Targeting GC-Rich Quadruplex Structures in the Human Genome with Small Molecule Ligands

Ying Xia

A thesis submitted for the degree of Doctor of Philosophy

University of East Anglia

School of Pharmacy

October 2022

Declaration

This thesis is submitted to the University of East Anglia for the Degree of Doctor of Philosophy and has not been previously submitted at this or any university assessment or for any other degree. Except where stated, and reference and acknowledgment are given, this work is original and has been carried out by the author alone.

Acknowledgments

Firstly, I would like to thank my supervisor Dr. Zoë A. E. Waller for her endless support for my Ph.D. study. In the past four years, she gave me all the patience, encouragement, and academic support in teaching me to be an independent researcher. Especially in my hardest time during the four years, I feel I am very lucky to have such excellent person as my supervisor. I would like to thank Zoë for proofreading my thesis, all the comments and advice were very useful.

Secondly, I would like to thank my secondary supervisor, Prof. Mark Searcey, for providing a synthetic lab during my first year. And also, thanks to Dr. Marco M.D Cominetti who supervised my synthesis project, and great help and support in NMR, HPLC, and mass spectrometry.

I would like to thank all collaboration groups. Dr. Hanna K. Maliszewska for the synthesis of metal complexes and Prof. Mike Hannon for the kind gift of Fe-cylinder. And Dr. Rianne Lord for providing compounds. Also, Prof. Shozeb Haider provided an opportunity the study TMPyP4 with RNA G-quadruplex.

I would like to thank Waller's group members for the past few years. Dr. Mahmoud AS Abdelhamid, Dr. Philip Spence, and Dr. Summer Rosonovski helped me a lot at the beginning of my Ph.D., showing me biophysical techniques and data analysis; Dr. Rupesh Chikhale for his encouragement and support over the past two years. I would like to thank Robert Yuan for the excellent work on FID screening experiments in **Chapter 5**. Also, I would like to thank Dr. Dilek Guneri, Effrosyni Alexandrou, and Fuad Mahamud for their kind support and company in the group at UCL.

In the end, I would like to thank my parents for their support and encouragement during the past four years.

Abstract

GC-rich DNA and RNA sequences can fold into a number of non-canonical structures such as G-quadruplexes and i-motifs. These structures have been shown to be able to affect many biological processes such as transcription, translation and replication. Targeting these sequences with small molecules can provide tools and leads for chemical intervention in these processes. This thesis describes the characterisation and interaction of compounds with GC-rich nucleic acids.

Chapter 1 is a general introduction to the research area of DNA secondary structures and focuses in particular on i-motif and G-quadruplex DNA, ligands, techniques used to study the folding and shape of these structures, and how to study melting curves of oligonucleotides.

Chapter 2 is the project of designing novel i-motif binding ligands and synthesis of a minor groove binder PyPyPyβDp. Another minor groove binder netropsin was also selected for DNA binding studies. Both minor groove binders were examined with i-motif, G-quadruplex and duplex DNAs using fluorescent intercalator displacement (FID) assay, fluorescence resonance energy transfer (FRET) melting and circular dichroism (CD) melting experiments. The results show these two minor groove binders can interact with the i-motif structure.

Chapter 3 is the work of exploring i-motif stabilizers using several biophysical techniques, including the FID assay, FRET melting and CD. Two NCI ligands 71795 and 19990 were found to stabilize the hTeloC i-motif. The metal complexes **22** and **23** can destabilize the i-motif structure. Iron cylinder was found to unfold the i-motif structure.

Chapter 4 is the work studying the interaction of TMPyP4 with RNA G-quadruplex. CD and UV titration experiments were used to examine the effect of TMPyP4 on PQS18-1 RNA G-quadruplex structure, and it can unfold the RNA structure. CD melting experiments were also used to investigate the thermal stability of PQS18-1 RNA G-quadruplex in the presence of ligands, and the results show TMPyP4 destabilizes the RNA G-quadruplex structure.

Chapter 5 begins with the characterization of GC-rich repetitive repeat sequences from the human genome. The sequences with the number of C and G-tracts from 3 to 5 were selected. The biophysical results show all these C-rich sequences can form intramolecular quadruplexes. With increased number of G-tracts, the proportion of parallel G-quadruplex topology also increased.

Chapter 6 provides a general conclusion, discussion about this work, and overviews for future works.

Chapter 7 describes the experimental procedures used in Chapters 2, 3, 4, and 5.

During the Ph.D. programme some of the work was published

1). Hanna K. Maliszewska, Carla Arnau del Valle, **Ying Xia**, María J. Marín, Zoë A. E. Waller, María Paz Muñoz, "Precious metal complexes of bis(pyridyl)allenes: Synthesis, catalytic and medicinal applications" *Dalton Trans.*, 2021,50, 16739-16750 (**Chapter 3**).

2). Susanta Haldar , Yashu Zhang , **Ying Xia**, Barira Islam, Sisi Liu, Francesco L Gervasio, Adrian J Mulholland, Zoë A E Waller, Dengguo Wei, Shozeb Haider, "Mechanistic Insights into the Ligand-Induced Unfolding of an RNA G-Quadruplex" *J. Am. Chem.* Soc. 2022, 144, 2, 935–950 (**Chapter 4**).

3). Baris Sergi, Ipek Bulut, **Ying Xia**, Zoë A E Waller, Yasemin Yildizhan, Ceyda Acilan, Rianne M Lord, "Understanding the Potential *In Vitro* Modes of Action of Bis(ß-diketonato) Oxovanadium(IV) Complexes" *ChemMedChem.,* 2021, 16, 15, 2402–2410 (not included in this thesis)

There are also manuscripts in preparation for publication

4) **Ying Xia**, M Comminetti, M Searcey, Z. A. E. Waller, "Minor groove binding compounds that bind i-motif DNA" (**Chapter 2**)

Table of Contents

List of Figures

Chapter 1: Introduction

Chapter 2: Towards design and synthesis of a novel i-motif ligand

Figure 2.11 *a). CD titration of 0 – 50 μM netropsin to hTeloG, CD signal of ligand was subtracted, b). CD melting of 50 μM netropsin to hTeloG, c). CD titration of 0 – 50 μM PyPyPyβDp to hTeloG, CD signal of ligand was subtracted, d). CD melting of 50 μM PyPyPyβDp to hTeloG. All experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 7.0.*

Figure 2.12 *a). CD titration of 0 – 50 μM netropsin to Hif1a, CD signal of ligand was subtracted b). CD melting of 50 μM netropsin to Hif1a, c). CD titration of 0 – 50 μM PyPyPyβDp to Hif1a, CD signal of ligand was subtracted, d). CD melting of 50 μM PyPyPyβDp to Hif1a. All experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 7.0.*

Figure 2.13 *a). CD titration of 0 – 50 μM netropsin to DAP, CD signal of ligand was subtracted b). CD melting of 50 μM netropsin to DAP, c). CD titration of 0 – 50 μM PyPyPyβDp to DAP, CD signal of ligand was subtracted, d). CD melting of 50 μM PyPyPyβDp to DAP. All experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 7.0.*

Figure 2.14 *a). CD titration of 0 – 50 μM netropsin to Nrf2, CD signal of ligand was subtracted b). CD melting of 50 μM netropsin to Nrf2, c). CD titration of 0 – 50 μM PyPyPyβDp to Nrf2, CD signal of ligand was subtracted, d). CD melting of 50 μM PyPyPyβDp to Nrf2. All experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 6.5.*

Figure 2.15 *a). CD titration of 0 – 50 μM netropsin to c-MYC, CD signal of ligand was subtracted b). CD melting of 50 μM netropsin to c-MYC, c). CD titration of 0 – 50 μM PyPyPyβDp to c-MYC, CD signal of ligand was subtracted, d). CD melting of 50 μM PyPyPyβDp to c-MYC. All experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 6.5.*

Chapter 3: Studies on the Effects of Small Molecules

Chapter 5: Repetitive GC-rich sequences study

1 μM, and the concentration of ligands was 2.5 μM (5 equivalents).

Figure 5.26 *Hits from FID screening, all the compounds showed over either positive or negative 10% TO displacement.*

Appendix

Appendix 1 *HPLC trace of PyPyPyβDp at 214 nm, using 5% MeOH: H2O, 0.05% TFA*

Appendix 2 *MALDI mass spectroscopy of PyPyPyβDp, samples were prepared at 1 mg/mL in MeOH.*

Appendix 4 *FRET melting curves for minor groove binders with DNA sequences, the concentration of DNA was 0.2 μM, the buffer is 10 mM sodium cacodylate and 100 mM potassium chloride at desired pH (pH 7.0 for DS, hTeloG, DAP, and Hif1a, pH 6.5 for Nrf2 and c-MYC, ph 5.5 for hTeloC), concentrations of minor groove binders were 1 μM and 10 μM.*

Appendix 5 *UV melting of DS, hTeloG, and Hif1a in the absence and presence of 50 μM of minor groove binders. [DNA] is 10 μM, and all experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 7.0. The data were fitted by sigmoidal curve using Origin 2022.*

Appendix 6 *UV melting of DAP, Nrf2, and c-MYC i-motif in the absence and presence of 50 μM of minor groove binders. [DNA] is 10 μM, and all experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 7.0 for DAP, pH 6.5 for Nrf2, and c-MYC. The data were fitted by sigmoidal curve using Origin 2022.*

Appendix 7 *CD titration and melting of NCI ligands with structures, all experiments were performed in 50 mM sodium cacodylate at pH 6.0, [hTeloC] is 10μM. The plot of normalized ellipticity at 288 nm against temperature gave sigmoidal curves.*

and 100 mM LiCl buffer at pH 7.0, the concentration of DNA is 2.5 μM.

List of Tables

Chapter 3: Studies on the Effects of Small Molecules

Chapter 5 Study of Repetitive GC-rich Sequences from Promoter Region of ALOX5

Abbreviations

A	Adenine
ALOX5	Arachidonate 5-lipoxygenase
AML	Acute myeloid leukaemia
BLAST	Basic Local Alignment Search Tool
C·C$^+$	cytosine-cytosine$^+$
C	Cytosine
CD	Circular dichroism
C-tract	Cytosine-tract
DAP	Death associated protein
DC$_{50}$	Half denaturing concentration
DIPEA	N, N-Diisopropylethylamine
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EtOH	Ethanol
FAM	6-Carboxyfluorescein
FID	Fluorescent intercalator displacement
FRET	Förster resonance energy transfer
G	Guanine
Hif1α	Hypoxia-inducible factor 1-alpha
HPLC	High performance liquid chromatography
hTelo	Human telomere
K$_d$	Dissociation constant
MALDI	Matrix-assisted laser desorption/ionization
MeOH	Methanol
MSMO1	Methylsterol Monooxygenase 1

NCBI	The National Centre for Biotechnology Information
NMR	Nuclear magnetic resonance
ANOVA	Analysis of Variance
PDB	Protein Data Bank
PEG	Polyethylene glycol
pH_T	Transitional pH
PQFS	Putative G-quadruplex forming sequences
T	Thymine
T_a	Annealing temperature
T_m	Melting temperature
ΔT_m	Change in melting temperature
TO	Thiazole orange
TDS	Thermal difference spectra
TAMRA	6-Carboxytetramethyl rhodamine
TFA	Trifluoracetic acid
TMPyP4	5,10,15,20-Tetrakis-(N-methyl-4-pyridyl) porphine
TSS	Transcription start site
UV	Ultraviolet
5′-UTR	5′ untranslated regions

Chapter 1

Introduction

1.1 DNA Structure

The structure of B-form DNA was first reported in 1953[1] by James Watson and Francis Crick, relying on the chemical analysis by Chargaff[2] and X-ray diffraction results of Franklin and Wilkins.[3,4] Their work proposed a right-handed double-helical structure consisting of two strands in an antiparallel orientation, where the 5′-end of one strand base pairs with the 3′ end of the other. DNA is comprised of four different bases: the pyrimidines, cytosine, and thymine, and the purines, guanine, and adenine. Hydrogen bonds form between these bases to form base pairs and this not only contributes to the stability of the DNA structure but also provides complementarity between the two strands.[1] Adenine base pairs by two hydrogen bonds with thymine and guanine base pairs using three hydrogen bonds with cytosine ***(Figure 1.1)***.[1]

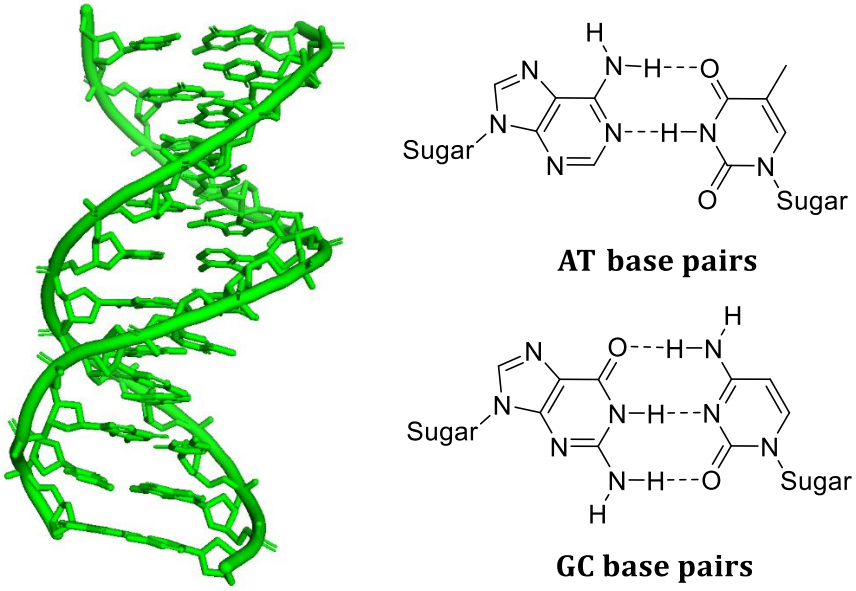


Figure 1.1 *Structure of the Dickerson-Drew dodecamer d(CGCGAATTCGCG) B-DNA (PDB ID: 1BNA), and Watson and Crick base pairing.*

Since then, much work has indicated that DNA can also fold into several other secondary structures and researchers across the world continue to study their characteristics and functions.[5–7] It has been assumed that the DNA within cells exists mostly as double-helical B-form DNA however, other double-helical structures have been identified such as A-DNA, a wide, short, and right-handed

DNA structure found in dehydrated DNA samples. It is not thought to exist under normal physiological environments.[8] Another example of double-helical conformation is Z-DNA, which has been found to form during gene expression in certain sequences and under certain specific conditions. It forms preferentially in alternating GC sequences under superhelical stress, but also can be further favoured by increasing the ionic strength. The ease of Z-DNA formation very much varies with sequence, e.g., $(CG)_n$ flips better than $(GGGC)_n$ and $(TG)_n$ i.e. it does not occur at random, regardless of the conditions, only specific sequences can form Z-DNA.[9,10] Unlike A- or B-form DNA, it has a unique left-handed conformation.[11] Under different conditions, the strands of DNA can adopt other different secondary structures such as hairpins,[12] triplexes,[13] cruciforms,[14] junctions[15], and quadruplexes (***Figure 1.2***).

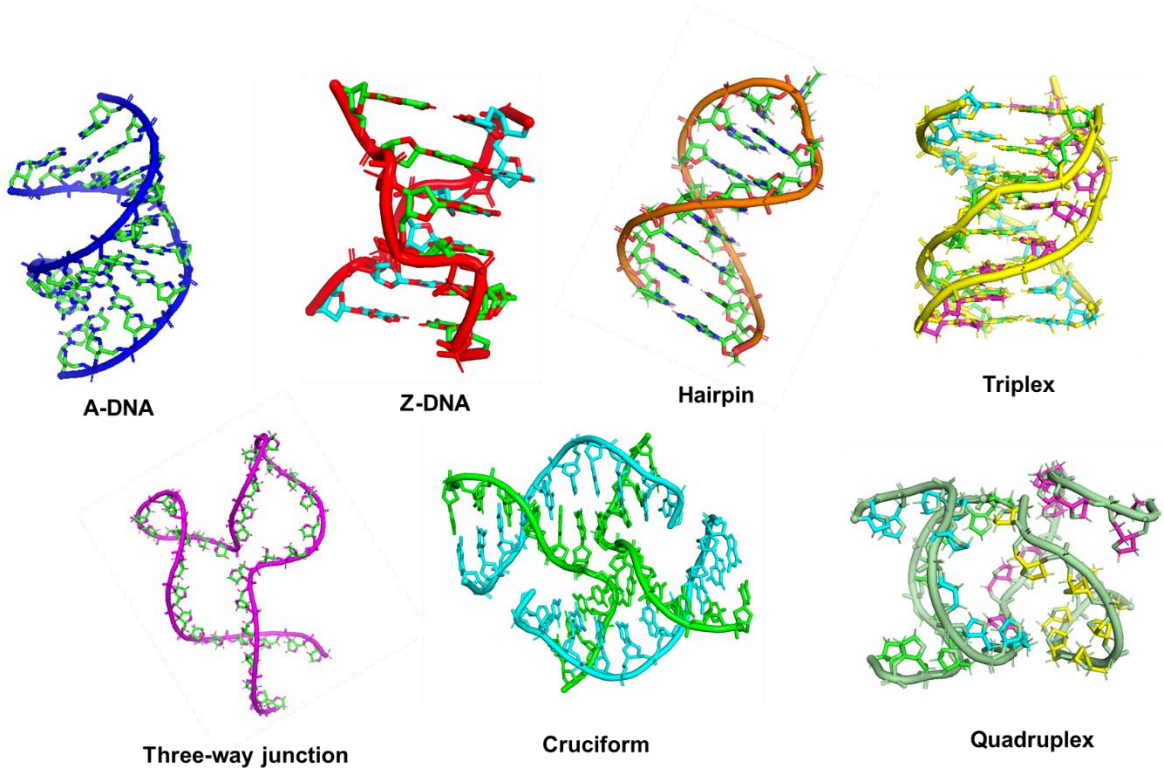


Figure 1.2 *Examples of a variety of DNA secondary structure types. A-DNA (PDB: 440D), Z-DNA (PDB:4R15), Hairpin (PDB: 1AC7), Triplex (PDB: 1D3X), three-way junction (PDB: 1SNJ), cruciform (PDB: 1P4Y), quadruplex (PDB: 1NYD).*

The ability of DNA sequences to form into different types of structures depends on the nucleotide sequences and environmental conditions. Different conformations

of DNA can result in changes in biological functions, for example, altering the binding of proteins and DNA-processing machinery.[16]

1.2 G-quadruplexes

The best-studied of all types of alternative DNA structures are G-quadruplexes, which are formed from guanine-rich sequences (***Figure 1.3***). G-quadruplexes are formed by stacking at least two planar G-quartets, which are created by Hoogsteen base pairs between four guanines,[17] The formation of G-quadruplex requires cations, which coordinate with the O6-oxygens from the guanines.[17]

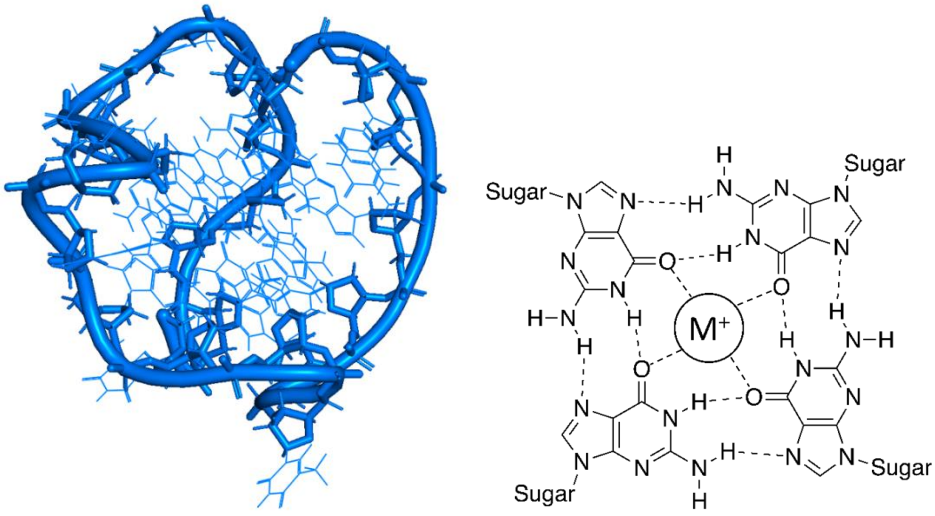


Figure 1.3 *Example of G-quadruplex structure (left, PDB: 1XAV) and a G-tetrad (right) stabilized by metal cation (M^+) within the core.*

G-quadruplexes are very well known to adopt different stable folds, e.g. parallel, antiparallel, and hybrid[18–21] (***Figure 1.4*** *a*), these contribute to different topologies of how the variable loop regions are presented.[22] The topologies of the G-quadruplex depend on the orientation of the guanines (*syn* and *anti* conformation of the guanine) within the G-tracts. In the parallel structure, all four strands are parallel, and the guanines are in the *anti* conformation. In the antiparallel and hybrid topologies, two and one G-tract are antiparallel respectively. The G-tracts are connected by loops, there are different types of loops including propeller,

lateral and diagonal (***Figure 1.4b***). Different conditions favour each of the G-quadruplex conformations, for example, G-quadruplexes with short loops always give parallel structures whereas sodium cations have been shown to often induce antiparallel G-quadruplex structures.[23] The combination of the directionality of the tracts and the length of the loops gives rise to different topologies within the G-quadruplex structure. For example, adjacent parallel G-tract strands are typically connected by propeller loops, lateral loops are able to link adjacent antiparallel G-backbones and diagonal loops are able to connect diagonal G-tract strands.

