

**Design, Synthesis, and Evaluation of Novel c-MYC G-
Quadruplex Stabilizing Ligands**

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Biochemie, Chemie und Pharmazie
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von

Vijaykumar D Nimbarte

aus Nagpur, Indien

Frankfurt am Main

2019

(D 30)

Vom Fachbereich Biochemie, Chemie und Pharmazie der Johann Wolfgang Goethe
Universität als Dissertation angenommen.

Dekan: Prof. Dr. Clemens Glaubitz

Gutachter: Prof. Dr. Harald Schwalbe

Prof. Dr. Alexander Heckel

Datum der Disputation: 18/06/2020

Dedicated to my beloved Parents...

वक्रतुण्ड महाकाय सूर्यकोटि समप्रभ ।
निर्विघ्नं कुरु मे देव सर्वकार्येषु सर्वदा ॥

This thesis was prepared under the supervision of Prof. Dr. Harald Schwalbe between March 2016 and Dec 2019 at the Institute for Organic Chemistry and Chemical Biology of the Goethe-University Frankfurt am Main.

TABLE OF CONTENT

Summary	9
Zusammenfassung	14
List of Abbreviation	21

Chapter 1

Introduction

1.1. Cancer -General introduction	24
1.2. DNA structures and DNA components	38
1.3. Biological functions of G4	45
1.4. G4 ligands	61
1.5. Summary and Thesis Aims	68
1.6. References	71

Chapter 2

A Fragment-Based Approach for the Development of G-Quadruplex Ligands

2.1. The Drug Discovery Process	88
2.2. Hit to lead optimization strategy	94
2.3. Project Objective	104
2.4. Results and Discussion	104
2.5. Initial Screening of Low-Molecular Weight Fragments	104
2.6. Design and Synthesis	108
2.7. SMILES (Simplified Molecular Input Line Entry System)	112
2.8. SWISS ADME: Profiling of the optimized and selected fragments	124
2.9. Fragments with Fluorescent indicator displacement (FID) assays	134
2.10. Molecular Docking studies	137
2.11. SWISS-ADME analysis of best lead fragments	139
2.12. DC ₅₀ and Ki values estimation for the top 5 fragments: FID assays	140
2.13. ¹ D-NMR: Ligand binding observed by NMR-spectroscopy	142
2.14. Cell Cytotoxicity and Cell cycle analysis	144

2.15. Western Blot analysis	144
2.16. Cell cycle analysis	145
2.17. Conclusion	146
2.18. Experimental Section	147
2.19. References	160

Chapter 3

Novel triazole linked indole derivatives as c-MYC G-quadruplex stabilizers *via* Click chemistry

3.1. Introduction	171
3.2. Bioisosterism	173
3.3. Project Objective	176
3.4. Results and Discussion	176
3.4.1 Design and Synthesis	176
3.5 Initial screening with fluorescent indicator displacement (FID) assays	178
3.6. Molecular Modelling	182
3.7. ¹ D-NMR: Ligand binding observed by NMR-spectroscopy	184
3.8. Cell Cytotoxicity and Cell cycle analysis	186
3.9. Western Blot analysis	186
3.10. Cell cycle analysis	187
3.11. Conclusion	188
3.12. Experimental Section	189
3.13. References	198

Chapter 4

Novel Azaindolin-2-one derived c-MYC Promoter G-Quadruplex Stabilizers

4.1. Introduction	204
4.2. Background and applications of azaindole	206
4.3. Project Objective	210
4.4. Results and Discussion	210
4.5. Design and Synthesis	210
4.6. Reaction mechanism	213

4.7. Knoevenagel condensation: <i>E</i> and <i>Z</i> isomer	217
4.8. Intramolecular hydrogen bond in <i>Z</i> -isomer	219
4.9. Initial screening with fluorescent indicator displacement (FID) assays	219
4.10. ¹ D-NMR: Ligand binding observed by NMR-spectroscopy	223
4.11. Cell Cytotoxicity	224
4.12. Western Blot analysis	225
4.13. Cell cycle analysis	225
4.14. Conclusion	226
4.15. Experimental Section	227
4.16. References	240
Acknowledgments	246
Publications	247
Conference contributions	249
Curriculum vitae	250

Summary

The work of this thesis focuses on the targeting of G-quadruplexes (G4s), wherein several specific and potential ligands were designed, synthesized and characterized for its structural and biological activity. G4s are nucleic acid secondary structures that may form in single-stranded guanine (G)-rich sequences under physiological conditions. Four Guanines (Gs) bind *via* Hoogsteen-type hydrogen bonds base pairing to yield G-quartets, which in turn stack on top of each other to form the G4. G4s are highly polymorphic, both in terms of strand stoichiometry (forming both inter and intramolecular structures) and strand orientation/topology. The presence of K⁺ cations specifically supports G4 formation and stability. In the human genome G4 DNA motifs have been found in telomeres, G-rich micro and mini-satellites, up-stream to oncogene promoters and within the ribosomal DNA (rDNA). Human G4 DNA motifs are over-expressed in recombinogenic regions, which are associated with genomic damage in cancer cells.

In the present work, we focus on lead identification with specificity towards the *c-MYC* promoter G4s. Drug discovery is a highly time consuming and costly process. Lead identification and development are key steps in the drug discovery program. Studies have suggested that a large number of commercially available drugs exhibit deep structural similarity to the lead compounds from which they were developed. Quality lead identification in terms of compounds with high potency and selectivity, favorable physicochemical parameters and *in vitro* Absorption Distribution Metabolism and Excretion (ADME) parameters are the foremost requirements for the success of the drug discovery process. We herein describe the fragment-based drug design approach for the development of pyrrolidine-substituted 5-nitroindole derivatives as a new class of G4 ligands that exhibit high affinity and selectivity for the *c-MYC* promoter G-quadruplex. This chapter focuses on the methodology explored whilst finding a suitable hit and its optimization with fragment expansion strategies, which undergo efficient G4 binding as, depicted in figure 1.

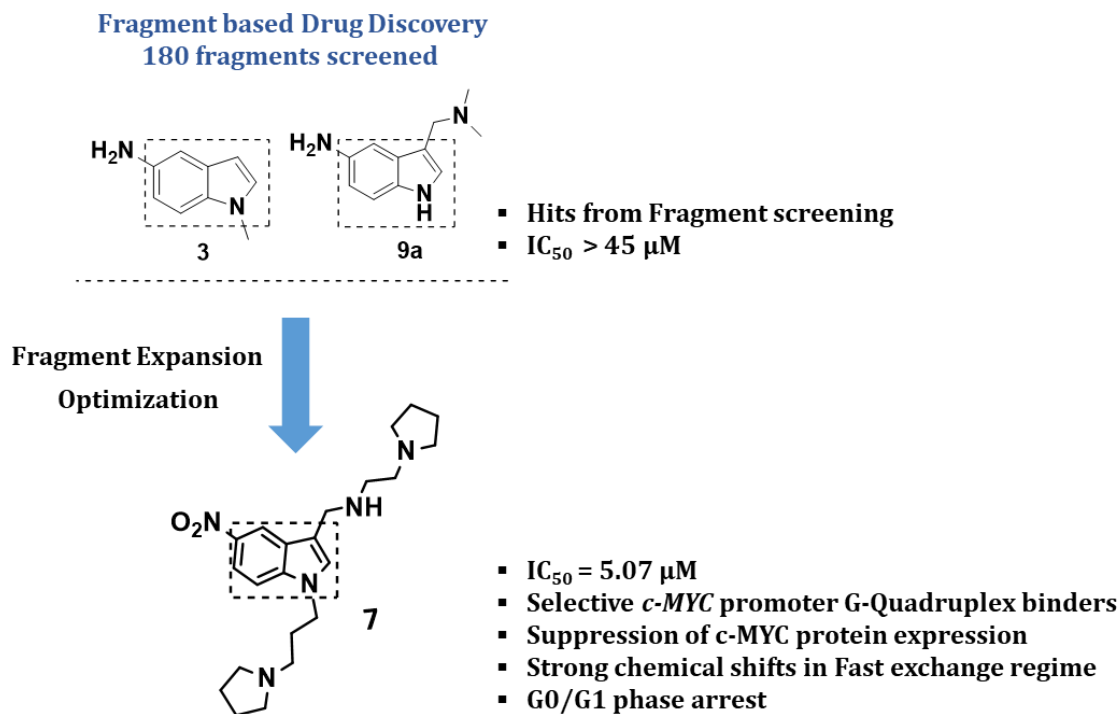


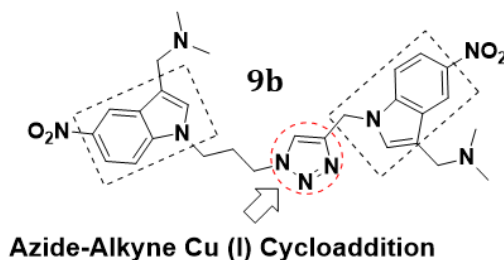
Figure 1. Fragment based drug design approach to effective *c-MYC* promoter G4 binders

To target G4 DNA, screenings of numerous heterocycles have been reported including indoles, 7-azaindoles, 1H-indazol-3-yl, benzothiazole, imidazo[1,5-a]pyridine, 2,6-diaminopyrimidin-4-ol, 1H-pyrazolo[4,3 d]pyrimidin-7-amine, morpholino, bis-indoles, 2-hydroxynaphthalene-1,4-dione, 1,4-dihydroxyanthracene-9,10-dione, benzofuran and piperonal derived from several alkaloids. In this part of the thesis, we set out to identify new binders targeting the *c-MYC* G-quadruplex starting from the indole fragment. Several synthetic strategies are reported to optimize and generate best hits starting from 5-nitro indole derivatives by introducing the secondary cationic linked pyrrolidine side chain. Interestingly, all improved versions of G4-indole fragments 5, 7 and 12 contain this 5-nitro functionality, which may aid in the electrostatic binding and contributes to hydrogen binding interactions of the ligands to G4 DNA. *In-silico* drug design, biological and biophysical analyses illustrated that the substituted 5-nitro indoles scaffolds show preferential affinity towards the *c-MYC* promoter G-quadruplex compared to other G-quadruplexes and double stranded DNA. *In vitro* cellular studies confirm that the substituted indole scaffolds downregulate *c-MYC* expression in cancer cells and have the potential to induce

cell cycle arrest in the G₀/G₁ phase. NMR analysis suggests that 5, 7, and 12 interacts in a fast exchange regime with the terminal G-quartets (5' and 3' end) in a 2:1 stoichiometry.

To further optimize the fragment generated in chapter II, a novel series of triazole linked indole derivatives as a potential G quadruplex stabilizers have been described in **chapter III**. The potential ligands can be obtained through an efficient, convergent, synthetic route in moderate to good yields. The synthesized triazole linked indole derivatives are selective towards *c-MYC* G₄-DNA vs. duplex-DNA. The planarity of the aromatic core and its ability to occupy more surface area by stacking over the G₄ greatly affect the ability of the compounds to stabilize the G₄. Further biophysical and biological studies revealed that the triazole linked nitro indoles are more promising than the amino indole derivatives as depicted in figure 2.

Peptidomimetics Approach



- $IC_{50} = 3.655 \mu M$
- **Selective *c-MYC* promoter G-Quadruplex binders**
- **Suppression of *c-MYC* protein expression**
- **Strong chemical shifts in Fast exchange regime**
- **G₀/G₁ phase arrest**

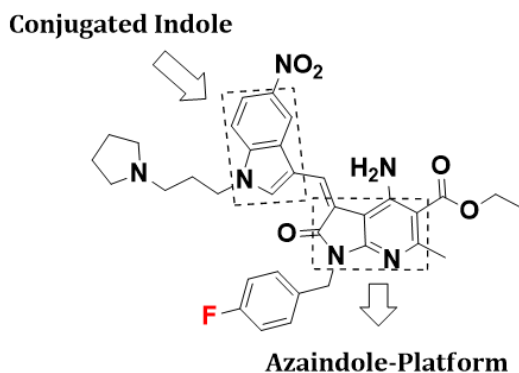
Figure 2. Click chemistry derived G₄ ligand

Additionally, the importance of the nitro functional group has been justified by molecular docking studies, where hydrogen-bonding interactions were observed in between the nitro group and the G₄ base pairs of the G-quadruplex. In biological findings, most of the synthesized triazole linked nitro indoles has found to be effective against human carcinoma (cervical) HeLa cell lines. Furthermore, western blot and cell cycle analysis confirms that the novel triazole linked 5-nitro indole derivatives (9b) could down-regulate *c-MYC* oncogene

expression in cancer cells *via* stabilizing its promoter quadruplex structure, arresting cell cycle in G0/G1 phase. NMR analysis suggests that 9b interacts in slow exchange regime with the terminal G-quartets (5' and 3'-end).

In **chapter IV** of the thesis, we have developed the synthetic strategies to generate more potent G4 ligands *via* Knoevenagel condensation. To investigate novel and selective G4 ligands for cancer chemotherapy, we designed and synthesized a series of azaindolin-2-one derivatives (11, 14, 15, 16 and 22) by attaching cationic pyrrolidine side chains and introducing a fluorine atom into the aromatic chromophore (Fig. 3). Fluorine atoms, with high electronegativity and small size, often exhibit unique properties in functional molecules. The electron-withdrawing effect of fluorine could reduce the electron density of the aromatic chromophore, which might favor a stronger interaction with the electron-rich π -system of the G-quartet. In addition, the introduction of fluorine atoms into small molecules might improve lipophilicity and thus the bioavailability. Fluorescent indicator displacement assay (FID) assays suggests that the synthesized azaindolin-2-one derivatives are selective towards *c-MYC* G4-DNA vs. duplex-DNA and showed potent anticancer activity against human carcinoma (cervical) HeLa cell lines. They down-regulate *c-MYC* expression in cancer cells *via* stabilizing its promoter quadruplex structure, arresting cell cycle in G0/G1 phase. Furthermore, NMR spectroscopy suggests that azaindolin-2-one conjugate interacts with terminal G-quartets as well as with the nearby G-rich tract (G13-G14-G15 and G8-G9-G10) of *c-MYC* quadruplex in intermediate exchange regime.

Knoevenagel condensation Approach



- $IC_{50} = 2.258 \mu M$
- Selective *c-MYC* promoter G-Quadruplex binders
- Suppression of *c-MYC* protein expression
- Strong chemical shifts in intermediate exchange regime
- G0/G1 phase arrest

Figure 3. Knoevenagel condensation derived Azaindolin-2-one derivative

Our present results increased our understanding of the effect of *c-MYC* G4 on oncogene transcriptional regulation. The studies performed in this thesis reveal that the pyrrolidine-substituted 5-nitroindole derivatives (chapter II), triazole linked nitro-indole derivatives (chapter III) and azaindolin-2-one conjugates (chapter IV) could become a potential lead compound for further development for cancer treatment by *c-MYC* promoter G4. This information can guide the optimization of G4-ligands, for increased binding affinity and selectivity toward G-quadruplex structures.

Zusammenfassung

Die vorliegende Dissertation fokussiert sich auf die G-Quadruplexe (G4s), hierfür wurden verschiedene spezifische und potentielle Liganden aufgrund ihrer strukturellen und biologischen Aktivität entworfen, synthetisiert und charakterisiert. G4s sind Nukleinsäure Sekundärstrukturen, die sich unter physiologischen Bedingungen in einzelsträngigen Guanin (G) reichen Sequenzen formen. Vier Guanine binden *via* Wasserstoffbrückenbindung (Hoogsteen-Typ) und ergeben G-Quartette, die sich wiederum aufeinanderstapeln um G4 zu formen. G4s sowohl in Bezug auf die Strangstöchiometrie (welche sowohl inter- als auch intramolekular Strukturen bildet) und Strangorientierung/-topologie hoch polymorph. Die Anwesenheit von Kaliumkationen unterstützt die spezifische Bildung und Stabilität von G4. Im menschlichen Genom wurden G4 DNA-Domäne in Teleomeren, G-reichen Mikro- und Minisatelliten, hochregulierten Onkogen Promotoren und innerhalb der ribosomalen DNA (rDNA) gefunden. Menschliche G4 DNA Domäne sind in rekombinanten Regionen überexprimiert, die mit genomischen Schäden in Krebszellen assoziiert sind.

In der vorliegenden Arbeit konzentrieren wir uns auf die Target Identifizierung mit Spezifität gegenüber dem c-MYC-Promotor G4s. Die Wirkstoffentdeckung ist ein zeitaufwendiger und kostspieliger Prozess. Hierbei sind die Identifikation und die Entwicklung wichtige Schritte im Rahmen der Entdeckung neuer Medikamente. Studien zeigen auf, dass eine große Anzahl kommerziell erhältlicher Medikamente eine tiefe strukturelle Ähnlichkeit mit Ihrer Leitverbindungen aufweisen, aus denen sie entwickelt wurden. Die Identifizierung von Verbindungen mit hoher Wirksamkeit und Selektivität und vorteilhafte physikalisch chemische Parameter sowie in vitro Absorption, Distribution, Metabolismus und Elimination (ADME) sind die wichtigsten Voraussetzungen für den Erfolg der Wirkstoffentwicklung. Wir beschreiben hier den fragmentbasierten Wirkstoffdesignansatz für die Entwicklung von Pyrrolidin substituierten 5-Nitroindolderivaten als neue Klasse von G4-Liganden, die eine hohe Affinität und Selektivität für den c-MYC-Promotor G-Quadruplex aufweisen. Dieser Abschnitt konzentriert sich auf die Methodik, die untersucht wurde um einen geeigneten Treffer zu finden und die Optimierung mit Hilfe von Fragment-

Erweiterungsstrategien zur Erhaltung eines effizienten G4 bindendes Molekül (Abbildung 1).

Um G4 DNA als Target zu anvisieren, wurden zahlreiche Screenings an Heterozyklen bereits berichtet, darunter Indole, 7-Azaindole, 1H-Indazol-3-yl, Benzothiazol, Imidazol[1,5-a]pyridin, 2,6-Diaminopyrimidin-4-ol, 1H-Pyrazol[4,3d]pyrimidin-7-amin, Morpholin, Bisindole, 2-Hydroxynaphthalin-1,4-dion, 1,4-Dihydroxyanthracen-9,10-dion, Benzofuran und Piperonal abgeleitet von verschiedenen Alkaloiden.

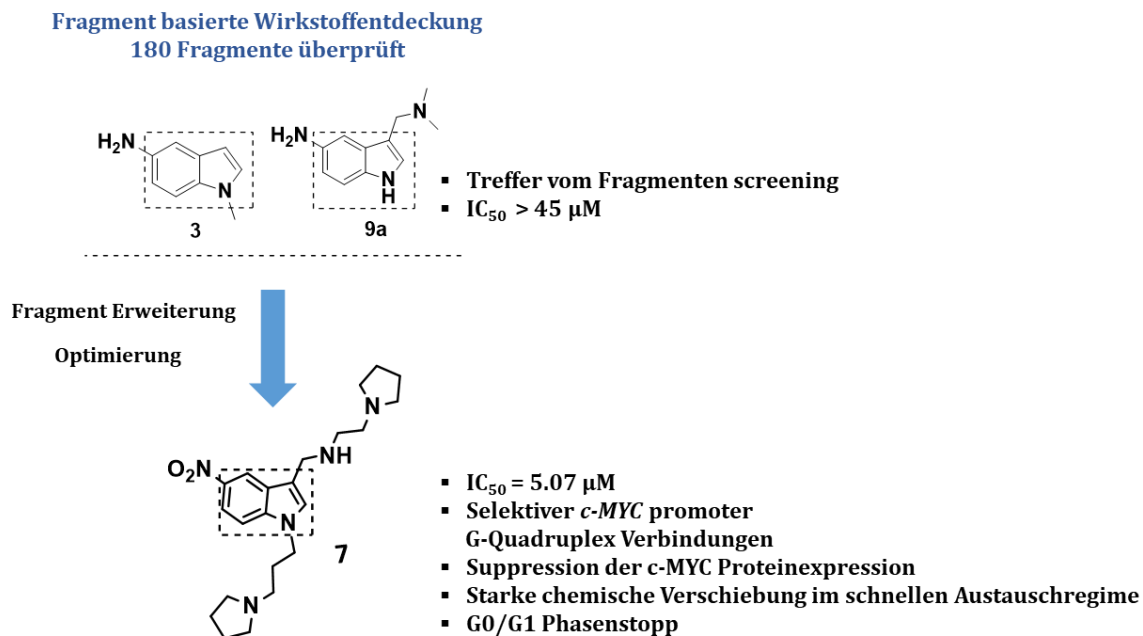


Abbildung 1. Fragment basierender Ansatz zum Wirkstoffdesign für wirksame *c-MYC*-Promotor-G4-Bindemittel.

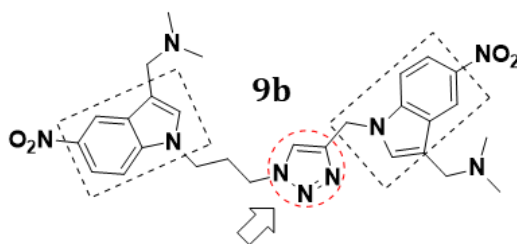
In diesem Abschnitt der Dissertation wollen wir neue potente *c-MYC* G-Quadruplex Moleküle identifizieren, ausgehend von der Leitstruktur dem Indolfragment. diverse Synthesestrategien gehen vom 5-Nitroindolderivat aus um die Einführung der sekundären kationisch verknüpften Pyrrolidinseitenkette zu ermöglichen. Interessanterweise, haben alle verbesserte Strukturen der G4 Indolfragmente 5, 7 und 12 eine 5-Nitrofunktionalität, die zum einen die elektrostatische Bindung unterstützt und zum anderen die Wasserstoffbrückenbindung der Liganden an die G4-DNA ermöglicht. *In silico*-Wirkstoffdesign sowie biologische und biophysikalische Analysen zeigten, dass das substituierte 5-Nitro-Indol Gerüst gegenüber anderen G-Quadruplexen und doppelsträngigen DNA eine bevorzugte Affinität zum *c-MYC*

Zusammenfassung

Promotor des G-Quadruplexen aufweist. *In vitro* Zellstudien bestätigen, dass die substituierten Indolgerüste die c-MYC Expression in Krebszellen runterreguliert und das Potential haben einen Zellzyklusstillstand in der G₀/G₁ Phase zu induzieren. Des Weiteren legt die NMR Analyse nahe, dass 5, 7 und 12 in einer 2:1 Stöchiometrie in einem schnellen Austauschregime mit den terminalen G-Quartetten (5' und 3' Ende) wechselwirken.

Zur Optimierung der erzeugten Verbindungen, welches im Kapitel II dargestellt worden sind, wurde in Kapitel III eine neue Reihe von Triazole verknüpften Indolderivaten als potentielle G-Quadruplex Stabilisatoren untersucht. Die potentiellen Liganden können durch einen effizienten konvergenten Syntheseweg in moderaten bis guten Ausbeuten erhalten werden. Die synthetisierten Triazole verknüpften Indolderivate sind selektiv gegenüber c-MYC G₄ DNA und Duplex DNA.

Peptidomimetischer ansatz



Azid-Alkin-Cu (I) -Cycloaddition

- $IC_{50} = 3.655 \mu M$
- **Selektiver c-MYC promoter G-Quadruplex Verbindungen**
- **Suppression der c-MYC Proteinexpression**
- **Starke chemische Verschiebung im schnellen Austauschregime**
- **G₀/G₁ Phasenstopp**

Abbildung 2. Click Chemie abgeleitet vom G₄ Liganden

Die Ebenheit des aromatischen Gerüsts und seine Fähigkeit mehr Oberfläche durch Stapeln über dem G₄ zu besetzen, beeinflussen die Fähigkeit der Verbindung bzw. die Stabilisierung der G₄ Bindung. Weitere biophysikalische und biologische Studien erwiesen, dass die Triazole verknüpften Nitroindole vielversprechender als die in Abbildung 2 dargestellten Amin-indolderivate sind. Darüber hinaus wurden die Bedeutsamkeit der funktionellen Nitrogruppe durch molekulare Docking Studien erwiesen, Wasserstoffbrückenbindungen zwischen der Nitrogruppe und den G₄-

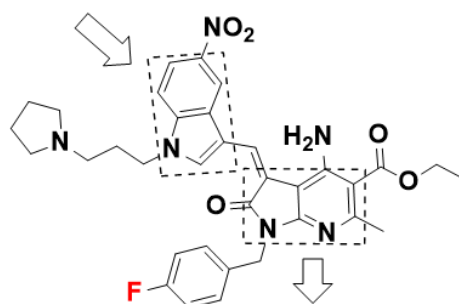
Zusammenfassung

Basenpaaren des G-Quadruplexen sind vorhanden. Nach den biologischen Erkenntnissen haben sich die meisten synthetisierten Triazole verknüpften Nitroindole als wirksam gegen HeLa Zelllinien von humanen Karzinomen (Gebärmutterhalskrebs) erwiesen. Ferner bestätigten Western Blot und Zellzyklus Analysen, dass die neuartigen verknüpften 5-Nitrolindolderivate (9b) die c-MYC Onkogen Expression in Krebszellen durch Stabilisierung Ihrer Promotor Quadruplex Struktur herunterregulieren und den Zellzyklus in der G₀/G₁ Phase zum Stillstand bringen könnte. Die NMR-Analysen legen nahe, dass 9b im langsamen Austausch mit den terminalen G Quartetten (5'- und 3'-Ende) wechselwirkt. Weitere Investigationen sind im Prozess um konkrete Ergebnisse zu erhalten.

In Kapitel IV dieser Arbeit haben wir die Synthesestrategie entwickelt mittels Knoevenagel Kondensation wirksamere G₄ Liganden zu erhalten. Zur Untersuchung neuartiger und selektiver G₄ Liganden für die Krebschemotherapie entwickelten und synthetisierten wir eine Reihe von Azaindol-2-on Derivate (11, 14, 15, 16 und 22), indem wir die kationische Pyrrolidinseitenketten anbrachten und ein Fluoratom in das aromatische Chromophor einführten (Abbildung 1)..

Knoevenagel Kondensation Ansatz

Konjugiertes Indol



Azaindole-Plattform

- IC₅₀ = 2.258 μM
- Selektiver *c-MYC* promoter G-Quadruplex Verbindungen
- Suppression der *c-MYC* Protein expression
- Starke chemische Verschiebung im Intermediaten Austauschregime
- G₀/G₁ Phasenstopp

Abbildung 3. Knoevenagel Kondensation derived Azaindolin-2-one derivative

Fluoratome mit hoher Elektronegativität und geringe Größe weisen häufig besondere Eigenschaften in funktionellen Molekülen auf. Der elektroziehende Effekt von Fluor

Zusammenfassung

konnte die Elektronendichte des aromatischen Chromophors verringern, welches eine stärkere Wechselwirkung mit dem elektronenreichen π -System des G-Quartetts begünstigt hat. Außerdem konnte die Einführung von Fluoratomen die Lipophilie erhöhen und die Bioverfügbarkeit verbessern. Fluoreszenzindikator Verdrängungsassay FID Assays legen nahe, dass die synthetisierten Azaindol-2-on Derivate gegenüber c-MYC G4 DNA und Duplex DNA selektiv sind und eine starke Antikrebsaktivität gegen humane Karzinom (Gebärmutterhalskrebs) HeLa Zelllinien aufweisen. Sie regulieren die c-MYC Expression in Krebszellen herunter, indem sie die Promotor Quadruplex Struktur stabilisieren und den Zellzyklus in der G0/G1 Phase beeinträchtigen. NMR-Spektroskopie unterstützte die Theorie dass das Azaindol-2-on Konjugat mit dem terminalen G Quartetten sowie mit den nahe gelegenen G reichen Trakt (G13-G14-G15 und G8-G9-G10) von c-MYC Quadruplex in intermediaten Austauschregime interagiert.

die in dieser Arbeit erreichten Ergebnissen konnten unser Verständnis für die Wirkung von c.MYC G4 auf die transkriptionelle Regulation von Onkogenen verbessern. Ebenfalls konnte gezeigt werdendass die pyrolidinsubstituierten 5-Nitroindolderivate (Kapitel II), triazol gebundene Nitroindolderivate (Kapitell III) und Azaindol-2-on Konjugate (Kapitel IV) zu einer potentiellen Leitstruktur für die weitere Entwicklung der Krebsbehandlung für c-MYC Promotor G4 werden könnte. Solche strukturelle Informationen können bei der Optimierung von G4 Liganden für eine erhöhte Bindungsaffinität und Selektivität gegenüber G-Quadruplexstrukturen sehr hilfreich.

List of abbreviations

^{13}C NMR	Carbon nuclear magnetic resonance
^1H NMR	Proton nuclear magnetic resonance
2D-COSY	Two dimensional correlation spectroscopy
A	Adenine
AlCl_3	Aluminum trichloride (Lewis Acid)
ATP	Adenosine triphosphate
BF_4	Tetrafluoroborate
br	Broad-IR
Bcl2	B-cell lymphoma 2
C	Cytosine
CD	Circular dichroism
Cs_2CO_3	Cesium carbonate
CD_3CN	Deuterated acetonitrile
c-myc	Oncogene DNA sequence
ct-DNA	Calf thymus DNA
d	Doublet
dd	Doublet of doublets
D_2O	Deuterated water
d_6 -DMSO	Deuterated dimethyl sulfoxide
DC_{50}	Concentration required to decrease fluorescence by 50 %
DFT	Density functional theory
DMF	Dimethylformamide
DCE	1, 2-Dichloroethane
DNA	Deoxyribose nucleic acid

DOSY	Diffusion ordered spectroscopy
DMSO	Dimethylsulfoxid
ds26	Self-complementary (26 base) duplex forming DNA
EI	Electron impact - mass spectrometry
EA	Ethyl acetate
Et ₃ N	Triethylamine
eq	Equivalentents
ESI	Electrospray ionization - mass spectrometry
FAM	6-carboxyfluorescein
F	Fluorine-
FID	Fluorescent indicator displacement
FRET	Förster resonance energy transfer
G	Guanine
G4	Guanine quartet
h	Hour
H	Hydrogen
HCl	Hydrochloric Acid
H ₃ PO ₄	Phosphoric acid
Hela	Henrietta Lacks (uterine cell variety; named for deceased patient- cancer cell)
htelo	Telomeric DNA sequence
imotif	intercalated motif
IC ₅₀	half maximal inhibitory concentration
IR	Infrared
J	coupling constant – NMR
K	Kelvin

K ⁺	potassium cation
Kb	Kilobase
KCl	potassium chloride
LiOH	lithium hydroxide
LiClO ₄	Lithium perchlorate
m	medium – IR
M	Molar
m	multiplet – NMR
m/z	mass to charge ratio
mdeg	Millidegrees
MeOD	deuterated methanol
min	Minute
M	Molar
mM	Millimolar
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetra-zolium bromide
Na ⁺	sodium cation
NH ₄ OH	Ammonium hydroxide
NaCl	Sodium chloride
nm	Nanometer
NMR	Nuclear magnetic resonance
NOESY	Nuclear overhauser effect spectroscopy
os	Circular
PAGE	polyacrylamide gel electrophoresis
PDB	Protein Database
POCl ₃	Phosphorus oxychloride

ppm	parts per million
PQS	putative quadruplex sequences
RNA	ribonucleic acid
s	Singlet
s	strong - IR
sc	Supercoiled
SPR	surface plasmon resonance
T	Thymine
T47D	breast cancer cell line
t	Triplet
TAMRA	6-carboxytetramethylrhodamine
THF	Tetrahydrofuran
TLC	thin layer chromatography
T _m	melting temperature
TO	thiazole orange
UV	Ultraviolet
UV-vis	ultraviolet-visible
vs	very strong -IR
w	weak -IR
δ	chemical shift – NMR
ϵ	molar absorption coefficient
λ	Wavelength
λ_{max}	maximum wavelength
μM	Micromolar

Chapter 1

General Introduction

Chapter 1 General Introduction

This thesis will explore the design and activity of a new class of G-quadruplex DNA recognition as potential anticancer agents. To contextualize why such agents are so important, existing structures of DNA and the agents that bind to them first need to be considered.

1.1 CANCER: A Collection of Related Diseases

Cancer is the name given to a collection of related diseases. In all types of cancer, some of the body's cells begin to divide without stopping and spread into surrounding tissues. Cancer can start almost anywhere in the human body, which is made up of trillions of cells. Normally, human cells grow and divide to form new cells as the body needs them. When cells grow old or become damaged, they die, and new cells take their place.¹

When cancer develops, however, this orderly process breaks down. As cells become more and more abnormal, old or damaged cells survive when they should die, and new cells form when they are not needed. These extra cells can divide without stopping and may form growths called tumors. Many cancers form solid tumors, which are masses of tissue. Cancers of the blood, such as leukemias, generally do not form solid tumors. Cancerous tumors are malignant, which means they can spread into, or invade, nearby tissues. In addition, as these tumors grow, some cancer cells can break off and travel to distant places in the body through the blood or the lymph system and form new tumors far from the original tumor. Unlike malignant tumors, benign tumors do not spread into, or invade, nearby tissues. Benign tumors can sometimes be quite large, however. When removed, they usually don't grow back, whereas malignant tumors sometimes do. Unlike most benign tumors elsewhere in the body, benign brain tumors can be life threatening.²

1.1.1 Differences between Cancer Cells and Normal Cells

Cancer cells differ from normal cells in many ways that allow them to grow out of control and become invasive. One important difference is that cancer cells are less specialized than normal cells. That is, whereas normal cells mature into very distinct cell types with specific functions, cancer cells do not. This is one reason that, unlike normal cells, cancer cells continue to divide without stopping. In addition, cancer cells are able to ignore signals that normally tell cells to stop dividing or that begin a process known as programmed cell death, or apoptosis, which the body uses to get rid of unneeded cells.³

Cancer cells may be able to influence the normal cells, molecules, and blood vessels that surround and feed a tumor—an area known as the microenvironment (Fig. 2A). For instance, cancer cells can induce nearby normal cells to form blood vessels that supply tumors with oxygen and nutrients, which they need to grow. These blood vessels also remove waste products from tumors (Fig. 2A & 2B). Cancer cells are also often able to evade the immune system, a network of organs, tissues, and specialized cells that protects the body from infections and other conditions (Fig. 2D). Although the immune system normally removes damaged or abnormal cells from the body, some cancer cells are able to “hide” from the immune system. Tumors can also use the immune system to stay alive and grow. For example, with the help of certain immune system cells that normally prevent a runaway immune response, cancer cells can actually keep the immune system from killing cancer cells.

Cancer is a genetic disease—that is, it is caused by changes to genes that control the way our cells function, especially how they grow and divide. Genetic changes that cause cancer can be inherited from our parents. They can also arise during a person’s lifetime as a result of errors that occur as cells divide or because of damage to DNA caused by certain environmental exposures (Fig. 1). Cancer-causing environmental exposures include substances, such as the chemicals in tobacco smoke, and radiation, such as ultraviolet rays from the sun. Each person’s cancer has a unique combination of genetic changes. As the cancer continues to grow, additional changes will occur. Even within the same tumor, different cells may have different genetic changes. In general, cancer cells have more genetic changes, such as mutations in DNA, than normal cells. Some of these changes may have nothing to do with the cancer; they may be the result of the cancer, rather than its cause. ⁴

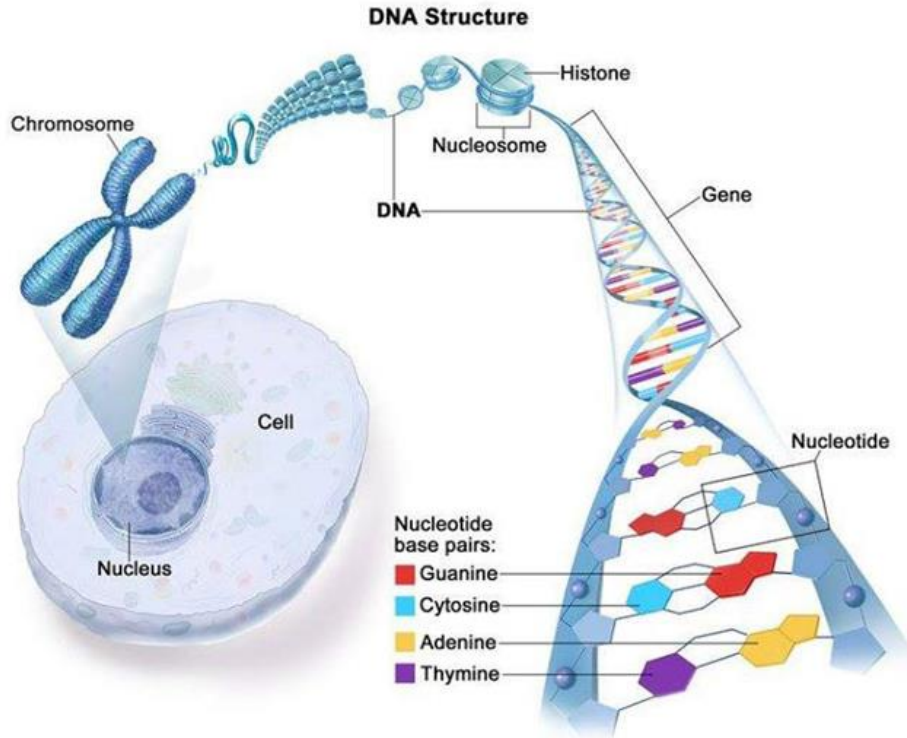


Figure 1. Cancer is caused by certain changes to genes, the basic physical units of inheritance. Genes are arranged in long strands of tightly packed DNA called chromosomes (Reproduced from reference ⁴).

1.1.2 "Drivers" of Cancer

The genetic changes that contribute to cancer tend to affect three main types of genes—proto-oncogenes, tumor suppressor genes, and DNA repair genes. These changes are sometimes called “drivers” of cancer. ⁵

- **DNA repair genes** are involved in fixing damaged DNA. Cells with mutations in these genes tend to develop additional mutations in other genes. Together, these mutations may cause the cells to become cancerous (Fig. 2B).
- **Tumor suppressor genes** are also involved in controlling cell growth and division. Cells with certain alterations in tumor suppressor genes may divide in an uncontrolled manner (Fig. 2B).
- **Proto-oncogenes** are involved in normal cell growth and division. However, when these genes are altered in certain ways or are more active than normal, they may become cancer-causing genes (or oncogenes), allowing cells to grow and survive when they should not (Fig. 2C).

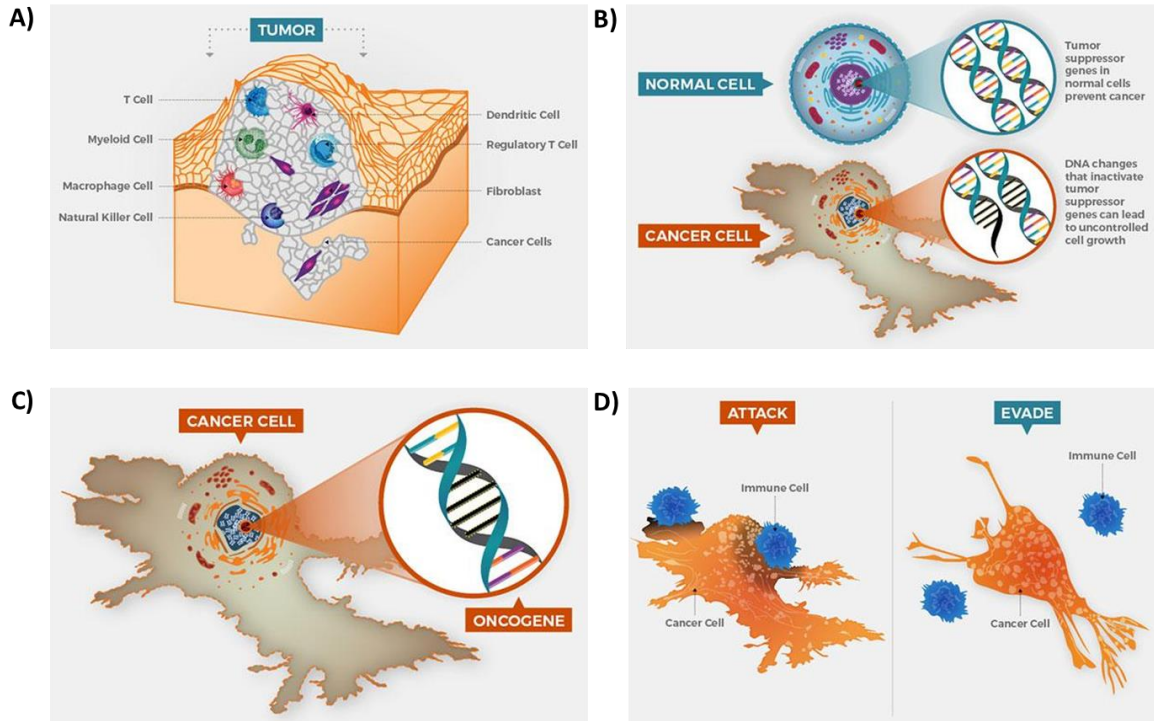


Figure 2. A) Within a tumor, cancer cells are surrounded by a variety of immune cells fibroblasts molecules, and blood vessels-what's known as tumor microenvironment, Cancer cells can turn the microenvironment, which in turn can affect how cancer grows and spread; B) In normal cells tumor suppressor genes prevent cancer by slowing or stopping cell growth. DNA changes that inactivates tumor suppressor genes can lead to uncontrolled cell growth and cancer; C) A DNA changes can cause genes involved in normal cell growth to become oncogene. Unlike normal cells oncogene cannot be turned off, so they cause uncontrolled cell growth; D) Immune system cells can detect and attack cancer cells. But some cancer cells can avoid detection or thwart an attack. Some cancer cells help the immune system better detect and kill cancer cells (Reproduced from reference ⁵ and <https://www.cancer.gov/about-cancer/causes-prevention/genetics>).

As scientists have learned more about the molecular changes that lead to cancer, they have found that certain mutations commonly occur in many types of cancer. Because of this, cancers are sometimes characterized by the types of genetic alterations that are believed to be driving them, not just by where they develop in the body and how the cancer cells look under the microscope.

1.1.3 Metastasis: When Cancer Spreads

A cancer that has spread from the place where it first started to another place in the body is called metastatic cancer. The process by which cancer cells spread to other parts of the body is called

metastasis. Metastatic cancer has the same name and the same type of cancer cells as the original, or primary cancer. For example, breast cancer that spreads to and forms a metastatic tumor in the lung is metastatic breast cancer, not lung cancer. Under a microscope, metastatic cancer cells generally look the same as cells of the original cancer (Fig 3).⁶

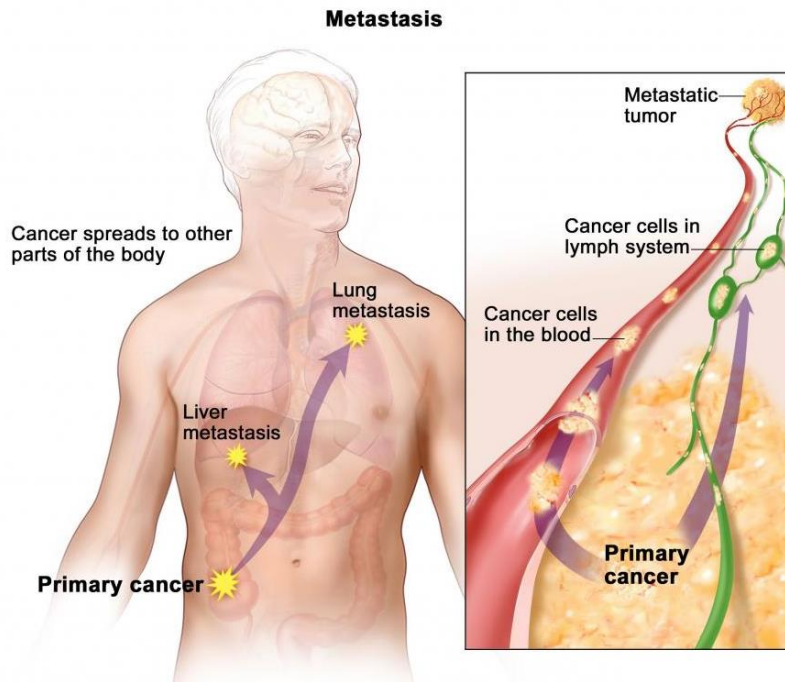


Figure 3. In metastasis, cancer cells break away from where they first formed (primary cancer), travel through the blood or lymph system, and form new tumors (metastatic tumors) in other parts of the body. The metastatic tumor is the same type of cancer as the primary tumor (Reproduced from reference⁶ <https://www.cancer.gov/about-cancer/causes-prevention/genetics>).

Moreover, metastatic cancer cells and cells of the original cancer usually have some molecular features in common, such as the presence of specific chromosome changes. Treatment may help prolong the lives of some people with metastatic cancer. In general, though, the primary goal of treatments for metastatic cancer is to control the growth of the cancer or to relieve symptoms caused by it. Metastatic tumors can cause severe damage to how the body functions, and most people who die of cancer is because of metastatic disease (Fig. 3).

1.1.4 Carcinoma

Carcinomas are the most common type of cancer.⁷ They are formed by epithelial cells, which are the cells that cover the inside and outside surfaces of the body. There are many types of epithelial

cells, which often have a column-like shape when viewed under a microscope. Carcinomas that begin in different epithelial cell types have specific names:

- **Adenocarcinoma:** Adenocarcinoma is a cancer that forms in epithelial cells that produce fluids or mucus. Tissues with this type of epithelial cell are sometimes called glandular tissues. Most cancers of the breast, colon, and prostate are adenocarcinomas.

Basal cell carcinoma is a cancer that begins in the lower or basal (base) layer of the epidermis, which is a person's outer layer of skin.

- **Squamous cell carcinoma:** Squamous cell carcinoma is a cancer that forms in squamous cells, which are epithelial cells that lie just beneath the outer surface of the skin. Squamous cells also line many other organs, including the stomach, intestines, lungs, bladder, and kidneys. Squamous cells look flat, like fish scales, when viewed under a microscope. Squamous cell carcinomas are sometimes called epidermoid carcinomas.
- **Transitional cell carcinoma:** Transitional cell carcinoma is a cancer that forms in a type of epithelial tissue called transitional epithelium, or urothelium. This tissue, which is made up of many layers of epithelial cells that can get bigger and smaller, is found in the linings of the bladder, ureters, and part of the kidneys (renal pelvis), and a few other organs. Some cancers of the bladder, ureters, and kidneys are transitional cell carcinomas.
- **Sarcomas:** Sarcomas are cancers that form in bone and soft tissues, including muscle, fat, blood vessels, lymph vessels, and fibrous tissue (such as tendons and ligaments). Osteosarcoma is the most common cancer of bone. The most common types of soft tissue sarcoma are leiomyosarcoma, Kaposi sarcoma, malignant fibrous histiocytoma, liposarcoma, and dermatofibrosarcoma protuberans.
- **Leukemia:** Cancers that begin in the blood-forming tissue of the bone marrow are called leukemias. These cancers do not form solid tumors. Instead, large numbers of abnormal white blood cells (leukemia cells and leukemic blast cells) build up in the blood and bone marrow, crowding out normal blood cells. The low level of normal blood cells can make it harder for the body to get oxygen to its tissues, control bleeding, or fight infections.

There are four common types of leukemia, which are grouped based on how quickly the disease gets worse (acute or chronic) and on the type of blood cell the cancer starts in (lymphoblastic or myeloid).

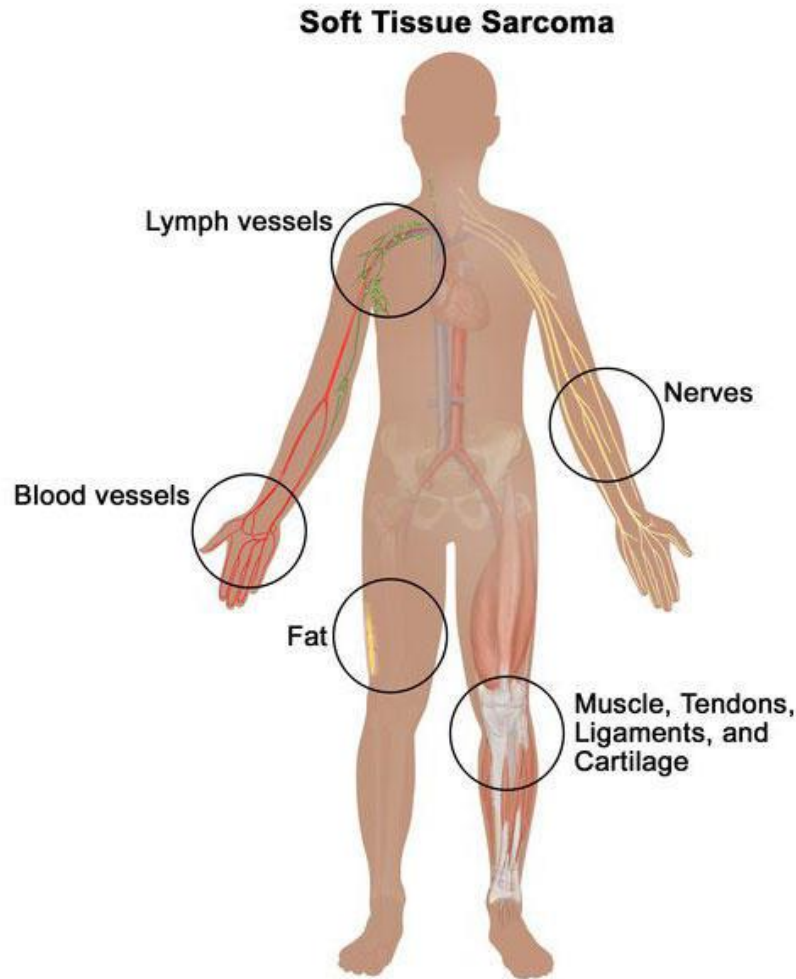


Figure 4. Soft tissue sarcoma forms in soft tissues of the body, including muscle, tendons, fat, blood vessels, lymph vessels, nerves, and tissue around joints (Reproduced from reference⁸ and from NCI website <https://www.cancer.gov/about-cancer>).

- **Lymphoma:** Lymphoma is cancer that begins in lymphocytes (T cells or B cells). These are disease-fighting white blood cells that are part of the immune system. In lymphoma, abnormal lymphocytes build up in lymph nodes and lymph vessels, as well as in other organs of the body.

There are two main types of lymphoma:

- **Hodgkin lymphoma:** People with this disease have abnormal lymphocytes that are called Reed-Sternberg cells. These cells usually form from B cells.

- **Non-Hodgkin lymphoma:** This is a large group of cancers that start in lymphocytes. The cancers can grow quickly or slowly and can form from B cells or T cells.
- **Multiple Myeloma:** Multiple myeloma is cancer that begins in plasma cells, another type of immune cell. The abnormal plasma cells, called myeloma cells, build up in the bone marrow and form tumors in bones all through the body. Multiple myeloma is also called plasma cell myeloma and Kahler disease.
- **Melanoma:** Melanoma is cancer that begins in cells that become melanocytes, which are specialized cells that make melanin (the pigment that gives skin its color). Most melanomas form on the skin, but melanomas can also form in other pigmented tissues, such as the eye.
- **Brain and Spinal Cord Tumors:** There are different types of brain and spinal cord tumors. These tumors are named based on the type of cell in which they formed and where the tumor first formed in the central nervous system. For example, an astrocytic tumor begins in star-shaped brain cells called astrocytes, which help keep nerve cells healthy. Brain tumors can be benign (not cancer) or malignant (cancer).

1.1.5 Other Types of Tumors

- **Germ Cell Tumors:** Germ cell tumors are a type of tumor that begins in the cells that give rise to sperm or eggs. These tumors can occur almost anywhere in the body and can be either benign or malignant.
- **Neuroendocrine Tumors:** Neuroendocrine tumors form from cells that release hormones into the blood in response to a signal from the nervous system. These tumors, which may make higher-than-normal amounts of hormones, can cause many different symptoms. Neuroendocrine tumors may be benign or malignant.
- **Carcinoid Tumors:** Carcinoid tumors are a type of neuroendocrine tumor. They are slow-growing tumors that are usually found in the gastrointestinal system (most often in the rectum and small intestine). Carcinoid tumors may spread to the liver or other sites in the body, and they may secrete substances such as serotonin or prostaglandins, causing carcinoid syndrome.

1.1.6 Tissue Changes that Are Not Cancer

Not every change in the body's tissues is cancer. Some tissue changes may develop into cancer if they are not treated, however. Here are some examples of tissue changes that are not cancer but, in some cases, are monitored:

- Hyperplasia occurs when cells within a tissue divide faster than normal and extra cells build up, or proliferate. However, the cells and the way the tissue is organized look normal under a microscope. Hyperplasia can be caused by several factors or conditions, including chronic irritation.⁹
- Dysplasia is a more serious condition than hyperplasia. In dysplasia, there is also a buildup of extra cells. But the cells look abnormal and there are changes in how the tissue is organized. In general, the more abnormal the cells and tissue look, the greater the chance that cancer will form.⁹

Some types of dysplasia may need to be monitored or treated. An example of dysplasia is an abnormal mole (called a dysplastic nevus) that forms on the skin. A dysplastic nevus can turn into melanoma, although most do not.

An even more serious condition is carcinoma in situ. Although it is sometimes called cancer, carcinoma in situ is not cancer because the abnormal cells do not spread beyond the original tissue. That is, they do not invade nearby tissue the way that cancer cells do. But, because some carcinomas in situ may become cancer, they are usually treated (Fig. 5).

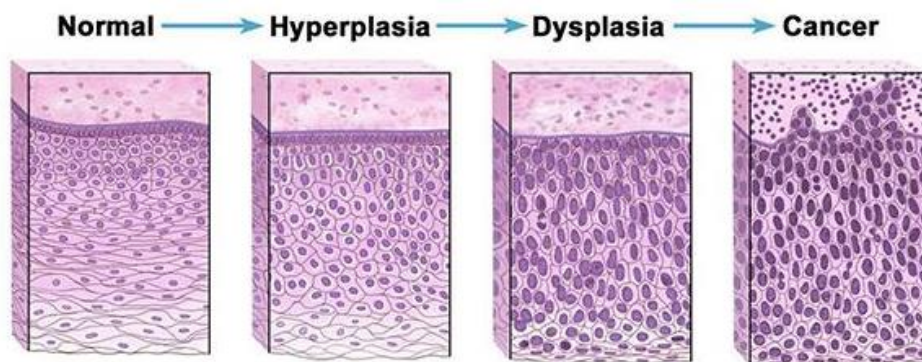


Figure 5. Normal cells may become cancer cells. Before cancer cells form in tissues of the body, the cells go through abnormal changes called hyperplasia and dysplasia. In hyperplasia, there is an increase in the number of cells in an organ or tissue that appear normal under a microscope. In dysplasia, the cells look abnormal under a microscope but are not cancer. Hyperplasia and dysplasia may or may not become cancer (reproduced from reference⁹ and from NCI website <https://www.cancer.gov/about-cancer>).

1.1.7 Development of Chemotherapy with Cell-Cycle Inhibitors

1.1.8 Cell Cycle and Cancer

Cell division is the fundamental process by which a parent cell divides, producing two identical daughter cells. The cell division cycle has four phases: G1 (Gap1), S (synthesis), G2 (Gap2), and M (mitotic). Cells enter the cell cycle in G1 phase through mitogenic signaling and prepare for DNA replication, which occurs in S phase. G2 is the preparation phase for mitosis. When cells stop proliferating they exit the cell cycle and become quiescent (Fig. 6). The cell cycle is highly coordinated by positive and negative regulators. Cell division is the fundamental process by which a parent cell divides, producing two identical daughter cells. The eukaryotic cell division cycle consists of two stages. First, during the long interphase, the chromosomes are replicated.¹⁰

They are then distributed between the two daughter cells during the mitotic phase (mitosis followed by cytokinesis). When they are not dividing, cells are in a quiescence state in the G0 phase. Cells are stimulated to enter the cell cycle by mitogenic signals. The cell cycle has four phases: G1 (Gap1), S (synthesis), G2 (Gap2), and M (mitotic). G1 is the preparation phase for DNA replication which occurs in S phase. G2 is the preparation phase for mitosis. When cells stop proliferating they exit the cell cycle and become quiescent. At the end of G1 phase, cells pass through the restriction point, when they become irreversibly committed to DNA replication and mitogenic signals are no longer required to complete the cell cycle.¹⁰ The four phases are coordinated and the proper progression through the cell cycle requires that a phase must not begin until the previous phase is successfully completed. The transition between two successive phases is irreversible. There are five major transitions:

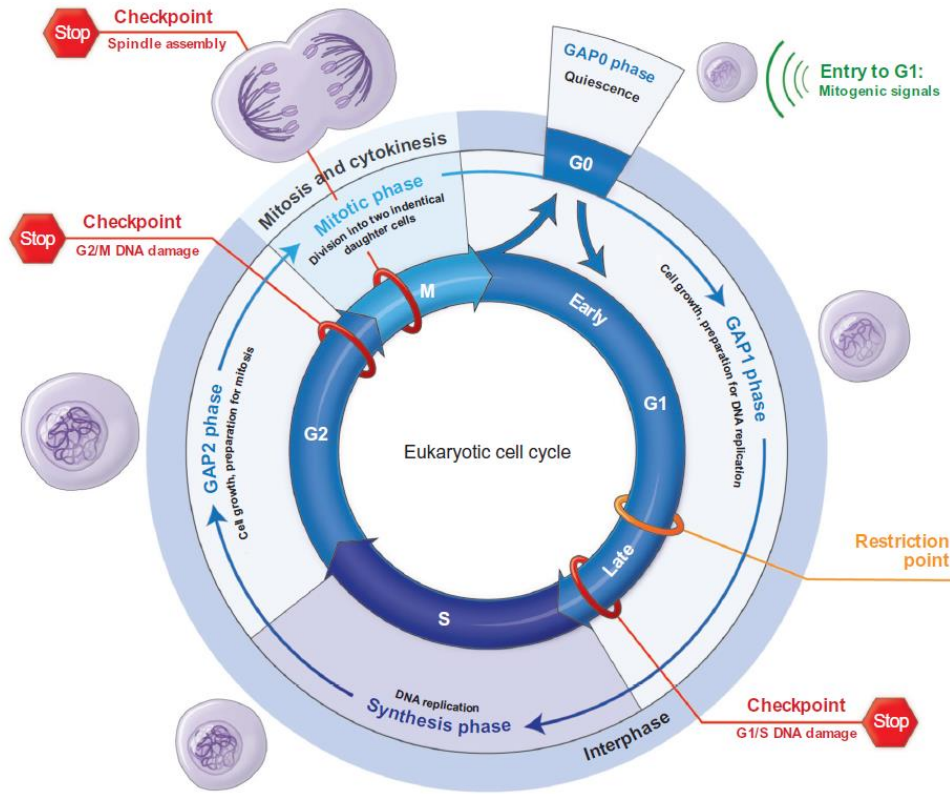


Figure 6. The cell cycle: different phases and major events (reproduced from Ref¹⁰).

G1a/G1b (the restriction point), G1/S (the initiation of DNA replication), G2/M (the entry into mitosis), metaphase/anaphase (the alignment of chromosomes on the mitotic spindle/sister-chromatid separation), and M/G1 (mitotic exit). Complex surveillance mechanisms act at several points in the cell cycle and trigger an arrest when an abnormality is detected. The DNA damage checkpoints¹¹ monitor the DNA integrity before the DNA replication (G1/S transition) and again before mitosis (G2/M transition). Progression to S phase and M phase, respectively, is blocked if DNA repair is needed, thus preventing the propagation of DNA mutations to daughter cells. The spindle assembly checkpoint¹² monitors the proper chromosome attachment to the mitotic spindle before anaphase (metaphase/anaphase transition). If incorrect attachments are detected, progression through mitosis is blocked thus preventing chromosome mis-segregation and the creation of aneuploid daughter cells.

The normal cell cycle consists of complex pathways that regulate the duplication of all molecules and organelles and their separation into two identical daughter cells. This progresses through four phases: G1 (gap), S (synthesis), G2 (gap), and M (mitosis) and is coordinated by cell-cycle

regulators that drive chromosome duplication, chromosome segregation, and cytokinesis. The accuracy of this process is controlled by cyclin-dependent kinases (CDK) and regulatory cyclins as well as checkpoint proteins that delay cell-cycle progression to detect errors and preserve genomic integrity. Although the replication machinery is highly conserved in higher eukaryotes, defects in these pathways can arise as a result of somatic mutations in genes encoding key regulatory proteins.¹³ This, along with defects in checkpoints, can lead to adaptations that result in cancer cell growth. Cell-cycle inhibitors target components that control DNA replication, or coordinate the DNA damage response (DDR) signaling networks, and the mitotic spindle. Inhibition of these processes induces cell cycle arrest and cell death. Current clinical development demonstrates the safety and tolerability of many single agents in adult and pediatric patients, but with limited efficacy, along with various side effects. Combination strategies of agents that induce cell-cycle arrest with cytotoxic therapies demonstrate greater antitumor effects in patients. This article reviews the rationale for dual combination treatments that support the idea that the activity of agents is mechanistically coupled to achieve greater efficacy. This approach is predicted to develop better regimens to improve survival and reduce toxicities in cancer patients.

Cell-Cycle Progression Human cells enter the cell cycle at G1 through various mitogenic stimuli. One such mechanism involves signaling through the Ras/Raf/MAPK pathway that increases cyclin D expression.¹⁴ The CDKs are a family of multifunctional enzymes that bind regulatory cyclins and modulate several protein substrates involved in the cell cycle. The retinoblastoma tumor suppressor gene product (Rb) is phosphorylated by cyclin D/CDK4/6 and cyclin E/CDK2, to govern the G1–S transition.¹⁴ The single replication of DNA during the cell cycle is ensured by loading of origin recognizing complexes onto the DNA at replication origins during G1 phase.¹⁵ Initiation and termination of DNA replication occurs during the S-phase in which 23 pairs of chromosomes are duplicated. The central DNA polymerase Pol δ , together with Pol ϵ and Pol α /primase, synthesize the daughter DNA strands at the eukaryotic replication fork.¹⁶ Pol δ and accessory proteins including proliferating cell nuclear antigen (PCNA) edit or proofread nascent DNA strands to monitor the correct incorporation of nucleotides.

The significance of the intra-S-phase checkpoint is the defense against mutagenic DNA synthesis in replicating cells upon genotoxic challenge. When DNA polymerases encounter errors including mutations and lesions, checkpoint regulation of S-phase progression via inhibition of DNA synthesis, occurs. This triggers ataxia telangiectasia and Rad3-related (ATR) activation¹⁷ to prevent extension of mismatched primers or the introduction of mutations and loss of genomic stability. Eukaryotic cells typically begin mitotic events, such as chromosome condensation,

when DNA replication is complete. During mitosis, the spindle assembly checkpoint restricts the onset of anaphase to facilitate the attachment of kinetochores to spindle microtubules and the tension setting during metaphase, to ensure proper segregation of chromosomes.¹⁸ This is followed by nuclear division and cytokinesis. Together, these checkpoints ensure the coordinated regulation of cell-cycle progression and maintain genetic fidelity between daughter cells.

DNA Damaging Effects of Chemotherapy Standard chemotherapy regimens are used to treat many types of solid tumors and hematologic malignancies. Cytotoxic agents induce DNA damage in cells and activate the intra-S-phase checkpoint to inhibit DNA synthesis,¹⁹ leading to apoptosis or cellular senescence. Chemotherapeutics have diverse actions and cause different forms of DNA damage. Alkylating agents (e.g., temozolomide, dacarbazine cyclophosphamide, ifosfamide) add an alkyl group to DNA binding proteins to prevent the linking of strands in the formation of the DNA double helix. This causes DNA strand breaks, the production of reactive oxygen species and inactivation of DNA polymerases.²⁰

Platinum-based agents (e.g., cisplatin, carboplatin, oxaliplatin) form DNA intrastand and interstrand crosslinks and DNA protein crosslinks that restrict DNA repair.²¹ Doxorubicin intercalates DNA, and disrupts the DNA repairing function of topoisomerase-II,²² preventing the DNA double helix from being resealed and thereby stopping the process of replication. Nucleoside analogues [e.g., gemcitabine, cytarabine, 5-fluorouracil (5-FU)] are incorporated into DNA strands and act as chain terminators to inhibit DNA polymerase. Anti-metabolites (e.g., methotrexate, 5-FU, and mercaptopurine) block essential enzymes necessary for DNA synthesis or also become incorporated into the DNA strand.

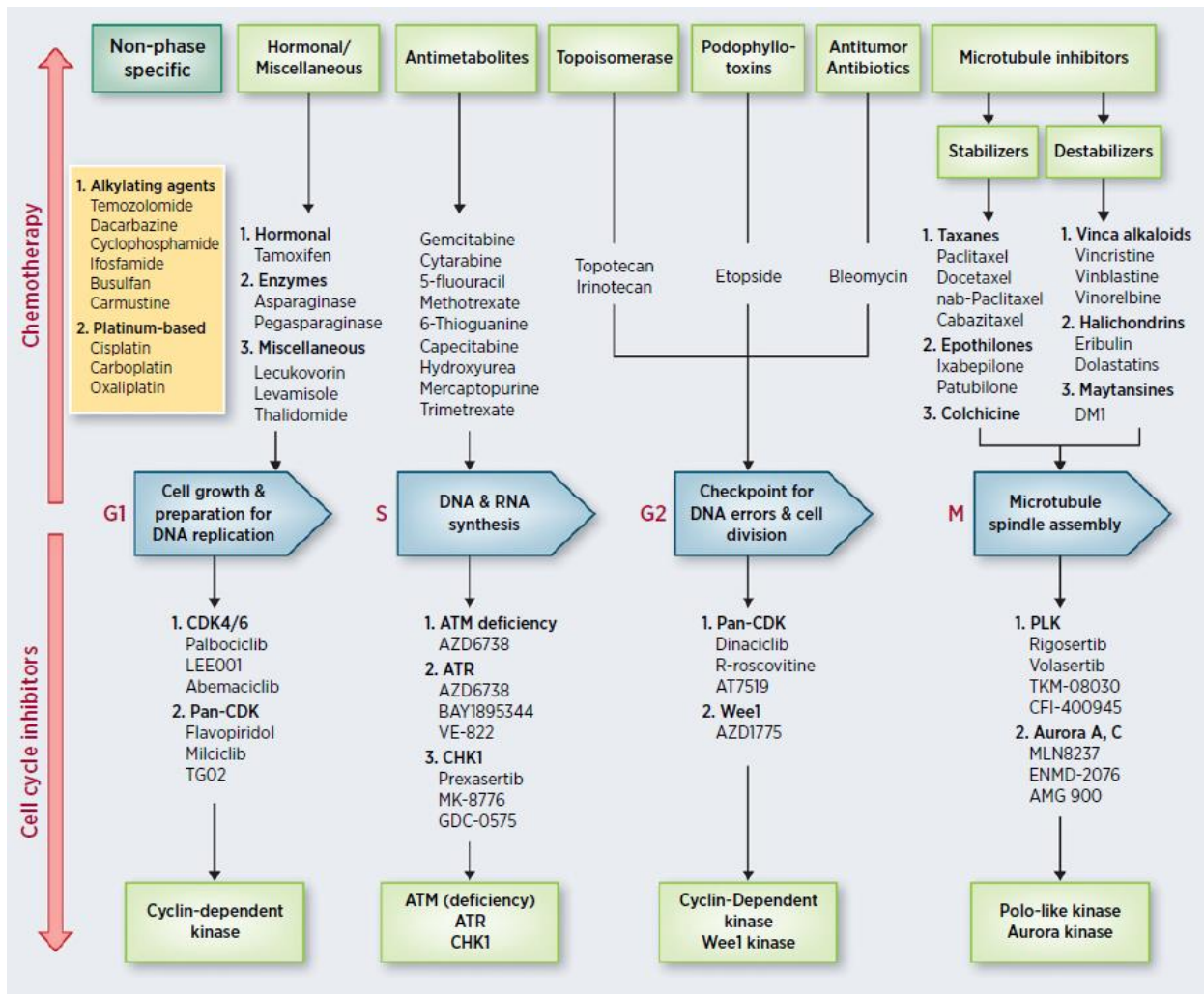


Figure 7. Therapies that target the cell cycle. The mechanism of action of some cytotoxic chemotherapies (top green boxes) may be cell-cycle independent (e.g., alkylating agents, yellow box) or rely on specific phases in the cell cycle (blue arrows) to induce cell-cycle arrest. Bottom, cell-cycle inhibitors target cell-cycle modulators (bottom green boxes) that act at specific phases to delay cell-cycle progression (reproduced from reference²³)

Microtubule inhibitors (vinca alkaloids and taxane) disrupt microtubule spindle formation during metaphase causing mitotic arrest (Fig. 7). Although chemotherapy is highly toxic to the cell cycle, cancer cells may adapt a variety of DNA repair mechanisms or rely on faulty checkpoints to reverse or escape chemical damage to DNA. Activation of DNA repair pathways in tumor cells can mediate base excision repair, mismatch repair, or DNA double-strand repair.²⁴ The recruitment of translesion synthesis (TLS) polymerases by ubiquitylated PCNA allows the repair of bulky lesions to compromise many platinum-based-induced lesions.²⁵ Defects at cell-

cycle checkpoints may allow some lesions to escape DNA polymerase selectivity and proofreading mechanisms, and become replicated. In addition, some base lesions such as O⁶-methylguanine, which are produced by alkylating agents and reactive oxygen species can introduce mutations during DNA synthesis in replicating cells upon genotoxic stress²⁶ and in such cases, drugs become ineffective. In another mechanism, the activation of proteins that are necessary for replication initiation following DNA damage (e.g., chromatin licensing and replication factor protein, Cdt1) can lead to DNA re-replication in neoplastic cells.²⁷ This initiates new rounds of replication of chromosome regions in a single cell cycle, resulting in an increase in aneuploid cell populations and gene amplifications. Several studies demonstrate that cells that have been exposed to DNA-damaging agents and topoisomerase blockers exhibit extended G₂ arrest to delay the onset of mitosis.²⁸ This allows more time for DNA repair processes prior to the start of mitosis, which can cause tumor cells to become refractory to chemotherapy. The fate of cancer cells with DNA damage induced by chemicals, radiation, oxidative, and replication stresses may be lethal or may allow cells to acquire mutations and epigenetic modifications. If abnormal cells are not eliminated, they may evolve to a more aggressive and malignant phenotype that is refractory to treatment.

1.2 DNA Structures, Recognition Sites and Binders

1.2.1 DNA structures and DNA components

Nucleic acids are constituted by the succession of elementary units called nucleotides linked by phosphodiester bounds. The nucleotides can be generated by the phosphorylation of nucleosides. Therefore, each nucleotide contains a nucleoside formed by a nucleobase covalently linked to a pentose through a glycosidic bond and a phosphate group (Fig. 8A). Four primary nucleobases appear in DNA: Adenine (A), Cytosine (C), Guanine (G) and Thymine (T) which is substituted by Uracil (U) in RNA (Fig. 8B). The pentose is 2-desoxy-D-ribose in DNA and D-ribose in RNA.

1.2.2 DNA structural polymorphism

1.2.3 Duplex DNA

In 1953, James Watson and Francis Crick discovered the first double helix DNA structure.^{29,30} DNA is usually represented by a right-handed double helix formed with two anti-parallel strands held

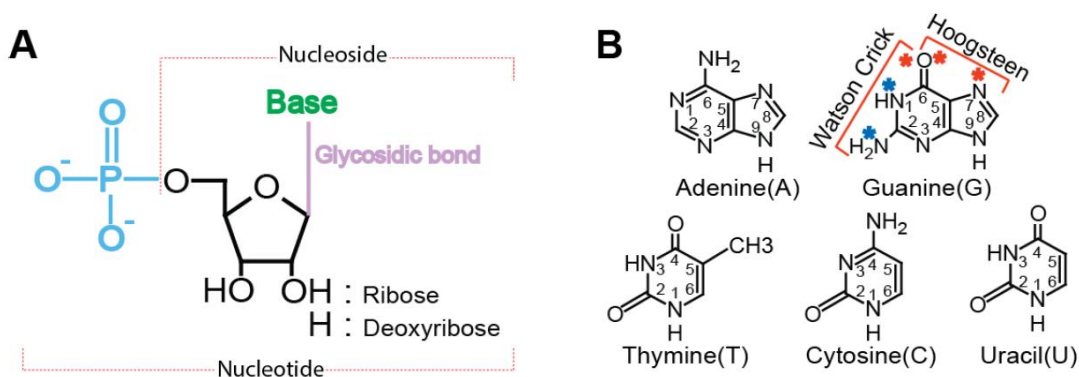


Figure 8. Nucleic acids components. (A) The chemical structures of nucleotide and nucleoside. Nucleobase in green, pentose in black and phosphate group in blue. (B) Schematic structures of the nucleobases constituting DNA and RNA with conventional numbering of atoms. H bond sites involved in Watson Crick and Hoogsteen base pairing (donor and acceptor sites showed on guanine nucleobase by blue and red asterisks respectively).

together by the complementary base pairing through two Hydrogen bonds between the bases A and T and Three between C and G. The helicoidal structure of DNA is stabilized by interactions of π - π stacking between consecutive bases. This DNA form called B form (Fig. 9A) is the canonical conformation predominantly found in the cells. The sugar pucker is always in *C2'-endo/C3'-exo* conformations and the glycosidic bond connecting the base to the sugar adopts an *anti*-conformation. Nevertheless, the B form is not the only base pairing arrangement that can occur between bases. Another right handed duplex DNA form can also be found, the A form (Fig. 9B) observed under conditions of low humidity differs from the B form by the sugar pucker which is in *C3'-endo* conformation. Besides the two first forms, a left-handed conformation called Z form was also highlighted.³¹ This structure is favored at high concentration of salts and requires an alternation of purine-pyrimidine dinucleotides especially C and G. The sugar puckering is in *C2'-endo/C3'-exo* and the glycosidic bond adopts an *anti*-conformation at pyrimidines. By contrast, the sugar puckering is in *C3'-endo/C2'-exo* and the glycosidic bond adopts a *syn* conformation at purines. The change to the *syn* position (Fig. 9B) induces a turning over of bases within the helix and gives to the backbone of the left-handed DNA a zig-zag appearance (Fig. 9A).

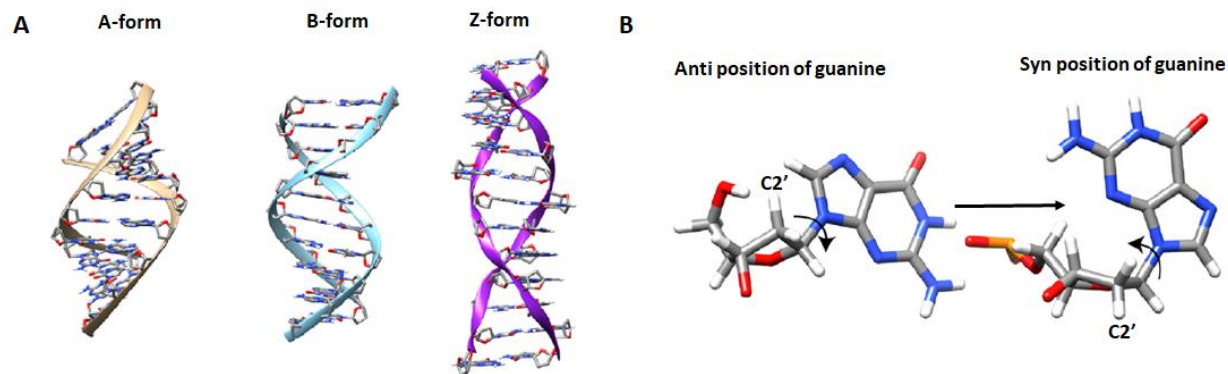


Figure 9. Duplex DNA structures. (A) Models of the A, B and Z DNA. B and A forms are right-handed helices, B form is characterized by a helical turn every 10 base pairs (3.4 nm) with 0.34 nm by base pair. A form is more compact with 11 base pairs per turn. Z form is left-handed with a zig-zag like backbone. (B) Syn and anti-conformations of glycosidic bond angles of guanines as in B and Z DNA. C2'-endo/C3'-exo and C2'-exo/C3'-endo sugar puckers of B and Z DNA. Images generated using UCSF Chimera.³²

1.2.4 Triplex DNA

In 1975, G. Felsenfeld et al observed a structure formed by a triple helix.³³ It is formed by the insertion of a third strand in the major groove of the duplex DNA (Fig. 10B). The formation of this structure requires an asymmetric double helix where one strand contains mostly pyrimidine bases while its complementary strand is mainly constituted by purine bases. The purine-rich strand is recognized by the third strand through non-canonical base pairing. These connections established *via* H bonds can explain the formation of the triplex. This non-canonical base pairing was highlighted by Karst Hoogsteen using X-ray diffraction and are called "Hoogsteen" (Fig. 10A).^{34,35}

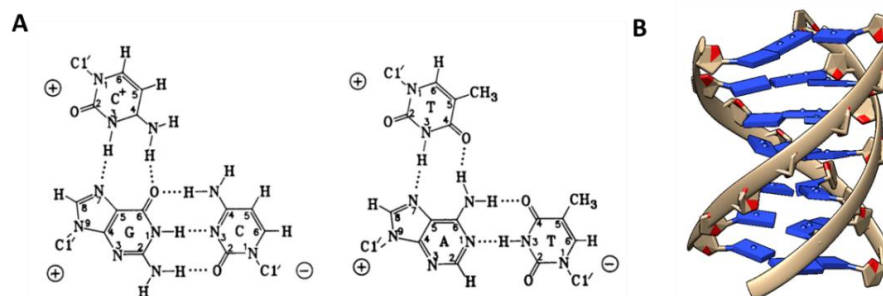


Figure 10. Triplex DNA structure. (A) Non-canonical base pairing of nucleotides within a triple helix of DNA. (B) Model of triplex DNA adapted from the structure pdb 149D obtained by NMR.³⁶ Images generated using UCSF Chimera.³²

1.2.5 Guanine quadruplex (G4) structures

Besides these DNA structures, another non-canonical base pairing can be established within guanine rich DNA sequences between guanines *via* four hydrogen bonds involving the "Watson Crick" and "Hoogsteen" edges. The association of four guanines by a network of eight H bonds generates a planar structure called a G-quartet.³⁷ The stacking of several G-quartets forms a G4 (Fig. 11).³⁸ These structures representing the main object of this work will be described more in detail in the rest of the manuscript. The study of G4 structures present a large interest in view of their structural diversity and the fact that they can be used as potential therapeutic targets.³⁹⁻⁴²

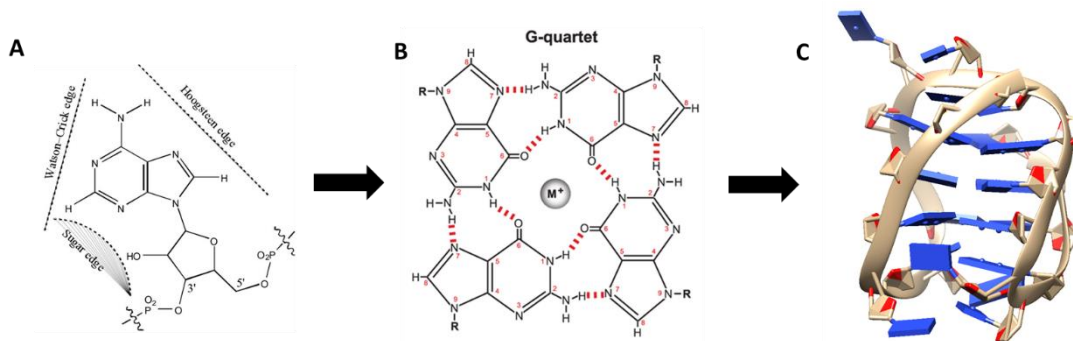


Figure 11. G4 DNA structure. (A) Schematic structure of guanine with the Hoogsteen and Watson Crick faces, the atoms marked with asterisks are involved in the H bonds. (B) Presentation of a G-quartet formed by the association of four guanines. (C) Model of G4 formed by the stacking of three G-quartets adapted from the Human telomeric sequence, $(AG_3[(T_2AG_3)]_3)$, code pdb 143D obtained by NMR.⁴³ Images generated using UCSF Chimera.³²

1.2.6 G4 structures

1.2.7 G-quartets

The planar structures formed by the association of four guanines are called G-quartets (Fig. 4B). The guanines are linked by two networks of H-bonds. The first H-bonds established between the oxygen of carbonyl group in position 6 (O6) and the Hydrogen of imino group in position 1 (N1), the second H bond is established between the nitrogen in position 7 (N7) and the hydrogen of amino group in position 2. The G-quartets are maintained by the presence of monovalent cations, mainly the metallic cations such as K^+ , and to a lesser degree Na^+ ⁴⁴ and NH_4^+ .⁴⁵ Other monovalent and divalent cations such as (Li^+ , Rb^+ , Sr^{2+} , Ca^{2+} and Pb^{2+}) can also be incorporated between G-quartets, to stabilize their formation.^{45,46} The cations interact with the four oxygen atoms belonging to the carbonyl group (O6) of the four guanines during the stacking of G-quartets.⁴⁷ The sugar pucker of the G forming G-quartet is usually in a C2'-endo/C3'-exo conformation. However, the glycosidic bond angle can adopt two different orientations *syn* and *anti*. The *syn/anti* conformations of guanines within a G-quartet define the groove dimension between two adjacent guanines. G4 present three types of grooves, they can be wide, narrow or medium.⁴⁸ The combination of these different conformations allows the specific structural polymorphism of G4; which can contain two or more G-quartets, arranged in intermolecular or intra-molecular structures, and different orientation of the strands.

1.2.8 Structural polymorphism of G4

The stacked G-quartets linked by the phosphate backbone constitute the invariant core of all G4 structures. Even if they are not connected by a phosphate backbone, the guanosines are able to form a helicoidal structure by self-assembly and stacking of G-quartets.⁴⁹ Therefore, a strong cooperation between three key factors including H bonding, Dipole interactions and π - π stacking interactions determines the structural diversity of G4. Different parameters have been described to explain the structural diversity of G4 mainly: the number of strands, their orientations, and the conformation of loops and the nature of cations.

1.2.9 Number of strands

According to the number of DNA molecules involved in the formation of a G4, two main categories can be described. Intermolecular G4 includes three categories; Tetra-molecular formed by the association of four strands, bimolecular formed by two strands and recently Zhou *et al.* were able

to form a G4 with three strands.⁵⁰ The intra-molecular or monomolecular G4 is formed by one strand containing several blocks of guanines able to fold back and form a G4 structure.

1.2.10 Orientation of strands

The orientation of each strand within the G4 is defined by the 5'-3' directionality of the phosphate backbone. Depending on the orientation of the aromatic base towards the sugar, the guanines can adopt two different conformations within the G-quartet, *syn* and *anti* (Fig. 12A). When all the guanines have the same conformation, the strands possess the same polarity and the structure called parallel. However, if the guanines present different GBA (Glycosidic Bond Angles) conformations, 16 conformations are in theory possible for each G-quartet that will govern the orientation of the strands.⁵¹ Therefore, 16 orientations are possible grouped in 4 categories: (a) A parallel conformation containing 4 strands oriented in the same direction 5'→3', (b) a hybrid conformation (3+1) presenting 3 strands in the same direction and the 4th strand in the opposite direction, (c) an anti-parallel conformation with 2 adjacent strands oriented in the same direction and the others in the opposite direction, and (d) another anti-parallel conformation where 2 adjacent strands are oriented in the opposite direction (Fig. 12B).

1.2.11 Conformation of loops

The intra-molecular and bimolecular structures generally present 2 or 3 loops respectively. The loops are the sequences located between the blocks of guanines and linked G-quartets. They present different conformations: (a) Diagonal loop connecting two anti-parallel strands, (b) lateral or edgewise loop connecting two anti-parallel adjacent strands, (c) propeller or chain reversal loop connecting two parallel adjacent strands and (d) V-shaped or snap-back loop connecting the wedges of G-quartet (Fig. 12C). The loops have an important impact on the conformational polymorphism and the stability of G4. The conformation of loops is remarkably dependent on the length of their sequences. In fact, loops containing one or several nucleotides have the ability to link more than two G-quartets.⁵² Short loops are usually external and favor a stable parallel conformation. However, longer loops give rise to anti-parallel conformation but with reduced stability of G4.⁵³⁻⁵⁶ The nucleotide type within loops plays also a role in G4 stability, the presence of adenine reduces remarkably their stability comparing to pyrimidines (C and T).⁵⁷

1.2.12 Nature of cations

The G4 interact with cations in two different manners: a specific interaction with the central ions inserted between G-quartets and an interaction with external ions which contribute to the partial screening of the negative charges of the phosphate backbone. The internal interaction was first evidenced by crystallography,^{58,59} where dehydrated cations are inserted between the carbonyl group of superior and inferior G-quartets. Recently, Sket *et al* demonstrated using NMR spectroscopy that smaller protons (1H) are able to exchange from one ammonium ion to another within d(G₃T₄G₄)₂ dimeric G4, while larger ammonium ions could not exchange through the central G-quartets.⁶⁰ According to cation type, different structures of G4 were highlighted and resolved both by NMR and crystallography, mainly the telomeric G4 which can adopt several conformations.

1.2.13 Intra-molecular G4 motifs

The formation of intra-molecular G4 requires DNA sequences containing at least 4 blocks of guanines with more than 2 guanines for each block separated with a variable number of nucleotides. This involves the formation of three single-stranded loops in the G4 structure in which the G-tracts would form continuous columns supporting the G-tetrad core, while the linkers would form loops connecting the corners of the G-tetrad core. The number of guanines in each block and the size of loops are critical for the stability of the structure. The measurements of thermodynamic stability according to the length of loops and the proximity of guanines within G4 led to the formulation of a consensus sequence capable of forming an intra-molecular G4 used in the search for putative quadruplex sequences (PQS) in the genome G₃-5NL1G₃-5NL₂G₃-5NL₃G₃-5.⁵³ To date, most of G4 structure resolved complies with this consensus. Nevertheless, it has been shown that several examples defy the description of this consensus.

1.2.14 Bulges in G4

As previously referred, exceptions to the consensus sequence have been reported by NMR and crystallography such as the sequence Pu24 from the human *c-myc* promoter⁶¹ and *c-kit* 87up from the human *c-kit* promoter.^{62,63} These two sequences show discontinuous disposition of guanines in one column of the G-tetrad core, despite the presence of four G-tracts each having at least three continuous guanines. The presence of bulges which are a projections of bases from the G-tetrad core have been observed. Recently, Phan *et al* have shown that many different bulges

can exist in G4 structure.⁶⁴ They vary in their sequence, size, position and numbers within the G4. Expanding this description should help to identify more potential G4 forming sequences.

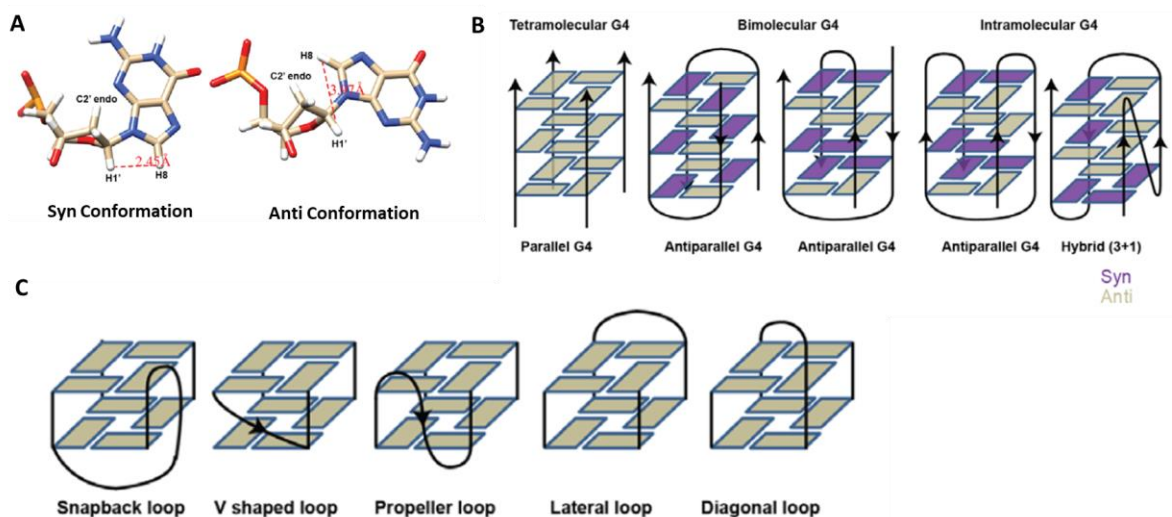


Figure 12. G4 polymorphism. (A) Conformations of guanine base towards sugar, anti (in the right) and syn (in the left). Average Distances of H8-H1' corresponding to each conformation are showed in red. (B) Schematic representation of different G4 conformations according to the molecularity (Tetra-, bi- and uni- or intra-molecular), strand orientations (Parallel, anti-parallel and hybrid (3+1)).⁶⁴ (C) Schematic representation of different loop conformations within G4 (Lateral, propeller, diagonal, V shaped and snap-back loops).⁶⁵

1.3 Biological functions of G4

1.3.1 G4 in silico evidence

The ability of certain guanine rich sequences to form spontaneously intramolecular G4 structures in physiological mimicking conditions suggests their existence *in vivo*. Besides these experimental observations, different *in silico* approaches were used to localize sequences susceptible of forming G4 within genomes of different species. Notably, the Quadparser algorithm⁶⁶ allows to identify sequences respecting the following motif $d(G_3+N_{1-7}G_3+N_{1-7}G_3+N_{1-7}G_3+)$. This program is usually used to identify putative quadruplex sequences (PQS) susceptible of forming G4 in human genome. Using Quadparser, more than 370000 sequences were detected. More recently, another tool representing an extension of Quadparser was developed.⁶⁷ Together with predictive algorithm Quadparser, this data repository contains a large database of predicted G4 (QuadDB) in the human and other genomes. By limiting the number of successive G to 5 nucleotides, the result obtained was nearly identical to that of Quadparser.⁶⁸ The limitation of

loop length to 7 nucleotides, in favor to reduce the complexity of data processing suggests that such sequences are more widespread, as stable G4 can be formed with longer loops.⁶⁹ The potential for G4 DNA formation (G4P) is based on a more qualitative description. The sequences show at least four blocks of three guanines in a window of more than 100 nucleotides, with a slippery interval of 20 nucleotides. The G4P is defined as the percentage of hits in the total number of analyzed windows. More recent analysis showed that the density of PQS is higher in promoters close to the transcription start sites (TSS).

The PQS density and the G4P are correlated with GC content of the genome. However, it is important to note that they are not correlated to the enrichment of CpG sites which requires the alternation of bases G and C rather than an asymmetry of G bases on one strand, and C bases on the other. Furthermore, the density of PQS is also correlated with the presence of DNase I hypersensitive sites (DHS). These sites are characteristic of regions undergoing transition by the opening of B-DNA and conformational changes in the promoters required for the activation of transcription. The G4 structures may have a role in *Cis*-Regulatory Elements (CREs) for 40 % of the human genome genes.⁷⁰ In addition, a correlation between G4P and gene function was highlighted. Several functions mainly G-protein coupled receptor, sensory perception, nucleosome assembly, nucleic acids fixation, Ubiquitin cycle, adhesion and cell division are characterized by a low G4P. Furthermore, a high G4P is associated with functions such as development, immune response, activity of transcription factors, cell signaling, muscle contraction, growth factors and cytokines. It is also interesting to note that the tumor suppressor genes are characterized by low G4P, as opposed to proto-oncogenes that have a high G4P.⁷¹ These studies suggest that G4 could play dynamic roles within the cell. Besides studies in human genome, bioinformatics allowed to identify 854 PQS in *Saccharomyces cerevisiae* using an algorithm with a slippery window of reasonable size (< 50 nucleotides).^{72,73} Among others, these predictive tools are limited. The precise prediction of structure and stability for a new sequence remain inaccurate.⁷⁴ The experimental data provided describe only the folding of the single strand, they do not take into consideration the presence of the complementary strand which favor the formation of DNA duplex. G4 forming requires the denaturing of DNA duplex, which occurs during replication, transcription or recombination phenomena.⁷⁵ Only telomeric overhangs may form G4 structures without need to dissociate a complementary strand. Moreover, most of experimental observations are acquired *in vitro* conditions. Therefore, they do not take in account the interactions of G4 with biological partners mainly proteins such as helicases⁷⁶ and nucleases like topoisomerase I^{77,78} and topoisomerase II.⁷⁹ Even though limited,

the predictions performed *in silico* on different genomes, namely human,⁶⁶ HIV⁸⁰ and *Saccharomyces cerevisiae*⁷³ genomes provide evidence that the G4 sequences are not randomly localized on genomes but they are specifically distributed in particular regions.

