	G-quadruplex
HPLC	High-performance liquid chromatography
IF	Immunofluorescence
iM	i-Motif
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
PBS	Phosphate buffered saline
PCA	Principal component analysis
PE	Petroleum ether
RT	Room temperature
STD	Saturation transfer difference
THF	Tetrahydrofuran

References

1. Bacolla, A.; Wells, R.D. Non-B DNA Conformations as Determinants of Mutagenesis and Human Disease. *Mol. Carcinog.* **2009**, *48*, 273–285. [CrossRef] [PubMed]
2. Abou Assi, H.; Garavís, M.; González, C.; Damha, M.J. I-Motif DNA: Structural Features and Significance to Cell Biology. *Nucleic Acids Res.* **2018**, *46*, 8038–8056. [CrossRef]
3. Gajarský, M.; Živković, M.L.; Stadlbauer, P.; Pagano, B.; Fiala, R.; Amato, J.; Tomáška, L.; Šponer, J.; Plavec, J.; Trantírek, L. Structure of a Stable G-Hairpin. *J. Am. Chem. Soc.* **2017**, *139*, 3591–3594. [CrossRef] [PubMed]
4. Spiegel, J.; Adhikari, S.; Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* **2020**, *2*, 123–136. [CrossRef]
5. Lane, A.N.; Chaires, J.B.; Gray, R.D.; Trent, J.O. Stability and Kinetics of G-Quadruplex Structures. *Nucleic Acids Res.* **2008**, *36*, 5482–5515. [CrossRef] [PubMed]
6. Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **2016**, *4*, 38. [CrossRef]
7. 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. Metal Ions in Life Sciences*; Sigel, A., Sigel, H., Sigel, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2016; pp. 203–258.
8. 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]
9. Guédin, A.; Gros, J.; Alberti, P.; Mergny, J.-L. How Long Is Too Long? Effects of Loop Size on G-Quadruplex Stability. *Nucleic Acids Res.* **2010**, *38*, 7858–7868. [CrossRef]
10. 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]
11. Marsico, G.; Chambers, V.S.; Sahakyan, A.B.; McCauley, P.; Boutell, J.M.; Di Antonio, M.; Balasubramanian, S. Whole Genome Experimental Maps of DNA G-Quadruplexes in Multiple Species. *Nucleic Acids Res.* **2019**, *47*, 3862–3874. [CrossRef]
12. 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]
13. Kim, N. The Interplay between G-Quadruplex and Transcription. *Curr. Med. Chem.* **2019**, *26*, 2898–2917. [CrossRef] [PubMed]
14. Brown, S.L.; Kendrick, S. The I-Motif as a Molecular Target: More Than a Complementary DNA Secondary Structure. *Pharmaceuticals* **2021**, *14*, 96. [CrossRef] [PubMed]
15. Dzatko, S.; Krafcikova, M.; Hänsel-Hertsch, R.; Fessl, T.; Fiala, R.; Loja, T.; Krafcik, D.; Mergny, J.-L.; Foldynova-Trantirkova, S.; Trantirek, L. Evaluation of the Stability of DNA I-Motifs in the Nuclei of Living Mammalian Cells. *Angew. Chem. Int. Ed.* **2018**, *57*, 2165–2169. [CrossRef]
16. Zeraati, M.; Langley, D.B.; Schofield, P.; Moye, A.L.; Rouet, R.; Hughes, W.E.; Bryan, T.M.; Dinger, M.E.; Christ, D. I-Motif DNA Structures Are Formed in the Nuclei of Human Cells. *Nat. Chem.* **2018**, *10*, 631–637. [CrossRef] [PubMed]
17. 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]
18. Kosiol, N.; Juranek, S.; Brossart, P.; Heine, A.; Paeschke, K. G-Quadruplexes: A Promising Target for Cancer Therapy. *Mol. Cancer* **2021**, *20*, 40. [CrossRef] [PubMed]