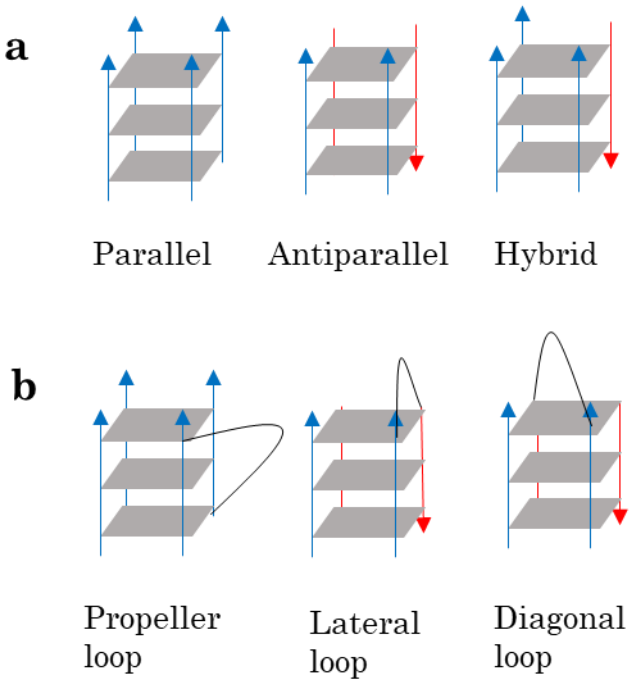


Figure 1.4 *Different possible strand orientations G-quadruplex and three example types of loops.*

The first report of Hoogsteen hydrogen-bonded guanine bases which form G-quadruplex was in 1962, by the observation that concentrated guanylic acid formed gels.[24] G-quadruplexes were later implicated in the inhibition of telomerase in 1991, by stabilization of G-quadruplex in the G-rich sequences of the human telomere.[25] Since then, studies into the biological functions of the G-quadruplex continue to attract much interest.[26] G-quadruplexes have been demonstrated to participate in key biological functions such as DNA replication,[27] transcription,[28] and translation.[29] Using a computational program to explore the

human genome, the Neidle group and the Balasubramanian group reported a large number of sequences that can potentially fold into G-quadruplexes in the human genome, with as many as 376,000 quadruplexes could potentially form.[3031] At the time, the authors indicated that it was unlikely that such a big number of quadruplexes could exist at the same time, and that there would be an equilibrium between quadruplexes and other types of DNA conformations. Since then, a study using next-generation sequencing methods showed that there may be over 700,000 potential G-quadruplex forming sequences in the human genome.[32] These findings all suggest undiscovered transcriptional functions of G-quadruplexes. Therefore, G-quadruplexes could be a potential target for the modulation of gene functions.

1.2.1 Cations Affect Stability of G-quadruplex

Metal ions are intrinsically associated with the formation and maintenance of the structure of nucleic acids in general.[33,34] Nucleic acids are highly polyanionic, in the formation of nucleic acid structures, the phosphate group attaches to the 5′ carbon of the sugar to form an ester bond that links nucleosides together, and each one has a negative charge, the cations such as sodium and potassium ions are required to neutralize the negative charges on the backbones so that nucleic acid molecules are enabled to form the compact native structures.[35,36,37]

Cations are particularly significant to the stability of G-quadruplex structures, as there is a set of four O6s from the composite guanines in the G-tetrad that readily coordinate with cations. Although potassium and sodium cations are both physiologically relevant ions, potassium is predominantly found within cells whereas sodium cations predominate in extracellular space. Plenty of research has investigated their effects on the stability of G-quadruplex, as well as other metal cations that have also been shown to affect the formation of G-quadruplex structures. The size of the cations and the charge affects the stability of the G-quadruplex structures. Potassium is largely regarded as the cation that favours G-quadruplex formation the most, with other cations following the order: K^+ >

$Ca^{2+} > Na^{+} > Mg^{2+} > Li^{+}$.[38] In the literature there are often references to lithium cations destabilizing G-quadruplex,[39] however, Li^{+} neither stabilizes nor destabilizes G-quadruplex structures.[40]

One of the factors that influence how cations affect the stability of the G-quadruplex is the ionic radius,[41] ***Table 1.1*** lists the effective cations interacting with G-quadruplex and their radii.

Table 1.1 *List of cations that affect the stability of G-quadruplex structures and their radii.*[41]

Cations	Coordination number	Angstroms (Å)
Monovalent cations		
K^{+}	6	1.33
Na^{+}	6	0.95
Li^{+}	6	0.60
$(NH)_4^{+}$	6	1.48
Tl^{+}	6	1.40
Rb^{+}	6	1.48
Cs^{+}	6	1.69
Divalent cations		
Co^{2+}	6	0.74
Mn^{2+}	6	0.80
Ni^{2+}	6	0.69
Mg^{2+}	6	0.65
Zn^{2+}	6	0.74
Cd^{2+}	6	0.97
Ca^{2+}	6	0.99
Sr^{2+}	6	1.13
Pb^{2+}	6	1.21
Ba^{2+}	6	1.35

Cations may coordinate in the plane of the G-tetrad, for example, Na^{+} fits the core exactly whereas Li^{+} is too small to bind tightly in the centre. Larger cations, such

as K^+ and Cs^+, are able to be coordinated between adjacent G-tetrads (i.e. between the planes).[42,43] Although most of the studies in the literature are about physiological monovalent cations with G-quadruplex such as K^+ and Na^+, it has been reported that G-rich sequences can also fold into a quadruplex structure in the presence of divalent cations. The crystal structure of the intermolecular G-quadruplex in the presence of Ca^{2+} and Na^+ was identified by the Neidle group, they observed an unequal distribution of Ca^{2+} in the G-quadruplex, indicating Ca^{2+} and Na^+ can interchange their positions.[44] The further study showed that Ca^{2+} induces a transition from antiparallel G-quadruplex to parallel structure of $d(G_4T_4G_4)$ sequence, and finally adopted the G-wire structure.[45] In the study of Sr^{2+} by the Chen group, through comparative experiments with monovalent cations, it was suggested that Sr^{2+} (T_m > *95°C)* stabilized the thermal stability of intermolecular G-quadruplex better than K^+ (T_m = 45/85°C) and Na^+ (T_m = 63°C),[46] the radius of Sr^{2+} is in between K^+ and Na^+.

Although G-quadruplexes have been very widely studied, they form on a strand of a duplex DNA sequence. In a genomic context wherever there is a G-rich sequence, there will also always be a complementary C-rich sequence. Such sequences are also important as they can also form quadruplex DNA secondary structures.

1.3 i-Motifs

i-Motifs are DNA structures formed from sequences rich in cytosine. Similar to G-quadruplexes, they are four-stranded but instead of having tetrads at the core, they are comprised of two hairpins in an antiparallel formation and locked together by cytosine-cytosine base pairs.[47] From 1961 it was known that hemi-protonated cytosine-cytosine base pairs were possible,[48] but it wasn't until 1993 that the structure of i-motif was firstly characterized by Gehring *et al.*,[47] using the cytosine-rich sequence d(TCCCCC) which formed the intercalated four-stranded structure under acidic conditions.[47] This structure was found to be comprised of two duplexes held together with hemi-protonated $C\text{-}C^+$ base pairs in an

antiparallel way (***Figure 1.5***).[47,49] After the first i-motif had been discovered, there were a number of reports of i-motif structures characterized by nuclear magnetic resonance (NMR) and X-ray crystallographic methods.[47,50]

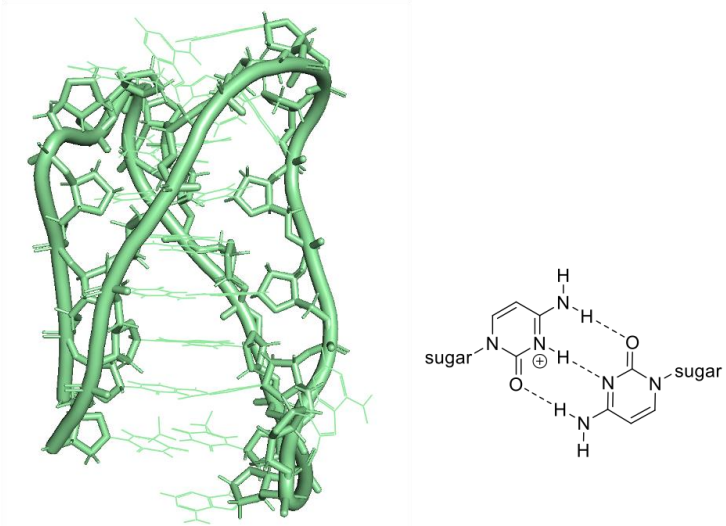


Figure 1.5 *Example of i-motif structure (left, PDB: 1ELN), and a hemi-protonated cytosine-cytosine base pair (right).*

The i-motif structure has two narrow grooves and two wide grooves. In the narrow grooves, the distance between the sugar-phosphate backbones in the strands is extremely short, the phosphate-phosphate distance can be extremely small and lead to destabilizing effects.[51] Although G-quadruplex structures have many different topologies (Section 1.2), all i-motifs to date are antiparallel. Differences between i-motif structures are still possible. i-Motifs are classified into two topologies, 3′E, and 5′E forms. The 3′E form is where the far most C-C^+ base pair is located at the 3′-end, while in the structure of 5′E, the outermost C-C^+ base pair is at the 5′-end.[52] It is possible to form intramolecular or intermolecular i-motifs. Sequences that have two tracts of cytosine can form bimolecular i-motifs and the first reported i-motif sequences were tetramolecular structures (***Figure 1.6***). The loops between C-tracts vary with each individual i-motif forming sequence and these affect the stability of the structures. When structures are comprised of one C-rich sequence, these are intramolecular and have loops of nucleotides that join the C-tracts together. Hurley's group previously classified intramolecular i-motif structures into "class I" and "class II" depending on the length of the loops.[53] i-

Motif structures with short loops have been regarded as class I, and those with long loops have been described as class II.

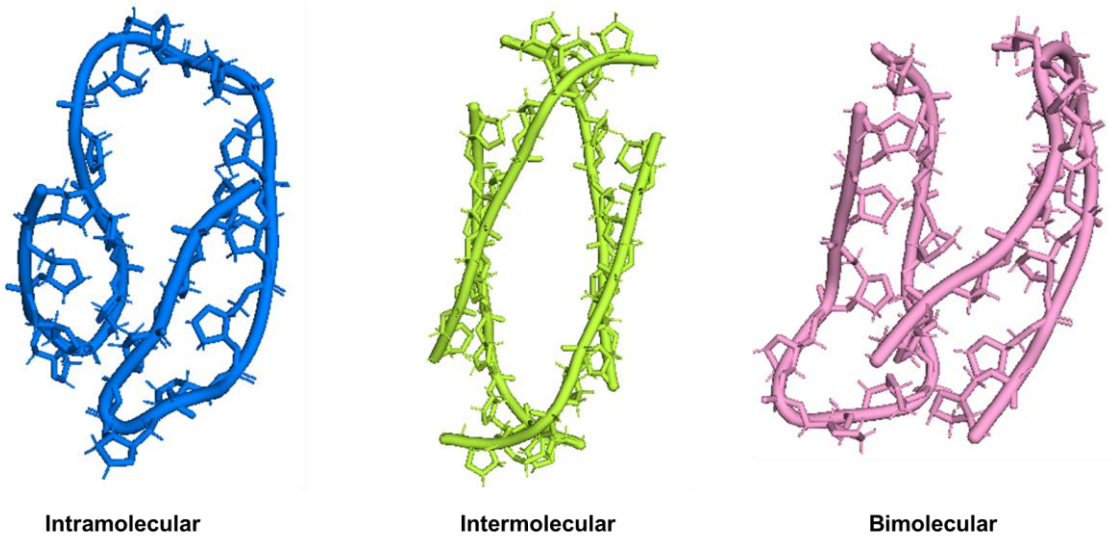


Figure 1.6 *The structures of intramolecular (left, PDB: 1A83). Intermolecular (middle, PDB: 1YBL) and bimolecular (right, PDB: 2MRZ)*

Early studies suggested that i-motif structures were most stable in acidic conditions, however, from more recent research, it has been indicated that i-motifs also exist and can be stable under neutral pH conditions.[54] Free cytosine-cytosine base pairing is most stable at the pH near the pK_a of cytosine N3 (pK_a = 4.6 for the free base). However, when within sequences of DNA, multiple stacks of base pairs can occur to form a core. As the stack of C-C$^+$ base pairs increases, the pK_a also increases, enabling some types of i-motif forming sequences to be stable at neutral pH.[55] For instance, in a study of the stability of the i-motif forming sequence from c-MYC over a pH range between 4.0 and 7.0, indicated the pK_a of cytosine also increased from 4.8 to 7.0.[55] i-Motif can also be stable at neutral pH depending on several environmental conditions including low temperatures (4°C),[56] the presence of copper [57] or silver cations,[58] under conditions of negative superhelicity[59] or molecular crowding.[60] Chemical modifications to the sugar-phosphate backbone can also encourage the formation of the i-motif structures at neutral pH e.g. use of 2'-deoxy-2'-fluoro-arabinocytidine has been shown to increase the stability of i-motif structure.[50]Although most i-motif structures show more stability under acidic environments, because of the requirements of hemi-protonation of cytosines,

C-C^+ base pairs do exist at neutral pH due to free H^+ ions taken up upon formation of folded i-motif structure.[61] There are many C-rich sequences identified that can form stable i-motif structures at neutral pH. For example, Hif1α, the first identified i-motif folded at neutral pH,[62] after this the Waller group has demonstrated a large number of C-rich sequences across the human genome that can form i-motif at neutral pH.[54] It was later shown by the Smith group that G-quadruplex and i-motif formation in human cells are interdependent and depend on cell cycle progression and chromatin accessibility.[63] This study also revealed that ligand-induced stabilization of G-quadruplex structure prevents formation of i-motif and vice versa. The equilibrium between duplex and quadruplex structures has been a fundamental question during studying quadruplex both *in vitro* and *in vivo*.[64] Under physiological conditions, most potential quadruplex-forming sequences normally exist as duplex structures.[64,65] This will depend on the sequences, as well as other conditions that may potentially shift the equilibrium towards duplex or quadruplex, such as the concentration of DNA, the type and concentration of stabilising cations and the pH.[66] As part of this previous study, it was found that many potentially stable i-motif forming sequences form in regions of the human genome that are associated with gene regulation. Therefore, it is indicated that i-motif could play a biological role in gene expression.

1.3.1 Biological Functions of i-Motifs

i-Motif forming sequences exist across different species of organisms, with examples in gene promoter regions and telomeres in both prokaryotes and eukaryotes.[67,68] Like the G-quadruplexes, i-motifs also play a role in regulating gene expression. There are several working models for i-motif function in regulating transcription:

1) i-Motif is used for transcription factor binding resulting in i-motif unfolding and transcriptional activation.
2) Small molecules destabilize i-motif, e.g., to a hairpin structure, causing loss of transcription factor binding and transcription repression.

3) Small molecule stabilization of i-motif blocks transcription factor binding sites and leads to transcriptional repression.

Although early studies mentioned the potential of i-motif relevant to biological functions, there was no direct evidence that i-motif could regulate transcription until 2014.[69] Through screening the National Cancer Institute Diversity Set library, two small molecules were identified to interact with the C-rich region of DNA from the promoter region of *Bcl-2* (5′-TTTTCTTTTCCCCCACGCCCTCTGCTTTGGGAACCCGGGAGGGGCGCTTACAGCCCCGCTCCCGCCCCCTTCCTCCCGCGCCCGCCCCT-3′). IMC-48 stabilises the i-motif, and IMC-76 destabilises the i-motif (***Figure 1.7***).[69] These ligands had opposing effects on Bcl-2 expression. IMC-48 shifts the equilibrium towards i-motif structure, whereas IMC-76 shifts the equilibrium to the hairpin species, meanwhile, these two ligands showed opposite effects on *Bcl-2* gene expression, where expression in BJAB cells was uninfluenced by IMC-76, but it showed downregulation effect at protein level; with the treatment of IMC-48, upregulation of gene expression was observed in BJAB cells.[69] Although other i-motif forming sequences had been studied before, this was the first example of small molecule targeting of i-motif and affecting gene expression.