1.3.2 Examples of G4 localization in genomes

1.3.3 Promoters of genes

The promoter is the region of the gene that is localized upstream to the transcription start site. They are often characterized by more richness in guanines and a nuclease sensitivity to DNase I, suggesting that B-DNA can be easily opened in these regions. According to the *in silico* analysis, more than 40 % of human genes contain at least one G4 motif within their promoters.⁷⁰ After the

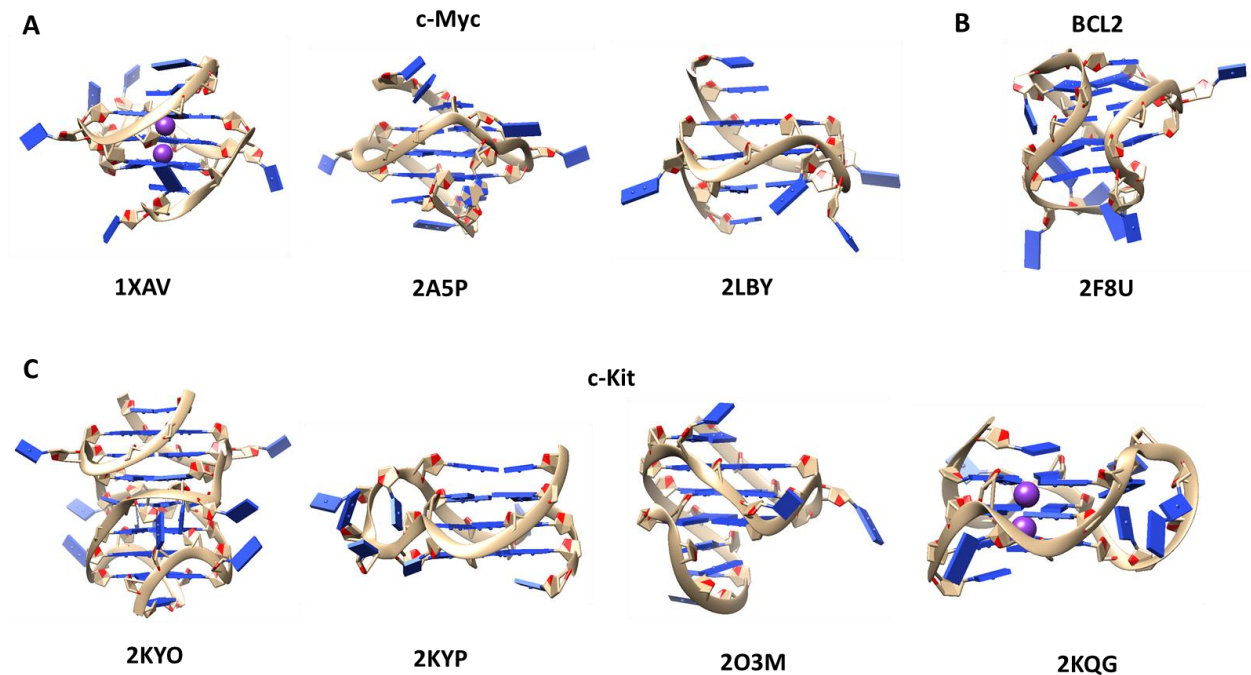


Figure 13: G4 in gene's promoters. Models of different G4 structures formed in genes promoter regions resolved by NMR. Each model is defined by its PDB ID in the bottom. A) Structures formed in c-myc promoter. 1XAV: Parallel G4 formed in K^+ .⁸¹ 2A5P: Parallel G4 formed in K^+ with all anti guanines and a snapback 3'-end syn guanine.⁶¹ 2LBY: Parallel G4 formed in K^+ .⁸⁸ B) Structures formed in BCL2 promoter. 2F8U: Hybrid G4 formed K^+ in containing two lateral loops and one side loop.⁸⁴ C) Structures formed in c-Kit promoter. 2KYO: Monomeric parallel G4 formed in K^+ .⁸⁹ 2KYP: Dimeric parallel G4 formed in K^+ .⁹⁰ 2O3M: Monomeric G4 formed in K^+ containing two single-residue double-chain-reversal loops, a two-residue loop, and a five-residue stem-loop.⁸⁹ 2KQG: Monomeric parallel propeller-type conformation G4 formed in K^+ .⁹¹ Images generated using UCSF Chimera.³²

telomeres, these regions are the most studied for their potential to form G4. The proto-oncogenes seem particularly to be enriched with PQS, contrary to tumor suppressor genes.⁷¹ The formation of intramolecular G4 have been studied *in vitro* for different sequences of these promoters, such as *c-myc* (Fig. 13A),^{61,81,82} *bcl2*^{83,84} (Fig. 13B), *c-kit*⁶² (Fig. 13C) and in the human genome. Other proto-oncogenes are less studied such as *Kras*,⁸⁵ *VEGF*.⁸⁶ These proto-oncogenes are involved in different process namely the self-sufficiency for growth signals, insensitivity to anti-growth signals, sustained angiogenesis, evasion of apoptosis, tissue invasion and metastasis and limitless replicative potential (Fig. 14).⁸⁷

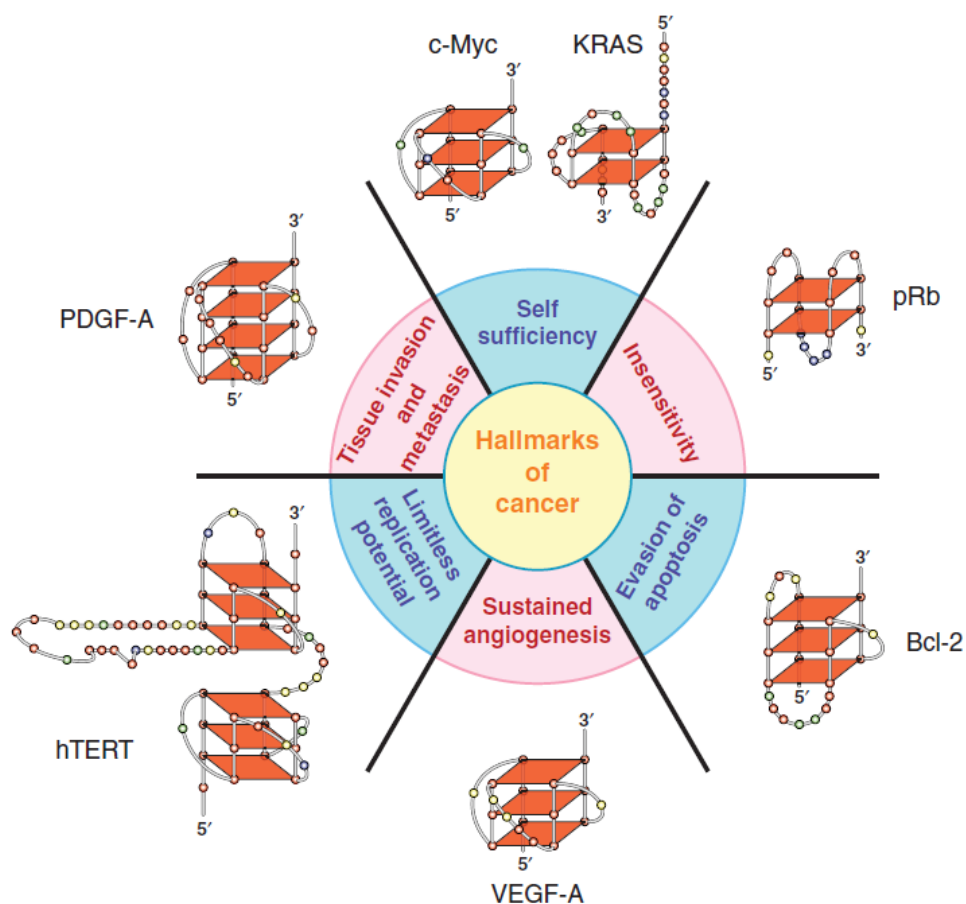


Figure 14. The six hallmarks of cancer shown with the associated G4 found in the promoter regions of genes, figure adapted from.⁹²

1.3.4 Minisatellites

Minisatellites are DNA segments within the genome located mainly near the ends of chromosomes that consist of repeating sequences from 5 to 100 nucleotides. They are known by

their hypermutability and genomic instability.⁹³ It was shown that they have as a generalized function to provide binding sites for recombination proteins in eukaryotes.⁹⁴

Certain Minisatellites sequences possess guanine rich repeated sequences and show their ability to form G4 structures. Amrane *et al.* solved the G4 structure formed by the conserved 26 nucleotides guanine rich fragment of the CEB25 motif in K⁺, which form a monomeric propeller-type parallel-stranded G4.⁹⁵ Another structure of CEB1 minisatellite forming a dimeric-G4 was recently resolved.⁹⁶ This formation of G4 is in agreement with the instability behavior observed in these genomic regions.

1.3.5 Immunoglobulin class switching regions

The genes coding the heavy chain of immunoglobulin undergo a biological mechanism called class switch recombination (CSR), which allows to bring the constant-regions portion close to the variable regions rearranged during B-lymphocyte differentiation. These regions are constituted of guanine rich degenerated repeats with a total length that varies between 1 and 10 kb. The ability of guanine rich motifs to form intermolecular G4 structures was showed for the first time in the CSR regions. The authors give their potential role at the telomeres, particularly during meiosis.⁹⁷

1.3.6 RNA Quadruplex

G4 forming by RNA sequences containing several blocks of guanines is possible in both coding and non-coding RNA. It has also been shown that some G4 RNA possess remarkable stability.⁹⁸⁻¹⁰⁰In addition, the formation of G4 RNA seems all more relevant since the RNA is single-stranded. Such structures could play key roles in their regulation and translation. G4 RNA can also be related to genomic stability and disease. The human fragile X syndrome is caused by the loss of FMRP protein ability to bind RNA or the repression of the gene FMR1 coding the protein. The *in vitro* selection of bound RNA by FMRP showed that the protein bind to G4-prone sequences.¹⁰¹ Certain mRNA coding for proteins involved in the biology of neuronal and synaptic development can also form G4 structures.^{101,102} Bioinformatics studies demonstrated that the 5' and 3' untranslated regions of mRNA are G-rich. It has been supposed that the guanine-rich UTRs can fold into G4, which may play a role in the regulation of translation and degradation of RNA.^{99,103,104} Recent studies reported that telomeric repeat-containing RNA (TERRA) folds into a parallel G4 conformation; additionally, the parallel RNA G4 tends to associate and form higher-order

structure.^{105,106} TERRA could play important roles in regulating telomerase and in chromatin remodeling during cell development.¹⁰⁷

1.3.7 Telomeres

The telomeres are a nucleoprotein complex localized at the end of each chromosome in eukaryotic cells.¹⁰⁸ They play key roles in the protection of genetic material especially towards the double-strand breaks and the erosion due to cell division, and also protect it from the degradation by nucleases and from recombination phenomena.^{109,110} In addition to this role in protecting the genome, telomeres may also have a key role in anchoring the nuclear membrane, in chromosome segregation during mitosis and meiosis as well as in suppressing the expression of genes in sub-telomeric regions.^{111,112} It has been shown that the ends of chromosomes consist of tandem repeats rich in guanines.¹¹³ The first telomeric sequence have been described in the ciliate *Tetrahymena thermophila* which is constituted by the repetition of the motif d(GGGGTT)^{114,115} while in mammalian cells the repeat unit is d(T₂AG₃). Human telomeric DNA consists of approximately 2–10 kb of the double stranded sequence and a single stranded 3' overhang of the G-rich strand, which is approximately 50-100 deoxynucleotides long.^{116,117} The yeast *Saccharomyces cerevisiae* is characterized by the repetition of the motif d(TG₂₋₃(TG)₁₋₆).¹¹⁸ In fact, the presence of repetitive guanine rich motifs forming the overhang at 3' extremity of telomeres suggests that G4 formation is a common feature shared by most telomere sequences.¹¹⁹⁻¹²³ These repeats form a complex with proteins of a wide range of functions. The DNA-protein complex is referred to as the telomeres.¹²⁴⁻¹²⁶ When the progressive erosion of telomeres over the generations is not compensated by an elongation mechanism, the proliferative capacity of the cells is limited, which prevents tumorigenesis. Indeed, in 85% of tumors, the ability of unlimited proliferation of tumor cells depends on the reactivation of a mechanism of telomere elongation, which usually occurs by the over-expression of telomerase.^{127,128} This makes telomerase a privileged target for the treatment of cancer.⁴⁰ The inhibitory effect of a G4 formed in the telomerase substrate on its activity *in vitro* has led to consider an indirect telomerase inhibition strategies by altering the structure of the substrate. Stabilizing a human telomeric G4 by a G4 ligand *in vitro* inhibits the elongation of a telomeric oligonucleotide by telomerase.¹²⁹⁻¹³¹ These experiments greatly stimulated the interest of G4 as therapeutic targets and the development of small molecules able to stabilize G4 specifically. Antibodies that bind specifically to G4 have also been developed: these antibodies are able to detect G4 formation in genomic DNA and RNA of human living cells (Fig. 13A, 13B).¹³²⁻¹³⁴ These observations in

agreement with the experimental data obtained *in vitro* represent the most direct proof for the existence of G4 *in vivo*. These studies also suggest a possible role of G4 in the protection of telomeres. Another alternative structure of the telomere is the T-loop (Telomeric loop) arising from the invasion of the telomeric duplex by the 3' end of the single stranded (Fig. 13C, 13D).¹³⁵ T-loop would hide the telomeric ends by association with the telomeric proteins and could contribute to the protection of the telomere function. *In vitro*, T-loop forming is mediated by the protein TRF2 which is a component of the Shelterin complex. TRF2 binds directly to the telomeric DNA.¹³⁶

1.3.8 G4 structures of human telomeric DNA

The 3' end of human telomeres finish with a single-stranded DNA G-rich overhang.¹¹⁶ Therefore, the DNA of human telomeres can easily fold into a quadruplex structure. However, this single-stranded is capped by proteins such hPOT1.^{137,138} Moreover, the human telomerase enzyme formed of the catalytic domain hTERT and the RNA template recognition domain hTRcap the 3' end in cancer cells.¹¹³ Since the demonstration that telomeres could be promising for cancer treatment, the structural features of the telomeric G4 present a great interest. Several high resolution structures have resolved both by NMR and X-ray crystallography. An anti-parallel G4 structure having a basket-like folding containing one diagonal loop and two lateral loops was obtained by NMR for the truncated 26 human telomeric d(AG₃(T₂AG₃)₃) sequence in sodium solution published in 1993 (Fig. 15: 143D).¹³⁹ Recently, a new anti-parallel (2+2) structure resolved in sodium conditions comprises a novel core arrangement from the known topology (Fig. 15: 2MBJ). Unlike the low polymorphism observed in sodium conditions, a set of polymorphic structures was observed under physiological mimicking conditions rich in potassium. Several structures were obtained from very close sequences.

Two different hybrid (3+1) G4 forms were observed with the sequences d(TAG₃(T₂AG₃)₃) and d(TAG₃(T₂AG₃)₃T₂) in potassium solution. Both structures contain the (3 + 1) G-tetrad core with one double-chain-reversal and two edgewise loops, but differ in the successive order of loop arrangements within the G4 scaffold¹⁴⁰ (Fig. 15: 2JSK, 2JSL, 2JSQ and 2JSM). The sequence A₃G₃(T₂AG₃)A₂ and d(T₂AG₃(T₂AG₃)₃T₂) form hybrid structures differing only in the arrangement of loops, strands orientation and capping structures (Fig. 15: 2JPZ, 2HY9).^{123,141} Another structure resolved for the structure T₂G₃(T₂AG₃)A presenting (3+1) strand fold topology, two edgewise loops and double-chain reversal loop (Fig. 15: 2GKU).^{122,141} The substitution of guanosines with 8-bromoguanosines at proper positions allowed to obtain more leaner NMR spectra and led to

resolve a new structure that is a mixed-parallel/anti-parallel G4 under physiological mimicking conditions (100 mM KCl and 3 mM NaCl)(Fig. 15: 2E4I).¹⁴² An anti-parallel structure very different from those already known was identified by NMR in a solution rich in potassium. It contains only two G-quartets and has a diagonal loop of five bases and two lateral loops (Fig. 15: 2KF7, 2KF8).¹⁴³ In addition to these folding observed in sodium or potassium optimized conditions, another all parallel propeller-type G4 structure was obtained under molecular crowding conditions (Fig. 15: 2LD8).¹⁴⁴ More generally, the high structural diversity of the human telomeric motif in both potassium and sodium conditions clearly show that a particular structure is not only dependent on external factors, but its sequence context is crucial, especially that of the 5' and 3' ends.

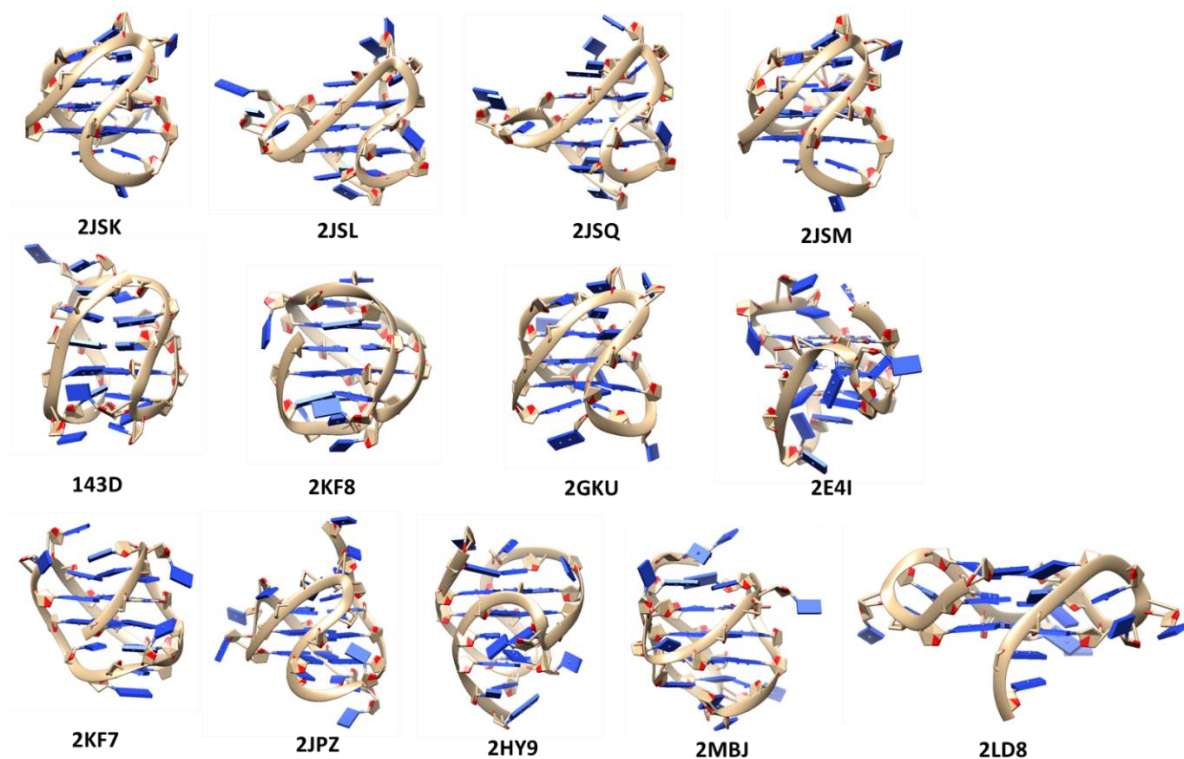


Figure 15. Human telomeric G4 polymorphism. Models of different structures adopted by the human telomeric sequence resolved by NMR. Each model is defined by its PDB ID in the bottom. 2JSK, 2JSL, 2JSQ and 2JSM: Hybrid G4 (3+1) in K⁺ containing two edgewise loops and double-chain reversal loop, two forms are observed from natural and mutated sequences (15BrG).¹⁴⁰ 143D: anti-parallel G4 in Na⁺, a unique basket type folding was observed with three stacked tetrads.¹³⁹ 2KF8, 2KF7: a two G-tetrad basket type G4 formed by a natural and mutated sequence (7BrG).¹⁴² 2GKU: Hybrid G4 (3+1) in K⁺.¹²² 2HY9: hybrid G4 containing a lateral and double-chain-reversal loops (113). 2JPZ: Hybrid G4.¹²³ 2MBJ: anti-parallel (2+2)

G4 in Na⁺.¹⁴⁵ 2E4I: mixed parallel-antiparallel G4 under physiological mimicking conditions (100 mM K⁺, 3 mM Na⁺) obtained with a mutated sequence contain several BrG substitutions.¹⁴² 2LD8: propeller-type all parallel stranded G4 under molecular crowding conditions.¹⁴⁴ Images generated using UCSF Chimera.³²

1.3.9 Biological G-Quadruplexes and their Function

G-Quadruplexes are an emerging target for anticancer complex development and are presently attracting the attention of many research groups trying to find the best specific binder. In order to target such a structure, the rationale for G-quadruplex formation and the conditions under which this occurs need to be examined. G-quadruplex structures have been found throughout the human genome in telomeric regions, oncogene promoter regions, immunoglobulin switch regions and areas where mutation is high.⁴¹ The number of putative quadruplex sequences (PQS) that exist in the genome has been studied by Huppert and Balasubramanian.⁶⁶ Together they developed the 'folding rule' with the formula $d(G_{3+N_{1-7}G_{3+N_{1-7}G_{3+N_{1-7}G_{3+}}})$ which is used in a computer generated quadparser algorithm.⁶⁶ The results after the analysis of the whole human genome with the quadparser algorithm found 376,000 PQS that may be capable of forming G quadruplexes.⁶⁶ Telomeres, consisting of a simple repeat sequence, were found to account for around 20,000 of the PQS found.⁶⁶ This indicates that there is likely to be many more quadruplexes found in other areas such as oncogene promoter regions.⁶⁶

1.3.10 Telomeric DNA

Chromosome ends are protected by a single stranded DNA sequence (telomere) that holds no genetic information, unlike the rest of the chromosomal DNA.¹¹⁰ Telomeres are necessary due to the nature of the replication process because after every successive cycle the DNA is shortened.¹⁴⁶ They also prevent the ends from being recognized as broken DNA and being 'stuck together' by specific repairing proteins.¹⁴⁶ When DNA unwinds in order to replicate, one strand becomes the leading strand and the other the lagging strand. This occurs because DNA polymerase III (an enzyme that builds the new strand in humans) synthesizes in the 5'→3' direction which is the same direction that the DNA is being unwound.¹⁴⁷ Therefore the strand being created in the direction of unwinding is the leading strand and the strand running in the opposite direction has to be synthesised in fragments (Okazaki fragments) hence called the lagging strand.¹⁴⁷

Human telomeric DNA is made up of a repeat base sequence, 5'-TTAGGG-3', which varies in length between 5 and 10 kilobases.^{148,149} The guanine rich nature of this segment of DNA means it is able to fold into the intramolecularly stable G-quadruplex structure and has been proven to do so

under cellular conditions.¹³⁴ The formation of G-quadruplexes is particularly useful in this area as it is believed that their formation inhibits the action of the enzyme telomerase which acts to extend the length of the telomere by adding on TTAGGG repeats to the ends of the DNA.¹⁵⁰

Telomerase is essential for the lifespan of frequently replicated cells such as white blood cells and embryonic stem cells. However, in tumour cells it becomes over-expressed and indeed it is up regulated in 85-90% of cancers.¹⁵⁰ This normally occurs when a mutated form of DNA has reached a critical length and instead of undergoing apoptosis telomerase is introduced and the system continues to replicate indefinitely (Fig. 16).¹⁵¹

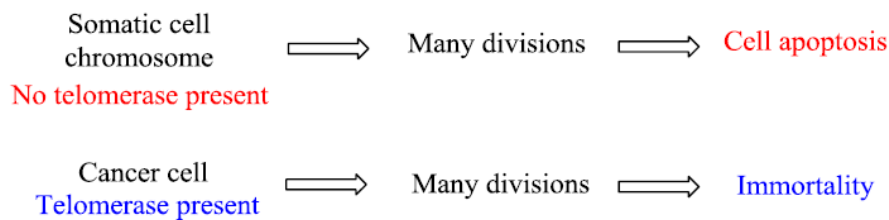


Figure 16. Summary of telomerase presence and its consequence.

1.3.11 Oncogene Promoter Regions

Other areas of DNA that possess G-quadruplex motifs can be found in oncogene promoter regions.¹⁵² This region of DNA can be found upstream of a gene (5' end) and is the control point for regulated transcription.⁹² Specific DNA sequences can be recognized by specific proteins leading to the activation of RNA polymerase and the creation of RNA from the coding part of the gene.⁹² Mutated genes that have the potential to cause cancer are called oncogenes. Their transformation leads to the abnormal regulation of cellular processes which can result in the loss of control of growth signals, the disruption of anti-growth signals, avoidance of apoptosis, continuous angiogenesis, unlimited replication abilities and metastasis.⁹² Stable G-quadruplexes are able to form in oncogene promoter regions and as a consequence they can be linked to the genes altered expression when associated with critical proteins (Fig 17).^{92,153}

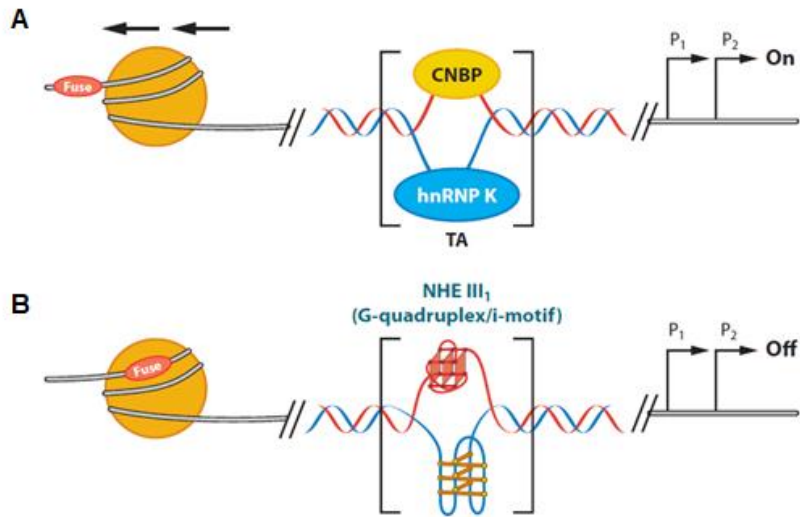


Figure 17. A. The binding of proteins CNBP (zinc finger protein 9) and hnRNP K (heterogeneous nuclear ribonucleoprotein K) to a promoter region causing transcription to be turned on, **B.** When the G-quadruplex and also i-motif are able to form in the G-rich sequences the protein can no longer bind and transcription is turned off. [Reproduced from Ref¹⁵⁴]

The structural diversity of G-quadruplex structures due to differing loop lengths, loop bases and folding patterns in the different oncogene gene promoter regions, potentially allows selective targeting to be accomplished.¹⁵³ Examples of specific oncogene promoter regions where G-quadruplex formation is beneficial to their regulation are shown in table 1.

Table 1 Examples of oncogene DNA regions where G-quadruplexes are able to form due to their guanine rich compositions and the transformations that occur due to deregulation of the oncogene.

Promoter Region	Effect of Transformation (Deregulation)	Quadruplex Effect
c-myc, c-kit and KRAS	Loss of control of growth signals	The formation and stabilization of G-quadruplexes in these areas causes gene silencing, therefore preventing the activation of a specific process such as uncontrolled cell growth. ⁹²
pRb	Disruption of anti-growth signals	If an oncogenic protein binds to this DNA region it can stop the tumour repressor protein from being transcribed which would normally prevent excessive cell growth. G-quadruplexes formed in this region help to prevent the binding of oncogenic proteins preventing insensitivity to tumour growth. ¹⁵³
Bcl-2	Avoidance of apoptosis	Up regulation can play a role in the resistance of conventional cancer treatments and deregulation of cell death. ¹⁵⁵ G-quadruplex formation upstream of the P1 promoter can help prevent Bcl-2 transcription. ¹⁵⁵
hTERT	Uncontrolled replication	Stabilization of G-quadruplexes formed in this area leads to the inhibition of gene expression. ¹⁵⁵ This directly inhibits telomerase expression instead of preventing actual telomerase binding to the telomeres. ¹⁵⁵
PDGF-A	Metastasis	This gene plays an essential role in cellular growth, proliferation, differentiation and development which when over expressed causes tumour growth. ¹⁵⁶ DNA G-quadruplexes formed in this region of the human gene promoter have also been found to inhibit transcriptional activity. ¹⁵⁶

The formation of quadruplexes which turn on and off transcription factors can be better understood when considering the well-studied oncogene c-myc.¹⁵⁴ C-myc expression has been linked to several cancers including breast, cervix, colon, small cell lung cancers, glioblastomas, osteosarcomas and myeloid leukemias.^{154,157} As a proto-oncogene it has a crucial role in the regulation of cellular processes controlling cell growth, cell cycle progression and apoptosis.¹⁵⁸ Strict regulation of c-myc transcription is essential for the role it plays in biological processes.¹⁵⁸ Deregulation can occur when there are alterations in upstream promoter region signaling pathways which in turn leads to an increase in transcription.¹⁵⁸

The formation of quadruplexes which turn on and off transcription factors can be better understood when considering the well-studied oncogene c-myc.¹⁵⁴ c-Myc expression has been linked to several cancers including breast, cervix, colon, small cell lung cancers, glioblastomas,

osteosarcomas and myeloid leukemias.^{154,157} As a proto-oncogene it has a crucial role in the regulation of cellular processes controlling cell growth, cell cycle progression and apoptosis.¹⁵⁸ Strict regulation of c-myc transcription is essential for the role it plays in biological processes.¹⁵⁸ Dereglulation can occur when there are alterations in upstream promoter region signaling pathways which in turn leads to an increase in transcription.¹⁵⁸

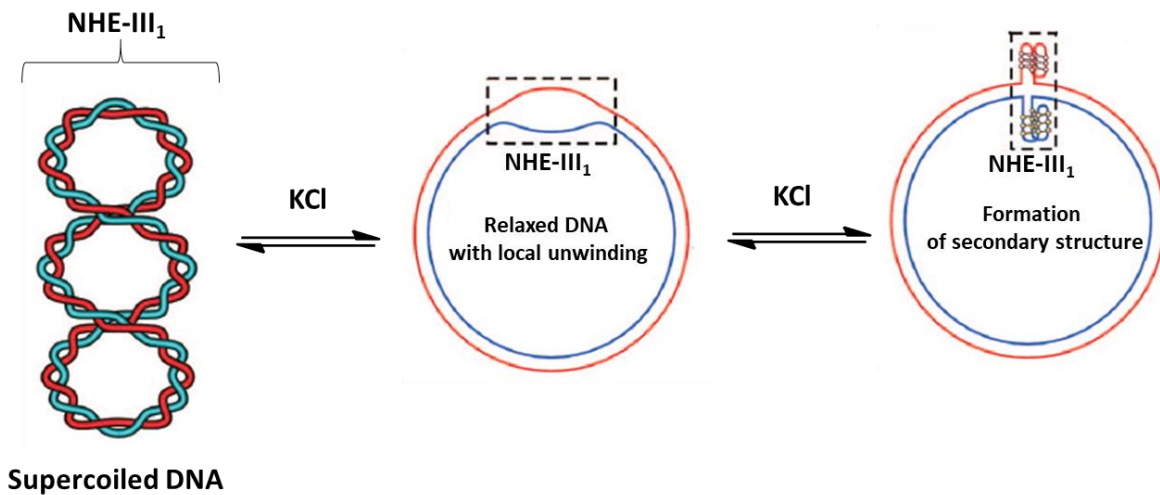


Figure 18. The unwinding of supercoiled DNA demonstrated with a Del4 wild type plasmid with NHE III₁ element showing the difference between the supercoiled and partially relaxed forms that occur during negative super helical stress. [reproduced from reference¹⁵⁹]

The c-myc promoter region DNA is able to exist in alternate forms to the normal double helical B-DNA structure.¹⁵⁹ The transition from B-DNA to single stranded DNA and other non-B-DNA forms such as G-quadruplexes and the i-motif can only occur when accompanied by localized unwinding or melting of the double helix.¹⁵⁹ Negative supercoiling stress under winding (Fig. 18) facilitates this unwinding (melting) process which is naturally formed behind the RNA polymerase complex during gene transcription.^{159,160} The area of most interest in the c-myc promoter region is the nuclease hypersensitive element III₁ region (NHE III₁) as it is here where the G-quadruplexes have the ability to form.¹⁵⁴ The slow equilibrium between duplex B-DNA, single stranded DNA and tetra stranded DNA is essential for controlling the binding of proteins in this area as up to 90% of c-myc transcription has been shown to be regulated by this part of the gene.

1.3.12 Proteins that Bind G-Quadruplexes

In the oncogene cases put forward so far the formation of the G-quadruplex structure prevents the binding of proteins, which require a single stranded binding site, and in doing so are able to alter transcription. There are however proteins that favour G-quadruplex binding over other types of DNA conformations. Interactions with some proteins help to stabilize the G-quadruplex structure and are reported to play a role in transcription and the control of cellular cycles. RecA is one such repair protein that catalyses DNA strand exchange reactions between homologous double stranded DNA and single stranded DNA (homologous recombination).¹⁶¹ The RecA protein is able to bind to single stranded DNA forming a nucleoprotein and it is this ATP dependent process that maintains genomic stability in both prokaryotic and eukaryotic cells.¹⁶¹ More than 50% of the total genomic DNA is made up of repeat DNA sequences many of which form the majority of non-B DNA type structures.¹⁶¹ Research into RecA-DNA interactions has been primarily focused on single stranded DNA that is able to form duplexes; understanding the interactions that evolve between proteins and non-B DNA forming sequences could be more biologically important.¹⁶¹ Further research has found that a labelled human telomeric sequence can preferentially bind to a RecA protein in the presence of K^+ ions in its G-quadruplex state over its single stranded form (Fig. 19).¹⁶¹

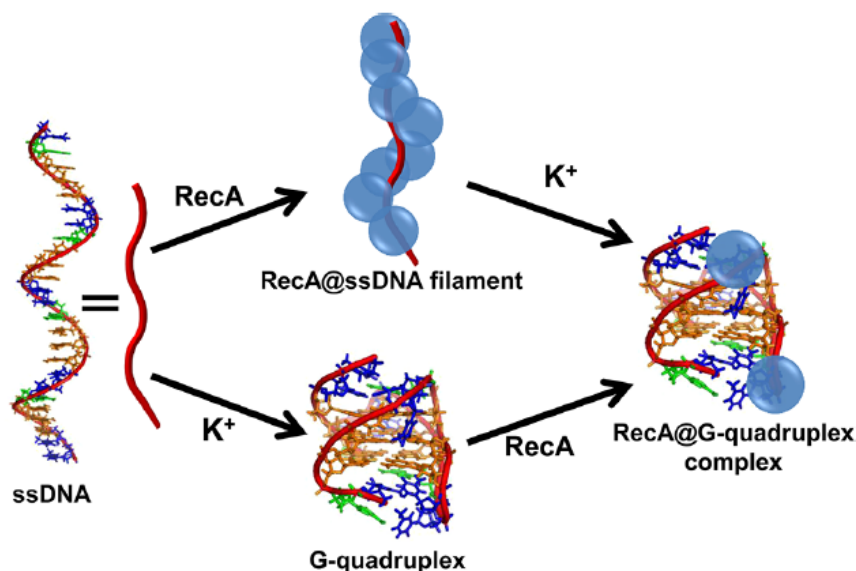


Figure 19. Escherichia coli RecA protein and dye-labelled human telomeric sequence (Cy5-5'-TAGGG-(TTAGGG)₃-TT-3'-Cy3) binding preferences with and without the presence of K^+ ions. [Reproduced from Ref ¹⁶¹]

The nucleoprotein structure formed between the RecA protein and the single stranded telomeric sequence (in the absence of K^+ ions) will dissociate in the presence of K^+ ions with the single stranded sequence folding into its G-quadruplex structure before again forming a nucleoprotein structure with the RecA protein.¹⁶¹

Another G-rich sequence of interest (5'-GGTTGGTGTGGTTGG-3') that folds into an antiparallel G-quadruplex structure, TBA (thrombin binding aptamer), is of particular interest as it is known to bind specifically to the protein thrombin (a coagulation protease) inhibiting it (Fig. 20).¹⁶² Under molecular crowding conditions in a cell mimicking environment (reflects osmotic pressure which changes in live cells where altering water activity can affect biomolecule hydration); binding experiments were conducted to examine the thermodynamics of TBA-thrombin interaction.¹⁶² The investigation revealed that both the binding affinity of the TBA-thrombin interaction and the G-quadruplex stability decrease when conducted in molecular crowding conditions.¹⁶³ This leads to the assumption that water is required for G-quadruplex-protein binding.¹⁶³

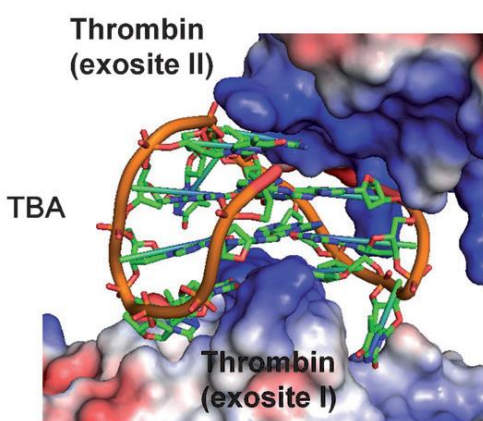


Figure 20. Structure showing the interaction between G-quadruplex (TBA) and thrombin. Binding occurs through two TT loops and one TGT loop of the quadruplex. [Reproduced from Ref ¹⁶³]

This type of binding system can be exploited in order to provide a sensor for thrombin at low detection limits.¹⁶³ Hemin is an iron-containing porphyrin (Fig. 21) which has been found to bind to TBA in an end stacking mode interacting through electrostatics and hydrophobic interactions, resulting in a complex with horseradish peroxidase type activity.¹⁶³ The complex formed between the DNA and hemin is known as a DNAzyme and has been used for the detection of metal ions, small molecules, DNA and proteins.¹⁶⁴

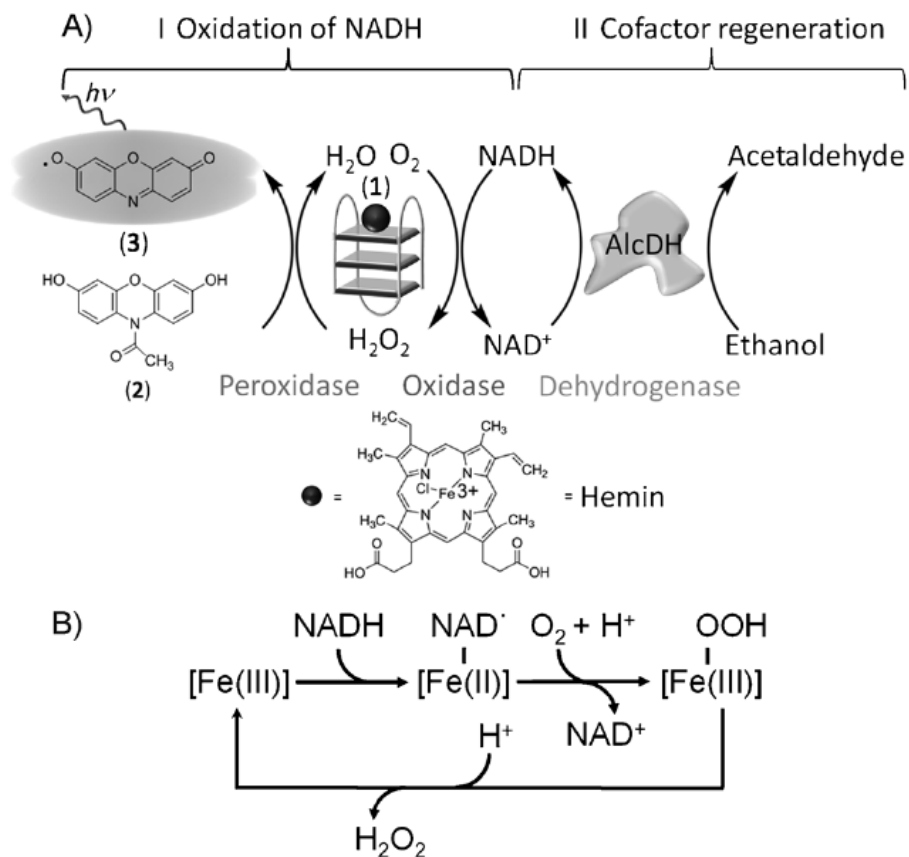


Figure 21. Structure of hemin, chloro (protoporphyrinato) iron (III). [Reproduced from Ref ¹⁶⁵]

The TBA-hemin system is particularly useful as it is able to catalyse the H₂O₂ mediated oxidation of potential fluorophores. In the presence of thrombin the activity of the TBA-hemin system is further increased giving rise to a strong fluorescent emission signal.¹⁶⁶ Advantages of fluorescence detection include low reagent concentrations as the method is highly sensitive, simple experiment set up and quick result turnaround.¹⁶⁶ This system therefore is very attractive in the generation of biological sensors.¹⁶⁶ The potential fluorophore of interest is that of thiamine which is itself non-fluorescent however after oxidation with H₂O₂ it transforms to the strongly fluorescent compound thiochrome (Fig 22).¹⁶⁶

Using this method of fluorescence detection for thrombin, concentrations as low as 1 pM can be recorded creating a highly sensitive thrombin aptasensor.¹⁶⁶ This type of interaction can also be exploited in nano device protein sensors.^{161,163}

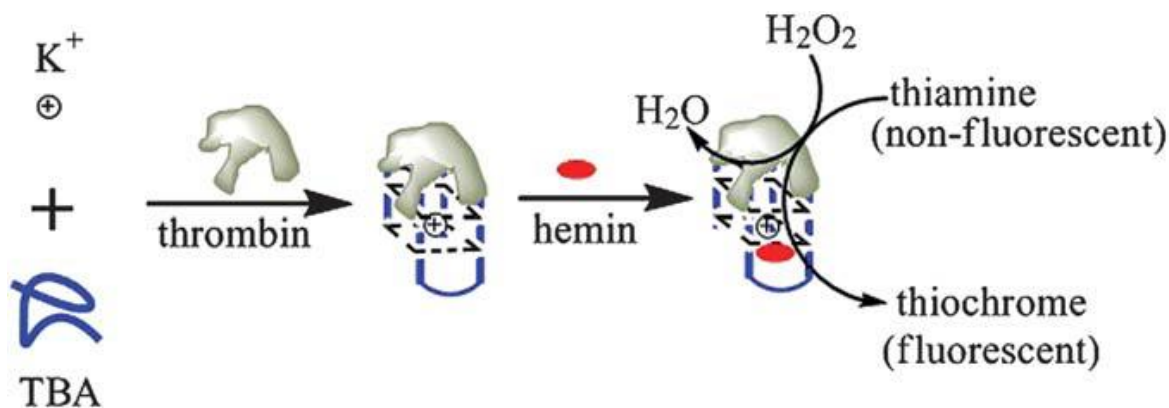


Figure 22. The production of fluorescent thiochrome through the hydrogen peroxide mediated oxidation step initiated by the formation of TBA-hemin complex and thrombin. This results in the fluorescent detection of thrombin. [Reproduced from Ref ¹⁶⁷]

1.4 G4 ligands

As previously mentioned, G4 are highly polymorphic and are susceptible to be formed in different key genomic regions, mainly in telomeres and gene promoters. Even if the existence of G4 *in vivo* is not absolutely demonstrated, the biological effects of selective G4 ligands have to be taken into consideration. As referred before, telomerase is up-regulated in 85% of cancer cells, this perturbation leads cancer cells to become immortal.^{119,168} Consequently, inhibiting telomerase would be a suitable approach for anti-cancer therapy.^{40,169-171} For blocking the active site of telomerase, different inhibitors have emerged.¹⁷² An alternative approach would be to target the substrate of telomerase, which is the 3' end of the single-stranded telomeric DNA overhang. To catalyze the elongation of the telomeric sequence by adding $d(T_2AG_3)$ units, telomerase requires the single-stranded DNA. Hence, the development of G4 ligands that stabilize or induce G4 formation from the 3' single-stranded overhang may prevent telomerase from catalyzing the elongation of telomeres which could be promising for cancer treatment.¹⁷³ The G4 stabilization by G4 ligands is often resulting from π - π stacking and electrostatic interactions.¹⁷⁴ Usually, the development of G4 ligands has been mainly based on polycyclic planar small compounds with positively charged substituents, these particularities lead to the compound stacking on top of the planar G-tetrads, as has been reported by several structural studies of G4-ligands complexes.¹⁷⁵⁻¹⁷⁷ To be more selective, the interacting aromatic surface of ligands have to be larger than DNA duplex binders in order to optimize G-quartet covering and to disfavor the DNA duplex binding. The positively charged ligand are able to interact more strongly with the negatively charged

phosphate backbone of DNA and contribute to stabilization. Hence, to develop a G4 ligand, it is not enough to investigate scaffolds which interact strongly with G4 structures. The main criterion for molecules to be considered as G4 ligand candidate is to show a high selectivity for G4 structures comparing to other DNA forms, particularly the duplex. Besides the targeting of the planar G-quartet, other structural features of G4 have to be taken into account particularly grooves and loops. The channel bearing monovalent cations such as K^+ , Na^+ and NH_4^+ stabilizing the G4 structure is also concerned by the interaction. Concisely, the G4 ligands can be ranked into different groups according to their modes of interaction: interactions of π or end stacking with the terminal G-quartet of G4, hypothetical intercalation of ligands between G-quartet, interactions with grooves and interactions with loops or combination of the two last modes. Among the interaction types, the intercalation of ligands between G-quartets seems unlikely since the unstacking of two G-quartets and the diversion of a cation coordinated between them is energetically unfavorable.⁴¹

1.4.1 π - π stacking interaction with G-quartets

Most of G4 ligands target the invariant part of the structure, the G-quartet, and also possess a large aromatic surface favorable to the interaction with the large hydrophobic surface constituted by the G-quartet. The binding is mediated by the electronic orbital π from the aromatic groups belonging to the ligand and the G-quartet. However, the conception of such molecules with large and planar surfaces containing several aromatic cycles affect their solubility in aqueous conditions. To improve the solubility of ligands, positive charges may be added to the molecule. As a consequence, the stabilization of G4 is based on the π stacking interaction between aromatic surface of the ligand and planar surface of the G-quartet and reinforced by electrostatic interactions between positive charges of the ligand and negative charges of the G4. Different ways can be employed to protonate the ligand mainly the quaternization of amine side-chains *via* in situ protonation, the aromatic N-methylation, the generation of neutral macrocyclic ligands and the chelation of metal cation such as Nickel and Manganese.¹⁷⁴

1.4.2 In situ protonated ligands

As we refer before, the G4 ligands have to conceive large flat aromatic systems prone to π stacking with G-quartet platform, while retaining reasonable water solubility. A classical way is to introduce a protonated side chain at neutral pH by usually adding amine groups around an aromatic core. Thus, the generated molecules are water soluble containing charges distant from aromatic core. Several G4 ligands were generated following this strategy such as Braco-19,

Quindoline, MMQ1, PIPER and a L2H2-6M(2)OTD which is a telomestatin derivative generated by adding protonated side chains to telomestatin (Fig. 23). Different structural studies and structures resolved by X-ray crystallography or NMR showing the interaction between these ligands and G4 have been done. Braco-19 shows an interaction dictated by hydrophobic π stacking interactions between the flat aromatic core of the ligand and two guanine residues of the accessible G-tetrad. Electrostatic interactions between the three protonable side chains of the ligand and the G4-grooves were also highlighted.¹⁷⁷ The main technique widely used to appreciate the affinity of G4 ligands for their targets is the FRET melting assay developed by Jean-Louis Mergny in 2000.¹⁷⁸ In this test, the binding affinity is correlated with the increase in melting temperature (T_m) of a fluorescently labeled G4. In this test, Braco-19 confers a stabilization of 27°C to the human telomeric quadruplex in the FRET-melting test. It exhibits a fair selectivity (31x) for G4 over duplexes as demonstrated by SPR and displays also an important inhibition effect on the telomerase enzyme.^{40,178,179} MMQ1 (Meta-quinacridine n-propylamine) is a pentacyclic quinacridine that displays a crescent shape likely to maximize the overlap with the guanines of the accessible G-quartets.

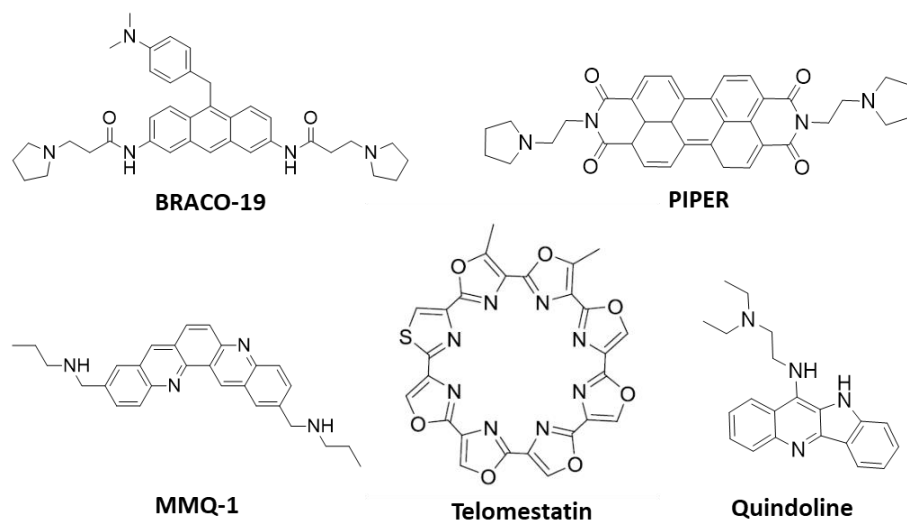


Figure 23. Chemical structures of G4 ligands generated by in situ protonation strategy. BRACO-19 is a 3, 6, 9-trisubstituted acridine. PIPER is N, N'-disubstituted perylene diimide. MMQ1 is a Meta-quinacridine n-propylamine and telomestatin.

The NMR structure of the complex formed between MMQ1 and the tetramolecular G4 d (TTAGGGT)₄ shows ligand binding on the terminal G-quartets of the G4 with a stoichiometry of 2:1. The interaction is mediated by π stacking of the aromatic surfaces and electrostatic

interactions between the side chains of the ligand and the loops of the G4 (Fig. 24).¹⁸⁰ PIPER ligand or perylene diimides is also a G4 ligand: NMR structural studies have demonstrated that PIPER can interact with a tetramolecular G4 by π stacking on the 3' terminal G-quartet. Interestingly, it is also able to be sandwiched between two G-quartet of different G4 inducing dimerization of d[TTAGGG]₄ G4. Moreover, PIPER has an important inhibitory effect on telomerase enzyme.¹⁸¹ NMR studies of the interaction between c-myc G4 and quindoline compound indicates that binding of quindoline increases the stability of the c-Myc G4 by more than 15 °C in its melting temperature at a stoichiometry of 2:1 as showed by 1D NMR-melting. The NMR structure resolved between the two partners demonstrates that two quindoline molecules bind c-myc G4 by π - π stacking generating a complex in which one quindoline covers the top guanine tetrad and another covers the bottom guanine tetrad.¹⁸⁰ Recently, Chung *et al* solved the structure of the complex formed between an intramolecular (3+1) human telomeric G4 and a telomestatin derivative by NMR. The ligand was observed to interact with the G4 by π stacking and electrostatic interactions (Fig. 24).¹⁸²

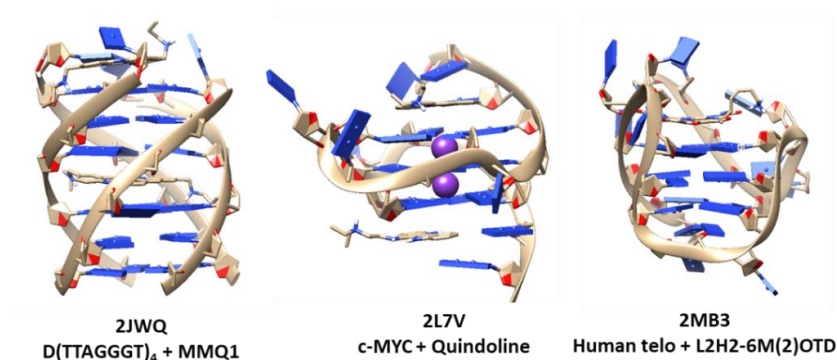


Figure 24. Models of G4-in situ protonated ligand complexes. Each model is defined by its PDB ID in the bottom. 2JWQ: Structure of the complex formed between d(TTAGGGT)₄ and MMQ1 ligand.¹⁸⁰ 2L7V: Structure of the complex formed between c-myc promoter sequence TGAG₃TG₃TAG₃TG₃TA₂ and Quindoline compound.¹⁸³ 2MB3: Structure of the complex formed between Human telomeric sequence T₂G₃T₂AG₃T₂AG₃T₂AG₃A and Telomestatin derivative compound L2H2-6M (2) OTD.¹⁸² Images generated using UCSF Chimera.³²

1.4.3 Aromatic N-methylated ligands

Another way of protonation is to use N-methylated ligands. Such strategy confers a good water solubility to the molecule without the need to add cationic side chains and increase the π stacking ability of the ligand by reducing the electron density of the aromatic part. The most known G4 ligands developed by this approach are the TMPyP4, 360A, Phen-DC3 and RHPS4 (Fig. 25). TMPyP4 is a tetra-cationic porphyrin that inhibits telomerase and possess a high affinity for G4

DNA with a stabilization of 17° Cas showed by FRET-melting, it is also able to regulate the expression of oncogenes such as c-myc and K-ras.^{182,184} Different structural studies were performed to describe the interaction mode of TMPyP4 ligand with G4. The NMR structure of the TMPyP4 complexed with c-myc G4 showed the stacking of the porphyrin onto the external G-quartet.⁶¹ The X-ray structure of the complex formed between TMPyP4 ligand and telomeric G4 surprisingly showed an external stacking onto TTA loops without any direct contact with G-quartets.¹⁸⁵ The pentacyclic acridine RHPS4 was identified as G4 ligand, this molecule interacts with the parallel telomeric G4 d(TTAGGGT)₄ by π stacking with a stoichiometry of 2:1.¹⁷⁶ Besides these ligands an important family of compounds based on the scaffold of PDC (bisquinolinium pyridodicarboxamide) was also developed.¹⁸⁶ The compound 360A is the most known, it confers a stabilization of 23°C for the intermolecular telomeric G4. It also has a very high affinity for the telomeric G4 d(AGGG(T₂AG₃)₃) and affinity for the DNA quadruplex 20 times higher than DNA duplex.^{169,170} To improve the G-quartet covering by PDC compounds, the structure of PDC was modified by the replacement of the central pyridine by a phenantroline generating PhenDC family compounds, the selectivity of these compounds is higher than PDC ligands.¹⁷¹ Recently, Chung *et al* solved by NMR the complex formed between Phen-DC3 and an intramolecular G4 derived from the c-myc promoter. Structural study revealed that Phen-DC3 interacts with the quadruplex through extensive π -stacking with guanine bases of the top G-tetrad (Fig. 26).¹⁸⁷

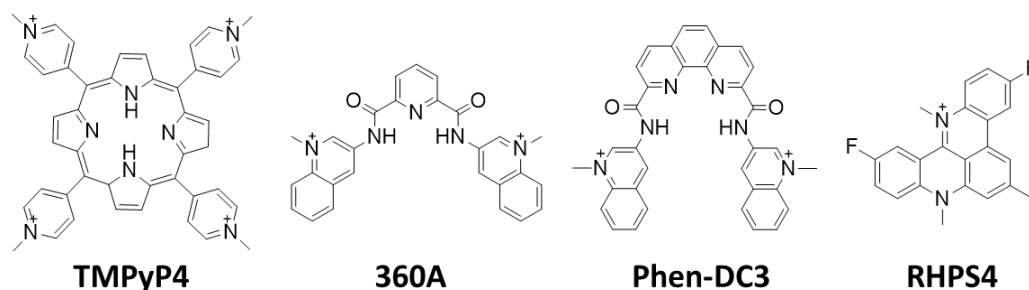


Figure 25. Schematic presentation of the chemical structures of G4 ligands generated by N-methylation strategy.

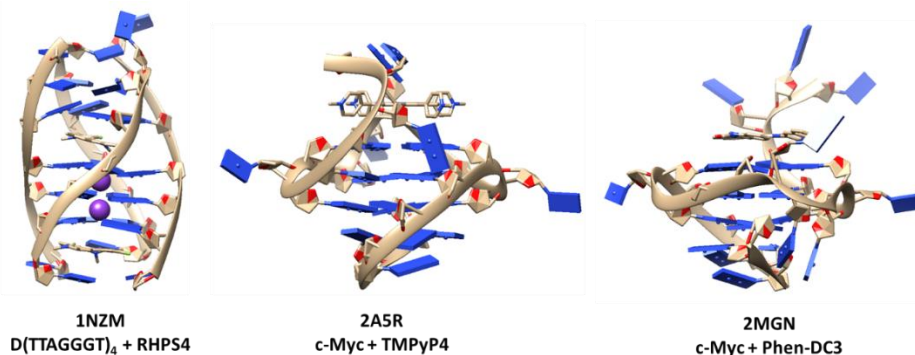


Figure 26. Models of G4-N-methylated ligand complexes. Each model is defined by its PDB ID in the bottom. 1NZM: Structure of the complex formed between d (TTAGGGT)₄ and RHPS4 ligand.¹⁷⁶ 2A5R: Structure of the complex formed between c-myc promoter sequence TGAG₃TG₂IGAG₃TG₄A₂G₂ and TMPyP4 compound.¹⁸⁸ 2MGN: Structure of the complex formed between c-myc promoter sequence TGAG₃TG₂TGAG₃TG₃GA₂G₂ and Phen-DC3 compound. Images generated using UCSF Chimera.³²

1.4.4 Neutral and negatively-charged macrocyclic ligands

Surprisingly, certain ligands possess a good affinity for G4 while they are not positively charged. The interaction of these compounds would be guided by covering the G-quartet with the ligand surface and π stacking interactions. The most famous ligand of this limited group is telomestatin which is a natural compound extracted from *Streptomyces annulatus*.¹⁸⁹ Telomestatin stabilizes human telomeric G4 and shows a high selectivity for G4 structures,¹⁷² it displays an important inhibition of telomerase activity. Other analogues has also been developed such as HXDV conferring to G4 structures a stabilization close to 20°C.¹⁷² Even more surprising is the observation that negatively charged compounds may have a good affinity for G-quadruplex. N-methyl mesoporphyrin IX (NMM) has a strong affinity for some quadruplex conformation despite being negatively charged (Fig. 27).

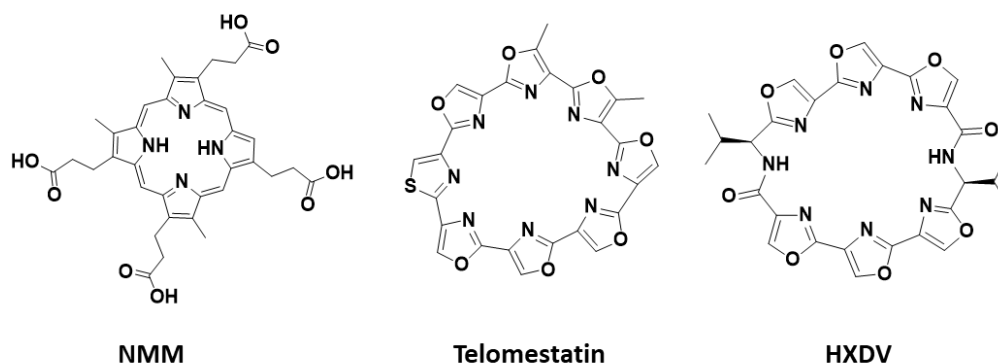


Figure 27. Schematic presentation of the chemical structures of G4 neutral and negatively-charged macrocyclic ligands. Telomestatin, HXDV and NMM as representatives.

1.4.5 Metallo-organic ligands

The design of metallo-organic molecules is also proposed for the development of G4 ligands. This approach is based on the insertion of a central metal cation which can be integrated on the ionic channel of the G- quadruplex. The cationic nature of these molecules helps to improve the association of the ligand with DNA G4 negatively charged, it allow also to optimize π stacking interaction between the two partners.¹⁹⁰ For instance, Ni-salphenes showed an important affinity for telomeric G4 conferring a stabilization higher than 30°C¹⁹¹ and Mn-porphyrine displaying a very high affinity for the telomeric G4 (Fig. 28).¹⁹²

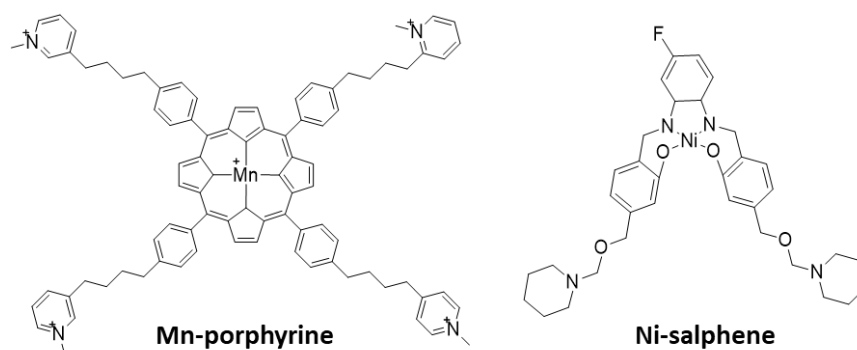


Figure 28. Schematic presentation of the chemical structures of some G4 metallo-organic ligands.

1.4.6 Ligands of grooves and loops

In comparison with duplex DNA, DNA G4 is characterized by the presence of 4 grooves and a large variety of loops. The recognition of grooves and loops confers a high selectivity to G4 DNA. Distamycin A (Fig. 29A) represents the most known ligand of this family.

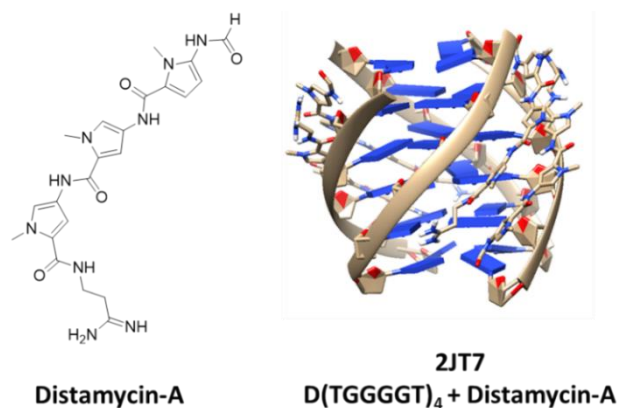


Figure 29. Model of G4-ligand complex showing groove interaction mode. A) Schematic presentation of the chemical structures of Distamycin A compound. **B)** PDB Model 2JT7: Structure of the complex formed between d(TGGGGT)₄ and Distamycin A. Image generated using UCSF Chimera.³²

NMR and thermodynamic studies demonstrate that distamycin A is able to interact with grooves of the intermolecular G4 d[TGGGGT]₄ with a stoichiometry of 4:1. Its crescent shape facilitates the insertion in grooves through the 4 hydrogen bonds established with guanines (Fig. 29B).^{193,194} Gai *et al* have shown the ability of DMSB cyanine dye (2, 20-diethyl-9-methyl selenocarbocyanine bromide) to interact with parallel-stranded G4 d[TGGGGT]₄ in the form of monomer and dimer in different sites. The dimer binding is a dual-site simultaneous binding mode, which is different from normal modes (such as end-stacking or pure groove-embedding). This kind of binding mode involves the recognition of two unique structural features of G4: a terminal G-tetrad and a groove.¹⁹⁵

1.5 Summary and Thesis Aims

G-quadruplexes (G4s) are nucleic acids secondary structures that may form in single-stranded guanine (G)-rich sequences under physiological conditions.^{196,197} Four Gs bind *via* Hoogsteen-type hydrogen bonds base-pairing to yield G-quartets, which in turn stack on top of each other to form the G4. G4s are highly polymorphic, both in terms of strand stoichiometry (forming both inter- and intramolecular structures) and strand orientation/topology. The presence of K⁺ cations specifically supports G4 formation and stability.^{198,199} In the human genome G4 DNA motifs have been found in telomeres, G-rich micro- and mini-satellites, up-stream to oncogene promoters and within the ribosomal DNA (rDNA).²⁰⁰⁻²⁰² Human G4 DNA motifs are over-expressed in recombinogenic regions,^{203,204} which are associated with genomic damage in cancer

cells. Additionally, these regions show mutational patterns that preserve the potential to form G4 DNA structures.²⁰⁵ The identification of G4 binding proteins²⁰⁶ and G4 visualization in cells with antibody-based technology^{37,207} have also provided convincing evidence of the existence of cellular G4s in vivo, prevalently in tumours compared to normal tissues.²⁰⁸ More recently, research on G4s has also focused on prokaryotes²⁰⁹ and viruses. Indeed, G4s have been found to control key viral steps⁴¹ and treatment with G4 ligands has shown to impair viral replication.²¹⁰

Due to the possibility to target pathogenic pathways by stabilizing G4 structures, several G4 binders have been developed. In particular, ligands targeting tumour mechanisms at the telomere and oncogene promoter level have been reported.²¹¹ These share an aromatic core and protonable side chains. Some of these compounds showed interesting antitumor properties; nevertheless, only quarfloxin proceeded into Phase II clinical trials. Unfortunately, its limited bioavailability prevented further progress.²¹¹ The success of tumour targeting through a G4-binding mechanism heavily relies on the sustainable identification structures.²⁰⁵ The identification of G4 binding proteins^{212,213} and G4 visualization in cells with antibody-based technology^{37,214} have also provided convincing evidence of the existence of cellular G4s in vivo, prevalently in tumours compared to normal tissues.²⁰⁸ More recently, research on G4s has also focused on prokaryotes²⁰⁹ and viruses. Indeed, G4s have been found to control key viral steps²⁰⁹ and treatment with G4 ligands has shown to impair viral replication.²¹⁰ Due to the possibility to target pathogenic pathways by stabilizing G4 structures, several G4 binders have been developed. In particular, ligands targeting tumour mechanisms at the telomere and oncogene promoter level have been reported.²¹¹ These share an aromatic core and protonable side chains. Some of these compounds showed interesting antitumor properties; nevertheless, only quarfloxin proceeded into Phase II clinical trials. Unfortunately, its limited bioavailability prevented further progress.¹⁷⁸

The work of this thesis focuses on the targeting of G-quadruplexes, where a simple binder has been designed that, has the potential to bind specifically to G-quadruplex forming DNA. Fragment based drug design approach will be used in these studies to generate hit and its optimization with fragment expansion strategy. The design incorporates indoles and azaindole heterocycles which is able to create a large planar π surface area with linked pyrrolidine side chains with aliphatic alkyl linkers (which impart cationic charge) to create a complex that is able to bind by stacking onto the terminal G quartet of a G-quadruplex structure. It is hoped that by attempting this fragment based drug design and screening novel series of small drug like molecules can be synthesized and enhanced selectivity will be obtained. The interactions of the

hit generated based on fragment screening approach with both duplex and quadruplex forming DNA will be investigated using fluorescent indicator displacements to determine selectivity using thiazole orange, followed by molecular docking analysis and NMR spectroscopy to monitor changes in the quadruplex spectrum as complex is added. The ability of the binder to interact with *c-MYC* promoter DNA quadruplexes will also be investigated along with cytotoxic assays, western blot analysis and cell cycle analysis on cervical cancer cell lines.

1.6 References

- (1) Hanahan, D.; Weinberg, R. A. Hallmarks of Cancer: The next Generation. *Cell*. 2011. <https://doi.org/10.1016/j.cell.2011.02.013>.
- (2) Hieger, I. Cancer Research. *Nature* 1966. <https://doi.org/10.1038/212665a0>.
- (3) Wodarz, A.; Gonzalez, C. Connecting Cancer to the Asymmetric Division of Stem Cells. *Cell*. 2006. <https://doi.org/10.1016/j.cell.2006.03.004>.
- (4) Sharma, S.; Kelly, T. K.; Jones, P. A. Epigenetics in Cancer. *Carcinogenesis*. 2009. <https://doi.org/10.1093/carcin/bgp220>.
- (5) Zhang, J.; Zhang, S. Discovery of Cancer Common and Specific Driver Gene Sets. *Nucleic Acids Res*. 2017. <https://doi.org/10.1093/nar/gkx089>.
- (6) Valastyan, S.; Weinberg, R. A. Tumor Metastasis: Molecular Insights and Evolving Paradigms. *Cell*. 2011. <https://doi.org/10.1016/j.cell.2011.09.024>.
- (7) Steeg, P. S. Tumor Metastasis: Mechanistic Insights and Clinical Challenges. *Nature Medicine*. 2006. <https://doi.org/10.1038/nm1469>.
- (8) Hunt, K. K.; Patel, S. R.; Pollack, A. Soft Tissue Sarcomas. In *Integrated Cancer Management: Surgery, Medical Oncology, and Radiation Oncology*; 1999. https://doi.org/10.5005/jp/books/12869_105.
- (9) Speight, P. M. Update on Oral Epithelial Dysplasia and Progression to Cancer. *Head Neck Pathol*. 2007. <https://doi.org/10.1007/s12105-007-0014-5>.
- (10) Shah, M. A.; Schwartz, G. K. Cell Cycle-Mediated Drug Resistance: An Emerging Concept in Cancer Therapy. *Clinical Cancer Research*. 2001.
- (11) Storchova, Z.; Pellman, D. From Polyploidy to Aneuploidy, Genome Instability and Cancer. *Nature Reviews Molecular Cell Biology*. 2004. <https://doi.org/10.1038/nrm1276>.
- (12) Castel, D.; Philippe, C.; Calmon, R.; Le Dret, L.; Truffaux, N.; Boddaert, N.; Pagès, M.; Taylor, K. R.; Saulnier, P.; Lacroix, L.; et al. Histone H3F3A and HIST1H3B K27M Mutations Define Two Subgroups of Diffuse Intrinsic Pontine Gliomas with Different Prognosis and Phenotypes. *Acta Neuropathol*. 2015. <https://doi.org/10.1007/s00401-015-1478-0>.
- (13) Hsieh, P.; Yamane, K. DNA Mismatch Repair: Molecular Mechanism, Cancer, and Ageing. *Mech. Ageing Dev*. 2008. <https://doi.org/10.1016/j.mad.2008.02.012>.
- (14) Roskoski, R. ERK1/2 MAP Kinases: Structure, Function, and Regulation. *Pharmacological Research*. 2012. <https://doi.org/10.1016/j.phrs.2012.04.005>.
- (15) Abbas, T.; Dutta, A. CRL4Cdt2: Master Coordinator of Cell Cycle Progression and Genome Stability. *Cell Cycle*. 2011. <https://doi.org/10.4161/cc.10.2.14530>.
- (16) Lee, M. Y. W. T.; Zhang, S.; Lin, S. H. S.; Chea, J.; Wang, X.; Leroy, C.; Wong, A.; Zhang, Z.; Lee, E. Y. C. Regulation of Human DNA Polymerase Delta in the Cellular Responses to DNA Damage. *Environmental and Molecular Mutagenesis*. 2012. <https://doi.org/10.1002/em.21743>.
- (17) Cimprich, K. A.; Cortez, D. ATR: An Essential Regulator of Genome Integrity. *Nature Reviews Molecular Cell Biology*. 2008. <https://doi.org/10.1038/nrm2450>.

- (18) Jaffe, A. B.; Hall, A. RHO GTPASES: Biochemistry and Biology. *Annu. Rev. Cell Dev. Biol.* 2005. <https://doi.org/10.1146/annurev.cellbio.21.020604.150721>.
- (19) Luo, Y.; Levenson, J. D. New Opportunities in Chemosensitization and Radiosensitization: Modulating the DNA-Damage Response. *Expert Review of Anticancer Therapy.* 2005. <https://doi.org/10.1586/14737140.5.2.333>.
- (20) Soll, J. M.; Sobol, R. W.; Mosammamarast, N. Regulation of DNA Alkylation Damage Repair: Lessons and Therapeutic Opportunities. *Trends in Biochemical Sciences.* 2017. <https://doi.org/10.1016/j.tibs.2016.10.001>.
- (21) Muniandy, P. A.; Liu, J.; Majumdar, A.; Liu, S. T.; Seidman, M. M. DNA Interstrand Crosslink Repair in Mammalian Cells: Step by Step. *Critical Reviews in Biochemistry and Molecular Biology.* 2010. <https://doi.org/10.3109/10409230903501819>.
- (22) Nitiss, J. L. Targeting DNA Topoisomerase II in Cancer Chemotherapy. *Nature Reviews Cancer.* 2009. <https://doi.org/10.1038/nrc2607>.
- (23) Mills, C. C.; Kolb, E. A.; Sampson, V. B. Development of Chemotherapy with Cell-Cycle Inhibitors for Adult and Pediatric Cancer Therapy. *Cancer Research.* 2018. <https://doi.org/10.1158/0008-5472.CAN-17-2782>.
- (24) Zhang, Y.; Rohde, L.; Wu, H. Involvement of Nucleotide Excision and Mismatch Repair Mechanisms in Double Strand Break Repair. *Curr. Genomics* 2009. <https://doi.org/10.2174/138920209788488544>.
- (25) Ghosal, G.; Chen, J. DNA Damage Tolerance: A Double-Edged Sword Guarding the Genome. *Translational Cancer Research.* 2013. <https://doi.org/10.3978/j.issn.2218-676X.2013.04.01>.
- (26) Marchetti, M. A.; Weinberger, M.; Murakami, Y.; Burhans, W. C.; Huberman, J. A. Production of Reactive Oxygen Species in Response to Replication Stress and Inappropriate Mitosis in Fission Yeast. *J. Cell Sci.* 2006. <https://doi.org/10.1242/jcs.02703>.
- (27) Blow, J. J.; Dutta, A. Preventing Re-Replication of Chromosomal DNA. *Nature Reviews Molecular Cell Biology.* 2005. <https://doi.org/10.1038/nrm1663>.
- (28) van den Heuvel, S. Cell-Cycle Regulation. *WormBook: the online review of C. elegans biology.* 2005. https://doi.org/10.4324/9780203339886_chapter_4.
- (29) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 1953. <https://doi.org/10.1038/171737a0>.
- (30) Watson, J. D.; Crick, F. H. C. Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* 1953. <https://doi.org/10.1038/171964b0>.
- (31) Wang, A. H. J.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; Van Boom, J. H.; Van Der Marel, G.; Rich, A. Molecular Structure of a Left-Handed Double Helical DNA Fragment at Atomic Resolution. *Nature* 1979. <https://doi.org/10.1038/282680a0>.
- (32) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* 2004. <https://doi.org/10.1002/jcc.20084>.
- (33) Felsenfeld, G.; Rich, A. Studies on the Formation of Two- and Three-Stranded

- Polyribonucleotides. *BBA - Biochim. Biophys. Acta* 1957. [https://doi.org/10.1016/0006-3002\(57\)90091-4](https://doi.org/10.1016/0006-3002(57)90091-4).
- (34) Hoogsteen, K. The Crystal and Molecular Structure of a Hydrogen-Bonded Complex between 1-Methylthymine and 9-Methyladenine. *Acta Crystallogr.* 1963. <https://doi.org/10.1107/s0365110x63002437>.
- (35) Hoogsteen, K. The Structure of Crystals Containing a Hydrogen-Bonded Complex of 1-Methylthymine and 9-Methyladenine. *Acta Crystallogr.* 1959. <https://doi.org/10.1107/s0365110x59002389>.
- (36) Radhakrishnan, I.; Patel, D. J. Solution Structure of a Pyrimidine-purine-pyrimidine DNA Triplex Containing T·AT, C⁺·GC and G·TA Triples. *Structure* 1994. [https://doi.org/10.1016/S0969-2126\(00\)00005-8](https://doi.org/10.1016/S0969-2126(00)00005-8).
- (37) GELLERT, M.; LIPSETT, M. N.; DAVIES, D. R. Helix Formation by Guanylic Acid. *Proc. Natl. Acad. Sci. U. S. A.* 1962. <https://doi.org/10.1073/pnas.48.12.2013>.
- (38) Arnott, S.; Chandrasekaran, K.; Marttila, C. M. Structures for Polyinosinic Acid and Polyguanylic Acid. *Biochem. J.* 1974. <https://doi.org/10.1042/bj1410537>.
- (39) Oganessian, L.; Bryan, T. M. Physiological Relevance of Telomeric G-Quadruplex Formation: A Potential Drug Target. *BioEssays*. 2007. <https://doi.org/10.1002/bies.20523>.
- (40) De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J. F.; Mergny, J. L. Targeting Telomeres and Telomerase. *Biochimie* 2008. <https://doi.org/10.1016/j.biochi.2007.07.011>.
- (41) Ou, T. M.; Lu, Y. J.; Tan, J. H.; Huang, Z. S.; Wong, K. Y.; Gu, L. Q. G-Quadruplexes: Targets in Anticancer Drug Design. *ChemMedChem*. 2008. <https://doi.org/10.1002/cmdc.200700300>.
- (42) Döchler, M. G-Quadruplexes: Targets and Tools in Anticancer Drug Design. *Journal of Drug Targeting*. 2012. <https://doi.org/10.3109/1061186X.2012.669384>.
- (43) Wang, Y.; Patel, D. J. Solution Structure of the Human Telomeric Repeat d[AG3(T2AG3)3] G-Tetraplex. *Structure* 1993. [https://doi.org/10.1016/0969-2126\(93\)90015-9](https://doi.org/10.1016/0969-2126(93)90015-9).
- (44) Mergny, J. L.; Phan, A. T.; Lacroix, L. Following G-Quartet Formation by UV-Spectroscopy. *FEBS Lett.* 1998. [https://doi.org/10.1016/S0014-5793\(98\)01043-6](https://doi.org/10.1016/S0014-5793(98)01043-6).
- (45) Wong, A.; Wu, G. Selective Binding of Monovalent Cations to the Stacking G-Quartet Structure Formed by Guanosine 5'-Monophosphate: A Solid-State NMR Study. *J. Am. Chem. Soc.* 2003. <https://doi.org/10.1021/ja0302174>.
- (46) Włodarczyk, A.; Grzybowski, P.; Patkowski, A.; Dobek, A. Effect of Ions on the Polymorphism, Effective Charge, and Stability of Human Telomeric DNA. Photon Correlation Spectroscopy and Circular Dichroism Studies. *J. Phys. Chem. B* 2005. <https://doi.org/10.1021/jp045274d>.
- (47) Pinnavaia, T. J.; Marshall, C. L.; Fisk, C. L.; Miles, H. T.; Becker, E. D. Alkali Metal Ion Specificity in the Solution Ordering of a Nucleotide, 5'-Guanosine Monophosphate. *Journal of the American Chemical Society*. 1978. <https://doi.org/10.1021/ja00479a070>.
- (48) Smith, F. W.; Feigon, J. Quadruplex Structure of Oxytricha Telomeric DNA Oligonucleotides.

- Nature* 1992. <https://doi.org/10.1038/356164a0>.
- (49) Wong, A.; Ida, R.; Spindler, L.; Wu, G. Disodium Guanosine 5'-Monophosphate Self-Associates into Nanoscale Cylinders at PH 8: A Combined Diffusion NMR Spectroscopy and Dynamic Light Scattering Study. *J. Am. Chem. Soc.* 2005. <https://doi.org/10.1021/ja042794d>.
- (50) Zhou, J.; Amrane, S.; Korkut, D. N.; Bourdoncle, A.; He, H. Z.; Ma, D. L.; Mergny, J. L. Combination of I-Motif and G-Quadruplex Structures within the Same Strand: Formation and Application. *Angew. Chemie - Int. Ed.* 2013. <https://doi.org/10.1002/anie.201301278>.
- (51) Webba Da Silva, M. Geometric Formalism for DNA Quadruplex Folding. *Chem. - A Eur. J.* 2007. <https://doi.org/10.1002/chem.200701255>.
- (52) Črnugelj, M.; Šket, P.; Plavec, J. Small Change in a G-Rich Sequence, a Dramatic Change in Topology: New Dimeric G-Quadruplex Folding Motif with Unique Loop Orientations. *J. Am. Chem. Soc.* 2003. <https://doi.org/10.1021/ja0348694>.
- (53) 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. <https://doi.org/10.1093/nar/gkq639>.
- (54) Guédin, A.; Alberti, P.; Mergny, J. L. Stability of Intramolecular Quadruplexes: Sequence Effects in the Central Loop. *Nucleic Acids Res.* 2009. <https://doi.org/10.1093/nar/gkp563>.
- (55) Rachwal, P. A.; Brown, T.; Fox, K. R. Sequence Effects of Single Base Loops in Intramolecular Quadruplex DNA. *FEBS Lett.* 2007. <https://doi.org/10.1016/j.febslet.2007.03.040>.
- (56) Rachwal, P. A.; Findlow, I. S.; Werner, J. M.; Brown, T.; Fox, K. R. Intramolecular DNA Quadruplexes with Different Arrangements of Short and Long Loops. *Nucleic Acids Res.* 2007. <https://doi.org/10.1093/nar/gkm316>.
- (57) Guédin, A.; De Cian, A.; Gros, J.; Lacroix, L.; Mergny, J. L. Sequence Effects in Single-Base Loops for Quadruplexes. *Biochimie* 2008. <https://doi.org/10.1016/j.biochi.2008.01.009>.
- (58) Laughlan, G.; Murchie, A. I. H.; Norman, D. G.; Moore, M. H.; Moody, P. C. E.; Lilley, D. M. J.; Luisi, B. The High-Resolution Crystal Structure of a Parallel-Stranded Guanine Tetraplex. *Science (80-.)*. 1994. <https://doi.org/10.1126/science.8036494>.
- (59) Phillips, K.; Dauter, Z.; Murchie, A. I. H.; Lilley, D. M. J.; Luisi, B. The Crystal Structure of a Parallel-Stranded Guanine Tetraplex at 0.95 Å Resolution. *J. Mol. Biol.* 1997. <https://doi.org/10.1006/jmbi.1997.1292>.
- (60) Šket, P.; Koźmiński, W.; Plavec, J. Is There Any Proton Exchange between Ammoniumions Localized within the d(G 3T 4G 4) 2 Quadruplex? *Acta Chim. Slov.* 2012.
- (61) Erratum to: Small-Molecule Interaction with a Five-Guanine-Tract G-Quadruplex Structure from the Human MYC Promoter. *Nature Chemical Biology.* 2005. <https://doi.org/10.1038/nchembio0905-234b>.
- (62) Todd, A. K.; Haider, S. M.; Parkinson, G. N.; Neidle, S. Sequence Occurrence and Structural Uniqueness of a G-Quadruplex in the Human c-Kit Promoter. *Nucleic Acids Res.* 2007. <https://doi.org/10.1093/nar/gkm609>.
- (63) Wei, D.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. Crystal Structure of a C-Kit Promoter Quadruplex Reveals the Structural Role of Metal Ions and Water Molecules in Maintaining

- Loop Conformation. *Nucleic Acids Res.* 2012. <https://doi.org/10.1093/nar/gks023>.
- (64) Mukundan, V. T.; Phan, A. T. Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.* 2013. <https://doi.org/10.1021/ja310251r>.
- (65) Yang, Q.; Xiang, J.; Yang, S.; Li, Q.; Zhou, Q.; Guan, A.; Zhang, X.; Zhang, H.; Tang, Y.; Xu, G. Verification of Specific G-Quadruplex Structure by Using a Novel Cyanine Dye Supramolecular Assembly: II. The Binding Characterization with Specific Intramolecular G-Quadruplex the Recognizing Mechanism. *Nucleic Acids Res.* 2009. <https://doi.org/10.1093/nar/gkp1045>.
- (66) Huppert, J. L.; Balasubramanian, S. Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Res.* 2005. <https://doi.org/10.1093/nar/gki609>.
- (67) Wong, H. M.; Stegle, O.; Rodgers, S.; Huppert, J. L. A Toolbox for Predicting G-Quadruplex Formation and Stability. *J. Nucleic Acids* 2010. <https://doi.org/10.4061/2010/564946>.
- (68) Todd, A. K.; Johnston, M.; Neidle, S. Highly Prevalent Putative Quadruplex Sequence Motifs in Human DNA. *Nucleic Acids Res.* 2005. <https://doi.org/10.1093/nar/gki553>.
- (69) Bourdoncle, A.; Torres, A. E.; Gosse, C.; Lacroix, L.; Vekhoff, P.; Le Saux, T.; Jullien, L.; Mergny, J. L. Quadruplex-Based Molecular Beacons as Tunable DNA Probes. *J. Am. Chem. Soc.* 2006. <https://doi.org/10.1021/ja0608040>.
- (70) Huppert, J. L.; Balasubramanian, S. G-Quadruplexes in Promoters throughout the Human Genome. *Nucleic Acids Res.* 2007. <https://doi.org/10.1093/nar/gkl1057>.
- (71) Eddy, J.; Maizels, N. Gene Function Correlates with Potential for G4 DNA Formation in the Human Genome. *Nucleic Acids Res.* 2006. <https://doi.org/10.1093/nar/gkl529>.
- (72) Hershman, S. G.; Chen, Q.; Lee, J. Y.; Kozak, M. L.; Yue, P.; Wang, L. S.; Johnson, F. B. Genomic Distribution and Functional Analyses of Potential G-Quadruplex-Forming Sequences in *Saccharomyces Cerevisiae*. *Nucleic Acids Res.* 2008. <https://doi.org/10.1093/nar/gkm986>.
- (73) Capra, J. A.; Paeschke, K.; Singh, M.; Zakian, V. A. G-Quadruplex DNA Sequences Are Evolutionarily Conserved and Associated with Distinct Genomic Features in *Saccharomyces Cerevisiae*. *PLoS Comput. Biol.* 2010. <https://doi.org/10.1371/journal.pcbi.1000861>.
- (74) Stegle, O.; Payet, L.; Mergny, J. L.; MacKay, D. J. C.; Huppert, J. L. Predicting and Understanding the Stability of G-Quadruplexes. In *Bioinformatics*; 2009. <https://doi.org/10.1093/bioinformatics/btp210>.
- (75) Phan, A. T. Human Telomeric DNA: G-Quadruplex, i-Motif and Watson-Crick Double Helix. *Nucleic Acids Res.* 2002. <https://doi.org/10.1093/nar/gkf597>.
- (76) Huber, M. D.; Duquette, M. L.; Shiels, J. C.; Maizels, N. A Conserved G4 DNA Binding Domain in RecQ Family Helicases. *J. Mol. Biol.* 2006. <https://doi.org/10.1016/j.jmb.2006.01.077>.
- (77) Marchand, C.; Pourquier, P.; Laco, G. S.; Jing, N.; Pommier, Y. Interaction of Human Nuclear Topoisomerase I with Guanosine Quartet-Forming and Guanosine-Rich Single-Stranded DNA and RNA Oligonucleotides. *J. Biol. Chem.* 2002. <https://doi.org/10.1074/jbc.M106372200>.

- (78) Arimondo, P. B. Interaction of Human DNA Topoisomerase I with G-Quartet Structures. *Nucleic Acids Res.* 2000. <https://doi.org/10.1093/nar/28.24.4832>.
- (79) Chung, I. K.; Mehta, V. B.; Spitzner, J. R.; Muller, M. T. Eukaryotic Topoisomerase II Cleavage of Parallel Stranded DNA Tetraplexes. *Nucleic Acids Res.* 1992. <https://doi.org/10.1093/nar/20.8.1973>.
- (80) Amrane, S.; Kerkour, A.; Bedrat, A.; Vialet, B.; Andreola, M. L.; Mergny, J. L. Topology of a DNA G-Quadruplex Structure Formed in the HIV-1 Promoter: A Potential Target for Anti-HIV Drug Development. *J. Am. Chem. Soc.* 2014. <https://doi.org/10.1021/ja501500c>.
- (81) 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. <https://doi.org/10.1021/bi048242p>.
- (82) Phan, A. T.; Modi, Y. S.; Patel, D. J. Propeller-Type Parallel-Stranded G-Quadruplexes in the Human c-Myc Promoter. *J. Am. Chem. Soc.* 2004. <https://doi.org/10.1021/ja048805k>.
- (83) Dai, J.; Dexheimer, T. S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R. A.; Yang, D. An Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel G-Strands Formed in the Human BCL-2 Promoter Region in Solution. *J. Am. Chem. Soc.* 2006. <https://doi.org/10.1021/ja055636a>.
- (84) 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. <https://doi.org/10.1093/nar/gkl610>.
- (85) Cogoi, S.; Xodo, L. E. G-Quadruplex Formation within the Promoter of the KRAS Proto-Oncogene and Its Effect on Transcription. *Nucleic Acids Res.* 2006. <https://doi.org/10.1093/nar/gkl286>.
- (86) Sun, D.; Guo, K.; Rusche, J. J.; Hurley, L. H. Facilitation of a Structural Transition in the Polypurine/Polypyrimidine Tract within the Proximal Promoter Region of the Human VEGF Gene by the Presence of Potassium and G-Quadruplex-Interactive Agents. *Nucleic Acids Res.* 2005. <https://doi.org/10.1093/nar/gki917>.
- (87) Huppert, J. L. Structure, Location and Interactions of G-Quadruplexes. *FEBS Journal.* 2010. <https://doi.org/10.1111/j.1742-4658.2010.07758.x>.
- (88) Mathad, R. I.; Hatzakis, E.; Dai, J.; Yang, D. C-MYC Promoter G-Quadruplex Formed at the 5'-End of NHE III 1 Element: Insights into Biological Relevance and Parallel-Stranded G-Quadruplex Stability. *Nucleic Acids Res.* 2011. <https://doi.org/10.1093/nar/gkr612>.
- (89) Phan, A. T.; Kuryavyi, V.; Burge, S.; Neidle, S.; Patel, D. J. Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-Kit Promoter. *J. Am. Chem. Soc.* 2007. <https://doi.org/10.1021/ja068739h>.
- (90) Kuryavyi, V.; Phan, A. T.; Patel, D. J. Solution Structures of All Parallel-Stranded Monomeric and Dimeric G-Quadruplex Scaffolds of the Human c-Kit2 Promoter. *Nucleic Acids Res.* 2010. <https://doi.org/10.1093/nar/gkq558>.
- (91) 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. <https://doi.org/10.1021/ja904007p>.

- (92) Brooks, T. A.; Kendrick, S.; Hurley, L. Making Sense of G-Quadruplex and i-Motif Functions in Oncogene Promoters. *FEBS Journal*. 2010. <https://doi.org/10.1111/j.1742-4658.2010.07759.x>.
- (93) Jeffreys, A. J.; Barber, R.; Bois, P.; Buard, J.; Dubrova, Y. E.; Grant, G.; Hollies, C. R. H.; May, C. A.; Neumann, R.; Panayi, M.; et al. Human Minisatellites, Repeat DNA Instability and Meiotic Recombination. In *Electrophoresis*; 1999. [https://doi.org/10.1002/\(SICI\)1522-2683\(19990101\)20:8<1665::AID-ELPS1665>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1522-2683(19990101)20:8<1665::AID-ELPS1665>3.0.CO;2-L).
- (94) Singh, L. Biological Significance of Minisatellites. *Electrophoresis* 1995. <https://doi.org/10.1002/elps.11501601262>.
- (95) Amrane, S.; Adrian, M.; Heddi, B.; Serero, A.; Nicolas, A.; Mergny, J. L.; Phan, A. T. Formation of Pearl-Necklace Monomorphic G-Quadruplexes in the Human CEB25 Minisatellite. *J. Am. Chem. Soc.* 2012. <https://doi.org/10.1021/ja208993r>.
- (96) Adrian, M.; Ang, D. J.; Lech, C. J.; Heddi, B.; Nicolas, A.; Phan, A. T. Structure and Conformational Dynamics of a Stacked Dimeric G-Quadruplex Formed by the Human CEB1 Minisatellite. *J. Am. Chem. Soc.* 2014. <https://doi.org/10.1021/ja4125274>.
- (97) Sen, D.; Gilbert, W. Formation of Parallel Four-Stranded Complexes by Guanine-Rich Motifs in DNA and Its Implications for Meiosis. *Nature* 1988, 334 (6180), 364–366. <https://doi.org/10.1038/334364a0>.
- (98) Cheong, C.; Moore, P. B. Solution Structure of an Unusually Stable RNA Tetraplex Containing G-and U-Quartet Structures. *Biochemistry* 1992. <https://doi.org/10.1021/bi00151a003>.
- (99) Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S. An RNA G-Quadruplex in the 5' UTR of the NRAS Proto-Oncogene Modulates Translation. *Nat. Chem. Biol.* 2007. <https://doi.org/10.1038/nchembio864>.
- (100) Saccà, B.; Lacroix, L.; Mergny, J. L. The Effect of Chemical Modifications on the Thermal Stability of Different G-Quadruplex-Forming Oligonucleotides. *Nucleic Acids Res.* 2005. <https://doi.org/10.1093/nar/gki257>.
- (101) Darnell, J. C.; Jensen, K. B.; Jin, P.; Brown, V.; Warren, S. T.; Darnell, R. B. Fragile X Mental Retardation Protein Targets G Quartet MRNAs Important for Neuronal Function. *Cell* 2001. [https://doi.org/10.1016/S0092-8674\(01\)00566-9](https://doi.org/10.1016/S0092-8674(01)00566-9).
- (102) Brown, V.; Jin, P.; Ceman, S.; Darnell, J. C.; O'Donnell, W. T.; Tenenbaum, S. A.; Jin, X.; Feng, Y.; Wilkinson, K. D.; Keene, J. D.; et al. Microarray Identification of FMRP-Associated Brain MRNAs and Altered mRNA Translational Profiles in Fragile X Syndrome. *Cell* 2001. [https://doi.org/10.1016/S0092-8674\(01\)00568-2](https://doi.org/10.1016/S0092-8674(01)00568-2).
- (103) Kikin, O.; Zappala, Z.; D'Antonio, L.; Bagga, P. S. GRSDB2 and GRS_UTRdb: Databases of Quadruplex Forming G-Rich Sequences in Pre-MRNAs and MRNAs. *Nucleic Acids Res.* 2008. <https://doi.org/10.1093/nar/gkm982>.
- (104) Huppert, J. L.; Bugaut, A.; Kumari, S.; Balasubramanian, S. G-Quadruplexes: The Beginning and End of UTRs. *Nucleic Acids Res.* 2008. <https://doi.org/10.1093/nar/gkn511>.
- (105) 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. <https://doi.org/10.1021/ja806592z>.

- (106) 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. <https://doi.org/10.1093/nar/gkq259>.
- (107) Luke, B.; Lingner, J. TERRA: Telomeric Repeat-Containing RNA. *EMBO Journal.* 2009. <https://doi.org/10.1038/emboj.2009.166>.
- (108) O'Sullivan, R. J.; Karlseder, J. Telomeres: Protecting Chromosomes against Genome Instability. *Nature Reviews Molecular Cell Biology.* 2010. <https://doi.org/10.1038/nrm2848>.
- (109) Olovnikov, A. M. A Theory of Marginotomy. The Incomplete Copying of Template Margin in Enzymic Synthesis of Polynucleotides and Biological Significance of the Phenomenon. *J. Theor. Biol.* 1973. [https://doi.org/10.1016/0022-5193\(73\)90198-7](https://doi.org/10.1016/0022-5193(73)90198-7).
- (110) Harley, C. B.; Futcher, A. B.; Greider, C. W. Telomeres Shorten during Ageing of Human Fibroblasts. *Nature* 1990. <https://doi.org/10.1038/345458a0>.
- (111) Pandita, T. K.; Hunt, C. R.; Sharma, G. G.; Yang, Q. Regulation of Telomere Movement by Telomere Chromatin Structure. *Cellular and Molecular Life Sciences.* 2007. <https://doi.org/10.1007/s00018-006-6465-0>.
- (112) Tham, W. H.; Zakian, V. A. Transcriptional Silencing at Saccharomyces Telomeres: Implications for Other Organisms. *Oncogene.* 2002. <https://doi.org/10.1038/sj/onc/1205078>.
- (113) Autexier, C.; Lue, N. F. The Structure and Function of Telomerase Reverse Transcriptase. *Annu. Rev. Biochem.* 2006. <https://doi.org/10.1146/annurev.biochem.75.103004.142412>.
- (114) Blackburn, E. H.; Gall, J. G. A Tandemly Repeated Sequence at the Termini of the Extrachromosomal Ribosomal RNA Genes in Tetrahymena. *J. Mol. Biol.* 1978. [https://doi.org/10.1016/0022-2836\(78\)90294-2](https://doi.org/10.1016/0022-2836(78)90294-2).
- (115) Jacob, N. K.; Skopp, R.; Price, C. M. G-Overhang Dynamics at Tetrahymena Telomeres. *EMBO J.* 2001. <https://doi.org/10.1093/emboj/20.15.4299>.
- (116) McElligott, R.; Wellinger, R. J. The Terminal DNA Structure of Mammalian Chromosomes. *EMBO J.* 1997. <https://doi.org/10.1093/emboj/16.12.3705>.
- (117) Chai, W.; Shay, J. W.; Wright, W. E. Human Telomeres Maintain Their Overhang Length at Senescence. *Mol. Cell. Biol.* 2005. <https://doi.org/10.1128/mcb.25.6.2158-2168.2005>.
- (118) Wellinger, R. J.; Wolf, A. J.; Zakian, V. A. Saccharomyces Telomeres Acquire Single-Strand TG1-3 Tails Late in S Phase. *Cell* 1993. [https://doi.org/10.1016/0092-8674\(93\)90049-V](https://doi.org/10.1016/0092-8674(93)90049-V).
- (119) Schultze, P.; Smith, F. W.; Feigon, J. Refined Solution Structure of the Dimeric Quadruplex Formed from the Oxytricha Telomeric Oligonucleotide d(GGGGTTTTGGGG). *Structure* 1994. [https://doi.org/10.1016/S0969-2126\(00\)00023-X](https://doi.org/10.1016/S0969-2126(00)00023-X).
- (120) Wang, Y.; Patel, D. J. Solution Structure of a Parallel-Stranded G-Quadruplex DNA. *J. Mol. Biol.* 1993. <https://doi.org/10.1006/jmbi.1993.1668>.
- (121) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. Human Telomeric Sequence Forms a Hybrid-Type Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel Strands in Potassium Solution. *Nucleic Acids Res.* 2006.