19. Amato, J.; Iaccarino, N.; Randazzo, A.; Novellino, E.; Pagano, B. Noncanonical DNA Secondary Structures as Drug Targets: The Prospect of the i-Motif. *ChemMedChem* **2014**, *9*, 2026–2030. [CrossRef] [PubMed]
20. Verma, A.; Yadav, V.K.; Basundra, R.; Kumar, A.; Chowdhury, S. Evidence of Genome-Wide G4 DNA-Mediated Gene Expression in Human Cancer Cells. *Nucleic Acids Res.* **2009**, *37*, 4194–4204. [CrossRef]

21. Shu, B.; Cao, J.; Kuang, G.; Qiu, J.; Zhang, M.; Zhang, Y.; Wang, M.; Li, X.; Kang, S.; Ou, T.-M.; et al. Syntheses and Evaluation of New Acridone Derivatives for Selective Binding of Oncogene C- Myc Promoter i-Motifs in Gene Transcriptional Regulation. *Chem. Commun.* **2018**, *54*, 2036–2039. [CrossRef]
22. Brown, R.V.; Wang, T.; Chappeta, V.R.; Wu, G.; Onel, B.; Chawla, R.; Quijada, H.; Camp, S.M.; Chiang, E.T.; Lassiter, Q.R.; et al. The Consequences of Overlapping G-Quadruplexes and i-Motifs in the Platelet-Derived Growth Factor Receptor β Core Promoter Nuclease Hypersensitive Element Can Explain the Unexpected Effects of Mutations and Provide Opportunities for Selective Targeting Of. *J. Am. Chem. Soc.* **2017**, *139*, 7456–7475. [CrossRef]
23. Satpathi, S.; Sappati, S.; Das, K.; Hazra, P. Structural Characteristics Requisite for the Ligand-Based Selective Detection of i-Motif DNA. *Org. Biomol. Chem.* **2019**, *17*, 5392–5399. [CrossRef]
24. Masoud, S.S.; Nagasawa, K. I-Motif-Binding Ligands and Their Effects on the Structure and Biological Functions of I-Motif. *Chem. Pharm. Bull.* **2018**, *66*, 1091–1103. [CrossRef]
25. Rostovtsev, V.V.; Green, L.G.; Fokin, V.V.; Sharpless, K.B. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596–2599. [CrossRef]
26. Saha, P.; Panda, D.; Dash, J. The Application of Click Chemistry for Targeting Quadruplex Nucleic Acids. *Chem. Commun.* **2019**, *55*, 731–750. [CrossRef]
27. Tron, G.C.; Pirali, T.; Billington, R.A.; Canonico, P.L.; Sorba, G.; Genazzani, A.A. Click Chemistry Reactions in Medicinal Chemistry: Applications of the 1,3-Dipolar Cycloaddition between Azides and Alkynes. *Med. Res. Rev.* **2008**, *28*, 278–308. [CrossRef] [PubMed]
28. Bonandi, E.; Christodoulou, M.S.; Fumagalli, G.; Perdicchia, D.; Rastelli, G.; Passarella, D. The 1,2,3-Triazole Ring as a Bioisostere in Medicinal Chemistry. *Drug Discov. Today* **2017**, *22*, 1572–1581. [CrossRef] [PubMed]
29. Moorhouse, A.D.; Santos, A.M.; Gunaratnam, M.; Moore, M.; Neidle, S.; Moses, J.E. Stabilization of G-Quadruplex DNA by Highly Selective Ligands via Click Chemistry. *J. Am. Chem. Soc.* **2006**, *128*, 15972–15973. [CrossRef]
30. Ritson, D.J.; Moses, J.E. A Fragment Based Click Chemistry Approach towards Hybrid G-Quadruplex Ligands: Design, Synthesis and Biophysical Evaluation. *Tetrahedron* **2012**, *68*, 197–203. [CrossRef]
31. Sparapani, S.; Haider, S.M.; Doria, F.; Gunaratnam, M.; Neidle, S. Rational Design of Acridine-Based Ligands with Selectivity for Human Telomeric Quadruplexes. *J. Am. Chem. Soc.* **2010**, *132*, 12263–12272. [CrossRef]
32. Drewe, W.C.; Neidle, S. Click Chemistry Assembly of G-Quadruplex Ligands Incorporating a Diarylurea Scaffold and Triazole Linkers. *Chem. Commun.* **2008**, 5295. [CrossRef]