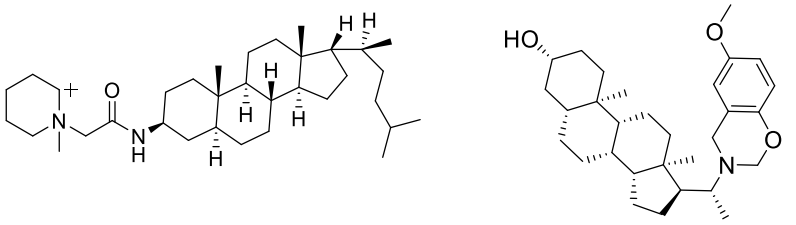


Figure 1.7 *Structures of IMC-48 and IMC-76 that interact with i-motif forming sequence from the promoter BCL-2.*

As some i-motif forming sequences are from promoter regions including oncogenes, and they are involved in regulating transcription, these i-motifs could be potential targets for shutting down oncogenic signalling. Previous studies have shown that we can find G-quadruplexes in nearly 43% of all gene promotors and 69% of cancer

gene promotors.[70] Similar studies on i-motif have been performed but it is clear that rules for i-motif formation in the genome are yet to be fully established. Regardless, as i-motif DNAs are complementary to G-quadruplexes, it is always worth considering them together. There are also i-motif forming sequences that exist in the telomeres. Cancer cells can proliferate with highly activated telomerase, which could be a critical feature of cancer, thus, the i-motif could be a potential therapeutic target for cancer.[71] Furthermore, from some more recent studies, it has been indicated that the i-motif DNAs can prevent DNA polymerase from progression,[72] and it is indicated that i-motif DNAs could interfere with DNA repair and replication.

1.3.2 i-Motif Forming Sequences

In addition to the example of modulation of gene expression in the promoter region of *BCL2*, many i-motif structures had already been identified in gene promotors of oncogenes such as *VEGF*,[73] *c-MYC*,[74] *Rb*,[75] and *KRAS*.[76] Sequences that fold into i-motifs had also been found in telomeric and centromeric regions[72] as well as repetitive regions of the genome such as the insulin-linked polymorphic region (ILPR).[77] There are lots of i-motif forming sequences that have been reported since, here we discuss in further detail the example i-motifs that are used as model sequences within this thesis. ***Table 1.2*** presents the i-motif sequence mentioned in this section.

Table 1.2 *i-Motif sequences are mentioned in this section.*

Name	Sequence 5′ → 3′
*hTelo*C	TAA-CCC-TAA-CCC-TAA-CCC-TAA-CCC
c-MYC	CCT-TCC-CCA-CCC-TCC-CCA-CCC-TCC-CCA
Hif1a	CGC-GCT-CCC-GCC-CCC-TCT-CCC-CTC-CCC-GCG-C
DAP	CCC-CCG-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CC
ATXN2L	CCC-CCC-CCC-CCC-CCC-CCC-CCC-CCC
MSMO1	CCC-CCG-CCC-CCG-CCC-CCG-CCC-CC
VEGF	CCC-GCC-CCC-GGC-CCG-CCC-C
Rb	CCG-CCC-AAA-ACC-CCC-C
KRAS	GCC-CGG-CCC-CCG-CTC-CTC-CCC-CGT-TGG-CCC-GGC-CCG-GTC-CCC-TCC-TTC-TCC-CCG

Human telomeric i-motif (*hTelo*)

The human telomeric i-motif is one of the first identified intermolecular i-motif sequences which was a tandem repeat sequence with a tract length of three cytosines, it forms a structure with a 3:3:3-nucleotide short loop topology.[78,52] The telomeres are non-coding regions at the ends of linear chromosomes that cap DNA and are comprised of tandem repeats of the sequences (TTAGGG). This special region plays an essential role in maintaining the stability and structure of a chromosome.[79] Due to the way DNA is replicated, the telomeres shorten at each round of DNA replication, known as end-replication problem.[80,81] This process provides a proxy for cell age, via the length of the sequences at the telomeres, as they shorten with each cell division. When cells reach a certain critical length of telomere, cell division stops and the cells become senescent, this is referred to the Hayflick limit and depending on the type of cell can be between 50 and 70 cell divisions. The telomeres can be lengthened using a reverse transcriptase called telomerase, which adds GGGTTA at the 3' terminus.[74] This extension of the telomeres by telomerase enables the cells to avoid the Hayflick limit.[82] Telomerase is not normally active in most somatic cells, however, it has been indicated to be

active in about 85% of cancers which can cause cancer cells to become immortal.[83,84] The *hTelo* i-motif has been shown to be stabilized by carboxyl-modified single-walled carbon nanotubes (SWNTs) and this has also been shown to inhibit telomerase, cause apoptosis, and initiate a DNA damage response.[85] Under slightly acidic conditions, it has been shown that the human telomeric sequence preferentially forms i-motif DNA, rather than the formation of the competing duplex.[83]

c-MYC i-motif

One of the well-studied i-motif forming sequences is from the nuclease hypersensitive element (NHE) III_1 region in the *c-MYC* gene, which has been suggested to be significant in the regulation of c-MYC expression[86] and controls 90% of transcription of *c-MYC*.[87] The factor of oncogene transcription *c-MYC* is expressed in many kinds of cancers; it plays a significant role in the rapid proliferation of cancer cells. The *c-MYC* gene has a broad range of functions in gene and cell cycle regulation, and the deregulation of the *c-MYC* gene is a feature in many types of cancers.[88]

Laurence Hurley's group has studied the i-motif forming sequence from the promoter region of *c-MYC* in detail and measured the transitional pH (pH_T) to be 6.6. The transitional pH is defined as the point at which half the i-motif is folded and half is unfolded (i.e. at the mid-point between folded and unfolded) and is now used widely as a measure of the stability of the i-motif.[89] Further studies on this i-motif forming sequence, have suggested that i-motif can form at neutral pH under conditions of negative superhelical stress.[90] However, the formation of the i-motif is quite different, for example, a "class II" variation of structure with a long loop topology only has six $C\text{-}C^+$ base pairs,[90] while the 5:5:5-nucleotide loop topology can form an i-motif with eight $C\text{-}C^+$ base pairs.[91]

Small molecules that stabilize *c-MYC* i-motif could be used as new agents for cancer therapy because they can potentially decrease oncogene expression. A study from researchers based at Sun Yat-sen University identified compounds

that stabilize the *c-MYC* i-motif.[92,93] One stabilizer is 2-methylacridin-9(10*H*)-one, an acridone derivative, analogues were subsequently generated to increase the selectivity for i-motif.[92] Although a number of compounds were investigated to stabilize i-motif structures, B19 was the best stabilizer compared to the others. This compound showed high binding affinity and selectivity to i-motif rather than G-quadruplex and double-stranded DNA. Furthermore, B19 also decreases transcriptional activation and reduces mRNA and protein expression, it can prevent the proliferation and metastasis of cervical cancer cells *in vitro*.[92] Another small molecule that binds to the *c-MYC* i-motif with high selectivity is a bisacridine derivative, A9 (***Figure 1.8***).[93] This molecule can target both the i-motif and G-quadruplex structures from the c-MYC promoter region with preference compared to the hairpin and duplex DNA. A9 prevented the proliferation of multiple cancer cells, including lung, liver, cervical, lymphoma, colon, and bone cancer cell lines.[93]

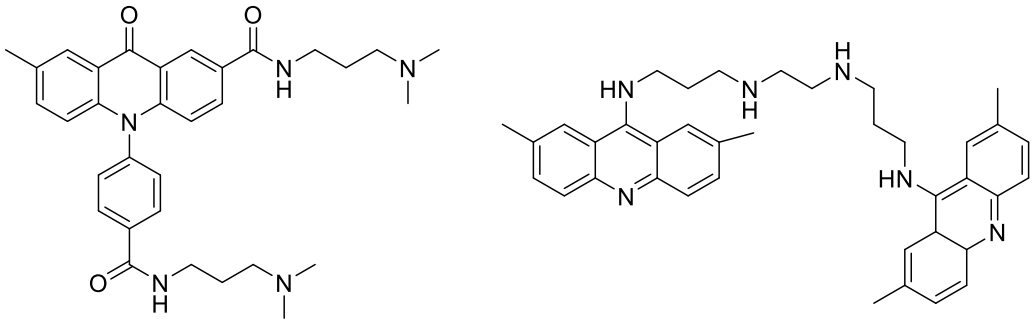


Figure 1.8 *Structures of example small molecules binds to the c-MYC i-motif.*

Hif1-α i-Motif

Hypoxia-inducible factor 1 (*Hif1*) is a transcription factor composed of hypoxia-inducible subunit alpha (Hif1α) and beta (Hif1ß).[94] *Hif1α* has been found overexpressed in many human cancers of intratumoral hypoxia.[95] It has also been identified during hypoxia, *Hif1α* would accumulate and translocate to the nucleus, then could form a dimer with Hif1ß.[96] The i-motif forming sequence from the promoter region of *Hif1α* was the first i-motif that showed unexcepted stability at neutral pH, the transitional pH was found to be 7.2,[62] close to physiological pH.

The unexpectedly high stability of the sequence from *Hif1α* challenged the belief that long looping regions (Class II) are required to form i-motif structures that are stable near physiological conditions. It was the first example of an i-motif forming sequence that was naturally stable at neutral pH, without any additional external stabilizing factors (such as cations or molecular crowding agents).

Neutral i-motif forming sequences

The Waller group previously built a study to investigate the stability of i-motifs with cytosine tract length (2 – 10).[54] From their research, C-rich sequences with a tract length of 5 or above were able to form i-motif structures and were stable at neutral pH. Following a computational search of the human genome for sequences that could fold into i-motifs with tract lengths of five or above the team discovered several native i-motif forming sequences that were found to be stable at neutral pH. This was the first time it was shown that there were multiple examples of i-motifs, found throughout the human genome, which were potentially stable under physiological pH, which include *DAP* (pH_T 7.0), *ATXN2L* (pH_T 7.0), *DUX4L* (pH_T 7.1), *PIM* (pH_T 7.0), *PLCB2* (pH_T 7.0), and *SNORD112* (pH_T 7.2) This finding suggested that the i-motif forming sequence from Hif1α was not the only special case that showed stability at neutral conditions, and across the human genome, there are thousands of potentially physiologically stable i-motif structures. Further study of these has indicated these sequences were not located randomly, they are from the specific regions in the genome associated with regulatory functions such as gene promoters, which indicated that the sequences may have a potential physiological function.[54] In this thesis we used three examples from this batch of neutral-stable i-motifs: the ones originally found in the promoter regions of *DAP* and *ATXN2L*, as these particular sequences are also found across multiple sites in the human genome, as well as the sequence from the promoter region of *MSMO1* which has a similar pattern of cytosine repeats as *DAP*.

1.4 i-Motif ligands

i-Motif DNA has been identified as a potential target that has a role in biological functions, therefore, it is important to identify potential i-motif interacting molecules as they may be used as probes to help investigate the biological functions and how they may be controlled or influenced by small molecules. Ligands can also be developed for therapeutics that target i-motif for cancer or other diseases and conditions.

Several factors for investigating ligands need to be considered, for example, is it possible to deduce a pharmacophore, or necessary chemical structure required for binding i-motif? What are the effects of ligands on the topology of i-motif? How can ligands affect the stability of i-motif structures? Herein, examples of i-motif ligands are shown in this section.

The first i-motif binding ligand discovered was the porphyrin TMPyP4 (***Figure 1.9***). TMPyP4 is quite an interesting but promiscuous ligand that can interact with both G-quadruplex and i-motif as well as double-stranded DNA.[97] The binding effect was initially identified by using human telomeric sequences $d(AATCCC)_4$ and $d(CCCAAT)_4$,[98] which TMPyP4 can bind and induce the formation of i-motif at pH 4.5. Modeling experiments showed that TMPyP4 binds to i-motif through stacking at the ends of the structure.[98] More recently, however, the Waller group performed fluorescence intercalator displacement (FID) assays of commonly used G-quadruplex binding ligands, including TMPyP4, to observe their interaction with i-motifs. The team used i-motif forming sequences from the promoter regions of *DAP* and *ATXN2L* as they are both stable at neutral pH. The DC_{50} (where the concentration showed 50% displacement of the intercalator thiazole orange) of TMPyP4 was determined 0.16 μM to *DAP*, indicating strong binding affinity. However, there were remarkable differences in the stabilization properties. UV melting experiments showed that TMPyP4 stabilizes *ATXN2L* with a change in melting temperature (ΔT_m) of +4.5°C but it significantly destabilizes *DAP* with ΔT_m -16.0 °C. It is clear that TMPyP4 has different properties not only depending on the type of secondary structure but also on the composite sequence. In **Chapter**

4 we discuss how the properties of TMPyP4 binding vary depending on the type of nucleic acid.

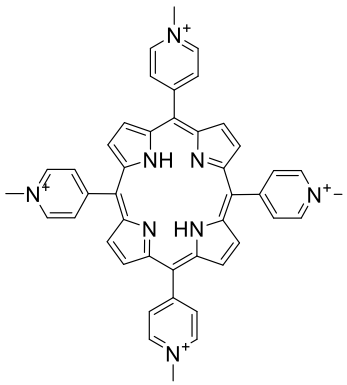
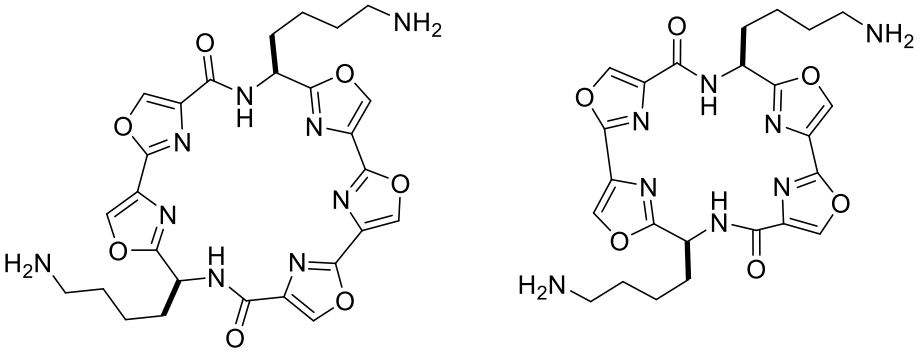


Figure 1.9 *Structure of porphyrin ligand TMPyP4.*

The Nagasawa group also looked at their G-quadruplex ligands for inspiration when thinking of targeting the i-motif. Their team had a number of macrocyclic poly-oxazole compounds that had been made and used as G-quadruplex ligands. They decided to test these compounds to see whether they could able to interact with i-motif.[99,100] Two analogues of telomestatin,[101–103] which is a well-known quadruplex stabilizer, penta-oxazole (L2H2-5OTD) and tetra-oxazoles (L2H2-4OTD) were made (***Figure 1.10***).[104] L2H2-4OTD did not show stabilization to G-quadruplex, however, it was shown to interact with i-motif by CD titration experiments. It was found that with increased additions of L2H2-4OTD to the human telomeric i-motif, the CD spectra of i-motif at 288 nm were shown to decrease and the peak was slightly shifted to a lower wavelength, indicating unfolding of i-motif structure. Using NMR, the authors indicated that the binding sites of L2H2-4OTD are on loop 1 and loop 3.[105]



L2H2-5OTD L2H2-4OTD

Figure 1.10 *Structures of L2H2-5OTD and L2H2 4OTD.*

Another early-described i-motif binding ligand is bis-acridine (BisA) [106] which was the first compound tested with human telomeric G-quadruplex and the i-motif by FRET melting (***Figure 1.11***). Through the FRET, Alberti's group measured the i-motif DNA melting temperature, they found BisA could stabilize G-quadruplexes and stabilize i-motif at pH 6.8 with ΔT_m +33°C. However, further work on BisA moved to the development of BisA as a G-quadruplexes binder rather than an i-motif binding compound.[106] Another acridine derivative monomeric acridine with two propylamine chains was also studied by the FRET melting experiment but showed no effect with i-motif.[107] Furthermore, work by Dr. Henry Day indicated that BisA actually precipitates DNA, rather than causing stabilization.[107]

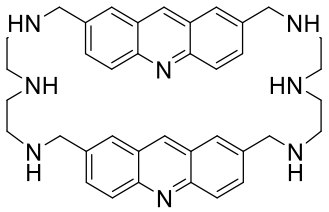


Figure 1.11 *Structure of acridine derivative BisA.*

Carboxyl-modified single-walled carbon nanotubes (SWNTs) are one of the most studied i-motif binders (***Figure 1.12***).[85,108,109] The SWNTs can stabilize the i-motif structure at pH 5.5 with the ΔT_m +22°C, it can also induce the formation of the *hTelo* i-motif DNA and inhibit the formation of duplexes even at pH 8.0.[110] SWNTs

can also induce the formation of i-motif at physiological pH under molecular crowding crowing conditions.[109] Due to the ligand size, the possible binding sites on the i-motif were proposed to be the 5′-end on the major groove.[111] It has been found that SWNTs inhibit telomeric activity by stabilizing the i-motif structure, using TRAP assay *in vitro* and *in vivo*.[111]

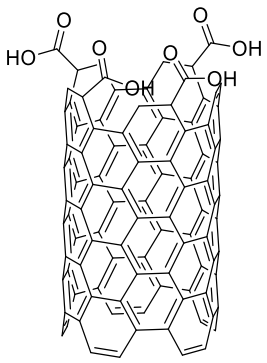


Figure 1.12 *Structure of SWNT.*

A perylene-neomycin conjugate is a specific human telomeric G-quadruplexes binding ligand (***Figure 1.13***), this conjugate was also compared with other G-quadruplexes, double-stranded and i-motif DNAs using a competition dialysis experiment.[112] The binding affinity of this ligand with i-motif was found to be weaker than that for G-quadruplex from the human telomere, but higher than double-stranded and other G-quadruplex DNAs. However, the competition dialysis experiments were all performed at pH 7.0, where the human telomeric i-motif forming sequence is normally unfolded, therefore, the method to measure binding affinity was more likely for a single-strand sequence instead of an i-motif structure.[112]

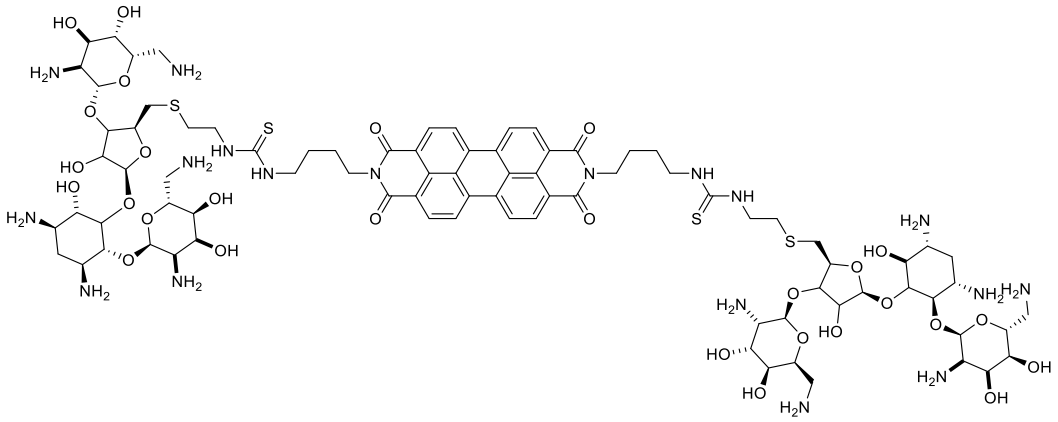


Figure 1.13 *Structure of a perylene-neomycin conjugate.*

Mitoxantrone (***Figure 1.14***) is an anticancer agent which has previously been studied by the Waller group for its interaction with i-motif DNA.[113] From FRET melting experiments, it appears that mitoxantrone stabilizes i-motif structures. Furthermore, i-motif structures can be formed in physiological pH (7.4) with increased additions of mitoxantrone. SPR analysis showed mitoxantrone's binding affinity to double-stranded DNA to be much weaker than its binding affinity to i-motif. In this study, CD indicated some interesting features, which the authors attributed to an induced circular dichroism effect. However, further work since then has suggested that mitoxantrone can destabilize i-motif.[97,114] Mitoxantrone was also found as a hit during a screen of an NCI library. This screen also discovered the ligands tilorone and tobramycin (***Figure 1.14***).[115]