<https://doi.org/10.1093/nar/gkl348>.

- (122) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. Structure of the Human Telomere in K⁺ Solution: An Intramolecular (3 + 1) G-Quadruplex Scaffold. *J. Am. Chem. Soc.* 2006. <https://doi.org/10.1021/ja062791w>.
- (123) Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. Structure of the Hybrid-2 Type Intramolecular Human Telomeric G-Quadruplex in K⁺ Solution: Insights into Structure Polymorphism of the Human Telomeric Sequence. *Nucleic Acids Res.* 2007. <https://doi.org/10.1093/nar/gkm522>.
- (124) Moyzis, R. K.; Buckingham, J. M.; Cram, L. S.; Dani, M.; Deaven, L. L.; Jones, M. D.; Meyne, J.; Ratliff, R. L.; Wu, J. R. A Highly Conserved Repetitive DNA Sequence, (TTAGGG)(n), Present at the Telomeres of Human Chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* 1988. <https://doi.org/10.1073/pnas.85.18.6622>.
- (125) Wright, W. E.; Tesmer, V. M.; Huffman, K. E.; Levene, S. D.; Shay, J. W. Normal Human Chromosomes Have Long G-Rich Telomeric Overhangs at One End. *Genes Dev.* 1997. <https://doi.org/10.1101/gad.11.21.2801>.
- (126) Blasco, M. A. Telomeres and Human Disease: Ageing, Cancer and Beyond. *Nature Reviews Genetics.* 2005. <https://doi.org/10.1038/nrg1656>.
- (127) Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L. C.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W. Specific Association of Human Telomerase Activity with Immortal Cells and Cancer. *Science (80-)*. 1994. <https://doi.org/10.1126/science.7605428>.
- (128) Palm, W.; de Lange, T. How Shelterin Protects Mammalian Telomeres. *Annu. Rev. Genet.* 2008. <https://doi.org/10.1146/annurev.genet.41.110306.130350>.
- (129) Sun, D.; Thompson, B.; Cathers, B. E.; Salazar, M.; Kerwin, S. M.; Trent, J. O.; Jenkins, T. C.; Neidle, S.; Hurley, L. H. Inhibition of Human Telomerase by a G-Quadruplex-Interactive Compound. *Journal of Medicinal Chemistry.* 1997. <https://doi.org/10.1021/jm970199z>.
- (130) Wang, Q.; Liu, J. Q.; Chen, Z.; Zheng, K. W.; Chen, C. Y.; Hao, Y. H.; Tan, Z. G-Quadruplex Formation at the 3' End of Telomere DNA Inhibits Its Extension by Telomerase, Polymerase and Unwinding by Helicase. *Nucleic Acids Res.* 2011. <https://doi.org/10.1093/nar/gkr164>.
- (131) Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M. Inhibition of Telomerase by G-Quartet DNA Structures. *Nature* 1991. <https://doi.org/10.1038/350718a0>.
- (132) Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. Telomere End-Binding Proteins Control the Formation of G-Quadruplex DNA Structures in Vivo. *Nat. Struct. Mol. Biol.* 2005. <https://doi.org/10.1038/nsmb982>.
- (133) 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. <https://doi.org/10.1038/nchem.1805>.
- (134) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* 2013. <https://doi.org/10.1038/nchem.1548>.

- (135) Griffith, J. D.; Comeau, L.; Rosenfield, S.; Stansel, R. M.; Bianchi, A.; Moss, H.; De Lange, T. Mammalian Telomeres End in a Large Duplex Loop. *Cell* 1999. [https://doi.org/10.1016/S0092-8674\(00\)80760-6](https://doi.org/10.1016/S0092-8674(00)80760-6).
- (136) Stansel, R. M.; De Lange, T.; Griffith, J. D. T-Loop Assembly in Vitro Involves Binding of TRF2 near the 3' Telomeric Overhang. *EMBO J.* 2001. <https://doi.org/10.1093/emboj/20.19.5532>.
- (137) Zaug, A. J.; Podell, E. R.; Cech, T. R. Human POT1 Disrupts Telomeric G-Quadruplexes Allowing Telomerase Extension in Vitro. *Proc. Natl. Acad. Sci. U. S. A.* 2005. <https://doi.org/10.1073/pnas.0504744102>.
- (138) Possemato, R.; Timmons, J. C.; Bauerlein, E. L.; Wada, N.; Baldwin, A.; Masutomi, K.; Hahn, W. C. Suppression of HPOT1 in Diploid Human Cells Results in an HERT-Dependent Alteration of Telomere Length Dynamics. *Mol. Cancer Res.* 2008. <https://doi.org/10.1158/1541-7786.MCR-08-0070>.
- (139) Wang, Y.; Patel, D. J. Solution Structure of the Human Telomeric Repeat d[AG3(T2AG3)3] G-Tetraplex. *Structure* 1993, 1 (4), 263–282. [https://doi.org/10.1016/0969-2126\(93\)90015-9](https://doi.org/10.1016/0969-2126(93)90015-9).
- (140) 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. <https://doi.org/10.1093/nar/gkm706>.
- (141) Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. Structure of the Intramolecular Human Telomeric G-Quadruplex in Potassium Solution: A Novel Adenine Triple Formation. *Nucleic Acids Res.* 2007. <https://doi.org/10.1093/nar/gkm009>.
- (142) Matsugami, A.; Xu, Y.; Noguchi, Y.; Sugiyama, H.; Katahira, M. Structure of a Human Telomeric DNA Sequence Stabilized by 8-Bromoguanosine Substitutions, as Determined by NMR in a K⁺ Solution. *FEBS J.* 2007. <https://doi.org/10.1111/j.1742-4658.2007.05881.x>.
- (143) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. Structure of the Human Telomere in K⁺ Solution: A Stable Basket-Type G-Quadruplex with Only Two G-Tetrad Layers. *J. Am. Chem. Soc.* 2009. <https://doi.org/10.1021/ja807503g>.
- (144) Heddi, B.; Phan, A. T. Structure of Human Telomeric DNA in Crowded Solution. *J. Am. Chem. Soc.* 2011. <https://doi.org/10.1021/ja200786q>.
- (145) Lim, K. W.; Ng, V. C. M.; Martín-Pintado, N.; Heddi, B.; Phan, A. T. Structure of the Human Telomere in Na⁺ Solution: An Antiparallel (2+2) G-Quadruplex Scaffold Reveals Additional Diversity. *Nucleic Acids Res.* 2013. <https://doi.org/10.1093/nar/gkt771>.
- (146) Thilagavathi, J.; Venkatesh, S.; Dada, R. Telomere Length in Reproduction. *Andrologia.* 2013. <https://doi.org/10.1111/and.12008>.
- (147) Frézal, J.; Bonaïti-Pellié, C. Introduction to Genetic Analysis. In *Principles and Prenatal Growth*; 1978. https://doi.org/10.1007/978-1-4684-0814-0_8.
- (148) Finkel, T.; Serrano, M.; Blasco, M. A. The Common Biology of Cancer and Ageing. *Nature.* 2007. <https://doi.org/10.1038/nature05985>.
- (149) Reed, J. E.; White, A. J. P.; Neidle, S.; Vilar, R. Effect of Metal Coordination on the Interaction

- of Substituted Phenanthroline and Pyridine Ligands with Quadruplex DNA. *Dalt. Trans.* 2009. <https://doi.org/10.1039/b820086f>.
- (150) Williamson, J. R. G-Quartets in Biology: Reprise. *Proceedings of the National Academy of Sciences of the United States of America.* 1993. <https://doi.org/10.1073/pnas.90.8.3124>.
- (151) Bodnar, A. G.; Ouellette, M.; Frolkis, M.; Holt, S. E.; Chiu, C. P.; Morin, G. B.; Harley, C. B.; Shay, J. W.; Lichtsteiner, S.; Wright, W. E. Extension of Life-Span by Introduction of Telomerase into Normal Human Cells. *Science (80-.)*. 1998. <https://doi.org/10.1126/science.279.5349.349>.
- (152) Georgiades, S. N.; Abd Karim, N. H.; Suntharalingam, K.; Vilar, R. Interaction of Metal Complexes with G-Quadruplex DNA. *Angewandte Chemie - International Edition.* 2010. <https://doi.org/10.1002/anie.200906363>.
- (153) Mikami-Terao, Y.; Akiyama, M.; Yuza, Y.; Yanagisawa, T.; Yamada, O.; Kawano, T.; Agawa, M.; Ida, H.; Yamada, H. Antitumor Activity of TMPyP4 Interacting G-Quadruplex in Retinoblastoma Cell Lines. *Exp. Eye Res.* 2009. <https://doi.org/10.1016/j.exer.2009.03.008>.
- (154) González, V.; Hurley, L. H. The C- MYC NHE III 1 : Function and Regulation . *Annu. Rev. Pharmacol. Toxicol.* 2009. <https://doi.org/10.1146/annurev.pharmtox.48.113006.094649>.
- (155) Le, V. H.; Nagesh, N.; Lewis, E. A. Bcl-2 Promoter Sequence G-Quadruplex Interactions with Three Planar and Non-Planar Cationic Porphyrins: TMPyP4, TMPyP3, and TMPyP2. *PLoS One* 2013. <https://doi.org/10.1371/journal.pone.0072462>.
- (156) Chen, Y.; Agrawal, P.; Brown, R. V.; Hatzakis, E.; Hurley, L.; Yang, D. The Major G-Quadruplex Formed in the Human Platelet-Derived Growth Factor Receptor β Promoter Adopts a Novel Broken-Strand Structure in K⁺ Solution. *J. Am. Chem. Soc.* 2012. <https://doi.org/10.1021/ja305764d>.
- (157) Nilsson, J. A.; Cleveland, J. L. Myc Pathways Provoking Cell Suicide and Cancer. *Oncogene.* 2003. <https://doi.org/10.1038/sj.onc.1207261>.
- (158) Nasi, S.; Ciarapica, R.; Jucker, R.; Rosati, J.; Soucek, L. Making Decisions through Myc. *FEBS Letters.* 2001. [https://doi.org/10.1016/S0014-5793\(01\)02118-4](https://doi.org/10.1016/S0014-5793(01)02118-4).
- (159) Sun, D.; Hurley, L. H. The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression. *J. Med. Chem.* 2009. <https://doi.org/10.1021/jm900055s>.
- (160) Liu, L. F.; Wang, J. C. Supercoiling of the DNA Template during Transcription. *Proc. Natl. Acad. Sci. U. S. A.* 1987. <https://doi.org/10.1073/pnas.84.20.7024>.
- (161) Tanaka, A.; Choi, J.; Kim, S. K.; Majima, T. Interaction of G-Quadruplex with RecA Protein Studied in Bulk Phase and at the Single-Molecule Level. *J. Phys. Chem. B* 2013. <https://doi.org/10.1021/jp4036277>.
- (162) Zavyalova, E.; Tagiltsev, G.; Reshetnikov, R.; Arutyunyan, A.; Kopylov, A. Cation Coordination Alters the Conformation of a Thrombin-Binding G-Quadruplex DNA Aptamer That Affects Inhibition of Thrombin. *Nucleic Acid Ther.* 2016. <https://doi.org/10.1089/nat.2016.0606>.

- (163) Nagatoishi, S.; Isono, N.; Tsumoto, K.; Sugimoto, N. Hydration Is Required in DNA G-Quadruplex-Protein Binding. *ChemBioChem* 2011. <https://doi.org/10.1002/cbic.201100264>.
- (164) Zhao, J.; Xin, M.; Cao, Y.; Yin, Y.; Shu, Y.; Ma, W. An Electrochemical Aptasensor for Thrombin Detection Based on the Recycling of Exonuclease III and Double-Stranded DNA-Templated Copper Nanoparticles Assisted Signal Amplification. *Anal. Chim. Acta* 2015. <https://doi.org/10.1016/j.aca.2014.12.026>.
- (165) Golub, E.; Freeman, R.; Willner, I. A Hemin/G-Quadruplex Acts as an NADH Oxidase and NADH Peroxidase Mimicking DNAzyme. *Angew. Chemie - Int. Ed.* 2011. <https://doi.org/10.1002/anie.201103853>.
- (166) Jing, P.; Xu, W.; Yi, H.; Wu, Y.; Bai, L.; Yuan, R. An Amplified Electrochemical Aptasensor for Thrombin Detection Based on Pseudobiozymic Fe₃O₄-Au Nanocomposites and Electroactive Hemin/G-Quadruplex as Signal Enhancers. *Analyst* 2014. <https://doi.org/10.1039/c3an02237d>.
- (167) Loo, A. H.; Bonanni, A.; Pumera, M. Impedimetric Thrombin Aptasensor Based on Chemically Modified Graphenes. *Nanoscale* 2012. <https://doi.org/10.1039/c1nr10966a>.
- (168) Pennarun, G.; Granotier, C.; Gauthier, L. R.; Gomez, D.; Hoffschir, F.; Mandine, E.; Riou, J. F.; Mergny, J. L.; Mailliet, P.; Boussin, F. D. Apoptosis Related to Telomere Instability and Cell Cycle Alterations in Human Glioma Cells Treated by New Highly Selective G-Quadruplex Ligands. *Oncogene* 2005. <https://doi.org/10.1038/sj.onc.1208468>.
- (169) Granotier, C.; Pennarun, G.; Riou, L.; Hoffschir, F.; Gauthier, L. R.; De Cian, A.; Gomez, D.; Mandine, E.; Riou, J. F.; Mergny, J. L.; et al. Preferential Binding of a G-Quadruplex Ligand to Human Chromosome Ends. *Nucleic Acids Res.* 2005. <https://doi.org/10.1093/nar/gki722>.
- (170) Lehn, J.-M.; Guittat, L.; Garestier, T.; Helene, C.; Lacroix, L.; Hounsou, C.; Vigneron, J.-P.; Hoarau, M.; Mergny, J.-L.; Arimondo, P. B.; et al. Telomerase Inhibitors Based on Quadruplex Ligands Selected by a Fluorescence Assay. *Proc. Natl. Acad. Sci.* 2002. <https://doi.org/10.1073/pnas.051620698>.
- (171) De Cian, A.; DeLemos, E.; Mergny, J. L.; Teulade-Fichou, M. P.; Monchaud, D. Highly Efficient G-Quadruplex Recognition by Bisquinolinium Compounds. *J. Am. Chem. Soc.* 2007. <https://doi.org/10.1021/ja067352b>.
- (172) Minhas, G. S.; Pilch, D. S.; Kerrigan, J. E.; LaVoie, E. J.; Rice, J. E. Synthesis and G-Quadruplex Stabilizing Properties of a Series of Oxazole-Containing Macrocycles. *Bioorganic Med. Chem. Lett.* 2006. <https://doi.org/10.1016/j.bmcl.2006.05.038>.
- (173) Neidle, S. Human Telomeric G-Quadruplex: The Current Status of Telomeric G-Quadruplexes as Therapeutic Targets in Human Cancer. *FEBS Journal.* 2010. <https://doi.org/10.1111/j.1742-4658.2009.07463.x>.
- (174) Monchaud, D.; Teulade-Fichou, M. P. A Hitchhiker's Guide to G-Quadruplex Ligands. *Org. Biomol. Chem.* 2008. <https://doi.org/10.1039/b714772b>.
- (175) Artese, A.; Costa, G.; Ortuso, F.; Parrotta, L.; Alcaro, S. Identification of New Natural DNA G-Quadruplex Binders Selected by a Structure-Based Virtual Screening Approach. *Molecules* 2013. <https://doi.org/10.3390/molecules181012051>.

- (176) Gavathiotis, E.; Heald, R. A.; Stevens, M. F. G.; Searle, M. S. Drug Recognition and Stabilisation of the Parallel-Stranded DNA Quadruplex d(TTAGGGT)₄ Containing the Human Telomeric Repeat. *J. Mol. Biol.* 2003. <https://doi.org/10.1016/j.jmb.2003.09.018>.
- (177) Campbell, N. H.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. Structural Basis of DNA Quadruplex Recognition by an Acridine Drug. *J. Am. Chem. Soc.* 2008. <https://doi.org/10.1021/ja8016973>.
- (178) Mergny, J. L.; Maurizot, J. C. Fluorescence Resonance Energy Transfer as a Probe for G-Quartet Formation by a Telomeric Repeat. *ChemBioChem* 2001. [https://doi.org/10.1002/1439-7633\(20010202\)2:2<124::AID-CBIC124>3.0.CO;2-L](https://doi.org/10.1002/1439-7633(20010202)2:2<124::AID-CBIC124>3.0.CO;2-L).
- (179) 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. <https://doi.org/10.1016/j.ymeth.2006.10.004>.
- (180) Hounsou, C.; Guittat, L.; Monchaud, D.; Jourdan, M.; Saettel, N.; Mergny, J. L.; Teulade-Fichou, M. P. G-Quadruplex Recognition by Quinacridines: A SAR, NMR, and Biological Study. *ChemMedChem* 2007. <https://doi.org/10.1002/cmdc.200600286>.
- (181) Fedoroff, O. Y.; Salazar, M.; Han, H.; Chemeris, V. V.; Kerwin, S. M.; Hurley, L. H. NMR-Based Model of a Telomerase-Inhibiting Compound Bound to G- Quadruplex DNA. *Biochemistry* 1998. <https://doi.org/10.1021/bi981330n>.
- (182) Chung, W. J.; Heddi, B.; Tera, M.; Iida, K.; Nagasawa, K.; Phan, A. T. Solution Structure of an Intramolecular (3 + 1) Human Telomeric G-Quadruplex Bound to a Telomestatin Derivative. *J. Am. Chem. Soc.* 2013. <https://doi.org/10.1021/ja405843r>.
- (183) Dai, J.; Carver, M.; Hurley, L. H.; Yang, D. Solution Structure of a 2:1 Quindoline-c-MYC G-Quadruplex: Insights into G-Quadruplex-Interactive Small Molecule Drug Design. *J. Am. Chem. Soc.* 2011. <https://doi.org/10.1021/ja205646q>.
- (184) Seenisamy, J.; Rezler, E. M.; Powell, T. J.; Tye, D.; Gokhale, V.; Joshi, C. S.; Siddiqui-Jain, A.; Hurley, L. H. The Dynamic Character of the G-Quadruplex Element in the c-MYC Promoter and Modification by TMPyP4. *J. Am. Chem. Soc.* 2004. <https://doi.org/10.1021/ja040022b>.
- (185) Parkinson, G. N.; Ghosh, R.; Neidle, S. Structural Basis for Binding of Porphyrin to Human Telomeres. *Biochemistry* 2007. <https://doi.org/10.1021/bi062244n>.
- (186) Lemarteleur, T.; Gomez, D.; Paterski, R.; Mandine, E.; Mailliet, P.; Riou, J. F. Stabilization of the C-Myc Gene Promoter Quadruplex by Specific Ligands' Inhibitors of Telomerase. *Biochem. Biophys. Res. Commun.* 2004. <https://doi.org/10.1016/j.bbrc.2004.08.150>.
- (187) Chung, W. J.; Heddi, B.; Hamon, F.; Teulade-Fichou, M. P.; Phan, A. T. Solution Structure of a G-Quadruplex Bound to the Bisquinolinium Compound Phen-DC3. *Angew. Chemie - Int. Ed.* 2014. <https://doi.org/10.1002/anie.201308063>.
- (188) Guo, K.; Pourpak, A.; Beetz-Rogers, K.; Gokhale, V.; Sun, D.; Hurley, L. H. Formation of Pseudosymmetrical G-Quadruplex and i-Motif Structures in the Proximal Promoter Region of the RET Oncogene. *J. Am. Chem. Soc.* 2007. <https://doi.org/10.1021/ja072185g>.
- (189) Shin-ya, K.; Wierzba, K.; Matsuo, K.; Ohtani, T.; Yamada, Y.; Furihata, K.; Hayakawa, Y.; Seto, H. Telomestatin, a Novel Telomerase Inhibitor from *Streptomyces Anulatus* [17]. *Journal of the American Chemical Society*. 2001. <https://doi.org/10.1021/ja005780q>.

- (190) Monchaud, D.; Yang, P.; Lacroix, L.; Teulade-Fichou, M. P.; Mergny, J. L. A Metal-Mediated Conformational Switch Controls G-Quadruplex Binding Affinity. *Angew. Chemie - Int. Ed.* 2008. <https://doi.org/10.1002/anie.200800468>.
- (191) Reed, J. E.; Arnal, A. A.; Neidle, S.; Vilar, R. Stabilization of G-Quadruplex DNA and Inhibition of Telomerase Activity by Square-Planar Nickel(II) Complexes. *J. Am. Chem. Soc.* 2006. <https://doi.org/10.1021/ja058509n>.
- (192) Dixon, I. M.; Lopez, F.; Tejera, A. M.; Estève, J. P.; Blasco, M. A.; Pratviel, G.; Meunier, B. A G-Quadruplex Ligand with 10000-Fold Selectivity over Duplex DNA. *J. Am. Chem. Soc.* 2007. <https://doi.org/10.1021/ja065591t>.
- (193) Martino, L.; Virno, A.; Pagano, B.; Virgilio, A.; Di Micco, S.; Galeone, A.; Giancola, C.; Bifulco, G.; Mayol, L.; Randazzo, A. Structural and Thermodynamic Studies of the Interaction of Distamycin A with the Parallel Quadruplex Structure [d(TGGGGT)]₄. *J. Am. Chem. Soc.* 2007. <https://doi.org/10.1021/ja075710k>.
- (194) Cocco, M. J.; Hanakahi, L. A.; Huber, M. D.; Maizels, N. Specific Interactions of Distamycin with G-Quadruplex DNA. *Nucleic Acids Res.* 2003. <https://doi.org/10.1093/nar/gkg392>.
- (195) Gai, W.; Yang, Q.; Xiang, J.; Jiang, W.; Li, Q.; Sun, H.; Guan, A.; Shang, Q.; Zhang, H.; Tang, Y. A Dual-Site Simultaneous Binding Mode in the Interaction between Parallel-Stranded G-Quadruplex [d(TGGGGT)]₄ and Cyanine Dye 2,20-Diethyl-9- Methylselenocarbocyanine Bromide. *Nucleic Acids Res.* 2013. <https://doi.org/10.1093/nar/gks1328>.
- (196) Watson, J. D.; Crick, F. H. C. THE STRUCTURE OF DNA References [Http://Symposium.Cshlp.Org/Content/18/123#related-Urls](http://Symposium.Cshlp.Org/Content/18/123#related-Urls) THE STRUCTURE OF DNA. *Cold Spring Harb Symp Quant Biol* 1953. <https://doi.org/10.1101/SQB.1953.018.01.020>.
- (197) Watson, J. D.; Crick, F. H. C. DNA - Informacao Importante (Confiavel). *Nature* 1953.
- (198) Gates, K. S. *Nucleic Acids in Chemistry and Biology* . Edited by C. Michael Blackburn and Michael J. Gait. Oxford and New York: Oxford University Press, 1996, Pp. Xix + 528. Price \$49.95. ISBN 0-19-963533-1. . *Acta Crystallogr. Sect. D Biol. Crystallogr.* 1999. <https://doi.org/10.1107/s0907444998011652>.
- (199) Rich, A.; Zhang, S. Z-DNA: The Long Road to Biological Function. *Nature Reviews Genetics.* 2003. <https://doi.org/10.1038/nrg1115>.
- (200) Zlatanova, J.; Van Holde, K. Binding to Four-Way Junction DNA: A Common Property of Architectural Proteins? *FASEB J.* 1998. <https://doi.org/10.1096/fasebj.12.6.421>.
- (201) Gottipati, P.; Cassel, T. N.; Savolainen, L.; Helleday, T. Transcription-Associated Recombination Is Dependent on Replication in Mammalian Cells. *Mol. Cell. Biol.* 2008. <https://doi.org/10.1128/mcb.00816-07>.
- (202) Micco, M.; Collie, G. W.; Dale, A. G.; Ohnmacht, S. A.; Pazitna, I.; Gunaratnam, M.; Reszka, A. P.; Neidle, S. Structure-Based Design and Evaluation of Naphthalene Diimide G-Quadruplex Ligands as Telomere Targeting Agents in Pancreatic Cancer Cells. *J. Med. Chem.* 2013. <https://doi.org/10.1021/jm301899y>.
- (203) Vuong, S.; Stefan, L.; Lejault, P.; Rousselin, Y.; Denat, F.; Monchaud, D. Identifying Three-Way DNA Junction-Specific Small-Molecules. *Biochimie* 2012. <https://doi.org/10.1016/j.biochi.2011.08.012>.

- (204) Sinden, R. R.; Potaman, V. N.; Oussatcheva, E. A.; Pearson, C. E.; Lyubchenko, Y. L.; Shlyakhtenko, L. S. Triplet Repeat DNA Structures and Human Genetic Disease: Dynamic Mutations from Dynamic DNA. *J. Biosci.* 2002. <https://doi.org/10.1007/BF02703683>.
- (205) Malina, J.; Hannon, M. J.; Brabec, V. Recognition of DNA Three-Way Junctions by Metallo-supramolecular Cylinders: Gel Electrophoresis Studies. *Chem. - A Eur. J.* 2007. <https://doi.org/10.1002/chem.200700159>.
- (206) Khuu, P. A.; Voth, A. R.; Hays, F. A.; Ho, P. S. The Stacked-X DNA Holliday Junction and Protein Recognition. *Journal of Molecular Recognition.* 2006. <https://doi.org/10.1002/jmr.765>.
- (207) Suntharalingam, K.; White, A. J. P.; Vilar, R. Synthesis, Structural Characterization, and Quadruplex DNA Binding Studies of Platinum(II)-Terpyridine Complexes. *Inorg. Chem.* 2009. <https://doi.org/10.1021/ic901319n>.
- (208) Reed, J. E.; Neidle, S.; Vilar, R. Stabilisation of Human Telomeric Quadruplex DNA and Inhibition of Telomerase by a Platinum-Phenanthroline Complex. *Chem. Commun.* 2007. <https://doi.org/10.1039/b709898g>.
- (209) Gu, J.; Leszczynski, J. Origin of Na⁺/K⁺ Selectivity of the Guanine Tetraplexes in Water: The Theoretical Rationale. *J. Phys. Chem. A* 2002. <https://doi.org/10.1021/jp012739g>.
- (210) Ernst, R. J.; Song, H.; Barton, J. K. DNA Mismatch Binding and Antiproliferative Activity of Rhodium Metalloinsertors. *J. Am. Chem. Soc.* 2009. <https://doi.org/10.1021/ja8081044>.
- (211) Vummidi, B. R.; Alzeer, J.; Luedtke, N. W. Fluorescent Probes for G-Quadruplex Structures. *ChemBioChem* 2013. <https://doi.org/10.1002/cbic.201200612>.
- (212) Hays, F. A.; Watson, J.; Ho, P. S. Caution! DNA Crossing: Crystal Structures of Holliday Junctions. *Journal of Biological Chemistry.* 2003. <https://doi.org/10.1074/jbc.R300033200>.
- (213) Egli, M. Nucleic Acid Crystallography: Current Progress. *Current Opinion in Chemical Biology.* 2004. <https://doi.org/10.1016/j.cbpa.2004.09.004>.
- (214) Lourenço, L. M. O.; Iglesias, B. A.; Pereira, P. M. R.; Girão, H.; Fernandes, R.; Neves, M. G. P. M. S.; Cavaleiro, J. A. S.; Tomé, J. P. C. Synthesis, Characterization and Biomolecule-Binding Properties of Novel Tetra-Platinum(II)-Thiopyridylporphyrins. *Dalt. Trans.* 2015. <https://doi.org/10.1039/c4dt02697g>.

Chapter 2

A Fragment-Based Approach for the Development of G-Quadruplex Ligands

Chapter 2 A Fragment-Based Approach for the Development of G-Quadruplex Ligands

2.1 The Drug Discovery Process

Drug discovery is the process through which potential new medicines are discovered. The process from initial hit to marketable drug can often take between 10-15 years and in 2014 the average cost to develop a new drug was estimated at US\$2.6 billion.¹ The two main approaches to discover new drugs are phenotypic drug discovery and target-based drug discovery. The general principle of early stage target-based drug discovery is depicted in figure 30.² Modern methods for the discovery of new drugs involve rational drug design and often begin with initial target identification. A target is usually a gene, protein or nucleic acid, which is connected to a particular disease. Identifying a 'drugable' target is important in order to develop a successful drug. A 'drugable' target is one that can be accessed by the potential drug compound and which produces a biological response upon binding of the drug molecule. The identified target must then be validated and shown to be critical in the disease process. As target, validation confirms that interaction with the target results in the desired effect on the disease this step is necessary to ensure that during the drug development process, there is the potential to access a safe and effective drug and therefore a drug discovery effort can be justified.



Figure 30. General process of the early stages of target-based drug discovery.²

The next step, hit identification, is to discover a molecule that binds to and interacts with the chosen target. A library of compounds is screened against the chosen target and any suitable hits can be grown and developed into larger lead-like compounds. There are a variety of screening techniques that can be employed such as high-throughput screening, fragment based screening or a virtual screen. Once a hit series has been defined, the 'hit to lead' stage of the development process is performed. Compounds in the hit series are optimised to improve potency and selectivity, often achieved through functional group modification. The identified lead compounds are then modified further to improve the physico-chemical and biological properties making the molecules more effective and safe, and therefore viable drug candidates. The optimised compounds are then taken through the drug development process starting with preclinical then clinical testing.²

2.1.1 Desirable Drug-like Properties of a Drug Candidate

The discovery of novel potent drug molecules is a major challenge and as described above is a lengthy and expensive process. In the USA only 1 in 1,000 drug candidates which begin preclinical testing advances through the process to clinical trials and only 10% of the drugs in phase 1 clinical trials will be approved by the Food and Drug Administration (FDA) and advanced onto the market.³ In 1991 adverse bioavailability and pharmacokinetic properties were the main causes of drug attrition within the pharmaceutical industry. However by 2000 the majority of drug candidates failed for reasons including toxicity and efficacy as well as commercial factors, and between 2011 and 2012 the main causes of drug attrition were still safety and efficacy.⁴⁻⁶ In addition to poor toxicity, safety indirectly causes drug failures by causing the compound dosage in humans to be limited, preventing adequate exposure. In 2014, Hay reported that the majority of drugs that failed in clinical trials between 2003 and 2011 did so in phase 3 or during submission to the national regulatory authority.³ In addition, Arrow smith reported that between 2011 and 2012 around 52% of drugs that failed in phase 3 clinical trials did so due to poor efficacy.⁵ For these reasons, the pharmacokinetic and physicochemical properties of a potential drug candidate should be considered alongside biological activity.

2.1.2 Guidelines for Desirable Physicochemical Properties

The link between physicochemical properties and aspects of “drug-likeness” of a molecule was first identified in 1997 by Lipinski who reported guidelines for the physicochemical properties of orally available drugs, known as the “rule of 5”. These guidelines stated that compounds with a molecular weight less than 500 Dalton’s, lipophilicity (LogP) less than 5 and no more than 5 H-bond donors or 10 H-bond acceptors were more likely to be absorbed within the body and therefore become a marketable drug.⁷ Over time these guidelines, intended to predict oral absorption, have been used to define the limits of drug like space, however they do not identify compounds with toxicological risks. As such, modifications of these guidelines have been reported to identify molecules in more general drug-like space. Gleeson reported that molecules with a clogP < 4 and a molecular weight < 400 Daltons (the GSK 4/400 rule) have more favourable ADMET profiles.⁸ ADMET refers to absorption, distribution, metabolism, excretion and toxicity. Compounds with issues relating to these pharmacokinetic properties often require further studies before approval by the national regulatory authority, emphasizing the importance of being able to predict these properties. Similarly Hughes reported the Pfizer 3/75 rule which indicated that compounds with clogP < 3 and polar surface area > 75 Å showed less toxicity and

therefore greater safety in pre-clinical trials.⁹ Polar surface area (PSA) is the sum of the surfaces of polar atoms in a molecule and this measurement has been shown to correlate with molecular transport properties such as intestinal absorption.¹⁰ Calculation of PSA is often time consuming and computationally demanding as the 3-dimensional geometry of the molecule needs to be considered and the surface determined, however in 2000 Ertl developed a method to calculate PSA based on topological information, TPSA.¹⁰ TPSA was based on the summation of surface contributions of the polar fragments within a molecule and allowed for a fast virtual bioavailability screen of a large number of compounds.

2.1.3 The Importance of Controlling Physico-chemical Properties

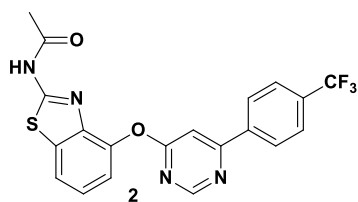
Molecular weight (MW) and lipophilicity are two influential properties on the success of a drug compound.¹¹ Molecular weight is often added during the optimization stage to increase the potency of a drug, however this may result in reduced solubility and permeability. As the drug must pass through a lipid bilayer in the cell membrane an increased lipophilicity is required to retain desirable permeability. Lipophilicity, logP, is defined as the logarithm of the ratio of concentrations of a solute between immiscible phases, often water and *n*-octanol, at equilibrium. In addition to improved permeability an advantage of increased lipophilicity is that binding of the drug compound to the desired target can improve, resulting in increased. To help determine desirable lead compounds and to quantify molecular properties, including MW, a measure known as Ligand Efficiency (LE) was reported. LE estimates binding efficiency of the drug to the target with respect to the number of heavy atoms, therefore correlating potency with molecular weight, equation 2-A.¹² LE does not consider lipophilicity, instead a Lipophilic Ligand Efficiency (LLE) value can be used, equation 2-B. A value of 5 or greater is suggested to lead to a drug candidate with desirable potency and lipophilicity and a reduced risk of toxicity.¹³ These concepts help to identify and control an increase in MW and/or lipophilicity that may become detrimental to the drug compound during optimisation.¹³

In addition, lipophilicity correlates with aqueous solubility, an important property of a drug compound that is required to achieve the desired concentration essential for pharmacological effect. As lipophilicity increases solubility decreases because higher lipophilicity represents less polar, hydrophobic, poorly soluble molecules.¹⁴ The human body metabolises lipophilic molecules to obtain more polar soluble compounds which can be excreted, therefore lipophilic compounds are required to be metabolically robust.¹⁵ As a result a careful balance between potency, lipophilicity, metabolic stability and solubility is required for a successful drug.

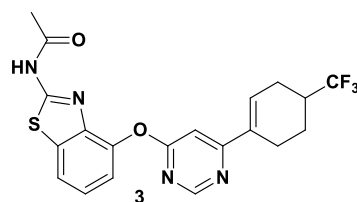
An additional property to consider when designing a drug is complexity. Over the last 5/6 years the desire to synthesize more diverse and architecturally complex molecules has emerged. There is a move away from the many flat sp^2 rich compounds produced in the pharmaceutical industry that could be readily accessed by powerful cross-coupling methods. Complex molecules are likely to be more “natural product like” in structure. As many potent drugs have been derived from natural products, creating more complex molecules was anticipated to increase the chance of finding appropriate bioactive compounds. Diverse, complex molecules will also allow the exploration of new areas of chemical space.¹⁶

Molecular complexity may be described by analyzing saturation and chirality within the molecule. Saturation allows more complex molecules to be created without increasing the molecular weight significantly.¹⁶ Increasing sp^3 character and designing out-of-plane substituents onto the molecule could allow more binding interactions between the drug candidate and target, often not possible with flat molecules, resulting in increased potency and selectivity. Complexity and saturation can be calculated by various methods. Lowering reported using the fraction of saturated carbons within a molecule (F_{sp3}) as a descriptor of saturation and therefore complexity, equation 3.¹⁶

Increasing the number of sp^3 centres

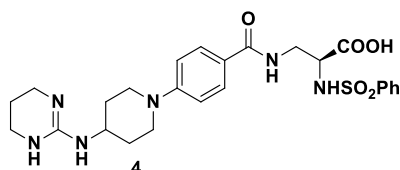


solubility (0.01 M HCl): <1 mg/mL

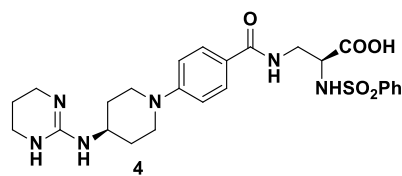


solubility (0.01 M HCl): <13 mg/mL

Increasing Chirality



aq.solubility: <0.1 mg/mL



aq.solubility: < 3.5 mg/mL

Figure 31. The effect of disrupting molecular planarity and symmetry on solubility.¹⁷

2.1.4 The Method of Fragment Based Drug Discovery

In 1981, Jencks reported that although small molecules bind weakly to a target they form high quality interactions.¹⁸ The fragment based methodology was then established by researchers at

Abbott who reported the discovery of high-affinity ligands for proteins by linking together small molecules that bound to the target.¹⁹ Subsequently, scientists at Astex Pharmaceuticals developed a pioneering structure guided fragment based approach to drug discovery using X-ray crystallography. This resulted in significant advances in the field and the fragment based approach to drug discovery emerged as a complementary method to high-throughput screening.²⁰

An important milestone for FBDD was reached in 2011 when the FDA approved PLX4032, marketed as Zelboraf, as a treatment for late stage melanoma, figure 3.²¹ PLX4032 was one of the first drugs approved that was developed using fragment based methodology. Twenty thousand compounds were screened, 7-azaindole **6** was selected for optimisation, leading to the development of PLX4720.²² Further optimisation by addition of an aryl group led to the development of PLX4032.

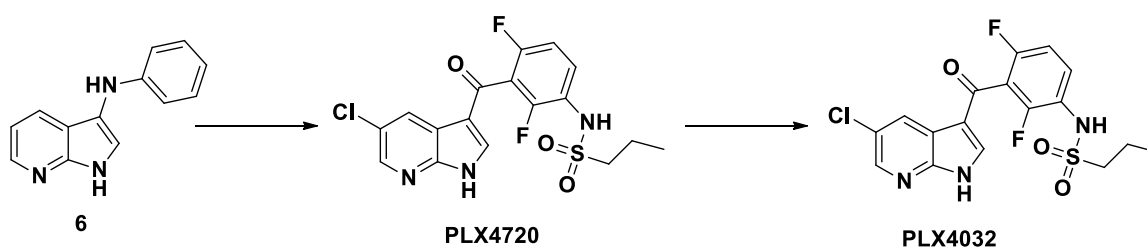


Figure 32. First example of an approved drug which was discovered using fragment based approaches.^{21,22}

FBDD involves fragment library design, fragment screening and fragment optimisation. When designing a fragment library it is important that the fragments have desirable physicochemical properties in order to identify a drug in drug-like space. Similar to the “rule of 5” for drug-like compounds researchers at Astex defined guidelines for desirable criteria for fragments, known as the “rule of 3”.²³

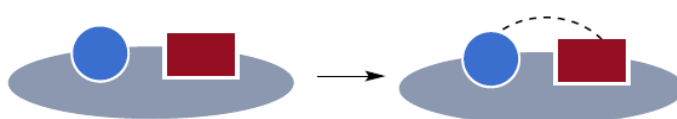
The following properties were suggested:

- MW < 300 daltons
- cLogP < 3
- H-bond donors ≤ 3
- H-bond acceptors ≤ 3

In addition, the number of rotatable bonds should be less than three and a polar surface area of less than 60 would be beneficial.²³ These are only recommended guidelines and in fact Klebe suggested that the number of hydrogen bond acceptors should be raised to six or fewer as this can improve solubility and provide additional binding elements.²⁴

Fragments hits can be optimised into lead-like compounds using methodology such as fragment linking or fragment growing.²⁵ Fragment linking involves connecting two or more low affinity fragments, which bind to the target in adjacent sites, to form a high affinity lead-like compound (Fig. 33). Fragments that bind to similar sites on the target and partially overlap may also be combined to improve fragment potency. In principle fragment linking is an appealing method, however in practice developing an appropriate linker to connect the fragments without altering their binding modes or finding synthetic methodology to access the merged fragment without affecting the binding affinity can be challenging.²⁵

A) Fragment Linking



B) Fragment Growing



Figure 33. An illustration of fragment optimisation strategies.

Fragment growing is a more general approach which involves the step-wise addition of substituents onto the core fragment hit to improve potency, in addition to pharmacological properties, and grow the fragment into a potent lead compound.²⁶ Fragment growing provides more options during optimisation compared to fragment linking as the fragment can be elaborated in a variety of directions.²⁶ Structural information of the target and fragment binding modes is advantageous when using both methods to facilitate the optimisation process.

2.2 Background of fragments used in these studies for Hit to lead optimization strategy

Heterocyclic chemistry is one of the most valuable sources of novel compounds with diverse biological activity, mainly because of the unique ability of the resulting compounds to mimic the structure of peptides and to bind reversibly to proteins.^{27,28} To medicinal chemists, the true utility of heterocyclic structures is the ability to synthesize one library based on one core scaffold and to screen it against a variety of different receptors, yielding several active compounds. Almost unlimited combinations of fused heterocyclic structures can be designed, resulting in novel polycyclic frameworks with the most diverse physical, chemical and biological properties. The fusion of several rings lead to geometrically well-defined rigid polycyclic structures and, thus, holds the promise of a high functional specialization resulting from the ability to orient substituents in three-dimensional space. Therefore, efficient methodologies resulting in polycyclic structures from biologically active heterocyclic templates are always of interest to both organic and medicinal chemists.

2.2.1 Indoles as Desirable Motifs for Use in Drug Discovery

Indole (**1**, Figure 34) is the parent substance of a large number of important compounds that occur in nature. Indole and the simple alkylindoles are colourless crystalline solids with a range of odours from naphthalene-like in the case of indole itself to faecal in the case of skatole (3-methylindole) (**2**) (Figure 1). Tryptophan²⁹ (2-amino-3-(3'-indolyl)propionic acid) (**3**, Figure 34) is one of the naturally occurring essential amino acids. Higher plants degrade tryptophan to heteroauxin (indole-3-acetic acid, **4**), a plant hormone (Fig. 34).

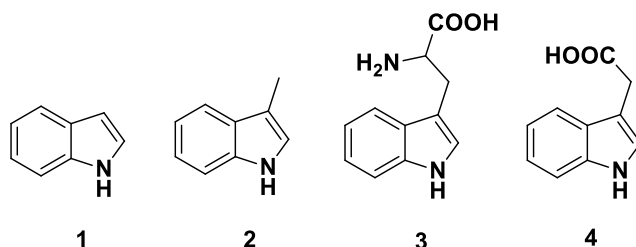


Figure 34. Indole derivatives

Tryptophan is an essential amino acid and as such is a constituent of most proteins; it also serves as a biosynthetic precursor for a wide variety of tryptamine-indole, and 2,3-dihydroindole-containing secondary metabolites. In animals, serotonin (5-hydroxytryptamine) is a very important neurotransmitter in the CNS, and in the cardiovascular and gastrointestinal systems.

The structurally similar hormone melatonin is thought to control the diurnal rhythm of physiological functions.

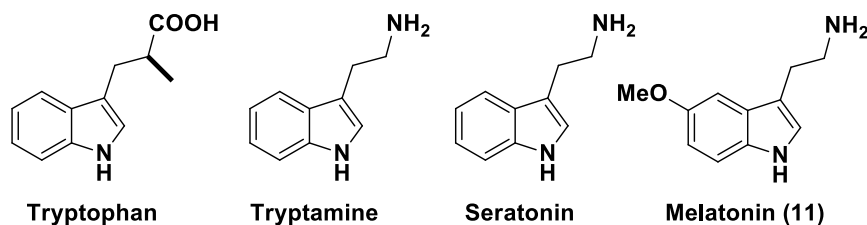


Figure 35. Structures of some naturally occurring indoles

2.2.2 Indoles as a G-Quadruplex Stabilizing agents

Diveshkumar *et al*³⁰ has reported the indolymethyleneindanone based scaffolds to specifically target the *c-MYC* and *c-KIT* G4 DNA. (Fig. 36).

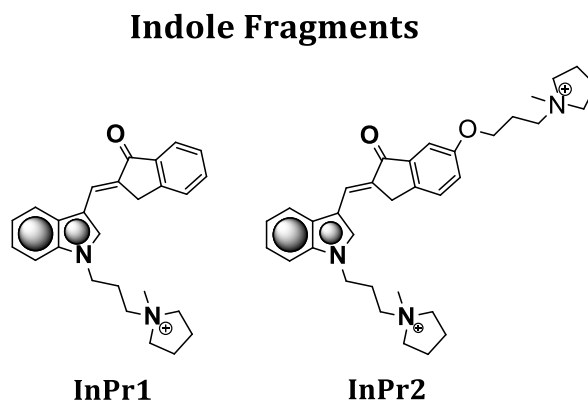


Figure 36. The synthesized indolymethyleneindanone based scaffolds ligands used for the stabilization of *c-MYC* G-Quadruplex

2.2.3 Quindoline Derivatives: Indole based G-Quadruplex Stabilizing agents

Quindoline (1) (Fig. 37) is a naturally occurring indolo [3,2-b]quinoline alkaloid. Quindoline was originally synthesized in 1906 and has since been prepared from cryptolepine. Quindoline was first extracted from the African plant *Cryptolepis sanguinolenta* (Periplocaceae) in 1978. This compound has antimicrobial, antimalarial, anti-inflammatory, and antihypolipidemic activities. In 2000, Neidle *et al.* first reported that the 2,10-disubstituted quindoline derivative 3 could bind to G-quadruplex structures in human telomeres, showing it to have modest cytotoxicity against

several cancer cells as well as inhibitory activity against telomerase.³¹ Quindoline scaffolds are built-up from indole skeleton, which inspire the usage of indole pharmacophore for further screening to optimize the fragments in our current strategy in this chapter.

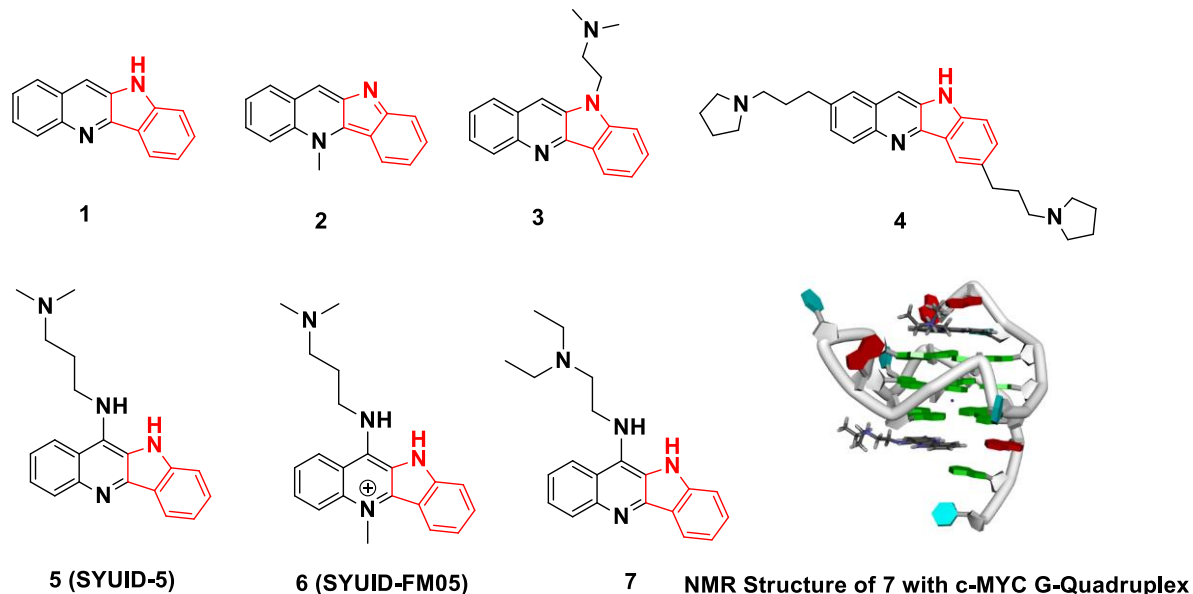


Figure 37. Structures of quindoline derivatives and binding mode with G-quadruplex and NMR structure of 7 with c-MYC promoter G-quadruplex

2.2.4 7- Azaindoles: Chemical and Biological Importance

The widespread usefulness of indoles in life sciences has stimulated the development of numerous methodologies for their synthesis, and a range of well-established classical methods is available. Typical examples include the Fischer indole synthesis, the Gassman synthesis, the Madelung cyclization, the Bischler indole synthesis, the Bartoli synthesis and the Batcho Leimgruber synthesis.³³⁻³⁵ Replacing one of the carbon atoms at positions 4 to 7 in the indole template with a nitrogen atom gives the so-called azaindoles which are frequently exploited as indole bioisosteres³⁶ and, although some examples exist in the nature, most of them are synthetic products. Indoles and Azaindoles belong to the fused [5,6]-member ring systems. They are often classified as purinomimetics or purine isosteres, and exhibit a wide range of biological activities and pharmacological properties.³⁷ Among the natural substances where the azaindole core is present, variolin,³⁸ grossularines,³⁹ and neocryptolepine,⁴⁰ can be mentioned, and it is also a part of synthetic analogues of naturally occurring alkaloids, such as 7-aza-rebeccamycin⁴¹ and 5-

aza-ellipticine.⁴² Their biological activities are based mostly on their affinity toward DNA, but also as topoisomerase inhibitors and as potential multi kinase inhibitors (Fig. 38).

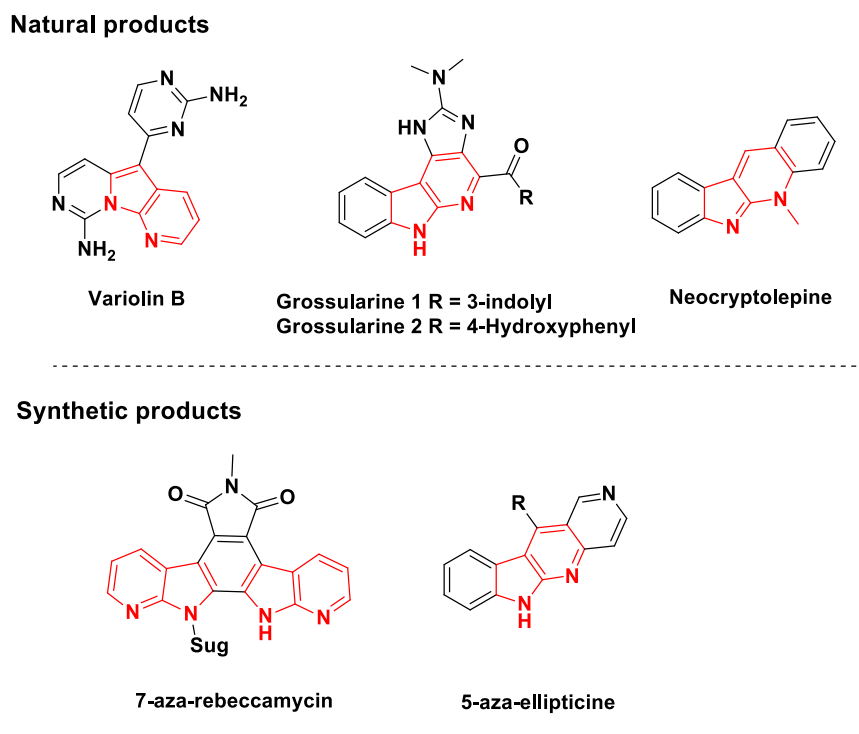


Figure 38. Hetero-condensed 7-azaindoles present in natural and synthetic compounds.

In the same way, highly functionalized 7-azaindoles have become major goals for medicinal chemistry studies. Some of them have been shown to be antagonists of the corticotropin-releasing hormone receptor (CRH1-R), which is involved in anxiety and depressive disorders,⁴³ and the substituted azaindole RWJ 68354 is a potent inhibitor of p38 kinase in vitro and in vivo and appears to be an attractive candidate for further preclinical evaluation (Fig.39).⁴⁴For all these reasons, they play a crucial role as potential lead compounds for the discovery of biologically active substances.

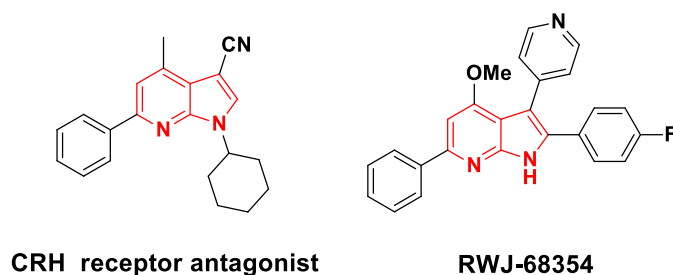


Figure 39. Biological compounds containing highly functionalized 7-azaindoles in their structure.

2.2.5 Indazole: Chemical and Biological Importance

Nitrogen-containing heterocycles are pharmacologically important scaffolds, and they are widely present in numerous commercially available drugs. As a crucial family of nitrogen containing heterocycles, the structurally diverse indazole analogues have received enormous attention in the past, as well as in recent years, because of their varieties of biological properties, such as anti-inflammatory, antimicrobial, anti-HIV, antihypertensive, and anticancer activities.^{45,46} More importantly, some indazole-based therapeutic agents, like pazopanib, axitinib, and niraparib have been approved for the treatment of cancers. Structurally, indazole, also called benzpyrazole or isoindazone, is an aromatic heterocyclic molecule in which a benzene ring is fused with a pyrazole ring. It exists in three tautomeric forms: 1H-indazole, 2H-indazole, and 3H-indazole (Fig. 40). 1H-Indazole and its derivatives are usually thermodynamically more stable than the corresponding 2H- or 3H forms and are therefore the predominant tautomers.⁴⁷ There is evidence that the indazole tautomer identity has an influence on biological properties.

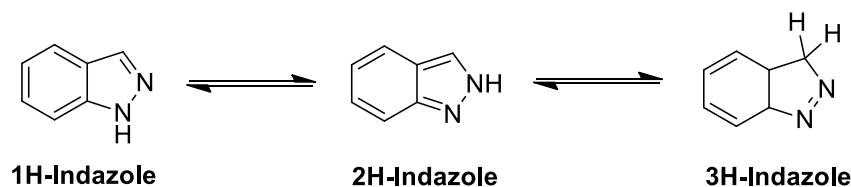


Figure 40. Tautomers of indazole.

2.2.6 Indazole derivatives as receptor tyrosine kinase inhibitors

The human receptor tyrosine kinase (RTK) family consists of 58 proteins divided into 20 subfamilies. These RTKs play a pivotal role in regulating cell proliferation, differentiation,

survival, apoptosis, adhesion, and migration. However, hyperactivation of RTKs could result in the development of a number of cancers. As a result, inhibition of RTK activity is becoming a common strategy for cancer therapy. Recently, the development of indazole derivatives targeting epidermal growth factor receptor (EGFR), fibroblast growth factor receptors (FGFRs), and vascular endothelial growth factor receptors (VEGFRs) has been widely reported.⁴⁸

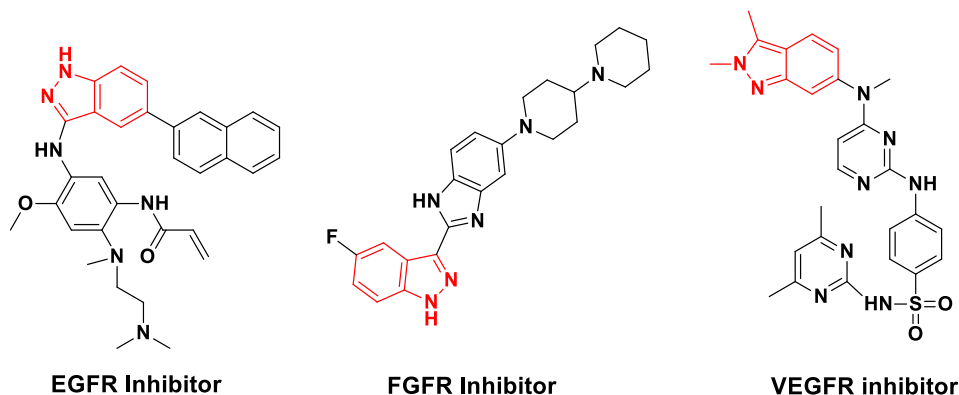


Figure 41. Biological compounds containing highly functionalized Indazole in their structure.

Indazole scaffolds has tremendous potential to act as an anticancer agent by targeting G4s sequences mainly because of its tautomeric efficiency. To explore the effect of indazole fragments on G4s DNA, we have reported the screening of indazole based scaffolds in this chapter.

2.2.7 Imidazo [1,5-a]pyridine and Benzimidazole: Chemical and Biological Importance

In recent years, different heterocyclic compounds have been identified through molecular biology, empirical screening, and rational drug design for the development of new anticancer agents.⁶ Nitrogen-bridgehead fused heterocycles containing an imidazole ring are common structural components in several pharmacologically important molecules that display a wide range of activities for diverse number of targets. Imidazopyridines exhibit anticancer activity through different molecular mechanisms such as vascular endothelial growth factor (VEGF)-receptor KDR inhibition and the induction of apoptosis.⁴⁹ Recently, reported synthesis of imidazopyridine-oxindole derivatives (4) as apoptosis inducing agents.⁵⁰

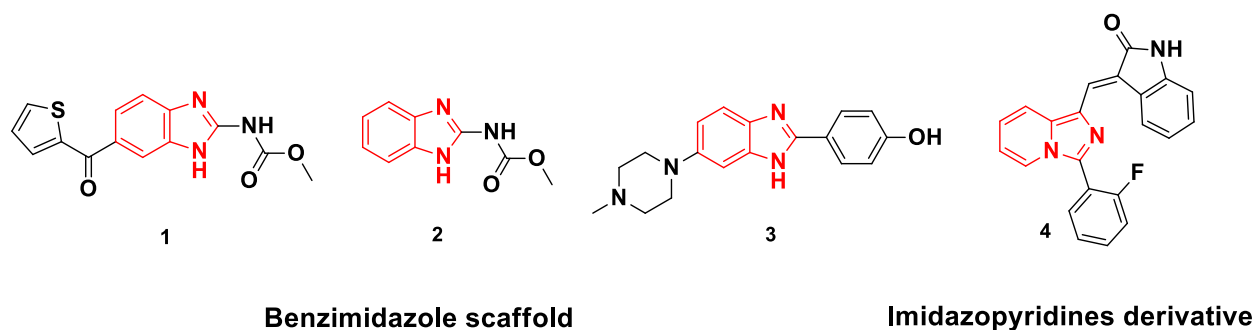


Figure 42. Lead structures of biologically active antitumor agents having benzimidazole scaffold: nocodazole (1), FB642 (2), Hoechst-33258 (3) and Imidazopyridines derivative (4)

2.2.8 2, 6-diaminopyrimidin-4-ol: Chemical and Biological Importance

Heterocyclic compounds contain at least one hetero atom in their rings, such as sulfur, nitrogen, or oxygen, and have a wide range of applications in both pharmacological and medicinal chemistry fields. Pyrimidine as one of the heteroaromatic systems often gets much interest because of its broad spectrum of biological and pharmacological activities. Pyrimidine is nothing but 1,3-diazine, similar to benzene and pyridine, and contains nitrogen at 1 and 3 positions. Pyrimidine and its fused pyrimidine derivatives are a class of heterocyclic scaffolds that demonstrates a broad range of biological and pharmacological activities such as anticancer, anxiolytic, antioxidant, antiviral, antifungal, anticonvulsant, antidepressant, and antibacterial activities. It has been reported in literature that the heterocyclic aromatic scaffolds that have pyrimidine molecules in their rings possess wide ranges of biological activities.

Some of the drugs available in the market that have pyrimidine analogs in their rings are used for treatment of several disorders. For example, stavudine is used as anti-HIV agent, fervernuline is used as antibiotic, minoxidil is used antihypertensive agent, 5-flourouracil is used in cancer treatment, and sulfamethiazine and trimethoprim are used in treatment against bacterial diseases. Also phenobarbitone is used as a sedative-hypotic as well as anticonvulsant, triflouridine and idoxuridine are used in treatment against various viral infections, and propylthiouracil is used as antithyroid agent. Triazolopyrimidines, pyridopyrimidines, pteridines, purines, furopyrimidines, pyrimidoazepines, pyrrolopyrimidines, and quinazolines are a class of fused pyrimidine scaffolds that exhibit good pharmacological properties. So, one can improve the activity of heterocyclic compounds by fusing the pyrimidine analogs with different heterocyclic moieties. Many pyrimidine and fused pyrimidines are vascular endothelial growth factor receptors, which have been used as cancer therapy against various types of cancers

as approved by the US Food and Drug Administration.⁵¹ Pazopanib, imatinib, dasatinib, nilotinib, uramustine, tegafur, cytarabine, and fluorouracil are a few examples of drugs that contain pyrimidine molecules as a core moiety used as anticancer agents (Fig. 43).

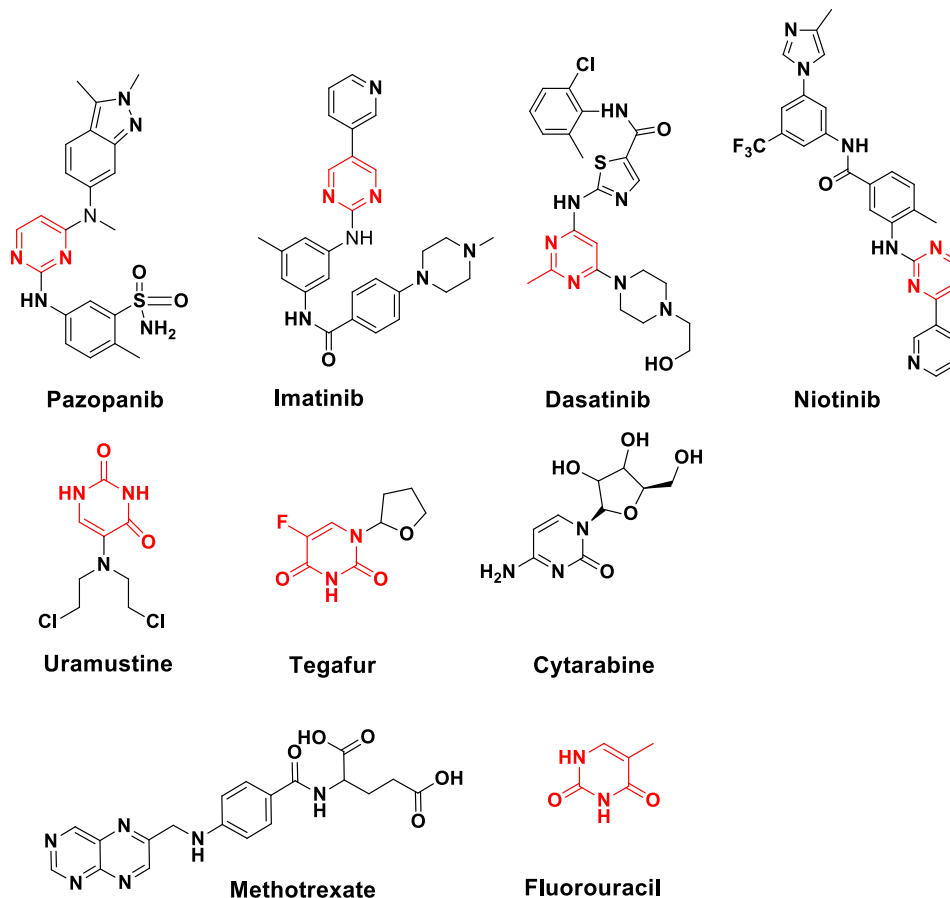


Figure 43. Drugs contained pyrimidine molecule as a core moiety.

2.2.9 Pyrazolopyrimidine, 2-hydroxynaphthalene-1,4-dione, 1,4-dihydroxyanthracene-9,10-dione, Benzofuran and Piperonal: Chemical and Biological Importance

Pyrazolopyrimidine are fused heterocyclic ring systems, which structurally can consider as bioisosteres of adenine, which is fundamental for every aspect of cell life. Pyrazolo [3,4-d]pyrimidines derivatives have been explored for their inhibitory activity towards various protein kinase enzymes and their role as anticancer agents (Fig.44).

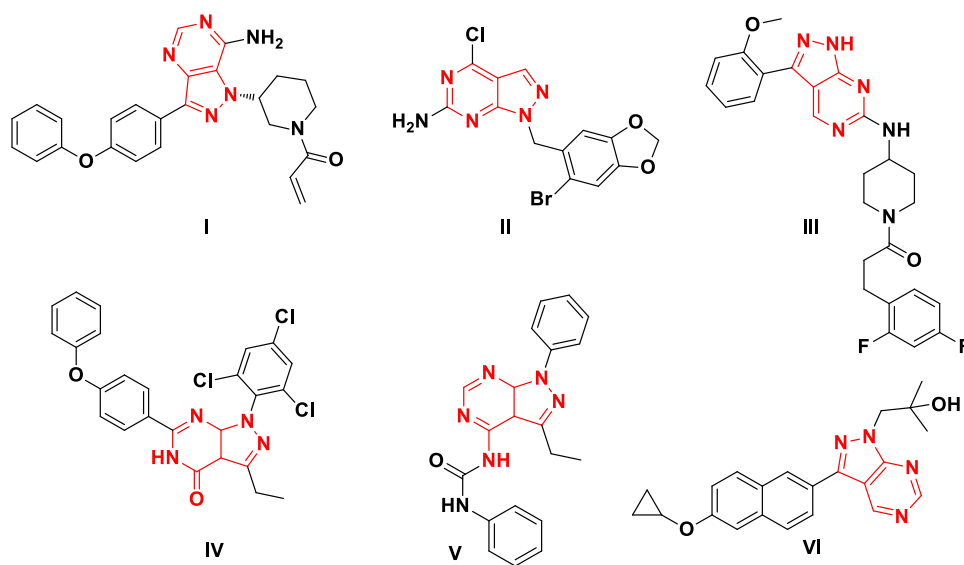


Figure 44. The structures of several Pyrazolopyrimidine.

2.2.10 Morpholino: Chemical and Biological Importance

Morpholine (1,4-tetrahydro-oxazine) is a simple heterocyclic compound that has great industrial importance and a wide range of applications. This chemical compound and its derivatives have been used as rubber additives, corrosion inhibitors, solvents, optical brighteners, antioxidants and in the manufacture of a range of drugs and herbicides (Fig. 45).⁵²

2.2.11 1,4-naphthoquinones: Chemical and Biological Importance

Quinones have been frequently exploited for the discovery of cellular mechanisms associated with cytotoxicity in various cancer cells. The redox properties of quinones can often induce apoptosis in cancer cells through oxidative stress induced by the *in-situ* creation of reactive oxygen species (ROS), while additional evidence proposes that some quinones can intercalate with DNA or inhibit proteins involved in DNA replication.⁵³ Quinones such as 1,4-naphthoquinones are abundant in nature⁵⁴ and naphthoquinone based natural products are known to possess anticancer activity. This pharmacophore is known to convey anticancer activity to some drugs such as streptonigrin,⁵⁵ mitomycin A.⁵⁶ 1, 4-Naphthoquinones are extensively distributed in nature and many well-known important anticancer drugs having a quinone moiety such as anthracyclines,⁵⁷ mitoxantrones,⁵⁸ and saintopin (Fig. 45) have shown tremendous anticancer activity.⁵⁸

Anthraquinone derivatives (AQs) 1,4-dihydroxyanthracene-9,10-dione represent an interesting scaffold to develop selective and multifunctional G4 ligands, with many potential applications, because of their well characterised DNA-binding properties,⁵⁹ fairly low redox potential and their ability to act as photosensitizers by one-electron oxidation. Structurally, AQs are strictly related to the anthracycline antibiotics like doxorubicin and daunomycin.⁶⁰

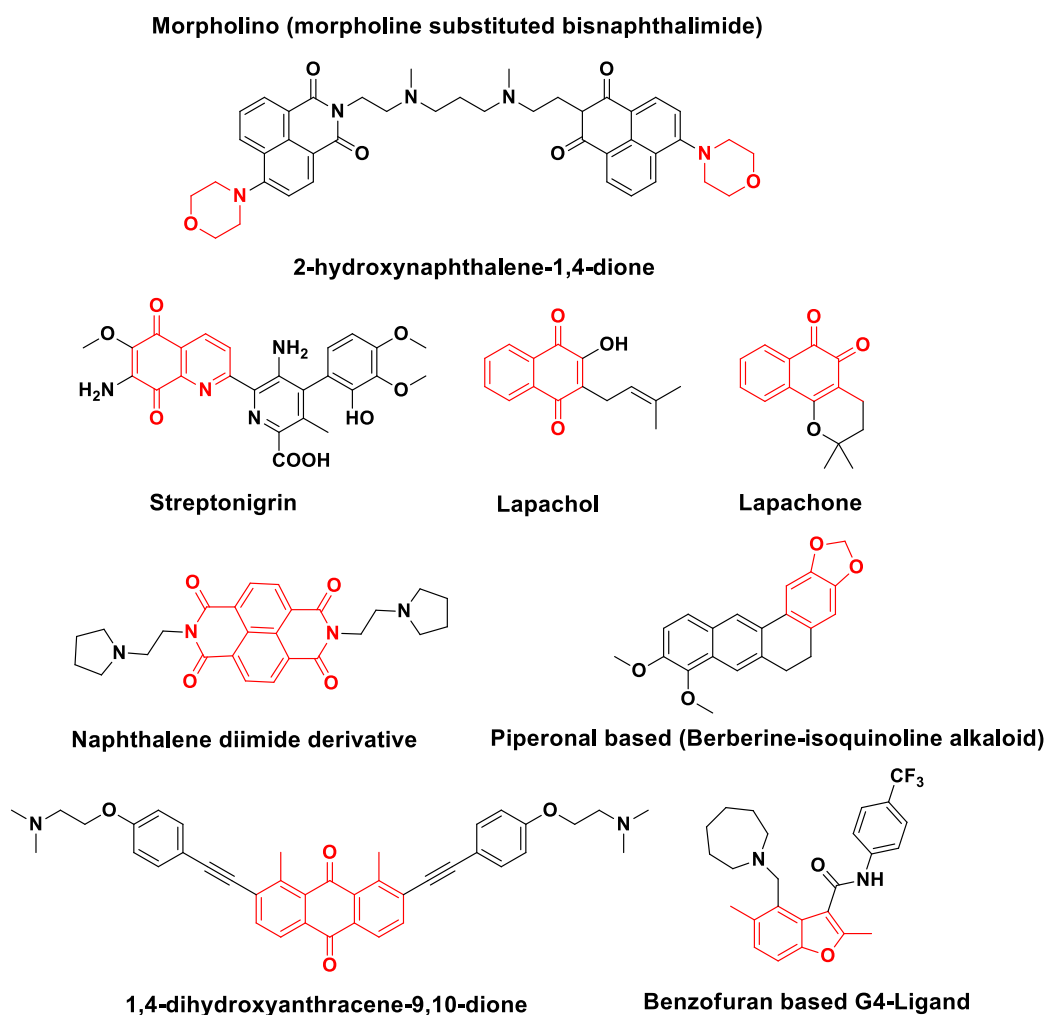


Figure 45. The structures of several biological active agents with morpholino, 2-hydroxynaphthalene-1,4-dione, 1,4-dihydroxyanthracene-9,10-dione, benzofuran and piperonal

In our current investigation, all fragments were selected by taking into consideration the features which were previously reported in literature to target G4s sequences.

2.3 Project Objective

The primary goal of chapter I in this thesis is to do an intensive screening of selected fragments obeying drug-likeness rules with Fluorescent indicator displacement (FID) assays. The fragment-based screening approach is used to identify the hit and based on our screening results identified hit were further optimized by fragment expansion strategy.

2.4 Results and Discussion

2.5 Initial Screening of Low-Molecular Weight Fragments

To identify new pharmacophores targeting G4s DNA, screenings of numerous heterocycles have been reported including indoles,⁶¹ 7-azaindoles,⁶² 1H-indazol-3-yl,⁶³ benzothiazole,⁶⁴ imidazo[1,5-a]pyridine,⁶⁵ 2,6-diaminopyrimidin-4-ol,⁶⁶ 1H-pyrazolo[4,3 d]pyrimidin-7-amine,⁶⁷ morpholino,⁶⁷ bis-indoles,^{68,69} 2-hydroxynaphthalene-1,4-dione,⁷⁰ 1,4-dihydroxyanthracene-9,10-dione, benzofuran and piperonal which were derived from several alkaloids.^{81,82} We initially screened around 130 structurally and chemically diverse fragment molecules, selected from an in-house library of commercial or in-lab synthesized compounds. Each member of this starting fragment family obeyed the principal criterion of fragment libraries (Table 4 & 5 in later part of this chapter). All the fragments were >95% pure. From our initial screening we got some promising hit fragment, which were then further optimized with fragment expansion strategies with feasible synthetic approaches which are reported in later half of this chapter.

2.5.1 Drug Likeness properties of screened fragments

The physico-chemical properties of all the fragments were assessed using the online SwissADME free web tool.⁷³ Based on Lipinski's rule of five,⁷⁴ used as a requirement for drug-like properties to predict oral bioavailability (O-B-A), the molecular property of a compound is described in order to estimate its pharmacokinetic parameters in the human body, including its absorption, distribution, metabolism and excretion. For most 'drug-like' molecules all together six phycochemical properties are taken into account for prediction of oral bioavailability radar, these are good permeability across the cell membrane (LIPO), solubility in aqueous media (INSOL), the saturation character (INSATU), the molecular flexibility as function of number of hydrogen bond acceptors (FLEX), a molecular weight below 500 (SIZE) a polar surface area characterizing drug absorption (POLAR) and less than 5 hydrogen bond donors. ⁷ Molecules violating more than one of these rules decrease the activity and selectivity of a likely drug candidate and therefore make

it unlikely to be orally active in humans. Significance of all the physico-chemical properties required for fragments to qualify as a drug-candidates are well explained in section 2.1.2, 2.1.3 and 2.14 of this chapter.

The physico-chemical properties of all the fragments were assessed using the online SwissADME free web tool.⁷³ The SwissADME free web tool is very efficient and useful for the analysis of pharmacokinetics, drug-likeness and medicinal chemistry parameters of the fragments before performing HTs screening with FID assays. The physico-chemical properties of the fragments screened in these chapters are mentioned in table 5, which suggests that all the optimized fragments are obeying the drug discovery rule and possess. The physico-chemical, pharmacokinetics, drug-likeness, lipophilicity and water solubilities properties are summarised in the later part of the chapter in table 6 and 7 for all the fragments used and further optimized *via* fragment based drug design strategy.

2.5.2 Initial screening of fragments with Fluorescent indicator displacement (FID) assays

To identify the hit we first screened for binding of commercial and synthesized 130 fragments to a DNA G4 derived from the *c-MYC* promoter by using fluorescent indicator displacement (FID) assays employing thiazole orange (TO). TO is a well-validated probe for screening G4-binding ligands.⁷⁵ It is highly fluorescent when bound and quenched after displacement (λ_{ex} . wavelength = 501 nm, λ_{em} . wavelength = 539 nm) (Fig. 46). FID assays were performed to investigate the binding affinity of selected fragments for *c-MYC*, *c-KIT1*, *BCL-2*, h-TELO, G4 and duplex ds24-DNA and ds27-DNA DNA sequences (Table 2). The dissociation constant for thiazole orange with the respective sequence of DNA used are depicted in figure 47 and table 3.

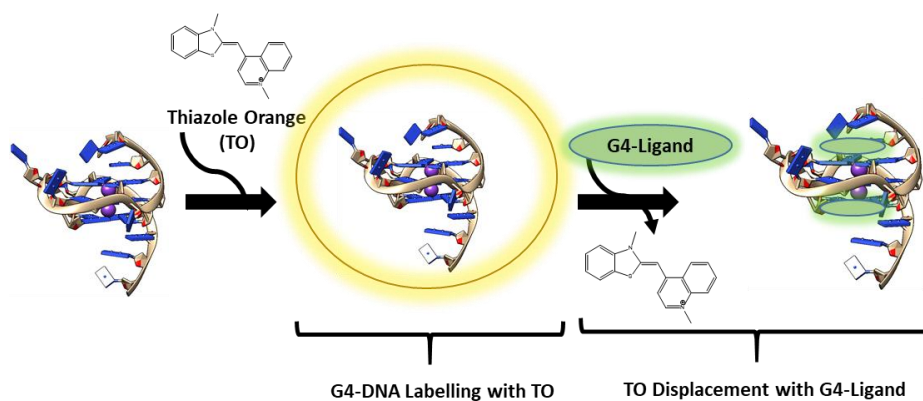


Figure 46. Schematic representation of fluorescent indicator displacement (FID) assays

Table 2. Sequence of G-Quadruplex and Duplex DNA used for thiazole displacement assays

<i>G₄</i> -DNA	Sequences (5'-....-3')
<i>c-MYC</i>	TGGGGAGGGGTGGGGAGGGTGGGGAAGG
<i>h-TELO</i>	TTGGGTTAGGGTTAGGGTTAGGGA
<i>BCL2</i>	AGGGGCGGGCGGGAGGAAGGGGCGGGAGCGGGGCTG
<i>c-KIT1</i>	GGGAGGGCGCTGGGAGGAGGG
dsDNA 27bp	TATAGCTATAAAAAAAAAATATAGCTATA
dsDNA 24bp	CGCGAATTCGCGCGCGAATTCGCG

Table 3. Dissociation constant (K_d) values of TO for G-quadruplex & Duplex-DNA Sequences.

	<i>c-MYC</i>	<i>h-TELO</i>	<i>BCL2</i>	<i>c-KIT1</i>	dsDNA 27bp	dsDNA 24bp
K_d Values	3.66 μ M	2.034 μ M	3.79 μ M	3.465 μ M	4.89 μ M	3.84 μ M

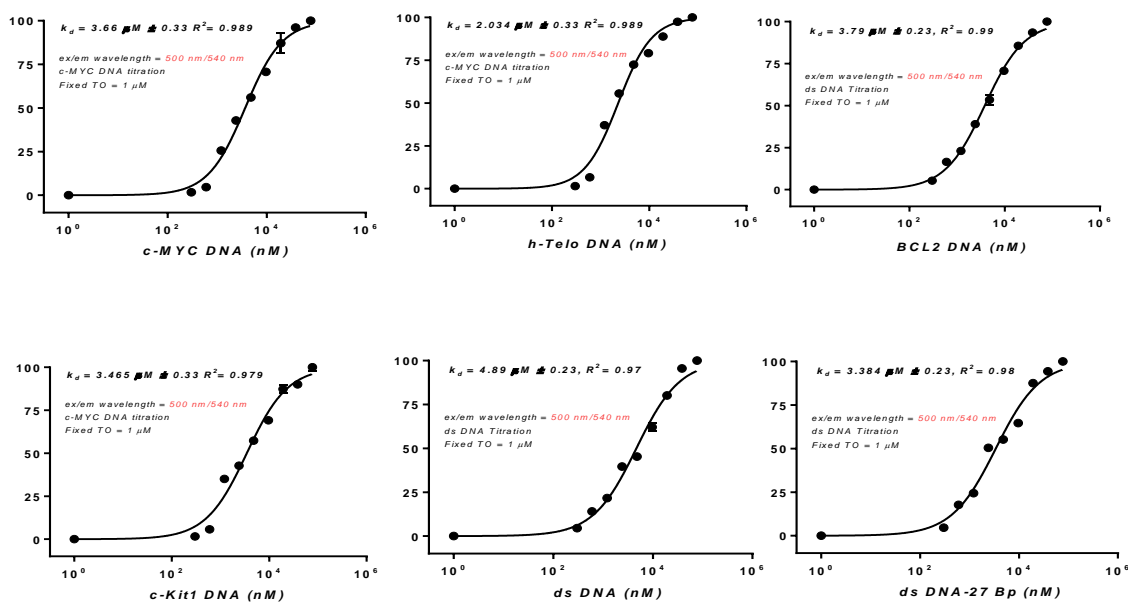


Figure 47. Dissociation constant (K_d) values of TO for *c-MYC*, *h-Telo*, *BCL-2* and *c-KIT-1* G-quadruplex & Duplex-DNA Sequences. **A)** Titration of fixed Thiazole Orange: $c(\text{TO}) = 1 \mu\text{M}$; $c(\text{cMYC})/c(\text{KIT1})/c(\text{BCL2})/c(\text{h-TELO}) = 0 \mu\text{M}$, 0.27 μM , 0.55 μM , 1.09 μM , 2.18 μM , 4.36 μM , 8.72 μM , 17.44 μM , 34.88 μM , 69.76 μM ; $c(\text{DMSO}) = 10\%$; 20 mM Na-caco, 140 mM KCl, pH = 7, rt **B)** Titration of ds-DNA - Thiazole Orange: $c(\text{TO})$

= 1 μ M; c(ds-DNA) = 0 μ M, 0.27 μ M, 0.55 μ M, 1.09 μ M, 2.18 μ M, 4.36 μ M, 8.72 μ M, 17.44 μ M, 34.88 μ M, 69.76 μ M ; c(DMSO) = 10%; rt, 20 mM Na cacodylate, 140 mM KCl, pH 7, r.t.

The high throughput screening of the 130 structurally and chemically diverse ligands were then performed by FID assays to check its affinity for c-MYC promoter G4 (Fig.48). Based on our results from the intensive HTs FID screening, we found that 5-aminogramine (labelled as A8 in Fig.50a in later part of this chapter). The identified 5-aminogramine fragment from our HTs FID screening was further optimized by the FBDD strategy.

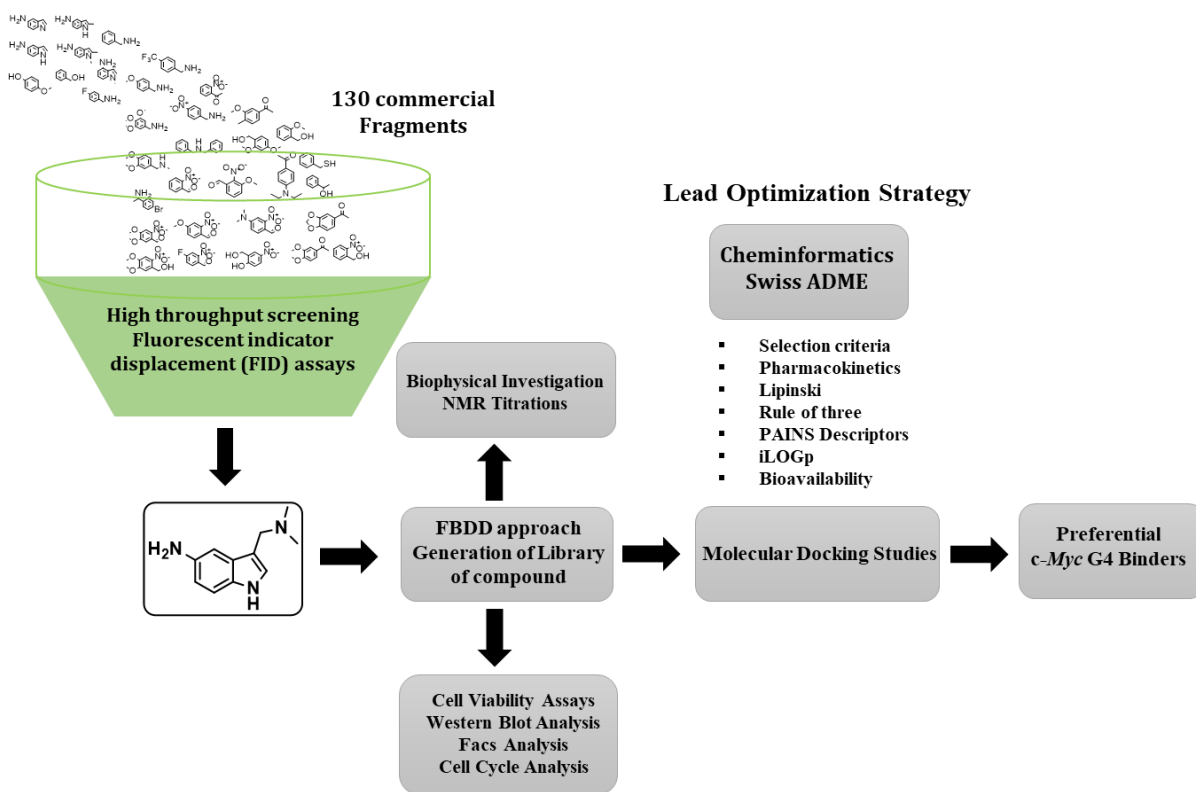


Figure 48. Schematic representation of high throughput screening to identify the hit and for lead optimization

The FID screening of all the fragment, which includes commercial as well as optimized fragments based on our FBDD strategy are presented in the later part of this chapter (Fig. 50a and 50b). Recently reported indolylmethyleneindanone, indenopyrimidine and bisbenzimidazole carboxamide derivatives of naphthiridine and phenanthroline can specifically stabilize c-MYC and c-KIT quadruplex DNAs with propeller topology (Fig. 49).^{76,77} Even though these ligands were specific toward a particular topology, weak stabilization and poor drug-likeness properties for

indenopyrimidine weakens the therapeutic index of these molecules. Taking into consideration the key feature of reported indolylmethyleneindanone with the extended cationic pyrrolidine side chain, which is the prerequisite for the effective binding to the G-quadruplex.

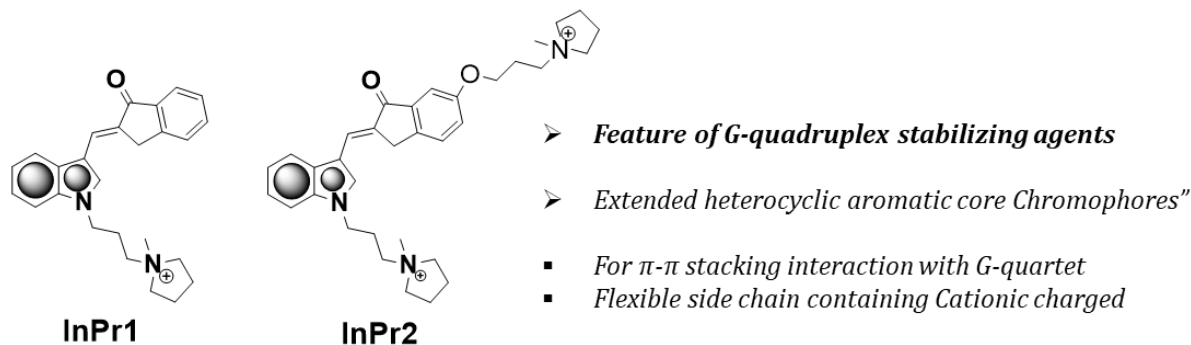


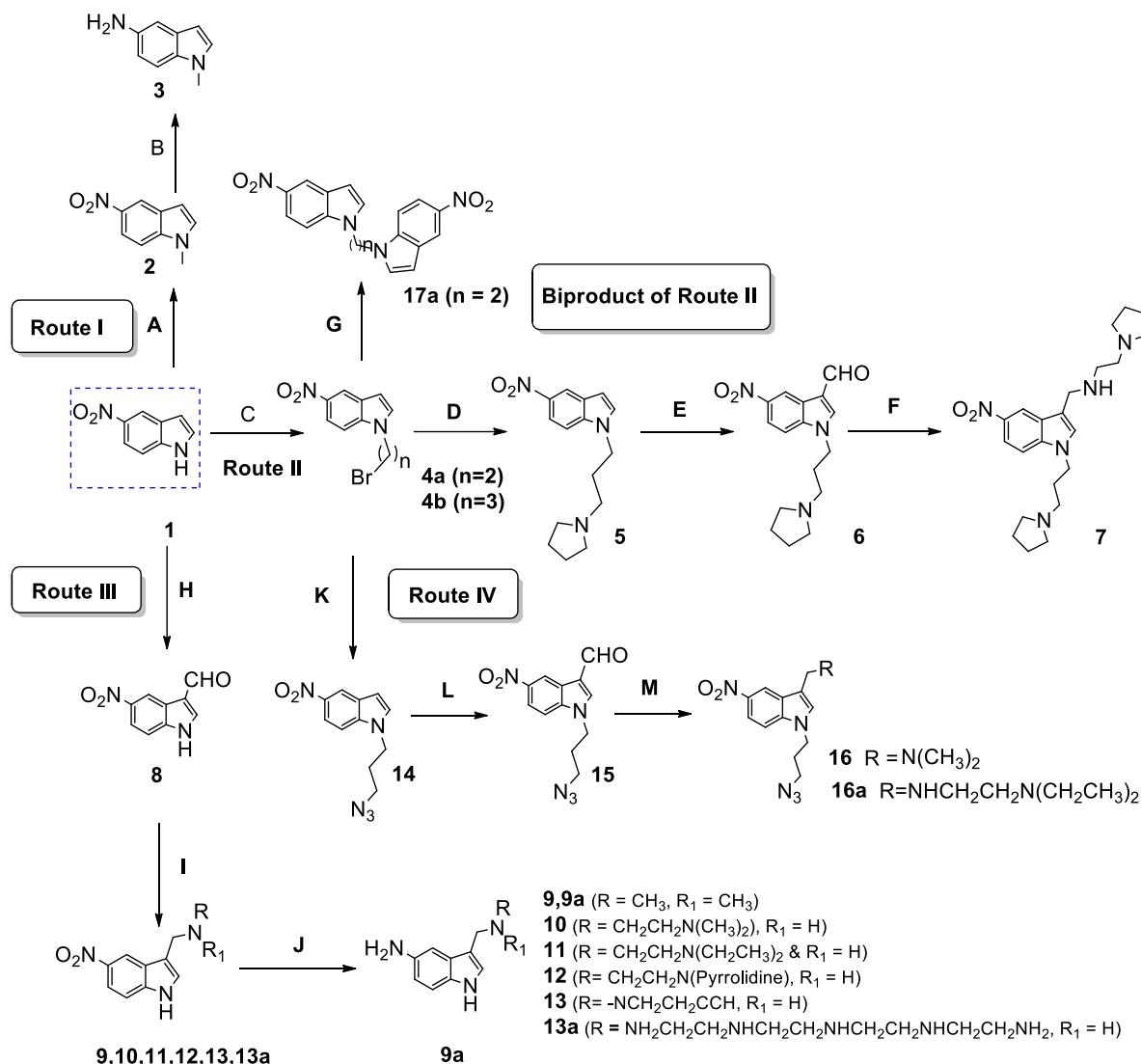
Figure 49. Chemical structures of indolylmethyleneindanone-based scaffolds developed for topology-specific stabilization of *c-MYC* and *c-KIT* promoter G-quadruplex DNAs.^{76,77}

In our current investigation, we did further hit optimization of Indole and several other heterocycles like azaindole, piperonal and indazole fragments with the feasible and synthetic strategies for the development of small drug-like. The Physico-chemical properties (Table 5), pharmacokinetic, drug-likeness, medicinal chemistry features (Table 6) and the lipophilicity and water solubility characteristics (Table 7) of the designed fragments are reported in the later part of this chapter.