33. Rodriguez, R.; Müller, S.; Yeoman, J.A.; Trentesaux, C.; Riou, J.-F.; Balasubramanian, S. A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-Damage Response at Telomeres. *J. Am. Chem. Soc.* **2008**, *130*, 15758–15759. [CrossRef]
34. Müller, S.; Sanders, D.A.; Di Antonio, M.; Matsis, S.; Riou, J.-F.; Rodriguez, R.; Balasubramanian, S. Pyridostatin Analogues Promote Telomere Dysfunction and Long-Term Growth Inhibition in Human Cancer Cells. *Org. Biomol. Chem.* **2012**, *10*, 6537–6546. [CrossRef]
35. Di Fonzo, S.; Amato, J.; D'Aria, F.; Caterino, M.; D'Amico, F.; Gessini, A.; Brady, J.W.; Cesàro, A.; Pagano, B.; Giancola, C. Ligand Binding to G-Quadruplex DNA: New Insights from Ultraviolet Resonance Raman Spectroscopy. *Phys. Chem. Chem. Phys.* **2020**, *22*, 8128–8140. [CrossRef]
36. Pagano, A.; Iaccarino, N.; Abdelhamid, M.A.S.; Brancaccio, D.; Garzarella, E.U.; Di Porzio, A.; Novellino, E.; Waller, Z.A.E.; Pagano, B.; Amato, J.; et al. Common G-Quadruplex Binding Agents Found to Interact With i-Motif-Forming DNA: Unexpected Multi-Target-Directed Compounds. *Front. Chem.* **2018**, *6*, 281. [CrossRef]
37. Hantzsch, A. Condensationsprodukte Aus Aldehydammoniak Und Ketonartigen Verbindungen. *Ber. Dtsch. Chem. Ges.* **1881**, *14*, 1637–1638. [CrossRef]
38. Hantzsch, A. Ueber Die Synthese Pyridinartiger Verbindungen Aus Acetessigäther Und Aldehydammoniak. *Justus Liebig's Ann. Chem.* **1882**, *215*, 1–82. [CrossRef]
39. Tamaddon, F.; Razmi, Z.; Jafari, A.A. Synthesis of 3,4-Dihydropyrimidin-2(1H)-Ones and 1,4-Dihydropyridines Using Ammonium Carbonate in Water. *Tetrahedron Lett.* **2010**, *51*, 1187–1189. [CrossRef]
40. Pagano, B.; Cosconati, S.; Gabelica, V.; Petraccone, L.; De Tito, S.; Marinelli, L.; La Pietra, V.; Saverio di Leva, F.; Lauri, I.; Trotta, R.; et al. State-of-the-Art Methodologies for the Discovery and Characterization of DNA G-Quadruplex Binders. *Curr. Pharm. Des.* **2012**, *18*, 1880–1899. [CrossRef] [PubMed]
41. Dai, J.; Chen, D.; Jones, R.A.; Hurley, L.H.; Yang, D. NMR Solution Structure of the Major G-Quadruplex Structure Formed in the Human BCL2 Promoter Region. *Nucleic Acids Res.* **2006**, *34*, 5133–5144. [CrossRef] [PubMed]
42. Kendrick, S.; Akiyama, Y.; Hecht, S.M.; Hurley, L.H. The I-Motif in the Bcl-2 P1 Promoter Forms an Unexpectedly Stable Structure with a Unique 8:5:7 Loop Folding Pattern. *J. Am. Chem. Soc.* **2009**, *131*, 17667–17676. [CrossRef] [PubMed]
43. 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]
44. Dai, J.; Hatzakis, E.; Hurley, L.H.; Yang, D. I-Motif Structures Formed in the Human c-MYC Promoter Are Highly Dynamic–Insights into Sequence Redundancy and I-Motif Stability. *PLoS ONE* **2010**, *5*, e11647. [CrossRef] [PubMed]
45. Dai, J.; Carver, M.; Yang, D. Polymorphism of Human Telomeric Quadruplex Structures. *Biochimie* **2008**, *90*, 1172–1183. [CrossRef]

46. 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]
47. Phan, A.T.; Guéron, M.; Leroy, J.-L. 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. [CrossRef]
48. Iaccarino, N.; Cheng, M.; Qiu, D.; Pagano, B.; Amato, J.; Di Porzio, A.; Zhou, J.; Randazzo, A.; Mergny, J.-L. Effects of Sequence and Base Composition on the CD and TDS Profiles of I-DNA. *Angew. Chem. Int. Ed.* **2021**, *60*, 10295–10303. [CrossRef]
49. Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]
50. Amato, J.; Morigi, R.; Pagano, B.; Pagano, A.; Ohnmacht, S.; De Magis, A.; Tiang, Y.-P.; Capranico, G.; Locatelli, A.; Graziadio, A.; et al. Toward the Development of Specific G-Quadruplex Binders: Synthesis, Biophysical, and Biological Studies of New Hydrazone Derivatives. *J. Med. Chem.* **2016**, *59*, 5706–5720. [CrossRef] [PubMed]
51. 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. [CrossRef] [PubMed]
52. Kaiser, C.E.; Van Ert, N.A.; Agrawal, P.; Chawla, R.; Yang, D.; Hurley, L.H. Insight into the Complexity of the I-Motif and G-Quadruplex DNA Structures Formed in the KRAS Promoter and Subsequent Drug-Induced Gene Repression. *J. Am. Chem. Soc.* **2017**, *139*, 8522–8536. [CrossRef]
53. Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem. Int. Ed.* **1999**, *38*, 1784–1788. [CrossRef]
54. Giancola, C.; Pagano, B. Energetics of Ligand Binding to G-Quadruplexes. *Top. Curr. Chem.* **2013**, *330*, 211–242. [PubMed]
55. King, J.J.; Irving, K.L.; Evans, C.W.; Chikhale, R.V.; Becker, R.; Morris, C.J.; Peña Martinez, C.D.; Schofield, P.; Christ, D.; Hurley, L.H.; et al. DNA G-Quadruplex and i-Motif Structure Formation Is Interdependent in Human Cells. *J. Am. Chem. Soc.* **2020**, *142*, 20600–20604. [CrossRef] [PubMed]
56. Böcker, R.H.; Guengerich, F.P. Oxidation of 4-Aryl- and 4-Alkyl-Substituted 2,6-Dimethyl-3,5-Bis(Alkoxycarbonyl)-1,4-Dihydropyridines by Human Liver Microsomes and Immunochemical Evidence for the Involvement of a Form of Cytochrome P-450. *J. Med. Chem.* **1986**, *29*, 1596–1603. [CrossRef]
57. Loev, B.; Snader, K.M. The Hantzsch Reaction. I. Oxidative Dealkylation of Certain Dihydropyridines. *J. Org. Chem.* **1965**, *30*, 1914–1916. [CrossRef]
58. Shahabi, D.; Amrollahi, M.A. Synthesis of Some Novel and Water-Soluble 2,4,6-Substituted 3,5-Dihydroxymethylpyridines. *Bulg. Chem. Commun.* **2014**, *46*, 264–268.
59. Amato, J.; Pagano, A.; Cosconati, S.; Amendola, G.; Fotticchia, I.; Iaccarino, N.; Marinello, J.; De Magis, A.; Capranico, G.; Novellino, E.; et al. Discovery of the First Dual G-Triplex/G-Quadruplex Stabilizing Compound: A New Opportunity in the Targeting of G-Rich DNA Structures? *Biochim. Biophys. Acta—Gen. Subj.* **2017**, *1861*, 1271–1280. [CrossRef] [PubMed]
60. 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]
61. Iaccarino, N.; Di Porzio, A.; Amato, J.; Pagano, B.; Brancaccio, D.; Novellino, E.; Leardi, R.; Randazzo, A. Assessing the Influence of pH and Cationic Strength on I-Motif DNA Structure. *Anal. Bioanal. Chem.* **2019**, *411*, 7473–7479. [CrossRef]
62. Musumeci, D.; Amato, J.; Zizza, P.; Platella, C.; Cosconati, S.; Cingolani, C.; Biroccio, A.; Novellino, E.; Randazzo, A.; Giancola, C.; et al. Tandem Application of Ligand-Based Virtual Screening and G4-OAS Assay to Identify Novel G-Quadruplex-Targeting Chemotypes. *Biochim. Biophys. Acta—Gen. Subj.* **2017**, *1861*, 1341–1352. [CrossRef] [PubMed]
63. Yang, S.Y.; Chang, E.Y.C.; Lim, J.; Kwan, H.H.; Monchaud, D.; Yip, S.; Stirling, P.C.; Wong, J.M.Y. G-Quadruplexes Mark Alternative Lengthening of Telomeres. *NAR Cancer* **2021**, *3*, zcab031. [CrossRef] [PubMed]