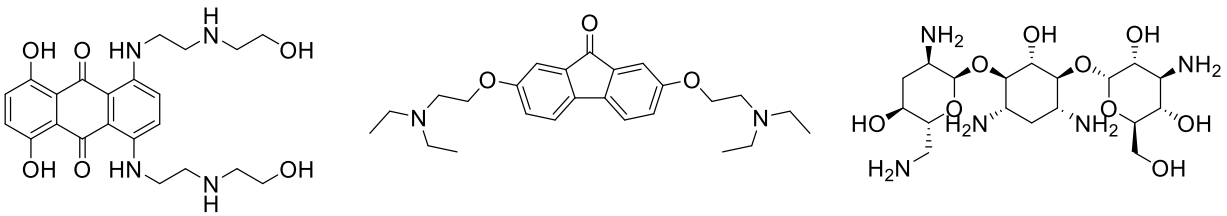


Figure 1.14 *Structure of mitoxantrone, tilorone, and tobramycin.*

Fisetin (Fis) is a bioactive plant flavanol that exists in various fruits, nuts, and vegetables (***Figure 1.15***).[116,117] The Fis flavanol shows multi-functional potential

therapeutic activities which include anti-diabetic,[118] anti-cancer[119], and neuroprotective effects[120] in human diseases. Bhowmik and co-workers previously tested the interaction of Fis against G-quadruplex DNA.[121] In subsequent work, they investigated the binding of Fis against different i-motif forming sequences from the promoter regions of oncogenes and the human telomere.[122] It is indicated that Fis is able to bind a *VEGF* i-motif forming sequence preferentially over other i-motifs. The interaction of Fis with the *VEGF* i-motif was shown to cause the transformation of the i-motif structure to a hairpin-like structure.[122]

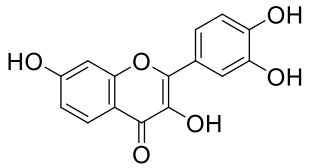


Figure 1.15 *Structure of fisetin.*

Similar to a lot of early i-motif binding compounds, Phenanthroline derivatives had previously been identified to interact with G-quadruplex and inhibit telomerase activity.[123] In two separate papers, the Wei group investigated the phenanthroline derivatives **1** – **6** (***Figure 1.16***) against i-motif. UV melting experiments showed ligands **1 - 3** stabilizing the i-motif *hTelo*C with ΔT_ms of +7.6, +7.2 and +10.1°C respectively at pH 5.5,[124] and ligands **4** – **6** significantly stabilized *hTelo*C i-motif with ΔT_m +14.6, +12.4 and +10.5°C respectively.[125] UV titration experiments suggested derivatives **1** – **6** might bind to i-motif via stacking on the C-C^+ base pairs.[125] However, these compounds are not specific to i-motif, as the binding affinity of i-motif (between 4 – 8 μM) is lower than G-quadruplex (1.6 - 2.5 μM) and the compounds also interact with duplex DNA (6.7 - 12 μM).[125]

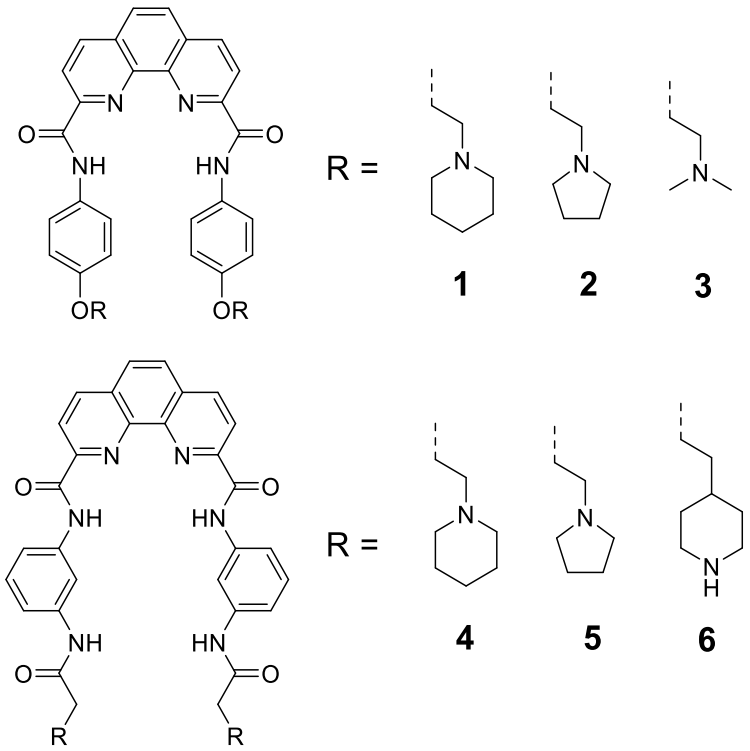


Figure 1.16 *Structures of phenanthroline derivatives.*

In the previous section, we gave examples of i-motif binding ligands that had been used to modulate biological function which has been discussed earlier in **Section 1.2.1**. In addition to the ligands IMC-48 and IMC-76 investigated by the Hurley group, the same team also reported a benzothiophenecarboxamide compound NSC 309874 (***Figure 1.17***) that can selectively bind to the platelet-derived growth factor receptor *(PDGFR)-β* i-motif and also can downregulate its promotor activity.[126]

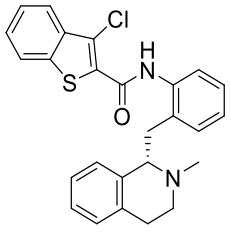


Figure 1.17 *Structures of NSC 309874.*

The Dash group also reported that the peptidomimetics PBP1 and PBP2 (***Figure 1.18***) can interact selectively with G-quadruplex and i-motif DNA. Through FRET

melting experiments, FID assays, and fluorescence binding titrations, it showed that PBP1 binds highly selectively to i-motif and PBP2 shows a preference to bind G-quadruplex.[127] Moreover, the team also reported that the addition of PBP1 and PBP2 could induce the formation of i-motif and G-quadruplex using single-molecule FRET. The cellular studies suggested that PBP1 can upregulate *Bcl-2* gene expression while PBP2 showed downregulation in *Bcl-2* gene expression in cancer cells.[127]

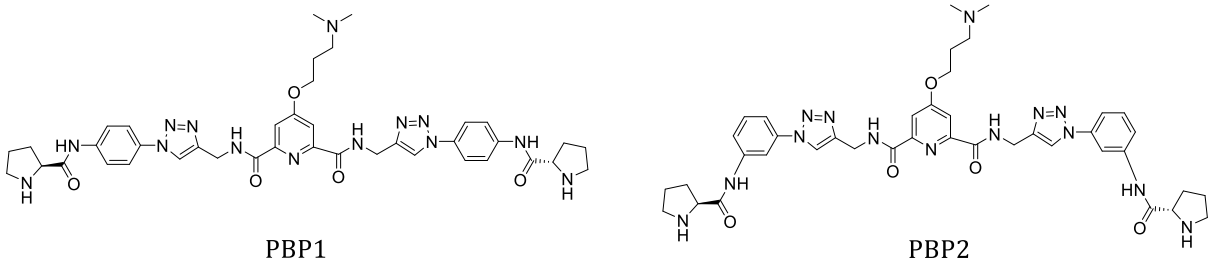
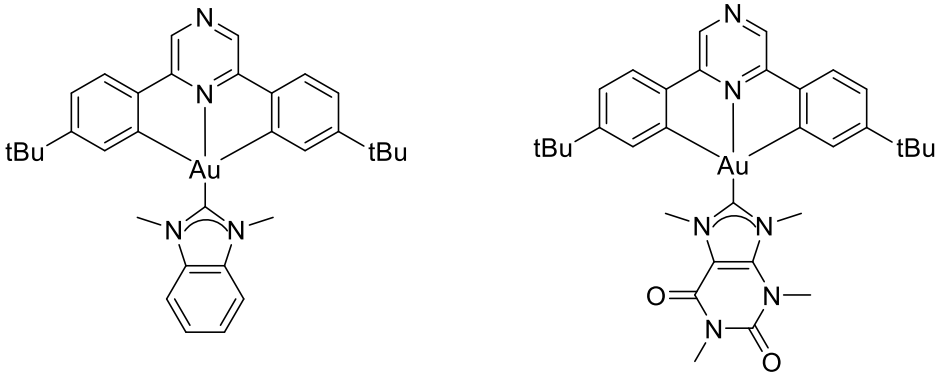


Figure 1.18 *Structures of peptidomimetic ligands PBP1 and PBP2.*

In addition to organic ligands, metal complexes are well known to interact with DNA, such as cis-platin, a metalating agent, which has been used in clinic for anti-cancer chemotherapy.[128] Cis-platin works by covalent attachment to guanines and can form both intra and intermolecular G-G cross-links in DNA. Since the discovery of cis-platin there have been a number of analogues that have been developed and there is much interest in metal-complexes and their interaction with DNA.[129] Some metal complexes can also interact with i-motif structures. A series of gold(III) complexes were synthesized by Bochmann and co-workers, these compounds are potential anticancer agents and were the first gold-based complexes to be shown to interact with i-motif DNA.[130] Complexes **7** and **8** (***Figure 1.19***) were identified to have good interactions with the i-motif from Hif1α and both the i-motif and G-quadruplex forming sequences from the human telomere (*hTelo*C and *hTelo*G). FRET melting experiments indicated that the ΔT_m showed significant stabilization temperatures (between 20 – 40°C) at 20 μM of ligands. Complex **7** showed a more stabilizing effect on *hTelo*C at pH 6.0 than at pH 5.5, whilst complex **8** did not show a big difference in interaction with *hTelo*C at both pH conditions.[130]



7 **8**

Figure 1.19 *Structures of Au(III) complexes.*

The fluorescent dye crystal violet (***Figure 1.20***) binding with i-motif was investigated to develop a fluorescent DNA logic gate.[131] This compound is a selective fluorescent probe for both G-quadruplex and i-motif over single-stranded and double-stranded DNA. When crystal violet interacts with single-stranded DNA structure, the fluorescence is lower whereas when it is in the presence of G-quadruplex or i-motif, its fluorescence is increased. Furthermore, the crystal violet fluorescence was found to be enhanced when the equilibrium of structures is shifted. I.e. formation of G-quadruplex and i-motif structures equates to higher fluorescence when the equilibrium shifts to the folded forms.[131]

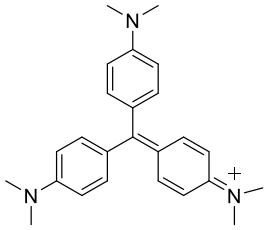


Figure 1.20 *Structure of crystal violet.*

Ruthenium complexes have also been studied with i-motifs (***Figure 1.21***). Some ruthenium complexes have previously been used as general DNA "light switches" because of their luminescence increase in the presence of DNA,[132,133], and such complexes have also been said to have potential as anticancer agents.[134] The complexes $[Ru(phen)_2(dppz)]^{2+}$ and $[Ru(bpy)_2(dppz)]^{2+}$ were identified to bind to G-quadruplex and can induce the formation of G-quadruplex.[135] In the examination with i-motif using CD, the CD signal showed a slight change with the addition of

these compounds at acidic conditions, but they do not affect the melting temperature of i-motif using UV melting experiments.[135] According to the luminescence experiments, these compounds showed a preference for G-quadruplex rather than i-motif.[135,136]

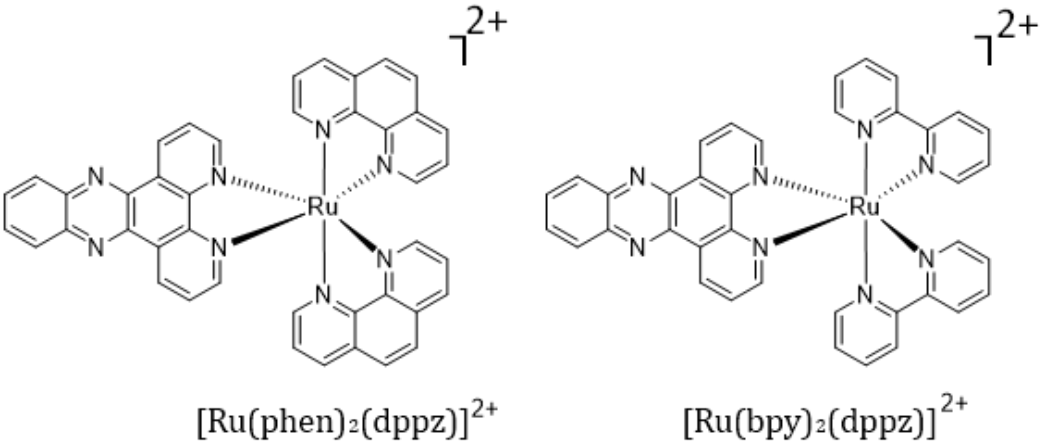


Figure 1.21 *Structures of ruthenium complexes*

Many i-motif binding ligands can be used as fluorescent probes, for example, crystal violet as described above has been used in DNA logic gates and other light-up effects. The Waller group has developed an FID displacement assay using thiazole orange,[137] and there are other fluorescent probes discovered to target i-motif DNA. The $[Ru(bqp)]^{2+}$ (***Figure 1.22***) was firstly developed by the Hammarstrom group[138], and its three isomers *mer*-$[Ru(bqp)]^{2+}$ (mer),[139,140] *cis,fac*-$[Ru(bqp)]^{2+}$ (cis) and *trans fac*-$[Ru(bqp)]^{2+}$ (trans)[141] were identified. The Waller group investigated the DNA binding properties of these three isomers using modeling, FID assay, electronic absorption, and emission lifetime measurement,[142] they found cis binds most types of DNA structure compared to mer and trans, furthermore, the single enantiomer cis showed preference to i-motif rather than G-quadruplex and double-stranded, it also showed a light switch-on effect in the presence of *hTelo*C i-motif,[142] therefore, it is a great example for the further development of light-switching compounds for DNA secondary probes.

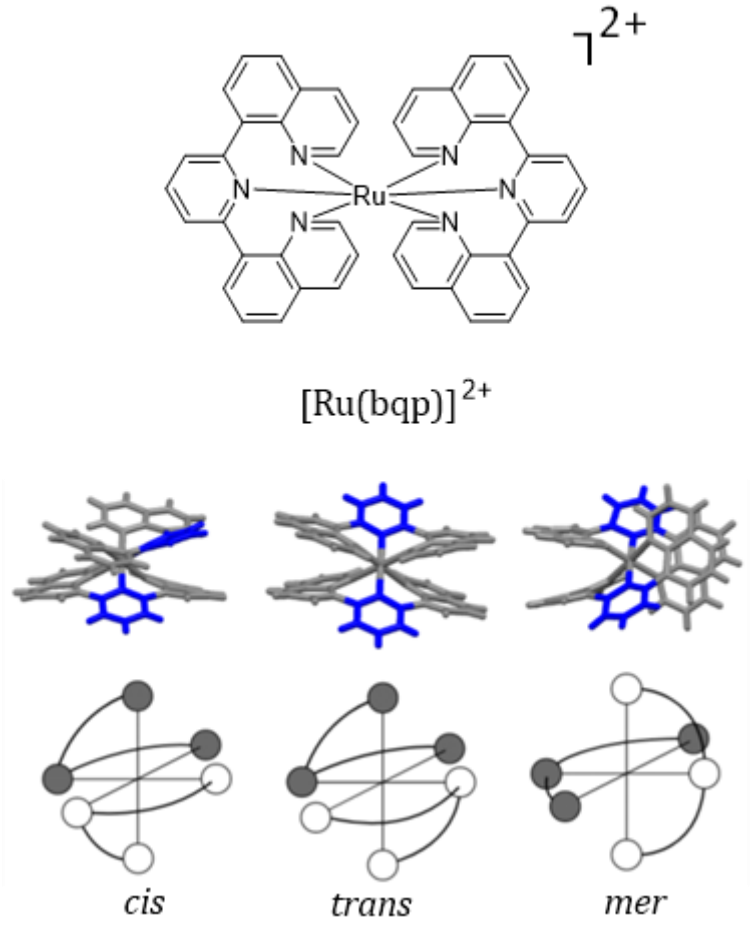


Figure 1.22 *Structures of different analogues of* $[Ru(bqp)]^{2+}$.

Investigation in i-motif ligands has attracted considerable interest in recent years, this area continues to expand and grow as more evidence shows i-motifs are biologically significant with potential as a target for therapeutics against genetic-based diseases. Nevertheless, there are only limited i-motif binding compounds that are specific for i-motif. Many described above also interact with G-quadruplex and/or double helical DNA. Many of the compounds that have been found have only low micromolar dissociation constants. There is much room in the field for tighter-binding ligands. Therefore, exploring specific i-motif binding ligands is still an important area of research.

1.5 Biophysical techniques for studying DNA-ligand interactions

In order to hunt for DNA-binding ligands, a number of biophysical techniques are required to investigate the interaction between DNA and ligands. For example, FID assays were developed by the Waller group using thiazole orange as a fluorescent probe to detect the binding of ligands to DNA.[115] The Balasubramanian group used fluorescence resonance energy transfer (FRET) melting and surface plasmon resonance (SPR) to study the binding properties of the ligands to G-quadruplex DNA.[143] Nuclear magnetic resonance (NMR) is a widely used technique to study DNA-ligand interaction, the Hurley group used this technique to identify the binding sites of TMPyP4 to i-motif.[98] Circular dichroism (CD)[144] spectroscopy/melting and ultraviolet (UV)[98] spectroscopy/melting are useful techniques to study DNA secondary structures and DNA-ligand interactions. Isothermal titration calorimetry (ITC) is also a powerful technique to study the interaction of DNA and compounds, Haider and co-workers performed ITC to determine the stoichiometry of TMPyP4 to G-quadruplex of 2:1.[145] Herein, some biophysical techniques were used in this thesis and will be discussed in detail, including FRET, FID, CD, and UV.

Fluorescence resonance energy transfer (FRET) and FRET melting

FRET is a physical phenomenon where the fluorescence of a donor fluorophore is transferred to an acceptor fluorophore when in sufficient proximity.[146] The process can be summarized as:

$$D^* + A \rightarrow D + A^*$$

Where D is the donor, A is the acceptor, and * is an electronically excited state.

This effect happens when sufficient energy matches between the excited states of two fluorophores, which is called resonance. The energy transfer relies on a no-

radiative mechanism between two fluorophores, and it is distance-dependent between fluorophores (~10 nm).[146] FRET efficiency (E_T) shows the percentage of the excitation photons that contribute to FRET. E_T can be calculated from the fluorescence change in the donor:

$$\textbf{\textit{Equation 1}}: E_T = 1 - \frac{FI_d}{FI_d^0}$$

Where FI_d is the fluorescence intensity of the donor, and FI_d^0 is the fluorescence intensity of the acceptor.