2.6 Design and Synthesis

We developed different synthetic strategies to generate drug-like G-quadruplex stabilizing ligands derived from 5-nitroindole (Scheme 1, 2 and 3). In route 1, scaffold 1-methyl-1H-indol-5-amine (**3**) is obtained in 60 % yield with Pd/C-catalyzed hydrogenation of 1-methyl-5-nitro-1H-indole (**2**), which is a nucleophilic substitution product of 5-nitro-1H-indole (**1**).⁷⁸ In route II, intermediates **4a-b** are generated *via* a nucleophilic substitution reaction of 5-nitro-1H-indole (**1**) with 1,3-dibromopropane and 1,2-dibromoethane, respectively, which generates the 5-nitroindole dimer 1,3-bis(5-nitro-1H-indol-1-yl)propane (**17a**) as a by-product.⁷⁹ In the next step, **4a** is protected with pyrrolidine to obtain 5-nitro-1-(3-(pyrrolidin-1-yl)propyl)-1H-indole (**5**).⁸⁰ **5** was treated with the Vilsmeier reagent⁸¹ to generate 5-nitro-1-(3-(pyrrolidin-1-yl)propyl)-1H-indole-3-carbaldehyde (**6**)⁸¹ to facilitate the one pot *in-situ* generation of N-((5-nitro-1-(3-(pyrrolidin-1-yl)propyl)-1H-indol-3-yl)methyl)-2-(pyrrolidin-1-yl)ethanamine (**7**).⁸²

In route III, intermediate 5-nitro-1H-indole-3-carbaldehyde (**8**) is obtained in 60 % yield via Vilsmeier-Haack reaction⁸¹ with 5-nitro-1H-

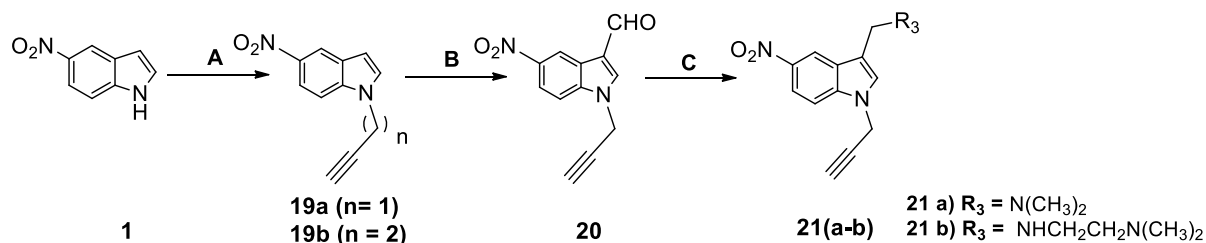


Scheme 1. A) NaH, DMF, rt, 1 h, CH₃I, r.t, 8 h, 95 %; B) Pd/C, H₂, MeOH, r.t, 3 h, 52%; C) K₂CO₃, DMF, 1,3-Dibromopropane, r.t, 4 h, 65 %; D) Dry ACN, Pyrrolidine, reflux, 4 h, 32%; E) POCl₃, DMF, 0 °C - r.t, 1 h, 46%; F) NaBH₄, MeOH, Substituted amine, r.t, 3h, 72%; G) K₂CO₃, DMF, rt, 3h, 62%; H) POCl₃, DMF, 0 °C - rt, 1 h, 30 %; I) NaBH₄, MeOH, Substituted amine, rt, 3h, 56 %, J) Pd/C, H₂, MeOH, RT, 3 h, 52%; K) NaN₃, DMF, 80 °C, reflux, 4 h, 48 %; L) POCl₃, DMF, 0 °C - rt, 1 h, 35 %; M) NaBH₄, MeOH, corresponding amine, rt, 3 h, 45 %.

indole (**1**) which is a key step to generate conjugates **9**, **10**, **11**, **12**, **13**, **13a** in quantitative yields with one pot *in-situ* reaction of aldehydes with substituted amines in presence of NaBH₄ as a reducing agent.⁸² This strategy is useful for the Pd/C-catalyzed hydrogenation⁷⁸ of N, N-dimethyl-

1-(5-nitro-1H-indol-3-yl)methanamine (**9**) to 3-((dimethylamino) methyl)-1H-indol-5-amine (**9a**). During the development of synthetic strategies, we found that the substituted 5-aminoindole derivatives are potentially unstable and it is difficult to protect them from air oxidation.⁸³ In addition to this, for the generation of 1-(1-(3-azidopropyl)-5-nitro-1H-indol-3-yl)-N,N dimethylmethanamine (**15**) and N1-((1-(3-azidopropyl)-5-nitro-1H-indol-3-yl)methyl)-N2,N2-diethylethane-1,2-diamine (**16**) in 40-45 % yields in route IV has been developed from the intermediate **4a**.⁸⁴

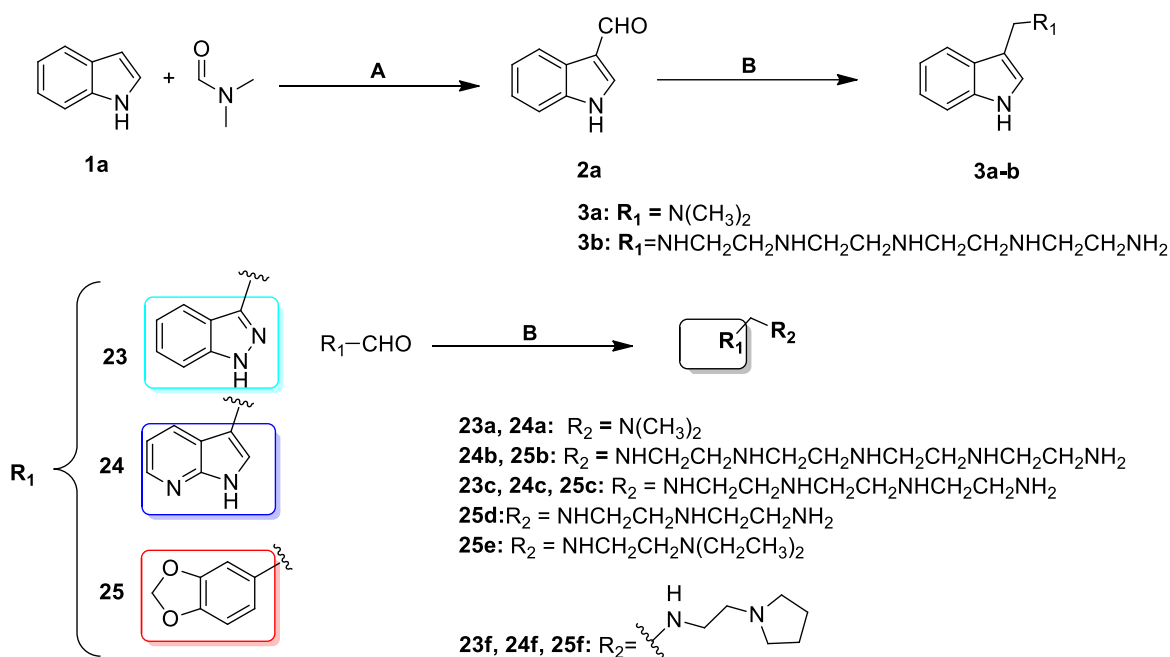
For this, intermediate **4a** is allowed to react with NaN_3 in DMF to reach 1-(3-azidopropyl)-5-nitro-1H-indole (**14**) *via* nucleophilic substitution reaction in 60 % yields. Precursor (**14**) is then treated with the Vilsmeier reagent⁸¹ to afford 1-(3-azidopropyl)-5-nitro-1H-indole-3-carbaldehyde (**15**) in 55 % yield.⁸¹ Conjugates **16a** and **16b** were obtained from **15** by one pot *in-situ* reaction of substituted amine with aldehyde in presence of NaBH_4 in methanol media with 56 % yield.⁸²



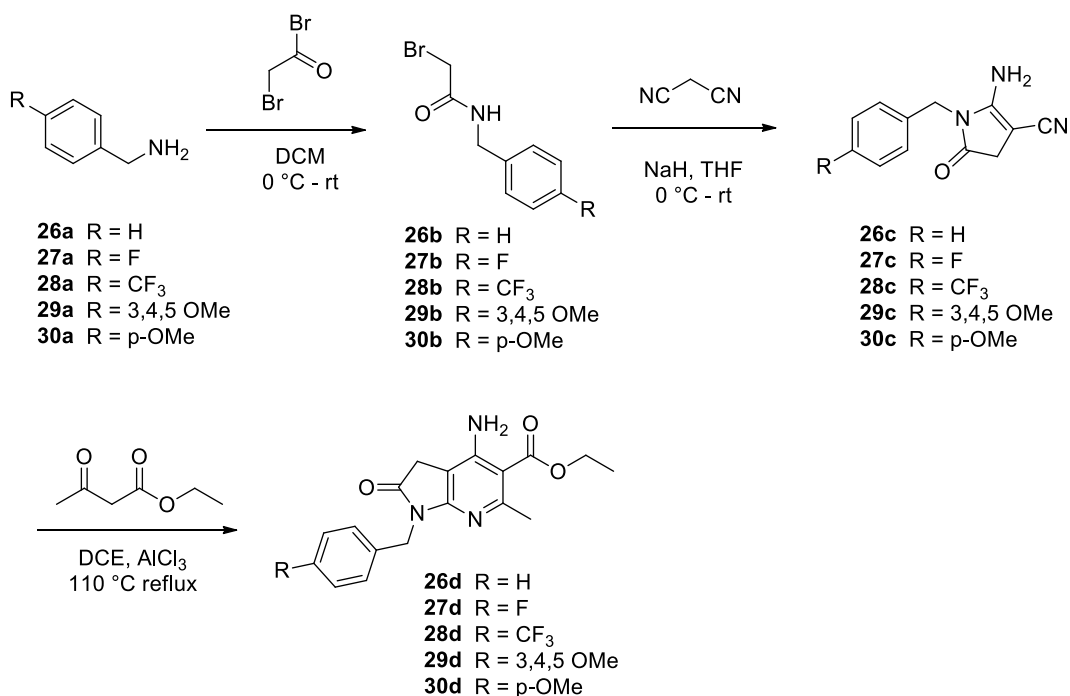
Scheme 2. A) K_2CO_3 , DMF, 3-bromo-1-propyne, r.t, 4 h, 88 %; B) POCl_3 , DMF, 0 °C - r.t, 1 h, 58 %; C) NaBH_4 , MeOH, Substituted amine, r.t, 3 h, 38 %

In Scheme 2 to afford N1,N1-dimethyl-N2-((5-nitro-1-(prop-2-yn-1-yl)-1H-indol-3-yl)methyl)ethane-1,2-diamine (**21a**)⁸² and N1,N1-diethyl-N2-((5-nitro-1-(prop-2-yn-1-yl)-1H-indol-3-yl)methyl)ethane-1,2-diamine (**21b**) were prepared *via* substitution reaction of propargyl bromide with **1** to obtain intermediate **19a-b** which pave route for the synthesis of 5-nitro-1-(prop-2-yn-1-yl)-1H-indole-3-carbaldehyde (**20**) with the Vilsmeier-Haack reaction.⁸¹ In addition to above in scheme 3, several other aldehydes (**2a**, **23**, **24**, **25**) were used to afford conjugates **3a-b**, **23a-c-f**, **24a-b-c-f**, **25b-e-f**, with one pot *in-situ* reaction of substituted amine with aldehyde in presence of NaBH_4 in methanol media with 56 % yield.

A Fragment-Based Approach for the Development of G-Quadruplex Ligands



Scheme 3. A) POCl_3 , DMF, $-10^\circ\text{C} > \text{r.t.}$, Reflux, 100°C , 1 h, 88 %; B) NaBH_4 , MeOH, Substituted amine, r.t, 3h, 25 -37 %



Scheme 4. A) DCM, $0^\circ\text{C} - \text{r.t.}$, 1 h, 95 %; B) NaH , THF, $0^\circ\text{C} - \text{r.t.}$, 52%; C) AlCl_3 , DCE, 110°C , reflux, 4 h, 48 %

Due to the great potential of pharmaceutical and biological applications, we aimed at exploring a concise and efficient approach (Scheme 4) to synthesize 7-azaindole derivatives. Here, we started from substituted N-benzyl-2-bromoacetamine (**26b**, **27b**, **28b**, **29b** and **30b**), which could be obtained from 2-bromoacetyl bromide **2** with various substituted phenylmethanamine (**26a**, **27a**, **28a**, **29a** and **30a**). Treatment of substituted N-alkyl- α -bromoacetamide with malononitrile **4** in THF in the presence of NaH, afforded substituted enamino γ -lactams (**26c**, **27c**, **28c**, **29c** and **30c**). 7-azaindole derivatives (**26d**, **27d**, **28d**, **29d** and **30d**) could be constructed through friedlander reaction of substituted enamino γ -lactams (**26c**, **27c**, **28c**, **29c** and **30c**) and ethyl acetoacetate with lewis acid ($AlCl_3$) under reflux condition. In total 50 compounds were synthesized, which includes substituted indoles, 5-nitroindoles, 7-azaindole and indazole derivatives.

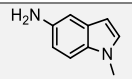
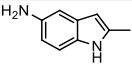
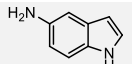
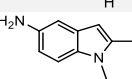
2.7 SMILES (Simplified Molecular Input Line Entry System)

For the investigation of drug-likeness of the screened fragments, firstly it is necessary to translate the computed fragments into a SMILES-code, by using an Online SMILES translator to run the drug-likeness calculations in SWISS-ADME filters web based tool (Table 4). In our current investigation to calculate all the physicochemical, pharmacokinetics, drug-likeness and lipophilicity of the fragments,

SMILES was developed and initialized in the 1980s by David Weininger.⁸⁵

Significance of SMILES: The simplified molecular-input line-entry system (**SMILES**) is a specification in the form of a line notation for describing the structure of chemical species using short ASCII strings. Molecules or reactions can now be easily describe by a code. It is also possible to show double bonds, use different isotopes or show ionic compounds.

Table 4. List of screened commercial and synthesized fragments with origin (commercial or synthesized) and SMILES

HIT	STRUCTURE	SMILE NOTATION	ORIGIN (SOURCE)
A1 (3)		<chem>CN1C=CC2=C1C=CC(N)=C2</chem>	Synthesized (3)
A2		<chem>CC1=CC2=C(N1)C=CC(N)=C2</chem>	Commercial
A3		<chem>NC1=CC2=C(NC=C2)C=C1</chem>	Synthesized (31)
A4		<chem>CN1C(C)=CC2=C1C=CC(N)=C2</chem>	Commercial

A Fragment-Based Approach for the Development of G-Quadruplex Ligands

A5		<chem>COCCN1C(C)=CC2=C1C=CC(N)=C2</chem>	Commercial
A6		<chem>NC1=CC2=C(C=C1)N(CC1=CC=CC=C1)C=C2</chem>	Commercial
A7		<chem>CCC1=NC2=C(C=CC(N)=C2)N1C</chem>	Commercial
A8 (9A)		<chem>CN(C)CC1=CNC2=C1C=C(N)C=C2</chem>	Synthesized (9a)
A9		<chem>NC1=CC2=C(C=C1)N(CC1=CC=CC=C1)N=C2O</chem>	Commercial
A10		<chem>NCC1=CNC2=C1C=C(N)C=C2</chem>	Commercial
A11		<chem>CN1C=NC2=C1C=CC(N)=C2</chem>	Commercial
A12		<chem>CN1C=C(C)C2=C1C=CC=C2N</chem>	Commercial
A13		<chem>C(C1=CNC2C=CC=CC12)C1=CNC2=C1C=CC=C2</chem>	Commercial
A14		<chem>CN1C(C)=C(CCN)C2=C1C=CC(N)=C2</chem>	Commercial
A15		<chem>O=CC1=NNC2=CC=CC=C12</chem>	Commercial
A16		<chem>CN1C(C)=C(CC2=C(C)SC=N2)C2=C1C=CC(N)=C2</chem>	Commercial
A17		<chem>[H]C(=O)C1=C(C)N(C)C2=C1C=C(N)C=C2</chem>	Commercial
A18		<chem>CN1C(C)=C(CC2=CSC(Cl)=N2)C2=C1C=CC(N)=C2</chem>	Commercial
A19		<chem>[H]C(=O)C1=CN(C)C2=C1C=C(C)C=C2</chem>	Commercial
A20		<chem>CN1C=C(C(O)=O)C2=C1C=CC(N)=C2</chem>	Commercial

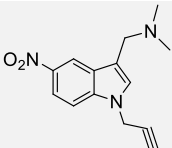
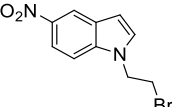
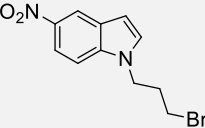
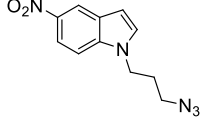
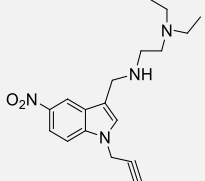
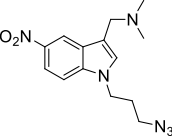
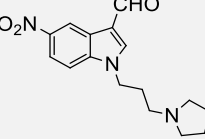
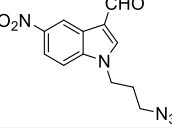
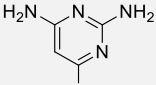
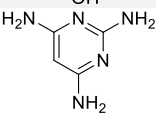
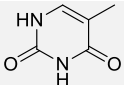
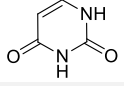
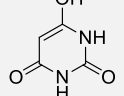
A Fragment-Based Approach for the Development of G-Quadruplex Ligands

A21		<chem>CN1C=C(CC(O)=O)C2=C1C=CC(=C2)N(=O)=O</chem>	Commercial
A22 (24c)		<chem>NCCNCCNCCNCC1=CNC2=C1C=CC=N2</chem>	Synthesized (24c)
A23		<chem>[H]C(=O)C1=CNC2=C1C=CC=N2</chem>	Commercial
A24		<chem>NC1=CC2=C(C=C1)N(C=C2)C1CCNCC1</chem>	Commercial
A25		<chem>NCC1=NN(CC2=CC=CC=C2)C2=C1C=C(C=C2)N(=O)=O</chem>	Commercial
A26		<chem>CN1C=C(CN)C2=C1C=CC(N)=C2</chem>	Commercial
A27		<chem>CN1C=NC2=C1C=CC(=C2)N(=O)=O</chem>	Commercial
A28		<chem>CN1C=CC2=C1C=CC(=C2)N(=O)=O</chem>	Commercial
A29		<chem>CN1C=C(CC2=CNC3=C2C=CC=C3)C2C=CC=CC12</chem>	Commercial
A30		<chem>CN1C(C)=C(CCN)C2=C1C=CC(N)=C2</chem>	Commercial
B1		<chem>C1=NC=C2C=CC=CN12</chem>	Commercial
B2 (19b)		<chem>C#CCCNC1=CC2=C1C=CC=C2</chem>	Synthesized (19b)
B3		<chem>[H]C(=O)C1=CN(C)C2=C1C=C(N)C=C2</chem>	Commercial
B4		<chem>CN1C(C)=C(CC2=CSC(F)=N2)C2=C1C=CC=C2</chem>	Commercial
B5		<chem>[H]C(=O)C1=CNC2=C1C=C(N)C=C2</chem>	Commercial
B6		<chem>CCN1C=C(C(O)=O)C2=C1C=CC(N)=C2</chem>	Commercial
B7		<chem>CCCN1C=C(CC(O)=O)C2=C1C=CC=C2</chem>	Commercial

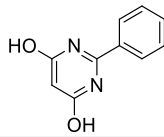
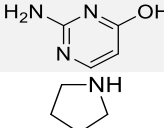

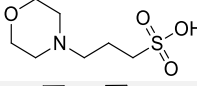
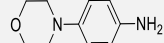
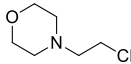

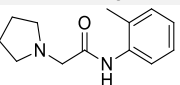
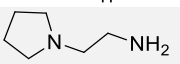
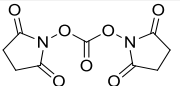
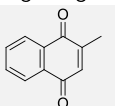
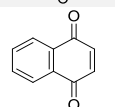
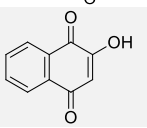
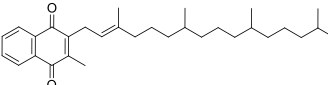
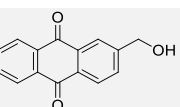
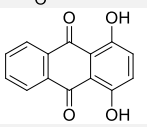
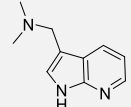
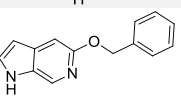
A Fragment-Based Approach for the Development of G-Quadruplex Ligands

B8		<chem>CN(C)CC1=CN(C)C2=C1C=C(OCC1=CC=CC=C1)C=C2</chem>	Commercial
B9		<chem>O=CC1=NC=C2C=CC=CN12</chem>	Commercial
B10		<chem>CN1C=CC2=C1C=CC(N)=C2</chem>	Commercial
B11 (9)		<chem>CN(C)CC1=CNC2=C1C=C(C=C2)N(=O)=O</chem>	Synthesized (9)
B12 (10)		<chem>CN(C)CNCC1=CNC2=C1C=C(C=C2)N(=O)=O</chem>	Synthesized (10)
B13 (12)		<chem>O=N(=O)C1=CC2=C(NC=C2CNCNC2CCCC2)C=C1</chem>	Synthesized (12)
B14 (11)		<chem>CCN(CC)CNCC1=CNC2=C1C=C(C=C2)N(=O)=O</chem>	Synthesized (11)
B15 (7)		<chem>O=N(=O)C1=CC2=C(C=C1)N(CCN1CCCC1)C=C2CNCNC1CCCC1</chem>	Synthesized (7)
B16 (17a)		<chem>O=N(=O)C1=CC2C=CN(CCCN3C=CC4=C3C=CC(=C4)N(=O)=O)C2C=C1</chem>	Commercial (17a)
B17		<chem>O=N(=O)C1=CC2C(C=CN2CCN2C=CC3=C2C=CC(=C3)N(=O)=O)C=C1</chem>	commercial
B18 (5)		<chem>O=N(=O)C1=CC2=C(C=C1)N(CCCN1CCCC1)C=C2</chem>	Synthesized (5)
B19 (19a)		<chem>O=N(=O)C1=CC2=C(C=C1)N(CC#C)C=C2</chem>	Synthesized (19a)
B20 (20)		<chem>[H]C(=O)C1=CN(CC#C)C2=C1C=C(C=C2)N(=O)=O</chem>	Synthesized (20)

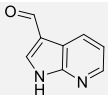
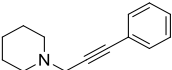
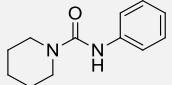
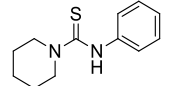
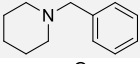
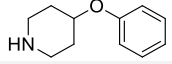
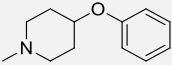
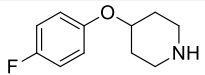
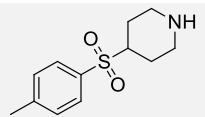
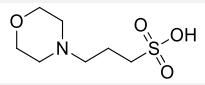
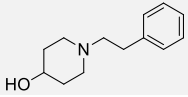
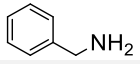
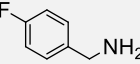
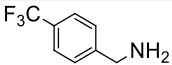
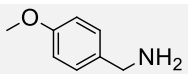
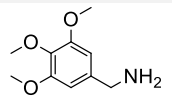
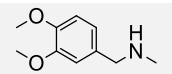
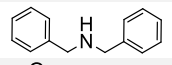
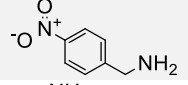
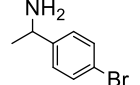
A Fragment-Based Approach for the Development of G-Quadruplex Ligands

B21 (21a)		<chem>CN(C)CC1=CN(CC#C)C2=C1C=C(C=C2)N(=O)=O</chem>	Synthesized (21a)
B22 (4b)		<chem>BrCCN1C=CC2=C1C=CC(=C2)N(=O)=O</chem>	Synthesized (4b)
B23 (4a)		<chem>BrCCCN1C=CC2=C1C=CC(=C2)N(=O)=O</chem>	Synthesized (4a)
B24 (14)		<chem>N=[N]=NCCCN1C=CC2=C1C=CC(=C2)N(=O)=O</chem>	Synthesized (14)
B25 (21b)		<chem>CCN(CC)CCNCC1=CN(CC#C)C2=C1C=C(C=C2)N(=O)=O</chem>	Synthesized (21b)
B26 (16)		<chem>CN(C)CC1=CN(CCCN=[N]#N)C2=C1C=C(C=C2)N(=O)=O</chem>	Synthesized (16)
B27 (6)		<chem>[H]C(=O)C1=CN(CCCN2CCCC2)C2=C1C=C(C=C2)N(=O)=O</chem>	Synthesized (6)
B28 (15)		<chem>[H]C(=O)C1=CN(CCCN=[N]#N)C2=C1C=C(C=C2)N(=O)=O</chem>	Synthesized (15)
B29		<chem>NC1=NC(N)=NC(O)=C1</chem>	Commercial
B30		<chem>NC1=CC(N)=NC(N)=N1</chem>	Commercial
C1		<chem>CC1=CNC(=O)NC1=O</chem>	Commercial
C2		<chem>O=C1NC=CC(=O)N1</chem>	Commercial
C3		<chem>OC1=CC(=O)NC(=O)N1</chem>	Commercial

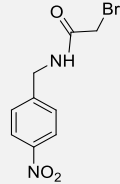
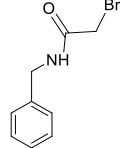
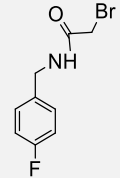
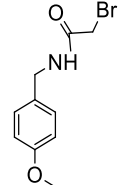
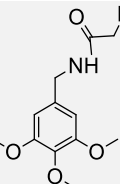
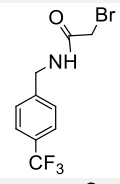
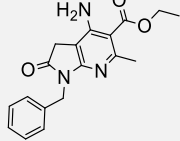
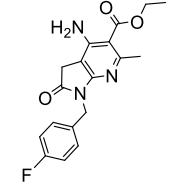
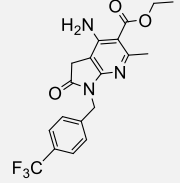
A Fragment-Based Approach for the Development of G-Quadruplex Ligands

C4		<chem>OC1=CC(O)=NC(=N1)C1=CC=CC=C1</chem>	Commercial
C5		<chem>NC1=NC(O)=CC=N1</chem>	Commercial
C6		<chem>C1CCNC1</chem>	Commercial
C8		<chem>NC1=NC=NC2=C1NN=C2 O.OS(=O)(=O)CCCN1CCOCC1</chem>	Commercial Commercial
C9		<chem>NC1=CC=C(C=C1)N1CCOCC1</chem>	Commercial
C10		<chem>C1CCN1CCOCC1</chem>	Commercial
C11		<chem>CN1CCCC1=O</chem>	Commercial
C12		<chem>CC1=CC=CC=C1NC(=O)CN1CCCC1</chem>	Commercial
C13		<chem>NCCN1CCCC1</chem>	Commercial
C14		<chem>O=C(ON1C(=O)CCC1=O)ON1C(=O)CCC1=O</chem>	Commercial
C15		<chem>CC1=CC(=O)C2=C(C=CC=C2)C1=O</chem>	Commercial
C16		<chem>O=C1C=CC(=O)C2=C1C=CC=C2</chem>	Commercial
C17		<chem>OC1=CC(=O)C2=C(C=CC=C2)C1=O</chem>	Commercial
C18		<chem>CC(C)CCCC(C)CCCC(C)CCC\C(C)=C\CC1=C(C)C(=O)C2=C(C=CC=C2)C1=O</chem>	Commercial
C19		<chem>OCC1=CC=C2C(=O)C3=C(C=CC=C3)C(=O)C2=C1</chem>	Commercial
C20		<chem>OC1=C2C(=O)C3=C(C=CC=C3)C(=O)C2=C(O)C=C1</chem>	Commercial
C21 (24a)		<chem>CN(C)CC1=CNC2=C1C=CC=N2</chem>	Synthesized (24a)
C22		<chem>C(OC1=NC=C2NC=CC2=C1)C1=CC=CC=C1</chem>	Commercial

A Fragment-Based Approach for the Development of G-Quadruplex Ligands

C23		<chem>O=CC1=CNC2=C1C=CC=N2</chem>	Commercial
C24		<chem>C(C#CC1=CC=CC=C1)N1CCCCC1</chem>	Commercial
C25		<chem>O=C(NC1=CC=CC=C1)N1CCCCC1</chem>	Commercial
C26		<chem>S=C(NC1=CC=CC=C1)N1CCCCC1</chem>	Commercial
C27		<chem>C(N1CCCCC1)C1=CC=CC=C1</chem>	Commercial
C28		<chem>C1CC(CCN1)OC1=CC=CC=C1</chem>	Commercial
C29		<chem>CN1CCC(CC1)OC1=CC=CC=C1</chem>	Commercial
C30		<chem>FC1=CC=C(OC2CCNCC2)C=C1</chem>	Commercial
D1		<chem>CC1=CC=C(C=C1)S(=O)(=O)C1CCNCC1</chem>	Commercial
D2		<chem>OS(=O)(=O)CCCN1CCOCC1</chem>	Commercial
D3		<chem>OC1CCN(CCC2=CC=CC=C2)CC1</chem>	Commercial
D4		<chem>NCC1=CC=CC=C1</chem>	Commercial
D5		<chem>NCC1=CC=C(F)C=C1</chem>	Commercial
D6		<chem>NCC1=CC=C(C=C1)C(F)(F)F</chem>	Commercial
D7		<chem>COC1=CC=C(CN)C=C1</chem>	Commercial
D8		<chem>COC1=CC(CN)=CC(OC)=C1OC</chem>	Commercial
D9		<chem>CNCC1=CC=C(OC)C(OC)=C1</chem>	Commercial
D10		<chem>C(NCC1=CC=CC=C1)C1=CC=CC=C1</chem>	Commercial
D11		<chem>NCC1=CC=C(C=C1)[N+](=[O-])=O</chem>	Commercial
D12		<chem>CC(N)C1=CC=C(Br)C=C1</chem>	Commercial

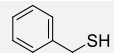
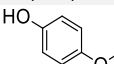
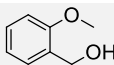
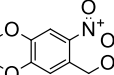
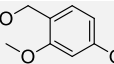
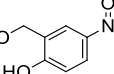
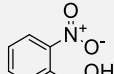
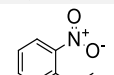
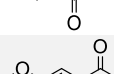
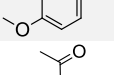
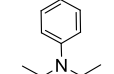
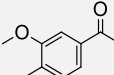
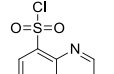
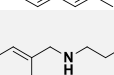
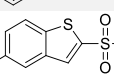
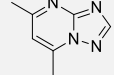
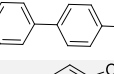
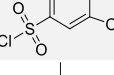
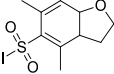
A Fragment-Based Approach for the Development of G-Quadruplex Ligands

D13		<chem>[O-][N+](=O)C1=CC=C(CNC(=O)CBr)C=C1</chem>	Commercial
D14 (26b)		<chem>BrCC(=O)NCC1=CC=CC=C1</chem>	Synthesized (26b)
D15 (27b)		<chem>FC1=CC=C(CNC(=O)CBr)C=C1</chem>	Synthesized (27b)
D16 (30b)		<chem>COC1=CC=C(CNC(=O)CBr)C=C1</chem>	Synthesized (30b)
D17 (29b)		<chem>COC1=CC(CNC(=O)CBr)=CC(OC)=C1OC</chem>	Synthesized (29b)
D18 (28b)		<chem>FC(F)(F)C1=CC=C(CNC(=O)CBr)C=C1</chem>	Synthesized (28b)
D19 (26d)		<chem>CCOC(=O)C1=C(N)C2=C(N=C1C)N(CC1=CC=CC=C1)C(=O)C2</chem>	Synthesized (26d)
D20 (27d)		<chem>CCOC(=O)C1=C(N)C2=C(N=C1C)N(CC1=CC=C(F)C=C1)C(=O)C2</chem>	Synthesized (27d)
D21 (28d)		<chem>CCOC(=O)C1=C(N)C2=C(N=C1C)N(CC1=CC=C(C=C1)C(F)(F)F)C(=O)C2</chem>	Synthesized (28d)

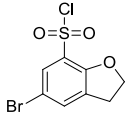
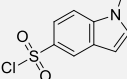
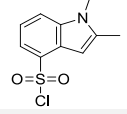
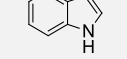
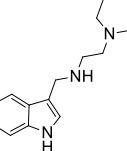
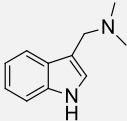
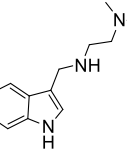
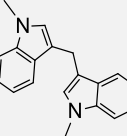
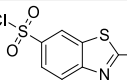
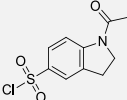
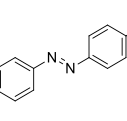
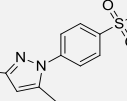
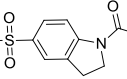
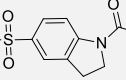
A Fragment-Based Approach for the Development of G-Quadruplex Ligands

D22 (29d)		<chem>CCOC(=O)C1=C(N)C2=C(N=C1C)N(CC1=CC(OC)=C(OC)C(OC)=C1)C(=O)C2</chem>	Synthesized (29d)
D23 (30d)		<chem>CCOC(=O)C1=C(N)C2=C(N=C1C)N(CC1=CC=C(OC)C=C1)C(=O)C2</chem>	Synthesized (30d)
D24 (26c)		<chem>NC1=C(CC(=O)N1)CC1=CC=CC=C1)C#N</chem>	Synthesized (26c)
D25 (27c)		<chem>NC1=C(CC(=O)N1)CC1=CC=C(F)C=C1)C#N</chem>	Synthesized (27c)
D26 (28c)		<chem>NC1=C(CC(=O)N1)CC1=CC=C(C=C1)C(F)(F)F)C#N</chem>	Synthesized (28c)
D27 (30c)		<chem>COC1=CC=C(CN2C(=O)CC(C#N)=C2N)C=C1</chem>	Synthesized (30c)
D28 (29c)		<chem>COC1=CC(CN2C(=O)CC(C#N)=C2N)=CC(OC)=C1OC</chem>	Synthesized (29c)
D29		<chem>[O-][N+](=O)C1=CC=CC=C1C=O</chem>	Commercial
D30		<chem>COC1=C(OC)C=C(C(=O)=C1)[N+](=[O-])=O</chem>	Commercial
E1		<chem>COC1=CC=C(C(=O)C=C1)[N+](=[O-])=O</chem>	Commercial
E2		<chem>COC1=C(C(=O)=CC=C1)[N+](=[O-])=O</chem>	Commercial
E3		<chem>CN(C)C1=CC=C(C(=O)C=C1)[N+](=[O-])=O</chem>	Commercial
E4		<chem>[O-][N+](=O)C1=CC(F)=CC=C1C=O</chem>	Commercial
E5		<chem>OCC1=CC=CC=C1</chem>	Commercial
E6		<chem>CC(O)C1=CC=CC=C1</chem>	Commercial

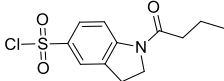
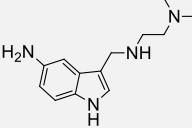
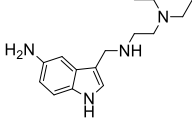

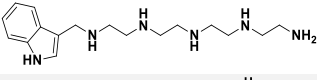
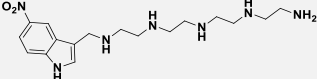
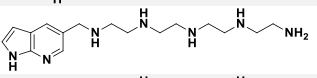
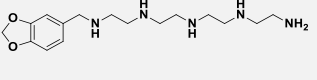
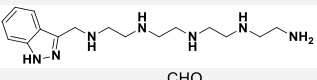






A Fragment-Based Approach for the Development of G-Quadruplex Ligands

E7		<chem>SCC1=CC=CC=C1</chem>	Commercial
E8		<chem>COC1=CC=C(O)C=C1</chem>	Commercial
E9		<chem>COC1=CC=CC=C1CO</chem>	Commercial
E10		<chem>COC1=C(OC)C=C(C(CO)=C1)[N+](=O)[O-]</chem>	Commercial
E11		<chem>COC1=CC=C(CO)C(OC)=C1</chem>	Commercial
E12		<chem>OCC1=CC(=CC=C1O)[N+](=O)[O-]</chem>	Commercial
E13		<chem>OCC1=CC=CC=C1[N+](=O)[O-]</chem>	Commercial
E14		<chem>CC(=O)C1=CC=CC=C1[N+](=O)[O-]</chem>	Commercial
E15		<chem>COC1=CC=C(C(C=O)C)C(C)=O</chem>	Commercial
E16		<chem>CCN(CC)C1=CC=C(C=C1)C(C)=O</chem>	Commercial
E17		<chem>COC1=CC(=CC=C1C)C(C)=O</chem>	Commercial
E18		<chem>CC1=CC2=CC=CC(=C2N=C1)S(=O)(=O)Cl</chem>	Commercial
E19		<chem>C(CN1CCCC1)NCC1=CC2=C(OCO2)C=C1</chem>	Commercial
E20		<chem>CC1=CC=C2SC(=CC2=C1)S(=O)(=O)Cl</chem>	Commercial
E21		<chem>CC1=NC2=NC=NN2C(C)=C1</chem>	Commercial
E22		<chem>CC1=CC=C(C=C1)C1=CC=C(C=C1)S(=O)(=O)Cl</chem>	Commercial
E23		<chem>ClS(=O)(=O)C1=CC=C2OCOC2=C1</chem>	Commercial
E24		<chem>CC1=C(C(C)=C(C)C2OC(C)CC12)S(=O)(=O)Cl</chem>	Commercial
E25		<chem>CC1=C(C(C)=C(C)C2OC(C)CC12)S(=O)(=O)Cl</chem>	Commercial

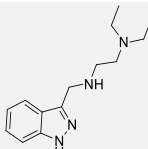
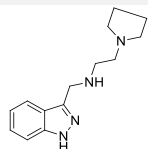
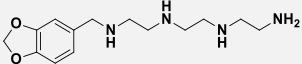
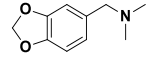
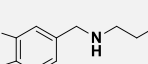
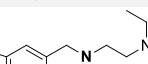
A Fragment-Based Approach for the Development of G-Quadruplex Ligands

E26		<chem>ClS(=O)(=O)C1=C2OCCC2=CC(Br)=C1</chem>	Commercial
E27		<chem>CN1C(C)=CC2=C1C=CC=C2S(Cl)(=O)=O</chem>	Commercial
E28		<chem>CN1C(C)=CC2=C1C=CC=C2S(Cl)(=O)=O</chem>	Commercial
E29		<chem>N1C=CC2=CC=CC=C12</chem>	Commercial
E30		<chem>CCN(CC)CCNCC1=CNC2=CC=CC=C12</chem>	Commercial
F1 (3a)		<chem>CN(C)CC1=CNC2=CC=CC=C12</chem>	Synthesized (3a)
F2		<chem>CN(C)CCNCC1=CNC2=CC=CC=C12</chem>	Commercial
F3		<chem>CN1C=C(CC2=CC3C=CC=CC3N2C)C2=C1C=CC=C2</chem>	Commercial
F4		<chem>CC1=NC2=CC=C(C=C2S1)S(Cl)(=O)=O</chem>	Commercial
F5		<chem>CC(=O)N1CCC2=C1C=CC(=C2)S(Cl)(=O)=O</chem>	Commercial
F6		<chem>CN(C)C1=CC=C(C=C1)\N=N\C1=CC=C(C=C1)S(Cl)(=O)=O</chem>	Commercial
F7		<chem>CC1=NN(C(C)=C1)C1=CC=C(C=C1)S(Cl)(=O)=O</chem>	Commercial
F8		<chem>CCC(=O)N1CCC2=C1C=CC(=C2)S(Cl)(=O)=O</chem>	Commercial
F9		<chem>OC(=O)N1CCC2=C1C=CC(=C2)S(Cl)(=O)=O</chem>	Commercial

A Fragment-Based Approach for the Development of G-Quadruplex Ligands

F10		<chem>CCCC(=O)N1CCC2=C1C=CC(=C2)S(Cl)(=O)=O</chem>	Commercial
F11		<chem>CN(C)CCNCC1=CNC2=CC=C(N)C=C12</chem>	Commercial
F12 (32)		<chem>CCN(CC)CCNCC1=CNC2=CC=C(N)C=C12</chem>	Synthesized (32)
F13		<chem>OCC1=CC=CC=C1[N+](C)C</chem>	Commercial
F14 (3b)		<chem>NCCNCCNCCNCC1=CNC2=CC=CC=C12</chem>	Synthesized (3b)
F15 (13a)		<chem>NCCNCCNCCNCC1=CNC2=CC=CC=C12</chem>	Synthesized (13a)
F16 (24b)		<chem>NCCNCCNCCNCC1=CC2=C(NC=C2)N=C1</chem>	Synthesized (24b)
F17 (25c)		<chem>NCCNCCNCCNCC1=CC2=C(OC2)C=C1</chem>	Synthesized (25c)
F18 (23c)		<chem>NCCNCCNCCNCC1=NNC2=CC=CC=C12</chem>	Synthesized (23c)
F19		<chem>O=CC1=CNC2=NC=CC=C12</chem>	Commercial
F20 (24a)		<chem>CN(C)CC1=CNC2=NC=CC=C12</chem>	Synthesized (24a)
F21		<chem>CN(C)CCNCC1=CNC2=NC=CC=C12</chem>	Commercial
F22		<chem>CCN(CC)CCNCC1=CNC2=NC=CC=C12</chem>	Commercial
F23 (24f)		<chem>C(CN1CCCC1)NCC1=CNC2=NC=CC=C12</chem>	Synthesized (24f)
F24		<chem>CN(C)CCNCC1=NNC2=CC=CC=C12</chem>	Commercial

A Fragment-Based Approach for the Development of G-Quadruplex Ligands

F25		<chem>CCN(CC)CCNCC1=NNC2=CC=CC=C12</chem>	Commercial
F26 (23f)		<chem>C(CN1CCCC1)NCC1=NNC2=CC=CC=C12</chem>	Synthesized (23f)
F27 (25c)		<chem>NCCNCCNCCNCC1=CC2=C(OCO2)C=C1</chem>	Synthesized (25c)
F28		<chem>CN(C)CC1=CC2=C(OCO2)C=C1</chem>	Commercial
F29		<chem>CN(C)CCNCC1=CC2=C(OCO2)C=C1</chem>	Commercial
F30 (25e)		<chem>CCN(CC)CCNCC1=CC2=C(OCO2)C=C1</chem>	Synthesized (25e)

2.8 SWISS ADME: Profiling of the optimized and selected fragments

The SWISS-ADME tool is used in our current investigations for the analysis of the physico-chemical, pharmacokinetics and medicinal chemistry properties of all the commercial and optimized fragments generated via FBDD fragment expansion strategy. The physico-chemical properties basically includes the criteria like molecular weight, number of rotatable bonds, Fraction Csp3, number of H-bond acceptors and H-bond donors all the these criteria are utmost important for the design of small drug-like fragments (Table 5). Furthermore, all the pharmacokinetics, drug-likeness and medicinal chemistry properties like probability of GI absorption, blood-brain-barrier permeability (BBB permeant), ability of fragments to involved in substrate reaction in human body *i.e* Pgp substrate reaction ability and the possible interactions of designed fragments with the intestinal CYP-P450 enzymes are analyzed, In addition to all the above parameters fragments were all also screened to check if there are any Lipinski violations and drug-likeness violations (Table 6). The lipophilicity and water solubility for example iLogP, WLogP, ESOL Class (solubility class), were also determined by SWISS-ADME filters and are reported in table 7.

Table 5. Physiochemical Properties of Fragments (SWISS-ADME-Filters)

	Formula	MW	#Heavy atoms	#Aromatic heavy atoms	Fraction Csp3	#Rotatable bonds	#H-bond acceptors	#H-bond donors	MR	TPSA
A1	C ₉ H ₁₀ N ₂	146.19	11	9	0.11	0	0	1	47.6	30.95
A2	C ₉ H ₁₀ N ₂	146.02	11	9	0.11	0	0	2	47.67	41.81
A3	C ₈ H ₈ N ₂	132.16	10	9	0	0	0	2	42.7	41.81
A4	C ₁₀ H ₁₂ N ₂	160.22	12	9	0.2	0	0	1	52.57	30.95
A5	C ₁₂ H ₁₆ N ₂ O	204.27	15	9	0.33	3	1	1	63.27	40.18
A6	C ₁₅ H ₁₄ N ₂	222.29	17	15	0.07	2	0	1	72.09	30.95
A7	C ₁₀ H ₁₃ N ₃	175.23	13	9	0.3	1	1	1	55.17	43.84
A8	C ₈ H ₈ N ₂	132.16	10	9	0	0	0	2	42.7	41.81
A9	C ₁₄ H ₁₃ N ₃ O	239.27	18	15	0.07	2	2	2	71.91	64.07
A10	C ₉ H ₁₁ N ₃	161.2	12	9	0.11	1	1	3	50.38	67.83
A11	C ₈ H ₉ N ₃	147.18	11	9	0.12	0	1	1	45.4	43.84
A12	C ₉ H ₁₀ N ₂	146.19	11	9	0.11	0	0	1	47.6	30.95
A13	C ₁₇ H ₁₆ N ₂	248.32	19	9	0.18	2	0	2	82.79	27.82
A14	C ₁₂ H ₁₇ N ₃	203.28	15	9	0.33	2	1	2	65.05	56.97
A15	C ₈ H ₆ N ₂ O	146.15	11	9	0	1	2	1	41.48	45.75
A16	C ₁₅ H ₁₇ N ₃ S	271.38	19	14	0.27	2	1	1	82.66	72.08
A17	C ₁₁ H ₁₂ N ₂ O	188.23	14	9	0.18	1	1	1	57.96	48.02
A18	C ₁₄ H ₁₄ C ₁ N ₃ S	291.8	19	14	0.21	2	1	1	82.71	72.08
A19	C ₁₁ H ₁₁ NO	173.21	13	9	0.18	1	1	0	53.55	22
A20	C ₁₀ H ₁₀ N ₂ O ₂	190.2	14	9	0.1	1	2	2	54.56	68.25
A21	C ₁₁ H ₁₂ N ₂ O ₂	204.23	15	9	0.18	2	2	2	59.15	68.25
A22	C ₁₈ H ₂₀ N ₂ O	280.36	21	15	0.22	5	2	1	86.75	28.26
A23	C ₈ H ₆ N ₂ O	146.15	7	5	0	1	1	1	26.18	32.86
A24	C ₁₃ H ₁₇ N ₃	215.29	16	9	0.38	1	1	2	71.44	42.98
A25	C ₁₄ H ₁₁ N ₃ O ₃	239.27	18	15	0.07	2	2	2	71.91	64.07
A26	C ₁₀ H ₁₃ N ₃	175.11	12	9	0.11	1	1	3	50.38	67.83
A27	C ₈ H ₇ N ₃ O ₂	147.18	11	9	0.12	0	1	1	45.4	43.84
A28	C ₈ H ₉ N ₃	147.18	11	9	0.12	0	1	1	45.4	43.84
A29	C ₁₇ H ₁₆ N ₂	248.32	19	9	0.18	2	0	2	82.79	27.82
A30	C ₁₂ H ₁₇ N ₃	203.28	15	9	0.33	2	1	2	65.05	56.97
B1	C ₇ H ₆ N ₂	118.14	9	9	0	0	0	0	35.19	17.3
B2	C ₁₅ H ₁₇ N ₃ S	271.38	19	14	0.27	2	1	1	82.66	72.08
B3	C ₁₀ H ₁₀ N ₂ O	174.2	13	9	0.1	1	1	1	52.99	48.02
B4	C ₁₄ H ₁₄ ClN ₃ S	291.8	19	14	0.21	2	1	1	82.71	72.08
B5	C ₉ H ₈ N ₂ O	160.17	12	9	0	1	1	2	48.09	58.88
B6	C ₁₀ H ₁₀ N ₂ O ₂	190.2	14	9	0.1	1	2	2	54.56	68.25
B7	C ₁₁ H ₁₂ N ₂ O ₂	204.23	15	9	0.18	2	2	2	59.15	68.25
B8	C ₁₉ H ₂₂ N ₂ O	294.39	22	15	0.26	5	2	0	91.65	17.4
B9	C ₈ H ₆ N ₂ O	146.15	11	9	0	1	2	0	40.58	34.37
B10	C ₉ H ₁₀ N ₂	146.19	11	9	0.11	0	0	1	47.6	30.95
B11	C ₁₁ H ₁₃ N ₃ O ₂	219.24	16	9	0.27	3	3	1	64.6	64.85
B12	C ₁₂ H ₁₆ N ₄ O ₂	248.28	18	9	0.33	5	4	2	72.21	76.88
B13	C ₁₅ H ₂₀ N ₄ O ₂	288.34	21	9	0.47	6	4	2	88.43	76.88
B14	C ₁₅ H ₂₂ N ₄ O ₂	290.36	21	9	0.47	8	4	2	86.63	76.88
B15	C ₂₁ H ₃₁ N ₅ O ₂	385.5	28	9	0.62	9	5	1	122.06	69.26
B16	C ₁₉ H ₁₈ N ₄ O ₄	366.37	27	9	0.26	6	4	0	109.36	99.81
B17	C ₁₈ H ₁₆ N ₄ O ₄	352.34	26	9	0.22	5	4	0	104.55	99.81
B18	C ₁₅ H ₁₉ N ₃ O ₂	273.33	20	9	0.47	5	3	0	85.56	53.99
B19	C ₁₁ H ₉ N ₂ O ₂	200.19	15	9	0.09	2	2	0	59.8	50.75
B20	C ₁₂ H ₈ N ₂ O ₃	228.2	17	9	0.08	3	3	0	65.19	67.82
B21	C ₁₄ H ₁₅ N ₃ O ₂	257.29	19	9	0.29	4	3	0	77.27	53.99
B22	C ₁₀ H ₉ BrN ₂ O ₂	269.09	15	9	0.2	3	2	0	64.7	50.75
B23	C ₁₁ H ₁₁ BrN ₂ O ₂	283.12	16	9	0.27	4	2	0	69.51	50.75
B24	C ₁₁ H ₁₂ N ₅ O ₂	246.25	18	9	0.27	5	5	1	68.53	86.96
B25	C ₁₈ H ₂₄ N ₄ O ₂	328.41	24	9	0.44	9	4	1	99.3	66.02
B26	C ₁₄ H ₁₈ N ₆ O ₂	302.33	22	9	0.43	7	6	0	83.93	103.74
B27	C ₁₆ H ₁₉ N ₃ O ₃	301.34	22	9	0.44	6	4	0	90.95	71.06
B28	C ₁₂ H ₁₁ N ₅ O ₃	273.25	20	9	0.25	6	6	0	71.84	117.57
B29	C ₄ H ₆ N ₄ O	126.12	9	6	0	0	3	3	32.86	98.05
B30	C ₄ H ₇ N ₅	125.13	9	6	0	0	2	3	35.25	103.84
C1	C ₅ H ₆ N ₂ O ₂	126.11	9	6	0.2	0	2	2	32.65	65.72
C2	C ₄ H ₄ N ₂ O ₂	112.09	8	6	0	0	2	2	27.68	65.72
C3	C ₄ H ₄ N ₂ O ₃	128.09	9	6	0	0	3	3	29.71	85.95
C4	C ₁₀ H ₈ N ₂ O ₂	188.18	14	12	0	1	4	2	51.51	66.24
C5	C ₄ H ₅ N ₃ O	111.1	8	6	0	0	3	2	28.46	72.03

A Fragment-Based Approach for the Development of G-Quadruplex Ligands

C6	C ₄ H ₉ N	71.12	5	0	1	0	1	1	25.94	12.03
C7	C ₅ H ₅ N ₅	135.13	10	9	0	0	3	2	36.09	80.48
C8	C ₇ H ₁₇ NO ₅ S	227.28	14	0	1	4	6	2	55.13	84.45
C9	C ₁₀ H ₁₄ N ₂ O	178.23	13	6	0.4	1	1	1	56.57	38.49
C10	C ₆ H ₁₂ ClNO	149.62	9	0	1	2	2	0	41.53	12.47
C11	C ₅ H ₉ NO	99.13	7	0	0.8	0	1	0	31.05	20.31
C12	C ₁₃ H ₁₈ N ₂ O	218.29	16	6	0.46	4	2	1	69.65	32.34
C13	C ₆ H ₁₄ N ₂	114.19	8	0	1	2	2	1	38.36	29.26
C14	C ₉ H ₈ N ₂ O ₇	256.17	18	0	0.44	4	7	0	58.33	110.29
C15	C ₁₁ H ₈ O ₂	172.18	13	6	0.09	0	2	0	49.05	34.14
C16	C ₁₀ H ₆ O ₂	158.15	12	6	0	0	2	0	44.24	34.14
C17	C ₁₀ H ₆ O ₃	174.15	13	6	0	0	3	1	45.81	54.37
C18	C ₃₁ H ₄₆ O ₂	450.7	33	6	0.61	14	2	0	144.72	34.14
C19	C ₁₅ H ₁₀ O ₃	238.24	18	12	0.07	1	3	1	65.88	54.37
C20	C ₁₄ H ₈ O ₄	240.21	18	12	0	0	4	2	63.8	74.6
C21	C ₁₀ H ₁₃ N ₃	175.23	13	9	0.3	2	2	1	53.57	31.92
C22	C ₁₄ H ₁₂ N ₂ O	224.26	17	15	0.07	3	2	1	67.07	37.91
C23	C ₈ H ₆ N ₂ O	146.15	11	9	0	1	2	1	41.48	45.75
C24	C ₁₄ H ₁₇ N	199.29	15	6	0.43	1	1	0	67.92	3.24
C25	C ₁₂ H ₁₆ N ₂ O	204.27	15	6	0.42	3	1	1	65.07	32.34
C26	C ₁₂ H ₁₆ N ₂ S	220.33	15	6	0.42	3	0	1	72.27	47.36
C27	C ₁₂ H ₁₇ N	175.27	13	6	0.5	2	1	0	60.14	3.24
C28	C ₁₁ H ₁₅ NO	177.24	13	6	0.45	2	2	1	56.76	21.26
C29	C ₁₂ H ₁₇ NO	191.27	14	6	0.5	2	2	0	61.67	12.47
C30	C ₁₁ H ₁₄ FNO	195.23	14	6	0.45	2	3	1	56.72	21.26
D1	C ₁₂ H ₁₇ NO ₂ S	239.33	16	6	0.5	2	3	1	68.33	54.55
D2	C ₇ H ₁₅ NO ₄ S	209.26	13	0	1	4	5	1	52.08	75.22
D3	C ₁₃ H ₁₉ NO	205.3	15	6	0.54	3	2	1	66.11	23.47
D4	C ₇ H ₉ N	107.15	8	6	0.14	1	1	1	34.12	26.02
D5	C ₇ H ₈ FN	125.14	9	6	0.14	1	2	1	34.07	26.02
D6	C ₈ H ₈ F ₃ N	175.15	12	6	0.25	2	4	1	39.12	26.02
D7	C ₈ H ₁₁ NO	137.18	10	6	0.25	2	2	1	40.61	35.25
D8	C ₁₀ H ₁₅ NO ₃	197.23	14	6	0.4	4	4	1	53.59	53.71
D9	C ₁₀ H ₁₅ NO ₂	181.23	13	6	0.4	4	3	1	52	30.49
D10	C ₁₄ H ₁₅ N	197.28	15	12	0.14	4	1	1	63.5	12.03
D11	C ₇ H ₈ N ₂ O ₂	152.15	11	6	0.14	2	3	1	42.94	71.84
D12	C ₈ H ₁₀ BrN	200.08	10	6	0.25	1	1	1	46.62	26.02
D13	C ₉ H ₉ BrN ₂ O ₃	273.08	15	6	0.22	5	3	1	60.72	74.92
D14	C ₉ H ₁₀ BrNO	228.09	12	6	0.22	4	1	1	51.89	29.1
D15	C ₉ H ₉ BrFNO	246.08	13	6	0.22	4	2	1	51.85	29.1
D16	C ₁₀ H ₁₂ BrNO ₂	258.11	14	6	0.3	5	2	1	58.39	38.33
D17	C ₁₂ H ₁₆ BrNO ₄	318.16	18	6	0.42	7	4	1	71.37	56.79
D18	C ₁₀ H ₉ BrF ₃ NO	296.08	16	6	0.3	5	4	1	56.9	29.1
D19	C ₁₈ H ₁₉ N ₃ O ₃	325.36	24	12	0.28	5	4	1	94.37	85.52
D20	C ₁₈ H ₁₈ FN ₃ O ₃	343.35	25	12	0.28	5	5	1	94.33	85.52
D21	C ₁₉ H ₁₈ F ₃ N ₃ O ₃	393.36	28	12	0.32	6	7	1	99.38	85.52
D22	C ₂₁ H ₂₅ N ₃ O ₆	415.44	30	12	0.38	8	7	1	113.85	113.21
D23	C ₁₉ H ₂₁ N ₃ O ₄	355.39	26	12	0.32	6	5	1	100.87	94.75
D24	C ₁₂ H ₁₁ N ₃ O	213.24	16	6	0.17	2	2	1	62.32	70.12
D25	C ₁₂ H ₁₀ FN ₃ O	231.23	17	6	0.17	2	3	1	62.28	70.12
D26	C ₁₃ H ₁₀ F ₃ N ₃ O	281.23	20	6	0.23	3	5	1	67.32	70.12
D27	C ₁₃ H ₁₃ N ₃ O ₂	243.26	18	6	0.23	3	3	1	68.81	79.35
D28	C ₁₅ H ₁₇ N ₃ O ₄	303.31	22	6	0.33	5	5	1	81.8	97.81
D29	C ₇ H ₅ NO ₃	151.12	11	6	0	2	3	0	40.65	62.89
D30	C ₉ H ₉ NO ₅	211.17	15	6	0.22	4	5	0	53.64	81.35
E1	C ₈ H ₇ NO ₄	181.15	13	6	0.12	3	4	0	47.14	72.12
E2	C ₈ H ₇ NO ₄	181.15	13	6	0.12	3	4	0	47.14	72.12
E3	C ₉ H ₁₀ N ₂ O ₃	194.19	14	6	0.22	3	3	0	54.86	66.13
E4	C ₇ H ₄ FNO ₃	169.11	12	6	0	2	4	0	40.61	62.89
E5	C ₇ H ₈ O	108.14	8	6	0.14	1	1	1	32.57	20.23

A Fragment-Based Approach for the Development of G-Quadruplex Ligands

E6	C ₈ H ₁₀ O	122.16	9	6	0.25	1	1	1	37.38	20.23
E7	C ₇ H ₈ S	124.2	8	6	0.14	1	0	0	39.34	38.8
E8	C ₇ H ₈ O ₂	124.14	9	6	0.14	1	2	1	34.96	29.46
E9	C ₈ H ₁₀ O ₂	138.16	10	6	0.25	2	2	1	39.06	29.46
E10	C ₉ H ₁₁ NO ₅	213.19	15	6	0.33	4	5	1	54.38	84.51
E11	C ₉ H ₁₂ O ₃	168.19	12	6	0.33	3	3	1	45.55	38.69
E12	C ₇ H ₇ NO ₄	169.13	12	6	0.14	2	4	2	43.41	86.28
E13	C ₇ H ₇ NO ₃	153.14	11	6	0.14	2	3	1	41.39	66.05
E14	C ₈ H ₇ NO ₃	165.15	12	6	0.12	2	3	0	45.46	62.89
E15	C ₁₀ H ₁₂ O ₃	180.2	13	6	0.3	3	3	0	49.62	35.53
E16	C ₁₂ H ₁₇ NO	191.27	14	6	0.42	4	1	0	60.46	20.31
E17	C ₁₀ H ₁₂ O ₂	164.2	12	6	0.3	2	2	0	48.09	26.3
E18	C ₁₀ H ₈ ClNO ₂ S	241.69	15	10	0.1	1	3	0	59.79	55.41
E19	C ₉ H ₇ ClO ₂ S ₂	246.73	14	9	0.11	1	2	0	59.87	70.76
E20	C ₇ H ₈ N ₄	148.17	11	9	0.29	0	3	0	40.71	43.08
E21	C ₁₃ H ₁₁ ClO ₂ S	266.74	17	12	0.08	2	2	0	69.93	42.52
E22	C ₇ H ₅ ClO ₄ S	220.63	13	6	0.14	1	4	0	45.59	60.98
E23	C ₁₃ H ₁₉ ClO ₃ S	290.81	18	0	0.69	1	3	0	74.31	51.75
E24	C ₁₃ H ₁₉ ClO ₃ S	290.81	18	0	0.69	1	3	0	74.31	51.75
E25	C ₈ H ₆ BrClO ₃ S	297.55	14	6	0.25	1	3	0	56.57	51.75
E26	C ₁₀ H ₁₀ ClNO ₂ S	243.71	15	9	0.2	1	2	0	61.25	47.45
E27	C ₈ H ₇ N	117.15	9	9	0	0	0	1	38.3	15.79
E28	C ₁₅ H ₂₃ N ₃	245.36	18	9	0.47	7	2	2	77.8	31.06
E29	C ₁₁ H ₁₄ N ₂	174.24	13	9	0.27	2	1	1	55.77	19.03
E30	C ₁₃ H ₁₉ N ₃	217.31	16	9	0.38	5	2	2	68.19	31.06
F1	C ₁₇ H ₁₄ N ₂	246.31	19	18	0.06	2	0	2	79.61	31.58
F2	C ₁₃ H ₁₉ N ₃	217.6	14	9	0.12	1	3	0	57.67	83.65
F3	C ₁₉ H ₁₈ N ₂	259.71	16	6	0.3	2	3	0	64.53	62.83
F4	C ₁₄ H ₁₄ ClN ₃ O ₂ S	323.8	21	12	0.14	4	4	0	84.35	70.48
F5	C ₁₁ H ₁₁ ClN ₂ O ₂ S	270.74	17	11	0.18	2	3	0	66.58	60.34
F6	C ₁₁ H ₁₂ ClNO ₃ S	273.74	17	6	0.36	3	3	0	69.33	62.83
F7	C ₉ H ₈ ClNO ₄ S	261.68	16	6	0.22	2	4	1	61.68	83.06
F8	C ₁₂ H ₁₄ ClNO ₃ S	287.76	18	6	0.42	4	3	0	74.14	62.83
F9	C ₁₅ H ₂₄ N ₄	260.38	19	9	0.47	7	2	3	82.21	57.08
F10	C ₁₇ H ₃₀ N ₆	318.46	23	9	0.53	13	5	6	95.63	89.93
F11	C ₁₅ H ₂₅ N ₅	275.39	20	9	0.47	10	4	5	83.22	77.9
F12	C ₁₆ H ₂₉ N ₇	319.45	23	9	0.56	13	6	6	93.43	102.82
F13	C ₁₆ H ₂₉ N ₅ O ₂	323.43	23	6	0.62	13	7	5	89.84	92.6
F14	C ₁₆ H ₂₉ N ₇	319.45	23	9	0.56	13	6	6	93.43	102.82
F15	C ₈ H ₆ N ₂ O	146.15	11	9	0	1	2	1	41.48	45.75
F16	C ₁₀ H ₁₃ N ₃	175.23	13	9	0.3	2	2	1	53.57	31.92
F17	C ₁₂ H ₁₈ N ₄	218.3	16	9	0.42	5	3	2	65.99	43.95
F18	C ₁₄ H ₂₂ N ₄	246.35	18	9	0.5	7	3	2	75.6	43.95
F19	C ₁₄ H ₂₀ N ₄	244.34	18	9	0.5	5	3	2	77.4	43.95
F20	C ₁₂ H ₁₈ N ₄	218.3	16	9	0.42	5	3	2	65.99	43.95
F21	C ₁₄ H ₂₂ N ₄	246.35	18	9	0.5	7	3	2	75.6	43.95
F22	C ₁₄ H ₂₀ N ₄	244.34	18	9	0.5	5	3	2	77.4	43.95
F23	C ₁₄ H ₂₄ N ₄ O ₂	280.37	20	6	0.57	10	6	4	77.43	80.57
F24	C ₁₀ H ₁₃ NO ₂	179.22	13	6	0.4	2	3	0	49.98	21.7
F25	C ₁₂ H ₁₈ N ₂ O ₂	222.28	16	6	0.5	5	4	1	62.4	33.73
F26	C ₁₄ H ₂₂ N ₂ O ₂	250.34	18	6	0.57	7	4	1	72.01	33.73
F27	C ₁₄ H ₂₀ N ₂ O ₂	248.32	18	6	0.57	5	4	1	73.81	33.73
F28	C ₁₀ H ₁₃ NO ₂	179.09	18	9	0.47	7	2	2	77.8	31.06
F29	C ₁₂ H ₁₈ N ₂ O ₂	222.14	13	9	0.27	2	1	1	55.77	19.03
F30	C ₁₄ H ₂₂ N ₂ O ₂	250.17	16	9	0.38	5	2	2	68.19	31.06

A Fragment-Based Approach for the Development of G-Quadruplex Ligands

Table 6. Pharmacokinetics, Drug-Likeliness and Medicinal Chemistry (SWISS-ADME-Filters)

	GI absorption	BBB permeant	Pgp substrate	CYP1A2 inhibitor	CYP2C19 inhibitor	CYP2C9 inhibitor	CYP2D6 inhibitor	CYP3A4 inhibitor	log Kp (cm/s)	Lipinski #violations	Lead-likeness #violations
A1	High	Yes	No	Yes	No	No	No	No	-5.74	0	1
A2	High	Yes	No	Yes	No	No	No	No	-5.94	0	1
A3	High	Yes	No	Yes	No	No	No	No	-6.61	0	1
A4	High	Yes	No	Yes	No	No	No	No	-5.76	0	1
A5	High	Yes	No	Yes	No	No	No	No	-6.47	0	1
A6	High	Yes	Yes	Yes	No	No	No	No	-9.64	0	1
A7	High	Yes	No	Yes	No	No	No	No	-6.24	0	1
A8	High	Yes	No	Yes	No	No	No	No	-6.61	0	1
A9	High	Yes	Yes	Yes	Yes	No	No	No	-6.01	0	1
A10	High	No	No	Yes	No	No	No	No	-7.12	0	1
A11	High	Yes	No	Yes	No	No	No	No	-6.66	0	1
A12	High	Yes	No	Yes	No	No	No	No	-6.25	0	1
A13	High	Yes	Yes	Yes	Yes	Yes	Yes	No	-5.39	0	1
A14	High	Yes	Yes	No	No	No	No	No	-6.8	0	1
A15	High	Yes	No	Yes	No	No	No	No	-6.23	0	1
A16	High	Yes	Yes	Yes	Yes	Yes	No	Yes	-5.75	0	0
A17	High	Yes	No	Yes	No	No	No	No	-6.61	0	1
A18	High	Yes	Yes	Yes	Yes	Yes	No	Yes	-5.47	0	1
A19	High	Yes	No	Yes	No	No	No	No	-6.06	0	1
A20	High	Yes	No	No	No	No	No	No	-6.52	0	1
A21	High	Yes	No	No	No	No	No	No	-6.99	0	1
A22	High	Yes	Yes	Yes	Yes	No	Yes	Yes	-5.63	0	0
A23	High	Yes	No	No	No	No	No	No	-6.76	0	1
A24	High	Yes	Yes	Yes	No	No	No	No	-6.65	0	1
A25	High	Yes	Yes	Yes	Yes	No	No	No	-6.01	0	1
A26	High	No	No	Yes	No	No	No	No	-7.12	0	1
A27	High	Yes	No	Yes	No	No	No	No	-6.66	0	1
A28	High	Yes	No	Yes	No	No	No	No	-6.66	0	1
A29	High	Yes	Yes	Yes	Yes	Yes	Yes	No	-5.39	0	1
A30	High	Yes	Yes	No	No	No	No	No	-6.8	0	1
B1	High	Yes	No	No	No	No	No	No	-5.71	0	1
B2	High	Yes	Yes	Yes	Yes	Yes	No	Yes	-5.75	0	0
B3	High	Yes	No	Yes	No	No	No	No	-6.81	0	1
B4	High	Yes	Yes	Yes	Yes	Yes	No	Yes	-5.47	0	1
B5	High	Yes	No	Yes	No	No	No	No	-6.69	0	1
B6	High	Yes	No	No	No	No	No	No	-6.52	0	1
B7	High	Yes	No	No	No	No	No	No	-6.99	0	1
B8	High	Yes	No	Yes	Yes	No	Yes	Yes	-5.75	0	0
B9	High	Yes	No	Yes	No	No	No	No	-6.03	0	1
B10	High	Yes	No	Yes	No	No	No	No	-5.74	0	1
B11	High	Yes	No	Yes	No	No	No	No	-6.42	0	1
B12	High	No	No	No	No	No	No	No	-6.83	0	1
B13	High	Yes	Yes	No	No	No	Yes	No	-6.8	0	0
B14	High	Yes	No	Yes	No	No	Yes	No	-6.64	0	1
B15	High	Yes	Yes	No	No	No	Yes	Yes	-7.1	0	2
B16	High	No	No	No	Yes	Yes	No	Yes	-6.11	0	1
B17	High	No	No	No	Yes	Yes	No	Yes	-6.28	0	1
B18	High	Yes	No	Yes	Yes	No	Yes	No	-6.09	0	0
B19	High	Yes	No	Yes	No	No	No	No	-6.07	0	1
B20	High	Yes	No	Yes	No	No	No	No	-6.62	0	1
B21	High	Yes	No	Yes	Yes	No	No	No	-6.53	0	0
B22	High	Yes	No	Yes	Yes	No	No	No	-6.15	0	0
B23	High	Yes	No	Yes	Yes	No	No	No	-5.98	0	0
B24	High	No	No	Yes	No	Yes	No	No	-5.85	0	1
B25	High	Yes	No	No	Yes	No	Yes	No	-6.76	0	1
B26	High	No	No	Yes	Yes	Yes	No	No	-6.11	0	0
B27	High	Yes	No	No	Yes	No	No	No	-6.64	0	0
B28	High	No	No	Yes	Yes	No	No	No	-6.2	0	0
B29	High	No	No	No	No	No	No	No	-7.43	0	1
B30	High	No	No	No	No	No	No	No	-7.66	0	1
C1	High	No	No	No	No	No	No	No	-7.51	0	1

A Fragment-Based Approach for the Development of G-Quadruplex Ligands

C2	High	No	No	No	No	No	No	No	-7.74	0	1
C3	High	No	No	No	No	No	No	No	-8	0	1
C4	High	Yes	No	Yes	No	No	No	No	-5.89	0	1
C5	High	No	No	No	No	No	No	No	-7.1	0	1
C6	Low	No	No	No	No	No	No	No	-6.41	0	1
C7	High	No	No	No	No	No	No	No	-7.43	0	1
C8	High	No	No	No	No	No	No	No	-10.26	0	1
C9	High	Yes	Yes	No	No	No	No	No	-6.9	0	1
C10	Low	No	No	No	No	No	No	No	-6.87	0	1
C11	Low	No	No	No	No	No	No	No	-7.29	0	1
C12	High	Yes	No	No	No	No	Yes	No	-6.28	0	1
C13	High	No	No	No	No	No	No	No	-6.94	0	1
C14	High	No	No	No	No	No	No	No	-8.75	0	0
C15	High	Yes	No	Yes	No	No	No	No	-5.79	0	1
C16	High	Yes	No	Yes	No	No	No	No	-6.05	0	1
C17	High	Yes	No	Yes	No	No	No	No	-6.38	0	1
C18	Low	No	Yes	Yes	No	No	No	No	-1.3	1	3
C19	High	Yes	No	Yes	No	No	No	No	-5.84	0	1
C20	High	Yes	No	Yes	No	No	No	Yes	-5.17	0	2
C21	High	Yes	No	Yes	No	No	No	No	-6.55	0	1
C22	High	Yes	Yes	Yes	Yes	No	Yes	Yes	-5.69	0	1
C23	High	Yes	No	Yes	No	No	No	No	-6.64	0	1
C24	High	Yes	No	No	No	No	Yes	No	-5.39	0	1
C25	High	Yes	No	No	No	No	No	No	-6.16	0	1
C26	High	Yes	No	Yes	Yes	No	Yes	No	-5.83	0	1
C27	Low	Yes	No	No	No	No	Yes	No	-5.52	0	1
C28	High	Yes	No	No	No	No	No	No	-5.95	0	1
C29	High	Yes	No	No	No	No	Yes	No	-5.71	0	1
C30	High	Yes	No	No	No	No	Yes	No	-5.99	0	1
D1	High	Yes	No	No	No	No	No	No	-6.59	0	1
D2	High	No	No	No	No	No	No	No	-9.81	0	1
D3	High	Yes	No	No	No	No	Yes	No	-6.07	0	1
D4	High	Yes	No	Yes	No	No	No	No	-6.18	0	1
D5	High	Yes	No	Yes	No	No	No	No	-6.21	0	1
D6	High	Yes	No	No	No	No	No	No	-5.96	0	1
D7	High	Yes	No	Yes	No	No	No	No	-6.54	0	1
D8	High	Yes	No	No	No	No	No	No	-7.28	0	1
D9	High	Yes	No	Yes	No	No	No	No	-6.79	0	1
D10	High	Yes	No	Yes	Yes	No	Yes	No	-5.61	0	1
D11	High	No	No	No	No	No	No	No	-6.48	0	1
D12	High	Yes	No	Yes	No	No	No	No	-6.19	0	1
D13	High	No	No	Yes	Yes	No	No	No	-6.79	0	0
D14	High	Yes	No	Yes	No	No	No	No	-6.28	0	1
D15	High	Yes	No	Yes	No	No	No	No	-6.43	0	1
D16	High	Yes	No	Yes	No	No	No	No	-6.6	0	0
D17	High	Yes	No	Yes	Yes	No	No	No	-7.01	0	0
D18	High	Yes	No	Yes	Yes	No	No	No	-6.18	0	0
D19	High	No	Yes	No	Yes	Yes	No	No	-6.72	0	0
D20	High	No	Yes	No	Yes	Yes	No	No	-6.76	0	0
D21	High	No	No	No	Yes	Yes	No	Yes	-6.51	0	1
D22	High	No	Yes	No	Yes	Yes	No	Yes	-7.33	0	2
D23	High	No	Yes	No	Yes	Yes	No	Yes	-6.93	0	1
D24	High	No	No	Yes	No	No	No	No	-7	0	1
D25	High	No	No	Yes	No	No	No	No	-7.04	0	1

A Fragment-Based Approach for the Development of G-Quadruplex Ligands

D26	High	Yes	No	Yes	Yes	No	No	No	-6.79	0	0
D27	High	No	No	Yes	No	No	No	No	-7.21	0	1
D28	High	No	Yes	No	No	No	No	No	-7.62	0	0
D29	High	Yes	No	Yes	No	No	No	No	-5.99	0	1
D30	High	No	No	No	No	No	No	No	-6.76	0	1
E1	High	No	No	No	No	No	No	No	-6.38	0	1
E2	High	No	No	No	No	No	No	No	-6.69	0	1
E3	High	Yes	No	Yes	No	No	No	No	-6.48	0	1
E4	High	Yes	No	No	No	No	No	No	-6.47	0	1
E5	High	Yes	No	Yes	No	No	No	No	-6.18	0	1
E6	High	Yes	No	Yes	No	No	No	No	-6.04	0	1
E7	High	Yes	No	Yes	No	No	No	No	-5.32	0	1
E8	High	Yes	No	No	No	No	No	No	-5.94	0	1
E9	High	Yes	No	Yes	No	No	No	No	-6.34	0	1
E10	High	No	No	No	No	No	No	No	-7.03	0	1
E11	High	Yes	No	No	No	No	No	No	-6.25	0	1
E12	High	No	No	No	No	No	No	No	-6.72	0	1
E13	High	No	No	No	No	No	No	No	-6.35	0	1
E14	High	Yes	No	Yes	No	No	No	No	-6.4	0	1
E15	High	Yes	No	Yes	No	No	No	No	-6.38	0	1
E16	High	Yes	No	Yes	No	No	No	No	-5.46	0	1
E17	High	Yes	No	Yes	No	No	No	No	-5.55	0	1
E18	High	Yes	No	Yes	Yes	No	No	No	-6.18	0	1
E19	High	Yes	No	Yes	Yes	Yes	No	No	-5.21	0	2
E20	High	Yes	No	Yes	No	No	No	No	-6.56	0	1
E21	High	Yes	No	Yes	Yes	Yes	No	No	-5.04	0	1
E22	High	Yes	No	Yes	No	No	No	No	-6.5	0	1
E23	High	Yes	No	No	Yes	No	No	No	-6.92	0	0
E24	High	Yes	No	No	Yes	No	No	No	-6.92	0	0
E25	High	Yes	No	Yes	Yes	No	No	No	-6.24	0	0
E26	High	Yes	No	Yes	Yes	No	No	No	-6.07	0	1
E27	High	Yes	No	Yes	No	No	No	No	-5.56	0	1
E28	High	Yes	Yes	Yes	No	No	Yes	No	-6.31	0	1
E29	High	Yes	No	Yes	No	No	No	No	-6.1	0	1
E30	High	Yes	Yes	Yes	No	No	Yes	No	-6.67	0	1
F1	High	Yes	Yes	Yes	Yes	No	Yes	Yes	-4.89	0	2
F2	High	No	No	Yes	Yes	Yes	No	No	-5.82	0	1
F3	High	Yes	No	Yes	Yes	No	No	No	-6.93	0	0
F4	High	No	No	Yes	Yes	Yes	No	No	-5.01	0	1
F5	High	Yes	No	Yes	Yes	Yes	No	No	-6.03	0	0
F6	High	Yes	No	Yes	Yes	No	No	No	-6.68	0	0
F7	High	No	No	Yes	No	No	No	No	-6.89	0	0
F8	High	Yes	No	Yes	Yes	No	No	No	-6.51	0	0
F9	High	Yes	Yes	No	No	No	No	No	-7.17	0	1
F10	High	Yes	Yes	No	No	No	Yes	No	-6.82	0	0
F11	High	No	Yes	No	No	No	Yes	No	-8.89	1	1
F12	High	No	Yes	No	No	No	Yes	No	-8.32	0	1
F13	High	No	Yes	No	No	No	No	No	-9.35	1	1
F14	High	No	Yes	No	No	No	No	No	-8.93	0	1
F15	High	No	Yes	No	No	No	No	No	-9.15	1	1
F16	High	Yes	No	Yes	No	No	No	No	-6.64	0	1

A Fragment-Based Approach for the Development of G-Quadruplex Ligands

F17	High	Yes	No	Yes	No	No	No	No	-6.55	0	1
F18	High	Yes	Yes	Yes	No	No	No	No	-7.12	0	1
F19	High	Yes	Yes	Yes	No	No	Yes	No	-6.77	0	1
F20	High	Yes	Yes	Yes	No	No	Yes	No	-6.93	0	1
F21	High	Yes	Yes	No	No	No	No	No	-6.92	0	1
F22	High	Yes	Yes	No	No	No	Yes	No	-6.57	0	1
F23	High	Yes	Yes	No	No	No	Yes	No	-6.74	0	1
F24	High	No	Yes	No	No	No	No	No	-8.37	0	1
F25	High	Yes	No	Yes	No	No	No	No	-6.14	0	1
F26	High	Yes	Yes	No	No	No	No	No	-6.7	0	1
F27	High	Yes	Yes	No	No	No	Yes	No	-6.35	0	1
F28	High	Yes	No	Yes	No	No	No	No	-6.18	0	1
F29	High	Yes	No	Yes	No	No	No	No	-6.04	0	1
F30	High	Yes	No	Yes	No	No	No	No	-5.32	0	1

Table 7. Lipophilicity and Water Solubility Properties of Fragments (SWISS-ADME-Filters)

	iLogP	xLogP3	WLogP	MLOGP	ESOL Class	Ali LogS	Ali Solubility (mol/l)	Ali Class
A1								
A2	1.58	2.04	1.77	1.23	Soluble	-2.32	4.81E-03	Soluble
A3	1.37	1.77	2.07	1.23	Soluble	-2.27	5.42E-03	Soluble
A4	1.08	0.7	1.76	0.91	Very soluble	-1.16	6.99E-02	Very soluble
A5	1.79	2.14	2.08	1.53	Soluble	-2.42	3.79E-03	Soluble
A6	2.26	1.51	2.19	1.22	Soluble	-1.96	1.09E-02	Very soluble
A7	2.25	-2.79	3.28	2.66	Highly soluble	2.69	4.94E+02	Highly soluble
A8	1.61	1.59	1.73	1.33	Soluble	-2.12	7.55E-03	Soluble
A9	1.08	0.7	1.76	0.91	Very soluble	-1.16	6.99E-02	Very soluble
A10	1.89	2.46	2.38	2.42	Soluble	-3.45	3.55E-04	Soluble
A11	0.83	0.23	1.06	0.35	Very soluble	-1.21	6.11E-02	Very soluble
A12	1.17	0.76	1.16	0.7	Very soluble	-1.26	5.49E-02	Very soluble
A13	1.59	1.32	1.77	1.23	Soluble	-1.57	2.69E-02	Very soluble
A14	2.2	3.42	2.93	2.88	Soluble	-3.68	2.07E-04	Soluble
A15	1.83	1.04	1.58	1.22	Soluble	-1.83	1.49E-02	Very soluble
A16	0.5	1.35	1.38	0.49	Soluble	-1.91	1.22E-02	Very soluble
A17	2.4	3.11	3.43	1.81	Soluble	-4.29	5.10E-05	Moderately soluble
A18	1.49	1.18	1.89	0.86	Soluble	-1.78	1.64E-02	Very soluble
A19					Moderately soluble			
A20	2.58	3.67	3.78	2.08	soluble	-4.87	1.34E-05	Moderately soluble
A21	1.97	1.83	2.3	1.47	Soluble	-1.91	1.22E-02	Very soluble
A22	0.83	1.32	1.47	0.8	Soluble	-2.35	4.42E-03	Soluble
A23	1.2	0.78	1.4	0.82	Very soluble	-1.79	1.61E-02	Very soluble
A24								
A25	3.02	3.35	3.5	2.51	Soluble	-3.62	2.39E-04	Soluble
A26	0.8	0.17	0.83	-0.56	Very soluble	-0.42	3.82E-01	Very soluble
A27	2.05	1.35	1.78	1.5	Soluble	-1.85	1.40E-02	Very soluble
A28	1.89	2.46	2.38	2.42	Soluble	-3.45	3.55E-04	Soluble
A29	0.83	0.23	1.06	0.35	Very soluble	-1.21	6.11E-02	Very soluble
A30	1.17	0.76	1.16	0.7	Very soluble	-1.26	5.49E-02	Very soluble
A31	1.17	0.76	1.16	0.7	Very soluble	-1.26	5.49E-02	Very soluble
A32	2.2	3.42	2.93	2.88	Soluble	-3.68	2.07E-04	Soluble
A33	1.83	1.04	1.58	1.22	Soluble	-1.83	1.49E-02	Very soluble
A34	1.63	1.85	1.33	0.72	Soluble	-1.83	1.46E-02	Very soluble
A35	2.4	3.11	3.43	1.81	Soluble	-4.29	5.10E-05	Moderately soluble
A36	1.36	0.78	1.58	0.57	Very soluble	-1.37	4.28E-02	Very soluble
A37					Moderately soluble			
A38	2.58	3.67	3.78	2.08	soluble	-4.87	1.34E-05	Moderately soluble
A39	0.83	0.83	1.57	0.27	Very soluble	-1.65	2.24E-02	Very soluble
A40	0.83	1.32	1.47	0.8	Soluble	-2.35	4.42E-03	Soluble
A41	1.2	0.78	1.4	0.82	Very soluble	-1.79	1.61E-02	Very soluble
A42								
A43	3.47	3.3	3.52	2.74	Soluble	-3.34	4.56E-04	Soluble
A44	1.03	1.64	1.15	0.09	Soluble	-1.97	1.06E-02	Very soluble
A45	1.58	2.04	1.77	1.23	Soluble	-2.32	4.81E-03	Soluble
A46	1.79	1.71	2.51	0.39	Soluble	-2.69	2.05E-03	Soluble
A47	1.69	1.38	2.06	0.26	Soluble	-2.6	2.53E-03	Soluble
A48	2.2	1.77	2.25	1.46	Soluble	-3	9.94E-04	Soluble
A49	2.68	2.02	2.88	0.65	Soluble	-3.26	5.47E-04	Soluble
A50	3.61	2.18	2.44	2.07	Soluble	-3.27	5.40E-04	Soluble
A51					Moderately soluble			
A52	2.62	3.41	4.15	1.75	soluble	-5.19	6.52E-06	Moderately soluble
A53	2.29	3.05	3.76	1.52	Soluble	-4.81	1.54E-05	Moderately soluble
A54								
A55	2.82	2.65	3.18	2.27	Soluble	-3.43	3.67E-04	Soluble
A56	1.96	2.04	2.79	1.95	Soluble	-2.73	1.85E-03	Soluble
A57	1.7	1.51	2.6	0.51	Soluble	-2.54	2.87E-03	Soluble

A Fragment-Based Approach for the Development of G-Quadruplex Ligands

B21	2.63	1.88	2.7	1.13	Soluble	-2.64	2.31E-03	Soluble
B22	1.91	2.52	3.47	1.36	Soluble	-3.23	5.86E-04	Soluble
B23	2.17	2.88	3.86	1.63	Soluble	-3.61	2.48E-04	Soluble
B24	2.13	2.75	3.01	0.24	Soluble	-4.23	5.87E-05	Moderately soluble
B25	3.59	2.18	3.07	1.3	Soluble	-3.2	6.31E-04	Soluble
B26	2.74	2.86	2.64	0.27	Soluble	-4.7	2.01E-05	Moderately soluble
B27	2.55	2.11	2.99	0.82	Soluble	-3.23	5.85E-04	Soluble
B28	1.94	2.49	2.54	-0.34	Soluble	-4.6	2.49E-05	Moderately soluble
B29	0.62	-0.51	-0.64	-1.06	Very soluble	-1.08	8.30E-02	Very soluble
B30	0.41	-0.84	-0.75	-1.06	Very soluble	-0.86	1.38E-01	Very soluble
C1	0.71	-0.62	-0.63	-0.39	Very soluble	-0.29	5.15E-01	Very soluble
C2	0.52	-1.07	-0.94	-0.8	Very soluble	0.18	1.51E+00	Highly soluble
C3	0.19	-1.29	-1.23	-0.96	Very soluble	-0.02	9.60E-01	Very soluble
C4	1.92	2.2	1.55	1.2	Soluble	-3.23	5.95E-04	Soluble
C5	0.76	-0.17	-0.23	-1.31	Very soluble	-0.89	1.30E-01	Very soluble
C6	1.51	0.46	-0.01	0.35	Very soluble	-0.28	5.23E-01	Very soluble
C7	0.15	-0.43	-0.06	-0.82	Very soluble	-0.8	1.60E-01	Very soluble
C8	0.74	-3.62	0.23	-1.43	Highly soluble	2.43	2.70E+02	Highly soluble
C9	1.71	0.68	0.73	0.83	Very soluble	-1.07	8.61E-02	Very soluble
C10	2.11	0.48	0.18	0.57	Very soluble	-0.31	4.88E-01	Very soluble
C11	1.48	-0.54	-0.14	0.1	Very soluble	0.58	3.83E+00	Highly soluble
C12	2.58	1.9	1.46	1.72	Soluble	-2.2	6.28E-03	Soluble
C13	1.71	0.08	-0.34	0.21	Very soluble	-0.25	5.64E-01	Very soluble
C14	1.19	-1.25	-1.49	-0.58	Very soluble	-0.57	2.69E-01	Very soluble
C15	1.74	2.2	2.01	1.2	Soluble	-2.55	2.81E-03	Soluble
C16	1.44	1.71	1.62	0.91	Soluble	-2.04	9.07E-03	Soluble
C17	1.4	1.38	1.51	0.03	Soluble	-2.12	7.50E-03	Soluble
C18	5.96	10.91	9.16	5.64	Poorly soluble	-11.59	2.57E-12	Insoluble
C19	1.95	2.69	1.8	1.24	Soluble	-3.48	3.28E-04	Soluble
C20	2.03	3.65	1.87	0.67	Moderately soluble	-4.91	1.24E-05	Moderately soluble
C21	1.81	1.15	1.47	1.06	Soluble	-1.41	3.85E-02	Very soluble
C22	2.3	2.78	2.99	1.93	Soluble	-3.23	5.86E-04	Soluble
C23	0.97	0.77	1.38	0.35	Very soluble	-1.31	4.89E-02	Very soluble
C24	3.09	3	2.22	3.12	Soluble	-2.73	1.85E-03	Soluble
C25	2.38	1.95	2.13	2.26	Soluble	-2.25	5.57E-03	Soluble
C26	2.67	2.55	2.3	2.33	Soluble	-3.19	6.43E-04	Soluble
C27	2.59	2.6	2.14	2.67	Soluble	-2.32	4.82E-03	Soluble
C28	2.3	2.02	1.44	1.73	Soluble	-2.09	8.06E-03	Soluble
C29	2.66	2.48	1.78	2.01	Soluble	-2.39	4.11E-03	Soluble
C30	2.48	2.12	2	2.14	Soluble	-2.2	6.34E-03	Soluble
D1	2.08	1.65	2.22	1.77	Soluble	-2.41	3.90E-03	Soluble
D2	0.84	-3.15	0.3	-0.62	Highly soluble	2.14	1.37E+02	Highly soluble
D3	2.54	2.09	1.3	2.02	Soluble	-2.21	6.13E-03	Soluble
D4	1.43	1.09	0.99	1.54	Very soluble	-1.23	5.91E-02	Very soluble
D5	1.56	1.2	1.55	1.97	Very soluble	-1.34	4.54E-02	Very soluble
D6	1.7	1.98	3.16	2.61	Soluble	-2.15	7.04E-03	Soluble
D7	1.7	0.84	1	1.21	Very soluble	-1.16	6.87E-02	Very soluble
D8	2.14	0.32	1.02	0.64	Very soluble	-1.01	9.75E-02	Very soluble
D9	2.47	0.86	1.27	1.22	Very soluble	-1.08	8.24E-02	Very soluble
D10	2.74	2.67	2.67	3.34	Soluble	-2.57	2.66E-03	Soluble
D11	1.15	1.06	0.9	0.34	Very soluble	-2.16	6.92E-03	Soluble
D12	2.24	1.87	2.14	2.61	Soluble	-2.04	9.16E-03	Soluble
D13	1.46	1.65	1.45	1.75	Soluble	-2.84	1.46E-03	Soluble
D14	1.8	1.99	1.55	2.03	Soluble	-2.23	5.92E-03	Soluble
D15	1.91	1.93	2.11	2.44	Soluble	-2.17	6.84E-03	Soluble
D16	2.2	1.8	1.55	1.7	Soluble	-2.22	5.97E-03	Soluble
D17	2.54	1.74	1.57	1.1	Soluble	-2.55	2.82E-03	Soluble
D18	2.09	2.71	3.72	2.99	Soluble	-2.97	1.06E-03	Soluble
D19	2.55	2.2	1.71	1.9	Soluble	-3.63	2.34E-04	Soluble
D20	2.66	2.3	2.27	2.29	Soluble	-3.73	1.85E-04	Soluble
D21	3.08	3.09	3.88	2.74	Moderately soluble	-4.55	2.80E-05	Moderately soluble
D22	3.01	2.12	1.74	1.03	Soluble	-4.13	7.44E-05	Moderately soluble

A Fragment-Based Approach for the Development of G-Quadruplex Ligands

D23	2.75	2.17	1.72	1.61	Soluble	-3.79	1.61E-04	Soluble
D24	1.64	0.84	0.58	0.57	Very soluble	-1.9	1.27E-02	Very soluble
D25	1.71	0.94	1.14	0.97	Very soluble	-2	1.00E-02	Very soluble
D26	1.9	1.72	2.75	1.5	Soluble	-2.81	1.55E-03	Soluble
D27	2	0.81	0.59	-0.13	Very soluble	-2.06	8.75E-03	Soluble
D28	2.22	0.75	0.61	-0.69	Soluble	-2.38	4.14E-03	Soluble
D29	0.89	1.74	1.41	1.07	Soluble	-2.68	2.10E-03	Soluble
D30	1.22	1.16	1.42	-0.2	Very soluble	-2.46	3.44E-03	Soluble
E1	1.09	1.45	1.42	0.03	Soluble	-2.57	2.69E-03	Soluble
E2	1.05	1	1.42	0.03	Very soluble	-2.1	7.88E-03	Soluble
E3	1.31	1.41	1.47	0.35	Soluble	-2.4	3.95E-03	Soluble
E4	0.88	1.22	1.97	1.51	Very soluble	-2.14	7.28E-03	Soluble
E5	1.66	1.1	1.03	1.54	Very soluble	-1.12	7.63E-02	Very soluble
E6	1.69	1.42	1.42	1.87	Very soluble	-1.45	3.55E-02	Very soluble
E7	1.79	2.45	1.96	2.61	Soluble	-2.91	1.23E-03	Soluble
E8	1.56	1.58	1.4	1.15	Soluble	-1.81	1.55E-02	Very soluble
E9	1.92	1.13	1.04	1.21	Very soluble	-1.34	4.55E-02	Very soluble
E10	1.61	0.81	0.95	-0.11	Very soluble	-2.17	6.82E-03	Soluble
E11	2.24	1.51	1.04	0.92	Soluble	-1.93	1.17E-02	Very soluble
E12	0.22	0.86	0.64	-0.22	Very soluble	-2.26	5.55E-03	Soluble
E13	0.99	1.24	0.94	0.34	Very soluble	-2.22	5.96E-03	Soluble
E14	1.1	1.28	1.8	0.59	Very soluble	-2.2	6.31E-03	Soluble
E15	2.19	1.44	1.91	1.13	Soluble	-1.79	1.62E-02	Very soluble
E16	2.5	2.83	2.74	2.31	Soluble	-2.91	1.22E-03	Soluble
E17	2.18	2.47	2.21	1.74	Soluble	-2.67	2.15E-03	Soluble
E18	1.95	2.24	3.55	1.55	Soluble	-3.04	9.14E-04	Soluble
E19	2.18	3.66	4.22	2.18	Moderately soluble	-4.84	1.46E-05	Moderately soluble
E20	1.65	0.9	0.74	0.47	Very soluble	-1.39	4.08E-02	Very soluble
E21	2.52	4.06	4.67	3.34	Moderately soluble	-4.66	2.20E-05	Moderately soluble
E22	2.05	1.61	2.42	0.67	Soluble	-2.5	3.14E-03	Soluble
E23	3.12	1.62	4.44	2.12	Soluble	-2.32	4.80E-03	Soluble
E24	3.12	1.62	4.44	2.12	Soluble	-2.32	4.80E-03	Soluble
E25	2.18	2.64	3.39	1.88	Soluble	-3.38	4.19E-04	Soluble
E26	2.17	2.42	3.5	1.59	Soluble	-3.06	8.73E-04	Soluble
E27	1.43	2.05	2.17	1.57	Soluble	-2.01	9.77E-03	Soluble
E28	2.99	2.09	2.45	1.75	Soluble	-2.37	4.24E-03	Soluble
E29	1.88	1.78	2.08	1.55	Soluble	-1.8	1.59E-02	Very soluble
E30	2.53	1.35	1.67	1.23	Soluble	-1.6	2.49E-02	Very soluble
F1	1.99	4.1	4.24	3	Moderately soluble	-4.47	3.40E-05	Moderately soluble
F2	2.02	2.8	3.61	1.27	Soluble	-4.21	6.12E-05	Moderately soluble
F3	1.99	1.34	2.22	1.23	Soluble	-2.26	5.48E-03	Soluble
F4	2.94	4.6	5.18	2.57	Moderately soluble	-5.8	1.57E-06	Moderately soluble
F5	2.33	2.71	3.5	1.8	Soluble	-3.63	2.34E-04	Soluble
F6	2.21	1.81	2.61	1.51	Soluble	-2.75	1.78E-03	Soluble
F7	1.49	1.42	2.35	0.92	Soluble	-2.77	1.70E-03	Soluble
F8	2.38	2.17	3	1.78	Soluble	-3.12	7.54E-04	Soluble
F9	2.14	0.77	1.26	0.64	Very soluble	-1.55	2.83E-02	Very soluble
F10	2.51	1.51	2.04	1.16	Soluble	-2.32	4.82E-03	Soluble
F11	3.11	-0.91	-0.17	-0.25	Very soluble	-0.5	3.20E-01	Very soluble
F12	2.52	-0.48	0.24	0.07	Very soluble	-0.69	2.05E-01	Very soluble
F13	2.61	-1.55	-0.77	-0.62	Very soluble	-0.1	7.91E-01	Very soluble
F14	3.75	-0.92	-0.92	-0.45	Very soluble	-0.54	2.88E-01	Very soluble
F15	2.31	-1.27	-0.77	-0.48	Very soluble	-0.39	4.05E-01	Very soluble
F16	0.97	0.77	1.38	0.35	Very soluble	-1.31	4.89E-02	Very soluble
F17	1.81	1.15	1.47	1.06	Soluble	-1.41	3.85E-02	Very soluble
F18	2.38	0.72	1.06	0.78	Very soluble	-1.22	6.01E-02	Very soluble
	2.81	1.45	1.84	1.31	Soluble	-1.98	1.05E-02	Very soluble

F19								
F20	2.32	1.21	1.22	1.31	Soluble	-1.73	1.86E-02	Very soluble
F21	2.11	1	1.06	0.92	Very soluble	-1.51	3.08E-02	Very soluble
F22	2.47	1.73	1.84	1.45	Soluble	-2.27	5.38E-03	Soluble
F23	2.32	1.48	1.22	1.45	Soluble	-2.01	9.77E-03	Soluble
F24	3.01	-0.5	-0.51	-0.15	Very soluble	-0.72	1.89E-01	Very soluble
F25	2.55	1.77	1.32	1.23	Soluble	-1.84	1.43E-02	Very soluble
F26	3.03	1.34	0.91	0.96	Soluble	-1.65	2.24E-02	Very soluble
F27	3.61	2.08	1.69	1.49	Soluble	-2.42	3.82E-03	Soluble
F28	2.33	2.71	3.5	1.8	Soluble	-3.63	2.34E-04	Soluble
F29	2.21	1.81	2.61	1.51	Soluble	-2.75	1.78E-03	Soluble
F30	1.49	1.42	2.35	0.92	Soluble	-2.77	1.70E-03	Soluble

From our investigation based on primary SWISS-ADME filters, we got the impression that all the designed and optimized fragments are obeying the basic requirements to be screen as small drug-like molecule to target G4 DNA sequences.