The E_T can be also calculated from the distance between two fluorophores, it is shown as:

$$\textbf{\textit{Equation 2}}: E_T = \frac{R_0^6}{R_0^6 + R^6}$$

Where R_0 is the Förster distance between two fluorophores where the E_T is 50%, and R is the distance between the donor and acceptor.

FRET needs both a donor fluorophore and an acceptor fluorophore to work, for most early FRET studies, the fluorophores used were 6-carboxyfluorescein (FAM) as the donor fluorophore to the 5′-end and tetramethylrhodamine (TAMRA) as the acceptor fluorophore to the 3′-end (***Figure 1.23***),[69] though many other donor pairs are possible. In terms of DNA structure, the fact that there may be a large change between folded and unfolded structures means potential distances between the ends of DNA will also change and can be monitored easily using FRET. DNA can be labelled at either end and when folded the fluorophores are close (and FRET may happen), whereas when they are unfolded, they are far apart (i.e., no FRET).

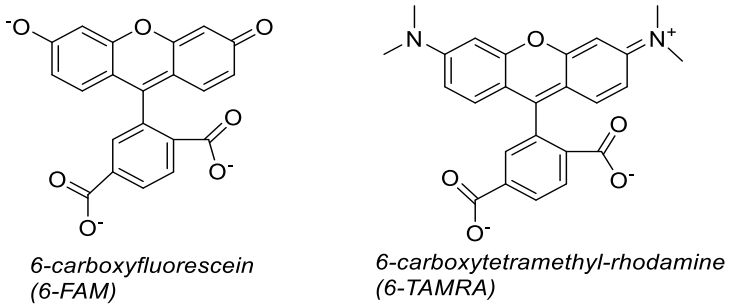


Figure 1.23 *The structures of 6-carboxyfluorescein and 6-carboxytetramethyl-rhodamine*

FRET can be characterized using the ratio between the donor and acceptor fluorophore emission intensities, which is known as a ratiometric method.[147] The ratiometric FRET is usually used for determining relative changes in FRET efficiency. The energy transfer between donor and acceptor intensity ratio can be calculated (relative FRET efficiency) using ***Equation 3***, Where I_A and I_D are the fluorescence intensity of acceptor and donor fluorophores, respectively.

$$\textbf{\textit{Equation 3}}\ E_{rel.} = \frac{I_A}{I_D + I_A}$$

FRET titrations can be used to monitor the folding of DNA on the addition of ligands or changes in conditions. For example, during a FRET titration experiment, the fluorophore attached to the DNA would be excited at a specific excitation wavelength and if the DNA is folded there would be high FRET. If a change in cations, pH, or ligand is triggered, and this starts to unfold the DNA, a change in fluorescence can be observed.[146] FRET can be used to monitor the folding or unfolding of DNA and has been used readily for high throughput ways of measuring DNA melting, using quantitative real-time PCR instruments. There are many different methods to measure the effect of ligands that bind to specific nucleic acid sequences, however, only a few of them can be used for screening large compound libraries. Laurence Hurley used a FRET-based folding assay to screen a big library of compounds in order to determine i-motif stabilizing ligands, although he used Black Hole Quencher (BHQ) as a quencher, the principle is still the same.[148] In addition to this FRET-based assay, FRET melting is a useful biophysical technique to study the changes in the melting temperature of DNA structures when ligands are added to the DNA.[149]

Jean-Louis Mergny firstly used FRET-based DNA melting to monitor the folding of an intramolecular i-motif.[150] FRET melting can be used to determine the melting temperature of DNA, as, at low temperatures, the DNA is folded (high FRET), and when there is an increase in temperature, the DNA will unfold and there will be less FRET. By monitoring either the donor or the acceptor

fluorescence, it is possible to observe a FRET melting curve. The change in fluorescence will give a sigmoidal curve, which can be used to determine melting temperature.[151] The melting and annealing profiles should be superimposable for a reversible interaction in which the reaction is at thermodynamic equilibrium throughout the process.

Quite often with nucleic acid structures, the melting and annealing profiles are not superimposable and there is a difference in the mid-point in the transitions between melting an annealing processes, a hysteresis. The presence of hysteresis indicates that the structural transition from folded to unfolded, is slow relative to the temperature gradient of the experiment. When there is a hysteresis, $T_{1/2}$ can be used as an approximation of the T_m as it will be different for the heating and the cooling profiles, depending on the rate of temperature changes.[152–154] The Fox group performed FRET melting experiments for different intermolecular G-quadruplex sequences, the melting experiments were repeated at different rates of heating and cooling (0.5°C/min and 0.05°C/min). The results for some sequences showed hysteresis of 11.4°C (T_m 57.1°C and T_a 45.7°C) at 0.5°C/min, while at 0.05°C/min, no hysteresis is observed.[153]

The basic thermodynamic parameters can be obtained from melting curves (**Equation 4**).

$$\textbf{\textit{Equation 4}}: \Delta G = \Delta H - T\Delta S = -RT\ln(K)$$

Where ΔG is the change of Gibbs free energy, ΔH is the change in enthalpy and ΔS is the change in entropy, T is the absolute temperature, ΔC_p is the heat capacity change.

From the equations, the thermodynamic parameters depend on temperature. ΔH and ΔS are interdependent, as ΔG measured from equilibrium constant (K). The nonzero C_p indicates ΔH and ΔS respond differently to temperature. Usually during G-quadruplex unfolding process, ΔH is positive at $T = T_m$, thus, at this temperature, ΔS is also positive.

Ligands can be added to the DNA and compared to the DNA melting temperature in the absence of ligands. If the melting temperature of the DNA with the ligand is higher than the control, so the change in melting temperature is positive, this indicates the ligand binds and stabilizes the DNA structure. If the melting temperature of DNA with the ligand is lower, then the change in melting temperature is negative, suggesting a destabilizing effect of the ligand on the DNA, this is indicative that the ligand preferentially binds to another structure, such as single strand or other folded structures.

Fluorescent intercalator displacement (FID)

Fluorescent intercalator displacement (FID) is a convenient technique to identify new DNA binding ligands, it is a high-throughput method that can be used to screen large libraries of compounds against different types of DNA structures. FID assays are performed using a compound that fluoresces when it is bound to DNA. As the compound is non-fluorescent when it is in the unbound state this means if it comes off the DNA, fluorescence is lost. This means adding ligands as competition can displace the fluorescent probe and be used as a measure of relative binding affinity. This technique[155,156] was previously developed for i-motif DNA by the Waller group using thiazole orange (TO) as a fluorescent probe.[137] TO itself is not fluorescent but adding DNA provides the binding site necessary, and the probe then fluoresces. In an assay, an equilibrium time is provided to allow the TO and DNA fluorescence enhancement to reach a stable state. This varies between different DNA structures. After this, testing compounds that could be potential i-motif binding ligands are added. If they bind to the DNA structure, it will replace the TO, and once TO no longer interacts with DNA, it loses its fluorescence. Thus, the results would show a change in fluorescence and from this change, the relative affinities of ligands can be calculated. FID can also be used to determine binding constants by measuring the loss of fluorescence derived from DNA bound ligand (with increased concentration of ligand).[157] ***Equation 5*** shows the calculation of TO displacement, where I_{TO+DNA} is the fluorescence intensity of thiazole orange

and DNA, $I_{TO+DNA+ligand}$ is the fluorescence intensity of thiazole orange and DNA and ligand.

Equation 5

$$\text{TO displacement (\%)} = \left(1 - \frac{I_{TO+DNA+ligand} - I_{TO}}{I_{TO+DNA} - I_{TO}}\right) \times 100\%$$

Circular dichroism (CD) spectroscopy

Circular dichroism (CD) spectroscopy is a widely used technique that uses polarized light, for characterizing nucleic acid secondary structures or proteins, and analysis of the conformation of nucleic acid or protein under varying conditions, it can also be used to study ligand interactions.[7,158–161] Compared to other biophysical techniques, CD spectroscopy has many advantages. CD measurement can be simple, fast, and inexpensive, allowing for many experiments to be performed. It is also sensitive, the concentration of biomolecules can be low (20 μg/mL),[162] which allows us to examine precious samples or samples with poor solubility. Furthermore, simple titration with conformational changing agents can be performed, this allows us to observe the real-time change in the structure with the addition of different conditions.[163]

CD measures the difference in absorbance between both the left and right circularly polarized components of radiation. A CD spectrum is typically described in ellipticity (θ). Different biomolecules absorb the light in different near and far UV-region. For example, in peptides and proteins, amino acids absorb the UV region between 250 – 290 nm under chiral environments. Nucleic acids absorb and give rise to different CD signals in the near and far region (180 – 320 nm) depending on the different types of structures, the CD intensity varies with sequences because of the different base chromophores and glycosylic angles, for example, the most studied B-form DNA gives a positive peak at the range of 260 – 280 nm and a negative peak at ~245 nm, and i-motif DNAs have a positive band at 288 nm and a negative band at 255 nm.[164] It is well-established that CD spectroscopy can provide important information about DNA secondary structure.

Certain elements (e.g. some π-π stacking geometries, loop types and glycosidic bond angles) may not show significant CD signal,[159] therefore, the use of CD spectra for determining the folding of quadruplexes may not conclusively prove the topology.

In the study of the interaction of nucleic acid and ligands, when adding a ligand to DNA, any change in the conformation of the nucleic acid will be observed if the ligand interacts with it. CD melting experiments can be performed to investigate the effect of the ligand on the thermal stability of DNA, the plot of ellipticity at a specific wavelength against temperature gives a melting curve that can be used to determine melting temperature.

Ultraviolet spectroscopy (UV)

UV is a type of non-ionizing radiation with a wavelength between 10 – 400 nm.[165] Due to the aromatic bases, DNA absorbs UV light when irradiated. Folded and unfolded DNA absorb differently due to the difference in π-π base stacking,[166] therefore, UV spectroscopy can be used to monitor DNA folding and unfolding processes.

UV absorbance for most types of DNAs can be recorded at 260 nm, because at this wavelength, DNA absorbs most strongly. Mergny and co-workers first used, the thermal difference spectrum (TDS, the spectrum of the difference between unfolded and folded species) to characterize the type of DNA structure formed.[167] This shows for each structure where the change in the spectrum occurs. For example, in G-quadruplex and i-motif structures, TDS shows a significant negative peak at 295 nm, suggesting the wavelength is selectively sensitive to the denaturation of quadruplex structure, this is not observed in other DNA structures, thus, 295 nm could be a typical wavelength for characterizing the quadruplex structure,[167] and TDS can be a useful tool to identify the type of DNA structure present in solution.[167]

As discussed above, UV absorbance in folded and unfolded DNA is different, therefore, a UV spectrometer can be used to perform DNA melting experiments. Through monitoring the change in absorbance at a specific wavelength (e.g., 295 nm for quadruplex) with increasing temperature, this technique can be used to describe the thermal stability of DNA structure under different conditions, such as different concentrations of DNA, cations, and pHs.[151]

1.5.1 Methods of Analysis of Thermal Melting Curves

Both melting (T_m) and annealing (T_a) temperatures can be determined from thermal curves, there are a number of ways to calculate T_m and T_a, herein, four commonly used methods including the baseline method, the first derivative, fraction folded analysis, and software fitting will be discussed.

The first method to determine the melting temperature is a manual way using baselines.[166] The melting and annealing temperature corresponds to the temperature at which half of the DNA folded, and half unfolded. The example ***Figure 1.24*** presents this method.

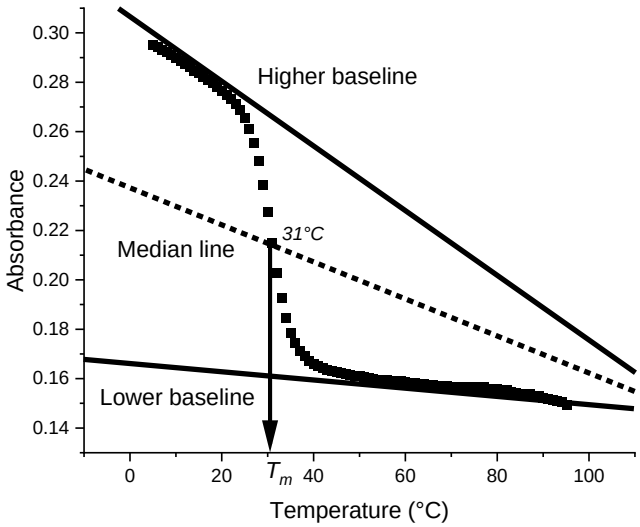


Figure 1.24 *Example of T_m determination. Baseline method, UV absorbance at 295 nm, 1 μM ZFN48-C in 10 mM sodium cacodylate 100 mM sodium chloride pH 7.0.*

The upper baseline corresponds to folded DNA, the bottom baseline corresponds to unfolded DNA, and the baselines are defined to be straight lines. The next step is to draw a median line between the two baselines and where the median intersects with the melting curve, the x-value of the intersection point can be used to determine the melting temperature. This method is the manual way, and although it is quite accurate, it would take a lot of time to find baselines and median lines, so it is not suitable for anything over medium throughput melting curves analysis.

The first derivative is also a widely used method to find the melting temperature.[168] It corresponds to the change of the melting curve from folded to unfolded, ***Figure 1.25*** gives an example of the first derivative, the lowest peak reflects the melting temperature. However, the drawback of the first derivative method is the quality of the data, and enough points need to be measured for this to work properly. For example, in CD melting analysis, where we have a large data interval, it is not the most accurate way to determine T_m, it can be used successfully when the data interval is lower, for example in our UV melting experiments data points are taken every 0.5 degrees. In spite of this, however, sometimes when we get a noisy melting curve, due to instrumental issues, if using the first derivative method, there can be lots of noise showing in the graph, which makes it difficult to find the peak point. When this occurs, sometimes it is better to look back to the original melting curve to determine T_m using another method.

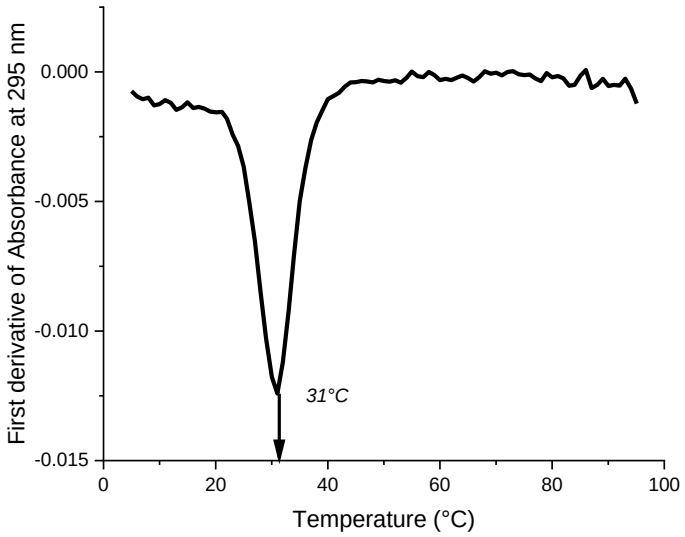


Figure 1.25 *Example of T_m determination. First derivative method, UV absorbance at 295 nm, 1 μM ZFN48-C in 10 mM sodium cacodylate 100 mM sodium chloride pH 7.0.*

The other method is to normalize an absorbance versus temperature plot into a fraction folded θ and temperature, the signal values were normalized between 0 to 1 from which the fraction folded (θ in the range of 0 to 1), was plotted against temperature from 5 – 95°C, when $\theta = 1$, $T >> T_m$, $\theta = 0$, $T << T_m$, $\theta = 0.5$, $T = T_m$. the example is presented in ***Figure 1.26***. This method works well for single transitions when the melting process is complete. It is difficult to convert to fraction folded when the melting is incomplete, also if the melting curve showed more than one transition, this method would be hard to determine several melting temperatures as they need to be individually determined, so cannot be performed in a general batch processing of data.

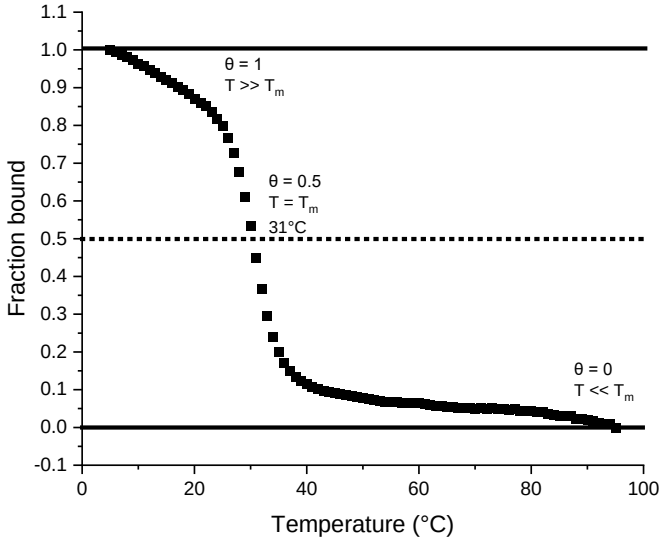


Figure 1.26 *Example of T_m determination. Fraction folded method, UV absorbance at 295 nm, 1 μM ZFN48-C in 10 mM sodium cacodylate 100 mM sodium chloride pH 7.0.*

Software methods of fitting the signal are another option, which is useful when the data interval is large (***Figure 1.27***). For example, within the Waller group CD melting is performed by measuring the whole CD spectra across a wide range of wavelengths. This has the drawback that the time taken to do this is long, so the data interval has been optimized to be a spectrum every five degrees. This enables monitoring the whole spectra, so analysis can enable choosing the best wavelength for plotting, for example, for i-motif DNA, we normally use the ellipticity at 288 nm for the plot of the curve, but when a ligand is added sometimes, we observe shifts in wavelength that are interesting and useful. As the data interval for these experiments is larger than normal, it means that the first derivative method for melting point analysis is inaccurate, and fitting the data is more appropriate.

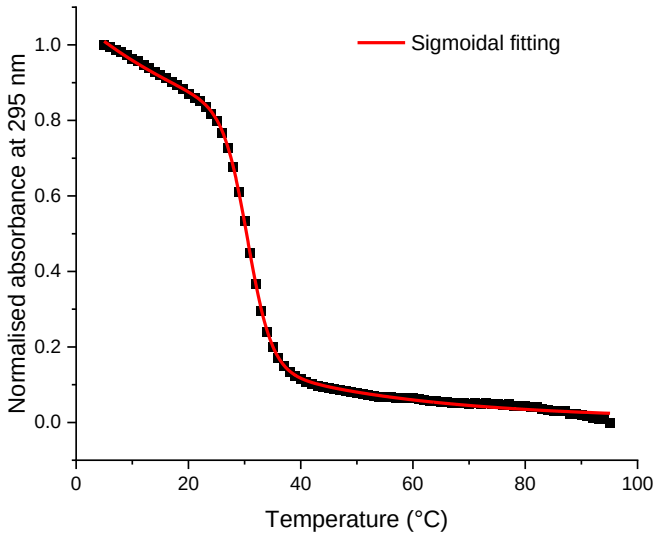


Figure 1.27 *Example of T_m determination. Fitting method, Bisigmoidal fitting for the plot of normalized ellipticity at 288 nm against temperature from 5 – 95°C, 1 μM ZFN48-C in 10 mM sodium cacodylate 100 mM sodium chloride pH 7.0.*

Fitting the melting curves enables the determination of the melting temperatures by solving the equations. E.g. for CD melting, normalized ellipticity at 288 nm, the characteristic wavelength for i-motif, can be plotted against temperature and then fitted using sigmoidal curves of Dose-Response or BiDose Response methods in OriginLab, where ***Top*** is the response at high dose, ***Bottom*** is the response to the untreated control, $\boldsymbol{T_m}$ is the dose giving a half-way response between ***Top*** and ***Bottom*** asymptotes, $\boldsymbol{logT_m}$ values reflect the X-axis which can be used to determine transitional pH or melting temperature. In the BiDose Response equation, there are two transitions, so $\boldsymbol{logT_m1}$ and $\boldsymbol{logT_m2}$ would determine pH_T1 and pH_T2 or T_m1 and T_m2. The equations of DoseResp (***Equation 6***) and BiDoseResp (***Equation 7***) are presented below:

***Equation 6**: Dose-Response method*

$$y = Bottom + \frac{Top-Bottom}{1 + 10^{(LogTm-x)}}$$

***Equation 7**: BiDose Response method*

$$y = Bottom + (Top\text{-}Bottom)\left[\frac{p}{1+10^{(LogTm1-x)h1}} + \frac{1-p}{1+10^{(LogTm2-x)h2}}\right]$$

Each analysis method can be used not only for thermal melting curves but can also be used for determining transitional pH (pH_T). The fitting method can be used to determine T_m from CD melting, especially when the curve has more than one transition, however, sometimes the curve cannot be fitted perfectly, therefore, we need to try different methods.

In this thesis, the biophysical techniques mentioned above were used in studying DNA secondary structures, e.g., using CD to characterize i-motif or G-quadruplex in different conditions like cations, also it can be used to determine transitional pH (pH_T). UV spectroscopy was used to determine the thermal stability of DNA structures under different conditions, such as different concentrations of DNA, cations, and pH. Also, these techniques were used to study the interaction between ligands and DNA structures. The medium-throughput FRET melting assay was used to study the effects of ligands on the thermal stability of DNA. The high-throughput FID assay was used to screen the library in order to discover binding hits for i-motif DNA. CD was used to observe the real-time change of ligands on DNA topology and UV was used to determine the binding affinity of a ligand to DNA.

1.6 Aims of this work

The overall purpose of this project was to explore new DNA-binding ligands, with a focus on i-motif DNA as there are not enough compounds to target this structure. Also, during the exploration of the ligands, we want to investigate the effects of ligands on the topology and stability of both G-quadruplex and i-motif DNA secondary structures, as many previous studies have only looked at double helical

DNA as a control DNA structure. As described in **Section 1.4**, many existing i-motif ligands do not show high selectivity to i-motif DNA, thus, we aim to design a strategy for specific binding ligands, this will be discussed in **Chapters 2** and **3** including the synthesis of a minor groove binder and exploring i-motif stabilizers for dual-purpose targeting combinations. TMPyP4 is a special ligand that interacts with both G-quadruplex and i-motif structures, **Chapter 4** aims to investigate its binding mechanism using biophysical methods, to support computational experiments and modelling. As described in **Section 1.3**, increased i-motif forming sequences have been identified to be stable under neutral pH, I aimed to characterize one of these i-motif forming sequences from the human genome and its potential biological context. **Chapter 5** aims to use bioinformatics to select potentially biologically important i-motifs, then characterize them using biophysical techniques and study their potential biological functions.

Chapter 2

Towards Design and Synthesis of a Novel i-Motif Ligand

2.1 Ligand Design

There are already a number of ligands that have been described to bind to i-motif structures, as described in **Chapter 1**, and new i-motif ligands continue to be discovered. However, there are limitations to the current i-motif ligands to date. Most of the ligands have low binding affinity and little selectivity to an i-motif structure specifically, so if we want to choose an i-motif forming sequence as a target for cancer therapy (or indeed any other diseases where i-motifs may be targeted), it's quite difficult to find a suitable ligand. There is a great need to develop a compound that can bind selectively to a specific i-motif structure. Here we present the idea of creating a dual-targeting ligand: one that targets a specific double helical DNA sequence and also an i-motif structure. To do this we propose to design a double helical minor groove binding ligand and attach it to a generic i-motif binding ligand. This type of ligand could be made specifically for a particular sequence of double helical DNA, which could be varied for targeting different i-motifs across the human genome. To do this successfully, we need to create a minor groove binder and combine it with an i-motif ligand that can bind and stabilize the i-motif structure. The synthesis of a proof-of-concept minor groove binder is described in **Section 2.3**, and i-motif ligand discovery is discussed in **Chapter 3**.

The i-motif structure has several characteristics that could be potential sites for ligand binding. In the i-motif structure, there are two minor (also known as narrow) and two major (also known as wide) grooves.[169] The two minor groves are extremely narrow, 3.1 Å compared to ~5.7 Å for B-DNA[51,170] and ligands that bind in this region are likely to destabilize the structure. The two wide grooves are wider (14.5–16.09 Å)[170] than B-form DNA and compounds that bind in this region could also be able to interact with the intercalated cytosine-cytosine base pairs. For example, the Su group has determined the binding mode of porphyrin compound TMPyP4 to i-motif structure, the coexistent binding modes were revealed as 80% of intercalation and 20 % of external groove binding.[171]

Other potential binding sites within i-motif structures include the loops.[169] As the loops vary between different i-motif forming sequences, these may be targeted to

develop selective ligands. If a ligand can be found or designed to bind a specific sequence in the loop of an i-motif, this can then provide some selectively to an i-motif-forming sequence.

Since DNA secondary structures are found to exist within the genome and have been implicated to be involved in gene expression, it is significant that potential ligands should have both high affinities and selectivity for one target if we want to develop them towards becoming a medicine. Therefore, we need to find ligands that can be selective for i-motif over other DNA secondary structures, given the current state of the art with regards to i-motif ligands in the literature, this is obviously challenging.

Dual probes similar to what we envision have been used to target G-quadruplex DNA. For example, the Tan group engineered a G-quadruplex-triggered fluorogenic hybridization (GTFH) probe (***Figure 2.1***) that selectively binds a specific G-rich region in *NRAS* mRNA. This probe was assembled with two parts: (1) the synthesized molecule LSCH-oa1 combined with the G-quadruplex light-on ligand ISCH-1 and an alkyne side chain and (2) an oligonucleotide complementary sequence to the selected *NRAS* G-rich sequence.[172]

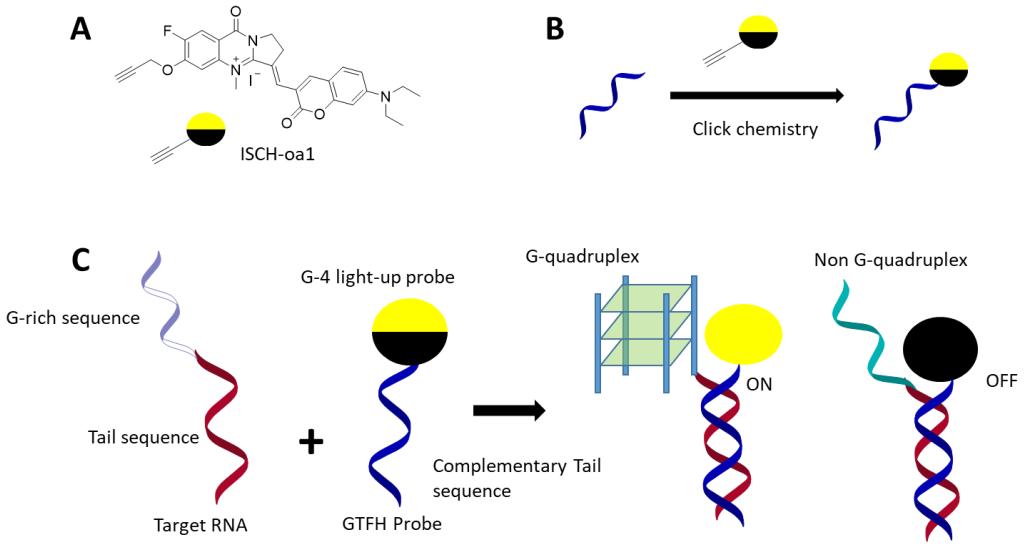


Figure 2.1 *Illustration of GTFH G-quadruplex probe. (A) Structure of ISCH-oa1. (B) GTFH probe preparation through click reaction. (C) Principle of designed GTFH G-quadruplex probe.*

The designed engineered G-quadruplex ligand suggests that it is possible to develop a highly specific and selective i-motif binding ligand using a similar approach. Inspired by the engineered G-quadruplex ligand, click chemistry can be used for designing and developing a specific i-motif ligand, three strategies are possible for the i-motif ligand design using this approach. S**trategy i** is to design a compound that can bind to the minor groove in a double helical DNA sequence and discover a small molecule that binds and stabilize the i-motif structure, and then combine the minor groove binder with the i-motif ligand to get a novel conjugate i-motif specific binder that targets only one type of i-motif (***Figure 2.2 i)***. Similar to the engineered G-quadruplex specific probe mentioned above, **strategy ii** would use a DNA sequence, which could bind the double-helical DNA as a tail to the i-motif structure, either to the single strand flanking the i-motif or even the double helix in the major groove. This could then be combined with an i-motif stabilizer, using a click reaction or other similar conjugation. The new conjugate would recognize and target an i-motif with a specific sequence (***Figure 2.2 ii)***. S**trategy iii** could be appropriate for some repetitive C-rich sequences, and tandem repeats, where multiple i-motifs could form alongside each other similar to "beads on a string." In these situations, two i-motif stabilizers could be linked, to enable the targeting of closely connected i-motifs (***Figure 2.2 iii)***. As a proof of concept, we started with **strategy i**, as this was the simplest methodologically.

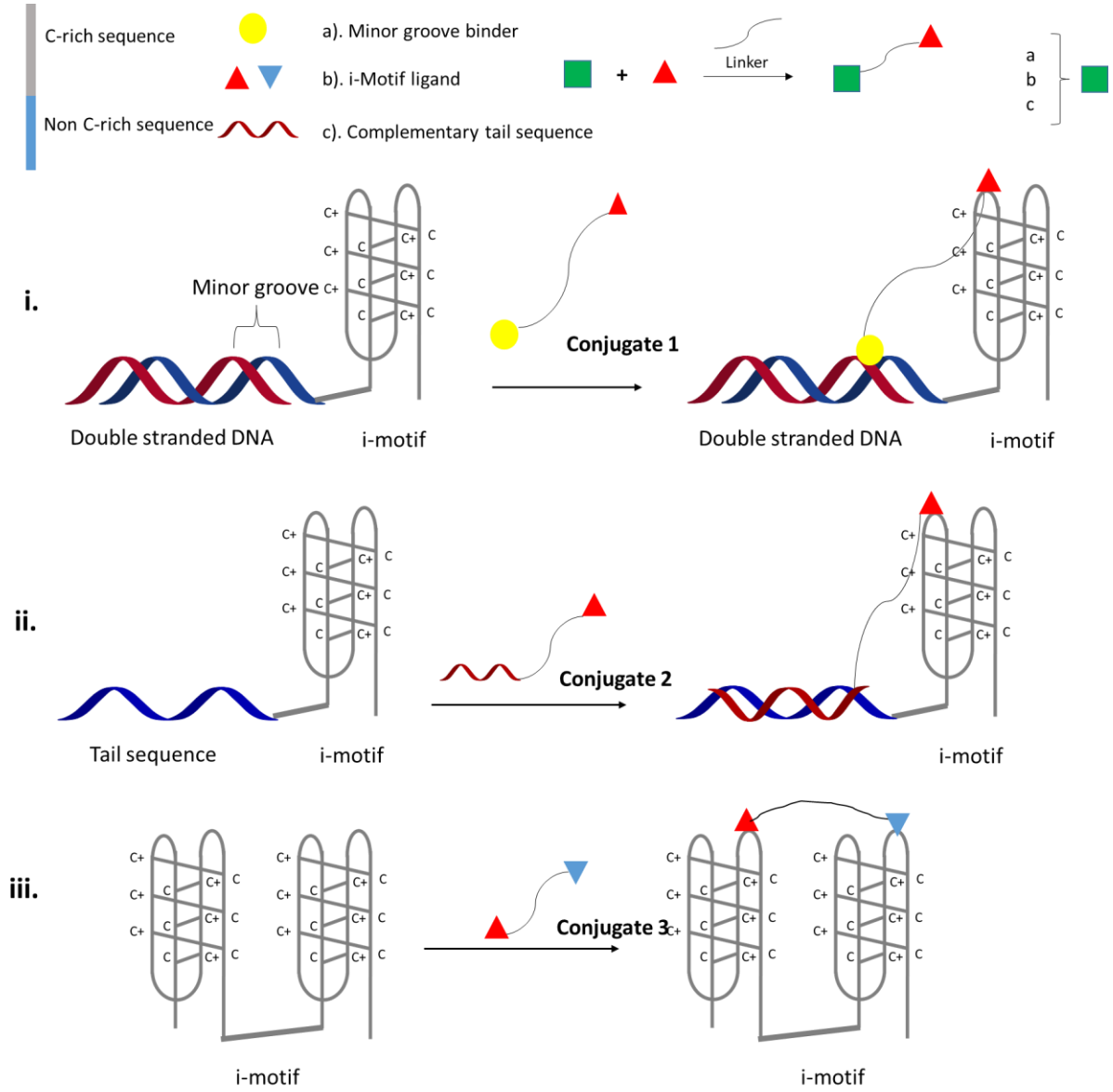


Figure 2.2 *This schematic picture shows three strategies to design a new i-motif selective ligand.* ***i****). Illustrated the combination of a minor groove binder and an i-motif stabilizer.* ***ii****). Illustrated the strategy of the combination of a DNA sequence and an i-motif ligand.* ***iii****). Described the click reaction between two i-motif ligands.*

Strategy i has two requirements: i-motif binding ligands and minor groove binders. In this chapter, work focused on the synthesis of a proof-of-concept minor groove binder and the investigation of its selective to double-stranded DNA.

2.2 Minor Groove Binders

Small molecules that bind specifically to DNA sequences in the human genome could potentially be used in human medicine.[173,174] The DNA minor groove is a potential target for a large number of non-covalent binding molecules.[175] Many small molecules have been discovered and synthesized that bind to the minor groove and are known as minor groove binders. Many of these minor groove-binding ligands target either AT or GC-rich sequences,[176,177] the Fox group investigated DNA sequence preferences of some classic minor groove using DNase I footprinting,[178] they identified the binding preference for distamycin A is AATT=TTTT>ATTA/TAAT>ATAT=TATA>TTAA, and for netropsin is AATT>TAAT=TTAA=ATAT>TATA, i.e. AT-selective minor groove binders bind less well to sequences that contain a TpA step.[178] All these compounds show several similar features, e.g., most of them are crescent-shaped molecules, and provide hydrogen bond donors on the inside edge; they all contain aromatic rings and have at least one positive charge.[176]

Some minor groove binders also have a high affinity and specificity for particular sequence types, and they also show promising biological effects such as antitumor, antibacterial, and antiviral properties.[179] Much past research on minor groove binders has focused on the development of anticancer agents.[180] Examples of some classic natural or synthesized minor groove binders are presented in ***Figure 2.3***.

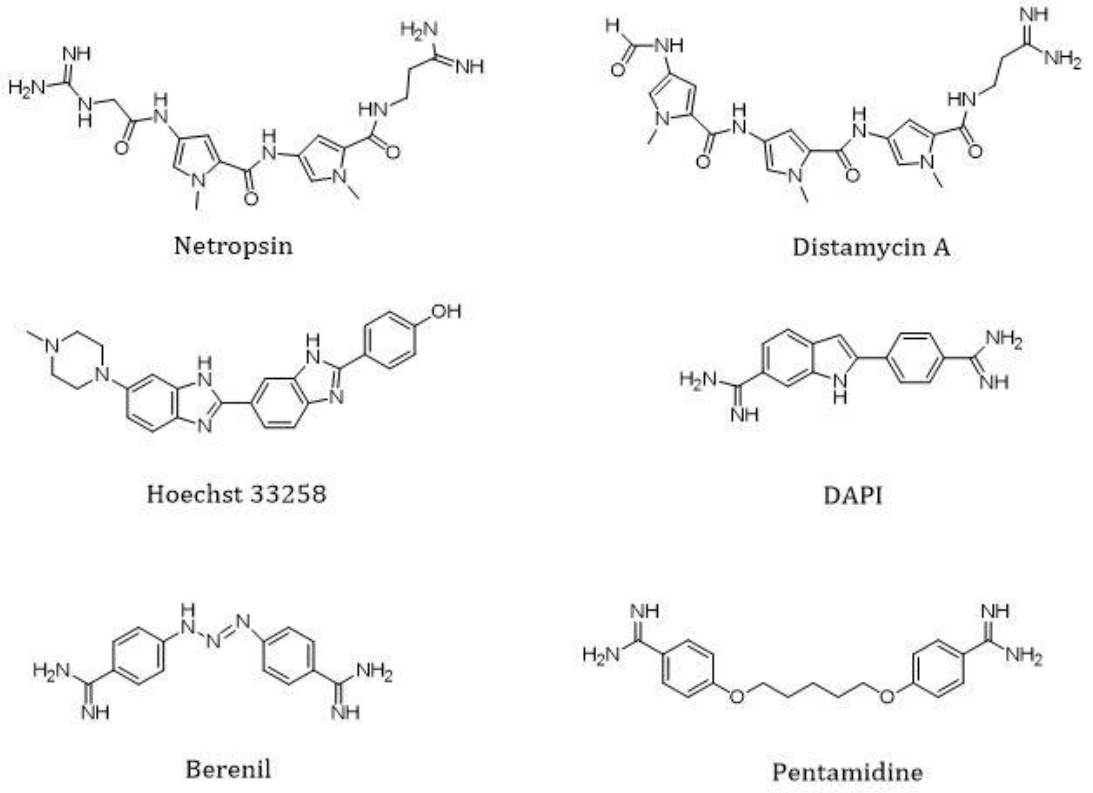


Figure 2.3 *Examples of classic minor groove binders.*

Distamycin A, an N-methypyrrole (Py) polyamide, is a natural antibiotic with antiviral and antibacterial activity, it is a classic minor groove binder.[181,182] This small molecule can bind to AT-rich sequences in the minor groove as 1:1 and 2:1 DNA-polyamide complexes. Distamycin A, which contains three pyrrole rings, binds to AT-rich regions of DNA of at least 4 base pairs in length, in addition, extended compound structures can increase the length of AT tract targeted. Another classic minor groove binder is netropsin, a pyrrole polyamide that contains two pyrrole rings linked by amide bonds, it also binds to AT-rich sequences.[183] Both distamycin and netropsin bind to AT-tracts which are four bases long. Both these minor groove binders can bind in a 1:1 mode with double helical DNA,[184,185] but only distamycin binds in a 2:1 mode.[186] The reason why netropsin doesn't bind in the 2:1 mode is because it has positive charges at both ends[187] (unlike distamycin). AT-selectivity of both ligands is due to the fact that the minor groove in AT-sequences is generally lower than GC, which gives a

smaller pocket for binding in AT-rich sequences compared to GC-rich sequences. Moreover, the binding sites size for distamycin is larger than that for netropsin. However, minor groove binders also have to navigate a potential steric clash with the 2-amino group of guanines in the minor groove in GC-rich sequences, so AT-rich sequences are generally preferred. This is the main reason for their AT selectivity: GC base pairs are sterically excluded.