2.9 Fragments screening with Fluorescent indicator displacement (FID) assays

The screening of all the primarily selected fragments and the optimized set of fragments are reported in this part of the thesis. Firstly, we did screening of all the commercial and synthesized fragments by FID assays to check the binding to a DNA G4 derived from the *c-MYC* promoter by using fluorescent indicator displacement (FID) employing thiazole orange (TO). The screening library comprised 180 structurally and chemically diverse ligands (130 commercial and 50 synthesized). Amongst them, the top 11 confirmed and profiled hits selected from the intensive FID screen for binding to *c-MYC* promoter G4 DNA (Fig. 50a and 50b) were used in molecular modelling with the solution NMR structure for the *c-MYC*-G4 (PDB entry **2L7V**; see below and table 10).⁸⁶

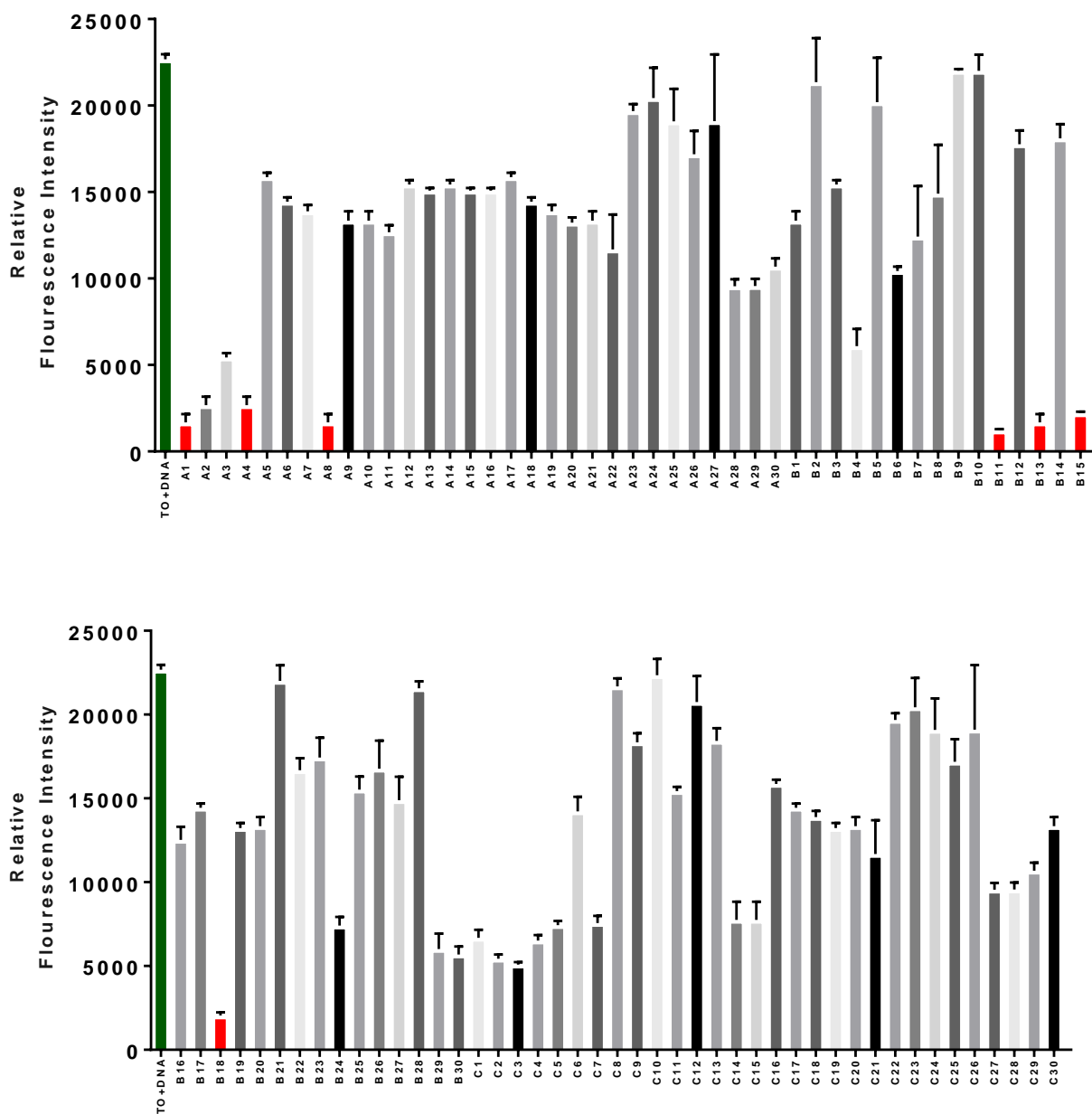


Figure 50a. Thiazole Displacement assays of screened fragments Series I (A), Series II (B), Series III (C) Series IV (D), Series V (E) Series VI (F); Experimental condition: 0.25 μM DNA, 0.5 μM Thiazole Orange, 10 % DMSO, 20 mM Na cacoc, 140 mM KCl, pH 7 (25 μL /well)

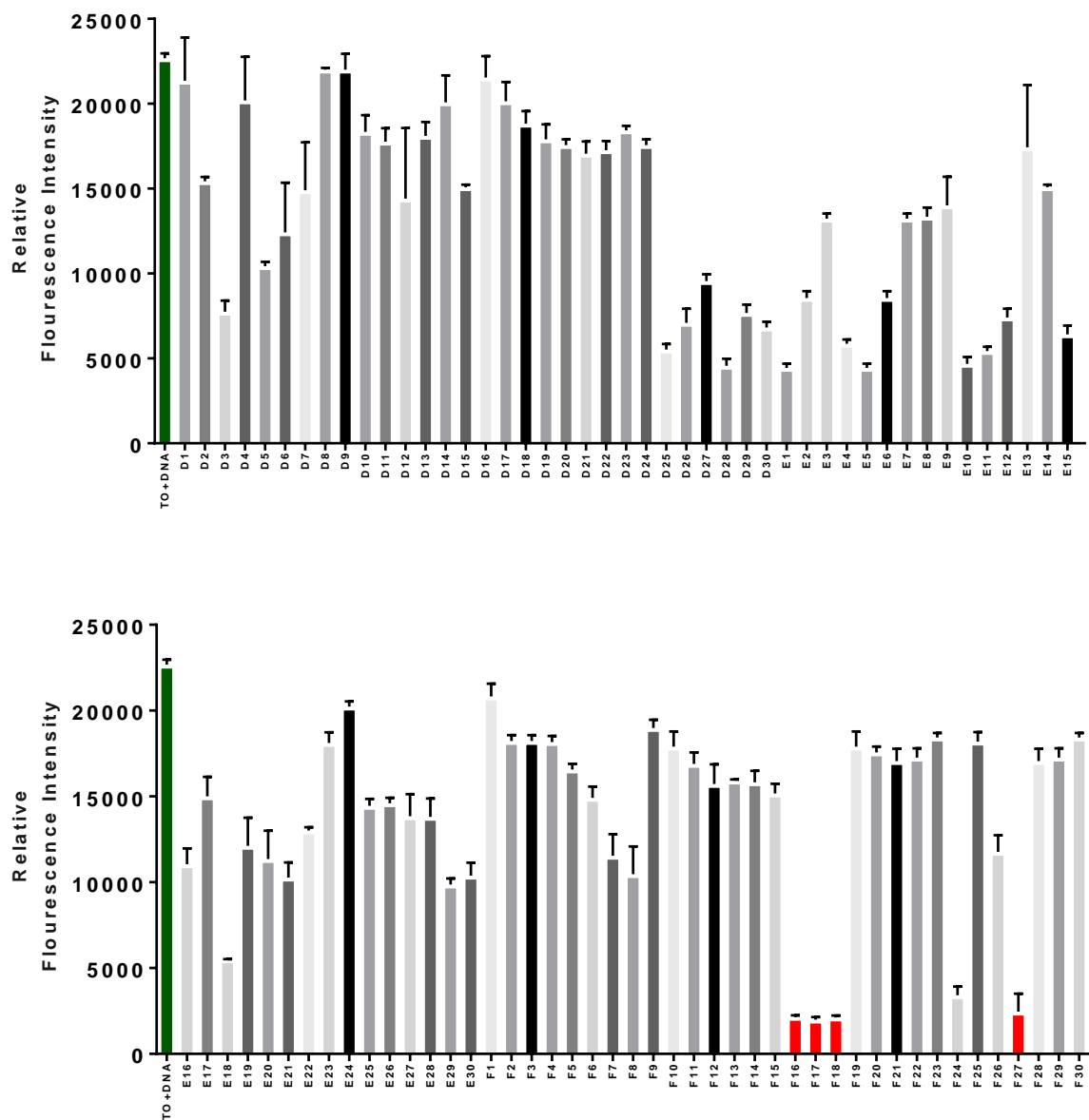


Figure 50b. Thiazole Displacement assays of screened fragments Series I (A), Series II (B), Series III (C), Series IV (D), Series V (E), Series VI (F); Experimental condition: 0.25 μM DNA, 0.5 μM Thiazole Orange, 10 % DMSO, 20 mM Na cac, 140 mM KCl, pH 7 (25 μL /well)

2.10 Molecular Docking studies

To gain molecular insight into the possible binding mode of the most promising scaffolds, molecular docking studies were performed with the *c-MYC* G4 (PDB: **2L7V**) at both 5'- and 3'-ends using Autodock 4.2 tools.⁸⁷ Prior to docking, all ligand conformations of best conjugates were optimized with Gaussian 3.0 with the semi empirical PM3 force field. Molecular docking for best 11 fragments generated from our primary FID screening were performed into the binding sites at the two ends of *c-MYC* G4-NMR structure of the 2:1 quindoline: *MYC* G4 complex in K⁺ solution (PDB ID **2L7V**).⁸⁶ Docking studies for best ligand in this series compound **7** & **5** gave similar docking scores for the 5'- and 3'-complexes at - 5.43 and - 5.48 kcal/mol, respectively (Table 8). Several similar binding poses predicted by the docking experiment for both the 5'- and 3'- sites.

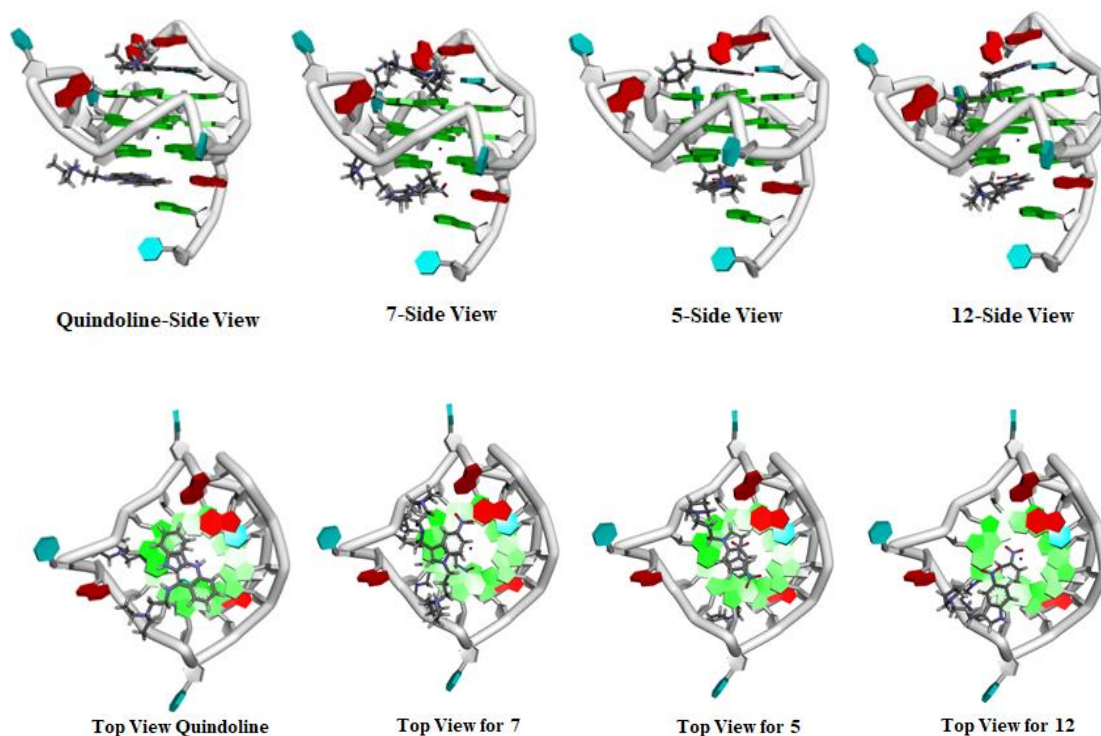
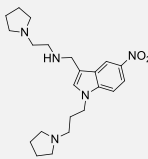
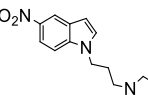
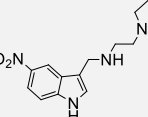
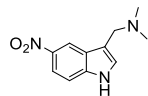
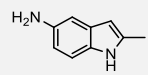
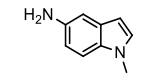
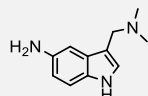
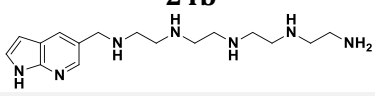
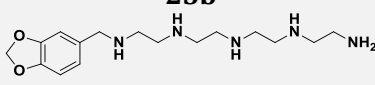
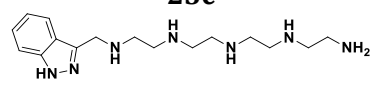


Figure 51. Binding pose for optimized fragments with 2L7V (2:1) Quindoline-*c-MYC* DNA (solution structure of *c-MYC* G-quadruplex pu22 (PDB ID: **2L7V**) suggested by Autodock 4.2.

The overall binding modes of the indole scaffolds resembles those of quindoline in the NMR structure of the 2:1 quindoline: *MYC* G4 complex.⁸⁶ Notably, the central aromatic core of scaffold **5**, **7** and **12** are stacking very well with both the 5- and 3-external tetrads, making extensive stacking interactions (Fig. 51). The positively charged pyrrolidine side chains of scaffolds **5**, **7** and

12 bind in the *c-MYC* G4 groove and form intermolecular salt bridges with phosphate groups on the nucleotide backbone.

Table 8. Ranking of ligands based upon their binding energies generated after docking analysis of 3, A2, 9a, 9, B12, 12, 7, 5, 24b, 25b & 23c

	Binding Energy (kcal/mol)					
7 	-5.65	-5.45	-5.13	-4.87	-4.65	-4.60
5 	-5.37	-5.08	-4.82	-4.48	-4.48	-4.29
12 	-5.45	-5.36	-4.75	-4.71	-4.67	-4.66
9 	-5.34	-4.76	-4.62	-4.61	-4.50	-4.25
A-2 	-4.62	-4.55	-4.49	-4.60	-4.60	-4.60
3 	-4.58	-4.50	-4.48	-4.40	-4.37	-4.33
9a 	-3.94	-3.89	-3.88	-3.84	-3.84	-3.83
24b 	-0.25	4.33	4.95	5.48	5.7	8.61
25b 	0.18	1.18	2.11	2.99	4.62	5.81
23c 	0.49	0.56	2.07	3.09	3.92	4.82

2.11 SWISS-ADME analysis of best lead fragments

The drug-likeness, Physico-chemical, pharmacokinetics, and lipophilicity properties calculated for the four best ligands **5**, **7**, **9a** and **12** along with parent compound **3**⁸⁸ by SWISS-ADME filters suggests that the $\text{miLog } P$ value of all fragments developed was found to be < 5 (3.88), suggesting that these molecules have a good permeability across the cell membrane. The molecular weights of all selected ligands are less than 500 g/mol, on average 234.34 g/mol. The number of hydrogen bond acceptors (O and N atoms) ($\text{NOHNH}=1$) and number of hydrogen bond donors (NH and OH; $n\text{ON}=2$) of best ligands were found to be less than 5, respectively, and were in good agreement with the Lipinski's rule of five (Table 6). While the ADME radar and the drug-likeness parameters are not optimal for **3**, the values for the developed compounds fit into the allowed region of the ADME radar (Fig.52). The above-determined molecular properties of all the fragments justify their use as a possible screening candidate for Fluorescent indicator displacement (FID) assays screening assays.⁸⁹

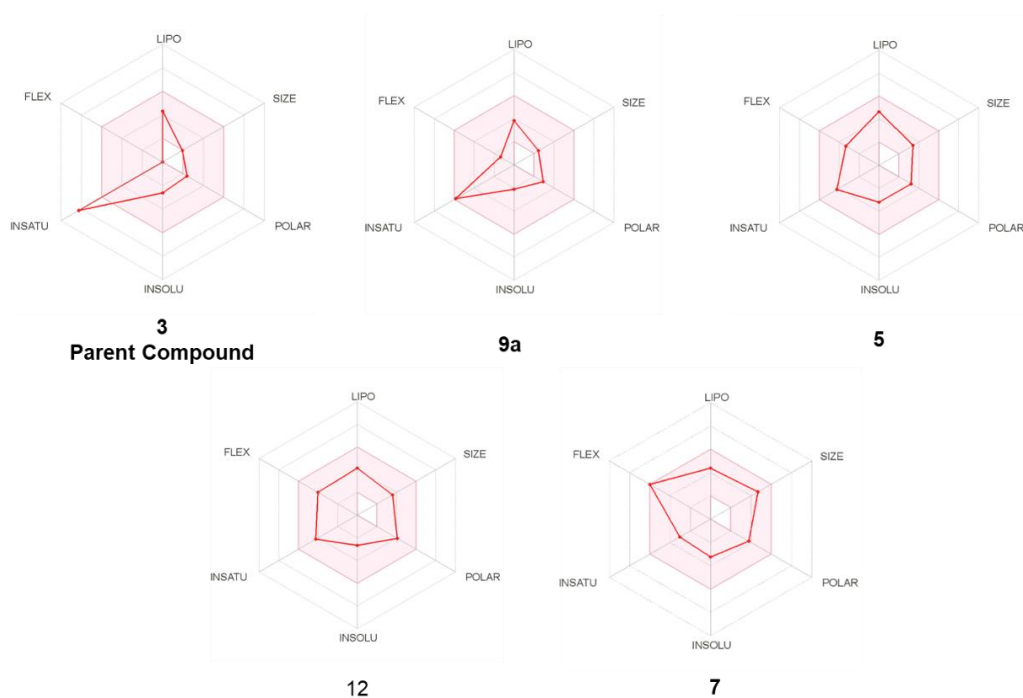


Figure 52. SWISS ADME Radar of **9a** and novel substituted 5-nitro indole derivatives **12**, **7** & **5**. Light red regions indicate preferred regions using the following borders: LIPO (Lipophilicity): $-0.7 < \text{XLOGP3} < +5.0$; SIZE: $150 \text{ g/mol} < \text{MW} < 500 \text{ g/mol}$; POLAR (Polarity): $20 \text{ \AA}^2 < \text{TPSA} < 130 \text{ \AA}^2$; INSOL (Insolubility): $0 < \text{LOGS (ESOL)} < 6$; INSATU (Insaturation): $0.25 < \text{Fraction CSP}^3 < 1$; FLEX (Flexibility): $0 < \text{Num.rotatable Bonds} < 9$ (Table 6).

2.12 DC₅₀ and Ki values estimation for the top 5 fragments: FID assays

The affinity of four best ligands **5**, **7**, **9**, **9a** and **12** along with parent compound **3**⁸⁸ for different folded G4s was further investigated by measuring the ability of the ligands to displace the bound thiazole-orange (TO) from pre-folded G-quadruplexes⁹⁰ and duplex oligonucleotide sequences and determining their K_i values. All the optimized hit molecules generated from the indole pharmacophore (**5**, **7**, **9a** and **12**) showed a very strong affinity towards *c-MYC* G4 DNA compared to the other G4s and compared to duplex DNA with K_i values ranging from 2-5 μM (Fig. 53 & 54, Table 9 & 10).

Table 9. DC₅₀ values for the most active ligands for Pu-22 *c-MYC* G-quadruplex sequence *c-MYC*; *c-KIT-1*; *BCL-2*; h-TELO; ds-DNA-24 & ds-DNA-27

DC ₅₀ (μM)	7	5	12	9	9a	3
Pu-22 c-MYC	3.34	5.08	6.89	4.45	3.826	3.70
c-KIT-1	11.42	8.931	9.70	8.931	10.143	8.60
BCL2	33.22	29.13	33.26	22.16	29.13	37.43
h-Telo- 22	138.99	142.65	129.86	105.10	129.89	138.77
ds-DNA-24	512.36	568.12	586.89	405.55	512.36	556.0
ds-DNA-27	2103.13	2345.12	2360.45	1670.56	2316.12	2466.0

Table 10. Ki values for the most active ligands based on their bound TO-displacement ability for Pu-22 *c-MYC* G-quadruplex sequence *c-MYC*; *c-KIT-1*; *BCL-2*; h-TELO; ds-DNA-24 & ds-DNA-27

Ki (μM)	7	5	12	9	9a	3
Pu-22 c-MYC	2.904	4.42	5.59	3.87	3.32	3.22
c-KIT-1	10.21	7.83	8.85	8.93	8.89	7.54
BCL2	29.37	25.75	29.17	19.43	25.75	32.43
h-Telo- 22	122.39	126.13	104.13	84.74	104.13	111.29
ds-DNA-24	464.13	515.23	532.86	368.12	464.13	504.53
ds-DNA-27	1861.13	2075.12	2088.3	1478.26	2049.66	2182.15

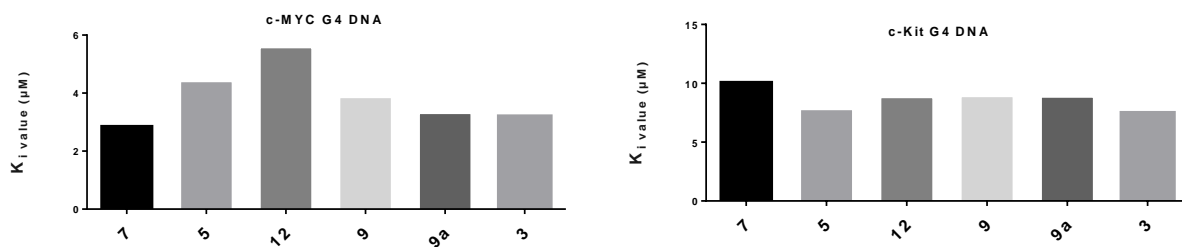


Figure 53. Ki value plots for the most active ligands for promoter G-Quadruplex sequences *c-MYC*, *c-KIT-1* based on Thiazole Displacement Assays ; Titration scheme of *c-MYC*: $c(\text{TO}) = 0,5 \mu\text{M}$; $c(\text{cMYC}) = 0,5 \mu\text{M}$; $c(\text{DMSO}) = 10\%$; $c(\text{Ligand}) = 256 \mu\text{M}, 128 \mu\text{M}, 64 \mu\text{M}, 32 \mu\text{M}, 16 \mu\text{M}, 8 \mu\text{M}, 4 \mu\text{M}, 2 \mu\text{M}, 1 \mu\text{M}, 0,5 \mu\text{M}, 0,25 \mu\text{M}, 0,125 \mu\text{M}, 0,0625 \mu\text{M}$; r.t; Titration of *c-KIT1*: $c(\text{TO}) = 0,5 \mu\text{M}$; $c(\text{c-KIT-1}) = 0,25 \mu\text{M}$; $c(\text{DMSO}) = 10\%$; $c(\text{Ligand}) = 1024 \mu\text{M}, 512 \mu\text{M}, 256 \mu\text{M}, 128 \mu\text{M}, 64 \mu\text{M}, 32 \mu\text{M}, 16 \mu\text{M}, 8 \mu\text{M}, 4 \mu\text{M}, 2 \mu\text{M}, 1 \mu\text{M}, 0,5 \mu\text{M}, 0,25 \mu\text{M}, 0,125 \mu\text{M}$; r.t

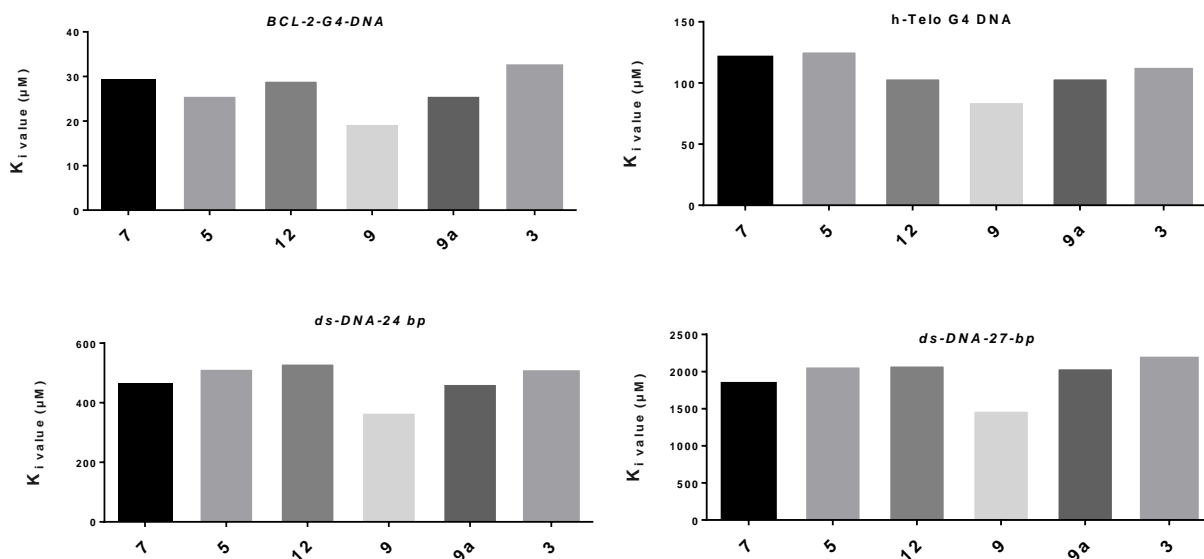


Figure 54. Ki value plots for the most active ligands for promoter G-Quadruplex sequences BCL-2, h-TELO, ds-DNA-24 and ds-DNA-27 based on Thiazole Displacement Assays; Titration scheme of BCL-2: $c(\text{TO}) = 0,5 \mu\text{M}$; $c(\text{BCL-2}) = 0,25 \mu\text{M}$; $c(\text{DMSO}) = 10\%$; $c(\text{Ligand}) = 1024 \mu\text{M}, 512 \mu\text{M}, 256 \mu\text{M}, 128 \mu\text{M}, 64 \mu\text{M}, 32 \mu\text{M}, 16 \mu\text{M}, 8 \mu\text{M}, 4 \mu\text{M}, 2 \mu\text{M}, 1 \mu\text{M}, 0,5 \mu\text{M}, 0,25 \mu\text{M}, 0,125 \mu\text{M}$; r.t; Titration of h-Telo-DNA: $c(\text{TO}) = 0,5 \mu\text{M}$; $c(\text{cMYC}) = 0,5 \mu\text{M}$; $c(\text{DMSO}) = 10\%$; $c(\text{Ligand}) = 5012 \mu\text{M}, 2056 \mu\text{M}, 1028 \mu\text{M}, 604 \mu\text{M}, 302 \mu\text{M}, 106 \mu\text{M}, 80 \mu\text{M}, 40 \mu\text{M}, 20 \mu\text{M}, 10 \mu\text{M}, 5 \mu\text{M}, 1 \mu\text{M}, 0,5 \mu\text{M}$; r.t; Titration of ds-DNA-27: $c(\text{TO}) = 0,5 \mu\text{M}$; $c(\text{cMYC}) = 0,5 \mu\text{M}$; $c(\text{DMSO}) = 10\%$; $80192 \mu\text{M}, 40096 \mu\text{M}, 20048 \mu\text{M}, 10024 \mu\text{M}, 5012 \mu\text{M}, 2056 \mu\text{M}, 1028 \mu\text{M}, 604 \mu\text{M}, 302 \mu\text{M}, 106 \mu\text{M}, 20 \mu\text{M}, 10 \mu\text{M}, 5 \mu\text{M}$; r.t; Titration of ds-DNA-24-Bp: $c(\text{TO}) = 0,5 \mu\text{M}$; $c(\text{cMYC}) = 0,25 \mu\text{M}$; $c(\text{DMSO}) = 10\%$; $c(\text{Ligand}) = 20048 \mu\text{M}, 10024 \mu\text{M}, 5012 \mu\text{M}, 2056 \mu\text{M}, 1028 \mu\text{M}, 604 \mu\text{M}, 302 \mu\text{M}, 106 \mu\text{M}, 80 \mu\text{M}, 40 \mu\text{M}, 20 \mu\text{M}, 10 \mu\text{M}, 5 \mu\text{M}$; r.t

2.13 ¹D-NMR: Ligand binding observed by NMR-spectroscopy

The best fragments **5**, **7**, **9a** and **12** along with parent compound **3**⁸⁸ were further studied by NMR spectroscopy for their interaction with *c-MYC* promoter G4. The interactions of **3**, **5**, **7**, **9a** and **12** with *c-MYC* G4 were further studied by NMR spectroscopy. Figure 55 shows NMR spectra of the DNA-ligand complexes at a ratio of 1:4 (DNA: ligand). In the spectrum of the DNA alone, signals (indicated by grey boxes) of minor conformations of the DNA are well visible. These signals disappear upon addition of ligand, indicating the stabilization of the major conformation upon binding. For ligands **3**, **9a** only minor chemical shift perturbations, and weak line broadening are observed corresponding to weak binding in fast exchange regime. This differs drastically for ligand **12** where severe line broadening in combination with chemical shift changes e.g. nicely visible for G13 of the DNA, indicate stronger interactions in the intermediate exchange regime. Strong chemical shift perturbations are observed for compound **5** and even more so for ligand **7** coupled to only slight line broadening, allowing a detailed examination of the interactions between the DNA and the ligand as shown for ligand **7** in figure 53. Already at a DNA:ligand ratio of 1:1, signals of minor conformations have disappeared and shifting of a significant number of signals rather than appearance of new signals indicate binding in fast exchange. Most prominent chemical shift perturbations with values (positive and negative) around 0.06 ppm at ratio 1:1 are observed for imino-signals of G19, G15 and G6, located at the 3'-tetrad of G4. Perturbations between 0.034 and 0.045 ppm are observed for the imino-signals of G4, G8 and G14, while all other perturbations are in magnitude smaller than 0.022 ppm. G4, G8 and G14 belong to the 5'-tetrad. These findings indicate that the ligand binds to both, the 3' and the 5' tetrad with slight preference for the 3'-tetrad. This trend is also seen at 1:2 ratio where CSPs for the respective 6 imino-signals are between 0.68 and 1.2 ppm in magnitude while they are less than 0.037 ppm for the other iminos.

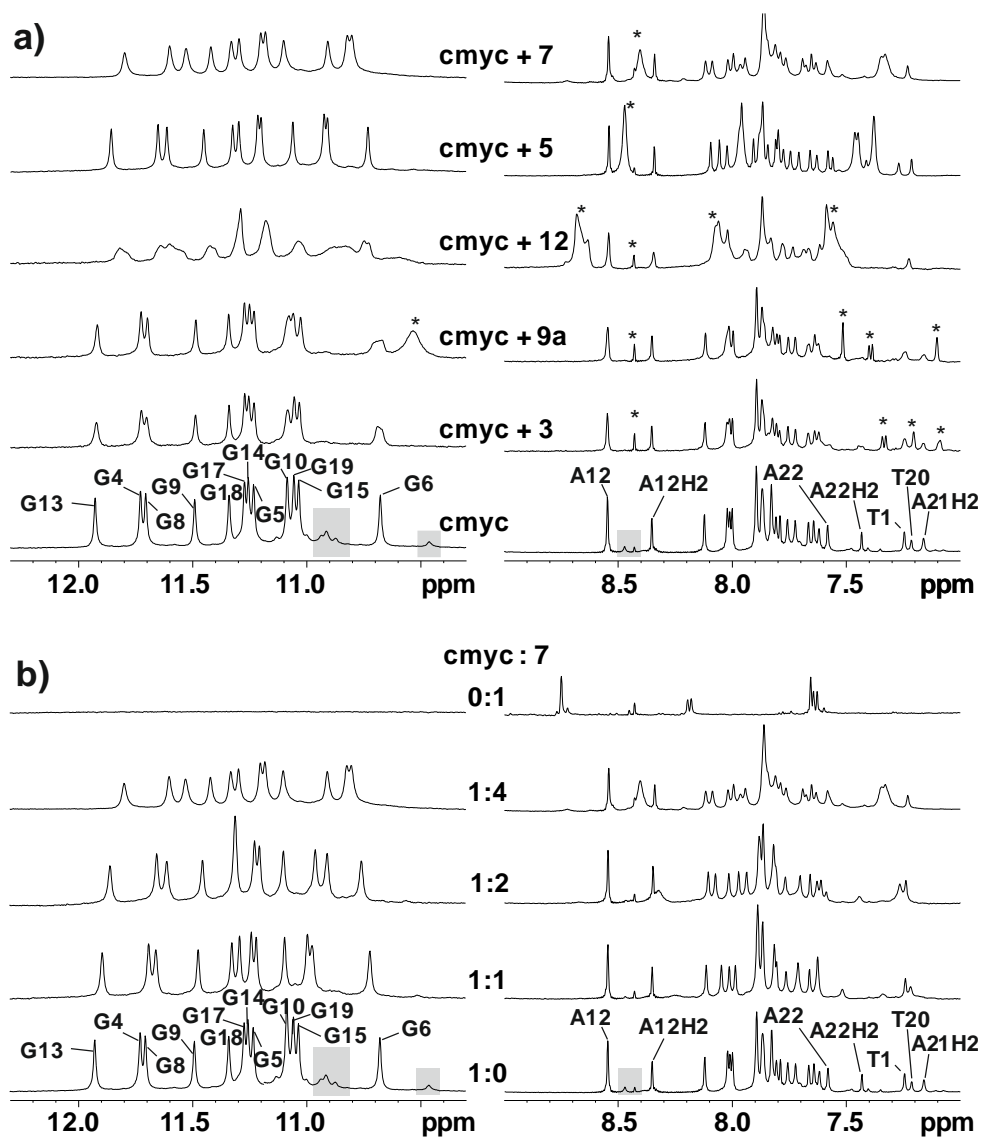


Figure 55. *c-MYC*-ligand interactions observed by ^1H proton NMR spectra (imino and aromatic region). a) Comparison of interactions of *c-MYC* with different ligands at a DNA: ligand ratio of 1:4. Signals clearly belonging to the respective ligand are indicated by asterisks. b) Interactions of ligand 7 with *c-MYC* at different DNA: ligand ratios. (a+b) assignments⁹¹ of imino-signals and clearly separated aromatic protons of loop and tail residues are indicated; grey boxes indicate signals of minor conformations; Experimental conditions:

Observation of binding or structural rearrangement of residues in loop and capping structures can be achieved by investigation of the aromatic region of the spectra. Signals of A12, A22, T1, T20 and A21 are reasonably well resolved to allow the assessment of behaviour upon binding. As A12 does not shift upon binding the loop does not change its environment upon binding. This is

different for T1, T20, A21 and A22. These residues located in the capping structures are perturbed significantly upon binding indicating rearrangement of the capping structures to accommodate the ligands. This rearrangement allowing to accommodate the ligand is also seen in the literature for *c-MYC*,⁸⁶ as well as for other G4s.⁹²

2.14 Cell Cytotoxicity and Cell cycle analysis

Substituted indole derivatives inhibit proliferation of cancer cells and induces cell cycle arrest:

The best fragments **5**, **7**, **9a** and **12** along with parent compound **3**⁸⁸ were further analysed *in vitro* by cell viability assays. HeLa cells were used to perform the cell proliferation assay, as they were not only having high levels of *c-MYC* expression but were also used for similar studies.⁹³ As described in the materials and methods, HeLa cells were exposed with various increasing concentrations (0-50 μM) of the candidate compounds and the parent fragment **3**⁸⁸ to delineate the effects of defined chemical modifications. The percentage of living cells was then measured with Alamar-Blue reagent, by monitoring the absorbance of pink colored and highly red fluorescent resorufin (7-hydroxy-3H-phenoxazin-3-one) which is basically the irreversible reduced product of Alamar-Blue, resazurin (7-Hydroxy-3H-phenoxazin-3-one 10-oxide).⁹⁴ Compound **7** and **5** exhibited significantly enhanced inhibitory effect on the cellular proliferation of HeLa cells showing an IC_{50} value of 5.08 μM & 5.89 μM respectively (Fig. 56A and Table 11) while the other three compounds have less effects on the inhibition of cellular proliferation.

Table 11. IC_{50} values derived for fragments screened *via* FID assays 9a, 9b, 10a and 10b in HeLa cell line after 72 h treatment.

G-Quadruplex Ligands											
	3	9a	12	7	5	A2	9	10	24b	25b	23c
IC_{50} (μM)	> 45	> 45	> 45	5.08	5.89	> 50	> 50	> 50	> 50	> 50	> 50

2.15 Western Blot analysis

Down-regulation of *c-MYC* expression in human cancer cells: To further investigate the effects of these fragments **5**, **7**, **9a** and **12** along with parent compound **3**⁸⁸ at the translational level, HeLa cells were treated with defined concentrations of the ligands for 24 h and the expression levels of *c-MYC* protein relative to the control cells were obtained using western blotting with an anti-*c-MYC* antibody. Protein expression levels calculated for ligand treated cells

from densitometry analysis of Western blots were normalized against untreated control cells. In HeLa cells treated with **12**, **7** & **5**, the c-MYC protein expression was downregulated by 30 % at 3 μM & 50 % at 10 μM , respectively (see 56B). The expression levels of the GAPDH, a housekeeping gene, was not affected by these ligands and similar to the control cells. These results suggest that the **12**, **7** & **5** can down regulate c-MYC expression at protein levels in HeLa cancer cell lines, which is in good agreement with the cell viability assay.

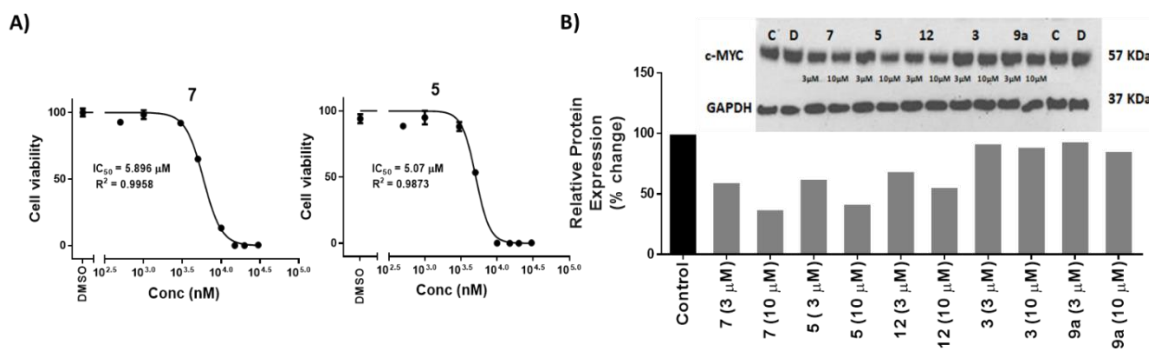


Figure 56. Effect of G₄-Ligands on the cellular expression of c-MYC downstream genes. **A)** IC₅₀ denotes in HeLa cell line after 72 h treatment (see panel A), MYC is the downregulation of protein level (see panel B) and NMR classifies the interactions seen by NMR-spectroscopy (see figure 55). **B)** Bar diagram representing the suppression of c-MYC protein expression in HeLa cells upon treatment with 3 & 10 μM of G₄-Ligands and Western blot analysis for c-MYC gene expression in of **3**, **5**, **7**, **9a** and **12** treated HeLa cells (C= HeLa cell alone control and D = DMSO control).

2.16 Cell cycle analysis

Cell cycle analysis were performed to investigate the effect of ligands on different phases of the cell cycle distribution in HeLa cells, we performed propidium iodide (AAD)-mediated cell cycle analysis experiments. Untreated control cells maintained a high percentage (68.1 %) of the total population in G₀/G₁ phase with 29.1 % in S/G₂phases, respectively. On exposure to compound **5** at 5 μM concentration, cells displayed a prominent G₁ phase arrest in flow cytometry (see figure 56). The G₀/G₁ cell population increased to 79.8 % upon incubation with compound **7** for 24 h at 5- μM concentration.

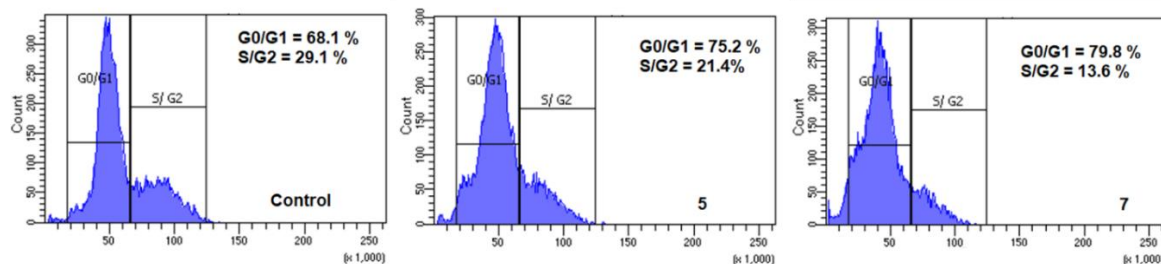


Figure 57. Cell cycle analysis of HeLa cells treated with Compound **5** and **7** at 5 μ M concentrations for 24 h.

The percentage of G0/G1 phase cells at 24 h incubation with Compound **5** is increased to 75.2 % (Fig. 57). The observed increase in G0/G1 population suggested that the compounds caused a significant block for S/G2-phase progression, most probably by hindering the DNA unwinding due to ligand-mediated stabilization of promoter *c-MYC* G-quadruplexes. The FACS analysis thus suggests that upon treatment with **5** and **7**, cells undergo G1 arrest with concomitant decrease of S-phase population thereby leading to cell death.

2.17 Conclusion

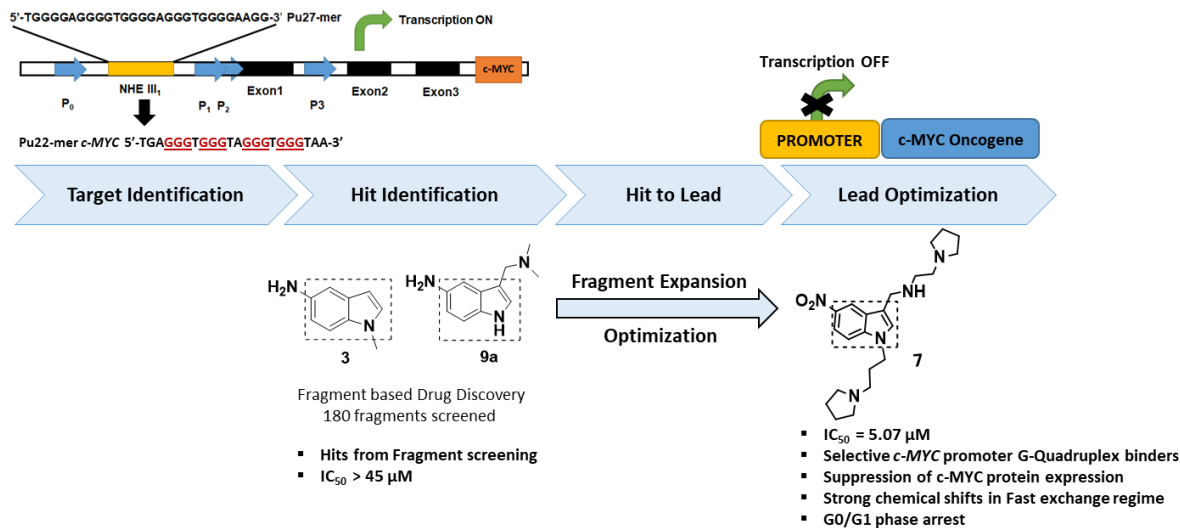


Figure 58. Fragment optimization strategy to preferential target promoter *c-MYC* G4s.

In the present study, we set out to identify new binders targeting the *c-MYC* G-quadruplex starting from the indole fragment. Several synthetic strategies were reported to optimize and generate best hits starting from 5-nitro indole derivatives by introducing the secondary cationic linked pyrrolidine side chain (Fig.58). Interestingly, all improved versions of G4-indole fragments **5**, **7**

and **12** contain this 5-nitro functionality, which may aid in the electrostatic binding and contributes to hydrogen binding interactions of the ligands to G4 DNA. *In-silico* drug design, biological and biophysical analyses illustrate that substituted 5-nitro indoles scaffolds show preferential affinity towards the *c-MYC* promoter G-quadruplex compared to other G-quadruplexes and double stranded DNA. In vitro cellular studies confirm that the substituted indole scaffolds downregulate *c-MYC* expression in cancer cells and have the potential to induce cell cycle arrest in the G0/G1 phase. NMR analysis suggests that **5**, **7**, and **12** interacts in a fast exchange regime with the terminal G-quartets (5' and 3'end) in a 2:1 stoichiometry.

2.18 Experimental Section

Synthesis

Chemistry. All solvents and reagents were purified by standard techniques or used as supplied from commercial sources (Sigma-Aldrich Corporation unless stated otherwise). All reactions were generally carried out under inert atmosphere unless otherwise noted. TLC was performed on Kieselgel 60 F254 plates, and spots were visualized under UV light. Products were purified by flash chromatography on silica gel (100-200 mesh). ¹H NMR spectra were recorded on 600 MHz instruments at 298 K. ¹³C NMR spectra were recorded at 151 MHz with proton decoupling. Chemical shifts are reported in parts per million (ppm) and are referred to the residual solvent peak. The following notations are used: singlet (s); doublet (d); triplet (t); quartet (q); multiplet (m); broad (br). Coupling constants are quoted in Hertz and are denoted as *J*. Mass spectra were recorded on a Micromass Q-ToF (ESI) spectrometer. Purity of the for key target compounds are detected by HPLC system [Waters e2695 (alliance), Column: GRACE-C18 (250 mm x 4.6 mm x 5 mm), Mobile phase: 1% TFA Buffer and ACN & Methanol, Flow rate: 1 mL/min] consisting of low pressure gradient pump plus auto sampler and Photo Diode Array (PDA) detector. The output signal was monitored and processed using Empower 2 software.

Generalized procedure for the reduction of nitro to amine: a solution of a nitro conjugate in 10 mL of ethanol was dropped into a suspension of 10% Pd/C (12.5 mg) saturated with H₂ in ethanol (20 ml). The mixture was stirred at room temperature for 3 h, according to TLC analysis (MeOH/DCM 1:3). The catalyst was recovered and the filtrate was evaporated under reduced pressure to dryness to give the corresponding amino-indole.

Generalized procedure for intermediates **6, **8**, **14**, **2a** and **20** with Vilsmeier Haack reaction (A):** POCl₃ (1.2 ml, 9.25 mmol, 1.5 eq.) as reaction solvent was added in dropwise manner to the dried DMF (10 ml) under inert condition at 0 °C. This mixture was allowed to stir at room

temperature for 1 hour followed with the addition of reactant (1 eq) **1**, **1a**, **5**, **13** and **19a**. The reaction mixture was then allowed to react for 1 hour by stirring at room temperature. The reaction mixture was then quenched with equal amount of ice water and 50 % NaOH was added until a pH of 9.

Generalized procedure for the condensation of substituted carboxaldehyde and corresponding amines (B): Substituted-carboxaldehyde **2a**, **1b**, **1c**, **1d** and **1e** (1eq) and substituted amines (3 eq) were dissolved in 50 mL of dry EtOH: CH₃CN (1:1). The resulting mixture stirred for 2 h at room temperature and then the solvent was concentrated under reduced pressure. The resultant residue was dissolved in 20 mL EtOH and then NaBH₄ (10 eq) was added to it in portion wise. The reaction mixture is allowed to stir for 24 h at room temperature, and then the excess of NaBH₄ was filtered off and the solvent was evaporated to dryness. The resultant solid was treated with deionized water and extracted with CH₂Cl₂ (3 x 50 mL). The organic phase was evaporated under reduced pressure. Generalized procedure for making HCl salt (C): The solid was dissolved in EtOH: dioxane (1/3) and precipitated with aqueous HCl 37% to obtain its hydrochloride salt. The precipitate was filtered and recrystallized with MeOH to give desired product

Generalized procedure for nucleophilic substitution of alkynes: 5-Nitro-1H-indole (2.01 g, 12.3 mmol) has added to a solution of anhydrous KOH (0.692 g, 12.3 mmol) in DMF (100 ml) at room temperature for 30 min. followed with the addition of 3-bromoprop-1-yne & 4-bromobut-1-yne (4 eq) and reaction mixture was allowed to stir overnight.

1-methyl-5-Amino-1H-indole (3): A solution of a 1-methyl-5-nitro-1H-indole (**2**) (5.87 mmol) is treated according to protocol A. Yield 96 %; brown solid. *R_f* (eluent MeOH/DCM 1:3) 0.44; Mp 87 °C; ¹H-NMR (600 MHz, DMSO-*d*₆) δ 7.09 (dd, *J* = 9.2, 5.8 Hz, 2H), 6.68 (s, 1H), 6.54 (d, *J* = 10.5 Hz, 1H), 6.10 (d, *J* = 2.9 Hz, 1H), 4.45 (s, 2H), 3.66 ppm (s, 3H). ¹³C NMR (151 MHz, DMSO-*d*₆): δ 211.30, 141.55, 130.56, 129.22, 111.82, 109.77, 103.59, 98.77, 32.65 ppm; HRMS: *m/z* calcd for C₉H₁₀N₂ [M+H]⁺ 146.19, found 146.19

5-nitro-1-(3-bromopropyl)-1H-Indole (4a): 5-Nitro-1H-indole (2.01 g, 12.3 mmol) was added to a solution of anhydrous KOH (0.692 g, 12.3 mmol) in DMF (100 ml) at room temperature for 30 min, followed by the addition of 1,3-dibromopropane (3.77 ml, 37.0 mmol). The reaction mixture was then allowed to stir overnight. Solvent was removed under reduced pressure, and the crude product was purified on silica as a stationary phase; Yield 46 % as yellow crystals; *R_f* = 0.65 (eluent EA/*n*-Hex 1:4); mp 102°C. ¹H NMR (600 MHz, CDCl₃): δ = 2.33-2.44 (m, 2H), 3.28-

3.35 (m, 2H), 4.40 (t, $J = 6.5$ Hz, 2H), 6.78 (d, $J = 3.1$ Hz, 1H), 7.67 (d, $J = 3.2$ Hz, 1H), 7.71 (d, $J = 9.1$ Hz, 1H), 8.04 (dd, $J = 2.3, 9.1$ Hz, 1H), 8.58 (d, $J = 2.2$ Hz, 1H). ^{13}C NMR (151 MHz, CDCl_3): $\delta = 30.05, 32.70, 44.53, 104.54, 109.33, 117.52, 118.42, 127.92, 131.22, 138.84, 141.79$ ppm; ESI-MS: m/z 282.0 [M+H]; HRMS: m/z calcd for $\text{C}_{11}\text{H}_{11}\text{BrN}_2\text{O}_2$ [M+H] + 282.0, found 282.0

1-(2-bromoethyl)-5-nitro-1H-indole (4b): 5-Nitro-1H-indole (2.01 g, 12.3 mmol) was added to a solution of anhydrous KOH (0.692 g, 12.3 mmol) in DMF (100 ml) at room temperature for 30 min, followed by the addition of 1, 2 -dibromoethane (3.77 ml, 37.0 mmol). The reaction mixture was then allowed to stir overnight. Solvent was removed under reduced pressure, and the crude product was purified on silica as a stationary phase; Yield 46 % as yellow crystals; $R_f = 0.65$ (eluent EA/n-Hex 1:4); mp 112 -114 °C. ^1H NMR (600 MHz, CDCl_3) δ 8.60 (d, $J = 1.9$ Hz, 1H), 8.18-8.10 (m, 1H), 7.37 (d, $J = 9.1$ Hz, 1H), 7.31 (d, $J = 3.2$ Hz, 1H), 6.72 (d, $J = 3.2$ Hz, 1H), 4.61-4.57 (m, 2H), 3.68 ppm (q, $J = 6.9$ Hz, 2H). ^{13}C NMR (151 MHz, CDCl_3): δ 142.07, 138.72, 131.15, 128.19, 118.50, 117.70, 107.70, 109.06, 104.85, 48.15, 29.73 ppm; ESI-MS: m/z 269.09 [M+H]; HRMS: m/z calcd for $\text{C}_{10}\text{H}_9\text{BrN}_2\text{O}_2$ [M+H] + 269.09, found 269.09

1, 3-bis (5-nitro-1H-indol-1-yl) propane (17a): Yield 46 % as yellow crystals; $R_f = 0.85$ (eluent EA/n-Hex 1:4); mp 112 - 114 °C. ^1H NMR (600 MHz, CDCl_3): δ 8.50 (t, $J = 5.0$ Hz, 2H), 8.09 (dd, $J = 9.1, 2.0$ Hz, 2H), 7.51 (d, $J = 3.4$ Hz, 2H), 7.43 - 7.39 (m, 2H), 7.20 - 7.12 (m, 2H), 6.73 (d, $J = 3.4$ Hz, 2H), 5.27 (dd, $J = 15.7, 1.1$ Hz, 2H), 4.91 (d, $J = 8.9$ Hz, 2H). ^{13}C NMR (151 MHz, CDCl_3): δ 30.05, 32.70, 44.53, 104.54, 109.33, 117.52, 118.42, 127.92, 131.22, 138.84, 141.79 ppm; ESI-MS: m/z 364.35 [M+H]⁺; HRMS: m/z calcd for $\text{C}_{19}\text{H}_{16}\text{N}_4\text{O}_4$ [M+H] + 364.35, found 364.35

5-nitro-1-(3-(pyrrolidin-1-yl) propyl)-1H-indole (5): Purified 5-nitro-1-(3-bromopropyl)-1H-Indole (4) (0.24 g, 0.88 mmol) was dissolved in dry ACN (6 mL/mmol) and to this pyrrolidine (3-10 equiv) was added, and the mixture was refluxed for 3-4 h. Solvent was concentrated under reduced pressure. The crude product was purified by column chromatography (eluent MeOH/DCM 1:3); using silica as a stationary phase; Yield 36 % as yellow viscous solid; R_f (eluent MeOH/ DCM 1:4) 0.35; mp 112-114 °C. ^1H NMR (600 MHz, $\text{DMSO}-d_6$): δ 8.58 (d, $J = 2.2$ Hz, 1H), 8.05 (dd, $J = 9.1, 2.3$ Hz, 1H), 7.74 (d, $J = 9.1$ Hz, 1H), 7.68 (d, $J = 3.1$ Hz, 1H), 6.79 (d, $J = 3.2$ Hz, 1H), 4.36 (t, $J = 7.0$ Hz, 2H), 3.03 (s, 4H), 2.92 (s, 2H), 2.08-2.14 (br, 2H), 1.85 ppm (s, 4H). ^{13}C NMR (151 MHz, $\text{DMSO}-d_6$): δ 140.79, 138.63, 132.38, 127.37, 117.58, 116.40, 110.34, 103.91, 53.31, 51.60, 43.43, 26.93, 22.71 ppm; ESI-MS: m/z 273.33 [M+H]⁺; HRMS: m/z calcd for $\text{C}_{15}\text{H}_{19}\text{N}_3\text{O}_2$ [M+H] + 273.15, found 273.15

5-nitro-1-(3-(pyrrolidin-1-yl)propyl)-1H-indole-3-carbaldehyde (6): The crude product is obtained from 5-nitro-1-(3-(pyrrolidin-1-yl)propyl)-1H-indole (5) by Vilsmeier Haack general protocol. Further purified on silica as a stationary phase with 1:4 MeOH: DCM. Yield 56 % as yellow solid; R_f (eluent MeOH/ DCM 1:4) 0.25; mp 132-134 °C. $^1\text{H-NMR}$ (600 MHz, DMSO- d_6): δ 9.99 (s, 1H), 8.93 (s, 1H), 8.60 (s, 1H), 8.18 (d, $J = 9.0$ Hz, 1H), 7.88 (d, $J = 9.1$ Hz, 1H), 4.41 (t, $J = 6.7$ Hz, 2H), 3.40 (s, 4H), 2.37-2.40 (br, 2H), 1.98 - 2.00 (br, 2H), 1.67 ppm (s, 4H). $^{13}\text{C NMR}$ (151 MHz, DMSO- d_6): δ 185.08, 143.87, 143.40, 140.06, 123.87, 119.06 - 118.89, 118.35, 117.11, 111.96, 53.28, 51.95, 44.88, 28.32, 23.08 ppm; HRMS: m/z calcd for $\text{C}_{16}\text{H}_{19}\text{N}_3\text{O}_3$ [M+H] $^+$ 301.14, found 301.14

N-((5-nitro-1-(3-(pyrrolidin-1-yl)propyl)-1H-indol-3-yl)methyl)-2-(pyrrolidin-1-yl)ethanamine (7): In a round bottom flask, 8-carboxaldehyde-5-nitro-1-[3-(1-pyrrolidinyl)propyl]-1H-Indole (6) (6) (6 g, 0.0199 mol) and 2-(pyrrolidin-1-yl)ethanamine (4.54 g, 0.039 mol) were reacted according to protocol B. The product was purified on silica as a stationary phase with 1:4 MeOH: DCM. Yield 46 % as yellow viscous solid; R_f (eluent MeOH/ DCM 1:4) 0.15; mp 102-104 °C; $^1\text{H-NMR}$ (600 MHz, DMSO- d_6): δ 8.63 (d, $J = 2.3$ Hz, 1H), 8.03 (dt, $J = 9.1, 2.1$ Hz, 1H), 7.66 (dd, $J = 9.1, 4.7$ Hz, 1H), 7.53 (s, 1H), 4.31 - 4.23 (m, 2H), 3.92 (s, 2H), 2.66 (dd, $J = 13.0, 6.4$ Hz, 2H), 2.36 (t, $J = 9.6$ Hz, 8H), 2.27 (tt, $J = 2.2, 11.1$ Hz, 5H), 1.98 - 1.87 (m, 4H), 1.75 - 1.55 (m, 6H); $^{13}\text{C NMR}$ (151 MHz, DMSO- d_6): δ 185.08, 143.87, 143.40, 140.06, 123.87, 119.06 - 118.89, 118.35, 117.11, 111.96, 53.28, 51.95, 44.88, 28.32, 23.08 ppm. HRMS: m/z calcd for $\text{C}_{22}\text{H}_{33}\text{N}_5\text{O}_2$ [M+H] $^+$ 399.26, found 399.26

5-Nitroindole-3-carboxaldehyde (8): The crude product was obtained from 5-nitro-Indole (1) by the Vilsmeier Haack general protocol. The product was purified on silica as a stationary phase; with 1:4 MeOH: DCM. Yield 85 % as yellow amorphous solid; R_f (eluent MeOH/ DCM 1:4) 0.55; M.p: 109-111 °C; $^1\text{H-NMR}$ (600 MHz, CDCl_3) δ 9.64 (s, 1H), 8.69 (d, $J = 1.8$ Hz, 1H), 7.73 (m, 2H), 7.18 ppm (d, $J = 6$ Hz, 1H), 2.50 - 2.62 (br, 1H); $^{13}\text{C-NMR}$: (151 MHz, CDCl_3) δ 183.97, 138.90, 118.13, 117.38, 111.91 ppm; HRMS (ESI): m/z calculated for $\text{C}_9\text{H}_6\text{N}_2\text{O}_3$: 190.04, found: 191.09

N,N-dimethyl-1-(5-nitro-1H-indol-3-yl)methanamine (9): 5-Nitroindole-3-carboxaldehyde (8) (1eq) and 3 equivalent of dimethylamine (40% aqueous, 60 mL) allowed to react according to the protocol B. The crude product is obtained as yellow crystalline solid and further purified on silica as a stationary phase with DCM: MeOH (9:1) over silica gel column. Yield 60 % as yellow solid; R_f (eluent MeOH/ DCM 1:4) 0.45; mp 114 - 116 °C. $^1\text{H-NMR}$ (600 MHz, DMSO- d_6) δ 11.71 (s, 1H), 8.62 (d, $J = 1.9$ Hz, 1H), 8.00 (dt, $J = 11.6, 4.2$ Hz, 2H), 7.56 - 7.51 (m, 1H), 3.68 (s, 2H), 2.22

ppm (s, 6H). ¹³C-NMR: (151 MHz, DMSO-*d*₆) δ 140.09, 139.47, 132.50, 128.31, 126.81, 116.34, 111.59, 69.59, 54.07, 44.65 ppm; HRMS (ESI): *m/z* calculated for C₁₁H₁₃N₃O₂: 219.10, found: 219.10

N,N-dimethyl-1-(5-amino-1H-indol-3-yl)methanamine (9a): Compound **9** was treated with protocol **A**. The crude product was purified with DCM: MeOH (9:1) over silica gel column. (**10a**) Yield 96 %; Brown solid. *R_f* (eluent MeOH/DCM 1:3) 0.41; Mp 87 - 89 °C; ¹H-NMR (600 MHz, DMSO-*d*₆) δ 10.38 (s, 1H), 7.04 - 6.99 (m, 2H), 6.76 (s, 1H), 6.47 (dd, *J* = 8.6, 1.9 Hz, 1H), 3.31-3.32 (b, 2H) 4.41 (s, 2H), 2.12 ppm (s, 6H). ¹³C-NMR: (151 MHz, DMSO-*d*₆) δ 211.06, 140.62, 129.95, 128.50, 124.07, 111.63, 111.20, 110.40, 102.17, 54.50, 44.83 ppm; HRMS (ESI): *m/z* calculated for C₁₁H₁₅N₃: 189.13, found: 189.13

N1,N1-dimethyl-N2-((5-nitro-1H-indol-3-yl)methyl)ethane-1,2-diamine (10): 5-Nitroindole-3-carboxaldehyde (**8**) (6 g, 62.5 mmol) and N1,N1-dimethylethane-1,2-diamine (40% aqueous, 60 mL) were allowed to react according to the protocol **B**. Yield 36 %, yellow solid; *R_f* (Eluent MeOH/DCM 1:3) 0.23; Mp 162 -164 °C. ¹H-NMR: (600 MHz, CDCl₃) δ 11.68 (s, 1H), 8.53 (dd, *J* = 8.6, 3.6 Hz, 1H), 8.02 -7.92 (m, 1H), 7.57- 7.45 (m, 2H), 3.77 (s, 2H), 3.46 (bs, 1H), 2.54 (t, *J* = 4.6 Hz, 2H), 2.46 (t, *J* = 6.3 Hz, 2H), 2.06 ppm (s, 6H). ¹³C-NMR: (151 MHz, CDCl₃) δ 140.27, 139.42, 128.91, 126.30, 116.41, 116.42, 116.10, 112.11, 51.00, 46.25, 45.35, 43.05, 11.35 ppm; HRMS (ESI): *m/z* calculated for C₁₃H₁₈N₄O₂: 262.14, found: [M+H]⁺ = 263.14

N1,N1-diethyl-N2-((5-nitro-1H-indol-3-yl)methyl)ethane-1,2-diamine (11): 5-Nitroindole-3-carboxaldehyde (**8**) (6 g, 62.5 mmol) and N1,N1-diethylethane-1,2-diamine (40% aqueous, 60 mL) were allowed to react according to the protocol **B**. Yield 36 %, yellow solid; *R_f*(eluent MeOH/DCM 1:3) 0.23; Mp 152- 154 °C. ¹H-NMR (600 MHz, DMSO-*d*₆) δ 11.84 (s, 1H), 8.67 (s, 1H), 7.97 (d, *J* = 8.0 Hz, 1H), 7.70 (s, 1H), 7.62 - 7.47 (m, 1H), 4.09 (s, 2H), 3.78 (s, 1H), 2.73 (m, 2H), 2.57 (s, 2H), 2.41 (t, *J* = 17.3 Hz, 4H), 1.01 - 0.82 ppm (m, 6H). ¹³C-NMR (151 MHz, DMSO-*d*₆) δ 140.12, 139.57, 127.96, 126.76, 116.94, 116.83, 116.38, 115.11, 111.58, 57.01, 50.58, 49.05, 44.99 ppm; HRMS (ESI): *m/z* calculated for C₁₅H₂₂N₄O₂: 290.17, found = 290.17

N-((5-nitro-1H-indol-3-yl)methyl)-2-(pyrrolidin-1-yl)ethanamine (12): 5-Nitroindole-3-carboxaldehyde (**8**) (6 g, 62.5 mmol) and 2-(pyrrolidin-1-yl)ethanamine (40% aqueous, 60 mL) were allowed to react according to the protocol **B**. Yield 26 %, yellow solid; *R_f*(eluent MeOH/DCM 1:3) 0.13; Mp 162 -164 °C. ¹H-NMR: (600 MHz, CDCl₃) δ 11.70 (bs, 1H), 8.67 (t, *J* = 9.1 Hz, 1H), 8.05-7.97 (m, 1H), 7.58 - 7.53 (m, 2H), 4.01 (s, 2H), 3.48 (bs, 1H), 2.73 (t, *J* = 6.4 Hz, 2H), 2.58 (t, *J* = 6.4 Hz, 2H), 2.49 - 2.40 (m, 2H), 1.71-1.62 (m, 4H), 1.41 ppm (s, 2H). ¹³C-NMR: (151 MHz, CDCl₃)

δ 140.11, 139.61, 127.82, 126.18, 116.83, 116.19, 115.60, 111.86, 54.35, 53.43, 46.61, 43.35, 26.46, 22.85 ppm; HRMS (ESI): m/z calculated for $C_{15}H_{20}N_4O_2$: 288.16, found = 288.16

1H-indole-3-carbaldehyde (2a): The crude product was obtained from 1H-indole (5) by vilsmeier Haack general protocol A. Crude product was purified with DCM: MeOH (9:1) over silica gel column. Yield 66 %, White crystalline solid; R_f (eluent MeOH/DCM 1:3) 0.33; M.p: 123 - 125 °C; 1H NMR (600 MHz, $CDCl_3$): δ 9.60 (s, 1H), 7.83 (d, J = 6Hz, 2H), 7.53 (d, J = 6Hz, 1H), 7.10 (d, J = 6Hz, 1H), 6.89 - 6.84 ppm (m, 2H); ^{13}C NMR (151 MHz, $CDCl_3$) δ 184.11, 136.63, 122.92, 121.57, 120.63, 111.50 ppm. (ESI): m/z calculated for C_9H_7NO :145.10, found: 146.02

1-(1H-indol-3-yl)-N,N-dimethylmethanamine (3a): Yield 46 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.43; M.p: 143 - 145 °C. 1H -NMR (600 MHz, $DMSO-d_6$): δ 10.92 (s, 1H), 7.61 (d, J = 7.9 Hz, 1H), 7.36 (d, J = 8.1 Hz, 1H), 7.21 (t, J = 5.1 Hz, 1H), 7.08 (t, J = 7.5 Hz, 1H), 6.98 (t, J = 7.4 Hz, 1H), 3.55 (s, 2H), 2.17 (s, 6H). ^{13}C NMR (151 MHz, $DMSO-d_6$) δ 136.32, 127.52, 124.38, 120.85, 118.98, 118.29, 111.50, 111.24, 54.36, 44.82 ppm. HRMS: m/z calcd for $C_{11}H_{14}N_2$ $[M+H]^+$ 174.12, found 174.12

N1-((1H-indol-3-yl)methyl)-N2-(2-((2-aminoethyl)amino)ethyl)ethane-1,2-diamine (3b): Yield 46 %, white solid; 1H -NMR (600 MHz, D_2O): δ 10.62 (s, 1H), 7.78 (d, J = 7.88 Hz, 1H), 7.61 (d, J = 2.48 Hz, 1H), 7.58 (s, 1H), 7.33 (t, J = 7.34 Hz, 1H), 7.26 (t, J = 7.38 Hz, 1H), 4.57 (s, 2H), 3.39-3.48 (m, 22H) ppm; ^{13}C NMR (151 MHz, D_2O): δ 42.20, 42.93, 43.77, 43.92, 44.24, 103.58, 112.32, 118.01, 120.29, 122.62, 126.13, 128.20, 136.19 ppm; HRMS (ESI): calcd for $C_{17}H_{30}N_6$ 318.26; found $[M+H]^+$ 319.26

1-(1H-indazol-3-yl)-N,N-dimethylmethanamine (23a): Yield 36 %, white solid; R_f (eluent MeOH/DCM 1:3) 0.43; M.p: 143 - 145 °C. 1H -NMR (600 MHz, $DMSO-d_6$): δ 12.93 (d, J = 75.7 Hz, 1H), 8.00 (s, 1H), 7.59 (s, 1H), 7.47 (d, J = 8.6 Hz, 1H), 7.29 (d, J = 8.6 Hz, 1H), 3.44 (s, 2H), 2.12 ppm (s, 6H); ^{13}C NMR (151 MHz, $DMSO-d_6$) δ 211.08, 133.18, 130.77, 127.53, 122.53, 119.87, 109.83, 63.28, 44.66 ppm; HRMS (ESI): calcd for $C_{10}H_{13}N_3$ 175.11; found $[M+H]^+$ 176.11

N1-((1H-indazol-3-yl)methyl)-N2-(2-((2-aminoethyl)amino)ethyl)ethane-1,2-diamine (23c): Yield 66 %, white solid; 1H -NMR (600 MHz, D_2O): δ 8.14 (d, J = 6.0 Hz, 1H), 7.92 (s, 1H), 7.65 (t, J = 7.6 Hz, 1H), 7.46 (d, J = 7.8 Hz, 1H), 4.39 (s, 2H), 3.56 -2.68 ppm (m, 23H); ^{13}C NMR (151 MHz, D_2O) δ 140.22, 134.64, 128.54, 123.62, 122.79, 51.97, 43.52, 42.90, 42.30, 42.24 ppm. HRMS (ESI): calcd for $C_{16}H_{29}N_7$ 319.25; found $[M+H]^+$ 320.25

N-((1H-indazol-3-yl)methyl)-2-(pyrrolidin-1-yl)ethanamine (23f): Yield 56 %, white solid; R_f (eluent MeOH/DCM 1:3) 0.23; M.p: 173 - 175 °C. $^1\text{H-NMR}$ (600 MHz, $\text{DMSO-}d_6$): δ 7.92 (s, 1H), 7.56 (s, 1H), 7.39 (d, $J = 8.6$ Hz, 1H), 7.25 (d, $J = 8.6$ Hz, 1H), 3.70 (d, $J = 5.7$ Hz, 2H), 2.50 (s, 2H), 2.48 - 2.38 (m, 2H), 2.37 - 2.22 (m, 2H), 1.90 (d, $J = 5.8$ Hz, 4H), 1.65 - 1.48 (m, 4H); $^{13}\text{C NMR}$ (151 MHz, $\text{DMSO-}d_6$): δ 132.60, 126.55, 118.69, 109.46, 55.33, 55.39, 53.70, 53.25, 53.16, 47.53, 23.07 ppm. HRMS (ESI): calcd for $\text{C}_{14}\text{H}_{20}\text{N}_4$ 244.17; found $[\text{M}+\text{H}]^+$ 245.17

N,N-dimethyl-1-(1H-pyrrolo[2,3-b]pyridin-3-yl)methanamine (24a): Yield 56 %, white solid; R_f (eluent MeOH/DCM 1:3) 0.13; M.p: 143 - 145 °C. $^1\text{H NMR}$ (600 MHz, $\text{DMSO-}d_6$): δ 11.39 (bs, 1H), 8.22 - 8.17 (m, 1H), 8.00 (dd, $J = 6.3, 5.2$ Hz, 1H), 7.34 (s, 1H), 7.06 - 7.02 (m, 1H), 3.55 (s, 2H), 2.16 (d, $J = 4.3$ Hz, 6H). $^{13}\text{C NMR}$ (151 MHz, $\text{DMSO-}d_6$): δ 185.17, 148.80, 145.17, 142.64, 127.37, 124.83, 119.78, 118.43, 115.43, 114.72, 110.20, 54.14, 44.33; HRMS (ESI): calcd for $\text{C}_{10}\text{H}_{13}\text{N}_3$ 175.11; found $[\text{M}+\text{H}]^+$ 176.11

N1-((1H-pyrrolo[2,3-b]pyridin-3-yl)methyl)-N2-(2-((2-aminoethyl)amino)ethyl)ethane-1,2-diamine (24b): Yield 36 %, white solid. $^1\text{H NMR}$ (600 MHz, D_2O): δ 8.74 (dd, $J = 8.0, 7.8$ Hz, 1H), 8.41 (m, 1H), 8.02 - 7.78 (m, 1H), 7.58 (t, $J = 6.7$ Hz, 1H), 4.56 (d, $J = 9.3$ Hz, 2H), 3.76 - 2.72 (m, 23H), 1.38 - 1.05 (m, 4H); $^{13}\text{C NMR}$ (151 MHz, D_2O): δ 138.22, 137.01, 133.41, 131.41, 131.98, 124.50, 116.58, 105.39, 52.57, 51.88, 50.33, 48.99, 44.65, 43.65, 43.46. HRMS (ESI): calcd for $\text{C}_{16}\text{H}_{29}\text{N}_7$ 319.25; found $[\text{M}+\text{H}]^+$ 320.26

N1-((1H-pyrrolo[2,3-b]pyridin-3-yl)methyl)-N2-(2-((2-aminoethyl)amino)ethyl)ethane-1,2-diamine (24c): Yield 36 %, white solid. $^1\text{H NMR}$ (600 MHz, D_2O): δ 8.76 (d, $J = 16.4$ Hz, 1H), 8.40 (d, $J = 15.1$ Hz, 1H), 7.90 (s, 1H), 7.61 (d, $J = 6.3$ Hz, 1H), 4.61 (dd, $J = 30.0, 13.2$ Hz, 6H), 3.63 - 2.47 ppm (m, 14H); HRMS (ESI): calcd for $\text{C}_{14}\text{H}_{24}\text{N}_6$ 276.21; found $[\text{M}+\text{H}]^+$ 277.21

N-((1H-pyrrolo[2,3-b]pyridin-3-yl)methyl)-2-(pyrrolidin-1-yl)ethanamine (24f): Yield 46 %, white solid. $^1\text{H NMR}$ (600 MHz, $\text{DMSO-}d_6$): δ 11.44 (bs, 1H), 8.19 (dd, $J = 6.6, 3.8$ Hz, 1H), 8.00 (dd, $J = 8.5, 7.6$ Hz, 1H), 7.92 - 7.85 (m, 1H), 7.41 - 7.31 (m, 1H), 7.08 - 6.92 (m, 1H), 3.85 (s, 2H), 3.72 (s, 2H), 2.67 - 2.61 (m, 2H), 2.46 - 2.33 (m, 2H), 2.31 - 2.23 (m, 2H), 1.70 - 1.53 ppm (m, 4H). $^{13}\text{C NMR}$ (151 MHz, $\text{DMSO-}d_6$) δ 148.80, 142.41, 127.45, 126.94, 124.88, 123.80, 119.57, 119.17, 114.83, 112.83, 112.87, 111.47, 85.47, 55.30, 53.42, 49.62, 22.71 ppm; HRMS (ESI): calcd for $\text{C}_{14}\text{H}_{20}\text{N}_4$ 244.17; found $[\text{M}+\text{H}]^+$ 245.17

N1-(2-aminoethyl)-N2-(2-((2-(benzo[d][1,3]dioxol-5-yl)methyl)amino)ethyl)amino)ethyl)ethane-1,2-diamine (25b): Yield 46 %, white solid. ^1H

NMR (600 MHz, D₂O) δ 6.91 (dd, J = 17.5, 6.4 Hz, 4H), 5.96 (s, 2H), 4.25 – 4.05 (m, 2H), 3.59 – 2.73 ppm (m, 21H); ¹³C NMR (151 MHz, D₂O) δ 147.80, 138.45, 133.80, 131.13, 124.69, 116.42, 105.97, 51.84, 44.65, 43.51, 42.46, 41.57, 40.07, 35.20, 18.06 ppm; HRMS (ESI): calcd for C₁₆H₂₉N₅O₂ 323.23; found [M+H]⁺ 324.23

N1-(benzo[d][1,3]dioxol-5-ylmethyl)-N2,N2-dimethylethane-1,2-diamine (25e): Yield 56 %, white solid; R_f (eluent MeOH/DCM 1:3) 0.13; M.p: 103 - 105 °C. ¹H NMR (600 MHz, DMSO-*d*₆): δ 6.91 (s, 1H), 6.79 (dd, J = 19.5, 7.9 Hz, 2H), 5.97 (s, 1H), 3.63 (s, 2H), 2.54-2.56 (m, 2H), 2.33 (t, J = 6.4 Hz, 4H), 2.11 (s, 6H). ¹³C NMR (151 MHz, DMSO-*d*₆): δ 147.23, 145.69, 134.46, 120.98, 108.32, 107.66, 100.51, 58.51, 52.48, 45.85, 45.08 ppm; HRMS (ESI): calcd for C₁₂H₁₈N₂O₂ 222.14; found 222.14

N-(benzo[d][1,3]dioxol-5-ylmethyl)-2-(pyrrolidin-1-yl)ethanamine (25f): Yield 46 %, white solid. ¹H NMR (600 MHz, D₂O): δ 7.05 – 6.92 (m, 4H), 6.04 (s, 2H), 4.26 (s, 2H), 3.57 (dt, J = 14.6, 6.6 Hz, 8H), 2.12 (s, 4H). ¹³C NMR (151 MHz, D₂O): δ 148.02, 147.67, 124.35, 123.55, 117.89, 114.96, 108.80, 109.14, 101.74, 54.83, 51.53, 49.47, 41.94, 22.39; HRMS (ESI): calcd for C₁₄H₂₀N₂O₂ 248.15; found [M+H]⁺ 249.15

N1-(2-aminoethyl)-N2-(2-(((5-nitro-1H-indol-3-yl)methyl)amino)ethyl)amino)ethyl)ethane-1,2-diamine (13a): Yield 46 %, Yellow solid. ¹H NMR (600 MHz, D₂O): δ 8.59 (s, 1H), 8.00 (d, J = 9.2 Hz, 1H), 7.67 (d, J = 3.8 Hz, 1H), 7.48 (d, J = 9.1 Hz, 1H), 4.58 – 4.40 (m, 2H), 3.68 (s, 2H), 3.62 – 2.73 (m, 19H), 1.10 (t, J = 7.1 Hz, 2H). ¹³C NMR (151 MHz, D₂O): δ 141.10, 131.11, 125.63, 117.90, 115.39, 112.0, 65.0, 50.77, 43.0, 43.06, 42.34, 35.27, 16.84 ppm; HRMS (ESI): calcd for C₁₇H₂₉N₇O₂ 363.24; found 363.24

1-(3-azidopropyl)-5-nitro-1H-indole (14): To a stirring mixture of **4** (1g, 3.225 mmol, 1 eq) in 15 ml Dry DMF about 3 eq of sodium azide (0.6386 g, 9.667 mmol) were added and allowed to reflux for 3-5 hrs. Followed by quenching with Ice water and extraction with ethyl acetate. Organic extract was then concentrated under reduced pressure, and the crude product was purified on silica as a stationary phase with Cyclohexane: Ethylacetate (9:1) over silica gel column. Yield 56 %, yellow solid; R_f (eluent EA/cyclo 1:3) 0.63; Mp 92 - 94 °C. ¹H NMR (600 MHz, DMSO-*d*₆): δ 10.00 (s, 1H), 8.94 (d, J = 2.1 Hz, 1H), 8.62 (s, 1H), 8.19 (dd, J = 9.1, 2.3 Hz, 1H), 7.90 (d, J = 9.1 Hz, 1H), 4.43 (t, J = 7.0 Hz, 2H), 3.40 (t, J = 6.6 Hz, 2H), 2.07-2.11 ppm (br, 2H); ¹³C NMR (151 MHz, DMSO-*d*₆): δ 185.13, 146.04-144.33, 143.35, 139.92, 123.97, 118.75, 118.26, 117.17, 111.90, 47.89, 44.23, 28.53 ppm; HRMS (ESI): calcd for C₁₁H₁₁N₅O₂ 245.09; found [M+H]⁺ 246.09

1-(3-azidopropyl)-5-nitro-1H-indole-3-carbaldehyde (15): The crude product was further purified with Cyclohexane: Ethylacetate (9:1) over silica gel column. Yield 66 %, yellow solid; R_f (eluent EA/cyclo 1:3) 0.23; Mp 102 - 104 °C. ^1H NMR (600 MHz, DMSO- d_6): δ 10.00 (s, 1H), 8.94 (d, $J = 2.1$ Hz, 1H), 8.62 (s, 1H), 8.19 (dd, $J = 9.1, 2.3$ Hz, 1H), 7.90 (d, $J = 9.1$ Hz, 1H), 4.43 (t, $J = 7.0$ Hz, 2H), 3.40 (t, $J = 6.6$ Hz, 2H), 2.07-2.11 ppm (br, 2H); ^{13}C NMR (151 MHz, DMSO- d_6): δ 185.13, 146.04, 144.33, 143.35, 139.92, 123.97, 118.75, 118.26, 117.17, 111.90, 47.89, 44.23, 28.53 ppm; HRMS (ESI): calcd for $\text{C}_{12}\text{H}_{11}\text{N}_5\text{O}_3$ 273.09; found 273.09

1-(1-(3-azidopropyl)-5-nitro-1H-indol-3-yl)-N,N-dimethylmethanamine (16a): compound **8** (1g, 3.663 mmol, 1 eq) and 3 eq of N1, N1-dimethylethane-1,2-diamine (1.284 g, 11.070 mmol) were allowed to react according to generalized procedure **B**. Crude product was then further purified with DCM: MeOH (9:1) over silica gel column. Yield 26 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.43; Mp 102 - 104 °C. ^1H -NMR (600 MHz, DMSO- d_6): δ 8.66 (d, $J = 1.8$ Hz, 1H), 8.03 (dd, $J = 9.1, 2.1$ Hz, 1H), 7.67 (d, $J = 9.1$ Hz, 1H), 7.56 (s, 1H), 4.29 (t, $J = 6.9$ Hz, 2H), 3.95 (s, 2H), 3.30 - 3.32 (bs, 4H), 2.64 (t, $J = 6.4$ Hz, 2H), 2.44 (q, $J = 7.1$ Hz, 2H), 1.99 - 2.03 (bs, 2H); ^{13}C NMR (151 MHz, DMSO- d_6): δ 140.35, 139.09, 126.55, 116.58, 110.24, 51.79, 47.98, 46.34, 43.59, 43.09, 28.90, 11.61 ppm. HRMS (ESI): calcd for $\text{C}_{14}\text{H}_{18}\text{N}_6\text{O}_2$ 302.15; found 302.15

N,N-dimethyl-N'-[[5-nitro-1-(3-Azidopropyl)-1H-indole]methyl]-1,2-Ethanediamine

(16b): The compound **8** (1g, 3.663 mmol, 1 eq) and about 3 eq of N1, N1-diethylethane-1,2-diamine (1.284 g, 11.070 mmol) were allowed to react according to generalized procedure **B**. Crude product was then further purified with DCM: MeOH (9:1) over silica gel column. Yield 46 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.33; Mp 162 - 164 °C. ^1H NMR (600 MHz, DMSO- d_6) δ 8.63 (d, $J = 2.3$ Hz, 1H), 8.60 (d, $J = 2.3$ Hz, 1H), 8.03 (dt, $J = 9.1, 2.1$ Hz, 2H), 7.66 (dd, $J = 9.1, 4.7$ Hz, 2H), 7.55 (d, $J = 5.6$ Hz, 1H), 7.54 (d, $J = 9.6$ Hz, 1H), 4.35 - 4.21 (m, 4H), 3.92 (s, 2H), 2.71 - 2.60 (m, 3H), 2.35 (dd, $J = 18.3, 2.7$ Hz, 2H), 2.24 (dd, $J = 14.1, 7.7$ Hz, 2H), 2.02 - 1.87 (m, 2H), 1.76 - 1.49 ppm (m, 4H). ^{13}C NMR (151 MHz, DMSO- d_6): δ 140.11, 139.22, 130.92, 130.39, 126.30, 116.70, 116.47, 116.12, 114.59, 110.20, 85.45, 55.25, 53.42, 53.03, 52.79, 51.84, 47.46, 43.62, 30.81, 28.75, 24.19, 23.06 ppm; HRMS (ESI): calcd for $\text{C}_{18}\text{H}_{27}\text{N}_7\text{O}_2$ 373.22; found 374.22

5-nitro-1-(prop-2-yn-1-yl)-1H-indole (19a): Yield 36 %, yellow solid; R_f (eluent EA/cyclo 1:3) 0.43; M.p: 113 - 115 °C. ^1H NMR (600 MHz, CDCl_3) δ 8.57 (d, $J = 2.2$ Hz, 1H), 8.17 - 8.11 (m, 1H), 7.43 (d, $J = 9.1$ Hz, 1H), 7.37 (d, $J = 3.3$ Hz, 1H), 6.71 (d, $J = 3.3$ Hz, 1H), 4.93 (d, $J = 2.6$ Hz, 2H), 2.48 (t, $J = 2.6$ Hz, 1H); ^{13}C NMR (151 MHz, CDCl_3) δ 142.04, 138.04, 130.46, 128.18, 118.25, 117.63, 109.37, 104.64, 76.33, 74.60, 36.27 ppm; HRMS (ESI): calcd for $\text{C}_{11}\text{H}_8\text{N}_2\text{O}_2$ 200.06; found 201.06

1-(but-3-yn-1-yl)-5-nitro-1H-indole (19b): Yield 26 %, yellow solid; R_f (eluent EA/cyclo 1:3) 0.33; M.p: 113 - 115 °C. $^1\text{H-NMR}$ (600 MHz, $\text{DMSO-}d_6$): δ 8.52 (t, $J = 2.7$ Hz, 1H), 8.08 - 8.03 (m, 1H), 7.32 (d, $J = 9.1$ Hz, 1H), 7.25 (t, $J = 2.7$ Hz, 1H), 6.62 (d, $J = 3.0$ Hz, 1H), 4.28 (q, $J = 6.9$ Hz, 2H), 2.65 (tt, $J = 5.4, 2.7$ Hz, 2H), 1.98 (dt, $J = 5.3, 2.7$ Hz, 1H); $^{13}\text{C-NMR}$ (151 MHz, $\text{DMSO-}d_6$) δ 131.54, 128.56, 118.1, 118.09, 109.41, 105.17, 80.2, 72.14, 46.39, 21.12 ppm; HRMS (ESI): calcd for $\text{C}_{12}\text{H}_{10}\text{N}_2\text{O}_2$ 214.07; found 215.07

5-nitro-1-(prop-2-yn-1-yl)-1H-indole-3-carbaldehyde (20): The crude product was purified on silica as a stationary phase; with 1:4 EA: CH. The product was a yellow solid. $^1\text{H-NMR}$ (600 MHz, $\text{DMSO-}d_6$): δ 10.01 (s, 1H), 8.93 (t, $J = 3.6$ Hz, 1H), 8.61 (s, 1H), 8.23 (dd, $J = 9.1, 2.4$ Hz, 1H), 7.87 (d, $J = 9.1$ Hz, 1H), 5.33 (d, $J = 2.5$ Hz, 2H), 3.61 (t, $J = 2.5$ Hz, 1H); $^{13}\text{C-NMR}$ (151 MHz, $\text{DMSO-}d_6$): δ 185.49, 143.43, 142.92, 139.37, 123.94, 118.99, 118.41, 117.24, 112.08, 77.32, 36.41 ppm

N,N-dimethyl-1-(5-nitro-1-(prop-2-yn-1-yl)-1H-indol-3-yl)methanamine (21a): 5-nitro-1-(prop-2-yn-1-yl)-1H-indole-3-carbaldehyde and 3 equivalent of dimethylamine (40% aqueous, 60 mL) were allowed to react according to the protocol B. $^1\text{H-NMR}$ (600 MHz, $\text{DMSO-}d_6$): δ 8.61 - 8.59 (d, $J = 2.1$ Hz, 1H), 8.11 (dd, $J = 9.1, 2.2$ Hz, 1H), 7.72 (s, 1H), 7.75 (d, $J = 9.1$ Hz, 1H), 5.17 (d, $J = 2.4$ Hz, 2H), 4.29 (s, 2H), 3.61 (bs, 1H), 2.17 - 2.18 (m, 6H). $^{13}\text{C NMR}$ (151 MHz, $\text{DMSO-}d_6$): δ 141.08, 139.16, 131.14, 127.62, 117.15, 116.99, 116.92, 115.37, 110.96, 78.78, 76.28, 53.96, 44.96, 35.93 ppm; HRMS (ESI): calcd for $\text{C}_{14}\text{H}_{15}\text{N}_3\text{O}_2$ 257.12; found 258.12

N1,N1-diethyl-N2-((5-nitro-1-(prop-2-yn-1-yl)-1H-indol-3-yl)methyl)ethane-1,2-diamine (21b): 5-nitro-1-(prop-2-yn-1-yl)-1H-indole-3-carbaldehyde and N1,N1-diethylethane-1,2-diamine (40% aqueous, 60 mL) were allowed to react according to the protocol B. $^1\text{H-NMR}$ (600 MHz, $\text{DMSO-}d_6$): δ 8.83 (d, $J = 2.1$ Hz, 1H), 8.13 (dd, $J = 9.1, 2.2$ Hz, 1H), 7.82 (s, 1H), 7.75 (d, $J = 9.1$ Hz, 1H), 5.25 (d, $J = 2.4$ Hz, 2H), 4.29 (s, 2H), 3.61 (bs, 2H), 3.01-2.93 (m, 2H), 2.84-2.74 (m, 2H), 2.67 (q, $J = 7.1$ Hz, 4H), 1.05 - 0.96 (m, 6H); $^{13}\text{C NMR}$ (151 MHz, $\text{DMSO-}d_6$): δ 140.81, 138.29, 132.66, 126.62, 117.03, 116.61, 110.87, 110.69, 78.7, 76.22, 48.79, 46.14, 43.39, 41.20, 35.69, 10.11 ppm; HRMS (ESI): calcd for $\text{C}_{18}\text{H}_{24}\text{N}_4\text{O}_2$ 328.19; found 329.19

N-benzyl-2-bromoacetamide (26b): White solid (Crude); $^1\text{H NMR}$ (600 MHz, $\text{DMSO-}d_6$) δ 8.81 (s, 1H, NH), 7.44 (dt, $J = 10.7, 5.1$ Hz, 2H), 7.32 (d, $J = 7.3$ Hz, 2H), 7.27 (s, 1H), 4.29 (d, $J = 5.9$ Hz, 2H), 3.92 (s, 2H). $^{13}\text{C NMR}$ (151 MHz, $\text{DMSO-}d_6$) δ 166.1, 138.8, 134.0, 128.9, 128.6, 128.5, 128.3, 127.3, 126.9, 42.5, 29.5.

2-bromo-N-(4-fluorobenzyl)acetamide (27b): (Crude); White solid; ^1H NMR (600 MHz, DMSO-*d*₆) δ 8.79 (s, 1H), 7.29 (d, *J* = 5.7 Hz, 2H), 7.16 (t, *J* = 8.5 Hz, 2H), 4.27 (d, *J* = 5.5 Hz, 2H), 3.90 (s, 2H); ^{13}C NMR (151 MHz, DMSO-*d*₆) δ 166.1, 135.1, 135.0, 129.3, 129.2, 115.2, 114.9, 41.8, 29.4.

2-bromo-N-(4-(trifluoromethyl) benzyl) acetamide (28b): White solid (Crude); ^1H NMR (600 MHz, DMSO-*d*₆) δ 8.90 (s, 1H), 7.70 (d, *J* = 7.9 Hz, 2H), 7.48 (d, *J* = 7.9 Hz, 2H), 4.39 (d, *J* = 5.7 Hz, 2H), 3.93 (s, 2H). ^{13}C NMR (151 MHz, DMSO-*d*₆) δ 166.3, 143.8, 127.9, 125.3, 125.3, 125.2, 125.2, 125.2, 42.1, 29.3.

2-bromo-N-(3,4,5-trimethoxybenzyl)acetamide (29b): White solid (Crude); ^1H NMR (600 MHz, DMSO-*d*₆) δ 6.70 (s, 2H), 6.57 (s, 1H), 4.74 (s, 2H), 3.73 (s, 6H), 3.63 (s, 3H), 3.93 (s, 2H); ^{13}C NMR (151 MHz, DMSO-*d*₆) δ 174.30, 167.39, 187.94, 156.98, 136.75, 132.53, 105.13, 60.50, 59.30, 55.70, 32.7

2-bromo-N-(4-methoxybenzyl)acetamide (30b): White solid (Crude); ^1H NMR (600 MHz, DMSO-*d*₆) δ 8.70 (s, 1H), 7.21 (t, *J* = 8.8 Hz, 2H), 6.95 – 6.84 (m, 2H), 4.23 (d, *J* = 5.9 Hz, 2H), 3.90 (s, 2H), 3.75 (s, 3H); ^{13}C NMR (151 MHz, DMSO-*d*₆) δ 174.30, 167.39, 187.94, 156.98, 136.75, 132.53, 105.13, 60.50, 59.30, 55.70, 32.7

2-amino-1-benzyl-5-oxo-4,5-dihydro-1H-pyrrole-3-carbonitrile (26c): Orange solid (Crude); ^1H NMR (600 MHz, DMSO-*d*₆) δ 7.71 (d, *J* = 8.1 Hz, 2H), 7.42 (d, *J* = 8.0 Hz, 2H), 7.27 (s, 2H), 4.84 (s, 2H), 3.30 (s, 2H); ^{13}C NMR (101 MHz, DMSO-*d*₆) δ 173.9, 157.5, 141.3, 127.6, 125.4, 125.4, 125.3, 123.3, 119.1, 47.3, 41.4, 34.3, 8.4.