The synthetic compound Hoechst 33258 and its derivatives are widely used as fluorescent cytological stains in the regulation of gene expression.[173,188–191] Similar to other minor groove-binding ligands, it binds preferentially to AT regions.[192] In addition to this, recent research from the Tokmakoff group, studying the binding of Hoechst 33258 with the sequence $d(CGCAAATTTGCG)_2$, indicated that the ligand also strongly destabilizes three CG base pairs adjacent to its binding site.[193–195]

DAPI and berenil have been shown to specifically bind to AT-rich sequences of double-stranded DNA, they bind to the minor groove with phenyl and indole rings interacting with a three-base pair ATT sequence.[196,197] Pentamidine, an aromatic diamidine, has a binding preference for AT-rich regions on minor grooves with at least five AT base pairs.[198]

To investigate and develop a minor groove binder with high binding affinity, we also need to know about specific structures for the recognition of base pairs. Peter Dervan and co-workers have spent decades investigating different types of minor-groove binding compounds. The key development came when they developed hairpin pyrrole–imidazole polyamides as a class of programmable small molecules that bind to the minor groove of DNA. These compounds essentially mimic a 2:1 binding mode by combining two small minor groove binders with a linker, so they bind in the minor groove as a hairpin. The affinities of these compounds were similar to transcription factors and have been shown to alter gene expression in cells by interfering with transcription factor/DNA binding sites. In their work towards specific DNA recognition, they identified ligand-based pairing rules in the minor groove for recognition by N-methylpyrrole (Py) and N-methylimidazole

polyamides (Im).[199] Dervan's group designed and investigated a number of compounds where Equilibrium association constants were determined at two five-base-pair sites for a series of four six-ring hairpin polyamides, in order to test the relative energetics of the four aromatic amino-acid pairings opposite G·C and A·T base pairs in the central position.[199] They demonstrated the correlation between DNA sequence and Py/Im polyamide sequence (***Table 2.1***).

Table 2.1 *The pairing codes for minor groove recognition by the Dervan hairpin polyamides, Py, Im, and Hp (N-methyl-3-hydroxypyrrole) show the favoured and disfavoured recognitions.*

Pair	GC	CG	AT	TA
Im/Py	+	–	–	–
Py/Im	–	+	–	–
Py/Py	–	–	+	+
Im/Im	–	–	–	–
Hp/Py	–	–	–	+
Py/Hp	–	–	+	–

Im/Py pairing specifically recognizes G-C base pairs; the Py/Im shows preference to target C-G base pairs; the Py/Py pairing targets A-T/T-A base pairs, and Im/Im does not show any preference for any of the DNA base pairs.[199] The same team then used the N-methyl-3-hydroxypyrrole (Hp) to pair across from Py to study thymine recognition.[200] They identified Hp/Py pairing showed to distinguish T-A from other base pairs, and Py/Hp recognizes the A-T base pair specifically.[200] The remarkable results guided the further development of minor groove-binding molecules.

In the study of the development of minor groove binders with high binding affinity, the Dervan group used a classic hairpin Im/Py polyamide PyPyPyPy-(R)^{H2N}γ-ImP**y**PyIm-ß-Dp (Dp is dimethylaminopropylamine) as the reference for the ligand design, which was expected to bind the 5′-TCTACA-3′ sequence.[201] They developed a new polyamide compound PyPyPyPy-(R)^{H2N}γ-ImP**y**-ß-Im-ß-Dp (***Figure 2.4***). The

Im in the Dervan's hairpin compounds can recognise guanine of the minor groove, as it can form a hydrogen bond to the 2-amino group of guanines.[202] The aromatic Py was replaced by ß-alanine (ß-Ala), and both polyamides were tested with the 5′-TCTACA-3′ sequence using quantitative DNase I footprinting titrations. The repeat distance between the units of these polyamides doesn't exactly match the repeat distance between the base pairs, thus they come out of alignment for longer binding molecules. Since ß-alanine is quite flexible and is therefore able to relax this constraint, furthermore, it prefers to interact with AT-sequence.[202] Hairpin polyamide **9** did not show a preference for the sequence, while the new polyamide **10** showed a 5 to 25-fold selectivity for the DNA sequence.[201] The combination with ß-Ala increases DNA-binding properties of N-terminal pyrrole polyamide,[201] thus incorporations with ß-Ala could be used for developing a good minor binder.

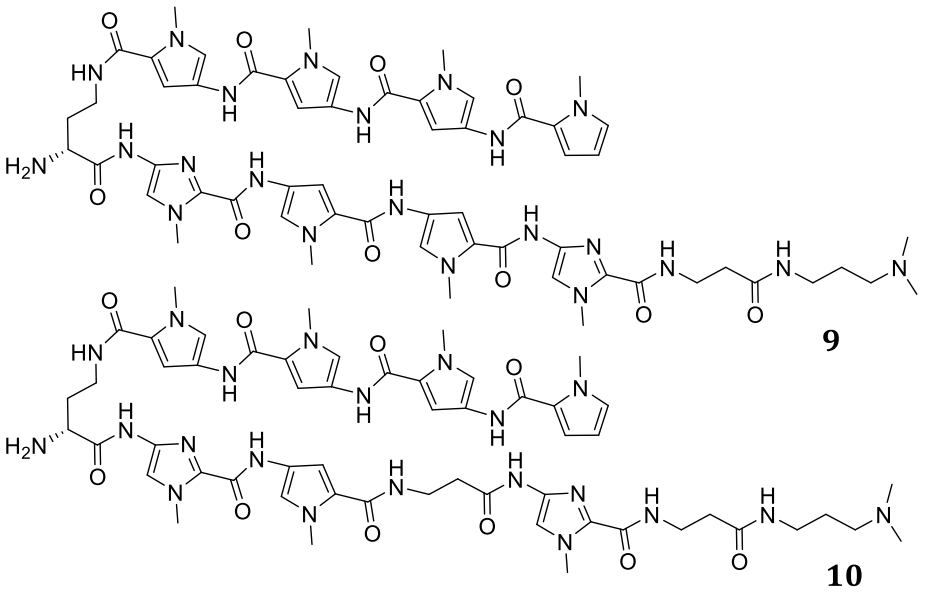


Figure 2.4 *Structure of polyamides. Compound 9: PyPyPyPy-(R)^{H2N}γ-ImPyPyPyIm-β-Dp, compound 10: PyPyPyPy-(R)^{H2N}γ-ImPy-β-Im-β-Dp.*

In our design, we envisaged a minor-groove binding compound coupled with an i-motif binding compound to form a conjugate capable of potentially targeting a single i-motif-forming sequence within the genome. As a proof of concept, we initially wanted to start with a simplified minor groove binding compound, and if successful, we could then design better ligands using Dervan's sequence-specific

compounds. In this chapter, we describe the synthesis of a distamycin derivative with the potential capability to be attached to a linker/i-motif ligand and its interactions with DNA.

2.3 Synthesis of a Minor Groove Binder

The simplified distamycin analogue compound **PyPyPyβDp** was firstly synthesized by the Gu Group in 2002.[203] The same team investigated the interaction between this ligand and duplex DNA using steady-state absorption and fluorescence spectroscopies and time-resolved fluorescence spectroscopy methods, and then they verified that **PyPyPyβDp** interacts and binds into the minor groove of double-stranded DNA.[186] This compound has the same arrangement of pyrroles as distamycin, and we would predict that it should have similar sequence binding properties. This compound has advantages synthetically in that it can be altered easily, does not take as long to make as the Dervan compounds and provides a simplified proof of concept that can be built upon for minor groove targeting. Gu and co-workers did not investigate the selectivity of the compound to the minor groove, so one of the first things required after synthesis would be to examine the specificity towards the minor groove, compared to competing DNA structures. Given our conjugate designs, this compound ideally should not bind i-motif DNA.

Gu and co-workers previously synthesized **PyPyPyβDp** in solution. Within the Searcey group, there is a lot of experience working with the synthesis of peptides using solid phase peptide synthesis, so we made this compound using Fmoc[204] solid phase peptide synthesis protocols; the scheme is presented in ***Figure 2.5***.

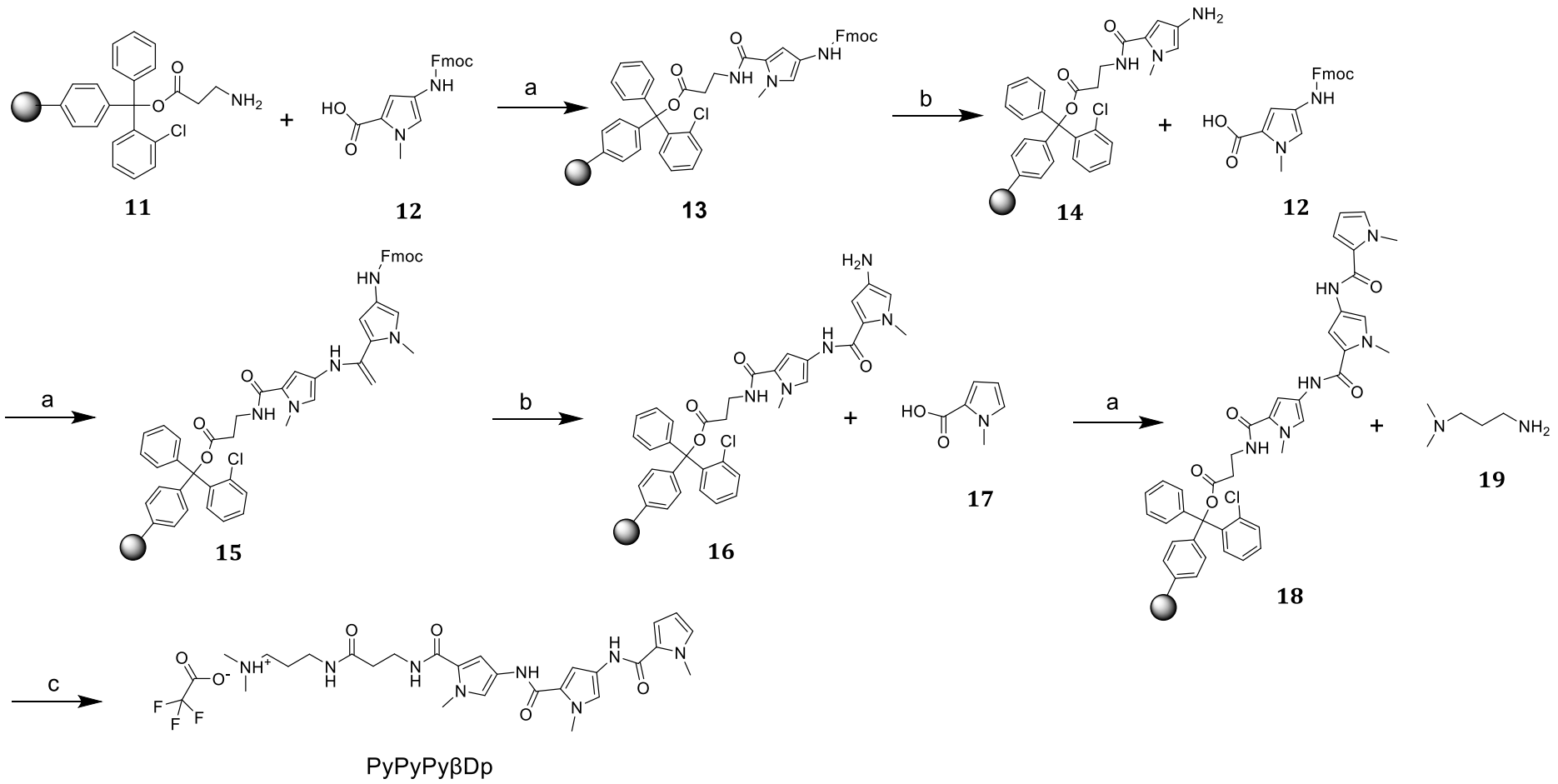


Figure 2.5 *Scheme depicting the solid-phase synthesis of* ***PyPyPyβDp****. Reagents and conditions: (a) Hexafluorophosphate Benzotriazole Tetramethyl Uronium (HBTU)/hydroxybenzotriazole (HOBt), and N,N-Diisopropylethylamine (DIPEA), room temperature, 6 hours. (b) 40% piperidine in Dimethylformamide (DMF), room temperature. (c) Cleavage of a resin base, 2:8 of trifluoroethanol and dichloromethane (DCM), room temperature.*

PyPyPyβDp was synthesized using a standard solid phase peptide synthesis F-moc-protocol on a *Syro II Fully Automated Parallel Peptide Synthesizer*, starting from 100 mg chlorotrityl chloride resin preloaded with ß alanine (**1**). The resin was transferred with immobilized precursor peptide to a fitted syringe and washed with DCM three times before coupling. Reagent **12** (996.6 mg, 2.7 mmol, 5 equivalents) was added into **11**, then, HBTU/HOBt (1706 / 608 mg, 4.5 mmol, dissolved in DMF), and DIPEA (2068 mg, 16 mmol, dissolved in DMF) were added and the reaction mixture was shaken for 6 hours at room temperature to give the intermediate **13**. Deprotection of the Fmoc group from the N-terminus of the resin-bound chain was achieved by adding 40% piperidine in DMF and was allowed to react for 10 min. The reagents were then drained and replenished to react again to give intermediate **14**. A further portion of **12** (996.6 mg, 2.7 mmol, 5 equivalents) was added into **14** and HBTU/HOBt (1706/608 mg, 4.5 mmol, dissolved in 10 mL), and DIPEA (2068 mg, 16 mmol, dissolved in 8 mL DMF) was used again for coupling and this mixture shaken at room temperature for 6 hours to give **15**. Removal of the Fmoc group was performed as described above, using 40% piperidine to give intermediate **16**. Building block **17** (156.4 mg, 1.2 mmol, 5 equivalents) was added to **16** and coupled using HBTU/HOBt (758/270 mg, 2 mmol, dissolved in 4.5 mL DMF), and DIPEA (1008 mg, 7.8 mmol, dissolved in DMF) as coupling reagents, the reaction was shaken at room temperature for 6 hours to give intermediate **18**. Finally, aliphatic amine **19** was added into the reaction coupled with HATU (1520 mg, 4 mmol, 5 equivalents). The final compound was cleaved from the resin using 1% trifluoroacetic acid in DCM. The sample was left in the TFA solution for 45 mins, drained and the reagents replenished four times in total, the final cleavage was left overnight. Purification of the final product **PyPyPyβDp** was achieved by reverse-phase preparative HPLC using 5% MeOH in H_2O and 0.05% TFA (***Appendix 1***) in 76% yield. MALDI mass spectrometry was consistent with the correct compound ***(Appendix 2)***. The mass calculated for **PyPyPyβDp** $C_{26}H_{37}N_8O_4$ was 525.29 and 525.2938 was found.

PyPyPyβDp was also characterized by ^1H NMR. The ^1H NMR was performed in methanol-D4 (***Figure 2.6***). Three singlet peaks at 3.93, 3.91, and 3.87 ppm

correspond to the methyl group attached to the N on an aromatic ring. Two methyl groups attached to the end of N showed one singlet peak at 2.84 ppm. The seven aromatic protons were shown in peaks between 7.2 – 6.1 ppm. The doublet of doublets at 6.1 ppm represents one proton at position 34, and the multiplet peaks at 6.89 and 6.86 ppm represent the protons at positions 35, and 36. The quintuple peak between 1.96 – 1.83 ppm represents the protons at position 13.

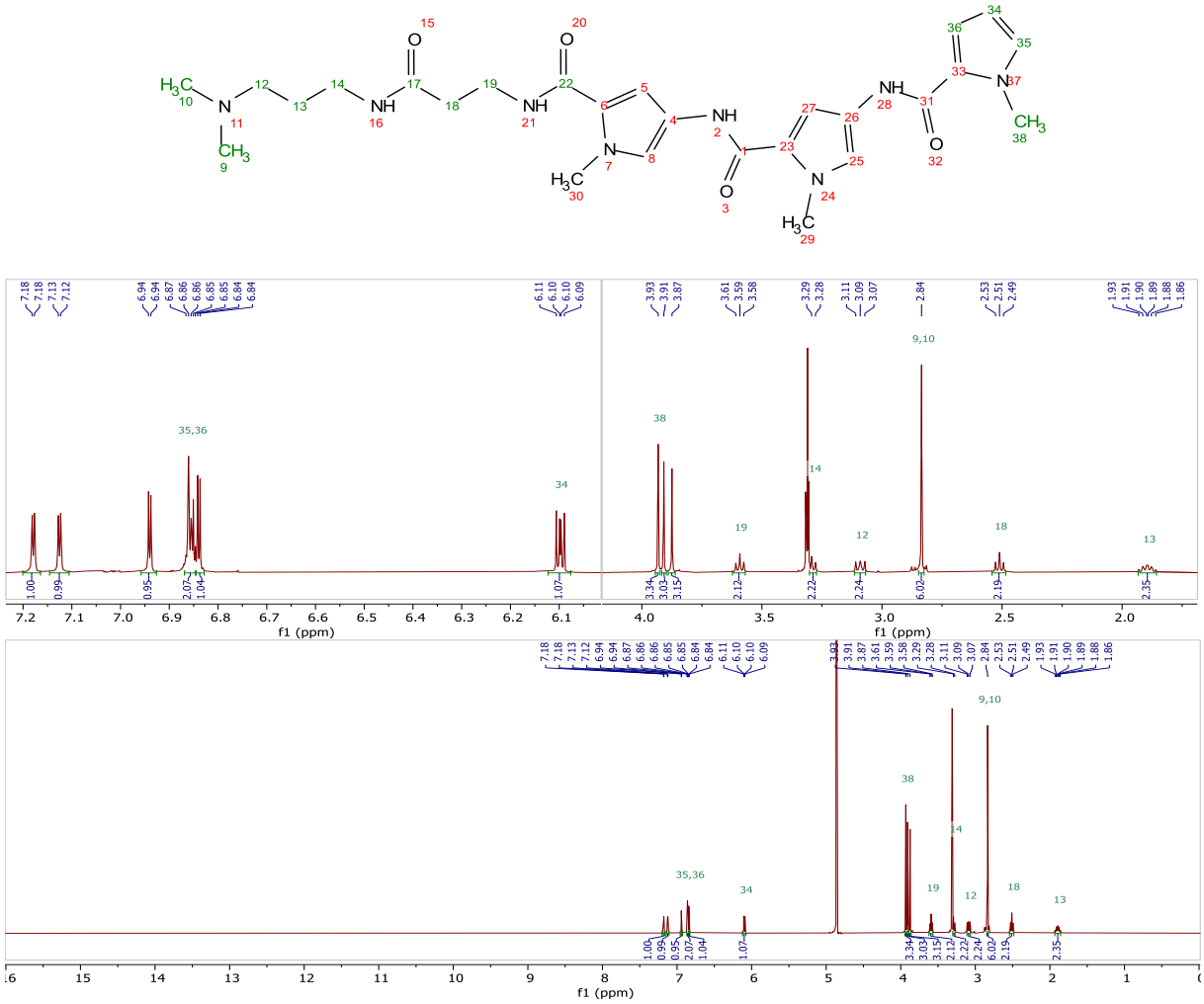
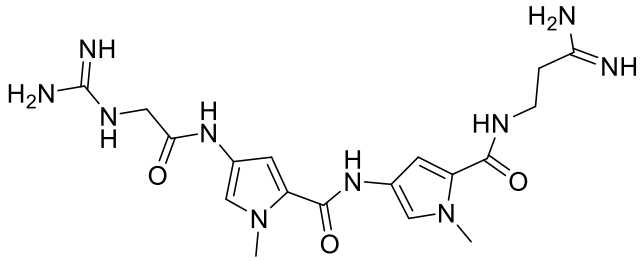


Figure 2.6 *^{1}H NMR of* PyPyPyβDp *with assignments.*

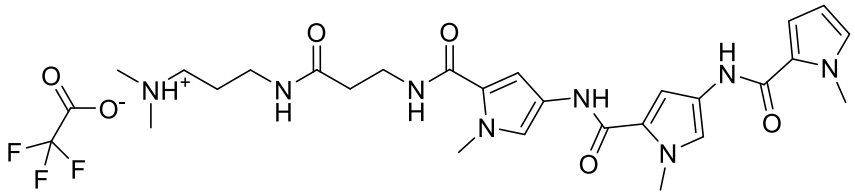
After the synthesis of **PyPyPyβDp**, the next step was to investigate its preference for double-stranded DNA structures, in the next section, several techniques were used to examine the compound.