2-amino-1-(4-fluorobenzyl)-5-oxo-4,5-dihydro-1H-pyrrole-3-carbonitrile (27c): Orange solid (Crude); ^1H NMR (600 MHz, DMSO-*d*₆) δ 7.27 (dd, *J* = 8.5, 5.6 Hz, 2H), 7.24 (s, 2H), 7.16 (t, *J* = 8.9 Hz, 2H), 4.71 (s, 2H), 3.27 (s, 2H); ^{13}C NMR (151 MHz, DMSO-*d*₆) δ 173.9, 157.6, 132.7, 129.1, 119.2, 115.3, 115.1, 47.2, 41.0, 34.3, 8.4.

2-amino-5-oxo-1-(4-(trifluoromethyl)benzyl)-4,5-dihydro-1H-pyrrole-3-carbonitrile (28c): Orange solid (Crude); ^1H NMR (600 MHz, DMSO-*d*₆) δ 7.71 (d, *J* = 8.1 Hz, 2H), 7.42 (d, *J* = 8.0 Hz, 2H), 7.27 (s, 2H), 4.84 (s, 2H), 3.30 (s, 2H); ^{13}C NMR (151 MHz, DMSO-*d*₆) δ 173.9, 157.5, 141.3, 127.6, 125.4, 125.4, 125.3, 123.3, 119.1, 47.3, 41.4, 34.3, 8.4.

2-amino-5-oxo-1-(3,4,5-trimethoxybenzyl)-4,5-dihydro-1H-pyrrole-3-carbonitrile (29c): Orange solid (Crude); ^1H NMR (600 MHz, DMSO-*d*₆) δ 7.23 (s, 2H), 7.19 (t, *J* = 5.8 Hz, 2H), 6.92 –

6.84 (m, 2H), 4.68 (s, 2H), 3.74 (s, 3H), 3.26 (s, 2H); ¹³C NMR (151 MHz, DMSO-*d*₆) δ 174.05, 158.64, 157.45, 130.61, 128.54, 118.97, 113.77, 54.83, 46.68, 34.35, 8.67

2-amino-1-(4-methoxybenzyl)-5-oxo-4,5-dihydro-1H-pyrrole-3-carbonitrile (30c):

Orange solid (Crude); ¹H NMR (600 MHz, DMSO-*d*₆) δ 7.21 (s, 2H), 6.60 (d, *J* = 8.5 Hz, 2H), 4.65 (d, *J* = 7.1 Hz, 2H), 4.45 (d, *J* = 6.1 Hz, 2H), 3.75 (s, 6H), 3.65 (s, 3H). ¹³C NMR (151 MHz, DMSO-*d*₆) δ 173.84, 157.46, 152.67, 137.00, 132.00, 118.98, 112.21, 104.89, 59.63, 47.02, 41.84, 34.37, 8.63

ethyl 4-amino-1-benzyl-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (26d):

White solid (Crude); ¹H NMR (600 MHz, DMSO-*d*₆) 7.27-7.39 (m, 5H), 6.58 (s br, 2H), 4.27 (q, *J* = 7.1 Hz, CH₂, 2H), 3.43 (s, CH₂, 2H), 2.47 (s, CH₃, 3H), 1.31 (t, *J* = 7.1 Hz, CH₃, 3H); ¹³C NMR (151 MHz, DMSO-*d*₆) δ 174.4, 173.9, 149.4, 137.3, 137.5, 136.5, 128.4, 128.3, 127.4, 127.2, 127.1, 126.8, 97.4, 60.5, 41.6, 34.3, 32.9, 25.6, 14.0.

ethyl 4-amino-1-(4-fluorobenzyl)-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (27d):

White solid (Crude); ¹H NMR (600 MHz, DMSO-*d*₆) δ 7.35 (dd, *J* = 8.3, 5.8 Hz, 2H, Aromatic H), 7.12 (t, *J* = 8.9 Hz, 2H; aromatic H), 6.62 (s, 1H, NH), 4.80 (s, 2H, CH₂), 4.29 (q, *J* = 7.1 Hz, 2H, CH₂), 3.43 (d, *J* = 12.8 Hz, 2H, CH₂), 2.47 (s, 3H, CH₃), 1.30 (t, *J* = 7.1 Hz, 3H, CH₃). ¹³C NMR (151 MHz, DMSO-*d*₆) δ 174.4, 167.6, 162.1, 160.5, 156.7, 149.4, 133.4, 129.6, 129.5, 115.1, 115.0, 106.9, 97.4, 60.5, 41.0, 33.0, 25.4, 14.0.

ethyl 4-amino-6-methyl-2-oxo-1-(4-(trifluoromethyl)benzyl)-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (28d):

White solid (Crude); ¹H NMR (600 MHz, DMSO-*d*₆) δ 7.66 (d, *J* = 8.1 Hz, 2H), 7.50 (d, *J* = 8.0 Hz, 2H), 6.60 (s, 2H), 4.90 (s, 2H), 4.29 (q, *J* = 7.1 Hz, 2H), 3.45 (s, 2H), 2.45 (s, 3H), 1.29 (s, 3H); ¹³C NMR (151 MHz, DMSO-*d*₆) δ 174.5, 167.7, 158.4, 156.8, 149.4, 142.0, 128.1, 125.3, 123.3, 106.9, 97.5, 60.4, 49.6, 41.2, 33.0, 30.0, 25.6, 14.0.

ethyl 4-amino-6-methyl-2-oxo-1-(3,4,5-trimethoxybenzyl)-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (29d):

White solid (Crude); ¹H NMR (600 MHz, DMSO-*d*₆) δ 6.69 (d, *J* = 11.6 Hz, 2H), 6.58 (bs, 2H), 4.74 (s, 2H), 4.31 (q, *J* = 7.1 Hz, 2H), 3.72 (d, *J* = 7.0 Hz, 6H), 3.66 – 3.59 (m, 3H), 3.43 (s, 3H), 1.32 (t, *J* = 7.1 Hz, 2H). ¹³C NMR (151 MHz, DMSO-*d*₆) δ 174.5, 167.7, 158.4, 156.8, 149.4, 142.0, 128.1, 125.3, 123.3, 106.9, 97.5, 60.4, 49.6, 41.2, 33.0, 30.0, 25.6, 14.0.

ethyl 4-amino-1-(4-methoxybenzyl)-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (30d):

White solid (Crude); ¹H NMR (600 MHz, DMSO-*d*₆) δ 7.46 (d, *J* = 7.7 Hz, 2H), 6.83 (d, *J* = 8.0 Hz, 2H), 5.74 (s, 2H), 4.92 (s, 2H), 4.38 (d, *J* = 6.9 Hz, 2H), 3.78 (s, 3H),

3.29 (s, 2H), 2.72 (s, 3H), 1.42 (t, $J = 6.8$ Hz, 3H). ^{13}C NMR (151 MHz, DMSO- d_6) δ 174.5, 167.7, 158.4, 156.8, 149.4, 142.0, 128.1, 125.3, 123.3, 106.9, 97.5, 60.4, 49.6, 41.2, 33.0, 30.0, 25.6, 14.0.

1H-indol-5-amine (31): brown (Crude): ^1H NMR (600 MHz, DMSO- d_6) δ 10.57 (s, 1H), 7.16 – 7.07 (m, 2H), 6.70 (d, $J = 2.0$ Hz, 1H), 6.51 (dd, $J = 8.5, 2.1$ Hz, 1H), 6.14 (d, $J = 2.2$ Hz, 1H), 4.41 (s, 2H). ^{13}C NMR (151 MHz, DMSO- d_6) δ 141.22, 129.63, 128.36, 124.57, 111.79, 111.49, 103.24, 99.46

N1-((5-amino-1H-indol-3-yl)methyl)-N2,N2-diethylethane-1,2-diamine(32): brown (Crude): ^1H NMR (600 MHz, DMSO- d_6) δ 10.32 (s, 1H), 7.03 (d, $J = 3.8$ Hz, 1H), 7.00 (d, $J = 2.3$ Hz, 2H), 6.71 (d, $J = 1.9$ Hz, 1H), 6.46 (dd, $J = 8.5, 2.0$ Hz, 2H), 3.71 (s, 2H), 3.17 (d, $J = 4.9$ Hz, 2H), 2.64 – 2.55 (m, 2H), 2.48 – 2.39 (m, 6H), 0.95 – 0.88 (m, 6H).

Experimental procedure for fragment screening with FID assays, ^1D -NMR titration, cells and culture conditions, cell proliferation assay, western blot analysis, c-MYC expression and immuno blotting, cell cycle analysis and molecular modelling are presented in chapter 4 of this thesis

2.19 References

- (1) Verdian Doghaei, A.; Housaindokht, M. R.; Bozorgmehr, M. R. Molecular Crowding Effects on Conformation and Stability of G-Quadruplex DNA Structure: Insights from Molecular Dynamics Simulation. *J. Theor. Biol.* 2015. <https://doi.org/10.1016/j.jtbi.2014.09.015>.
- (2) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.* 2006. <https://doi.org/10.1093/nar/gkl655>.
- (3) Zhang, Z.; Dai, J.; Veliath, E.; Jones, R. A.; Yang, D. Structure of a Two-G-Tetrad Intramolecular G-Quadruplex Formed by Variant Human Telomeric Sequence in K⁺ Solution: Insights into the Interconversion of Human Telomeric G-Quadruplex Structures. *Nucleic Acids Res.* 2009. <https://doi.org/10.1093/nar/gkp1029>.
- (4) Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S. High-Throughput Sequencing of DNA G-Quadruplex Structures in the Human Genome. *Nat. Biotechnol.* 2015. <https://doi.org/10.1038/nbt.3295>.
- (5) Hänsel-Hertsch, R.; Beraldi, D.; Lensing, S. V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; et al. G-Quadruplex Structures Mark Human Regulatory Chromatin. *Nat. Genet.* 2016. <https://doi.org/10.1038/ng.3662>.
- (6) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* 2013. <https://doi.org/10.1038/nchem.1548>.
- (7) Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-Quadruplexes in Gene Promoters: A Novel Anticancer Strategy? *Nat. Rev. Drug Discov.* 2011, 10 (4), 261–275. <https://doi.org/10.1038/nrd3428>.
- (8) Wu, Y.; Brosh, R. M. G-Quadruplex Nucleic Acids and Human Disease. *FEBS Journal.* 2010. <https://doi.org/10.1111/j.1742-4658.2010.07760.x>.
- (9) Neidle, S. Quadruplex Nucleic Acids as Targets for Anticancer Therapeutics. *Nature Reviews Chemistry.* 2017. <https://doi.org/10.1038/s41570-017-0041>.
- (10) Valton, A. L.; Prioleau, M. N. G-Quadruplexes in DNA Replication: A Problem or a Necessity? *Trends in Genetics.* 2016. <https://doi.org/10.1016/j.tig.2016.09.004>.
- (11) Mendoza, O.; Bourdoncle, A.; Boulé, J. B.; Brosh, R. M.; Mergny, J. L. G-Quadruplexes and Helicases. *Nucleic Acids Research.* 2016. <https://doi.org/10.1093/nar/gkw079>.
- (12) Rhodes, D.; Lipps, H. J. Survey and Summary G-Quadruplexes and Their Regulatory Roles in Biology. *Nucleic Acids Research.* 2015. <https://doi.org/10.1093/nar/gkv862>.

- (13) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress *c-MYC* Transcription. *Proc. Natl. Acad. Sci.* 2002. <https://doi.org/10.1073/pnas.182256799>.
- (14) Sneppen, K.; Zocchi, G. DNA and RNA. In *Physics in Molecular Biology*; 2010. <https://doi.org/10.1017/cbo9780511755699.005>.
- (15) Pelengaris, S.; Khan, M.; Evan, G. *c-MYC*: More than Just a Matter of Life and Death. *Nature Reviews Cancer.* 2002. <https://doi.org/10.1038/nrc904>.
- (16) Risitano, A.; Fox, K. R. Influence of Loop Size on the Stability of Intramolecular DNA Quadruplexes. *Nucleic Acids Res.* 2004. <https://doi.org/10.1093/nar/gkh598>.
- (17) Cevec, M.; Plavec, J. Role of Loop Residues and Cations on the Formation and Stability of Dimeric DNA G-Quadruplexes. *Biochemistry* 2005. <https://doi.org/10.1021/bi0514414>.
- (18) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. Human Telomeric Sequence Forms a Hybrid-Type Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel Strands in Potassium Solution. *Nucleic Acids Res.* 2006. <https://doi.org/10.1093/nar/gkl348>.
- (19) Agarwal, T.; Roy, S.; Chakraborty, T. K.; Maiti, S. Selective Targeting of G-Quadruplex Using Furan-Based Cyclic Homooligopeptides: Effect on *c-MYC* Expression. *Biochemistry* 2010. <https://doi.org/10.1021/bi1005927>.
- (20) Brown, R. V.; Danford, F. L.; Gokhale, V.; Hurley, L. H.; Brooks, T. A. Demonstration That Drug-Targeted down-Regulation of MYC in Non-Hodgkins Lymphoma Is Directly Mediated through the Promoter G-Quadruplex. *J. Biol. Chem.* 2011. <https://doi.org/10.1074/jbc.M111.274720>.
- (21) Boddupally, P. V. L.; Hahn, S.; Beman, C.; De, B.; Brooks, T. A.; Gokhale, V.; Hurley, L. H. Anticancer Activity and Cellular Repression of *c-MYC* by the G-Quadruplex-Stabilizing 11-Piperazinylquinoline Is Not Dependent on Direct Targeting of the G-Quadruplex in the *c-MYC* Promoter. *J. Med. Chem.* 2012. <https://doi.org/10.1021/jm300282c>.
- (22) Diveshkumar, K. V.; Sakrikar, S.; Harikrishna, S.; Dhamodharan, V.; Pradeepkumar, P. I. Targeting Promoter G-Quadruplex DNAs by Indenopyrimidine-Based Ligands. *ChemMedChem* 2014. <https://doi.org/10.1002/cmdc.201402394>.
- (23) Dhamodharan, V.; Harikrishna, S.; Bhasikuttan, A. C.; Pradeepkumar, P. I. Topology Specific Stabilization of Promoter over Telomeric G-Quadruplex Dnas by Bisbenzimidazole Carboxamide Derivatives. *ACS Chem. Biol.* 2015. <https://doi.org/10.1021/cb5008597>.
- (24) Lipinski, C. A. Lead- and Drug-like Compounds: The Rule-of-Five Revolution. *Drug Discovery Today: Technologies.* 2004. <https://doi.org/10.1016/j.ddtec.2004.11.007>.

- (25) 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. *Bioorganic Med. Chem. Lett.* 2006. <https://doi.org/10.1016/j.bmcl.2006.06.067>.
- (26) Nasiri, H. R.; Bell, N. M.; McLuckie, K. I. E.; Husby, J.; Abell, C.; Neidle, S.; Balasubramanian, S. Targeting a *c-MYC* G-Quadruplex DNA with a Fragment Library. *Chem. Commun.* 2014. <https://doi.org/10.1039/c3cc48390h>.
- (27) Carabet, L. A.; Rennie, P. S.; Cherkasov, A. Therapeutic Inhibition of *Myc* in Cancer. Structural Bases and Computer-Aided Drug Discovery Approaches. *International Journal of Molecular Sciences.* 2019. <https://doi.org/10.3390/ijms20010120>.
- (28) Del Villar-Guerra, R.; Gray, R. D.; Trent, J. O.; Chaires, J. B. A Rapid Fluorescent Indicator Displacement Assay and Principal Component/Cluster Data Analysis for Determination of Ligand-Nucleic Acid Structural Selectivity. *Nucleic Acids Res.* 2018, 46 (7), e41. <https://doi.org/10.1093/nar/gky019>.
- (29) Carvalho, J.; Nottelet, P.; Mergny, J. L.; Queiroz, J. A.; Salgado, G. F.; Cruz, C. Study of the Interaction between Indole-Based Compounds and Biologically Relevant G-Quadruplexes. *Biochimie* 2017. <https://doi.org/10.1016/j.biochi.2017.02.005>.
- (30) Caldwell, J. J.; Davies, T. G.; Donald, A.; McHardy, T.; Rowlands, M. G.; Aherne, G. W.; Hunter, L. K.; Taylor, K.; Ruddle, R.; Raynaud, F. I.; et al. Identification of 4-(4-Aminopiperidin-1-yl)-7H-Pyrrolo[2,3-d]Pyrimidines as Selective Inhibitors of Protein Kinase B through Fragment Elaboration. *J. Med. Chem.* 2008. <https://doi.org/10.1021/jm701437d>.
- (31) Gaikwad, D. D.; Chapolikar, A. D.; Devkate, C. G.; Warad, K. D.; Tayade, A. P.; Pawar, R. P.; Domb, A. J. Synthesis of Indazole Motifs and Their Medicinal Importance: An Overview. *European Journal of Medicinal Chemistry.* 2015. <https://doi.org/10.1016/j.ejmech.2014.11.029>.
- (32) Buchholz, I.; Karg, B.; Dickerhoff, J.; Sievers-Engler, A.; Lämmerhofer, M.; Weisz, K. Selective Targeting of G-Quadruplex Structures by a Benzothiazole-Based Binding Motif. *Chem. - A Eur. J.* 2017. <https://doi.org/10.1002/chem.201700298>.
- (33) Chen, S. Bin; Tan, J. H.; Ou, T. M.; Huang, S. L.; An, L. K.; Luo, H. Bin; Li, D.; Gu, L. Q.; Huang, Z. S. Pharmacophore-Based Discovery of Triaryl-Substituted Imidazole as New Telomeric G-Quadruplex Ligand. *Bioorganic Med. Chem. Lett.* 2011. <https://doi.org/10.1016/j.bmcl.2010.12.019>.
- (34) Seganish, W. M.; Fischmann, T. O.; Sherborne, B.; Matasi, J.; Lavey, B.; McElroy, W. T.; Tulshian, D.; Tata, J.; Sondey, C.; Garlisi, C. G.; et al. Discovery and Structure Enabled Synthesis of

2,6-Diaminopyrimidin-4-One IRAK4 Inhibitors. *ACS Med. Chem. Lett.* 2015. <https://doi.org/10.1021/acsmchemlett.5b00279>.

(35) Alcaro, S.; Artese, A.; Iley, J. N.; Missailidis, S.; Ortuso, F.; Parrotta, L.; Pasceri, R.; Paduano, F.; Sissi, C.; Trapasso, R.; et al. Rational Design, Synthesis, Biophysical and Antiproliferative Evaluation of Fluorenone Derivatives with DNA G-Quadruplex Binding Properties. *ChemMedChem* 2010. <https://doi.org/10.1002/cmdc.200900541>.

(36) Dash, J.; Nath Das, R.; Hegde, N.; Pantoå, G. D.; Shirude, P. S.; Balasubramanian, S. Synthesis of Bis-Indole Carboxamides as G-Quadruplex Stabilizing and Inducing Ligands. *Chem. - A Eur. J.* 2012. <https://doi.org/10.1002/chem.201102556>.

(37) 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. <https://doi.org/10.3389/fchem.2014.00054>.

(38) Doria, F.; Nadai, M.; Sattin, G.; Pasotti, L.; Richter, S. N.; Freccero, M. Water Soluble Extended Naphthalene Diimides as PH Fluorescent Sensors and G-Quadruplex Ligands. *Org. Biomol. Chem.* 2012. <https://doi.org/10.1039/c2ob07006e>.

(39) Franceschin, M.; Rossetti, L.; D'Ambrosio, A.; Schirripa, S.; Bianco, A.; Ortaggi, G.; Savino, M.; Schultes, C.; Neidle, S. Natural and Synthetic G-Quadruplex Interactive Berberine Derivatives. *Bioorganic Med. Chem. Lett.* 2006. <https://doi.org/10.1016/j.bmcl.2005.12.001>.

(40) Bessi, I.; Bazzicalupi, C.; Richter, C.; Jonker, H. R. A.; Saxena, K.; Sissi, C.; Chioccioli, M.; Bianco, S.; Bilia, A. R.; Schwalbe, H.; et al. Spectroscopic, Molecular Modeling, and NMR-Spectroscopic Investigation of the Binding Mode of the Natural Alkaloids Berberine and Sanguinarine to Human Telomeric G-Quadruplex DNA. *ACS Chem. Biol.* 2012. <https://doi.org/10.1021/cb300096g>.

(41) Li, F.; Frett, B.; Li, H. Y. Selective Reduction of Halogenated Nitroarenes with Hydrazine Hydrate in the Presence of Pd/C. *Synlett* 2014. <https://doi.org/10.1055/s-0033-1339025>.

(42) Egger, J.; Weckerle, C.; Cutting, B.; Schwardt, O.; Rabbani, S.; Lemme, K.; Ernst, B. Nanomolar E-Selectin Antagonists with Prolonged Half-Lives by a Fragment-Based Approach. *J. Am. Chem. Soc.* 2013. <https://doi.org/10.1021/ja4029582>.

(43) Auberson, Y. P.; Troxler, T.; Zhang, X.; Yang, C. R.; Feuerbach, D.; Liu, Y. C.; Lagu, B.; Perrone, M.; Lei, L.; Shen, X.; et al. From Ergolines to Indoles: Improved Inhibitors of the Human H3 Receptor for the Treatment of Narcolepsy. *ChemMedChem* 2015. <https://doi.org/10.1002/cmdc.201402418>.

- (44) Jarrahpour, A.; Zarei, M. The Vilsmeier Reagent: A Useful and Versatile Reagent for the Synthesis of 2-Azetidinones. *Tetrahedron* 2009. <https://doi.org/10.1016/j.tet.2009.02.005>.
- (45) Hanna-Elias, A.; Manallack, D. T.; Berque-Bestel, I.; Irving, H. R.; Coupar, I. M.; Iskander, M. N. Synthesis and Preliminary Screening of Novel Indole-3-Methanamines as 5-HT₄ Receptor Ligands. *Eur. J. Med. Chem.* 2009. <https://doi.org/10.1016/j.ejmech.2009.01.015>.
- (46) Kim, J. Y.; Lee, K.; Kim, Y.; Kim, C. K.; Lee, K. Production of Dyestuffs from Indole Derivatives by Naphthalene Dioxygenase and Toluene Dioxygenase. *Lett. Appl. Microbiol.* 2003. <https://doi.org/10.1046/j.1472-765X.2003.0279.x>.
<https://doi.org/10.1021/jm0302039>.
- (47) Shafakat Ali, N.; Dar, B.; Pradhan, V.; Farooqui, M. Chemistry and Biology of Indoles and Indazoles: A Mini-Review. *Mini-Reviews Med. Chem.* **2013**. <https://doi.org/10.2174/1389557511313120009>.
- (48) Zhao, H.; Huang, D.; Caflisch, A. Discovery of Tyrosine Kinase Inhibitors by Docking into an Inactive Kinase Conformation Generated by Molecular Dynamics. *ChemMedChem* **2012**. <https://doi.org/10.1002/cmdc.201200331>.
- (49) Ferrara, N. Vascular Endothelial Growth Factor as a Target for Anticancer Therapy. *Oncologist* **2004**. https://doi.org/10.1634/theoncologist.9-suppl_1-2.
- (50) Chen, G.; Liu, Z.; Zhang, Y.; Shan, X.; Jiang, L.; Zhao, Y.; He, W.; Feng, Z.; Yang, S.; Liang, G. Synthesis and Anti-Inflammatory Evaluation of Novel Benzimidazole and Imidazopyridine Derivatives. *ACS Med. Chem. Lett.* **2013**. <https://doi.org/10.1021/ml300282t>.
- (51) Frey, R. R.; Curtin, M. L.; Albert, D. H.; Glaser, K. B.; Pease, L. J.; Soni, N. B.; Bouska, J. J.; Reuter, D.; Stewart, K. D.; Marcotte, P.; et al. 7-Aminopyrazolo[1,5-a]Pyrimidines as Potent Multitargeted Receptor Tyrosine Kinase Inhibitors. *J. Med. Chem.* **2008**. <https://doi.org/10.1021/jm701397k>.
- (52) Kielhorn, J.; Rosner, G. Morpholine. *Environ. Heal. Criteria* **1996**. <https://doi.org/10.1016/b978-081551401-5.50372-5>.
- (53) Bolton, J. L.; Trush, M. A.; Penning, T. M.; Dryhurst, G.; Monks, T. J. Role of Quinones in Toxicology. *Chemical Research in Toxicology*. 2000. <https://doi.org/10.1021/tx9902082>.
- (54) Kumagai, Y.; Shinkai, Y.; Miura, T.; Cho, A. K. The Chemical Biology of Naphthoquinones and Its Environmental Implications. *Annu. Rev. Pharmacol. Toxicol.* **2012**. <https://doi.org/10.1146/annurev-pharmtox-010611-134517>.
- (55) Bolzán, A. D.; Bianchi, M. S. Genotoxicity of Streptonigrin: A Review. *Mutation Research - Reviews in Mutation Research*. 2001. [https://doi.org/10.1016/S1383-5742\(00\)00062-4](https://doi.org/10.1016/S1383-5742(00)00062-4).

- (56) Bradner, W. T. Mitomycin C: A Clinical Update. *Cancer Treat. Rev.* **2001**. <https://doi.org/10.1053/ctrv.2000.0202>.
- (57) Minotti, G.; Menna, P.; Salvatorelli, E.; Cairo, G.; Gianni, L. Anthracyclines: Molecular Advances and Pharmacologic Developments in Antitumor Activity and Cardiotoxicity. *Pharmacological Reviews*. 2004. <https://doi.org/10.1124/pr.56.2.6>.
- (58) Verma, R. Anti-Cancer Activities of 1,4-Naphthoquinones: A QSAR Study. *Anticancer Agents Med. Chem.* **2008**. <https://doi.org/10.2174/187152006778226512>.
- (59) Read, M. A.; Wood, A. A.; Harrison, J. R.; Gowan, S. M.; Kelland, L. R.; Dosanjh, H. S.; Neidle, S. Molecular Modeling Studies on G-Quadruplex Complexes of Telomerase Inhibitors: Structure-Activity Relationships. *J. Med. Chem.* **1999**. <https://doi.org/10.1021/jm990287e>.
- (60) Gewirtz, D. A. A Critical Evaluation of the Mechanisms of Action Proposed for the Antitumor Effects of the Anthracycline Antibiotics Adriamycin and Daunorubicin. *Biochem. Pharmacol.* **1999**. [https://doi.org/10.1016/S0006-2952\(98\)00307-4](https://doi.org/10.1016/S0006-2952(98)00307-4).
- (61) Carvalho, J.; Nottelet, P.; Mergny, J. L.; Queiroz, J. A.; Salgado, G. F.; Cruz, C. Study of the Interaction between Indole-Based Compounds and Biologically Relevant G-Quadruplexes. *Biochimie* **2017**. <https://doi.org/10.1016/j.biochi.2017.02.005>.
- (62) Caldwell, J. J.; Davies, T. G.; Donald, A.; McHardy, T.; Rowlands, M. G.; Aherne, G. W.; Hunter, L. K.; Taylor, K.; Ruddle, R.; Raynaud, F. I.; et al. Identification of 4-(4-Aminopiperidin-1-yl)-7H-Pyrrolo[2,3-d]Pyrimidines as Selective Inhibitors of Protein Kinase B through Fragment Elaboration. *J. Med. Chem.* **2008**. <https://doi.org/10.1021/jm701437d>.
- (63) Gaikwad, D. D.; Chapolikar, A. D.; Devkate, C. G.; Warad, K. D.; Tayade, A. P.; Pawar, R. P.; Domb, A. J. Synthesis of Indazole Motifs and Their Medicinal Importance: An Overview. *European Journal of Medicinal Chemistry*. 2015. <https://doi.org/10.1016/j.ejmech.2014.11.029>.
- (64) Buchholz, I.; Karg, B.; Dickerhoff, J.; Sievers-Engler, A.; Lämmerhofer, M.; Weisz, K. Selective Targeting of G-Quadruplex Structures by a Benzothiazole-Based Binding Motif. *Chem. - A Eur. J.* **2017**. <https://doi.org/10.1002/chem.201700298>.
- (65) Chen, S. Bin; Tan, J. H.; Ou, T. M.; Huang, S. L.; An, L. K.; Luo, H. Bin; Li, D.; Gu, L. Q.; Huang, Z. S. Pharmacophore-Based Discovery of Triaryl-Substituted Imidazole as New Telomeric G-Quadruplex Ligand. *Bioorganic Med. Chem. Lett.* **2011**. <https://doi.org/10.1016/j.bmcl.2010.12.019>.
- (66) Seganish, W. M.; Fischmann, T. O.; Sherborne, B.; Matasi, J.; Lavey, B.; McElroy, W. T.; Tulshian, D.; Tata, J.; Sondey, C.; Garlisi, C. G.; et al. Discovery and Structure Enabled Synthesis of 2,6-Diaminopyrimidin-4-One IRAK4 Inhibitors. *ACS Med. Chem. Lett.* **2015**.

<https://doi.org/10.1021/acsmchemlett.5b00279>.

(67) Alcaro, S.; Artese, A.; Iley, J. N.; Missailidis, S.; Ortuso, F.; Parrotta, L.; Pasceri, R.; Paduano, F.; Sissi, C.; Trapasso, R.; et al. Rational Design, Synthesis, Biophysical and Antiproliferative Evaluation of Fluorenone Derivatives with DNA G-Quadruplex Binding Properties. *ChemMedChem* **2010**. <https://doi.org/10.1002/cmdc.200900541>.

(68) Dash, J.; Nath Das, R.; Hegde, N.; Pantoå, G. D.; Shirude, P. S.; Balasubramanian, S. Synthesis of Bis-Indole Carboxamides as G-Quadruplex Stabilizing and Inducing Ligands. *Chem. - A Eur. J.* **2012**. <https://doi.org/10.1002/chem.201102556>.

(69) 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**. <https://doi.org/10.3389/fchem.2014.00054>.

(70) Doria, F.; Nadai, M.; Sattin, G.; Pasotti, L.; Richter, S. N.; Freccero, M. Water Soluble Extended Naphthalene Diimides as PH Fluorescent Sensors and G-Quadruplex Ligands. *Org. Biomol. Chem.* **2012**. <https://doi.org/10.1039/c2ob07006e>.

(71) Franceschin, M.; Rossetti, L.; D'Ambrosio, A.; Schirripa, S.; Bianco, A.; Ortaggi, G.; Savino, M.; Schultes, C.; Neidle, S. Natural and Synthetic G-Quadruplex Interactive Berberine Derivatives. *Bioorganic Med. Chem. Lett.* **2006**. <https://doi.org/10.1016/j.bmcl.2005.12.001>.

(72) Bessi, I.; Bazzicalupi, C.; Richter, C.; Jonker, H. R. A.; Saxena, K.; Sissi, C.; Chioccioli, M.; Bianco, S.; Bilia, A. R.; Schwalbe, H.; et al. Spectroscopic, Molecular Modeling, and NMR-Spectroscopic Investigation of the Binding Mode of the Natural Alkaloids Berberine and Sanguinarine to Human Telomeric G-Quadruplex DNA. *ACS Chem. Biol.* **2012**. <https://doi.org/10.1021/cb300096g>.

(73) Daina, A.; Michielin, O.; Zoete, V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules. *Sci. Rep.* **2017**. <https://doi.org/10.1038/srep42717>.

(74) Lipinski, C. A. Lead- and Drug-like Compounds: The Rule-of-Five Revolution. *Drug Discovery Today: Technologies.* 2004. <https://doi.org/10.1016/j.ddtec.2004.11.007>.

(75) Tran, P. L. T.; Largy, E.; Hamon, F.; Teulade-Fichou, M. P.; Mergny, J. L. Fluorescence Intercalator Displacement Assay for Screening G4 Ligands towards a Variety of G-Quadruplex Structures. *Biochimie* **2011**. <https://doi.org/10.1016/j.biochi.2011.05.011>.

(76) Diveshkumar, K. V.; Sakrikar, S.; Harikrishna, S.; Dhamodharan, V.; Pradeepkumar, P. I. Targeting Promoter G-Quadruplex DNAs by Indenopyrimidine-Based Ligands. *ChemMedChem*

2014. <https://doi.org/10.1002/cmdc.201402394>.

(77) Dhamodharan, V.; Harikrishna, S.; Bhasikuttan, A. C.; Pradeepkumar, P. I. Topology Specific Stabilization of Promoter over Telomeric G-Quadruplex Dnas by Bisbenzimidazole Carboxamide Derivatives. *ACS Chem. Biol.* **2015.** <https://doi.org/10.1021/cb5008597>.

(78) Li, F.; Frett, B.; Li, H. Y. Selective Reduction of Halogenated Nitroarenes with Hydrazine Hydrate in the Presence of Pd/C. *Synlett* **2014.** <https://doi.org/10.1055/s-0033-1339025>.

(79) Egger, J.; Weckerle, C.; Cutting, B.; Schwardt, O.; Rabbani, S.; Lemme, K.; Ernst, B. Nanomolar E-Selectin Antagonists with Prolonged Half-Lives by a Fragment-Based Approach. *J. Am. Chem. Soc.* **2013.** <https://doi.org/10.1021/ja4029582>.

(80) Auberson, Y. P.; Troxler, T.; Zhang, X.; Yang, C. R.; Feuerbach, D.; Liu, Y. C.; Lagu, B.; Perrone, M.; Lei, L.; Shen, X.; et al. From Ergolines to Indoles: Improved Inhibitors of the Human H3 Receptor for the Treatment of Narcolepsy. *ChemMedChem* **2015.** <https://doi.org/10.1002/cmdc.201402418>.

(81) Jarrahpour, A.; Zarei, M. The Vilsmeier Reagent: A Useful and Versatile Reagent for the Synthesis of 2-Azetidinones. *Tetrahedron* **2009.** <https://doi.org/10.1016/j.tet.2009.02.005>.

(82) Hanna-Elias, A.; Manallack, D. T.; Berque-Bestel, I.; Irving, H. R.; Coupar, I. M.; Iskander, M. N. Synthesis and Preliminary Screening of Novel Indole-3-Methanamines as 5-HT₄ Receptor Ligands. *Eur. J. Med. Chem.* **2009.** <https://doi.org/10.1016/j.ejmech.2009.01.015>.

(83) Kim, J. Y.; Lee, K.; Kim, Y.; Kim, C. K.; Lee, K. Production of Dyestuffs from Indole Derivatives by Naphthalene Dioxygenase and Toluene Dioxygenase. *Lett. Appl. Microbiol.* **2003.** <https://doi.org/10.1046/j.1472-765X.2003.01279.x>.

(84) Gigant, N.; Claveau, E.; Bouyssou, P.; Gillaizeau, I. Diversity-Oriented Synthesis of Polycyclic Diazinic Scaffolds. *Org. Lett.* **2012.** <https://doi.org/10.1021/ol203364b>.

(85) Lo, Y. C.; Rensi, S. E.; Torng, W.; Altman, R. B. Machine Learning in Chemoinformatics and Drug Discovery. *Drug Discovery Today.* 2018. <https://doi.org/10.1016/j.drudis.2018.05.010>.

(86) Dai, J.; Carver, M.; Hurley, L. H.; Yang, D. Solution Structure of a 2:1 Quindoline-c-MYC G-Quadruplex: Insights into G-Quadruplex-Interactive Small Molecule Drug Design. *J. Am. Chem. Soc.* **2011.** <https://doi.org/10.1021/ja205646q>.

(87) Forli, W.; Halliday, S.; Belew, R.; Olson, A. AutoDock Version 4.2. *Citeseer* **2012.**

(88) Nasiri, H. R.; Bell, N. M.; McLuckie, K. I. E.; Husby, J.; Abell, C.; Neidle, S.; Balasubramanian, S. Targeting a C-MYC G-Quadruplex DNA with a Fragment Library. *Chem. Commun.* **2014.** <https://doi.org/10.1039/c3cc48390h>.

(89) 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. *Bioorganic Med. Chem. Lett.* **2006**.
<https://doi.org/10.1016/j.bmcl.2006.06.067>.

(90) Gray, R. D.; Trent, J. O.; Chaires, J. B. Folding and Unfolding Pathways of the Human Telomeric G-Quadruplex. *J. Mol. Biol.* **2014**. <https://doi.org/10.1016/j.jmb.2014.01.009>.

(91) 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**. <https://doi.org/10.1021/bi048242p>.

(92) Wirmer-Bartoschek, J.; Bendel, L. E.; Jonker, H. R. A.; Grün, J. T.; Papi, F.; Bazzicalupi, C.; Messori, L.; Gratteri, P.; Schwalbe, H. Solution NMR Structure of a Ligand/Hybrid-2-G-Quadruplex Complex Reveals Rearrangements That Affect Ligand Binding. *Angew. Chemie - Int. Ed.* **2017**. <https://doi.org/10.1002/anie.201702135>.

(93) Das, T.; Panda, D.; Saha, P.; Dash, J. Small Molecule Driven Stabilization of Promoter G-Quadruplexes and Transcriptional Regulation of c-MYC. *Bioconjug. Chem.* **2018**. <https://doi.org/10.1021/acs.bioconjchem.8b00338>.

(94) O'Brien, J.; Wilson, I.; Orton, T.; Pognan, F. Investigation of the Alamar Blue (Resazurin) Fluorescent Dye for the Assessment of Mammalian Cell Cytotoxicity. *Eur. J. Biochem.* **2000**. <https://doi.org/10.1046/j.1432-1327.2000.01606.x>.

Chapter 3

**Novel triazole linked indole derivatives as *c-MYC*
G-quadruplex stabilizers *via* Click chemistry**

Chapter 3 Novel triazole linked indole derivatives as c-MYC G-quadruplex stabilizers via click chemistry

3.1 Introduction

It is well established that nucleic acid sequences rich in guanosine (G) residues can form secondary structures known as G quadruplexes (G4s),¹ which are assimilated in nucleic acid sequences, has ignited a new role for DNA in biology. The interest in G4 DNA structures has increased enormously in recent years, due to evidence that guanine rich (G-rich) sequence are located in biologically functional regions of the human genome such as oncogene promoters and telomeres.²⁻⁴ The G4s are reported to be found in the promoter region of oncogenes such as *c-KIT*,⁵ *VEGF*,⁶ *BCL-2*,⁷ *KRAS*,⁸ and *c-MYC*, whose overexpression is related to a number of human cancer like breast, colon, osteosarcoma, myeloid leukemias.⁹ G4 motifs are over-represented also in the terminal region of chromosomes (telomeres), where they may be involved in maintaining chromosome stability.¹⁰

In particular, up to 80% of all solid tumors (including gastrointestinal, ovarian and breast tumors) overexpress c-MYC.¹¹ It has been suggested that ligands that bind selectively with promoter G-quadruplexes could in principle down-regulate the expression of these genes and thus have an anticancer effect.¹²⁻¹⁴ Over the past few years, a large number of peptidomimetics ligands have been synthesized and tested for their potential to interact with *c-MYC* promoter G4 structures.¹⁵ Some of these G4 ligands have been reported for their ability to interfere with the binding of c-MYC-associated proteins.¹⁶ The design of G4 binding small molecule scaffolds has been predominately based on the fundamental principles of having a heteroaromatic core with a planar conformation to enable bindings in G4 grooves and to loops.

A number of triazole linked compounds have been reported as potent and selective G4-small molecules have been reported that can stabilize the *c-MYC* G-quadruplex such as trisubstituted acridines,¹⁷ 11-triazole substituted benzofuro[3,2-b]quinolone,¹⁸ Phenanthroline-2,9-bistriazoles¹⁹ and phenanthroline-bis-oxazole,²⁰ bis-triazole,²¹ and 1,3-biphenylene-bistriazole²² has shown the significant G4 stabilization potential (Fig. 59). Nevertheless, few of these ligands have been demonstrated to exhibit specificity toward the *c-MYC* G quadruplex over other G4 sequences. However, the diversity of G-quadruplex structures could be used to enhance the selectivity of G-quadruplex ligands they tend to non-selectively bind to various types of nucleic acid structures. Therefore, we were greatly interested in seeking a ligand that can selectively target the *c-MYC* G-quadruplex.

It is hypothesized that higher selectivity for G4 over duplex DNA can be achieved by extending the heteropolyaromatic scaffold to increase π - π overlap with the larger surface area of a terminal G-quartet.²³ This is mainly due to the fact that selectivity for the larger surface area of a terminal G-quartet increases with ligand size and the number of side-chain substituents. In this context, alkynes are one of the commonly used click chemistry transformations, has gained considerable interest in designing G-quadruplex binding ligands.^{24,25} This reaction is operationally simple, regioselective, modular and bio-orthogonal, providing high yields of the products.²⁶ Cu(I)-catalyzed Huisgen 1,3-dipolar cycloaddition of azides and The heteroaromatic 1,2,3-triazole motif is an appealing peptide bond isostere because of their similar structural and electronic characteristics to those of a peptide bond.²⁷ Endowed with these unique features, this peptidomimetics Cu(I) catalyzed azide-alkyne cycloaddition approach have been viewed as ideal reactions for chemical targeting of G-quadruplexes.²⁵

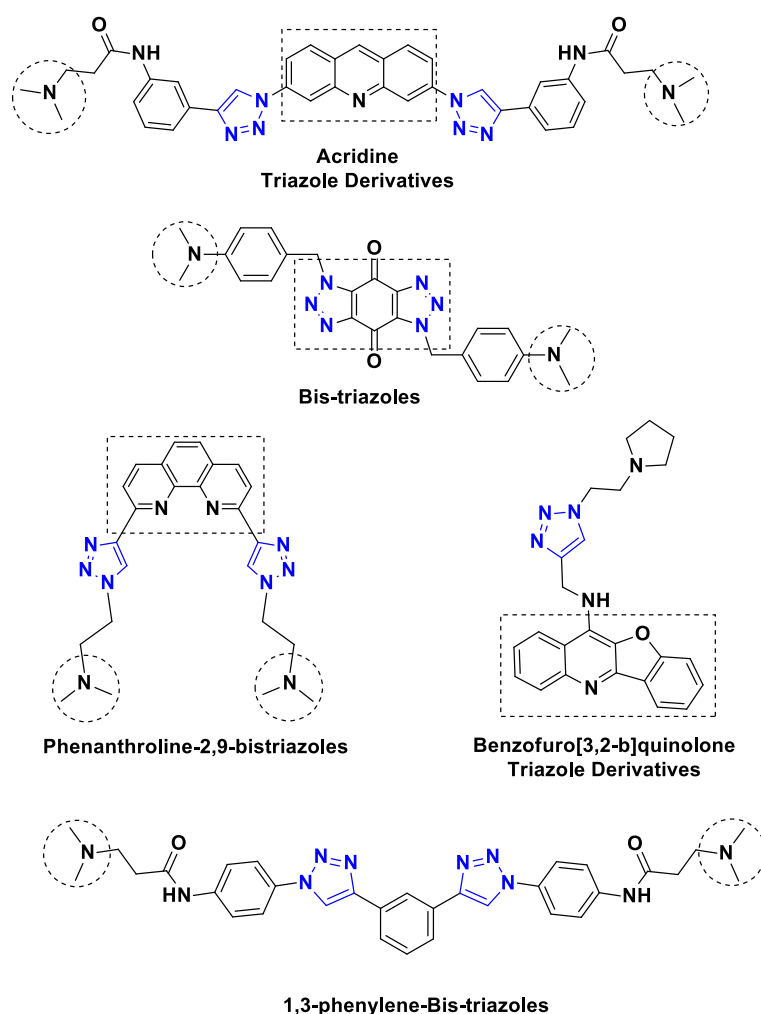


Figure 59. Triazole based G-quadruplex binding and stabilizing ligands reported in the literature

3.2 Bioisosterism

Isosteres were first defined in 1919 by Langmuir, as atoms or groups possessing same number of valence electrons.²⁸ An extension of this concept was provided by Grimms and Erlenmeyer who redefined isosters as atoms, ions, and molecules in which the peripheral layers of electrons are same.²⁹ Pioneering work by Friedman and Thornber in early 1950s led to further evolution of the concept of bioisosters and their consideration as atoms or groups which exhibit similar physicochemical properties and biological activity.^{30,31} Classical examples of bioisosters in nature are amino acids such as serine and cysteine, tyrosine and histidine, pyrimidine and purine base (Fig.60). Bioisosters are classified into two main categories, classical and non-classical bioisosters. Classical bioisosters are those atoms or groups which have similar valence electrons or ring equivalents (Table 12). Non-classical bioisosters are those which do not have any similarity in their valence electrons or ring equivalents (Table 13).^{32,33}

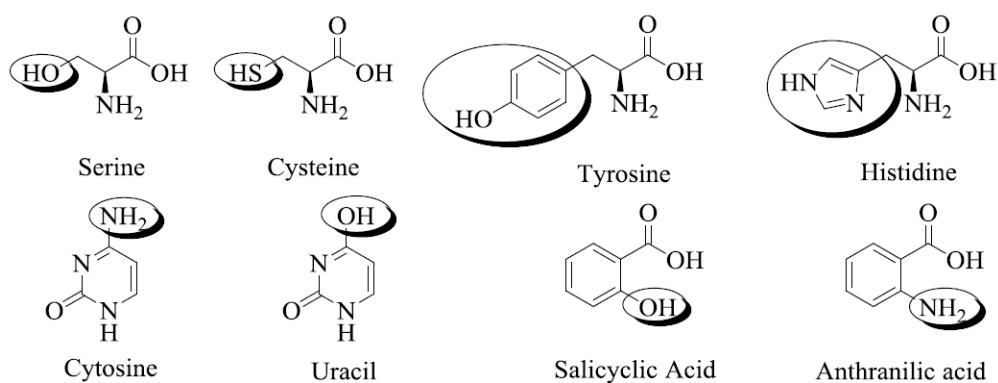


Figure 60. Bioisosters in nature.

Table 12. Classical Bioisosters groups and atoms

Monovalent	Divalent	Trivalent	Tetravalent
-OH, -NH ₂ , -CH ₃ , -OR	-CH ₂	=CH-	=C=
-F, -Cl, -Br, -I, -SH, -PH ₂	-O-	=N-	=Si=
-Si ₃ , -SR	-S-	=P-	=N ⁺ =
	-Se-	=As-	=P ⁺ =
	-Te-	-Sb-	=As ⁺ =
			=Sb ⁺ =

Table 13. Non-classical Bioisosters groups and atoms

-CO-	-COOH	-SO ₂ NH ₂	-H	-CONH	-CONH ₂
-CO ₂ -	-SO ₃ H	-PO(OH)NH ₂	-F	-ROCO	-CSNH ₂
-SO ₂ -	-Tetrazole	=N-	-OH	-Benzimidaole	
-SO ₂ NR	-SO ₂ NH ₂	=P-	-CH ₂ OH		
-CON-	=N-	=As-	-NHCONH ₂		
	-CF ₃	-Sb-	-NH-CS-NH ₂		
	-CN				

Bioisosterism is the most common approach applied in pharmaceutical industry for lead evolution. Applicability of these principle to develop therapeutically active new substances against distinct classes of targets, development of me-too types of drugs and also as a tool for molecular modification have made bioisosterism quite popular among medicinal chemists in industry.³³ Many successful stories of lead optimization using this approach have further enhanced its stature in drug design.³⁴ The principles of bioisosterism continue to be employed for two main reasons.³⁵

- Enhancement in potency via isoform selectivity.
- Favourable pharmacokinetic parameters.

Several similar success stories of lead optimization *via* bioisosterism point to the relevance of the approach to modern drug development. Many of the leads that have been developed as such have emerged as blockbuster drugs. Appropriate use of bioisosterism can be assisted by theoretical predictability or modeling of parameters and adequate knowledge of pharmacodynamics.

3.2.1 Importance of click chemistry in medicinal chemistry

A click reaction satisfies many criteria: it can be applicable modularly and widely in scope, obtain high chemical yields, produce minimal byproducts that can be removed by chromatographic methods, and be stereospecific (but not necessarily enantioselective) when applicable. In addition, it has simple reaction conditions, involves readily available starting materials, reagents, and a benign solvent (preferably water), and allows simple isolation of products by crystallization or distillation but not preparative chromatography. In practice, click reactions tend to have large negative free energies and hence involve carbon-heteroatom bond forming processes. Thus, unlike many conventional synthetic reactions, the power of click chemistry lies in its simplicity and ease of use.

3.2.2 The physicochemical properties of 1,2,3-triazoles

1,2,3-Triazoles are π -electron-deficient and exhibit both basic and acidic properties. Due to their poor basicity, the 1,2,3- triazole ring is not protonated at physiological pH. In addition, they possess a strong dipole moment (5.2 -5.6 Debye),³⁶ an aromatic character and a good hydrogen-bond-accepting ability. Furthermore, they are very stable to both metabolic³⁷ and chemical degradations, being rather inert to severe hydrolytic, oxidizing and reducing conditions, even at high temperatures.

3.2.3 The application of 1,2,3-triazoles in drug discovery

The synthetic 1,2,3-triazoles show diverse biological activities, such as anticancer,³⁸ antifungal and antibacterial,³⁹⁻⁴¹ antituberculosis,^{42,43} and antiviral.^{44,45} Thus, click chemistry has attracted extensive interest in almost all aspects of drug discovery, such as lead discovery through combinatorial chemistry, target-templated in vitro chemistry, and proteomics and DNA research by using bioconjugation reactions.

3.2.4 1,2,3-Triazole ring as the non-classic bioisostere of amide

Among a variety of bioisosteres of the amide moiety 1,2,3- triazole has attracted increasing attention in drug discovery. The similitude of the two moieties can be seen in the size (distances between substituents are 3.8 -3.9 Å in amides and 5.0 - 5.1 Å in 1, 2, 3-triazoles, the dipolar moment (amide \sim 4 Debye, 1,2,3-triazole \sim 5 Debye), and the H-bond acceptor capacity (Figure 1).⁴⁶ The 1,2,3-triazole rings, with sp^2 -hybridized nitrogen atoms N(2) and N(3), can function as weak hydrogen-bond acceptor due to the lone pair electrons. Moreover, the strong dipole moment of 1,2,3-triazole ring polarizes the C(5) proton to such a degree that it can function as a hydrogen-bond donor similar to the amide NH (Fig. 61). Furthermore, the 1,2,3-triazole ring has a large dipole that could align with that of the other amides in a given peptide secondary structure.⁴⁷ However, unlike amides 1,2,3-triazoles are extremely stable to hydrolysis. Because of the ability of 1,2,3-triazoles to function as rigid linking units that mimic the atom replacement and electronic properties of a peptide bond without the susceptibility of hydrolysis, many known 1,2,3-triazoles possess biological activity. A number of promising agents with the 1,2,3-triazole moiety instead of the amide bond retain the biological activity.

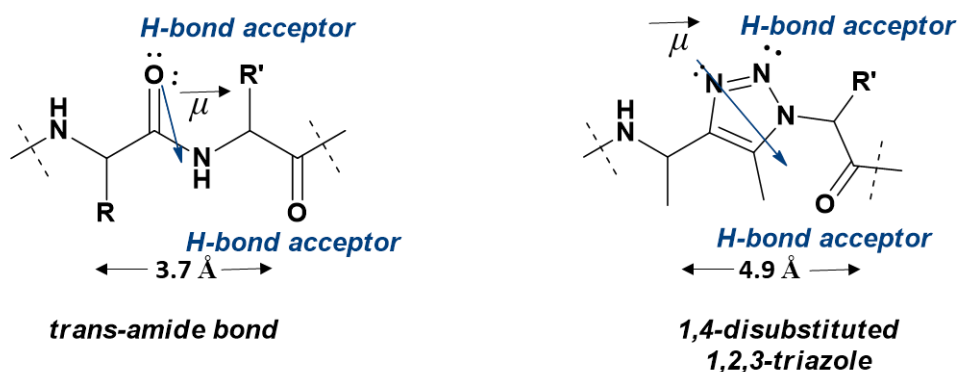


Figure 61. Hypothesis on how 1,2,3-triazole acts as non-classic bioisosteres of amides (Reproduced from reference ⁴⁰)

3.3 Project Objective

Main objective of this chapter is to generate feasible bioconjugation strategies for fragment expansion to optimize the identified hit from our primary screening of fragments with FID screenings. From our initial fluorescence displacement (FID) screening of 130 chemically diverse fragments, we found that fragment 5-aminogranine (labeled as A8/9a in table 4 and fig. 50a of chapter 1) could stabilize the *c-MYC* promoter G4, results into down-regulation of the *c-MYC* transcription and expression (see Fig. 56B, compound in chapter 2) at higher concentration. Our further study showed that identified fragments have no significant binding effect on the *c-MYC* G4. However, the G4 stabilizing effect and binding affinity of 12 & 15 are not good enough for further in depth studies. Therefore, further hit optimization from these identified fragments were performed by click chemistry peptidomimetics approach.

3.4 Results and Discussion

3.4.1 Design and Synthesis

To generate the library of triazole linked indole conjugate simple synthetic strategies were reported in this chapter. In scheme 1, scaffold 1-methyl-1H-indol-5-amine (12) is obtained in 60 % yield with Pd/C, $\text{NH}_2\text{NH}_2 \cdot \text{H}_2\text{O}$ hydrogenation of 1-methyl-5-nitro-1H-indole (11), which is a nucleophilic substitution product of 5-nitro-1H-indole (1).⁴⁸ The intermediate 5-nitro-1H-indole-3-carbaldehyde (13) is obtained in 60 % yield *via* Vilsmeier-Haack reaction⁴⁹ with 5-nitro-1H-indole (1) which is a key step to generate conjugates 14 in quantitative yields with one pot *in-situ* reaction of aldehydes with substituted amines in presence of NaBH_4 as a reducing agent.⁵⁰

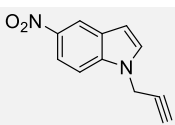
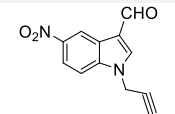
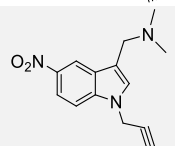
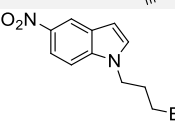
(3) with Vilsmeier-Haack reaction.⁴⁹ Conjugate 4 is obtained from 5-nitro-1-(prop-2-yn-1-yl)-1H-indole-3-carbaldehyde (3) by one pot *in-situ* reaction of substituted amine with aldehyde in presence of NaBH₄ in methanol media with 56 % yield. In scheme 2, precursors 2, 4 and 8, 8a and amine with aldehyde in presence of NaBH₄ in methanol media with 56 % yield. In scheme 2, precursors 2, 4 and 8, 8a and 8b are allowed undergo copper (I) catalyzed 1, 3-dipolar cycloaddition click reaction to afford 9, 9a, 9b and 9c respectively. This strategy is useful for Pd/C, NH₂NH₂.H₂O-catalyzed hydrogenation⁴⁸ of 9, 9a and 9b to 10, 10a and 10b respectively.

Conjugate 4 is obtained from 5-nitro-1-(prop-2-yn-1-yl)-1H-indole-3-carbaldehyde (3) by one pot *in-situ* reaction of substituted amine with aldehyde in presence of NaBH₄ in methanol media with 56 % yield. In scheme 2, precursors 2, 4 and 8, 8a and amine with aldehyde in presence of NaBH₄ in methanol media with 56 % yield. In scheme 2, precursors 2, 4 and 8, 8a and 8b are allowed undergo copper (I) catalyzed 1, 3-dipolar cycloaddition click reaction to afford 9, 9a, 9b and 9c respectively. This strategy is useful for Pd/C, NH₂NH₂.H₂O-catalyzed hydrogenation⁴⁸ of 9, 9a and 9b to 10, 10a and 10b respectively.

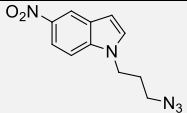
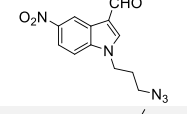
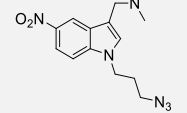
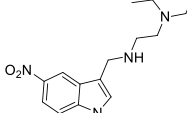
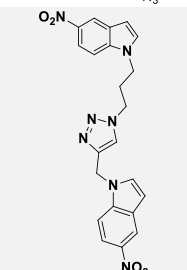
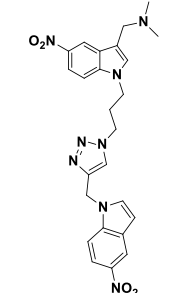
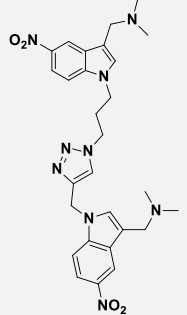
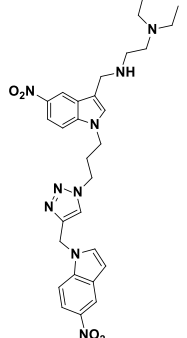
3.5 Initial screening with Fluorescent indicator displacement (FID) assays

The ability of all seven triazole derivatives compounds (9, 9a, 9b, 9c, 10, 10a and 10b) and all the intermediates to FID assays were performed to investigate the binding affinity of selected fragments for *c-MYC*, *c-KIT1*, *BCL-2*, h-TELO, G4 and duplex ds24-DNA and ds27-DNA DNA sequences (Table 2 in chapter 2). In this assay, we monitored the thiazole displacement ability by increasing concentrations of the ligands.

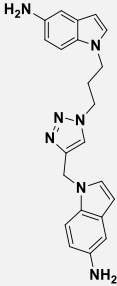
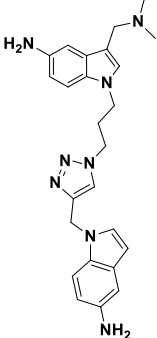
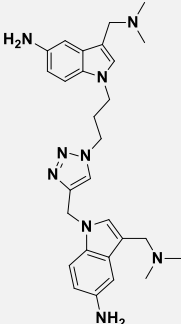
Table 14. List of synthesized fragments screened based on Thiazole Displacement Assays

Hits	Structures	SMILE NOTATION	FID AFFINITY
2		<chem>O=N(=O)C1=CC2=C(C=C1)N(CC#C)C=C2</chem>	++
3		<chem>[H]C(=O)C1=CN(CC#C)C2=C1C=C(C=C2)N(=O)=O</chem>	++
4		<chem>CN(C)CC1=CNC2=C1C=C(N)C=C2</chem>	++
5		<chem>BrCCCN1C=CC2=C1C=CC(=C2)N(=O)=O</chem>	++

Novel triazole linked indole derivatives as c-MYC G-quadruplex stabilizers via click chemistry

6		<chem>N=[N]=NCCCN1C=CC2=C1C=CC(=C2)N(=O)=O</chem>	++
7		<chem>[H]C(=O)C1=CN(CCCN=[N]#N)C2=C1C=C(C=C2)N(=O)=O</chem>	++
8		<chem>CN(C)CC1=CN(CCCN=[N]#N)C2=C1C=C(C=C2)N(=O)=O</chem>	++
8a		<chem>CCN(CC)CCNCC1=CN(CCCN=[N]#N)C2=C1C=C(C=C2)N(=O)=O</chem>	++
9		<chem>[O-][N+](=O)C1=CC2=C(C=C1)N(CC1=CN(CCCN3C=CC4=CC(=CC=C34)[N+][O-])=O)N=N1)C=C2</chem>	++++
9a		<chem>CN(C)CC1=CN(CCCN2C=C(CN3C=CC4=C3C=CC(=C4)[N+][O-])=O)N=N2)C2=CC=C(C=C12)[N+][O-]=O</chem>	+++++
9b		<chem>CN(C)CC1=CN(CCCN2C=C(CN3C=C(CN(C)C)C4=C3C=CC(=C4)[N+][O-])=O)N=N2)C2=CC=C(C=C12)[N+][O-]=O</chem>	+++++
9c		<chem>CCN(CC)CCNCC1=CN(CCCN2C=C(CN3C=CC4=C3C=CC(=C4)[N+][O-])=O)N=N2)C2=CC=C(C=C12)[N+][O-]=O</chem>	+

Novel triazole linked indole derivatives as c-MYC G-quadruplex stabilizers via click chemistry

10		<chem>NC1=CC2=C(C=C1)N(CC1=CN(CCCN3C=CC4=CC(N)=CC=C34)N=N1)C=C2</chem>	++++
10a		<chem>CN(C)CC1=CN(CCCN2C=C(CN3C=CC4=C3C=CC(N)=C4)N=N2)C2=CC=C(N)C=C12</chem>	++++
10b		<chem>CN(C)CC1=CN(CCCN2C=C(CN3C=C(CN(C)C)C4=C3C=CC(N)=C4)N=N2)C2=CC=C(N)C=C12</chem>	++++

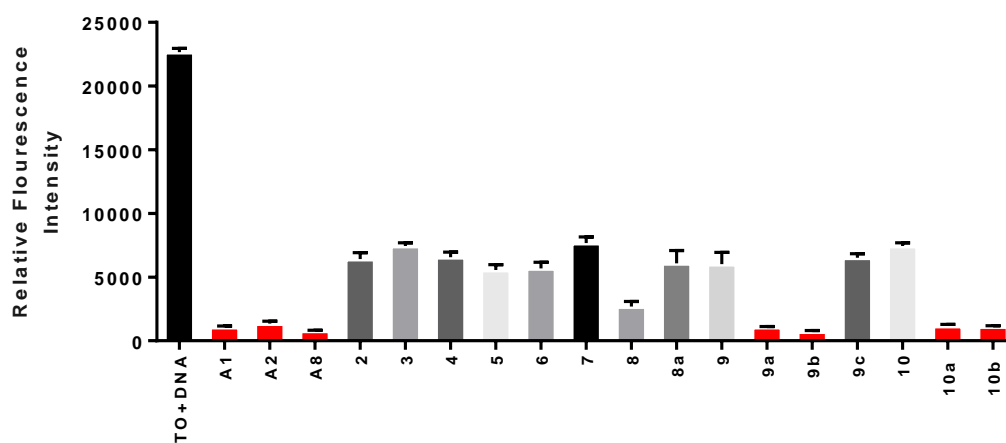


Figure 62. Thiazole displacement potential of the commercial ligand 12, A2 and 15 and synthesized G4-ligands. Best ligands are highlighted in red color. Experimental condition: 0.25 μ M DNA, 0.5 μ M Thiazole Orange, 10 % DMSO, 20 mM Na cacoc, 140 mM KCl, pH 7 (25 μ L/well)

Additionally, ligands were also screened towards a double-stranded DNA in order to assess the G4 vs ds-DNA selectivity. In our findings, as anticipated triazole conjugates 9a, 9b, 10a and 10b generated from parent fragment⁵¹ showed a very strong affinity towards *c-MYC* G4 DNA with K_i values ranging from 2.0-4.0 μM . Most compounds showed the preferential affinity towards human parallel promoter *c-MYC* G4 DNA sequences in comparison to other promoter, telomeric G-quadruplex and duplex oligonucleotide sequences (see Table 15, figure 63 & 64).

Table 15. K_i values for the most active ligands for G-quadruplex and Duplex DNA sequences

K_i (μM)	9a	9b	10a	10b	15	12
c-MYC	2.11	2.42	8.59	8.87	3.32	3.22
c-KIT-1	20.21	27.83	28.85	28.93	8.89	7.54
BCL2	39.37	35.75	39.17	29.43	25.75	32.43
h-Telo- 22	102.39	106.13	114.13	104.74	104.13	111.29
ds-DNA-24	1064.13	1015.23	532.86	668.12	464.13	504.53
ds-DNA-27	1661.13	1675.12	1688.3	1578.26	2049.66	2182.15

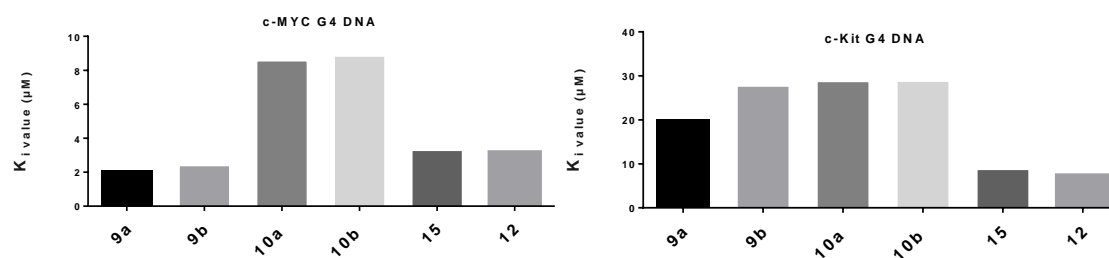


Figure 63. K_i value plots for the most active ligands for promoter G-Quadruplex sequences *c-MYC*, *c-KIT-1* based on Thiazole Displacement Assays ; Titration scheme of *c-MYC*: $c(\text{TO}) = 0,5 \mu\text{M}$; $c(\text{cMYC}) = 0,5 \mu\text{M}$; $c(\text{DMSO}) = 10\%$; $c(\text{Ligand}) = 256 \mu\text{M}, 128 \mu\text{M}, 64 \mu\text{M}, 32 \mu\text{M}, 16 \mu\text{M}, 8 \mu\text{M}, 4 \mu\text{M}, 2 \mu\text{M}, 1 \mu\text{M}, 0,5 \mu\text{M}, 0,25 \mu\text{M}, 0,125 \mu\text{M}, 0,0625 \mu\text{M}$; r.t; Titration of *c-KIT1*: $c(\text{TO}) = 0,5 \mu\text{M}$; $c(\text{c-KIT-1}) = 0,25 \mu\text{M}$; $c(\text{DMSO}) = 10\%$; $c(\text{Ligand}) = 1024 \mu\text{M}, 512 \mu\text{M}, 256 \mu\text{M}, 128 \mu\text{M}, 64 \mu\text{M}, 32 \mu\text{M}, 16 \mu\text{M}, 8 \mu\text{M}, 4 \mu\text{M}, 2 \mu\text{M}, 1 \mu\text{M}, 0,5 \mu\text{M}, 0,25 \mu\text{M}, 0,125 \mu\text{M}$; r.t

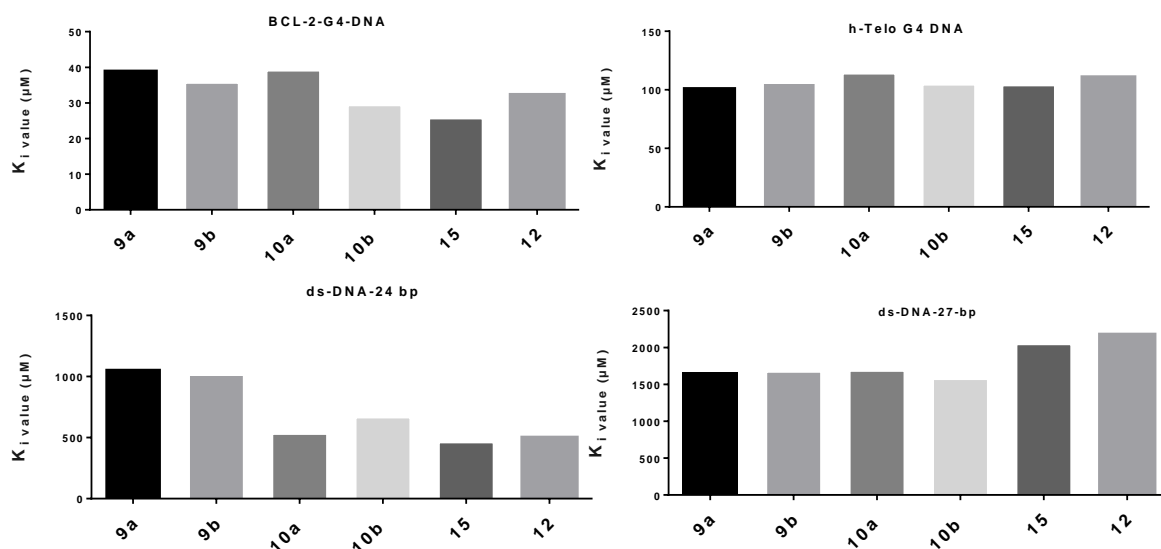


Figure 64. Ki value plots for the most active ligands for promoter G-Quadruplex sequences *BCL-2*, h-TELO, ds-DNA-24 and ds-DNA-27 based on Thiazole Displacement Assays; Titration scheme of *BCL-2*: c(TO) = 0,5 µM; c(*BCL-2*) = 0,25 µM; c(DMSO) = 10%; c(Ligand) = 1024 µM, 512 µM, 256 µM, 128 µM, 64 µM, 32 µM, 16 µM, 8 µM, 4 µM, 2 µM, 1 µM, 0,5 µM, 0,25 µM, 0,125 µM; r.t; Titration of h-Telo-DNA: c(TO) = 0,5 µM; h-Telo = 0,5 µM; c(DMSO) = 10%; c(Ligand) = 5012 µM, 2056 µM, 1028 µM, 604 µM, 302 µM, 106 µM, 80 µM, 40 µM, 20 µM, 10 µM, 5 µM, 1 µM, 0,5 µM; r.t; Titration of ds-DNA-27: c(TO) = 0,5 µM; (ds-DNA-27) = 0,5 µM; c(DMSO) = 10%; 80192 µM, 40096 µM, 20048 µM, 10024 µM, 5012 µM, 2056 µM, 1028 µM, 604 µM, 302 µM, 106 µM, 20 µM, 10 µM, 5 µM; r.t; Titration of ds-DNA-24-Bp: c(TO) = 0.5 µM; c(cMYC) = 0.25 µM; c(DMSO) = 10%; c(Ligand) = 20048 µM, 10024 µM, 5012 µM, 2056 µM, 1028 µM, 604 µM, 302 µM, 106 µM, 80 µM, 40 µM, 20 µM, 10 µM, 5 µM; r.t

3.6 Molecular Modelling

The structure-activity relationship was further explored as described below. First, the number of alkyl amino side chains had a significant impact on stabilization activity. The triazole linked nitro indoles with two *N,N*-dimethylmethanamine side chains, such as **9b** and **10b**, showed a significant increased stabilization ability with the *c-MYC* promoter G-quadruplex compared to conjugates with one alkyl amino side chain (**9a** and **9b**). In addition, the length of alkyl bridge between two linked to access the more surface area over G-Quartet is very important and type of alkyl amino side chain were also important factors for the stabilization activity of the tested compound. At the 3-position, of 5-nitroindole compounds with an *N¹,N¹,N²*-diethylethane-1,2-diamine (**9c**) showed weak stabilization activity for the G-quadruplex than the two *N,N*-dimethylmethanamine side chains (**9b**) and one *N,N*-dimethylmethanamine side chains (**9a**). At the 5-position,

compounds **9**, **9a**, and **9b**, with nitro functional group are relatively much better *c*-MYC promoter G4 ligands in compared to that of amino conjugates **10**, **10a** and **10b**.

Amongst them the best confirmed and profiled hits **9a** and **9b** were used in molecular modelling studies with the solution NMR structure for the *c*-MYC-G4 (PDB entry 2L7V).⁵² In order to elucidate the binding mode of best peptidomimetics ligands **9a** were performed with the *c*-MYC G4 (PDB:2L7V) at both 5'- and 3'-ends using Autodock 4.2 tools.⁵³ All ligands (*i.e.* **9a** and **9b**) were optimized and prepared for docking simulation with Gaussian 3.0 with semi empirical PM3 force field.⁵⁴ Molecular docking for best ligands were performed into the binding sites at the two ends of *c*-MYC G4-NMR structure of the 2:1 quindoline: MYC G4 complex in K⁺ solution (PDB: 2L7V).⁵² The best energy-scored (BE) pose of ligands **9a** and **9b** for both the 5'- and 3'-sites are predicted by AutoDock 4.2 are shown in the figure 65 and table 16.

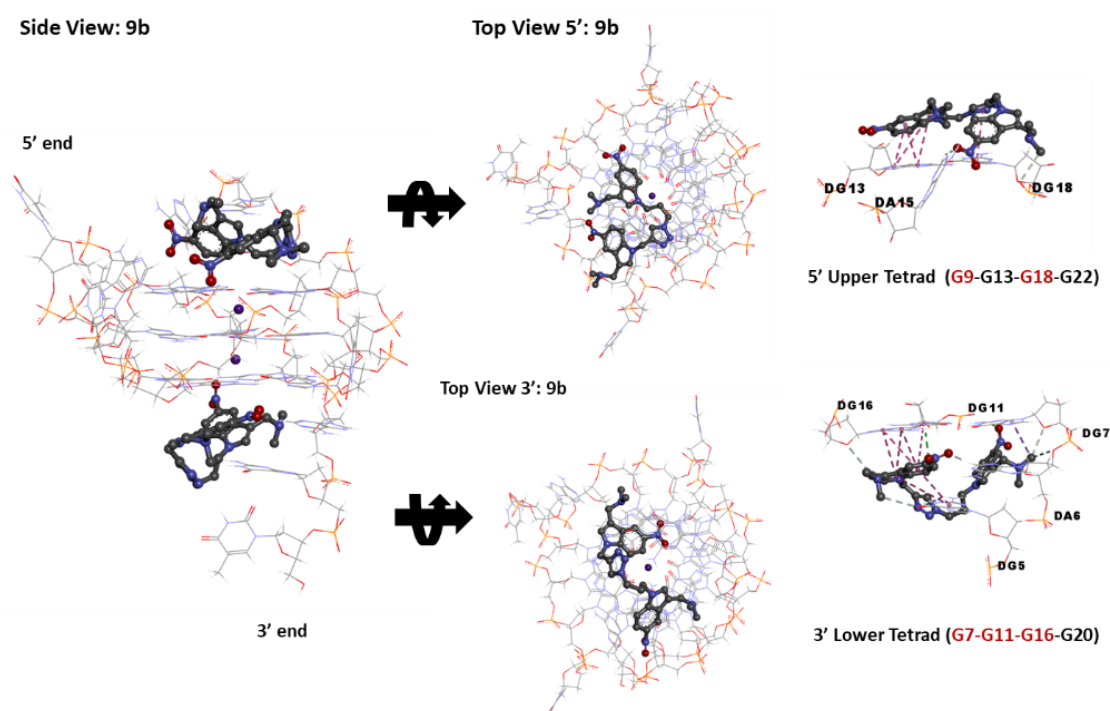


Figure 65. Best energy-scored (BE) predicted by AutoDock 4.2 of the most promising compound **9b** against the *c*-MYC G-quadruplex Pu22 (PDB ID: 2L7V)⁵²; compound **9b** is shown in ball and sticks model; Hydrogen bond interactions (highlighted in green color) and Pi-Pi stacking interactions (highlighted in pink color) are depicted as dashed lines. Compound **9b** is stacking at the 5' upper and 3' lower tetrads and interacting with G9, G18, G7, G11 and G16 (highlighted in red color).

Given the interesting results of biophysical experiments for compound **9a** and **9b** against multiple G4 sequences (Table 15) and particularly against the *c*-MYC G4, we have focused our

analysis on the predicted BE pose of compound **9a** and **9b** with *c*-MYC G4 (PDB entry **2L7V**) reported in Figure 65. Looking at the binding mode of drug-like triazole indole scaffolds **9b**, we can nicely see that the central aromatic core of scaffold **9a** and **9b** are stacking very well with both the 5'- and 3'-external tetrads, making extensive stacking π - π interactions with aromatic core. Interestingly, the positively charged dimethylamine side chain of scaffolds **9a** and **9b** is involved in the intermolecular salt bridges and some π -sigma stacking interactions with aromatic core and phosphate groups on the nucleotide backbone. Some interesting conventional hydrogen interactions were observed in between nitro group and aromatic core of nucleotide backbone (Fig. 65). These findings led us to deeper analyses of the binding mode of **9b** against *c*-MYC G4 through NMR spectroscopy.

Table 16. Ranking of ligands based upon their binding energies generated after docking analysis of **12**, **15**, **9a**, **9a** and **9b**

	Binding Energy (kcal/mol)					
9a	-6.13	-6.09	-6.03	-6.01	-5.91	-5.85
9b	-6.16	-6.10	-6.05	-6.03	-5.85	-5.66
9c	-5.16	-5.10	-5.19	-5.03	-5.03	-5.02
10a	-5.06	-5.10	-5.05	-5.03	-4.85	-4.66
10b	-4.16	-4.10	-4.05	-4.03	-3.85	-3.66

3.7 ¹D-NMR: Ligand binding observed by NMR-spectroscopy

The interactions of click conjugate **9b** with *c*-MYC G4 were further studied by NMR spectroscopy. Figure 66 shows NMR spectra of the DNA-ligand complexes at a ratio of 1:4 (DNA: ligand). In the spectrum of the DNA alone, signals (indicated by grey boxes) of minor conformations of the DNA are well visible. These signals disappear upon addition of ligand, indicating the stabilization of the major conformation upon binding. ¹H NMR investigation were used to know the interaction modes and dynamic binding of the MycG4 with click conjugate **9b** under the physiologically relevant K⁺ conditions. In ¹H NMR spectra (Fig. 66), upon addition of **9b** into the free MycG4 solution, a new set of sharp, distinct imino proton peaks appeared immediately at lower ligand equivalence 0–0.7, while the imino proton peaks of free MycG4 remained, indicating a slow-exchange binding of BMVC to the MycG4 on the NMR timescale, characteristic of high-affinity binding. Further 2D NMR investigations are under process to understand the mode of interaction of click conjugate with MycG4 DNA.

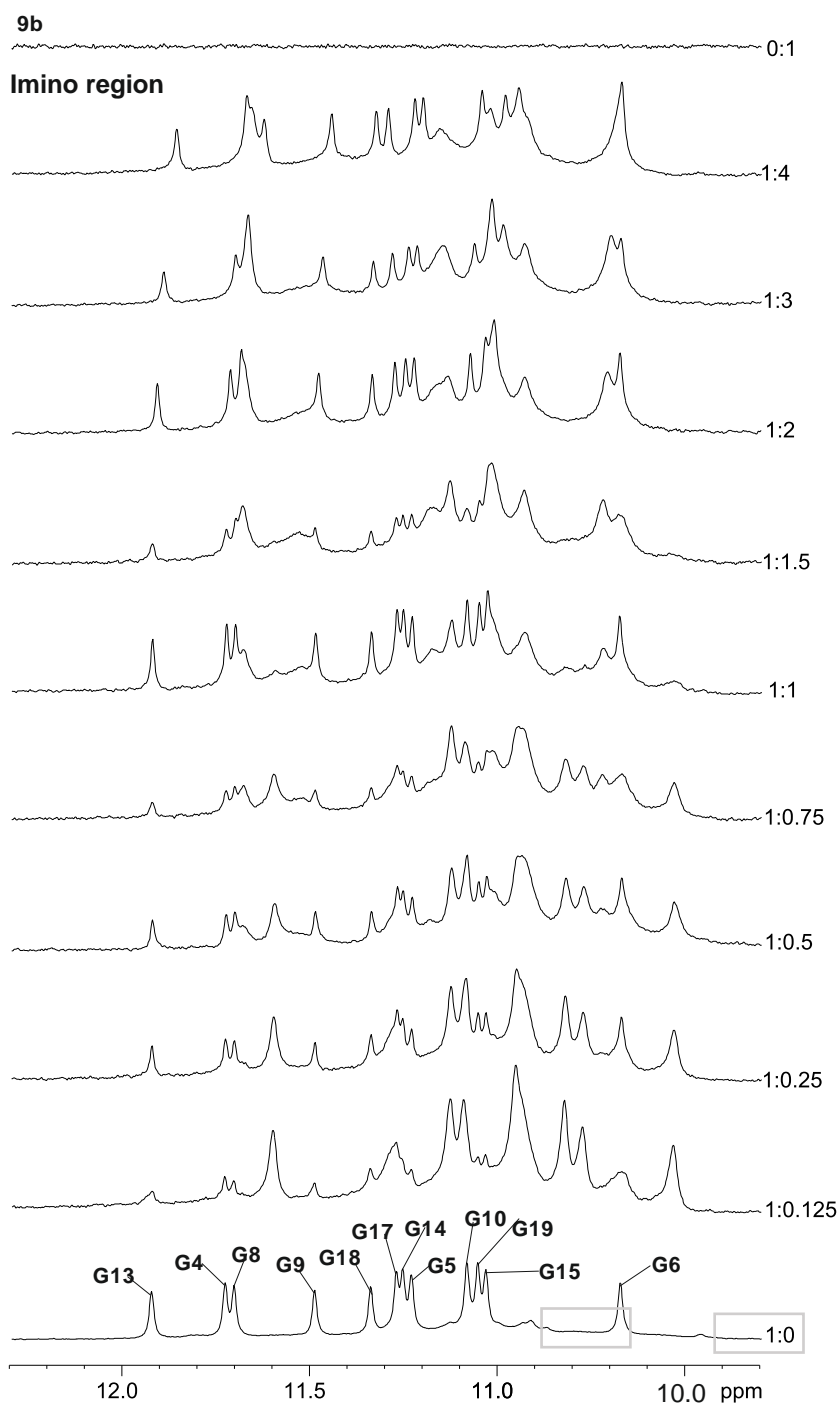


Figure 66. *c*-MYC: (9b) interactions observed by ^1H proton NMR spectra (imino). a) Comparison of interactions of *c*-MYC with 9b at a DNA: ligand ratio of 1:4. Signals clearly belonging to the respective ligand are indicated by asterisks. The spectra were recorded at 298 K, 600 MHz. Experimental conditions: 0.1mM *c*-MYC in 25mM KPi , 70mM KCl, pH 7, 0.025mM DSS and 10% DMSO- d_6 .

3.8 Cell Cytotoxicity

On the basis of all the data from the above experiments, we studied whether the most significant triazole derivatives could affect c-MYC transcription in cells. Firstly, promising ligands were analyzed *in vitro* by cell viability assays. HeLa cells were used to perform the cell proliferation assay, as they were not only having high levels of c-MYC expression but were also used for similar studies.⁵⁵ As described in the materials and methods, HeLa cells were exposed with various increasing concentrations (0-50 μM) of the candidate compounds and the parent building fragments **9a**, **9b**, **10a** and **10b** respectively for noting the effects of defined modifications. The percentage of living cells was then measured with Alamar-Blue reagent.⁵⁶ Best selected hits showed dose-dependent cytotoxic effects in HeLa cell line. Dose-response curves, the relative IC_{50} values, and the corresponding linear regression coefficients (R^2), for each G4-ligand, are reported. Interestingly, **9a** and **9b** showed comparable IC_{50} values. Compound **9a** and **9b** exhibited significantly enhanced inhibitory effect on the cellular proliferation of HeLa cells (Fig. 67c & Table 17). Likely, the new hits need to be further optimized in terms of target recognition to better exploit their potential activity on the target sequences.

Table 17. IC_{50} values derived for fragments screened *via* Thiazole displacement assays **9a**, **9b**, **10a** and **10b** in HeLa cell line after 72 h treatment.

	G-Quadruplex Ligands			
	9a	9b	10a	10b
IC_{50} (μM)	4.294	3.655	> 45	> 45

3.9 Western Blot analysis

Down-regulation of c-MYC expression in human cancer cells: To observe the effects of these ligands (**9a**, **9b**, **10a** and **10b**) at the translational level, HeLa cells were treated with defined similar concentrations of the ligands for 24 hr and monitored. The expression levels of c-MYC protein relative to the control cells, with western blot using anti-c-MYC antibody (Fig. 67 A-B). The protein expressions calculated for ligand treated cells from densitometry analysis of western blots were normalized against untreated control cells. In **9a** and **9b** treated HeLa cells; the c-MYC protein expression was downregulated by 40 % at 3 μM & 60 % at 10 μM respectively (Fig. 67a & 67B). Since **9b** is found to be the best G4-ligand in triazole

series, which could down-regulate c-MYC transcription and expression in cells, we next evaluated whether **9b** could affect cancer cell proliferation.

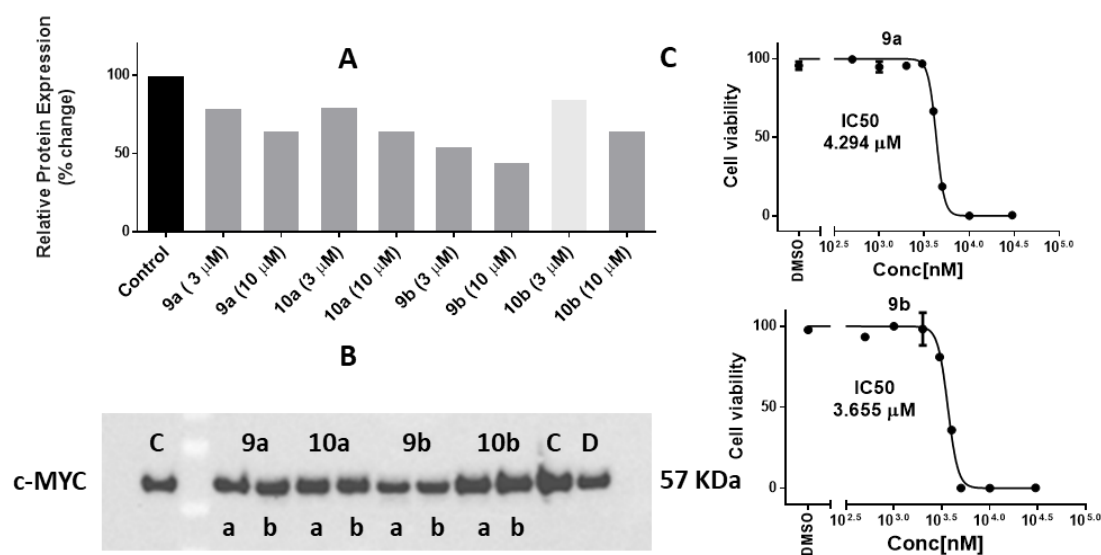


Figure 67. A) Bar diagram representing fold expression of *c*-MYC protein expression in HeLa cells B) Western blot analysis for *c*-MYC gene expression in of **9a**, **9b**, **10a**, and **10b** treated HeLa cells (a=3 μ M and b = 10 μ M); C & D) Dose-response and IC₅₀ values derived for best ligands 9a and 9b in HeLa cell line after 72 h treatment.

3.10 Cell Cycle Analysis

To investigate the effect of **9b** on different phases of untreated control cells maintained a normal percentage (68.1 %) of the total population in G₀/G₁ phase with 29.1 % in S/G₂ phases, respectively. cell cycle distribution in Hela cells, we have performed 7-amino actinomycin D (7-AAD)-based cell cycle analysis experiments. On exposure to **9b** at 4 μ M concentration, cells displayed a prominent G₁ phase arrest in flow cytometry (Fig. 68). The G₀/G₁ cell population increased to 74.0 % upon incubation with **9b** for 24 hr at 4 μ M concentration. The observed increase in G₀/G₁ population suggested that the compounds caused a significant block in S/G₂-phase progression. Most probably by hindering the DNA unwinding due to ligand-mediated stabilization of promoter *c*-MYC G-quadruplexes. The FACS analysis thus suggests that upon treatment with **9b**, cells undergo G₀/G₁ arrest with concomitant decrease of S/G₂-phase population thereby leading to cell death. The above results suggested that **9b** could inhibit proliferation of cancer cells, which could possibly be related to its binding with the *c*-MYC G₄ of the *c*-MYC promoter and repressing *c*-MYC oncogene transcription.

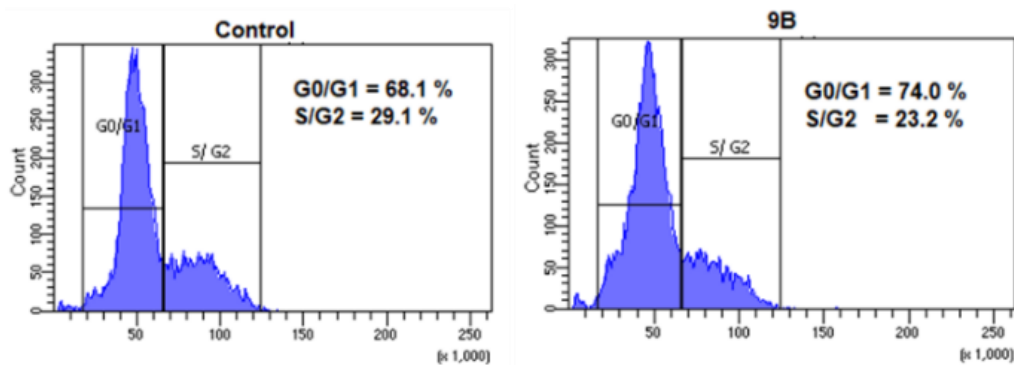


Figure 68. Cell cycle analysis of HeLa cells treated with ligand **9b** at 4 μM concentrations for 24 h.

3.11 Conclusion

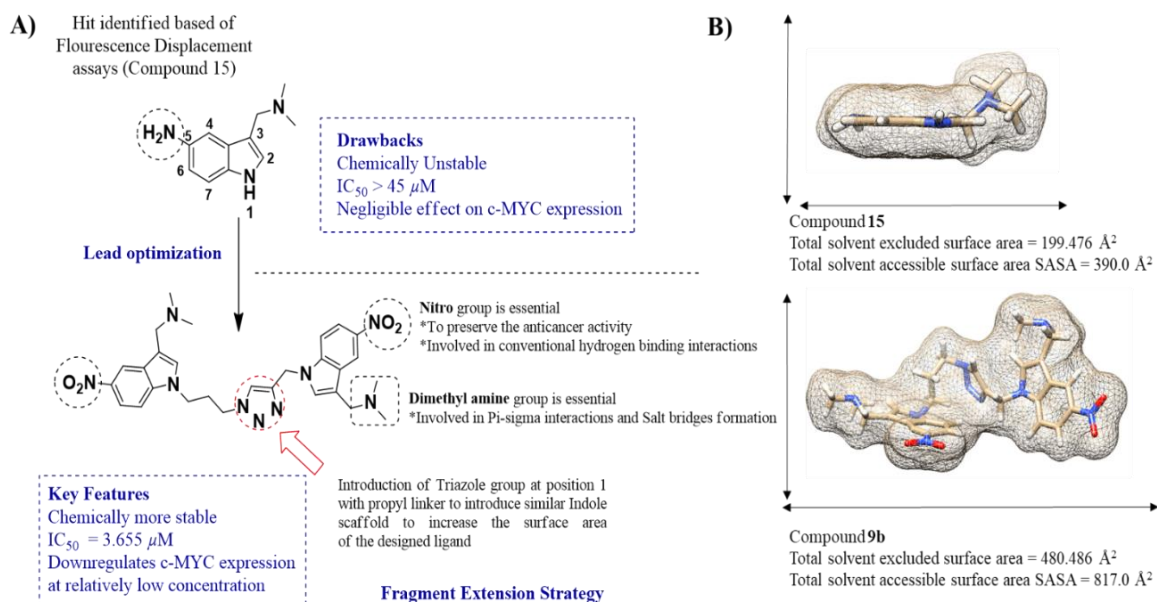


Figure 69. A) Hit Identified based of the FID assays and lead optimization strategy with structure activity relationship; B) Fragment extension strategy *via* click chemistry to generate triazole linked indole conjugates **9b**

A new series of triazole linked indole derivatives as a potential G quadruplex stabilizers have been described. The target ligands can be obtained through an efficient, convergent, synthetic route in moderate to good yields. In FID assays four ligands show specific binding to the *c-MYC* promoter G-quadruplex sequences compared to that of the other G-quadruplex sequences (*c-KIT-1*, *BCL2*, *h-Telo*). The synthesized triazole linked indole derivatives **9a**, **9b**, **10a** and **10b** are selective towards *c-MYC* G4-DNA vs. duplex-DNA. The planarity of the aromatic core, and its ability to occupy more surface area by stacking over the G-quartet greatly affect the ability of the compounds to stabilize the G-quadruplex. In addition to that in our biophysical and biological evaluation we found that the triazole linked nitro indoles are

more promising than the amino indole derivatives. Further the importance of the nitro functional group has been justified by molecular docking studies, in which interesting conventional hydrogen bonding interactions were observed in between the nitro group and the guanine base pairs of the G-quadruplex structure. In cell viability assays, most of the synthesized triazole linked nitro indoles (**9a** and **9b**) has shown potent cell growth inhibitory activity against human carcinoma (cervical) HeLa cell lines. In vitro cellular studies confirms that the novel triazole linked 5-nitro indole derivatives down-regulate *c-MYC* expression in cancer cells *via* stabilizing its promoter quadruplex structure and have the potential to induce cell cycle arrest in G0/G1 phase.

The weak correlation between FID assays and cell viability assay results suggests that the synthesized ligands may have additional mechanisms of action besides targeting G-quadruplexes. Docking studies give us a clear idea about the nature of the ligand-quadruplex DNA interactions and rationalize the moderate binding affinity the ligands shown in the FID and which was further confirmed by the NMR investigations, NMR analysis suggests that **9b** interacts in slow exchange regime with the terminal G-quartets (5' and 3'-end). Further investigation are under process to reach concrete conclusion. Our present results increased our understanding of the effect of c-MYC G4 on oncogene transcriptional regulation. The present study also suggests that the triazole linked indole derivative **9b** could become a potential lead compound for further development for cancer treatment with a new target of the *c-MYC* promoter G-Quadruplex. This information can guide the optimization of ligands **9a** and **9b**, for increased binding affinity and selectivity toward G-quadruplex structures.

3.12 Experimental Section

Chemistry: All solvents and reagents were purified by standard techniques or used as supplied from commercial sources (Sigma-Aldrich Corporation unless stated otherwise). All reactions were generally carried out under inert atmosphere unless otherwise noted. TLC was performed on Kieselgel 60 F254 plates, and spots were visualized under UV light. Products were purified by flash chromatography on silica gel (100-200 mesh). ¹H NMR spectra were recorded at 600 MHz instruments at 298 K. ¹³C NMR spectra were recorded on either 151 MHz with complete proton decoupling. Chemical shifts are reported in parts per million (ppm) and are referred to the residual solvent peak. The following notations are used: singlet (s); doublet (d); triplet (t); quartet (q); multiplet (m); broad (br). Coupling constants are quoted in Hertz and are denoted as J. Mass spectra were recorded on a Micromass Q-Tof (ESI) spectrometer.

Generalized procedure for the reduction of nitro to amine: a solution of a nitro conjugate in 10 mL of ethanol was dropped into a suspension of 10% Pd/C (12.5 mg) saturated with H₂ in ethanol (20 ml). The mixture was stirred at room temperature for 3 h, according to TLC analysis (MeOH/DCM 1:3). The catalyst was recovered and the filtrate was evaporated under reduced pressure to dryness to give the corresponding amino-indole.

Generalized procedure for Intermediate 3 & 7 with Vilsmeier Haack reaction (A): POCl₃ (1.2 ml, 9.25 mmol, 1.5 eq.) as reaction solvent was added in dropwise manner to the dried DMF (10 ml) under inert condition at 0 °C. This mixture was allowed to stir at room temperature for 1 hour followed with the addition of reactant (1 eq) **2** and **6**. The reaction mixture was then allowed to react for 1 hour by stirring at room temperature. The reaction mixture was then quenched with equal amount of ice water and 50 % NaOH was added until a pH of 9.

Generalized procedure for the condensation of substituted carboxaldehyde and corresponding amines (B): Substituted-carboxaldehyde **2a**, **1b**, **1c**, **1d** and **1e** (1eq) and substituted amines (3 eq) were dissolved in 50 mL of dry EtOH: CH₃CN (1:1). The resulting mixture stirred for 2 h at room temperature and then the solvent was concentrated under reduced pressure. The resultant residue was dissolved in 20 mL EtOH and then NaBH₄ (10 eq) was added to it in portion wise. The reaction mixture is allowed to stir for 24 h at room temperature, and then the excess of NaBH₄ was filtered off and the solvent was evaporated to dryness. The resultant solid was treated with deionized water and extracted with CH₂Cl₂ (3 x 50 mL). The organic phase was evaporated under reduced pressure. Generalized procedure for making HCl salt (C): The solid was dissolved in EtOH: dioxane (1/3) and precipitated with aqueous HCl 37% to obtain its hydrochloride salt. The precipitate was filtered and recrystallized with MeOH to give desired product.

Generalized procedure for nucleophilic substitution of alkynes: 5-Nitro-1H-indole (2.01 g, 12.3 mmol) has added to a solution of anhydrous KOH (0.692 g, 12.3 mmol) in DMF (100 ml) at room temperature for 30 min. followed with the addition of 3-bromoprop-1-yne & 4-bromobut-1-yne (4 eq) and reaction mixture was allowed to stir overnight.

Generalized procedure for Click Chemistry: Corresponding azide **8** & **8a** (1 eq.) reacted with corresponding alkyne **2** & **4** (**10**, 40.97 mg, 0.195 mmol, 1.1 eq.) by adding CuSO₄ · 5H₂O (18.22 mg, 0.089 mmol, 0.5 eq), Na-ascorbate (18.73 mg, 0.089, 0.5 eq), 10ml t-BuOH and 10ml H₂O. The reaction lasted 8 hours at room temperature. After completion of the reaction. The reaction mixture was extracted with ethyl acetate and water. The separated organic phase dissolved in ethyl acetate was concentrated under reduced pressure by Rotavapor and

dried with MgSO₄. Finally, through the column chromatography G-quadruplex stabilization agent **9**, **9a**, **9b** and **9c** was purified.

5-nitro-1-(prop-2-yn-1-yl)-1H-indole (2): Yield 36 %, yellow solid; *R_f* (eluent EA/cyclo 1:3) 0.43; M.p: 113 - 115 °C. ¹H NMR (600 MHz, CDCl₃) δ 8.57 (d, *J* = 2.2 Hz, 1H), 8.17 - 8.11 (m, 1H), 7.43 (d, *J* = 9.1 Hz, 1H), 7.37 (d, *J* = 3.3 Hz, 1H), 6.71 (d, *J* = 3.3 Hz, 1H), 4.93 (d, *J* = 2.6 Hz, 2H), 2.48 (t, *J* = 2.6 Hz, 1H); ¹³C NMR (151 MHz, CDCl₃) δ 142.04, 138.04, 130.46, 128.18, 118.25, 117.63, 109.37, 104.64, 76.33, 74.60, 36.27 ppm; HRMS (ESI): calcd for C₁₁H₈N₂O₂ 200.06; found 201.06

5-nitro-1-(prop-2-yn-1-yl)-1H-indole-3-carbaldehyde (3): crude product was purified on silica as a stationary phase; with 1:4 EA: CH. The product was a yellow solid. ¹H-NMR (600 MHz, DMSO-*d*₆): δ 10.01 (s, 1H), 8.93 (t, *J* = 3.6 Hz, 1H), 8.61 (s, 1H), 8.23 (dd, *J* = 9.1, 2.4 Hz, 1H), 7.87 (d, *J* = 9.1 Hz, 1H), 5.33 (d, *J* = 2.5 Hz, 2H), 3.61 (t, *J* = 2.5 Hz, 1H); ¹³C NMR (151 MHz, DMSO-*d*₆): δ 185.49, 143.43, 142.92, 139.37, 123.94, 118.99, 118.41, 117.24, 112.08, 77.32, 36.41ppm

N,N-dimethyl-1-(5-nitro-1-(prop-2-yn-1-yl)-1H-indol-3-yl)methanamine (4): 5-nitro-1-(prop-2-yn-1-yl)-1H-indole-3-carbaldehyde and 3 equivalent of dimethylamine (40% aqueous, 60 mL) allowed to react according to the protocol **B**. ¹H-NMR (600 MHz, DMSO-*d*₆): δ 8.61 - 8.59 (d, *J* = 2.1 Hz, 1H), 8.11 (dd, *J* = 9.1, 2.2 Hz, 1H), 7.72 (s, 1H), 7.75 (d, *J* = 9.1 Hz, 1H), 5.17 (d, *J* = 2.4 Hz, 2H), 4.29 (s, 2H), 3.61 (bs, 1H), 2.17 - 2.18 (m, 6H). ¹³C NMR (151 MHz, DMSO-*d*₆): δ 141.08, 139.16, 131.14, 127.62, 117.15, 116.99, 116.92, 115.37, 110.96, 78.78, 76.28, 53.96, 44.96, 35.93; HRMS (ESI): calcd for C₁₄H₁₅N₃O₂ 257.12; found 258.12

5-nitro-1-(3-bromopropyl)-1H-Indole (5): 5-Nitro-1H-indole (2.01 g, 12.3 mmol) was added to a solution of anhydrous KOH (0.692 g, 12.3 mmol) in DMF (100 ml) at room temperature for 30 min. Followed with the addition of 1, 3-dibromopropane (3.77 ml, 37.0 mmol) and reaction mixture was then allowed to stir for overnight. Solvent was removed under reduced pressure, and the crude product was purified on silica as a stationary phase; Yield 46 % as yellow crystals; *R_f* = 0.65 (eluent EA/n-Hex 1:4); mp 102°C. ¹H NMR (600 MHz, CDCl₃): δ 2.33-2.44 (m, 2H), 3.28-3.35 (m, 2H), 4.40 (t, *J* = 6.5 Hz, 2H), 6.78 (d, *J* = 3.1 Hz, 1H), 7.67 (d, *J* = 3.2 Hz, 1H), 7.71 (d, *J* = 9.1 Hz, 1H), 8.04 (dd, *J* = 2.3, 9.1 Hz, 1H), 8.58 (d, *J* = 2.2 Hz, 1H). ¹³C NMR (151 MHz, CDCl₃): δ = 30.05, 32.70, 44.53, 104.54, 109.33, 117.52, 118.42, 127.92, 131.22, 138.84, 141.79 ppm; ESI-MS: *m/z* 282.0 [M+H]; HRMS: *m/z* calcd for C₁₁H₁₁BrN₂O₂ [M+H] + 282.0, found 282.0

1-(3-azidopropyl)-5-nitro-1H-indole (6): To a stirring mixture of **4** (1g, 3.225 mmol, 1 eq) in 15 ml Dry DMF about 3 eq of sodium azide (0.6386 g, 9.667 mmol) has been added and allowed to reflux for 3-5 hrs. Followed by quenching with Ice water and extraction with ethyl acetate. Organic extract was then concentrated under reduced pressure, and the crude product was purified on silica as a stationary phase with Cyclohexane: Ethylacetate (9:1) over silica gel column. Yield 56 %, yellow solid; R_f (eluent EA/cyclo 1:3) 0.63; Mp 92 - 94 °C. ^1H NMR (600 MHz, DMSO- d_6): δ 10.00 (s, 1H), 8.94 (d, $J = 2.1$ Hz, 1H), 8.62 (s, 1H), 8.19 (dd, $J = 9.1, 2.3$ Hz, 1H), 7.90 (d, $J = 9.1$ Hz, 1H), 4.43 (t, $J = 7.0$ Hz, 2H), 3.40 (t, $J = 6.6$ Hz, 2H), 3.34 (H₂O), 2.07-2.11 ppm (br, 2H); ^{13}C NMR (151 MHz, DMSO- d_6): δ 185.13, 146.04-144.33, 143.35, 139.92, 123.97, 118.75, 118.26, 117.17, 111.90, 47.89, 44.23, 28.53 ppm; HRMS (ESI): calcd for C₁₁H₁₁N₅O₂ 245.09; found [M+H] 246.09

1-(3-azidopropyl)-5-nitro-1H-indole-3-carbaldehyde (7): Crude product further purified with Cyclohexane: Ethylacetate (9:1) over silica gel column. Yield 66 %, yellow solid; R_f (eluent EA/cyclo 1:3) 0.23; Mp 102 - 104 °C. ^1H NMR (600 MHz, DMSO- d_6): δ 10.00 (s, 1H), 8.94 (d, $J = 2.1$ Hz, 1H), 8.62 (s, 1H), 8.19 (dd, $J = 9.1, 2.3$ Hz, 1H), 7.90 (d, $J = 9.1$ Hz, 1H), 4.43 (t, $J = 7.0$ Hz, 2H), 3.40 (t, $J = 6.6$ Hz, 2H), 2.07-2.11 ppm (br, 2H); ^{13}C NMR (151 MHz, DMSO- d_6): δ 185.13, 146.04, 144.33, 143.35, 139.92, 123.97, 118.75, 118.26, 117.17, 111.90, 47.89, 44.23, 28.53 ppm; HRMS (ESI): calcd for C₁₂H₁₁N₅O₃ 273.09; found 273.09

1-(1-(3-azidopropyl)-5-nitro-1H-indol-3-yl)-N,N-dimethylmethanamine (8): compound **8** (1g, 3.663 mmol, 1 eq) and 3 eq of N1, N1-dimethylethane-1,2-diamine (1.284 g, 11.070 mmol) were allowed to react according to generalized procedure **B**. Crude product was then further purified with DCM: MeOH (9:1) over silica gel column. Yield 26 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.43; Mp 102 - 104 °C. ^1H -NMR (600 MHz, DMSO- d_6): δ 8.66 (d, $J = 1.8$ Hz, 1H), 8.03 (dd, $J = 9.1, 2.1$ Hz, 1H), 7.67 (d, $J = 9.1$ Hz, 1H), 7.56 (s, 1H), 4.29 (t, $J = 6.9$ Hz, 2H), 3.95 (s, 2H), 3.30 - 3.32 (bs, 5H), 2.64 (t, $J = 6.4$ Hz, 2H), 2.44 (q, $J = 7.1$ Hz, 4H), 1.99 - 2.03 (bs, 2H), 0.91 ppm (t, $J = 7.1$ Hz, 6H); ^{13}C NMR (151 MHz, DMSO- d_6): δ 140.35, 139.09, 126.55, 116.58, 110.24, 51.79, 47.98, 46.34, 43.59, 43.09, 28.90, 11.61 ppm. HRMS (ESI): calcd for C₁₄H₁₈N₆O₂ 302.15; found 302.15

N,N-dimethyl-N'-[[5-nitro-1-(3-Azidopropyl)-1H-indole]methyl]-1,2-Ethanediamine (8a): compound **8** (1g, 3.663 mmol, 1 eq) and about 3 eq of N1, N1-diethylethane-1,2-diamine (1.284 g, 11.070 mmol) allowed to react according to generalized procedure **B**. Crude product was then further purified with DCM: MeOH (9:1) over silica gel column. Yield 46 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.33; Mp 162 - 164 °C. ^1H NMR (600 MHz, DMSO) δ 8.57 (d, $J = 10.6$ Hz, 2H), 8.09 (s, 1H), 8.01 (dd, $J = 27.3, 9.1$ Hz, 2H), 7.80 (d, $J = 9.1$ Hz, 1H), 7.71 (d, $J =$

3.1 Hz, 1H), 7.62 (d, $J = 3.0$ Hz, 1H), 7.57 (d, $J = 9.1$ Hz, 1H), 6.76 (dd, $J = 15.1, 3.1$ Hz, 2H), 5.58 (s, 2H), 4.31 (dt, $J = 22.5, 7.0$ Hz, 4H), 2.33 (p, $J = 6.9$ Hz, 2H). ^{13}C NMR (151 MHz, DMSO- d_6): δ 140.11, 139.22, 130.92, 130.39, 126.30, 116.70, 116.47, 116.12, 114.59, 110.20, 85.45, 55.25, 53.42, 53.03, 52.79, 51.84, 47.46, 43.62, 30.81, 28.75, 24.19, 23.06 ppm; HRMS (ESI): calcd for $\text{C}_{18}\text{H}_{27}\text{N}_7\text{O}_2$ 373.22; found 374.22

5-nitro-1-(3-(4-((5-nitro-1H-indol-1-yl)methyl)-1H-1,2,3-triazol-1-yl)propyl)-1H-indole (9): Crude product was then further purified with DCM: MeOH (9:1) over silica gel column. Yield 46 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.33; Mp 162 - 164 °C. ^1H NMR (600 MHz, DMSO) δ 8.63 (d, $J = 2.3$ Hz, 1H), 8.60 (d, $J = 2.3$ Hz, 1H), 8.03 (dt, $J = 9.1, 2.1$ Hz, 2H), 7.66 (dd, $J = 9.1, 4.7$ Hz, 2H), 7.55 (d, $J = 5.6$ Hz, 1H), 7.54 (d, $J = 9.6$ Hz, 1H), 4.35 - 4.21 (m, 4H), 3.92 (s, 2H), 2.71 - 2.60 (m, 3H), 2.35 (dd, $J = 18.3, 2.7$ Hz, 2H), 2.24 (dd, $J = 14.1, 7.7$ Hz, 6H), 2.02 - 1.87 (m, 4H), 1.76 - 1.49 ppm (m, 4H). ^{13}C NMR (151 MHz, DMSO- d_6): δ 142.83, 140.46, 138.57, 132.67, 131.93, 127.17, 123.65, 123.16, 116.27, 110.62, 110.10, 104.38, 104.21, 103.47, 46.82, 42.96, 41.08, 30.16 ppm; HRMS (ESI): calcd for $\text{C}_{22}\text{H}_{19}\text{N}_7\text{O}_4$ 445.1; found 445.1493

N,N-dimethyl-1-(5-nitro-1-(3-(4-((5-nitro-1H-indol-1-yl)methyl)-1H-1,2,3-triazol-1-yl)propyl)-1H-indol-3-yl)methanamine (9a): Crude product was then further purified with DCM: MeOH (9:1) over silica gel column. Yield 46 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.33; Mp 162 - 164 °C. ^1H NMR (600 MHz, DMSO- d_6) δ 8.63 (d, $J = 2.3$ Hz, 1H), 8.60 (d, $J = 2.3$ Hz, 1H), 8.03 (dt, $J = 9.1, 2.1$ Hz, 2H), 7.66 (dd, $J = 9.1, 4.7$ Hz, 2H), 7.55 (d, $J = 5.6$ Hz, 1H), 7.54 (d, $J = 9.6$ Hz, 1H), 4.35 - 4.21 (m, 4H), 3.92 (s, 2H), 2.71 - 2.60 (m, 3H), 2.35 (dd, $J = 18.3, 2.7$ Hz, 2H), 2.24 (dd, $J = 14.1, 7.7$ Hz, 6H), 2.02 - 1.87 (m, 4H), 1.76 - 1.49 ppm (m, 4H). ^{13}C NMR (151 MHz, DMSO- d_6): δ 142.84, 140.76, 140.45, 138.80, 138.23, 132.53, 131.32, 127.44, 126.82, 123.25, 117.40, 116.78, 116.44, 110.66, 110.03, 104.03, 53.40, 46.67, 44.66, 43.19, 41.25, 29.87, 26.10 ppm; HRMS (ESI): calcd for $\text{C}_{25}\text{H}_{26}\text{N}_8\text{O}_4$ 502.22; found 502.20715

1-(1-(3-(4-((3-((dimethylamino)methyl)-5-nitro-1H-indol-1-yl)methyl)-1H-1,2,3-triazol-1-yl)propyl)-5-nitro-1H-indol-3-yl)-N,N-dimethylmethanamine (9b): Crude product was then further purified with DCM: MeOH (9:1) over silica gel column. Yield 46 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.33; Mp 162 - 164 °C. ^1H NMR (600 MHz, DMSO- d_6) δ 8.58 (dd, $J = 7.4, 2.3$ Hz, 2H), 8.07 (s, 1H), 8.03 (dt, $J = 10.8, 3.7$ Hz, 1H), 7.95 (dt, $J = 9.2, 4.6$ Hz, 1H), 7.75 (t, $J = 8.6$ Hz, 1H), 7.58 (s, 1H), 7.52 - 7.48 (m, 2H), 5.52 (s, 2H), 4.36 - 4.29 (m, 2H), 4.21 (q, $J = 7.3$ Hz, 2H), 3.55 (t, $J = 18.5$ Hz, 4H), 2.35 - 2.27 (m, 2H), 2.16 ppm (s, 12H). ^{13}C NMR (151 MHz, DMSO- d_6): δ 142.91, 140.45, 139.06, 133.55, 131.27, 131.12, 126.99,

123.36, 121.90, 116.48, 116.38, 111.07, 110.60, 109.97, 69.23, 53.66, 47.25, 46.63, 44.73, 42.90, 40.85, 30.09 ppm; HRMS (ESI): calcd for C₂₈H₃₃N₉O₄ 559.27; found 559.290

N1,N1-diethyl-N2-((5-nitro-1-(3-(4-((5-nitro-1H-indol-1-yl)methyl)-1H-1,2,3-triazol-1-yl)propyl)-1H-indol-3-yl)methyl)ethane-1,2-diamine (9c): Crude product was then further purified with DCM: MeOH (9:1) over silica gel column. Yield 46 %, yellow solid; *R_f* (eluent MeOH/DCM 1:3) 0.33; Mp 162 - 164 °C. ¹H NMR (600 MHz, DMSO-*d*₆) δ 8.61 (d, *J* = 2.1 Hz, 1H), 8.56 (d, *J* = 2.1 Hz, 2H), 8.07 (s, 1H), 8.02 (dd, *J* = 9.1, 2.2 Hz, 1H), 7.95 (dd, *J* = 9.1, 2.2 Hz, 1H), 7.78 (d, *J* = 9.1 Hz, 1H), 7.69 (d, *J* = 3.2 Hz, 1H), 7.49 (d, *J* = 8.7 Hz, 2H), 6.76 (d, *J* = 3.0 Hz, 1H), 5.56 (s, 2H), 4.31 (t, *J* = 7.1 Hz, 2H), 4.22 (t, *J* = 7.0 Hz, 2H), 3.87 (s, 2H), 3.87 (s, 2H), 3.36 (H₂O), 2.57 (t, *J* = 6.4 Hz, 2H), 2.45 (t, *J* = 6.4 Hz, 2H), 2.39 (q, *J* = 7.1 Hz, 4H), 2.29 (dd, *J* = 14.0, 7.0 Hz, 2H), 0.87 (t, *J* = 7.1 Hz, 6H). ¹³C NMR (151 MHz, DMSO-*d*₆): δ 142.95, 141.83 - 140.99, 140.59, 139.00, 138.50, 132.49, 130.11, 128.87 - 128.58, 127.07, 123.38, 119.38 - 118.70, 117.52, 117.36 - 116.39, 111.67 - 111.63, 110.35, 103.94, 52.25, 46.93, 46.68, 46.50, 43.93, 42.99, 41.27, 30.15, 11.75 ppm; HRMS (ESI): calcd for C₂₉H₃₅N₉O₄ 573.28; found 574.2884

1-(3-(4-((5-amino-1H-indol-1-yl)methyl)-1H-1,2,3-triazol-1-yl)propyl)-1H-indol-5-amine (10): Crude product was then further purified with DCM: MeOH (9:1) over silica gel column. Yield 46 %, yellow solid; *R_f* (eluent MeOH/DCM 1:3) 0.33; Mp 162 - 164 °C. ¹H NMR (600 MHz, DMSO-*d*₆) δ 7.99 (s, 1H), 7.22 (t, *J* = 6.5 Hz, 2H), 7.12 (d, *J* = 3.0 Hz, 1H), 7.04 (d, *J* = 8.6 Hz, 1H), 6.69 (s, 2H), 6.52 (dd, *J* = 10.9, 4.2 Hz, 2H), 5.31 (s, 2H), 4.53 (s, 4H), 4.27 (t, *J* = 7.1 Hz, 2H), 4.03 (t, *J* = 6.9 Hz, 2H), 2.23 (p, *J* = 6.9 Hz, 2H). ¹³C NMR (151 MHz, DMSO-*d*₆): δ 142.83, 140.46, 138.57, 132.67, 131.93, 127.17, 123.65, 123.16, 116.27, 110.62, 110.10, 104.38, 104.21, 103.47, 46.82, 42.96, 41.08, 30.16 ppm; HRMS (ESI): calcd for C₂₂H₂₃N₇ 385.20; found 385.2009

1-(3-(4-((5-amino-1H-indol-1-yl)methyl)-1H-1,2,3-triazol-1-yl)propyl)-3-((dimethylamino)methyl)-1H-indol-5-amine (10a): Crude product was then further purified with DCM: MeOH (9:1) over silica gel column. Yield 46 %, yellow solid; *R_f* (eluent MeOH/DCM 1:3) 0.33; Mp 162 - 164 °C. ¹H NMR (600 MHz, DMSO-*d*₆) δ 8.63 (d, *J* = 2.3 Hz, 1H), 8.60 (d, *J* = 2.3 Hz, 1H), 8.03 (dt, *J* = 9.1, 2.1 Hz, 2H), 7.66 (dd, *J* = 9.1, 4.7 Hz, 2H), 7.55 (d, *J* = 5.6 Hz, 1H), 7.54 (d, *J* = 9.6 Hz, 1H), 4.35 - 4.21 (m, 4H), 3.92 (s, 2H), 2.71 - 2.60 (m, 3H), 2.35 (dd, *J* = 18.3, 2.7 Hz, 2H), 2.24 (dd, *J* = 14.1, 7.7 Hz, 6H), 2.02 - 1.87 (m, 4H), 1.76 - 1.49 ppm (m, 4H). ¹³C NMR (151 MHz, DMSO-*d*₆): δ 140.11, 139.22, 130.92, 130.39, 126.30, 116.70, 116.47, 116.12, 114.59, 110.20, 85.45, 55.25, 53.42, 53.03, 52.79, 51.84, 47.46, 43.62, 30.81, 28.75, 24.19, 23.06 ppm; HRMS (ESI): calcd for C₂₅H₃₀N₈ 442.26; found 465.2485

1-(3-(4-((5-amino-3-((dimethylamino)methyl)-1H-indol-1-yl)methyl)-1H-1,2,3-triazol-1-yl)propyl)-3-((dimethylamino)methyl)-1H-indol-5-amine (10b): Crude product was then further purified with DCM: MeOH (9:1) over silica gel column. Yield 46 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.33; Mp 162 - 164 °C. ^1H NMR (600 MHz, DMSO- d_6) δ 7.96 (d, J = 4.6 Hz, 1H), 7.18 (dd, J = 8.1, 4.0 Hz, 1H), 7.09 (d, J = 4.2 Hz, 1H), 7.01 (d, J = 2.6 Hz, 1H), 6.99 (s, 1H), 6.76 (dd, J = 4.3, 1.9 Hz, 2H), 6.53 - 6.48 (m, 2H), 5.26 (s, 2H), 4.36 - 4.17 (m, 4H), 3.99 (t, J = 6.9 Hz, 2H), 3.12-3.14 (br, 4H), 2.16 (d, J = 2.4 Hz, 2H), 2.14 - 2.11 ppm (m, 12H). ^{13}C NMR (151 MHz, DMSO- d_6): δ 143.58, 141.28, 129.58, 128.71, 127.36, 123.01, 116.37, 111.79, 109.97, 109.48, 102.84, 54.61, 54.13, 53.66, 47.02, 44.56, 42.50, 40.51, 30.09 ppm; HRMS (ESI): calcd for $\text{C}_{28}\text{H}_{37}\text{N}_9$ 499.32; found $\text{C}_{28}\text{H}_{37}\text{N}_9\text{Na}$ 522.306

1-methyl-5-Amino-1H-indole (12): a solution of a 1-methyl-5-nitro-1H-indole (2) (5.87 mmol) is treated according to protocol A. Yield 96 %; brown solid. R_f (eluent MeOH/DCM 1:3) 0.44; Mp 87 °C; ^1H -NMR (600 MHz, DMSO- d_6) δ 7.09 (dd, J = 9.2, 5.8 Hz, 2H), 6.68 (s, 1H), 6.54 (d, J = 10.5 Hz, 1H), 6.10 (d, J = 2.9 Hz, 1H), 4.45 (s, 2H), 3.66 ppm (s, 3H). ^{13}C NMR (151 MHz, DMSO- d_6): δ 211.30, 141.55, 130.56, 129.22, 111.82, 109.77, 103.59, 98.77, 32.65 ppm; HRMS: m/z calcd for $\text{C}_9\text{H}_{10}\text{N}_2$ $[\text{M}+\text{H}]^+$ 146.19, found 146.19

5-Nitroindole-3-carboxaldehyde (13): crude product was obtained from 5-nitro-Indole (1) by vilsmeier Haack general protocol. Further purified on silica as a stationary phase; with 1:4 MeOH: DCM. Yield 85 % as yellow amorphous solid; R_f (eluent MeOH/ DCM 1:4) 0.55; M.p: 109-111 °C; ^1H -NMR (600 MHz, CDCl_3) δ 9.64 (s, 1H), 8.69 (d, J = 1.8 Hz, 1H), 7.73 (m, 2H), 7.18 ppm (d, J = 6 Hz, 1H), 2.50 - 2.62 (br, 1H); ^{13}C -NMR: (151 MHz, CDCl_3) δ [ppm] 183.97, 138.90, 118.13, 117.38, 111.91 ppm; HRMS (ESI): m/z calculated for $\text{C}_9\text{H}_6\text{N}_2\text{O}_3$: 190.04, found: 191.09

N,N-dimethyl-1-(5-nitro-1H-indol-3-yl)methanamine (14): 5-Nitroindole-3-carboxaldehyde (13) (1eq) and 3 equivalent of dimethylamine (40% aqueous, 60 mL) allowed to react according to the protocol B. Crude product is obtained as yellow crystalline solid and further purified on silica as a stationary phase with DCM: MeOH (9:1) over silica gel column. Yield 60 % as yellow solid; R_f (eluent MeOH/ DCM 1:4) 0.45; mp 114 - 116 °C. ^1H -NMR (600 MHz, DMSO- d_6) δ 11.71 (s, 1H), 8.62 (d, J = 1.9 Hz, 1H), 8.00 (dt, J = 11.6, 4.2 Hz, 2H), 7.56 - 7.51 (m, 1H), 3.68 (s, 2H), 2.22 ppm (s, 6H). ^{13}C -NMR: (151 MHz, DMSO- d_6) δ 140.09, 139.47, 132.50, 128.31, 126.81, 116.34, 111.59, 69.59, 54.07, 44.65 ppm; HRMS (ESI): m/z calculated for $\text{C}_{11}\text{H}_{13}\text{N}_3\text{O}_2$: 219.10, found: 219.10

N,N-dimethyl-1-(5-amino-1H-indol-3-yl)methanamine (15): Compound 9 has been treated with protocol A. The crude product was purified with DCM: MeOH (9:1) over silica gel

column. (**10a**) Yield 96 %; Brown solid. R_f (eluent MeOH/DCM 1:3) 0.41; Mp 87 - 89 °C; $^1\text{H-NMR}$ (600 MHz, $\text{DMSO-}d_6$) δ 10.38 (s, 1H), 7.04 - 6.99 (m, 2H), 6.76 (s, 1H), 6.47 (dd, $J = 8.6, 1.9$ Hz, 1H), 3.31-3.32 (b, 2H) 4.41 (s, 2H), 2.12 ppm (s, 6H). $^{13}\text{C-NMR}$: (151 MHz, $\text{DMSO-}d_6$) δ 211.06, 140.62, 129.95, 128.50, 124.07, 111.63, 111.20, 110.40, 102.17, 54.50, 44.83 ppm; HRMS (ESI): m/z calculated for $\text{C}_{11}\text{H}_{15}\text{N}_3$: 189.13, found: 189.13.

Experimental procedure for fragment screening with FID assays, $^1\text{D-NMR}$ titration, cells and culture conditions, cell proliferation assay, western blot analysis, c-MYC expression and immuno blotting, cell cycle analysis and molecular modelling are presented in chapter 4 of this thesis

3.13 References

- (1) Sneppen, K.; Zocchi, G. DNA and RNA. In *Physics in Molecular Biology*; 2010. <https://doi.org/10.1017/cbo9780511755699.005>.
- (2) Huppert, J. L. Hunting G-Quadruplexes. *Biochimie*. 2008. <https://doi.org/10.1016/j.biochi.2008.01.014>.
- (3) Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-Quadruplexes in Gene Promoters: A Novel Anticancer Strategy? *Nat. Rev. Drug Discov.* **2011**, *10* (4), 261–275. <https://doi.org/10.1038/nrd3428>.
- (4) Yuan, L.; Tian, T.; Chen, Y.; Yan, S.; Xing, X.; Zhang, Z.; Zhai, Q.; Xu, L.; Wang, S.; Weng, X.; et al. Existence of G-Quadruplex Structures in Promoter Region of Oncogenes Confirmed by G-Quadruplex DNA Cross-Linking Strategy. *Sci. Rep.* **2013**. <https://doi.org/10.1038/srep01811>.
- (5) 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**. <https://doi.org/10.1021/ja904007p>.
- (6) Sun, D.; Guo, K.; Rusche, J. J.; Hurley, L. H. Facilitation of a Structural Transition in the Polypurine/Polypyrimidine Tract within the Proximal Promoter Region of the Human VEGF Gene by the Presence of Potassium and G-Quadruplex-Interactive Agents. *Nucleic Acids Res.* **2005**. <https://doi.org/10.1093/nar/gki917>.
- (7) Onyshchenko, M. I.; Gaynutdinov, T. I.; Englund, E. A.; Appella, D. H.; Neumann, R. D.; Panyutin, I. G. Stabilization of G-Quadruplex in the BCL2 Promoter Region in Double-Stranded DNA by Invading Short PNAs. *Nucleic Acids Res.* **2009**. <https://doi.org/10.1093/nar/gkp840>.
- (8) Cogoi, S.; Xodo, L. E. G-Quadruplex Formation within the Promoter of the KRAS Proto-Oncogene and Its Effect on Transcription. *Nucleic Acids Res.* **2006**. <https://doi.org/10.1093/nar/gkl286>.
- (9) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription. *Proc. Natl. Acad. Sci.* **2002**. <https://doi.org/10.1073/pnas.182256799>.
- (10) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* **2013**. <https://doi.org/10.1038/nchem.1548>.
- (11) Brooks, T. A.; Hurley, L. H. The Role of Supercoiling in Transcriptional Control of MYC

- and Its Importance in Molecular Therapeutics. *Nature Reviews Cancer*. 2009. <https://doi.org/10.1038/nrc2733>.
- (12) Lemarteleur, T.; Gomez, D.; Paterski, R.; Mandine, E.; Mailliet, P.; Riou, J. F. Stabilization of the C-Myc Gene Promoter Quadruplex by Specific Ligands' Inhibitors of Telomerase. *Biochem. Biophys. Res. Commun.* **2004**. <https://doi.org/10.1016/j.bbrc.2004.08.150>.
- (13) Li, Q.; Xiang, J.; Li, X.; Chen, L.; Xu, X.; Tang, Y.; Zhou, Q.; Li, L.; Zhang, H.; Sun, H.; et al. Stabilizing Parallel G-Quadruplex DNA by a New Class of Ligands: Two Non-Planar Alkaloids through Interaction in Lateral Grooves. *Biochimie* **2009**. <https://doi.org/10.1016/j.biochi.2009.03.007>.
- (14) Diveshkumar, K. V.; Sakrikar, S.; Rosu, F.; Harikrishna, S.; Gabelica, V.; Pradeepkumar, P. I. Specific Stabilization of C-MYC and c-KIT G-Quadruplex DNA Structures by Indolylmethyleneindanone Scaffolds. *Biochemistry* **2016**, *55* (25), 3571–3585. <https://doi.org/10.1021/acs.biochem.6b00120>.
- (15) Chauhan, A.; Paladhi, S.; Debnath, M.; Mandal, S.; Das, R. N.; Bhowmik, S.; Dash, J. A Small Molecule Peptidomimetic That Binds to C-KIT1 G-Quadruplex and Exhibits Antiproliferative Properties in Cancer Cells. *Bioorganic Med. Chem.* **2014**. <https://doi.org/10.1016/j.bmc.2014.05.060>.
- (16) Asamitsu, S.; Obata, S.; Yu, Z.; Bando, T.; Sugiyama, H. Recent Progress of Targeted G-Quadruplex-Preferred Ligands toward Cancer Therapy. *Molecules*. 2019. <https://doi.org/10.3390/molecules24030429>.
- (17) Moore, M. J. B.; Schultes, C. M.; Cuesta, J.; Cuenca, F.; Gunaratnam, M.; Tanious, F. A.; Wilson, W. D.; Neidle, S. Trisubstituted Acridines as G-Quadruplex Telomere Targeting Agents. Effects of Extensions of the 3,6- and 9-Side Chains on Quadruplex Binding, Telomerase Activity, and Cell Proliferation. *J. Med. Chem.* **2006**. <https://doi.org/10.1021/jm050555a>.
- (18) Zeng, D. Y.; Kuang, G. T.; Wang, S. K.; Peng, W.; Lin, S. L.; Zhang, Q.; Su, X. X.; Hu, M. H.; Wang, H.; Tan, J. H.; et al. Discovery of Novel 11-Triazole Substituted Benzofuro[3,2-b]Quinolone Derivatives as c-Myc G-Quadruplex Specific Stabilizers via Click Chemistry. *J. Med. Chem.* **2017**. <https://doi.org/10.1021/acs.jmedchem.7b00016>.
- (19) Nielsen, M. C.; Larsen, A. F.; Abdikadir, F. H.; Ulven, T. Phenanthroline-2,9-Bistriazoles as Selective G-Quadruplex Ligands. *Eur. J. Med. Chem.* **2014**. <https://doi.org/10.1016/j.ejmech.2013.11.027>.
- (20) Medeiros-Silva, J.; Guédin, A.; Salgado, G. F.; Mergny, J. L.; Queiroz, J. A.; Cabrita, E. J.; Cruz, C. Phenanthroline-Bis-Oxazole Ligands for Binding and Stabilization of G-Quadruplexes. *Biochim. Biophys. Acta - Gen. Subj.* **2017**.

- <https://doi.org/10.1016/j.bbagen.2016.11.024>.
- (21) Ritson, D. J.; Moses, J. E. A Fragment Based Click Chemistry Approach towards Hybrid G-Quadruplex Ligands: Design, Synthesis and Biophysical Evaluation. *Tetrahedron* **2012**. <https://doi.org/10.1016/j.tet.2011.10.066>.
- (22) Jain, A. K.; Reddy, V. V.; Paul, A.; Muniyappa, K.; Bhattacharya, S. Synthesis and Evaluation of a Novel Class of G-Quadruplex-Stabilizing Small Molecules Based on the 1,3-Phenylene-Bis(Piperazinyl Benzimidazole) System. *Biochemistry* **2009**. <https://doi.org/10.1021/bi9003815>.
- (23) Ali, A.; Bansal, M.; Bhattacharya, S. Ligand 5,10,15,20-Tetra(N-Methyl-4-Pyridyl)Porphine (TMPyP4) Prefers the Parallel Propeller-Type Human Telomeric G-Quadruplex DNA over Its Other Polymorphs. *J. Phys. Chem. B* **2015**. <https://doi.org/10.1021/jp505792z>.
- (24) Droumaguet, L.; Wang, C.; Hanni, Q.; Leigh, K. D.; Hein, D. A.; Fokin, J. E.; Holub, V. V.; Kirshenbaum, J. M.; Meldal, K.; Tornøe, M.; et al. (6) For Selected References to CuAAC Applications in 2010, See: (A) Sharma, P.; Moses. *J. Am. Chem. Soc* **2010**. <https://doi.org/10.1021/jo101495k>.
- (25) Xu, Y.; Suzuki, Y.; Komiyama, M. Click Chemistry for the Identification of G-Quadruplex Structures: Discovery of a DNA-RNA g-Quadruplex. *Angew. Chemie - Int. Ed.* **2009**. <https://doi.org/10.1002/anie.200806306>.
- (26) Saha, P.; Panda, D.; Dash, J. The Application of Click Chemistry for Targeting Quadruplex Nucleic Acids. *Chem. Commun.* **2019**, 55 (6), 731–750. <https://doi.org/10.1039/C8CC07107A>.
- (27) Angell, Y. L.; Burgess, K. Peptidomimetics via Copper-Catalyzed Azide-Alkyne Cycloadditions. *Chemical Society Reviews*. 2007. <https://doi.org/10.1039/b701444a>.
- (28) Langmuir, I. Isomorphism, Isosterism and Covalence. *J. Am. Chem. Soc.* **1919**. <https://doi.org/10.1021/ja02231a009>.
- (29) Blass, B. E. *Basic Principles of Drug Discovery and Development*; 2015. <https://doi.org/10.1016/C2012-0-06670-7>.
- (30) Friedman, H. Influence of Isosteric Replacements upon Biological Activity. *Symp. Chem. Correl.* **1951**.
- (31) Thornber, C. W. Isosterism and Molecular Modification in Drug Design. *Chemical Society Reviews*. 1979. <https://doi.org/10.1039/CS9790800563>.
- (32) Lima, L.; Barreiro, E. Bioisosterism: A Useful Strategy for Molecular Modification and Drug Design. *Curr. Med. Chem.* **2012**. <https://doi.org/10.2174/0929867053363540>.
- (33) Patani, G. A.; LaVoie, E. J. Bioisosterism: A Rational Approach in Drug Design. *Chem. Rev.*

- 1996.** <https://doi.org/10.1021/cr950066q>.
- (34) Carini, D. J.; Duncia, J. V.; Aldrich, P. E.; Chiu, A. T.; Johnson, A. L.; Pierce, M. E.; Price, W. A.; Santella, J. B.; Wells, G. J.; Wexler, R. R.; et al. Nonpeptide Angiotensin II Receptor Antagonists: The Discovery of a Series of N-(Biphenylmethyl)Imidazoles as Potent, Orally Active Antihypertensives. *J. Med. Chem.* **1991**. <https://doi.org/10.1021/jm00112a031>.
- (35) Hamada, Y.; Kiso, Y. The Application of Bioisosteres in Drug Design for Novel Drug Discovery: Focusing on Acid Protease Inhibitors. *Expert Opinion on Drug Discovery*. 2012. <https://doi.org/10.1517/17460441.2012.712513>.
- (36) Purcell, W. P.; Singer, J. A. Electronic and Molecular Structure of Selected Unsubstituted and Dimethyl Amides from Measurements of Electric Moments and Nuclear Magnetic Resonance. *J. Phys. Chem.* **1967**. <https://doi.org/10.1021/j100872a024>.
- (37) Chemama, M.; Fonvielle, M.; Arthur, M.; Valéry, J. M.; Etheve-Quellejeu, M. Synthesis of Stable Aminoacyl-TRNA Analogues Containing Triazole as a Bioisoster of Esters. *Chem. - A Eur. J.* **2009**. <https://doi.org/10.1002/chem.200801563>.
- (38) Dijkgraaf, I.; Rijnders, A. Y.; Soede, A.; Dechesne, A. C.; Van Esse, G. W.; Brouwer, A. J.; Corstens, F. H. M.; Boerman, O. C.; Rijkers, D. T. S.; Liskamp, R. M. J. Synthesis of DOTA-Conjugated Multivalent Cyclic-RGD Peptide Dendrimers via 1,3-Dipolar Cycloaddition and Their Biological Evaluation: Implications for Tumor Targeting and Tumor Imaging Purposes. *Org. Biomol. Chem.* **2007**. <https://doi.org/10.1039/b615940k>.
- (39) Vatmurge, N. S.; Hazra, B. G.; Pore, V. S.; Shirazi, F.; Chavan, P. S.; Deshpande, M. V. Synthesis and Antimicrobial Activity of β -Lactam-Bile Acid Conjugates Linked via Triazole. *Bioorganic Med. Chem. Lett.* **2008**. <https://doi.org/10.1016/j.bmcl.2008.01.102>.
- (40) Vatmurge, N. S.; Hazra, B. G.; Pore, V. S.; Shirazi, F.; Deshpande, M. V.; Kadreppa, S.; Chattopadhyay, S.; Gonnade, R. G. Synthesis and Biological Evaluation of Bile Acid Dimers Linked with 1,2,3-Triazole and Bis- β -Lactam. *Org. Biomol. Chem.* **2008**. <https://doi.org/10.1039/b809221d>.
- (41) Kategaonkar, A. H.; Shinde, P. V.; Kategaonkar, A. H.; Pasale, S. K.; Shingate, B. B.; Shingare, M. S. Synthesis and Biological Evaluation of New 2-Chloro-3-((4-Phenyl-1H-1,2,3-Triazol-1-yl)methyl)Quinoline Derivatives via Click Chemistry Approach. *Eur. J. Med. Chem.* **2010**. <https://doi.org/10.1016/j.ejmech.2010.04.002>.
- (42) Somu, R. V.; Boshoff, H.; Qiao, C.; Bennett, E. M.; Barry, C. E.; Aldrich, C. C. Rationally-Designed Nucleoside Antibiotics That Inhibit Siderophore Biosynthesis of Mycobacterium Tuberculosis. *J. Med. Chem.* **2006**.

- <https://doi.org/10.1021/jm051060o>.
- (43) Costa, M. S.; Boechat, N.; Rangel, É. A.; da Silva, F. de C.; de Souza, A. M. T.; Rodrigues, C. R.; Castro, H. C.; Junior, I. N.; Lourenço, M. C. S.; Wardell, S. M. S. V.; et al. Synthesis, Tuberculosis Inhibitory Activity, and SAR Study of N-Substituted-Phenyl-1,2,3-Triazole Derivatives. *Bioorganic Med. Chem.* **2006**. <https://doi.org/10.1016/j.bmc.2006.08.019>.
- (44) Whiting, M.; Tripp, J. C.; Lin, Y. C.; Lindstrom, W.; Olson, A. J.; Elder, J. H.; Sharpless, K. B.; Fokin, V. V. Rapid Discovery and Structure-Activity Profiling of Novel Inhibitors of Human Immunodeficiency Virus Type 1 Protease Enabled by the Copper(I)-Catalyzed Synthesis of 1,2,3-Triazoles and Their Further Functionalization. *J. Med. Chem.* **2006**. <https://doi.org/10.1021/jm060754+>.
- (45) Saito, Y.; Escuret, V.; Durantel, D.; Zoulim, F.; Schinazi, R. F.; Agrofoglio, L. A. Synthesis of 1,2,3-Triazolo-Carbanucleoside Analogues of Ribavirin Targeting an HCV in Replicon. *Bioorganic Med. Chem.* **2003**. [https://doi.org/10.1016/S0968-0896\(03\)00349-3](https://doi.org/10.1016/S0968-0896(03)00349-3).
- (46) 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. *Medicinal Research Reviews.* 2008. <https://doi.org/10.1002/med.20107>.
- (47) Palmer, M. H.; Findlay, R. H.; Gaskell, A. J. Electronic Charge Distribution and Moments of Five- and Six-Membered Heterocycles. *J. Chem. Soc. Perkin Trans. 2* **1974**. <https://doi.org/10.1039/p29740000420>.
- (48) Li, F.; Frett, B.; Li, H. Y. Selective Reduction of Halogenated Nitroarenes with Hydrazine Hydrate in the Presence of Pd/C. *Synlett* **2014**. <https://doi.org/10.1055/s-0033-1339025>.
- (49) Jarrahpour, A.; Zarei, M. The Vilsmeier Reagent: A Useful and Versatile Reagent for the Synthesis of 2-Azetidinones. *Tetrahedron* **2009**. <https://doi.org/10.1016/j.tet.2009.02.005>.
- (50) Hanna-Elias, A.; Manallack, D. T.; Berque-Bestel, I.; Irving, H. R.; Coupar, I. M.; Iskander, M. N. Synthesis and Preliminary Screening of Novel Indole-3-Methanamines as 5-HT₄ Receptor Ligands. *Eur. J. Med. Chem.* **2009**. <https://doi.org/10.1016/j.ejmech.2009.01.015>.
- (51) Nasiri, H. R.; Bell, N. M.; McLuckie, K. I. E.; Husby, J.; Abell, C.; Neidle, S.; Balasubramanian, S. Targeting a C-MYC G-Quadruplex DNA with a Fragment Library. *Chem. Commun.* **2014**. <https://doi.org/10.1039/c3cc48390h>.

- (52) Dai, J.; Carver, M.; Hurley, L. H.; Yang, D. Solution Structure of a 2:1 Quindoline-c-MYC G-Quadruplex: Insights into G-Quadruplex-Interactive Small Molecule Drug Design. *J. Am. Chem. Soc.* **2011**. <https://doi.org/10.1021/ja205646q>.
- (53) Forli, W.; Halliday, S.; Belew, R.; Olson, A. AutoDock Version 4.2. *Citeseer* **2012**.
- (54) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. Gaussian09. *Gaussian 09*. 2009.
- (55) Das, T.; Panda, D.; Saha, P.; Dash, J. Small Molecule Driven Stabilization of Promoter G-Quadruplexes and Transcriptional Regulation of c-MYC. *Bioconjug. Chem.* **2018**. <https://doi.org/10.1021/acs.bioconjchem.8b00338>.
- (56) O'Brien, J.; Wilson, I.; Orton, T.; Pognan, F. Investigation of the Alamar Blue (Resazurin) Fluorescent Dye for the Assessment of Mammalian Cell Cytotoxicity. *Eur. J. Biochem.* **2000**. <https://doi.org/10.1046/j.1432-1327.2000.01606.x>.

Chapter 4

Novel Azaindolin-2-one derived *c-MYC* Promoter G-Quadruplex Stabilizers

Chapter 4 Novel Azaindolin-2-one derived c-MYC Promoter G-Quadruplex Stabilizers

4.1 Introduction

Guanine (G)-rich DNA sequences can form four-stranded structures that are called G-quadruplexes (G4s). These non-canonical DNA structures are over-represented in key regulatory regions within the human genome.¹ The occurrence of G4s in gene regulatory regions have made them an attractive target for drug development to trap quadruplexes in cellular DNA, downregulate oncogene transcription, promote telomere disruption, and induce growth arrest in cancer cells.² G4 have significant biological relevance in the regulation of human genes.³⁻⁵ They are found particularly in the promoter regions of oncogenes including *c-MYC*, *c-KIT*, *BCL-2*, and notably in telomeres.^{6,7}

Recent research revealed that numerous potential G4-forming sequences exist in the human genome. They can act as a target for a range of small drug-like molecules but may cause issues with a potential drug selectivity problem.⁴ It has been proposed that the G4 present in NHE III₁ in the *c-MYC* gene is crucial for transcriptional silencing^{8,9} and G4-stabilizing ligands can potentially down-regulate *c-MYC* transcription.⁸ Consequently, the c-MYC protein-dependent proliferation was inhibited, leading to inhibition of cancer cell growth.¹⁰ The design of G4 binding small molecule scaffolds has been predominately based on the fundamental principles of having a heteropolyaromatic core with a planar conformation to enable binding in G4 grooves and to loops. Recently, from molecular dynamics (MD) simulation, researchers have suggested that fused aromatic ligand stacking on the G-quartet has a positive impact on G-quartet stabilization, and helps the overall G4 stabilization.¹¹ On the other hand, crystallographic, NMR structures and MD studies of G4 have showed that the G-quartet surface is not completely coplanar and are highly dynamic in nature.^{11,12} Based on these considerations, unfused aromatic ligands with adaptive structural feature may be more suitable to be accommodated on the dynamic G-quartet surface rather than polycyclic compounds with rigid aromatic rings. Furthermore, this adaptive structural feature can prevent ligand from intercalating into the duplex DNA, and is consequently important for its strong quadruplex selectivity.

However, in contrast to fused aromatic compounds, much less is known about how planarity of an unfused aromatic ligand influence its G-quadruplex binding properties. In this context natural products are a prime source of innovative molecular fragments and are privileged scaffolds for drug discovery and development.¹³ Natural alkaloid Schizocommunin¹⁴ (Figure 68) comprises a conjugated system of quinazolinone and isatin moieties^{15,16}, fused indolylmethyleneindanone¹⁷ and SYUIQ-5¹⁸ have shown significant G4 stabilization potential, they tend to selectively bind to

various h-TELO, *c-MYC* and *c-KIT* promoter G4s. It is hypothesized that higher selectivity for G4 over duplex DNA can be achieved by extending the heteropolyaromatic scaffold to increase π - π stacking overlap with the larger surface area of a terminal G-quartet.¹⁹ The crescent-shaped, large, nearly planar chromophore led us to explore a novel scaffold for the design of selective G-quadruplex ligands.

To explore new and selective G-quadruplex ligands for cancer chemotherapy,²⁰ we designed and synthesized a series of schizocommunin derivatives by attaching cationic amino side chains and introducing a fluorine atom into the aromatic chromophore. Fluorine atoms, with high electronegativity and small size, often exhibit unique properties in functional molecules.²¹ The electron-withdrawing effect of fluorine could reduce the electron density of the aromatic chromophore, which might favor a stronger interaction with the electron-rich π -system of the G-quartet.²² In addition, the introduction of fluorine atoms into small molecules might improve liposolubility and bioavailability. The structure-activity relationships and activity mechanism were also investigated. The results revealed that derivatives could selectively stabilize and bind to the telomeric G-quadruplex in vitro and in cells.

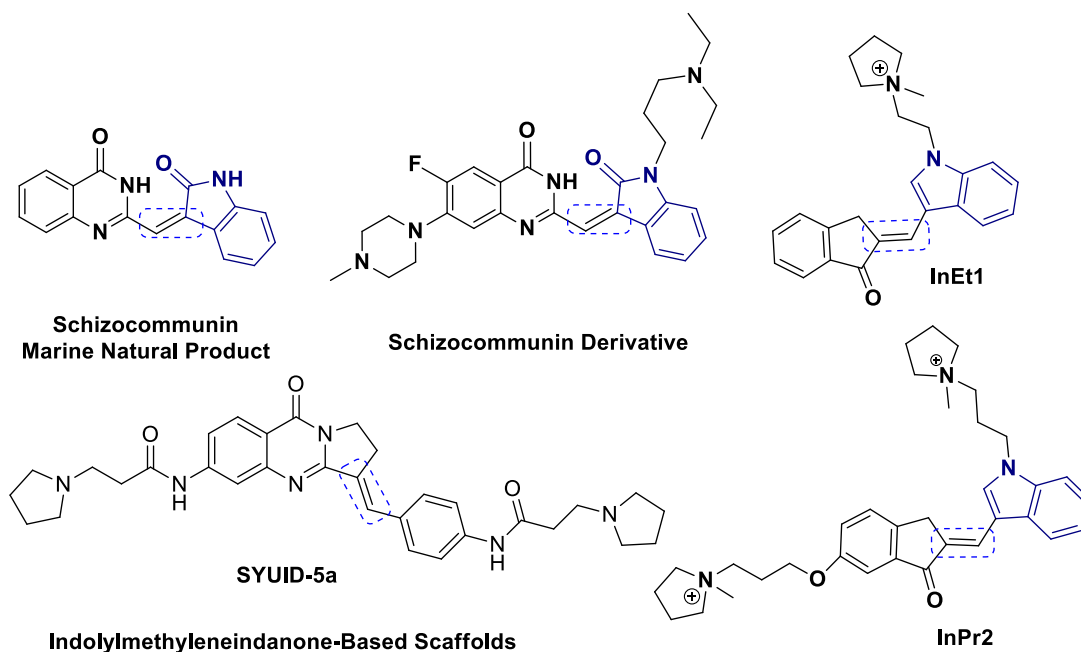


Figure 70. Structures of schizocommunin and indolymethyleneindanone based scaffolds reported in literature to achieve topology specific stabilization of *c-MYC*, *c-KIT* promoter G-Quadruplexes and *h-TELO* G-quadruplex binding and stabilizing ligands.

4.2 Background and applications of azaindole

Azaindoles, also named as pyrrolopyridines, represent a class of condensed hetero-bicycles that are described as valuable targets in drug discovery as indole bioisosteres. Replacing one of the carbon atoms at positions 4-7 in the indole template with a nitrogen atom gives the so-called 4-, 5-, 6-, or 7-azaindoles, respectively, as shown in figure 71.



Figure 71. Indole and azaindole frameworks.

Structurally, azaindoles consist of two fused heterocycles with opposite electronic features, one electron-rich pyrrole and the other electron-deficient pyridine rings, on one heteroaromatic skeleton.

4.2.1 Synthesis methods towards Azaindole

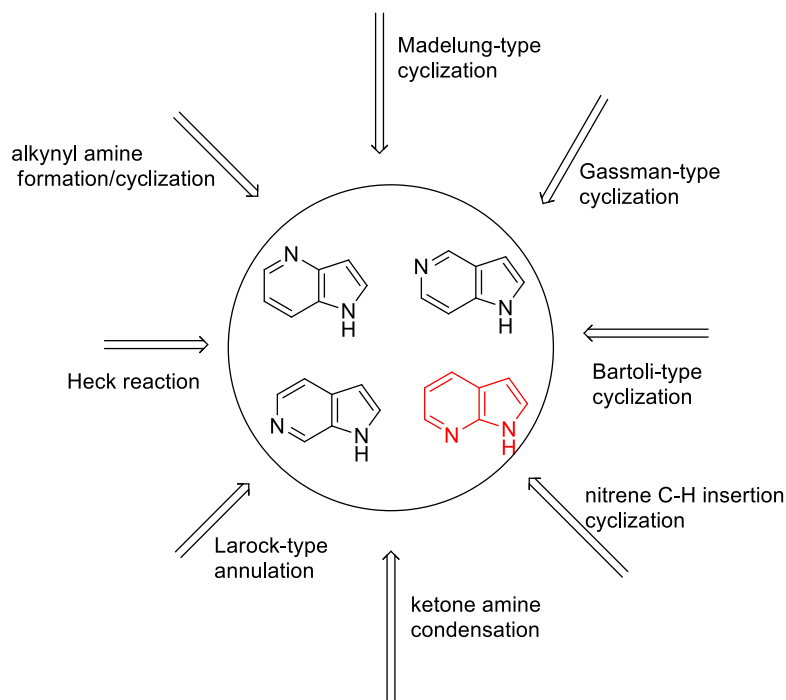


Figure 72. Representative synthesis methods toward azaindoles.

The conventional synthetic methods towards azaindole are Madelung-type cyclization,²³ Gassman-type cyclization,²⁴ Bartoli-type cyclization,²⁵ intramolecular nitrene C-H insertion cyclization,²⁶ ketone amine condensation,²⁷ electrophilic cyclization *via* a Pictet-Spengler reaction,²⁸ and Friedel-Crafts type cyclization followed by dehydration.²⁹ These traditional methods suffer from poor yields, limited reaction scope, and harsh reaction conditions; hence, very few functional groups are compatible.²³ More recently, Pd-catalyzed alkynyl amine formation/cyclization,³⁰ Heck reaction,³¹ Larock-type annulation,³² Ar-Pd-X-mediated cyclization of alkynyl amines,³³ ring-opening of a spiro pyridone-cyclopropane followed by cross-couplings,³⁴ and a double Buchwald-Hartwig C-N coupling³⁵ overcome the drawbacks of traditional methods owing to the mild reaction conditions and high functional group compatibility of Pd chemistry (Fig. 72). However, they still suffer from low yields and the limited availability of the appropriate starting materials. Here, we have developed a 4 steps method to synthesize 7-azaindole in high yield (80% -85%). Compared with previous methods, the reaction conditions of our method are mild, and the reaction steps are less, which is very important for mass production in the future.

Azaheterocyclic compounds contain nitrogen atom in the ring. They are of vital importance in the race to improve our understanding of basic chemistry, which underlies nearly all of the important life-processes and a large proportion of transformations leading to and creating the increasingly sophisticated products, which enhance our society today. A large numbers of Azaheterocyclic compounds are well known and this number is increasing rapidly because they hold a special place among pharmaceutically significant natural products and synthetic compounds. The chemistry of azaheterocyclic compounds is as logical as that of aliphatic or aromatic in character, depending on their electronic constitution. Their study is of great interest both from the theoretical as well as practical standpoint. Azaheterocyclic compounds are very widely distributed in nature and are essential to life in various ways. Compounds such as alkaloids, antibiotics, essential amino acids, vitamins, haemoglobin, hormones and a large number of synthetic drugs and dyes contain azaheterocyclic ring systems. Knowledge of azaheterocyclic chemistry is useful in biosynthesis as well as in drug metabolism. There are also a large number of azaheterocyclic compounds with other important practical applications as antioxidants, vulcanization accelerators, copolymers, solvents photographic sensitizer and developers, dyestuffs and many are valuable intermediates in synthesis. Azaheterocycles are omnipresent extremely in all branches of chemistry and biochemistry as well as in our lives.

4.2.2 Anticancer activity

The uncontrolled growth of cells in the body, started due to certain stimuli, lays the foundation of cancer, anticancer drugs either kill cancer cell or modify their growth. Cancer or neoplastic disease may be regarded as a family of related disorders. A common feature in different forms of cancer is an abnormal and uncontrolled cell division, frequently at a rate greater than that of most normal body cells. Among the heterocyclic compounds, five member heterocyclic moieties fused with aromatic ring system with nitrogen atom possess wide spectrum of pharmacological activity. Heterocycles like indole, azaindoles, pyrimidine, pyridine, quinoline etc. are an integral part of huge number of natural and synthetic compounds and play important roles in the biological system. For developing the suitable leads for anticancer drugs introduction of appropriate substituent's at C-3 of indole, C-5 of pyrimidine, C-2 of pyridine and quinoline is required.

4.2.3 7-Azaindole – Promising building block for anticancer therapy

7-Azaindole (Fig. 73) got an increasing interest in the last years for research, because of its biological effects. Azaindole derivatives play a huge role as Kinase inhibitors.³⁶ Azaindole derivatives were observed to have an effect on cancer cell cycles. So they could probably be used as anti-cancer drug where these derivatives can stop the abnormal growth of such cells. With all these expected effects, this scaffold was chosen to be the starting material for synthesizing G-quadruplex stabilizing ligands.

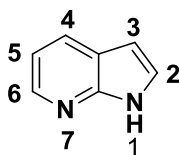


Figure 73. The chemical structure of 7-azaindole, which was used as scaffold in various reactions to obtain an anti-cancer, drug.

4.2.4 Importance of Azaindolin-2-one heterocycle as Cancer therapeutics

Sunitinib is a new multitargeted oral anti-angiogenic and antitumor drug that has been recently approved against gastrointestinal stromal tumors (GIST) and advanced renal cell carcinoma (RCC).³⁷ It is in clinical studies for the treatment of other solid tumors, such as pancreatic neuroendocrine tumors,³⁸ meningioma,³⁹ metastatic breast cancer⁴⁰ and non-small cell lung cancer.⁴¹ Recently, structural modifications mainly at the 3- and 5-positions of the indolin-2-one

ring of Sunitinib have made considerable progress in the ability to increase antitumor activity through inhibition on different receptors.⁴² As early lead compounds discovered, Z24 and LK-B030 (Fig. 74) bearing a (piperidin-1-yl)methyl and a (3-dimethylamino)propyl group at the N-1 position, respectively, display a broad spectrum of antitumor activity by inhibiting angiogenesis in new blood vessels.⁴³ More recently, a series of novel 5-halogenated-7-azaindolin-2-one derivatives have been reported and amongst these compounds IMB-1501 to have better in vitro activity than Sunitinib against the entire tested cancer cell lines.⁴⁴ Taken in to consideration the importance of 7-azaindolin-2-one as a potential platform for further modification and generation of G-quadruplex stabilizing ligands, we planned to explore other possibilities for diversification of the 7-azaindolin-2-one group with efficient Knoevenagel condensation to introduce flexible side-chain with several heterocycles. Thus, a series of novel substituted 7-azaindolin-2-one derivatives were designed, synthesized and evaluated for their antitumor activity in this study. Our primary objective was to optimize the potency of these compounds against a set of solid tumors and contribute to the development of novel c-MYC promoter G-quadruplex stabilizing agents to downregulate c-myc oncogene at transcriptional and translational level.

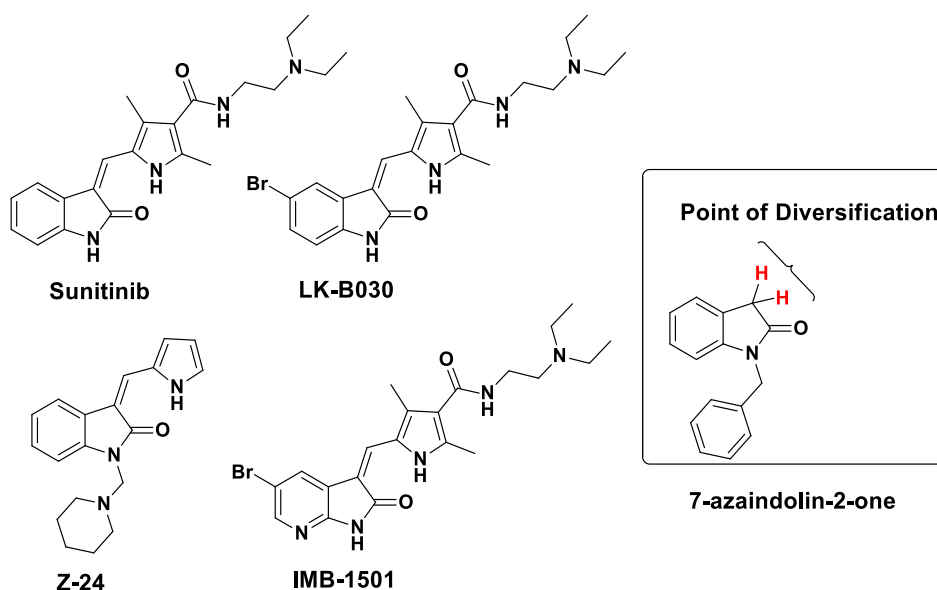


Figure 74. Structures of Sunitinib, Z-24, LK-B030 and IMB-1501 and an azaindole derivative with its acidic protons (in red) with the carbonyl as neighbour.

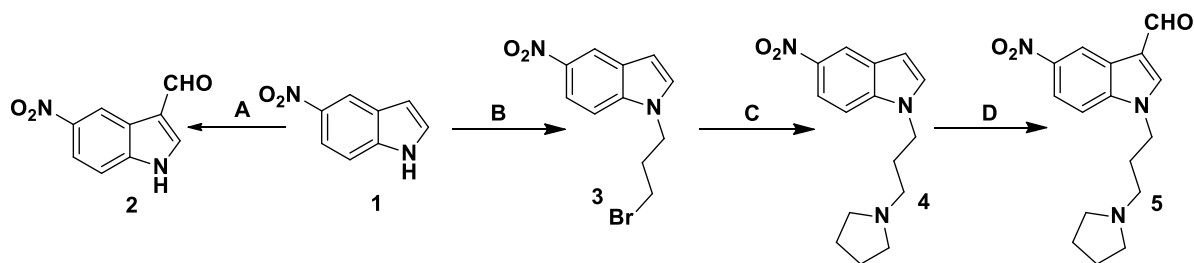
4.3 Project objective

The main objective of this chapter is to generate potent anticancer agents based on Azaindolin-2-one *via* Knoevenagel condensation reaction. Azaindolin-2-one scaffold is unexplored and it can bind effectively to G4 tetrads and gives us the platform for its derivatization to generate library of azaindolin-2-one conjugates to evaluate their binding affinity with *c*-MYC promoter G-quadruplex sequences.

4.4 Results and Discussions

4.5 Design and Synthesis

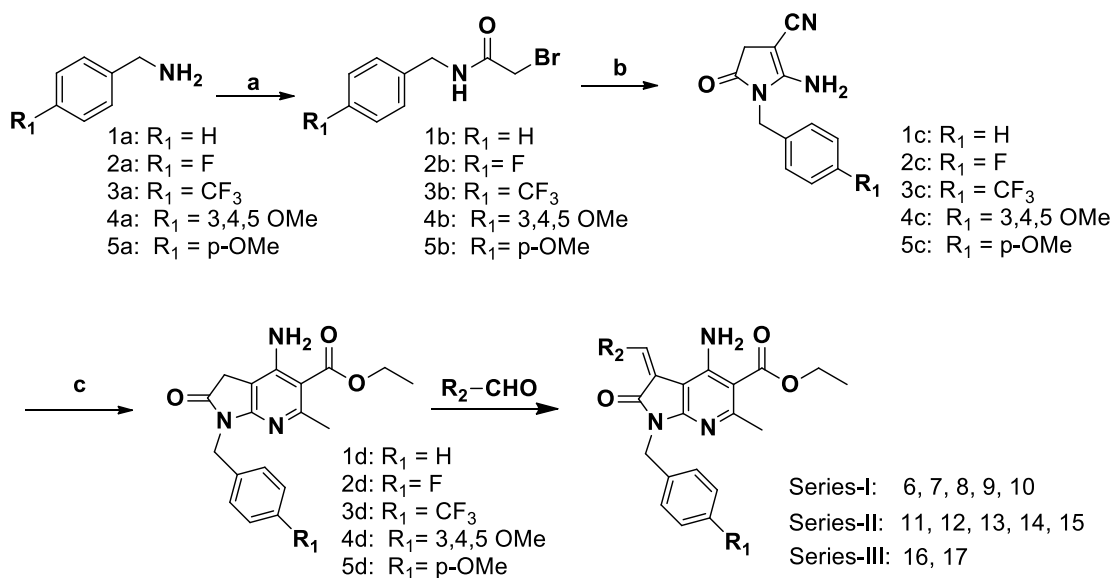
In scheme I, intermediate 5-nitro-1H-indole-3-carbaldehyde (**2**) is obtained in 60 % yield *via* Vilsmeier-Haack reaction⁴⁵ with 5-nitro-1H-indole (**1**) which is a key step to generate conjugate **11** in quantitative yields with one pot *in-situ* Knoevenagel condensation reaction of aldehydes with substituted azaindolin-2-one in presence of piperidine as a reducing agent.¹⁷ Intermediates 1-(3-bromopropyl)-5-nitro-1H-indole (**3**) are generated *via* a nucleophilic substitution reaction of 5-nitro-1H-indole (**1**) with 1,3-dibromopropane.⁴⁶ In the next step, 1-(3-bromopropyl)-5-nitro-1H-indole (**3**) is refluxed with pyrrolidine to generate the 5-nitro-1-(3-(pyrrolidin-1-yl)propyl)-1H-indole (**4**).⁴⁷ To afford 5-nitro-1-(3-(pyrrolidin-1-yl)propyl)-1H-indole-3-carbaldehyde (**6**)⁴⁵, intermediate 5-nitro-1-(3-(pyrrolidin-1-yl)propyl)-1H-indole (**5**) was treated with the Vilsmeier reagent.⁴⁵



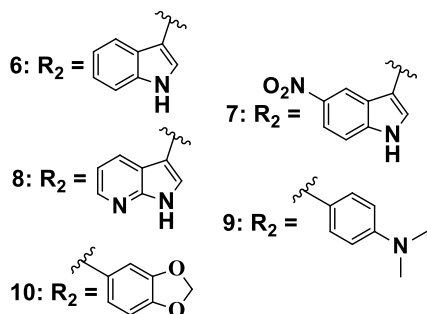
Scheme I. A) POCl₃, Dry DMF, 0 °C - r.t, 0.5 h, 65 %; B) K₂CO₃, DMF, 1,3-Dibromopropane, r.t, 4 h, 65 %; C) Dry ACN, pyrrolidine, reflux, 4 h, 32%; D) POCl₃, DMF, 0 °C - r.t, 1 h, 46%

In scheme II, various substituted *N*-alkyl- α -bromoacetamide (**1b-5b**) were generated from the precursor's **1a-5a** *via* nucleophilic substitution reaction with 2-bromoacetyl bromide. These intermediates (**1b-5b**) were then allowed to react with malononitrile in the presence of sodium

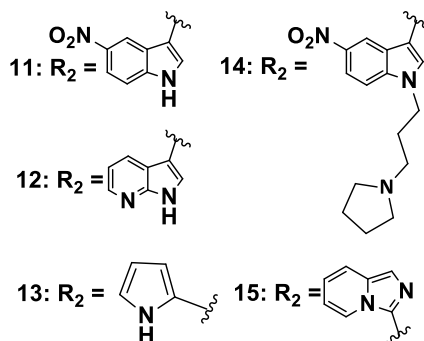
hydride (NaH) as a base in tetrahydrofuran (THF) leads to the formation of enamino γ -lactams (**1c-5c**) in excellent yields of 85-90%.⁴⁸ Intermediates (**1c-5c**) were then allowed to undergo lewis acid promoted Friedlander condensation reactions with ethyl 3-oxobutanoate in presence of 99.9% AlCl₃ (lewis acid) to give our desired substituted 2-oxo-2, 3-dihydro-1H-pyrrolo [2, 3-b] pyridine-5-carboxylate (**1d-5d**) in 48-52 % yield.⁴⁹ These intermediates (**1d-5d**) were then allowed to undergo base-catalyzed Knoevenagel condensation reaction with the various aldehydes to afford our desired substituted azaindolin-2-one derivatives (**6-17**) reported in scheme II.^{50,51}



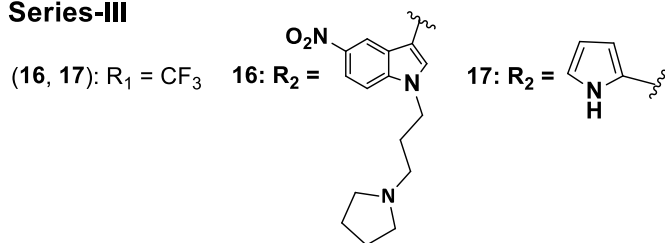
Series-I (6, 7, 8, 9, 10): R₁ = H



Series-II (11,12, 13, 14, 15, 22): R₁ = F



Series-III



Scheme II. A) 2-bromoacetyl bromide, DCM, 0 °C - r.t, 1 h, 95 %; B) NaH, THF, 0 °C - r.t, 52%; C) ethyl 3-oxobutanoate, AlCl₃, DCE, 110 °C, reflux, 4 h, 48 -52 %; D) catalytic piperidine, MeOH, reflux, 100 °C, 42-51%

4.6 Reaction mechanism of important transformations: Scheme II

4.6.1 Nucleophilic addition step to reach intermediates 1-5b

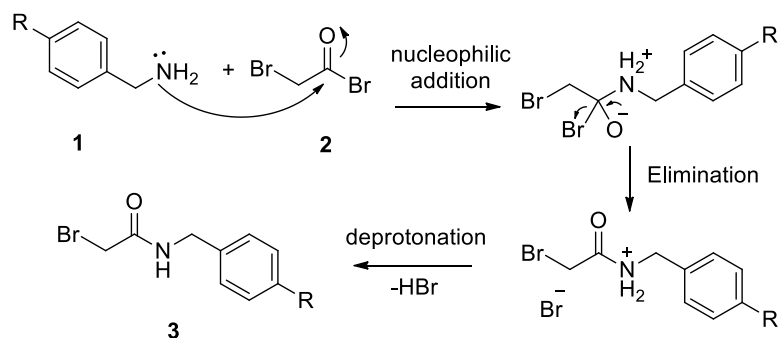


Figure 75. Representation of mechanism.

4.6.2 From substituted-2-bromoacetamine 1-5b to lactams 1-5c

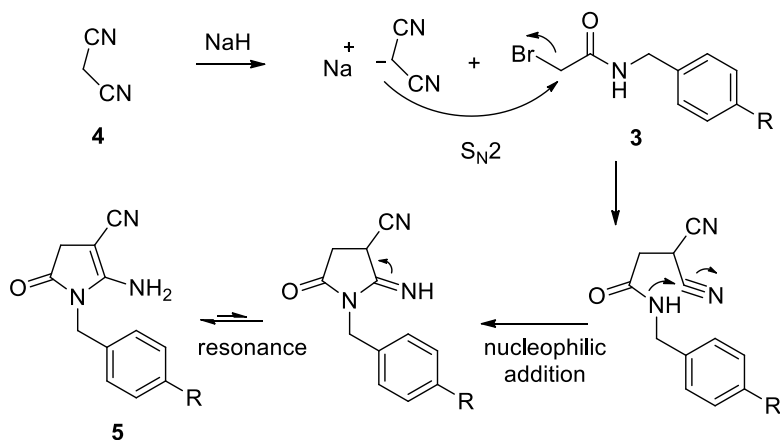


Figure 76. Representation of mechanism.

4.6.3 Friedlander reaction mechanism from 5 to 7

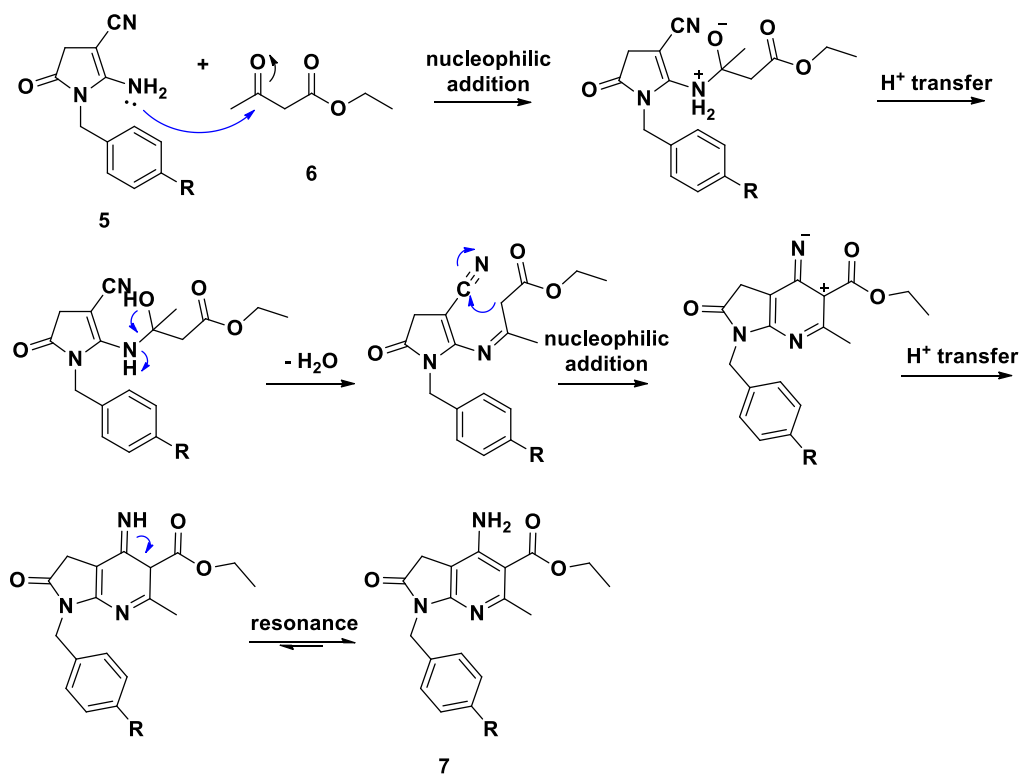
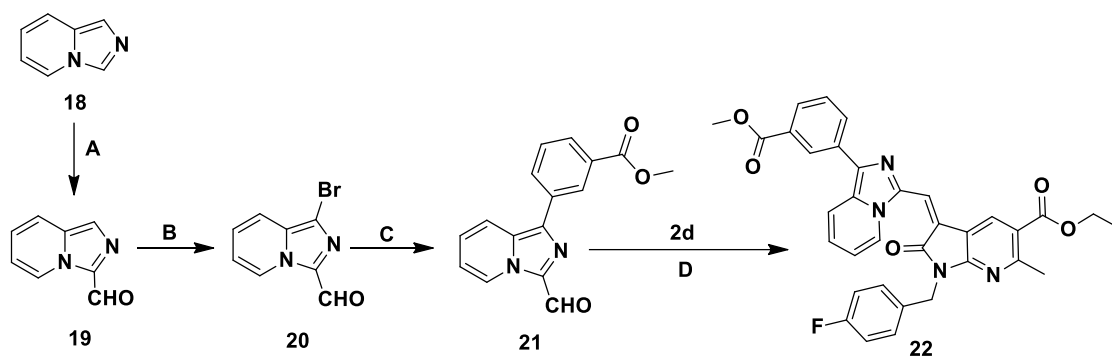


Figure 77. Representation of mechanism.

Scheme III: To afford substituted azaindolin-2-one derivative (**22**), Intermediate imidazo[1,5-a]pyridine-3-carbaldehyde (**20**) is obtained in 55 % yield *via* Vilsmeier-Haack reaction⁴⁵ with imidazo[1,5-a]pyridine (**18**). Followed by bromination of imidazo [1, 5-a] pyridine-3-carbaldehyde (**19**) to get 1-bromoimidazo [1, 5-a] pyridine-3-carbaldehyde (**20**) in 48 % yield *via* electrophilic addition reaction with *N*-bromosuccinimide (NBS) in dimethylformamide (DMF) medium. Intermediate 1-bromoimidazo[1,5-a]pyridine-3-carbaldehyde (**20**) was then treated with substituted boronic acid for Suzuki coupling reaction in presence of tetrakis(triphenylphosphine)palladium(0) as a catalyst and cesium carbonate (Cs₂CO₃) base under N₂-atm in dry DMF to afford methyl 3-(3-formylimidazo[1,5-a]pyridin-1-yl)benzoate (**21**) in 35 % yield. The substituted azaindolin-2-one (**22**) was synthesised 42 % yield by base-catalyzed Knoevenagel condensation reaction of methyl 3-(3-formylimidazo[1,5-a]pyridin-1-yl)benzoate (**21**) and ethyl 4-amino-1-(4-fluorobenzyl)-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (**2d**).



Scheme III. A) POCl_3 , dry DMF, 0 °C - r.t, 0.5 h, 55 %; B) NBS, dry DMF, N_2 -atm, r.t, 3-4 h, 48 %; C) CsCO_3 , $\text{Pd}(\text{PPh}_3)_4$, N_2 -atm, dry DMF, r.t, 3-4 h, 35 %; D) catalytic piperidine, MeOH, reflux, 100 °C, 40%

4.6.4 Vilsmeier-Haak-Reaction for compound 18 to reach compound 19: Scheme III

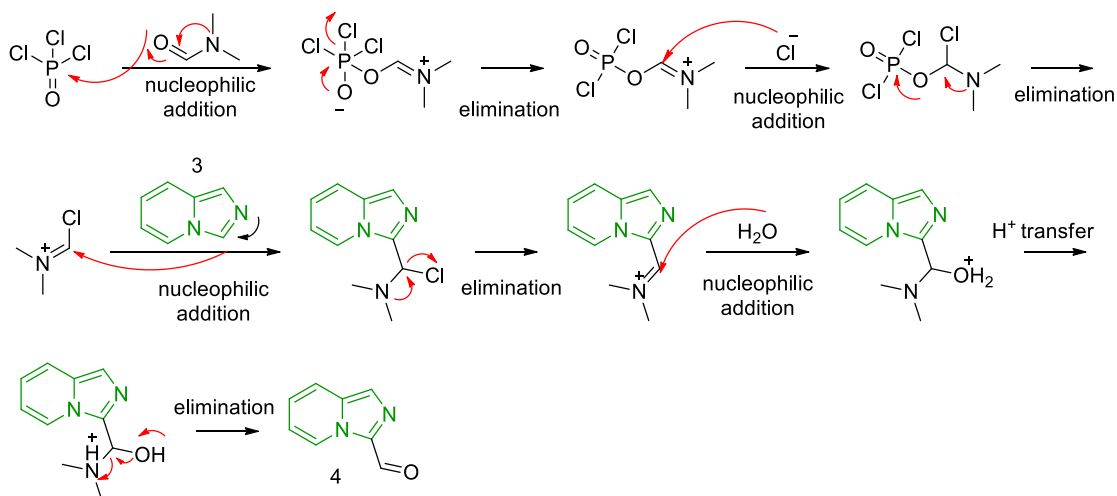


Figure 78. Representation of mechanism

4.6.5 Bromination reaction mechanism for the conversion of 19 to 20

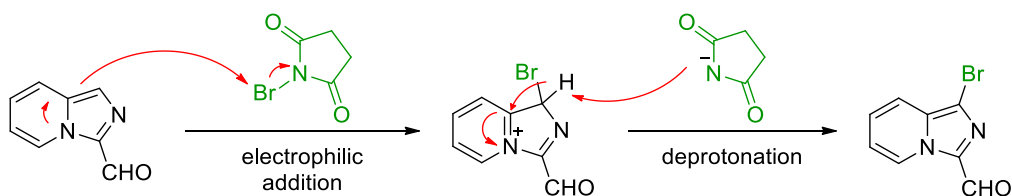


Figure 79. Representation of mechanism.

4.6.6 Knoevenagel condensation mechanism for scheme II and scheme III

Enamine alkylation step to reach conjugate 15 is depicted in figure 80.

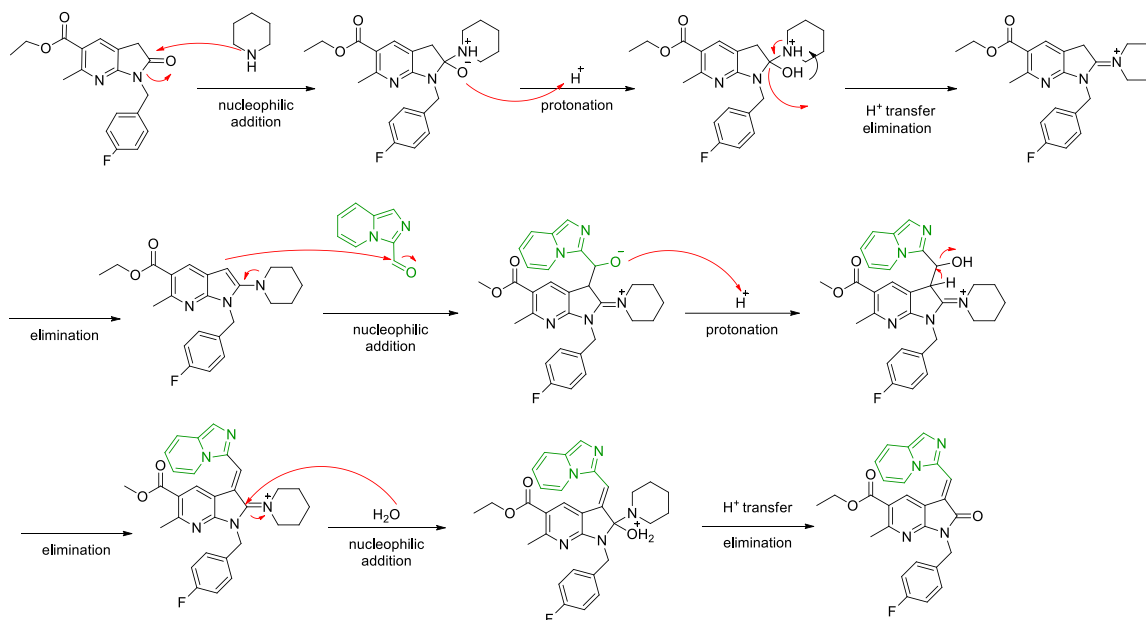


Figure 80. Representation of mechanism.

4.6.7 Palladium (0) catalyzed Suzuki cross coupling reaction mechanism for the conversion of 20 to 21

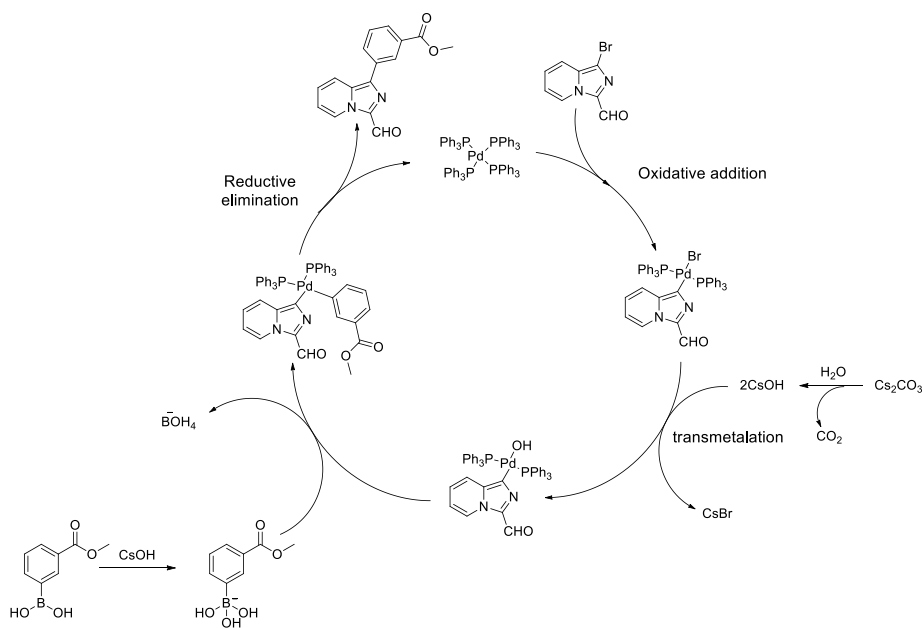
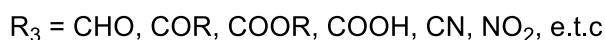
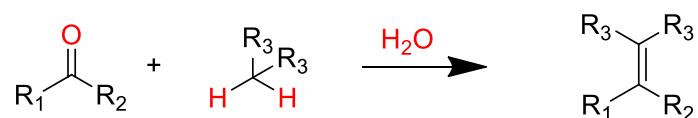


Figure 81. Representation of mechanism.

4.7 Knoevenagel condensation: *E* and *Z* isomer

4.7.1 Knoevenagel Condensation reaction

The Knoevenagel condensation reaction is analogous to the aldol reaction. Under the effect of a weak base catalyst, a dehydration condensation of a compound containing active methylene with aldehyde or ketone can occur, α , β -unsaturated carbonyl compounds and their analogs are thus formed (scheme II and III). Depending on the type of basic catalyst used, there are two possible mechanisms. One is that the condensation of aldehydes or ketones with amines to form imines which added then with carbanions. Another mechanism (Hann-Lapworth mechanism) is under the action of alkali, diethyl malonate and aldol aldehyde through addition reaction to form a, β -hydroxy dicarbonyl compound as the intermediate, and then water is eliminated to obtain the final condensation product as depicted in scheme IV.



Scheme IV: Knoevenagel condensation reaction.

For the following reactions, reflux conditions were chosen, M. The configuration was checked with a NOESY (Nuclear Overhauser Effect Spectroscopy) spectrum of one product. The proximity of the olefinic proton towards the NH_2 was checked (Fig. 82)

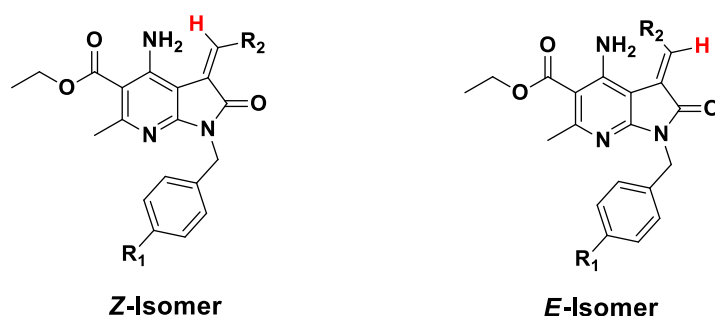


Figure 82. The two possible configurations of the products. When it has the *Z*-configuration, a cross peak from the hydrogen in **red** to the -NH_2 would be visible in a NOESY-spectrum. ($\text{R}_1 = \text{H, F, CF}_3$; $\text{R}_2 = \text{Pyrrole, Azaindole, Indole}$).

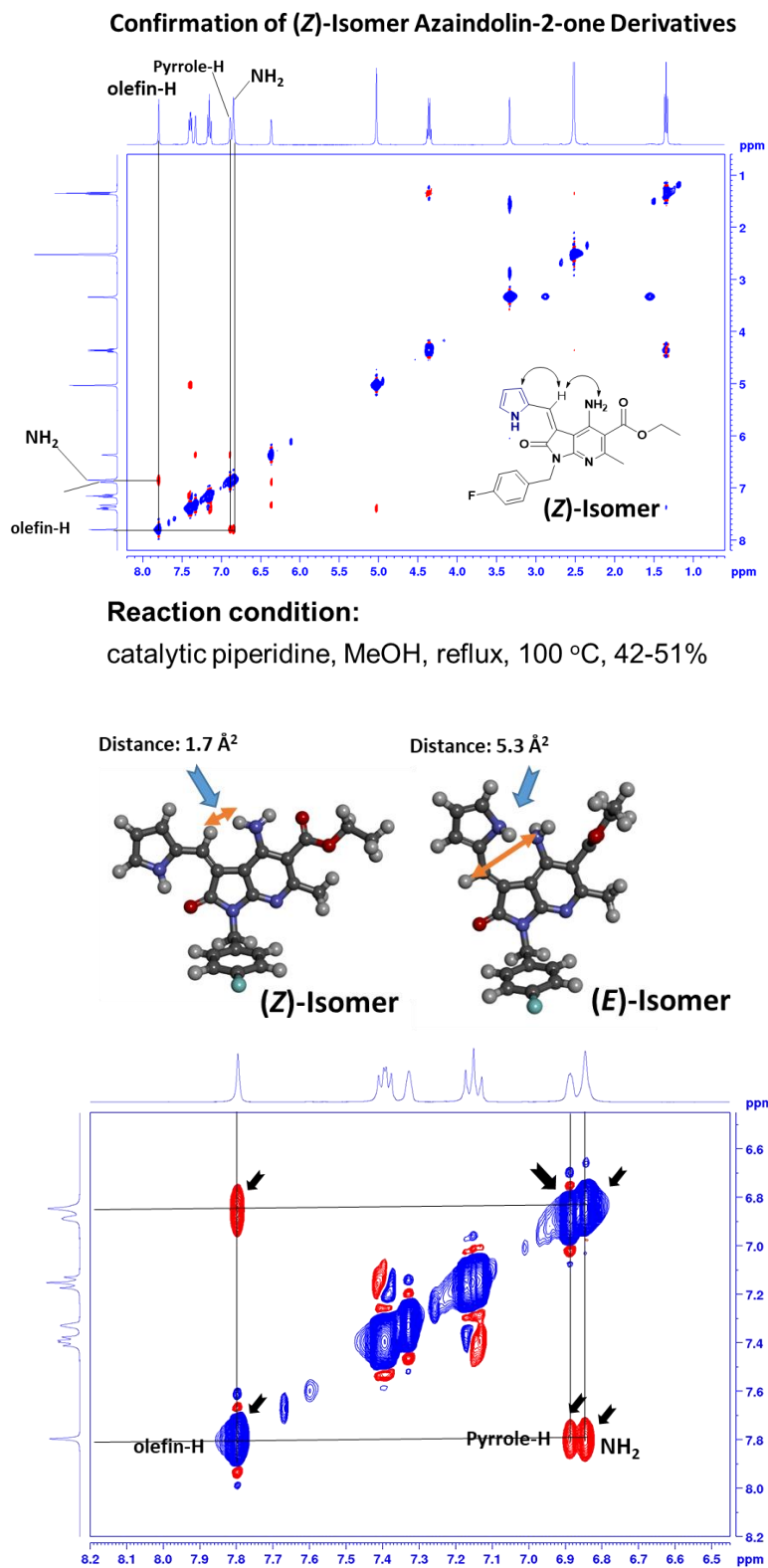


Figure 83. A cross peak was observed in the NOESY spectrum, so the products which were synthesized with this procedure have the Z-configuration

4.8 Intramolecular hydrogen bond in *Z*-isomer

The intramolecular hydrogen bond is an important structural factor that can rigidify molecules and maintain a large, nearly planar structure, which may allow ligands to effectively stack with the G-quartet.^{52,53} The crescent-shaped, large, nearly planar chromophore led us to explore a novel scaffold for the design of selective G quadruplex ligands. In case of *Z*-isomer, this important feature of intramolecular hydrogen bond has been observed (Fig. 84).

Intramolecular hydrogen bonds in (*Z*) isomer of Azaindolin-2-one Derivatives

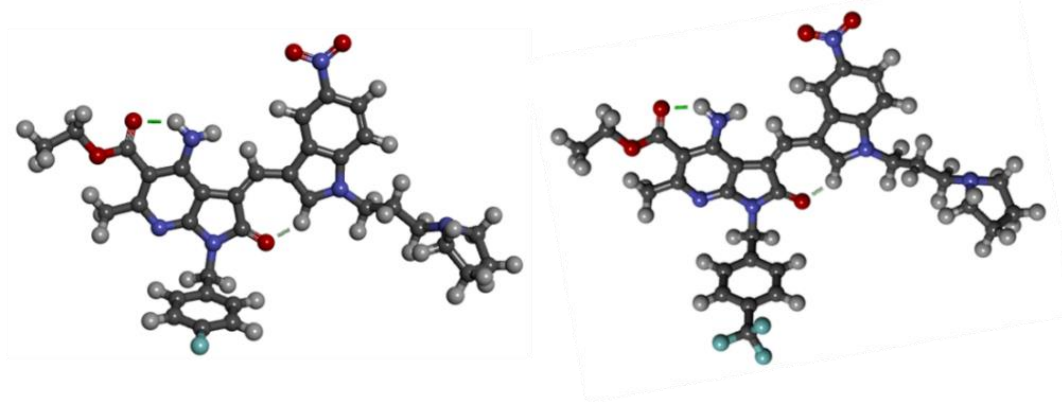
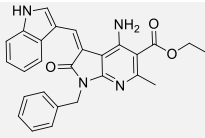
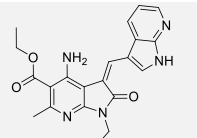
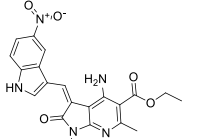
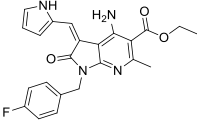
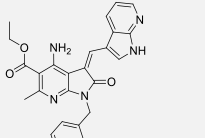
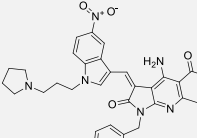
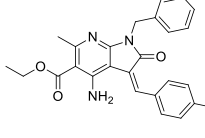
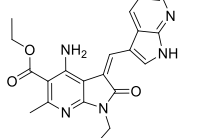
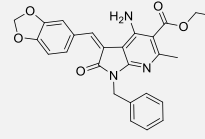
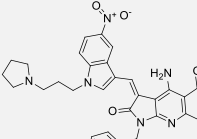
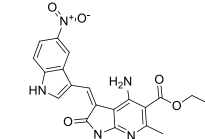
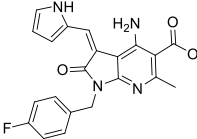
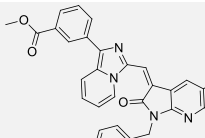


Figure 84. Intramolecular hydrogen bond (highlighted in green color) observed in case of *Z*-isomers visualized by Discovery suite studio

4.9 Initial screening of azaindolin-2-one derived conjugates with Fluorescent indicator displacement (FID) assays

The ability of all key azaindolin-2-one derived conjugates generated from scheme II and III were then investigated to evaluate by FID assays for *c*-MYC promoter sequence (Fig. 85 and Table 18). From primary FID screening best azaindolin-2-one derived conjugates (**14**, **16**, **18**, **19** and **22**) were then further investigated for their ability to stabilize the stabilize several G4-forming sequences (thiazole-orange labeled *c*-MYC, *c*-KIT1, *BCL-2* G4-sequences. In this assay, we monitored the thiazole displacement ability by increasing concentrations of the ligands. Additionally, ligands were tested also towards a double-stranded DNA in order to assess the G4 vs ds-DNA selectivity. In our findings, we found that azaindolin-2-one derived conjugates (**14** and **16**) showed a very strong affinity towards *c*-MYC G4 DNA with K_i values ranging from 2.0-4.0 μ M. Most compounds showed the preferential affinity towards human parallel promoter *c*-MYC G4 DNA sequences in comparison to other promoter, telomeric G-quadruplex and duplex oligonucleotide sequences (Table 19, Fig. 86 & 87).

Table 18. List of synthesized fragments screened based on Thiazole Displacement Assays

Hit	Best hits	Fid affinity	Hit	Best hits	Fid affinity
6		++	12		++
7		++	13		++
8		++	14		+++++
9		++	15		++++
10		++	16		+++++
11		+++	17		++
22		+++			

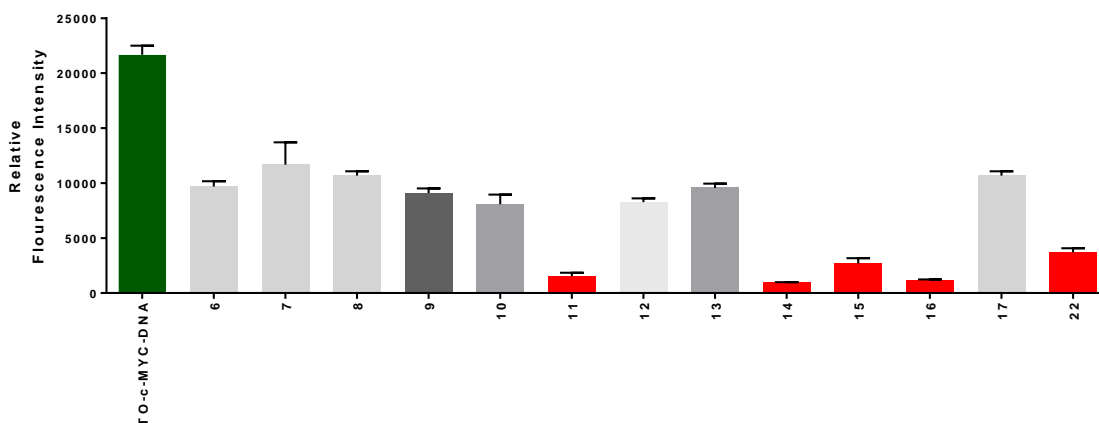


Figure 85. Thiazole displacement potential of the azaindolin-2-one derived G4-ligands. Best ligands are highlighted in red color Experimental condition: 0.25 μM DNA, 0.5 μM Thiazole Orange, 10 % DMSO, 20 mM Na cac, 140 mM KCl, pH 7 (25 μL /well)

Table 19. K_i values for the most active ligands for G-quadruplex and Duplex DNA

K_i (μM)	14	16	15	11	22
c-MYC	2.01	4.96	8.59	8.23	8.12
c-KIT-1	24.21	27.83	28.85	28.93	28.89
BCL2	33.37	31.75	43.17	30.43	29.75
h-Telo- 22	81.39	83.13	81.13	98.74	108.13
DNA-27	5861.13	5075.12	2088.3	1478.26	2149.66

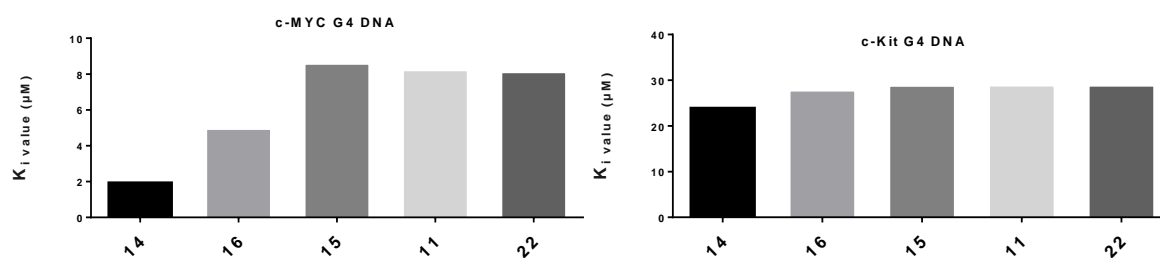


Figure 86. K_i value plots for the most active ligands for promoter G-Quadruplex sequences *c-MYC*, *c-KIT-1* based on Thiazole Displacement Assays ; Titration scheme of *c-MYC*: $c(\text{TO}) = 0,5 \mu\text{M}$; $c(\text{cMYC}) = 0,5 \mu\text{M}$; $c(\text{DMSO}) = 10\%$; $c(\text{Ligand}) = 256 \mu\text{M}$, $128 \mu\text{M}$, $64 \mu\text{M}$, $32 \mu\text{M}$, $16 \mu\text{M}$, $8 \mu\text{M}$, $4 \mu\text{M}$, $2 \mu\text{M}$, $1 \mu\text{M}$, $0,5 \mu\text{M}$, $0,25 \mu\text{M}$, $0,125 \mu\text{M}$, $0,0625 \mu\text{M}$; r.t; Titration of *c-KIT1*: $c(\text{TO}) = 0,5 \mu\text{M}$; $c(\text{c-KIT-1}) = 0,25 \mu\text{M}$; $c(\text{DMSO}) = 10\%$; $c(\text{Ligand}) = 1024 \mu\text{M}$, $512 \mu\text{M}$, $256 \mu\text{M}$, $128 \mu\text{M}$, $64 \mu\text{M}$, $32 \mu\text{M}$, $16 \mu\text{M}$, $8 \mu\text{M}$, $4 \mu\text{M}$, $2 \mu\text{M}$, $1 \mu\text{M}$, $0,5 \mu\text{M}$, $0,25 \mu\text{M}$, $0,125 \mu\text{M}$; r.t

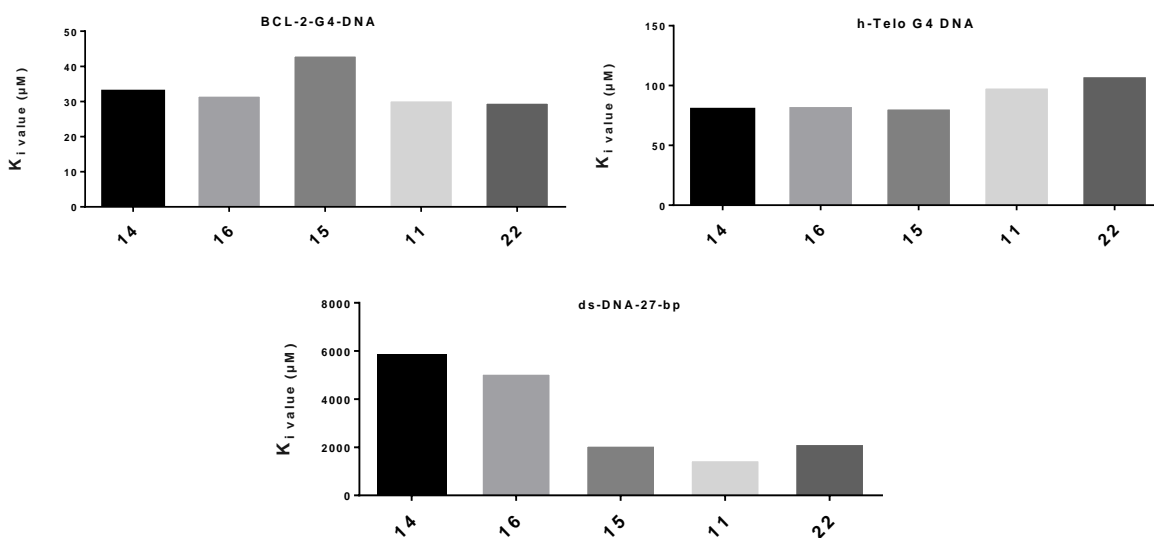


Figure 87. K_i value plots for the most active ligands for promoter G-Quadruplex sequences *BCL-2*, *h-TELO*, and *ds-DNA-27* based on Thiazole Displacement Assays; Titration scheme of *BCL-2*: $c(\text{TO}) = 0,5 \mu\text{M}$; $c(\text{BCL-2}) = 0,25 \mu\text{M}$; $c(\text{DMSO}) = 10\%$; $c(\text{Ligand}) = 1024 \mu\text{M}$, $512 \mu\text{M}$, $256 \mu\text{M}$, $128 \mu\text{M}$, $64 \mu\text{M}$, $32 \mu\text{M}$, $16 \mu\text{M}$, $8 \mu\text{M}$, $4 \mu\text{M}$, $2 \mu\text{M}$, $1 \mu\text{M}$, $0,5 \mu\text{M}$, $0,25 \mu\text{M}$, $0,125 \mu\text{M}$; r.t; Titration of *h-Telo-DNA*: $c(\text{TO}) = 0,5 \mu\text{M}$; *h-Telo* = $0,5 \mu\text{M}$; $c(\text{DMSO}) = 10\%$; $c(\text{Ligand}) = 5012 \mu\text{M}$, $2056 \mu\text{M}$, $1028 \mu\text{M}$, $604 \mu\text{M}$, $302 \mu\text{M}$, $106 \mu\text{M}$, $80 \mu\text{M}$, $40 \mu\text{M}$, $20 \mu\text{M}$, $10 \mu\text{M}$, $5 \mu\text{M}$, $1 \mu\text{M}$, $0,5 \mu\text{M}$; r.t; Titration of *ds-DNA-27*: $c(\text{TO}) = 0,5 \mu\text{M}$; $(\text{ds-DNA-27}) = 0,5 \mu\text{M}$; $c(\text{DMSO}) = 10\%$; $80192 \mu\text{M}$, $40096 \mu\text{M}$, $20048 \mu\text{M}$, $10024 \mu\text{M}$, $5012 \mu\text{M}$, $2056 \mu\text{M}$, $1028 \mu\text{M}$, $604 \mu\text{M}$, $302 \mu\text{M}$, $106 \mu\text{M}$, $20 \mu\text{M}$, $10 \mu\text{M}$, $5 \mu\text{M}$; r.t; Titration of *ds-DNA-24-Bp*: $c(\text{TO}) = 0,5 \mu\text{M}$; $c(\text{cMYC}) = 0,25 \mu\text{M}$; $c(\text{DMSO}) = 10\%$; $c(\text{Ligand}) = 20048 \mu\text{M}$, $10024 \mu\text{M}$, $5012 \mu\text{M}$, $2056 \mu\text{M}$, $1028 \mu\text{M}$, $604 \mu\text{M}$, $302 \mu\text{M}$, $106 \mu\text{M}$, $80 \mu\text{M}$, $40 \mu\text{M}$, $20 \mu\text{M}$, $10 \mu\text{M}$, $5 \mu\text{M}$; r.t

4.10 ^1H -NMR: Ligand binding observed by NMR-spectroscopy

The interactions of azaindolin-2-one conjugate **14** with *c*-MYC G4 were further studied by NMR spectroscopy. Figure 88a and 88b shows NMR spectra of the DNA-ligand complexes at a ratio of 1:4 (DNA: ligand). In the spectrum of the DNA alone, signals (indicated by grey boxes) of minor conformations of the DNA are well visible. These signals disappear upon addition of ligand, indicating the stabilization of the major conformation upon binding. For ligand **14** strong chemical shift perturbations and line broadening are observed, further addition of ligand up to a ratio of 4:1 [**14**]:[*c*-MYC] leads to stronger line broadening, suggesting binding in an intermediate exchange. Observation of binding or structural rearrangement of residues in loop and capping structures can be achieved by investigation of the aromatic region of the spectra. Signals of A12, A22, T1, T20 and A21 are reasonably well resolved to allow the assessment of behavior upon binding. As A12 does not shift upon binding the loop does not change its environment upon binding. This is different for T1, T20, A21 and A22. These residues located in the capping structures are perturbed significantly upon binding indicating rearrangement of the capping structures to accommodate the ligands. This rearrangement allowing to accommodate the ligand is also seen in the literature for *c*-MYC,⁵⁴ as well as for other G4s.⁵⁵

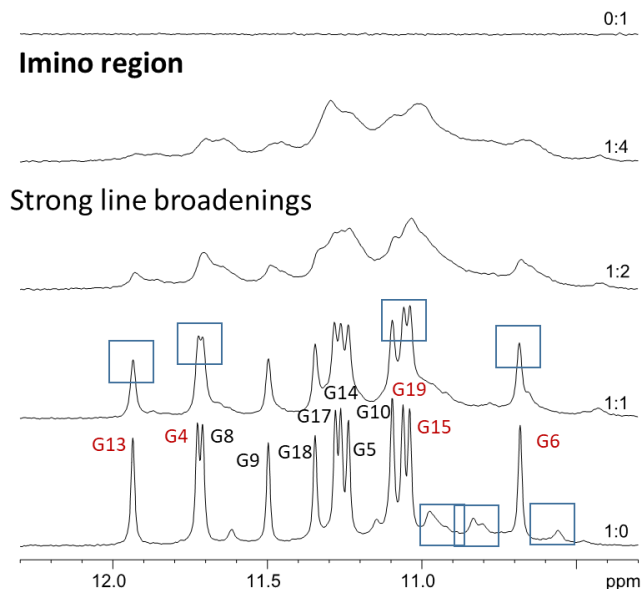


Figure 88a. Imino region of 1D ^1H NMR spectrum of the *c*-MYC with increasing [**14**]: [DNA] ratio. Experimental conditions: 298 K, 600 MHz, 100 μM DNA in 25 mM Tris-HCl (pH 7.4), 100 mM KCl buffer, 10% d_6 -DMSO/90% H_2O .

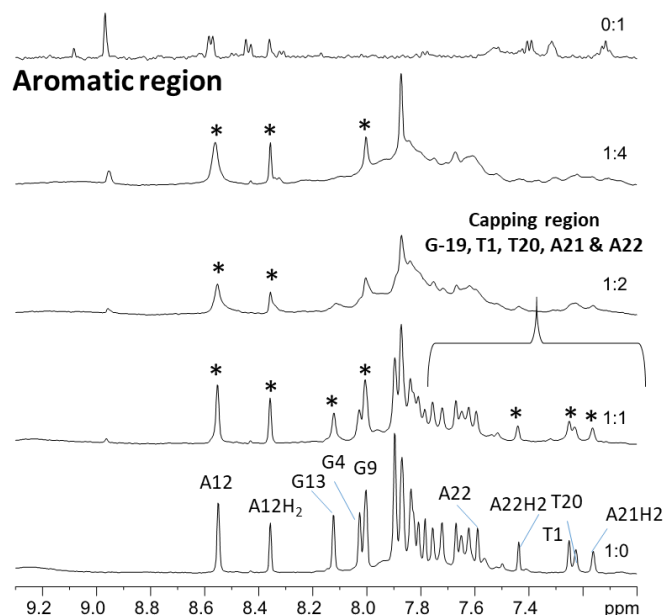


Figure 88b. Aromatic region of ^1H NMR spectrum of the *c*-MYC with increasing [14]: [DNA] ratio. Experimental conditions: 298 K, 600 MHz, 100 μM DNA in 25 mM Tris-HCl (pH 7.4), 100 mM KCl buffer, 10% d_6 -DMSO/90% H_2O .

4.11 Cell Cytotoxicity

Based on all the data from the above experiments, we studied whether the most significant triazole derivatives could affect *c*-MYC transcription in cells. Firstly, promising ligands were analyzed *in vitro* by cell viability assays. HeLa cells were used to perform the cell proliferation assay, as they were not only having high levels of *c*-MYC expression but were also used for similar studies.⁵⁶ As described in the materials and methods, HeLa cells were exposed with various increasing concentrations (0-50 μM) of the candidate compounds and the **11**, **14**, **15**, and **16** respectively for noting the effects of defined modifications. The percentage of living cells was then measured with alamar-Blue reagent.⁵⁷ Best-selected hits showed dose-dependent cytotoxic effects in HeLa cell line. Dose-response curves, the relative IC_{50} values, and the corresponding linear regression coefficients (R^2), for each G4-ligand, are reported. Interestingly, conjugate **14** is amongst the most potent in terms IC_{50} values. Conjugates **14** and **16** were found to be the most potent against HeLa cervical cancer cells (Fig. 90C & Table 20).

Table 20. IC₅₀ values derived for **11**, **14**, **15** and **16** in HeLa cell line after 72 h treatment.

	G-Quadruplex Ligands			
	11	14	15	16
IC ₅₀ (μ M)	5.575	2.258	6.520	3.865

4.12 Western Blot analysis

Down-regulation of c-MYC expression in human cancer cells: In order to investigate the effect of conjugate **14** and **16** at the translational level, HeLa cells were treated with defined similar concentrations of the ligands for 24 hr and monitored. To further investigate the effects of these fragments **5**, **7**, **9a** and **12** along with parent compound **3**⁵⁸ at the translational level, HeLa cells were treated with defined concentrations of the ligands for 24 h and the expression levels of c-MYC protein relative to the control cells were obtained using western blotting with an anti-c-MYC antibody (Fig. 90 A-B). The protein expressions calculated for ligand treated cells from densitometry analysis of western blots were normalized against untreated control cells. In **14** and **16** treated HeLa cells, the c-MYC protein expression was downregulated by 40 % at 3 μ M & 60 % at 10 μ M respectively (Fig. 85a & 85b). Quantitative gene expression data are often normalized to the expression levels of GAPDH "housekeeping" genes.

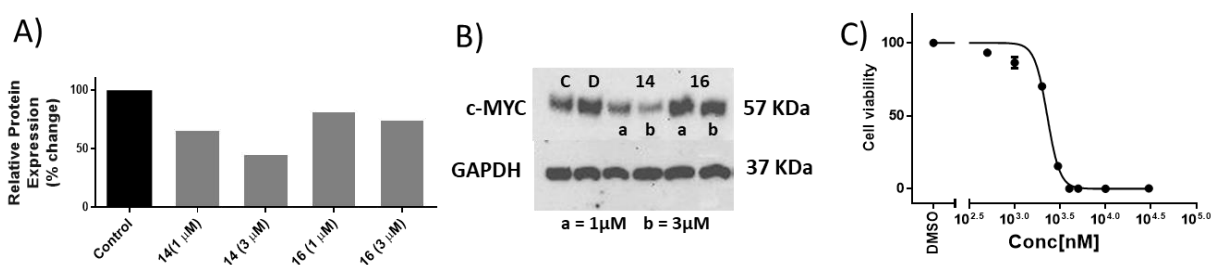


Figure 90. A) Bar diagram representing fold expression of c-MYC protein expression in HeLa cells B) Western blot analysis for c-MYC gene expression in of 14 and 16 treated HeLa cells (a=3 μ M and b= 10 μ M); C) Dose-response and IC₅₀ values derived for best ligand 14 in HeLa cell line after 72 h treatment.

4.13 Cell Cycle Analysis

In our investigation we found that azaindolin-2-one conjugate **14** is the best G4-ligand in this series, which could down-regulate c-MYC transcription and expression in cells, we next evaluated

whether **14** could affect cancer cell proliferation. On exposure to **14** at 2.5 μ M concentration, cells displayed a prominent G1 phase arrest in flow cytometry (Fig. 91). The G0/G1 cell population increased to 73.8 % upon incubation with conjugate **14** for 24 hr at 2.5 μ M concentration. The observed increase in G0/G1 population suggested that the compounds caused a significant block in S/G2-phase progression. Most probably by interfering in DNA unwinding due to ligand-mediated stabilization of promoter *c*-MYC G-quadruplexes. The FACS analysis thus suggests that upon treatment with conjugate **14**, cells undergo G0/G1 arrest with concomitant decrease of S/G2-phase population thereby leading to cell death. The above results suggested that conjugate **14** could inhibit proliferation of cancer cells, which could possibly be related to its binding with the *c*-MYC G4 of the *c*-MYC promoter and repressing *c*-MYC oncogene transcription.

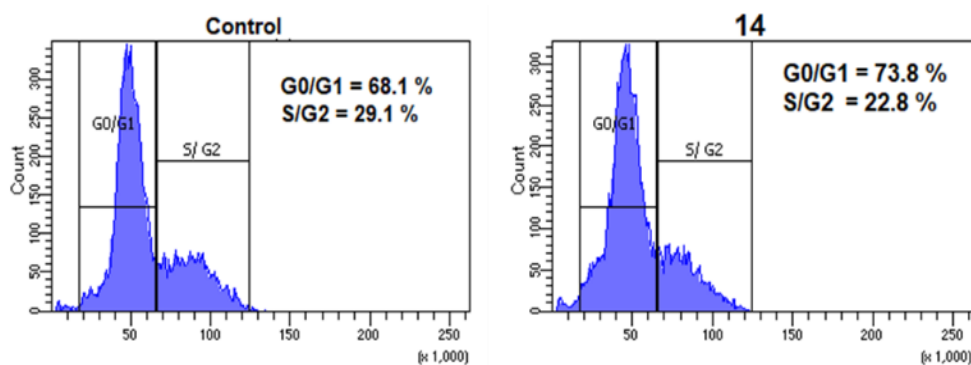


Figure 91. Cell cycle analysis of HeLa cells treated with ligand **14** at 2.5 μ M concentrations for 24 h.

4.14 Conclusion

A new series of azaindolin-2-one derivatives as potential G quadruplex stabilizers have been described. The target ligands can be obtained through an efficient, convergent, synthetic route in moderate to good yields. In FID assays azaindolin-2-one conjugates (**11**, **14**, **15**, **16**, and **22**) show specific binding to the *c*-MYC promoter G-quadruplex sequences compared to that of the other G-quadruplex sequences (*c*-KIT-1, *BCL2*, h-Telo). The synthesized azaindolin-2-one derivatives **11**, **14**, **15**, **16**, and **22** are selective towards *c*-MYC G4-DNA vs. duplex-DNA. The planarity of the aromatic core and its ability to occupy more surface area by stacking over the G-quartet greatly affect the ability of the compounds to stabilize the G-quadruplex. In addition to that, in our biophysical and biological evaluation, we found that the indole linked azaindolin-2-one derivatives (**14** & **16**) are more promising than the other heterocycle conjugates. In cell viability assays, the indole linked azaindolin-2-one derivatives (**14** & **16**) has shown potent cell growth

inhibitory activity against human carcinoma (cervical) HeLa cell lines. In vitro cellular studies confirms that the indole linked azaindolin-2-one conjugates could down-regulate *c-MYC* expression in cancer cells *via* stabilizing its promoter quadruplex structure and have the potential to induce cell cycle arrest in G₀/G₁ phase. NMR analysis suggests that an azaindolin-2-one conjugate (14) interacts in an intermediate exchange regime with the terminal G-quartets (5' and 3'-end) and strong chemical shift perturbations were observed. Our present results increased our understanding of the effect of c-MYC G₄ on oncogene transcriptional regulation. The present study also suggested that the azaindolin-2-one conjugates 14 could become a potential lead compound for further development for cancer treatment with a new target of the *c-MYC* promoter G-Quadruplex. This information can guide the optimization of ligands 14 and 16, for increased binding affinity and selectivity toward G-quadruplex structures.

4.15 Experimental Section

Chemistry

All solvents and reagents were purified by standard techniques or used as supplied from commercial sources (Sigma-Aldrich Corporation unless stated otherwise). All reactions were generally carried out under inert atmosphere unless otherwise noted. TLC was performed on Kieselgel 60 F254 plates, and spots were visualized under UV light. Products were purified by flash chromatography on silica gel (100-200 mesh). ¹H NMR spectra were recorded at 600 MHz instruments at 298 K. ¹³C NMR spectra were recorded on either 151 MHz with complete proton decoupling. Chemical shifts are reported in parts per million (ppm) and are referred to the residual solvent peak. The following notations are used: singlet (s); doublet (d); triplet (t); quartet (q); multiplet (m); broad (br). Coupling constants are quoted in Hertz and are denoted as J. Mass spectra were recorded on a Micromass Q-ToF (ESI) spectrometer.

General Procedure for the Synthesis of substituted-(benzylamino)-3-bromopropan-2-one (1b, 2b, 3b, 4b and 5b): 2-bromoacetyl bromide (1.40 ml, 16.1 mmol and 1.0 eq) was added dropwise to a solution of substituted-phenylmethanamine (2.30 ml, 21.1 mmol, and 1.0 eq) in DCM (10 mL) at 0 °C. The mixture was allowed to warm up to room temperature and stirred for another 2 h. After completion of the reaction (TLC showed no starting material), the reaction was quenched with H₂O (5 mL). The precipitate was filtered off, the organic phase was washed with H₂O, dried over Na₂SO₄ and evaporated to afford the product as white solid.

General Procedure for the Synthesis of Substituted N-alkyl- α -bromoacetamide (1c, 2c, 3c, 4c and 5c): Malononitrile (3.47 g, 52.5 mmol, 2.5 eq.) was added in small fraction to a suspension

of NaH (1.22 g, 50.8 mmol, and 2.5 eq.) in THF (40 mL) at 0 °C. The mixture was allowed to warm up to room temperature and stirred for 30 minutes. After addition of substituted-(benzylamino)-3-bromopropan-2-one (4.80 g, 21.0 mmol, 1.0 eq.), the mixture was allowed to stirred overnight at room temperature. After completion of the reaction, the reaction was quenched with ammonium chloride aqueous solution (40 ml) and extracted with ethyl acetate for two times. The organic phase was combined, dried over Na₂SO₄ and concentrated under reduced pressure. The residue was purified by column chromatography to give product as white solid.

General Procedure for the Synthesis of ethyl 1-benzyl-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (1d, 2d, 3d, 4d and 5d): Ethyl 3-oxobutanoate (1.60 ml, 12.5 mmol, 2.2 eq.) was added to a suspension of aluminium trichloride (1.50 g, 11.2 mmol, 2.2 eq.) in 1,2-dichloroethane. The mixture was reflux at 110 °C for 5 h. 2-amino-1-benzyl-5-oxo-4,5-dihydro-1H-pyrrole-3-carbonitrile (1.20 g, 5.60 mmol) was added to the above reaction mixture and stirred at this temperature for 5 h. The reaction was quenched by adding an ammonium chloride aqueous solution (40 mL) with ice cooling and then extracted with ethyl acetate. The organic phase was washed with brine and dried over Na₂SO₄. The solvent was concentrated under reduced pressure to give a white solid.

General Procedure for the Synthesis of Knoevenagel condensation products (6, 7, 8, 9, 10,11, 12, 13, 14, 15, 16, 17 and 22) One equivalent of azaindolin-2-one precursor (1d, 2d, 3d) for Knoevenagel condensation reaction was dissolve in MeOH, a catalytic amount of piperidine was added in to the reaction mixture and was allow to stir for 30 min. Then 1 equivalent of substituted aldehydes was slowly added to the reaction mixture. The reaction mixture was heated to reflux for 2 h. The reaction mixture was then allowed to cool to the room temperature and was then quenched by adding an ammonium chloride aqueous solution (40 mL) with ice cooling and then extracted with ethyl acetate. The organic phase was washed with brine and dried over Na₂SO₄. The solvent was concentrated under reduced pressure to get Knoevenagel condensation product.

Procedure for the synthesis of imidazo [1,5-a]pyridine-3-carbaldehyde (19): To a stirring dry DMF (3 mL) about 2 mL of POCl₃ were added in dropwise manner. Afterwards reaction was allowed to stir at room temperature for 30 min followed with addition of Imidazo [1,5-a]pyridine (3, 0.500 g, 4.23 mmol, 1 eq) in 5 mL of DME dropwise manner. The reaction was kept at room temperature under N₂-atmosphere for 2 h. The reaction mixture was quenched with ice water and then neutralized with 1 N NaOH to pH 8-9. The organic layer was extracted with ethyl acetate

and dried with MgSO_4 . The rotary evaporator was used to concentrate the product with under reduced pressure. After silica gel column chromatography, the product was obtained. As eluent the mixture of ethyl acetate/ cyclohexene was used.

Procedure for the synthesis of 1-bromoimidazo [1,5-a]pyridine-3-carbaldehyde (20):

Under nitrogen atmosphere at room temperature (0.151 g, 0.846 mmol, 1 eq) N-Bromosuccinimide was added to a solution of imidazo[1,5-a]pyridine-3-carbaldehyde (0.151 g, 0.846 mmol, 1 eq) in 20 ml dry DMF. The reaction mixture was then allowed to stir at room temperature for 3-4 hr. The reaction mixture was further quenched with ice cold water. The solid obtained was then filtered and rinsed with water and dried overnight under reduced pressure. The product was further purified by column chromatography (EA: Cyclohexene 1:1).

Procedure for the synthesis of methyl 3-(3formylimidazol[1,5-a]pyridine-1-yl)benzoate (21):

Under nitrogen atmosphere 60 mg (0.33 mmol, 1.5 eq) of (3-(methoxycarbonyl)phenyl)boronic acid, 50 mg (0.22 mmol, 1 eq) of 1-bromoimidazo[1,5-a]pyridine-3-carbaldehyde, 181 mg (0.557 mmol, 2.5 eq) of Cs_2CO_3 and 20 mg (0.017 mmol, 5 %) of tetrakis(triphenylphosphine)palladium(0) were dissolved in 2 ml distilled water and 20 ml dioxane. The reaction mixture was allowed refluxed for 4 h under inert atmosphere. The reaction was quenched by adding an ammonium chloride aqueous solution (40 mL) with ice cooling and then extracted with ethyl acetate. The organic phase was washed with brine and dried over Na_2SO_4 . The dried organic phase was concentrated under reduced pressure to give a white solid. The crude product was then further purified by column (ethyl acetate/ cyclohexene).

Spectral information for scheme I is mentioned in Chapter II

N-benzyl-2-bromoacetamide (1b): White solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO}-d_6$) δ [ppm] 8.81 (s, 1H), 7.44 (dt, $J = 10.7, 5.1$ Hz, 2H), 7.32 (d, $J = 7.3$ Hz, 2H), 7.27 (s, 1H), 4.29 (d, $J = 5.9$ Hz, 2H), 3.92 (s, 2H); ^{13}C NMR (151 MHz, $\text{DMSO}-d_6$) δ [ppm] 166.1, 138.8, 134.0, 128.9, 128.6, 128.5, 128.3, 127.3, 126.9, 42.5, 29.5.

2-bromo-N-(4-fluorobenzyl)acetamide (2b): White solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO}-d_6$) δ [ppm] 8.79 (s, 1H), 7.29 (d, $J = 5.7$ Hz, 2H), 7.16 (t, $J = 8.5$ Hz, 2H), 4.27 (d, $J = 5.5$ Hz, 2H), 3.90 (s, 2H); ^{13}C NMR (151 MHz, $\text{DMSO}-d_6$) δ [ppm] 166.1, 135.1, 135.0, 129.3, 129.2, 115.2, 114.9, 41.8, 29.4.

2-bromo-N-(4-(trifluoromethyl) benzyl) acetamide (3b): White solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 8.90 (s, 1H), 7.70 (d, $J = 7.9$ Hz, 2H), 7.48 (d, $J = 7.9$ Hz, 2H), 4.39 (d, $J = 5.7$ Hz, 2H), 3.93 (s, 2H); ^{13}C NMR (151 MHz, $\text{DMSO-}d_6$) δ [ppm] 166.3, 143.8, 127.9, 125.3, 125.3, 125.2, 125.2, 42.1, 29.3.

2-bromo-N-(3,4,5-trimethoxybenzyl)acetamide (4b): White solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 6.70 (s, 2H), 6.57 (s, 1H), 4.74 (s, 2H), 3.73 (s, 6H), 3.63 (s, 3H), 3.93 (s, 2H); ^{13}C NMR (151 MHz, $\text{DMSO-}d_6$) δ [ppm] 174.30, 167.39, 187.94, 156.98, 136.75, 132.53, 105.13, 60.50, 59.30, 55.70, 32.7.

2-bromo-N-(4-methoxybenzyl)acetamide (5b): White solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 8.70 (s, 1H), 7.21 (t, $J = 8.8$ Hz, 2H), 6.95 – 6.84 (m, 2H), 4.23 (d, $J = 5.9$ Hz, 2H), 3.90 (s, 2H), 3.75 (s, 3H); ^{13}C NMR (151 MHz, $\text{DMSO-}d_6$) δ [ppm] 174.30, 167.39, 187.94, 156.98, 136.75, 132.53, 105.13, 60.50, 59.30, 55.70, 32.7.

2-amino-1-benzyl-5-oxo-4,5-dihydro-1H-pyrrole-3-carbonitrile (1c): Orange solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 7.71 (d, $J = 8.1$ Hz, 2H), 7.42 (d, $J = 8.0$ Hz, 2H), 7.27 (s, 2H), 4.84 (s, 2H), 3.30 (s, 2H); ^{13}C NMR (101 MHz, $\text{DMSO-}d_6$) δ [ppm] 173.9, 157.5, 141.3, 127.6, 125.4, 125.4, 125.3, 123.3, 119.1, 47.3, 41.4, 34.3, 8.4.

2-amino-1-(4-fluorobenzyl)-5-oxo-4,5-dihydro-1H-pyrrole-3-carbonitrile (2c): Orange solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 7.27 (dd, $J = 8.5, 5.6$ Hz, 2H), 7.24 (s, 2H), 7.16 (t, $J = 8.9$ Hz, 2H), 4.71 (s, 2H), 3.27 (s, 2H); ^{13}C NMR (151 MHz, $\text{DMSO-}d_6$) δ [ppm] 173.9, 157.6, 132.7, 129.1, 119.2, 115.3, 115.1, 47.2, 41.0, 34.3, 8.4.

2-amino-5-oxo-1-(4-(trifluoromethyl)benzyl)-4,5-dihydro-1H-pyrrole-3-carbonitrile (3c): Orange solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 7.71 (d, $J = 8.1$ Hz, 2H), 7.42 (d, $J = 8.0$ Hz, 2H), 7.27 (s, 2H), 4.84 (s, 2H), 3.30 (s, 2H); ^{13}C NMR (151 MHz, $\text{DMSO-}d_6$) δ [ppm] 173.9, 157.5, 141.3, 127.6, 125.4, 125.4, 125.3, 123.3, 119.1, 47.3, 41.4, 34.3, 8.4.

2-amino-5-oxo-1-(3,4,5-trimethoxybenzyl)-4,5-dihydro-1H-pyrrole-3-carbonitrile(4c): Orange solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 7.23 (s, 2H), 7.19 (t, $J = 5.8$ Hz, 2H), 6.92 – 6.84 (m, 2H), 4.68 (s, 2H), 3.74 (s, 3H), 3.26 (s, 2H); ^{13}C NMR (151 MHz, $\text{DMSO-}d_6$) δ [ppm] 174.05, 158.64, 157.45, 130.61, 128.54, 118.97, 113.77, 54.83, 46.68, 34.35, 8.67.

2-amino-1-(4-methoxybenzyl)-5-oxo-4,5-dihydro-1H-pyrrole-3-carbonitrile (5c): Orange solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 7.21 (s, 2H), 6.60 (d, $J = 8.5$ Hz, 2H), 4.65 (d, $J = 7.1$ Hz, 2H), 4.45 (d, $J = 6.1$ Hz, 2H), 3.75 (s, 6H), 3.65 (s, 3H); ^{13}C NMR (151 MHz, $\text{DMSO-}d_6$) δ [ppm] 173.84, 157.46, 152.67, 137.00, 132.00, 118.98, 112.21, 104.89, 59.63, 47.02, 41.84, 34.37, 8.63.

Ethyl 4-amino-1-benzyl-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (1d): White solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 7.27-7.39 (m, 5H), 6.58 (s br, 2H), 4.27 (q, $J = 7.1$ Hz, 2H), 3.43 (s, 2H), 2.47 (s, 3H), 1.31 (t, $J = 7.1$ Hz, 3H); ^{13}C NMR (151 MHz, $\text{DMSO-}d_6$) δ [ppm] 174.4, 173.9, 149.4, 137.3, 137.5, 136.5, 128.4, 128.3, 127.4, 127.2, 127.1, 126.8, 97.4, 60.5, 41.6, 34.3, 32.9, 25.6, 14.0.

ethyl 4-amino-1-(4-fluorobenzyl)-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (2d): White solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 7.35 δ [ppm] (dd, $J = 8.3, 5.8$ Hz, 2H), 7.12 (t, $J = 8.9$ Hz, 2H), 6.62 (s, 1H), 4.80 (s, 2H), 4.29 (q, $J = 7.1$ Hz, 2H), 3.43 (d, $J = 12.8$ Hz, 2H), 2.47 (s, 3H), 1.30 (t, $J = 7.1$ Hz, 3H); ^{13}C NMR (151 MHz, $\text{DMSO-}d_6$) δ [ppm] 174.4, 167.6, 162.1, 160.5, 156.7, 149.4, 133.4, 129.6, 129.5, 115.1, 115.0, 106.9, 97.4, 60.5, 41.0, 33.0, 25.4, 14.0.

ethyl 4-amino-6-methyl-2-oxo-1-(4-(trifluoromethyl)benzyl)-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (3d): White solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 7.66 (d, $J = 8.1$ Hz, 2H), 7.50 (d, $J = 8.0$ Hz, 2H), 6.60 (s, 2H), 4.90 (s, 2H), 4.29 (q, $J = 7.1$ Hz, 2H), 3.45 (s, 2H), 2.45 (s, 3H), 1.29 (s, 3H); ^{13}C NMR (151 MHz, $\text{DMSO-}d_6$) δ [ppm] 174.5, 167.7, 158.4, 156.8, 149.4, 142.0, 128.1, 125.3, 123.3, 106.9, 97.5, 60.4, 49.6, 41.2, 33.0, 30.0, 25.6, 14.0.

ethyl 4-amino-6-methyl-2-oxo-1-(3,4,5-trimethoxybenzyl)-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (4d): White solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 6.69 (d, $J = 11.6$ Hz, 2H), 6.58 (bs, 2H), 4.74 (s, 2H), 4.31 (q, $J = 7.1$ Hz, 2H), 3.72 (d, $J = 7.0$ Hz, 6H), 3.66 – 3.59 (m, 3H), 3.43 (s, 3H), 1.32 (t, $J = 7.1$ Hz, 2H); ^{13}C NMR (151 MHz, $\text{DMSO-}d_6$) δ [ppm] 174.5, 167.7, 158.4, 156.8, 149.4, 142.0, 128.1, 125.3, 123.3, 106.9, 97.5, 60.4, 49.6, 41.2, 33.0, 30.0, 25.6, 14.0.

ethyl 4-amino-1-(4-methoxybenzyl)-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (5d): White solid (Crude) used directly in the next step; ^1H NMR (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 7.46 (d, $J = 7.7$ Hz, 2H), 6.83 (d, $J = 8.0$ Hz, 2H), 5.74 (s, 2H), 4.92 (s, 2H),

4.38 (d, $J = 6.9$ Hz, 2H), 3.78 (s, 3H), 3.29 (s, 2H), 2.72 (s, 3H), 1.42 (t, $J = 6.8$ Hz, 3H); ^{13}C NMR (151 MHz, DMSO- d_6) δ [ppm] 174.5, 167.7, 158.4, 156.8, 149.4, 142.0, 128.1, 125.3, 123.3, 106.9, 97.5, 60.4, 49.6, 41.2, 33.0, 30.0, 25.6, 14.0.

(Z)-ethyl 3-((1H-indol-3-yl)methylene)-4-amino-1-benzyl-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (6): Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 46 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.33; Mp 112 - 114 °C. ^1H NMR (600 MHz, DMSO- d_6) δ [ppm] 11.23 (s br, 1H), 9.08 (d, $J = 4$ Hz, 1H), 7.62 (d, $J = 8$ Hz, 1H), 7.17-7.10 (m, 5H), 7.08-6.88 (m, 6H), 4.87 (s, 2H), 4.08 (q, $J = 8$ Hz, 2H), 2.37 (s, 3H), 1.09 (t, $J = 8$ Hz, 3H); m/z 452.76 [M+H]

(Z)-Ethyl-4-amino-1-benzyl-6-methyl-3-((5-nitro-1H-indol-3-yl)methylene)-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (7): Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 56 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.26; Mp 102 - 104 °C. ^1H -NMR (600 MHz, CDCl_3) δ [ppm] 9.32 (s, 1H), 8.80 (s, 1H), 7.98-7.92 (dd, $J = 12$ Hz, 1H), 7.71 (s, 1H), 7.34-7.26 (m, 2H), 7.09-7.03 (m, 5H), 6.60 (s br, 2H), 4.93 (s, 2H), 4.23 (q, $J = 6$ Hz, 2H), 2.39 (s, 3H), 1.25 (t, $J = 6$ Hz, 3H); m/z calculated for $\text{C}_{27}\text{H}_{23}\text{N}_5\text{O}_5$: 497.32, found: 498.04

(Z)-ethyl 3-((1H-pyrrolo[2,3-b]pyridin-3-yl)methylene)-4-amino-1-benzyl-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (8): Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 26 %, greenish yellow solid; R_f (eluent MeOH/DCM 1:3) 0.33; Mp 122-124 °C. ^1H -NMR (600 MHz, CDCl_3) δ [ppm] 9.28 (s, 1H), 8.27 (s, 1H), 8.16 (dd, $J = 18$ Hz, 2H), 7.92 (d, $J = 6$ Hz, 1H), 7.82 (d, $J = 24$ Hz, 1H), 7.08-7.05 (m, 5H), 7.01 (m, 2H), 4.99 (s, 2H), 4.19 (q, $J = 6$ Hz, 2H), 2.50 (s, 3H), 1.20 (t, $J = 6$ Hz, 3H); HRMS (ESI): m/z calculated for $\text{C}_{26}\text{H}_{23}\text{N}_5\text{O}_3$: 453.31, found: 454.07

(Z)-ethyl 4-amino-1-benzyl-3-(4-(dimethylamino)benzylidene)-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (9): Crude product further purified with DCM:MeOH (9:1) over silica gel column. Yield 16 %, white solid; R_f (eluent MeOH/DCM 1:3) 0.31; Mp 132 - 134 °C. ^1H -NMR: (600 MHz, CDCl_3) δ 9.32 (s, 1H), 8.13 (d, $J = 12$ Hz, 1H), 7.85 (s, 1H), 7.53 (d, $J = 6$ Hz, 2H), 7.31-7.7.28 (m, 3H), 7.20-7.19 (m, 3H), 6.72 (d, $J = 6$ Hz, 2H), 5.09 (s, 2H), 4.36 (q, $J = 6$ Hz, 2H), 3.05 (s, 6H), 2.71 (s, 3H), 1.40 (t, $J = 6$ Hz, 3H); ^{13}C -NMR (150 MHz, CDCl_3) δ 137.64, 133.21, 128.83, 128.49, 111.34, 60.90, 45.58, 40.20, 31.02, 14.41; m/z calculated for $\text{C}_{27}\text{H}_{28}\text{N}_4\text{O}_3$: 456.31, found: 457.10

(Z)-ethyl 4-amino-3-(benzo[d][1,3]dioxol-5-ylmethylene)-1-benzyl-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (10): Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 46 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.23; Mp 92-94 °C. $^1\text{H-NMR}$ (600 MHz, DMSO- d_6) δ [ppm] 7.96 (d, J = 1.75 Hz, 1H), 7.81 (s, 1H), 7.56-7.52 (m, 2H), 7.30 (s, 2H), 6.90-6.71 (m, 5H), 6.05 (s, 2H), 5.17 (s, 2H), 4.36 (q, J = 7.5 Hz, 2H), 2.76 (s, 3H), 1.40 (t, J = 7.5 Hz, 3H); $^{13}\text{C-NMR}$ (150 MHz, DMSO- d_6) δ [ppm] 137.32, 128.39, 127.42, 110.85, 107.80, 101.52, 60.80, 25.52, 13.92; HRMS (ESI): m/z calculated for $\text{C}_{26}\text{H}_{23}\text{N}_3\text{O}_5$: 457.99, found: 458.06

(Z)-ethyl 4-amino-1-(4-fluorobenzyl)-6-methyl-3-((5-nitro-1H-indol-3-yl)methylene)-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (11): Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 46 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.23; Mp 108 - 110 °C. $^1\text{H NMR}$ (600 MHz, DMSO- d_6) δ [ppm] 12.6 (br, 1H), 9.44 (s, 1H), 9.170 (s, 1H), 8.16-8.18 (d, J = 8.1 Hz, 1H), 8.01 (s, 1H), 7.84-7.86 (d, J = 8.1 Hz, 1H), 7.38-7.42 (t, J = 8.2 Hz, 2H), 7.12-7.16 (t, J = 7.8 Hz, 2H), 7.07 (s, 2H), 5.01 (s, 2H), 4.36 (q, J = 7.5 Hz, 2H), 2.76 (s, 3H), 1.40 (t, J = 7.5 Hz, 3H); HRMS (ESI): m/z calculated for $\text{C}_{27}\text{H}_{22}\text{FN}_5\text{O}_5$: 515.16, found: $[\text{M}+\text{H}]^+$ 516.16

(Z)-ethyl 3-((1H-pyrrolo[2,3-b]pyridin-3-yl)methylene)-4-amino-1-(4-fluorobenzyl)-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (12): Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 26 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.23; Mp 106 - 108 °C. $^1\text{H NMR}$ (600 MHz, DMSO- d_6) δ [ppm] 12.6 (br, 1H) 9.40 (s, 1H), 8.57 - 8.46 (m, 1H), 8.36 (d, J = 3.2 Hz, 1H), 7.96 (s, 1H), 7.41 (s, 2H), 7.29 (s, 1H), 7.15 (t, J = 8.4 Hz, 2H), 6.95 (s, 2H), 5.01 (s, 2H), 4.49 - 4.26 (m, 2H), 3.26 (s, 3H), 1.36 (t, J = 6.8 Hz, 3H); HRMS (ESI): m/z calculated for $\text{C}_{26}\text{H}_{22}\text{FN}_5\text{O}_3$: 471.16, found: $[\text{M}+\text{H}]^+$ 472.51

(Z)-Ethyl-3-((1H-pyrrol-2-yl)methylene)-4-amino-1-(4-fluorobenzyl)-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (13): Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 35 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.23; Mp 71-72 °C. $^1\text{H-NMR}$ (600 MHz, CDCl_3) δ [ppm] 13.36 (s, 1H), 7.59 (dd, J = 6 Hz, 2H), 7.39 (s, 1H), 7.37 (s, 1H), 7.10 (dd, J = 6 Hz, 2H), 6.83 (s, 1H), 6.49 (d, J = 6 Hz, 1H), 6.34 (s, 2H), 5.21 (s, 2H), 4.53 (q, J = 6 Hz, 2H), 2.83 (s, 3H), 1.55 (t, J = 6 Hz, 3H); $^{13}\text{C-NMR}$ (150 MHz, CDCl_3) δ [ppm] 133.08, 130.40, 130.35, 126.55, 125.20, 115.52, 111.81, 61.14, 41.89, 27.14, 14.42; HRMS (ESI): m/z calculated for $\text{C}_{23}\text{H}_{21}\text{N}_4\text{O}_3\text{F}$: 420.26, found: 421.05

(Z)-ethyl 4-amino-1-(4-fluorobenzyl)-6-methyl-3-((5-nitro-1-(3-(pyrrolidin-1-yl)propyl)-1H-indol-3-yl)methylene)-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (14): Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 23 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.23; Mp 12 - 114 °C. ^1H NMR (600 MHz, DMSO- d_6) δ [ppm] 9.44 (s, 1H), 9.170 (s, 1H), 8.16-8.18 (d, J = 8.1 Hz, 1H), 8.01 (s, 1H), 7.84-7.86 (d, J = 8.1 Hz, 1H), 7.38-7.42 (t, J = 8.2 Hz, 2H), 7.12-7.16 (t, J = 7.8 Hz, 2H), 7.07 (s, 2H), 5.01 (s, 2H), 4.45-4.49 (t, 2H, J = 8.0 Hz), 4.33-4.39 (m, 3H), 2.45-2.42 (m, 6H), 1.99 -2.02 (m, 2H), 1.71 (br, 4H), 1.35-1.37 (m, 3H), 1.25-1.27 (m, 2H); ^{13}C NMR (151 MHz, DMSO- d_6) δ [ppm] 175.16, 173.16, 168.06, 166.05, 160.57, 156.99, 153.49, 147.87, 141.87, 138.72, 138.55, 133.69, 129.46, 125.91, 119.64, 117.59, 116.47, 115.08, 112.29, 111.08, 108.02, 98.02, 60.55, 53.08, 51.86, 44.33, 40.83, 28.53, 25.35, 23.05, 13.08; HRMS (ESI): m/z calculated for $\text{C}_{34}\text{H}_{35}\text{FN}_6\text{O}_5$; ESI: $[\text{M}+\text{H}]^+$ peak 627

(Z)-ethyl 1-(4-fluorobenzyl)-3-(imidazo[1,5-a]pyridin-3-ylmethylene)-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (15): Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 41 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.23; Mp 102 - 104 °C. ^1H NMR (600 MHz, DMSO- d_6): δ [ppm] 10.16 (s, 1H), 8.36 (d, J = 9.0 Hz, 1H), 8.22 - 8.16 (m, 1H), 7.99 - 7.62 (m, 2H), 7.38 - 7.27 (m, 2H), 7.24 - 7.10 (m, 2H), 6.88 - 6.71 (m, 2H), 6.22 (dd, J = 12.9, 6.2 Hz, 1H), 4.12 (q, J = 7.1 Hz, 2H), 3.61 (s, 2H), 2.04 (s, 3H), 1.26 (t, J = 7.1 Hz, 3H); HRMS (ESI): m/z calculated for $\text{C}_{26}\text{H}_{21}\text{FN}_4\text{O}_3$; MS (MALDI+) m/z = 457.17794 $[\text{M}+\text{H}]^+$

(Z)-ethyl 4-amino-6-methyl-3-((5-nitro-1-(3-(pyrrolidin-1-yl)propyl)-1H-indol-3-yl)methylene)-2-oxo-1-(4-(trifluoromethyl)benzyl)-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (16): Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 66 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.23; Mp 122-124 °C. ^1H NMR (600 MHz, DMSO- d_6) δ [ppm] 9.44 (s, 1H), 9.170 (s, 1H), 8.16-8.18 (d, J = 8.1 Hz, 1H, Aromatic H), 8.01 (s, 1H, Aromatic H), 7.84-7.86 (d, J = 8.1 Hz, 1H-Aromatic), 7.38-7.42 (t, J = 8.2 Hz, 2H, Aromatic H), 7.12-7.16 (t, J = 7.8 Hz, 2H, Aromatic), 7.07 (s, 2H), 5.01 (s, 2H), 4.45-4.49 (t, 2H, J = 8.0 Hz), 4.33-4.39 (m, 3H), 2.45-2.42 (m, 6H), 1.99 -2.02 (m, 2H), 1.71 (br, 4H), 1.35-1.37 (m, 3H), 1.25-1.27 (m, 2H), 1.06-1.08 (m, 1H); ^{13}C NMR (151 MHz, DMSO- d_6) δ [ppm] 175.16, 173.16, 168.06, 166.05, 160.57, 156.99, 153.49, 147.87, 141.87, 138.72, 138.55, 133.69, 129.46, 125.91, 119.64, 117.59, 116.47, 115.08, 112.29, 111.08, 108.02, 98.02, 60.55, 53.08, 51.86, 44.33, 40.83, 28.53, 25.35, 23.05, 13.08; HRMS (ESI): m/z calculated for $\text{C}_{34}\text{H}_{35}\text{FN}_6\text{O}_5$; ESI: $[\text{M}+\text{H}]^+$ peak 627

(Z)-ethyl 3-((1H-pyrrol-2-yl)methylene)-4-amino-6-methyl-2-oxo-1-(4-(trifluoromethyl)benzyl)-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (17):

Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 66 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.39; Mp 86-88 °C. $^1\text{H-NMR}$ (600 MHz, CDCl_3) δ [ppm] 13.19 (s, 1H), 7.56 (s, 5H), 7.14 (s, 1H), 6.73 (s, 1H), 6.38 (d, $J=6$ Hz, 2H), 6.24 (s, 2H), 5.19 (s, 2H), 4.41 (q, $J=6$ Hz, 2H), 2.69 (s, 3H), 1.43 (t, $J=6$ Hz, 3H); $^{13}\text{C-NMR}$ (150 MHz, CDCl_3) δ [ppm] 133.08, 130.40, 130.35, 126.55, 125.20, 115.52, 111.81, 61.14, 41.89, 27.14, 14.42 ppm; HRMS (ESI): m/z calculated for $\text{C}_{24}\text{H}_{21}\text{N}_4\text{O}_3\text{F}_3$: 470.28, found: 470.99

imidazo[1,5-a]pyridine-3-carbaldehyde (19): Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 66 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.53; Mp 82-84 °C. $^1\text{H NMR}$ (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 9.97 (s, 1H), 8.66 - 8.62 (td, $J=8.6$ Hz, 1H), 8.56 (s, 1H), 8.15 - 8.11 (qd, $J=8.2$ Hz, 1H), 7.46 - 7.39 (dq, $J=7.4$ Hz, 1H), 7.12 - 7.06 (td, $J=7.1$ Hz, 1H); $^{13}\text{C NMR}$ (151 MHz, $\text{DMSO-}d_6$) δ [ppm] 183.6, 15.6, 139.4, 126.2, 123.0, 119.4, 118.8, 114.6; HRMS (ESI): m/z calculated for $\text{C}_8\text{H}_6\text{N}_2\text{O}$: ESI: 147.06 $[\text{M}+\text{H}]^+$

1-bromoimidazo[1,5-a]pyridine-3-carbaldehyde (20): Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 66 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.43; Mp 92-94 °C. $^1\text{H NMR}$ (600 MHz, $\text{DMSO-}d_6$): δ [ppm] 9.98 (s, 1H), 8.24 - 8.23 (dt, $J=8.2$ Hz, 1H), 8.04 - 8.03 (dt, $J=8.0$ Hz, 1H), 7.27 - 7.24 (qd, $J=7.26$ Hz, 1H), 6.98 - 6.95 (td, $J=7.0$ Hz, 1H); $^{13}\text{C NMR}$ (151 MHz, $\text{DMSO-}d_6$) δ 183.6, 150.6, 131.1, 126.4, 1253.4, 124.4, 118.8, 114.6; HRMS (ESI): m/z calculated for $\text{C}_8\text{H}_5\text{BrN}_2\text{O}$: MS MALDI $m/z = 224.98591$

methyl 3-(3formylimidazol[1,5-a]pyridine-1-yl)benzoate (21): Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 66 %, yellow solid; R_f (eluent MeOH/DCM 1:3) 0.63; Mp 103- 105°C. $^1\text{H NMR}$ (600 MHz, CDCl_3) δ [ppm] 10.18 (s, 1H), 8.37 (dt, $J=9.1, 1.2$ Hz, 1H), 8.26 - 8.14 (m, 1H), 7.81-7.59 (m, 4H), 7.34 - 7.28 (m, 1H), 6.85 (td, $J=6.9, 1.1$ Hz, 1H), 3.62 (d, $J=2.1$ Hz, 3H); $^{13}\text{C NMR}$ (151 MHz, $\text{DMSO-}d_6$) δ [ppm] 183.6, 165.9, 150.6, 132.9, 132.4, 131.1, 130.6, 130.1, 129.9, 126.2, 124.4, 51.5; MS (ESI) $m/z = 281.08$ $[\text{M}+\text{H}]^+$

(Z)-ethyl 1-(4-fluorobenzyl)-3-((1-(3-(methoxycarbonyl)phenyl)imidazo[1,5-a]pyridin-3-yl)methylene)-6-methyl-2-oxo-2,3-dihydro-1H-pyrrolo[2,3-b]pyridine-5-carboxylate (22): Crude product further purified with DCM: MeOH (9:1) over silica gel column. Yield 23 %, orange solid; R_f (eluent MeOH/DCM 1:3) 0.29; Mp 102-104 °C. $^1\text{H NMR}$ (600 MHz, $\text{DMSO-}d_6$) δ [ppm] 10.54 (dd, $J=15.6, 6.6$ Hz, 1H), 8.12 (dd, $J=16.1, 8.5$ Hz, 2H), 8.05 - 7.98 (m, 2H), 7.92 -

7.76 (m, 2H), 7.55 (t, $J = 16.9$ Hz, 1H), 7.44 - 7.36 (m, 2H), 7.34 - 7.11 (m, 1H), 6.97 (dd, $J = 15.5$, 8.4 Hz, 1H), 6.54 - 6.43 (m, 2H), 4.96 (s, 2H), 4.28 (q, $J = 7.1$ Hz, 2H), 3.61 - 3.53 (m, 3H), 2.47 (s, 3H), 1.29 (t, $J = 7.1$ Hz, 3H); HRMS (ESI): m/z calculated for $C_{34}H_{27}FN_4O_5$: MS MALDI $m/z = 591.98$ $[M+H]^+$

Fragment screening with Thiazole Displacement Assays

K_d determination of Thiazole orange

The FID assay was performed using the procedure described earlier.⁵⁹ Dissociation constant (K_d) for TO binding to *c-MYC*, *c-KIT1*, *h-TELO*, *BCL2* ds24 & ds27 DNA has been determined and reported in (Table 3 and Fig. 47 of chapter 2) using the following conditions 0.25 μ M DNA, 0.5 μ M Thiazole Orange, 10 % DMSO, 20 mM Na cacoc, 140 mM KCl, pH 7 (25 μ L/well). All fragment molecules were 95% pure and obtained from commercial sources or were synthesized in-house. For assay optimization sufficient negative and positive controls were used, DMSO only (10 % v/v) wells, which contained no small molecule, were used as a negative control, while positive control wells consisted target DNA and the intercalator TO.

For screening, 1.25 μ L of each fragment from its original 100 mM DMSO stock plate was transferred to a 384 well assay plate (low volume flat bottom black NBS treated, Corning 3820) with each 384 well plate containing 180 fragments with negative and positive controls. To the fragments were added 23.75 μ L of the annealed *MYC* oligo containing 0.25 μ M DNA, 0.5 μ M Thiazole Orange, 20 mM Na cacoc, 140 mM KCl, pH 7 and the plate incubated for 30 min at room temperature. The fluorescent measurements were taken at 25 °C using an excitation filter of 510 nm and an emission filter of 540 nm using an Infinite 200 Pro Micro Plate Reader (Tecan i-control). Experiments were performed in triplicate and were repeated three times. The fragments were ranked according to their TO displacement effect and those fragments showing $\geq 95\%$ displacement were subjected to a dose response, under the original screening conditions. The 50% displacement value (DC_{50}) and subsequent K_i , were calculated for ligands from the fluorescence intensity (F) at the emission maxima, using the following Equation (1) and Equation (2) respectively.

Equation (1)

$$\text{TO-Displacement \% FID} = 100 + [100 + F/F_0]$$

$$F = F_{(\text{Ligand} + \text{DNA} + \text{TO})} - F_{(\text{Buffer} + \text{DNA})} - F_{(\text{Ligand} + \text{DNA})}$$

$$F_0 = F_{(\text{DNA} + \text{TO})} - F_{(\text{Buffer} + \text{TO})}$$

Equation 2.....

$$K_i = \frac{DC_{50}}{1 + \frac{[TO]}{K_d}}$$

Where, K_d is the dissociation binding constant of TO with G4 DNA and duplex DNA sequences; K_i is inhibition constant of ligand with G4 DNA; $[TO]$ = concentration of TO; $[DC_{50}]$ = Concentration of ligand for 50% displacement of bound TO.

¹D-NMR Titration

All NMR measurements were recorded at 600MHz at 298K. Samples contained 0.1mM *c-MYC* in 25mM KP_i , 70mM KCl, pH 7 0.025mM DSS and 10% DMSO- d_6 . Due to poor solubility of the ligands in the absence of DNA, a single sample for each DNA: ligand ratio was prepared to keep the DMSO-concentration constant. For concentration the titration, the DNA was provided as a 100 μ M solution in 25 mM Tris·HCl buffer (pH 7.4) with 100 mM KCl in 10% d_6 -DMSO/90% H_2O . Small amounts of the ligand stock solution in 100% d_6 -DMSO were added directly into the NMR tube (10 % d_6 -DMSO at the end of the titration). 2, 2-dimethyl-2-silapentane-5-sulphonate (DSS) was used as internal reference. Watergate W5 pulse sequence with gradients (37) was used for water suppression.

Cells and culture conditions

HeLa cells (human cervical cancer cells) were grown in Dulbecco's modified Eagle's medium (DMEM, Gibco® Life Technologies) supplemented with 10% fetal bovine serum (PAN Biotech), and 1% penicillin/streptomycin (Invitrogen). HeLa cells were maintained in 25 or 75 cm^2 flasks in a humidified atmosphere containing 5% CO_2 , at 37°C. For cell passaging, biotase (Invitrogen) was used for detaching the cells and the standard Neubauer chamber for counting the cells and seeding accordingly.

Cell proliferation assay

HeLa cells were used for performing the cell viability assays. A day before the assay cells were seeded in a 96-well plate (Nunclon 96 Flat Bottom Transparent), with a density of 1000 cells per

well in 100 μ l. Ligand dilutions were prepared in DMSO, vortexed and stores in -20 °C dilution range varied between 0, 3, 10, 30, 100, 300, 1000, 3000, 9000, 30000, 100000 nm depending upon the potency of the respective ligands. The working concentrations were made in the culture medium and added to the cells, to a final DMSO concentration of 0.1 %. Cells treated with the ligands were placed back in the 5% CO₂ incubator. After 3 days of ligand treatment cell viability assay was performed using Alamar Blue reagent (Thermo Scientific) following the instructions of the manufacturer. The fluorescence signal of the Alamar Blue was measured with infinite 200 Pro Micro Plate Reader (Tecan i-control) with the excitation wavelength 540 nm & emission wavelength 590 nm. Ligand treatments were performed in triplicates and were repeated three times. The cytotoxicity was evaluated based on the percentage of cell survival in a dose-dependent manner with regard to the negative un-treated control. Percentage cell viability was calculated by using the following equation:

$$\% \text{ of cell viability} = \frac{\text{O.D. of treated cells}}{\text{O.D. of untreated cells}} \times 100$$

The final IC₅₀ values were calculated by using the GraphPad Prism 6.0 software.

c-MYC expression and immuno blotting

For monitoring the expression levels of c-MYC after ligand, treatment Western blot analysis was performed. Cells treated for 24 hours with different ligands in varying concentrations were harvested with biotase, and suspended in PBS buffer containing EDTA free protease inhibitor. Followed by cell lysis with ultrasonicator and quantification of the cell lysate with Roti-Quant. Equal amounts of the cell lysates were loaded onto the pre-casted NuPAGE 4-12% Bis-Tris Gels and with the MES SDS running buffer (50 mM MES, 50 mM Tris base, 0.1% SDS, 1 mM EDTA, pH 7.3) gel electrophoresis was performed. Proteins on the gel were transferred onto the PVDF membrane (activated by brief incubation in methanol) for immunoblotting. After the protein transfer, the membrane was blocked for an hour in 5% non-fat milk in TBS buffer. The blot was then probed either with c-MYC rabbit monoclonal antibody (1:2,000, #5605, Cell signaling) in 1X TBS with 5% BSA or with anti-GAPDH primary monoclonal antibody (1:2,000, Cell signaling) in 1X PBS with 5% non-fat dry milk, and incubated overnight at 4°C, on a gel rocking platform shaker. To get rid off the unbound primary antibody, the membrane was washed three times with PBST (PBS with 0.05% Tween 20) for 10 min/wash. Further, the membrane was probed either with horseradish peroxidase (HRP) -conjugated affinipure goat anti-rabbit IgG (H+L) sera in PBS (1:5,000, Dianova) or with HRP-conjugated goat anti-mouse secondary antibody in PBS (1:5,000,

Dianova) respectively, for an hour. Followed by three washing steps with PBS Tween (0.05%) (10 min/wash). Subsequently, the membrane was developed with Pierce™ ECL chemiluminescent HRP substrate, and the chemiluminescence signals were detected using digital Lumi-imager (Roche). Relative intensities of the ECL signals were then determined using Image J software.

Cell Cycle Analysis

Exponentially growing HeLa cells were seeded in 12-well plates at a density of 2×10^5 cells/ml and allowed to grow in DMEM complete media (Thermofisher) for 24 hours. Cells were then treated with ligands ($5 \mu\text{M}$) in fresh DMEM media for 24 hours. Drug media was then washed with 1x PBS, trypsinized with 1x trypsin (Thermofisher), pelleted down by centrifugation at 1000 rpm and washed with 1x PBS. 400 μL permeabilisation buffer (0.1 M Phosphate/citrate (tablets), 0.15 M NaCl, 5 mM EDTA, 0.5% BSA, 0.02% Saponin) were added to 2×10^5 cells, mixed gently with 3 μg 7-ADD (BD Biosciences) and incubated for 30 min at 4°C in the dark. Cells were then washed once PBS and resuspended in 400 μL PBS for FACS analysis. Cell distribution among cell cycle phases were acquired on BD LSR Fortessa analyser (BD Biosciences, San Diego, CA, USA) flow cytometer (Yellow Green filter 561nm). A total of 10,000 events were recorded.

Molecular Modelling

The binding sites at the two ends of the *c-MYC* G4 were defined by using the NMR structure of the 2:1 Quindoline: *c-MYC* G4 complex (PDB ID 2L7V).⁵⁴ During docking, the DNA was fixed while the ligand was flexible. Before running docking structure of ligands were optimized by Gaussian 3.0 tool with semi empirical PM3 force field. The docking simulations were performed by Autodock 4.2 tool.⁶⁰

4.16 References

- (1) Neidle, S. Quadruplex Nucleic Acids as Targets for Anticancer Therapeutics. *Nature Reviews Chemistry*. 2017. <https://doi.org/10.1038/s41570-017-0041>.
- (2) Lam, E. Y. N.; Beraldi, D.; Tannahill, D.; Balasubramanian, S. G-Quadruplex Structures Are Stable and Detectable in Human Genomic DNA. *Nat. Commun.* **2013**. <https://doi.org/10.1038/ncomms2792>.
- (3) Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S. High-Throughput Sequencing of DNA G-Quadruplex Structures in the Human Genome. *Nat. Biotechnol.* **2015**. <https://doi.org/10.1038/nbt.3295>.
- (4) Hänsel-Hertsch, R.; Beraldi, D.; Lensing, S. V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; et al. G-Quadruplex Structures Mark Human Regulatory Chromatin. *Nat. Genet.* **2016**. <https://doi.org/10.1038/ng.3662>.
- (5) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* **2013**. <https://doi.org/10.1038/nchem.1548>.
- (6) Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-Quadruplexes in Gene Promoters: A Novel Anticancer Strategy? *Nat. Rev. Drug Discov.* **2011**, *10* (4), 261–275. <https://doi.org/10.1038/nrd3428>.
- (7) Wu, Y.; Brosh, R. M. G-Quadruplex Nucleic Acids and Human Disease. *FEBS Journal*. 2010. <https://doi.org/10.1111/j.1742-4658.2010.07760.x>.
- (8) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription. *Proc. Natl. Acad. Sci.* **2002**. <https://doi.org/10.1073/pnas.182256799>.
- (9) Sneppen, K.; Zocchi, G. DNA and RNA. In *Physics in Molecular Biology*; 2010. <https://doi.org/10.1017/cbo9780511755699.005>.
- (10) Pelengaris, S.; Khan, M.; Evan, G. C-MYC: More than Just a Matter of Life and Death. *Nature Reviews Cancer*. 2002. <https://doi.org/10.1038/nrc904>.
- (11) Haider, S. Computational Methods to Study G-Quadruplex–Ligand Complexes. *Journal of the Indian Institute of Science*. 2018. <https://doi.org/10.1007/s41745-018-0083-3>.
- (12) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal Structure of Parallel Quadruplexes from Human Telomeric DNA. *Nature* **2002**. <https://doi.org/10.1038/nature755>.
- (13) Rodrigues, T.; Reker, D.; Schneider, P.; Schneider, G. Counting on Natural Products for Drug Design. *Nature Chemistry*. 2016. <https://doi.org/10.1038/nchem.2479>.

- (14) Che, T.; Chen, S. Bin; Tu, J. L.; Wang, B.; Wang, Y. Q.; Zhang, Y.; Wang, J.; Wang, Z. Q.; Zhang, Z. P.; Ou, T. M.; et al. Discovery of Novel Schizocommunin Derivatives as Telomeric G-Quadruplex Ligands That Trigger Telomere Dysfunction and the Deoxyribonucleic Acid (DNA) Damage Response. *J. Med. Chem.* **2018**. <https://doi.org/10.1021/acs.jmedchem.7b01615>.
- (15) Hosoe, T.; Nozawa, K.; Kawahara, N.; Fukushima, K.; Nishimura, K.; Miyaji, M.; Kawai, K. I. Isolation of a New Potent Cytotoxic Pigment along with Indigotin from the Pathogenic Basidiomycetous Fungus *Schizophyllum Commune*. *Mycopathologia* **1999**. <https://doi.org/10.1023/A:1007082619328>.
- (16) Uehata, K.; Kimura, N.; Hasegawa, K.; Arai, S.; Nishida, M.; Hosoe, T.; Kawai, K. I.; Nishida, A. Total Synthesis of Schizocommunin and Revision of Its Structure. *J. Nat. Prod.* **2013**. <https://doi.org/10.1021/np400263f>.
- (17) Diveshkumar, K. V.; Sakrikar, S.; Rosu, F.; Harikrishna, S.; Gabelica, V.; Pradeepkumar, P. I. Specific Stabilization of C-MYC and c-KIT G-Quadruplex DNA Structures by Indolylmethyleneindanone Scaffolds. *Biochemistry* **2016**, *55* (25), 3571–3585. <https://doi.org/10.1021/acs.biochem.6b00120>.
- (18) Shen, F. H.; Jin, J.; Li, J.; Wang, Y.; Zhu, S. H.; Lu, Y. J.; Ou, T. M.; Huang, Z. S.; Huang, M.; Huang, Z. Y. The G-Quadruplex Ligand, SYUIQ-FM05, Targets Proto-Oncogene c-Kit Transcription and Induces Apoptosis in K562 Cells. *Pharm. Biol.* **2013**. <https://doi.org/10.3109/13880209.2012.738424>.
- (19) Ali, A.; Bansal, M.; Bhattacharya, S. Ligand 5,10,15,20-Tetra(N-Methyl-4-Pyridyl)Porphine (TMPyP4) Prefers the Parallel Propeller-Type Human Telomeric G-Quadruplex DNA over Its Other Polymorphs. *J. Phys. Chem. B* **2015**. <https://doi.org/10.1021/jp505792z>.
- (20) Asamitsu, S.; Obata, S.; Yu, Z.; Bando, T.; Sugiyama, H. Recent Progress of Targeted G-Quadruplex-Preferred Ligands toward Cancer Therapy. *Molecules*. 2019. <https://doi.org/10.3390/molecules24030429>.
- (21) Hagmann, W. K. The Many Roles for Fluorine in Medicinal Chemistry. *Journal of Medicinal Chemistry*. 2008. <https://doi.org/10.1021/jm800219f>.
- (22) Lu, Y. J.; Ou, T. M.; Tan, J. H.; Hou, J. Q.; Shao, W. Y.; Peng, D.; Sun, N.; Wang, X. D.; Wu, W. Bin; Bu, X. Z.; et al. 5-N-Methylated Quindoline Derivatives as Telomeric G-Quadruplex Stabilizing Ligands: Effects of 5-N Positive Charge on Quadruplex Binding Affinity and Cell Proliferation. *J. Med. Chem.* **2008**. <https://doi.org/10.1021/jm800497p>.
- (23) Song, J. J.; Reeves, J. T.; Gallou, F.; Tan, Z.; Yee, N. K.; Senanayake, C. H. Organometallic

- Methods for the Synthesis and Functionalization of Azaindoles. *Chem. Soc. Rev.* **2007**.
<https://doi.org/10.1039/b607868k>.
- (24) Debenham, S. D.; Chan, A.; Liu, K.; Price, K.; Wood, H. B. A Convenient One-Pot Synthesis of 4-, 6-, and 7-Azaindoles from Aminopyridines. *Tetrahedron Lett.* **2005**.
<https://doi.org/10.1016/j.tetlet.2005.02.011>.
- (25) Zhang, Z.; Yang, Z.; Meanwell, N. A.; Kadow, J. F.; Wang, T. A General Method for the Preparation of 4- and 6-Azaindoles. *J. Org. Chem.* **2002**.
<https://doi.org/10.1021/jo0111614>.
- (26) Gallou, F.; Reeves, J. T.; Tan, Z.; Song, J. J.; Yee, N. K.; Campbell, S.; Jones, P. J.; Senanayake, C. H. Regioselective Halogenation of 6-Azaindoles: Efficient Synthesis of 3-Halo-2,3-Disubstituted-6-Azaindole Derivatives. *Synlett* **2005**. <https://doi.org/10.1055/s-2005-872680>.
- (27) Estel, L.; Marsais, F.; Quéguiner, G. Metalation/SRN1 Coupling in Heterocyclic Synthesis. A Convenient Methodology for Ring Functionalization. *J. Org. Chem.* **1988**.
<https://doi.org/10.1021/jo00247a016>.
- (28) Larghi, E. L.; Amongero, M.; Bracca, A. B. J.; Kaufman, T. S. The Intermolecular Pictet-Spengler Condensation with Chiral Carbonyl Derivatives in the Stereoselective Syntheses of Optically-Active Isoquinoline and Indole Alkaloids. *Arkivoc* **2005**.
<https://doi.org/10.3998/ark.5550190.0006.c09>.
- (29) Fang, Y.-Q.; Yuen, J.; Lautens, M. A General Modular Method of Azaindole and Thienopyrrole Synthesis via Pd-Catalyzed Tandem Couplings of Gem-Dichloroolefins. *ChemInform* **2007**. <https://doi.org/10.1002/chin.200750161>.
- (30) Rodriguez, A. L.; Koradin, C.; Dohle, W.; Knochel, P. Versatile Indole Synthesis by a 5-Endo-Dig Cyclization Mediated by Potassium or Cesium Bases. *Angew. Chemie - Int. Ed.* **2000**.
[https://doi.org/10.1002/1521-3773\(20000717\)39:14<2488::AID-ANIE2488>3.0.CO;2-E](https://doi.org/10.1002/1521-3773(20000717)39:14<2488::AID-ANIE2488>3.0.CO;2-E).
- (31) Nazaré, M.; Schneider, C.; Lindenschmidt, A.; Will, D. W. A Flexible, Palladium-Catalyzed Indole and Azaindole Synthesis by Direct Annulation of Chloroanilines and Chloroaminopyridines with Ketones. *Angew. Chemie - Int. Ed.* **2004**.
<https://doi.org/10.1002/anie.200460122>.
- (32) Ujjainwalla, F.; Warner, D. Synthesis of 5-, 6- and 7-Azaindoles via Palladium-Catalyzed Heteroannulation of Internal Alkynes. *Tetrahedron Lett.* **1998**.
[https://doi.org/10.1016/S0040-4039\(98\)01069-7](https://doi.org/10.1016/S0040-4039(98)01069-7).

- (33) Cacchi, S.; Fabrizi, G.; Parisi, L. M. The Aminopalladation-Reductive Elimination Process as a Tool for the Solution-Phase Synthesis of 2,3-Disubstituted Azaindole Libraries. *J. Comb. Chem.* **2005**. <https://doi.org/10.1021/cc049821g>.
- (34) Young, I. S.; Kerr, M. A. Three-Component Homo 3 + 2 Dipolar Cycloaddition. A Diversity-Oriented Synthesis of Tetrahydro-1,2-Oxazines and FR900482 Skeletal Congeners. *Org. Lett.* **2004**. <https://doi.org/10.1021/ol0362919>.
- (35) Willis, M. C.; Brace, G. N.; Holmes, I. P. Palladium-Catalyzed Tandem Alkenyl and Aryl C-N Bond Formation: A Cascade N-Annulation Route to 1-Functionalized Indoles. *Angew. Chemie - Int. Ed.* **2005**. <https://doi.org/10.1002/anie.200461598>.
- (36) Irie, T.; Sawa, M. 7-Azaindole: A Versatile Scaffold for Developing Kinase Inhibitors. *Chem. Pharm. Bull.* **2018**. <https://doi.org/10.1248/cpb.c17-00380>.
- (37) Yuan, R.; Kay, A.; Berg, W. J.; Lebwohl, D. Targeting Tumorigenesis: Development and Use of MTOR Inhibitors in Cancer Therapy. *J. Hematol. Oncol.* **2009**. <https://doi.org/10.1186/1756-8722-2-45>.
- (38) Raymond, E.; Dahan, L.; Raoul, J. L.; Bang, Y. J.; Borbath, I.; Lombard-Bohas, C.; Valle, J.; Metrakos, P.; Smith, D.; Vinik, A.; et al. Sunitinib Malate for the Treatment of Pancreatic Neuroendocrine Tumors. *N. Engl. J. Med.* **2011**. <https://doi.org/10.1056/NEJMoa1003825>.
- (39) Kaley, T. J.; Wen, P.; Schiff, D.; Ligon, K.; Haidar, S.; Karimi, S.; Lassman, A. B.; Nolan, C. P.; De Angelis, L. M.; Gavrilovic, I.; et al. Phase II Trial of Sunitinib for Recurrent and Progressive Atypical and Anaplastic Meningioma. *Neuro. Oncol.* **2015**. <https://doi.org/10.1093/neuonc/nou148>.
- (40) Ebos, J. M. L.; Lee, C. R.; Cruz-Munoz, W.; Bjarnason, G. A.; Christensen, J. G.; Kerbel, R. S. Accelerated Metastasis after Short-Term Treatment with a Potent Inhibitor of Tumor Angiogenesis. *Cancer Cell* **2009**. <https://doi.org/10.1016/j.ccr.2009.01.021>.
- (41) Gandhi, L.; Rodríguez-Abreu, D.; Gadgeel, S.; Esteban, E.; Felip, E.; De Angelis, F.; Domine, M.; Clingan, P.; Hochmair, M. J.; Powell, S. F.; et al. Pembrolizumab plus Chemotherapy in Metastatic Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **2018**. <https://doi.org/10.1056/NEJMoa1801005>.
- (42) Sudta, P.; Kirk, N.; Bezos, A.; Gurlica, A.; Mitchell, R.; Weber, T.; Willis, A. C.; Prabpai, S.; Kongsaree, P.; Parish, C. R.; et al. Synthesis, Structural Characterisation, and Preliminary Evaluation of Non-Indolin-2-One-Based Angiogenesis Inhibitors Related to Sunitinib (Sutent®). *Aust. J. Chem.* **2013**. <https://doi.org/10.1071/CH13219>.

- (43) Zhang, J.; Shen, W.; Li, X.; Chai, Y.; Li, S.; Lv, K.; Guo, H.; Liu, M. Synthesis and Antitumor Activity of 5-Bromo-7-Azaindolin-2-One Derivatives Containing a 2,4-Dimethyl-1H-Pyrrole-3-Carboxamide Moiety. *Molecules* **2016**. <https://doi.org/10.3390/molecules21121674>.
- (44) Shah, S.; Lee, C.; Choi, H.; Gautam, J.; Jang, H.; Kim, G. J.; Lee, Y. J.; Chaudhary, C. L.; Park, S. W.; Nam, T. G.; et al. 5-Hydroxy-7-Azaindolin-2-One, a Novel Hybrid of Pyridinol and Sunitinib: Design, Synthesis and Cytotoxicity against Cancer Cells. *Org. Biomol. Chem.* **2016**. <https://doi.org/10.1039/c6ob00406g>.
- (45) Jarrahpour, A.; Zarei, M. The Vilsmeier Reagent: A Useful and Versatile Reagent for the Synthesis of 2-Azetidinones. *Tetrahedron* **2009**. <https://doi.org/10.1016/j.tet.2009.02.005>.
- (46) Egger, J.; Weckerle, C.; Cutting, B.; Schwaradt, O.; Rabbani, S.; Lemme, K.; Ernst, B. Nanomolar E-Selectin Antagonists with Prolonged Half-Lives by a Fragment-Based Approach. *J. Am. Chem. Soc.* **2013**. <https://doi.org/10.1021/ja4029582>.
- (47) Auberson, Y. P.; Troxler, T.; Zhang, X.; Yang, C. R.; Feuerbach, D.; Liu, Y. C.; Lagu, B.; Perrone, M.; Lei, L.; Shen, X.; et al. From Ergolines to Indoles: Improved Inhibitors of the Human H3 Receptor for the Treatment of Narcolepsy. *ChemMedChem* **2015**. <https://doi.org/10.1002/cmdc.201402418>.
- (48) Oukli, N.; Comesse, S.; Chafi, N.; Oulyadi, H.; Daïch, A. Unexpected Reversible Nitrogen Atom Transfer in the Synthesis of Polysubstituted Imides and 7-Aza-Hexahydroindolones via Enaminonitrile γ -Lactams. *Tetrahedron Lett.* **2009**. <https://doi.org/10.1016/j.tetlet.2009.01.072>.
- (49) Riesgo, E. C.; Jin, X.; Thummel, R. P. Introduction of Benzo[h]Quinoline and 1,10-Phenanthroline Subunits by Friedländer Methodology. *J. Org. Chem.* **1996**. <https://doi.org/10.1021/jo952164h>.
- (50) Siegel, J. S.; Tobe, Y.; Merino, P.; Collier, S. J.; Singh, S. K.; Zysman-Colman, E.; Alonso, D. A.; Nájera, C.; Abou-Hadeed, K.; Hansen, H.-J.; et al. Knoevenagel Condensation. In *Monocyclic Arenes, Quasiarenes, and Annulenes*; 2010. <https://doi.org/10.1055/sos-sd-045-00621>.
- (51) Khan, F. A.; Dash, J.; Satapathy, R.; Upadhyay, S. K. Hydrotalcite Catalysis in Ionic Liquid Medium: A Recyclable Reaction System for Heterogeneous Knoevenagel and Nitroaldol Condensation. *Tetrahedron Lett.* **2004**. <https://doi.org/10.1016/j.tetlet.2004.02.103>.
- (52) Hou, J. Q.; Tan, J. H.; Wang, X. X.; Chen, S. Bin; Huang, S. Y.; Yan, J. W.; Chen, S. H.; Ou, T. M.; Luo, H. Bin; Li, D.; et al. Impact of Planarity of Unfused Aromatic Molecules on G-

- Quadruplex Binding: Learning from Isaindigotone Derivatives. *Org. Biomol. Chem.* **2011**. <https://doi.org/10.1039/c1ob05884c>.
- (53) Li, Z.; Tan, J. H.; He, J. H.; Long, Y.; Ou, T. M.; Li, D.; Gu, L. Q.; Huang, Z. S. Disubstituted Quinazoline Derivatives as a New Type of Highly Selective Ligands for Telomeric G-Quadruplex DNA. *Eur. J. Med. Chem.* **2012**. <https://doi.org/10.1016/j.ejmech.2011.10.057>.
- (54) Dai, J.; Carver, M.; Hurley, L. H.; Yang, D. Solution Structure of a 2:1 Quindoline-c-MYC G-Quadruplex: Insights into G-Quadruplex-Interactive Small Molecule Drug Design. *J. Am. Chem. Soc.* **2011**. <https://doi.org/10.1021/ja205646q>.
- (55) Wirmer-Bartoschek, J.; Bendel, L. E.; Jonker, H. R. A.; Grün, J. T.; Papi, F.; Bazzicalupi, C.; Messori, L.; Gratteri, P.; Schwalbe, H. Solution NMR Structure of a Ligand/Hybrid-2-G-Quadruplex Complex Reveals Rearrangements That Affect Ligand Binding. *Angew. Chemie - Int. Ed.* **2017**. <https://doi.org/10.1002/anie.201702135>.
- (56) Das, T.; Panda, D.; Saha, P.; Dash, J. Small Molecule Driven Stabilization of Promoter G-Quadruplexes and Transcriptional Regulation of c-MYC. *Bioconjug. Chem.* **2018**. <https://doi.org/10.1021/acs.bioconjchem.8b00338>.
- (57) O'Brien, J.; Wilson, I.; Orton, T.; Pognan, F. Investigation of the Alamar Blue (Resazurin) Fluorescent Dye for the Assessment of Mammalian Cell Cytotoxicity. *Eur. J. Biochem.* **2000**. <https://doi.org/10.1046/j.1432-1327.2000.01606.x>.
- (58) Nasiri, H. R.; Bell, N. M.; McLuckie, K. I. E.; Husby, J.; Abell, C.; Neidle, S.; Balasubramanian, S. Targeting a C-MYC G-Quadruplex DNA with a Fragment Library. *Chem. Commun.* **2014**. <https://doi.org/10.1039/c3cc48390h>.
- (59) Tran, P. L. T.; Largy, E.; Hamon, F.; Teulade-Fichou, M. P.; Mergny, J. L. Fluorescence Intercalator Displacement Assay for Screening G4 Ligands towards a Variety of G-Quadruplex Structures. *Biochimie* **2011**. <https://doi.org/10.1016/j.biochi.2011.05.011>.
- (60) Forli, W.; Halliday, S.; Belew, R.; Olson, A. AutoDock Version 4.2. *Citeseer* **2012**.

Acknowledgments

I would like to thank my Ph.D. supervisor Professor Dr. Harald Schwalbe for the opportunity to work within his group. My project was very interesting and enabled me to learn new skills with his support, guidance, and encouragement. I am also thankful for the administrative staff of Goethe University, Frankfurt am Main for all the necessary support. Thank you Kerstin and Anna Paulus for your support in the beginning of my Ph.D.

The past and present members of the Schwalbe group has been supportive and helpful throughout my time in the group and I would like to thank them all. I'd like to thank Dr. Julia, Sam, Elke and Kerstin for all of her help and support in the lab. A special thanks go out to Dr. Irene, Isam, Andreas, Robin, Betül and Alix who has been a great friend during my time in the Ph.D. group. I am also thankful to the Goethe university analytical facility staff for their support and service. Special thanks to Horizon room Linda, Andi, Sven, György and Gerd for bearing my quite stay in the group for a long time.

I am very grateful for the support and kindness that I received from my family, who have always encouraged and motivated me to be the best that I can be. I am thankful to my niece and nephews (Anu, adu, and Vallu) who have put up with my moods during a stress-filled doctorate research years. Special acknowledgment to my brave and courageous parents for all the support. I am grateful to my late sister in law for all beautiful memories during my doctorate research you will always stay in my heart forever.

Finally, I am grateful to the Goethe University Frankfurt am Main and the EU ITN-projects: Hyperpol and Haemmetabolomics, DFG-Normalverfahren G-quadruplex and the state of Hesse for BMRZ for funding my research.

I'd especially like to thank Dr. Santosh Gande and Dr. Sridhar for their incredible contribution in all my research projects. My sincere gratitude to Robin, Melissa, and Isam for helping me with German translation of my thesis summary.

Finally, I would like to thank everybody who was important to the successful realization of the thesis, as well as expressing my apology that I could not mention personally one by one. If anything contained in this thesis is deemed meritorious, it most assuredly can be attributed to these individuals; I alone take credit for the less than meritorious content.

Thank You All!!!!!!

Publications

Research publications during Ph.D. Studies

1. V.D Nimbarte, J.B Wirmer-Bartoschek, S.L. Gande, I. Alshamleh, H.R Nasiri, F Schnütgen, H Serve, H. Schwalbe; Novel Indole Derivatives as potent and selective *c-MYC* promoter G-Quadruplex binders down-regulating *c-MYC* Expression in Human Cancer Cells; 2020; under preparation.
2. I Elamri, M Radloff, K.F Hohmann, V.D Nimbarte, H. R. Nasiri, M Bolte, S Safarian, H Michel, H Schwalbe; Synthesis and Biological Screening of New Lawson Derivatives as Selective Substrate-Based Inhibitors of Cytochrome bo₃ Ubiquinol Oxidase from *Escherichia coli*; *ChemMedChem*; 2020; doi.org/10.1002/cmdc.201900707
3. I Saidi, V.D. Nimbarte, H Schwalbe, P Waffo-Teguo, A.H Harrath, L Mansour; Anti-tyrosinase, anti-cholinesterase and cytotoxic activities of extracts and phytochemicals from the Tunisian *Citharexylum spinosum* L: Molecular docking and SAR analysis; *Bioorganic Chemistry*; 2020; Manuscript number: BIOORG_2019_1953 (under major revision)

Research publications from previous experience

4. S Chortani, V.D Nimbarte, M Horchani, H.B Jannet, A Romdhane.; Synthesis, biological evaluation and molecular docking analysis of novel benzopyrimidinone derivatives as potential anti-tyrosinase agents; *Bioorganic Chemistry*; 2019; doi.org/10.1016/j.bioorg.2019.103270
5. A Ibrahim, V.D Nimbarte; Isocostic Acid, a Promising Bioactive Agent from the Essential Oil of *Inula viscosa* (L.): Insights from Drug Likeness Properties, Molecular Docking and SAR Analysis; *Chemistry and Biodiversity*, 2019; doi: 10.1002/cbdv.201800648
6. M Debbabi, V.D Nimbarte; Design and synthesis of novel potent anticoagulant and anti-tyrosinase pyranopyrimidines and pyranotriazolopyrimidines: Insights from molecular docking and SAR analysis; *Bioorganic Chemistry*, 2018, doi: 10.1016/j.bioorg.2018.10.004
7. M.P.N Rao, V.D Nimbarte; Synthesis of Imidazo-thiadiazole linked Indolinone Conjugates and evaluated their Microtubule network disrupting and Apoptosis Inducing ability; *Bioorganic Chemistry*, 2017, doi: 10.1016/j.bioorg.2017.11.021
8. A.O Aliyene, V.D Nimbarte; Bi(OTf)₃-Catalysed Access to 2,3-Substituted Isoindolinones and Tricyclic N,O-Acetals by Trapping of Bis-*N*-Acyliminium Species in a Tandem Process; *European journal of organic Chemistry*, 2016, doi.org/10.1002/ejoc.201600530

9. E.I Blidi, L Namoune, A Bridoux, V. D Nimbarte, A.M Lawson, S Comesse, A Daïch; Expeditious Synthesis of the Topoisomerase-I Inhibitors Isoindolo[2,1-*b*]-isoquinolin-7(5*H*)-one and the Alkaloid Rosettacin based on Aryl Radical Cyclization of Enamide Generated by using *N*-Acyliiminium Chemistry; Synthesis, 2015, doi: 10.1055/s-0034-1378811
10. A Kamal, S Reddy, T Vishnuvardhan, V.D Nimbarte, A.V Subba Rao, S Vunnam, S Nagula; Synthesis of 2-aryl-1,2,4-oxadiazolo-benzimidazoles: Tubulin polymerization inhibitors and apoptotic inducing agents; Bioorg. Med. Chem. 2015, doi.org/10.1016/j.bmc.2015.05.060
11. A Kamal, B Moku, V. L Nayak, T.B Shaik, F. Shaikh, V.D Nimbarte; Design and Synthesis of Imidazo[2,1-*b*]thiazole-Chalcone Conjugates: Microtubule-Destabilizing Agents; ChemMedChem 2014, doi: 10.1002/cmdc.201402310
12. V.D Nimbarte, M Hadianawala, S Shrivastava, V. G. M Naidu, K. R Atcha; Design, Synthesis and Biological Evaluation of 4-(1-(4-(Sulphanilamide)phenyl) -3-(methyl)-1*H*-pyrazol-5-yl)dine Urea and *N*-Acyl Derivatives as Potent Soluble Epoxide Hydrolase (sEH) Inhibitor; Med. Chem. Res. 2013, doi:10.1007/s00044-013-0817-8, 2013.
13. A Kamal, N.V Reddy, V.L Reddy, V. S. Prasad, V. D Nimbarte, V Srinivasulu, M Vishnuvardhan, C.S Reddy; Synthesis and Biological Evaluation of Benzo[*b*]furans as Inhibitors of Tubulin Polymerization and Apoptosis Inducers; ChemMedChem, 2014, doi: 10.1002/cmdc.201300366
14. A. Kamal, M. P Rao, P Das, S. P Polepalli , V. D Nimbarte, K Mullagiri, J Kovvuri , N Jain; Synthesis and Biological Evaluation of Imidazo [2,1-*b*][1,3,4]thiadiazole-Oxindole Conjugates as potent tubulin polymerization Inhibitors; ChemMedChem 2014, doi: 10.1002/cmdc.201400069
15. V.D. Nimbarte; Design, synthesis and biological evaluation of substituted 9-oxo-1,2-dihydro-pyrrolo[2,1-*b*]quinazolin-3(9*H*)-ylidene)methyl)piperidine-1-carboxamide derivatives as dual COX-2 and (sEH) inhibitor; Der Pharma Chemica 2013, 5, 280-287.
16. V. D Nimbarte, Murtuza, H; Atcha, K. R.; Dual Inhibitors of COX-2 and Soluble Epoxide Hydrolase (sEH); J. Mod. Med. Chem. 2013, doi.org/10.12970/2308-8044.2013.01.02.5
17. K.R. Atcha, V. D Nimbarte, M.S. Hadianawala.; Design and *in-silico* studies of 4-(1-(4-Sulphonamide/ methylsulphonyl) phenyl)-3-(trifluoromethyl/methyl)-1*H*-pyrazl-5-yl)-dine urea and amide derivatives as novel potent dual active agents: Selective COX-2 and sEH Inhibitors; Biologically Active Molecule; 2012, 11, 209-211. (ISBN: 978-93-82062-03-5).

Conference contributions

1. V. D Nimbarte, H Schwalbe; Synthesis of isotope labelled biomolecules for DNP application, oral presentation at Europol meeting at University of Birmingham, United Kingdom, Feb 2017
2. V. D Nimbarte, H Schwalbe; Synthesis of isotope labelled biomolecules for DNP application, oral presentation at Europol meeting at DTU Copenhagen, Denmark, Feb 2018