2.4 Interaction between Minor Groove Binders and DNAs

The purpose of this project is to find if our minor groove binder **PyPyPyβDp** binds to double helical structure with high selectivity. The first step is to know its binding properties and selectivity profiles with other types of DNAs, for example, G-quadruplex and i-motif structures, therefore, **PyPyPyβDp** was tested with different DNA sequences using biophysical techniques. As a control DNA structure, we were interested in general B-form DNA binding, so used a previously used double helical sequence. However, the sequence that would be best for this minor groove binder **PyPyPyβDp** should contain at least four consecutive AT-base pairs and avoid the presence of TpA steps. With the benefit of hindsight, the choice of sequence for testing duplex binding in Table 2.2 is not appropriate for an AT-selective minor groove ligand. For determining the binding of **PyPyPyβDp** to duplex sequence, further experiments are required to be performed in the future with additional AT-rich sequences. In addition, to our best knowledge, minor groove binders have never been studied with i-motif structures, so this would be the first time they have been assessed. The other classic existing well-studied minor groove binder, **netropsin**[205] (***Figure 2.7***) was also chosen for testing with DNAs. We intended also to compare with Distamycin, but there were no commercially available sources of the compound in stock at the time and the material within the Searcey group had degraded.



Netropsin



PyPyPyβDp

Figure 2.7 *Structures of netropsin and PyPyPyβDp used for biophysical examinations.*

We studied the binding of the compounds against a double-stranded (DS) DNA sequence (for FRET melting experiments, HEG linker is used to connect both DS and DS complementary sequences), the G-quadruplex from the human telomere (*hTelo*G), and i-motif forming sequences from the human telomere (*hTelo*C) and the promoter regions of *Nrf2*, *DAP*, *ATXN2L*, *Hif1α* and *c-MYC* (***Table 2.2***).

Table 2.2 *Custom oligonucleotide sequences used in the study of minor groove binders.*

Name	Sequence 5′ → 3′
*hTelo*C$_{FRET}$	FAM-TAA-CCC-TAA-CCC-TAA-CCC-TAA-CCC-TAMRA
*hTelo*G$_{FRET}$	FAM-GGG-TTA-GGG-TTA-GGG-TTA-GGG-TAMRA
DS$_{FRET}$	FAM-TAT-AGC-TAT-A-HEG(18)-TAT-AGC-TAT-A-TAMRA
Hif1a$_{FRET}$	FAM-CGC-GCT-CCC-GCC-CCC-TCT-CCC-CTC-CCC-GCG-C-TAMRA
DAP$_{FRET}$	FAM-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CC-TAMRA
ATXN2L$_{FRET}$	FAM-CCC-CCC-CCC-CCC-CCC-CCC-CCC-CCC-TAMRA
Nrf2$_{FRET}$	FAM-CCC-TCC-CGC-CCT-TGC-TCC-CTT-CCC-TAMRA
c-MYC$_{FRET}$	FAM-CCT-TCC-CCA-CCC-TCC-CCA-CCC-TCC-CCA-TAMRA
*hTelo*C	TAA-CCC-TAA-CCC-TAA-CCC-TAA-CCC
*hTelo*G	GGG-TTA-GGG-TTA-GGG-TTA-GGG
DS	GGC-ATA-GTG-CGT-GGG-CGT-TAG-C
DS$_{comp}$	GCT-AAC-GCC-CAC-GCA-CTA-TGC-C
Hif1a	CGC-GCT-CCC-GCC-CCC-TCT-CCC-CTC-CCC-GCG-C
DAP	CCC-CCG-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CC
ATXN2L	CCC-CCC-CCC-CCC-CCC-CCC-CCC-CCC
Nrf2	CCC-TCC-CGC-CCT-TGC-TCC-CTT-CCC
c-MYC	CCT-TCC-CCA-CCC-TCC-CCA-CCC-TCC-CCA

Three biophysical techniques were used to investigate the interactions of the two minor groove binders with DNA structures. FRET melting experiments were used to observe a change in thermal stability upon the addition of compounds, FID assays were used to determine the relative binding affinities of the ligands, and CD titrations and melting experiments were used to investigate the effect of the ligand on the structure, topology, and the thermal stability of DNA. In the FID assay, *ATXN2L* was not used, as it is not a good candidate for this technique because of the low fluorescence enhancement observed after the equilibration of thiazole orange.[206] To keep our experiments consistent, all the samples were prepared in 10 mM sodium cacodylate and 100 mM potassium chloride buffer at desired pH (*hTelo*C in pH 5.5, *Nrf2*, and *c-MYC* in pH 6.5, DS, *hTelo*G, *DAP*, *ATXN2L* and *Hif1a* in pH 7.0). This buffer condition was used because sodium cacodylate is chemically stable and the pH does not vary with temperature.[110,150]

It is a commonly used buffer for studying nucleic acids, the pH in this solution can be maintained in the range of 5.0 – 7.4 and the DNA would be stable under such conditions. Additional potassium chloride was used to introduce K^+ to mimic the physiological cationic conditions.

Biophysical results

Netropsin and **PyPyPyβDp** were firstly tested with DNA sequences using FID assays (as described in **Section 1.5**). The experiments were performed with 0.5 μM of DNA with the ligands over a concentration range from 1 – 100 μM, all the experiments were repeated in triplicate (percentage of TO displacement are presented in **Appendix 3**). From the results (***Figure 2.8***), there is no significant displacement for both **netropsin** and **PyPyPyβDp** at low concentrations (1, 2.5, and 5 μM), however, with increasing the concentration of ligands, the displacement by netropsin is observed. However, **PyPyPyβDp** did not show a big difference in displacement for most of the DNA sequences, compared to the change in TO displacement by **netropsin**. After increased concentration up to 50 and 100 μM, high negative displacement appeared to *c-MYC* i-motif, indicating enhanced fluorescence. The reason could be a high concentration of **PyPyPyβDp** shifts the equilibrium of single-stranded structure to i-motif structure, therefore, more i-motif species folded and more TO would bind to DNA, then increased fluorescence. **PyPyPyβDp** may interact differently with DNA structures, in a different binding mode to TO, which would mean if it does bind, it does not displace the ligand, or **PyPyPyβDp** may bind to TO instead of DNA structures. However, such a high percentage of TO displacement only happened at large ratios where the ligands were in excess of concentration to DNA, the top concentration of ligand was 100 μM, which was 200-fold compared to DNA (0.5 μM).

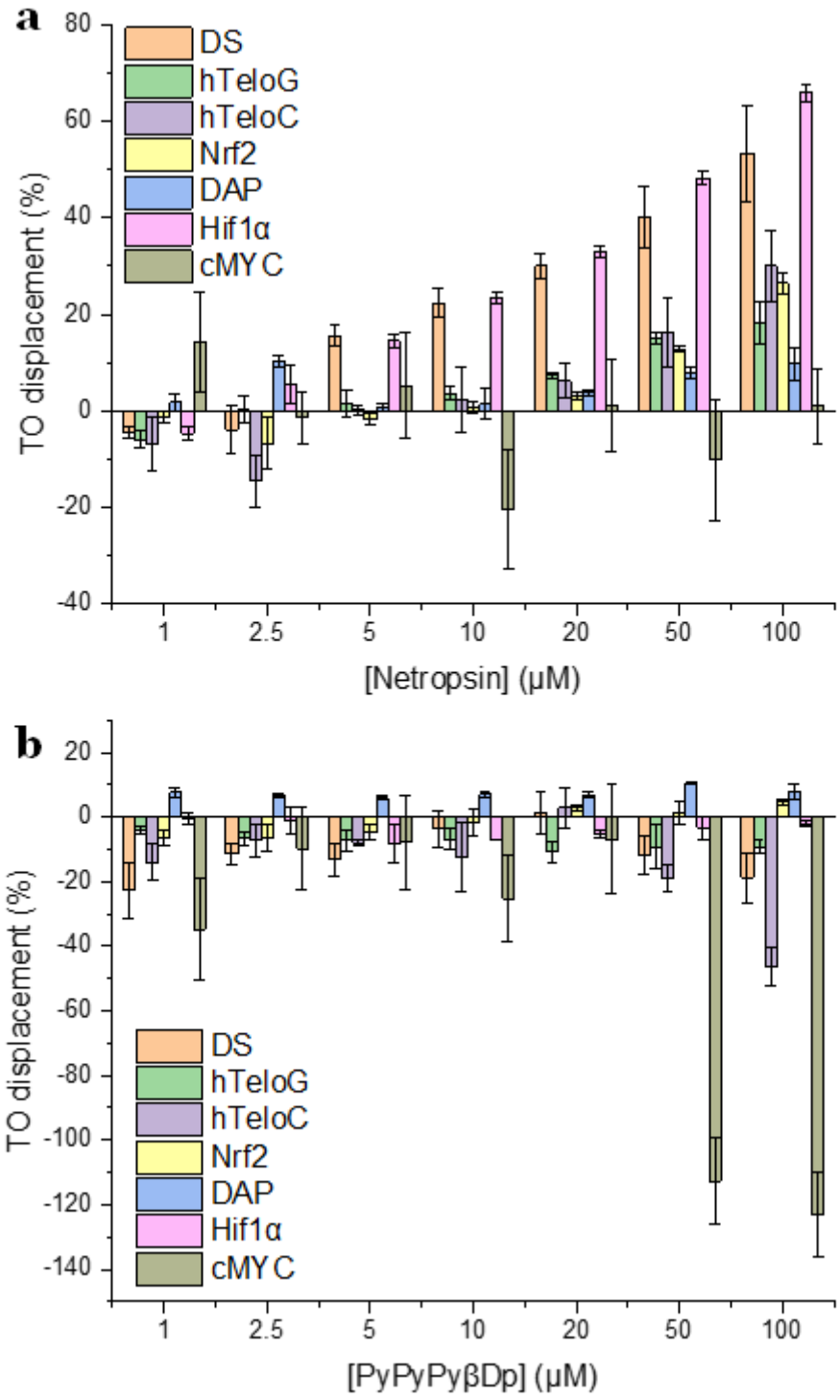


Figure 2.8 *Percentage of thiazole orange displacement of netropsin (top) and PyPyPyβDp (bottom) with different DNA sequences, all experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride buffer, [DNA] = 0.5 μM.*

Then both minor groove binders were examined using FRET melting experiments to investigate their effects on the thermal stability of the DNAs. The FRET melting experiments were performed with the concentration of both 1 μM and 10 μM of ligands, each experiment was performed in triplicate, and the change in melting temperature (ΔT_m) was calculated using the first derivative of each melting curve. The change in melting temperature values is shown in ***Table 2.3***. The FRET melting curves are presented in **Appendix 4**.

Table 2.3 *T_ms and ΔT_ms of netropsin and PyPyPyβDp with selected DNA sequences measured by FRET melting in the presence of 1 μM and 10 μM of ligands, buffer containing 10 mM sodium cacodylate, and 100 mM potassium chloride. Errors show the standard deviation from triplicate experiments.*

		ΔT_m (°C)			
	T_m(°C)	Netropsin		PyPyPyβDp	
Sequence		1μM (5 eq)	10μM (50 eq)	1μM (5 eq)	10μM (50 eq)
DS	64.2 ± 0.3	1.1 ± 0.5	4.5 ± 0.2	-0.3 ± 0.4	-0.3 ± 1.9
*hTelo*G	63.2 ± 0.5	0.2 ± 0.8	4.4 ± 1.6	-0.4 ± 1.0	0.8 ± 1.5
*hTelo*C	46.5 ± 0.2	-0.5 ± 0.2	0 ± 0.4	-0.1 ± 0.2	-0.2 ± 0.3
Nrf2	60.9±0.6/66.9 ±0.7	0.2 ± 0.8/0.4 ± 1.2	-2.0 ± 0.9/-9.8 ± 3.0	0 ± 0.6/0.7 ± 0.7	0 ± 0.4/-3.3 ±1.6
DAP	28.1 ± 0.5	NA	-1.2 ± 0.1	-0.6 ± 0.9	0.6 ± -1.2
ATXN2L	30.1 ± 0.4	-0.4 ± 0.3	-0.1 ± 0.3	-0.2 ± 0.2	-0.5 ± 0.3
Hif1a	57.6 ± 0.6	0.9 ± 0.6	-1.4 ± 0.4	3.9 ± 0.7	2.3 ± 0.7
c-MYC	27.8 ± 0.2/66.2±0.5	-0.5 ± 0.1/-3.7±0.8	-12.2 ± 1.3	-0.2 ± 0.1/-3.5±0.7	-0.2 ± 0.1/0.2±0.3

Based simply on the change in melting temperature values, at 1 µM (5 eq) of ligands, **netropsin** showed the ability to slightly stabilize double-stranded DNA (ΔT_m = +1.1°C), but no change for the rest of the sequences, on the other hand, at the same concentration of ligand, **PyPyPyβDp** displayed stabilization to Hif1α with ΔT_m +3.9°C, but had no significant effect on the rest of the DNA sequences. at 10 µM (50 eq), netropsin shows stabilization to double-stranded (ΔT_m = +4.5°C) and hTeloG (ΔT_m = +4.4°C), and destabilization to Nrf2 (ΔT_m = -2/-9.8°C), *DAP* (ΔT_m = -1.2°C) and *Hif1α* (ΔT_m = -1.4°C). There was a large destabilizing effect observed for *c-MYC* in the presence of 10 µM (50 eq) **netropsin** (ΔT_m = -12.2°C), furthermore, **netropsin** changed the shape of this melting curve, the DNA in the absence of netropsin showed two clear transitions, which is indicative of two different DNA species that could be folded, but the curve with **netropsin** has only one transition, suggesting also some sort of interaction happened (***Figure 2.9***). Analogous experiments with **PyPyPyβDp** did not switch the shape in the same manner.

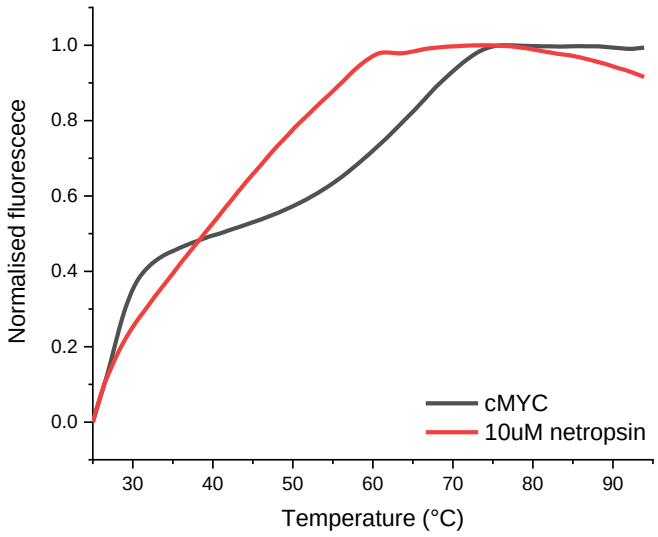


Figure 2.9 *Example of FRET melting curve of 0.2 µM c-MYC with 10 µM (50 equivalents) netropsin in 10 mM sodium cacodylate and 100 mM potassium chloride buffer at pH 6.5. The black line is DNA control, the red line is DNA with the ligand.*

While **PyPyPyβDp** does not show a significant change in melting temperature with double-stranded DNA, it only stabilizes the i-motif forming sequence from *Hif1α* (ΔT_m = +2.3°C at 10 μM). This high concentration may show other effects other than specific binding, as aggregation effects may also occur. We would like to see if there is a real DNA-ligand interaction in that ligands can interact with specific binding sites like loops or grooves, and this would happen at small ratios of ligand to DNA or with the same equivalent of ligand to DNA.[206] Therefore, the explanation for the observed stabilization and destabilization could be the ligands interact with the labelled fluorophores instead of DNA structures, or some non-specific interactions occurred that have an impact on melting temperature.

Biophysical techniques such as FRET melting, and FID assay give a quantitative observation of how the ligands affect the thermal stability or bind to DNA structures, there is, however, a limit for these two techniques that both techniques work based on the folded DNA conformation.[207] Small molecules could likely interact with a particular site of the sequence rather than the folded DNA structure, the conformation of DNA secondary structure may be re-modelled during the interaction. Therefore, need to use other techniques to support the results e.g., CD.[208]

Based on the results of the FRET and FID assay, *Hif1α*, *DAP*, *Nrf2*, and *c-MYC* i-motifs were chosen for further study with **netropsin** and **PyPyPyβDp** using CD spectroscopy. Double-stranded and *hTelo*G were also examined to compare with i-motif. DS, *hTelo*G, *Hif1α*, and *DAP* were performed at pH 7.0, and *Nrf2* and *c-MYC* were prepared at pH 6.5. CD was used first to confirm that these four sequences formed featured structures under experimental conditions at desired pHs. Under these conditions, ligands were titrated into DNA up to a concentration of 50 μM All the CD titrations were corrected by subtracting the CD signal of the ligand alone. This was an important procedure to perform as part of the processing, to account for any signals from ligand alone, however, the CD of the ligand may not be the same in the presence of the nucleic acid, so this could cause artifacts.

CD signal of double-stranded DNA showed a positive peak between 270 – 280 nm, and a negative one at ~250 nm, indicating a double helical structure formed at pH 7.0. Upon the addition of **netropsin**, there was a significant decrease in CD signal at ~280 nm, and a bathochromic shift from ~280 nm to ~287 nm was also observed (***Figure 2.10 a***), however, no isoelliptic point was observed, which is indicative of not just one bound species formed. CD melting experiments also showed both destabilizing and stabilizing effects on the stability of the double-stranded structure with ΔT_{m1} of -5.3°C and ΔT_{m2} of +1.7°C, respectively (***Figure 2.10 b***). However, **PyPyPyβDp** did not show a large CD signal change in CD titrations (***Figure 2.10 c***). In CD melting experiments (***Figure 2.10 d***), it is clear that **PyPyPyβDp** had a small effect on the melting curve, indicating some binding, also **PyPyPyβDp** showed destabilizing to DNA (ΔT_m = -1.0°C).

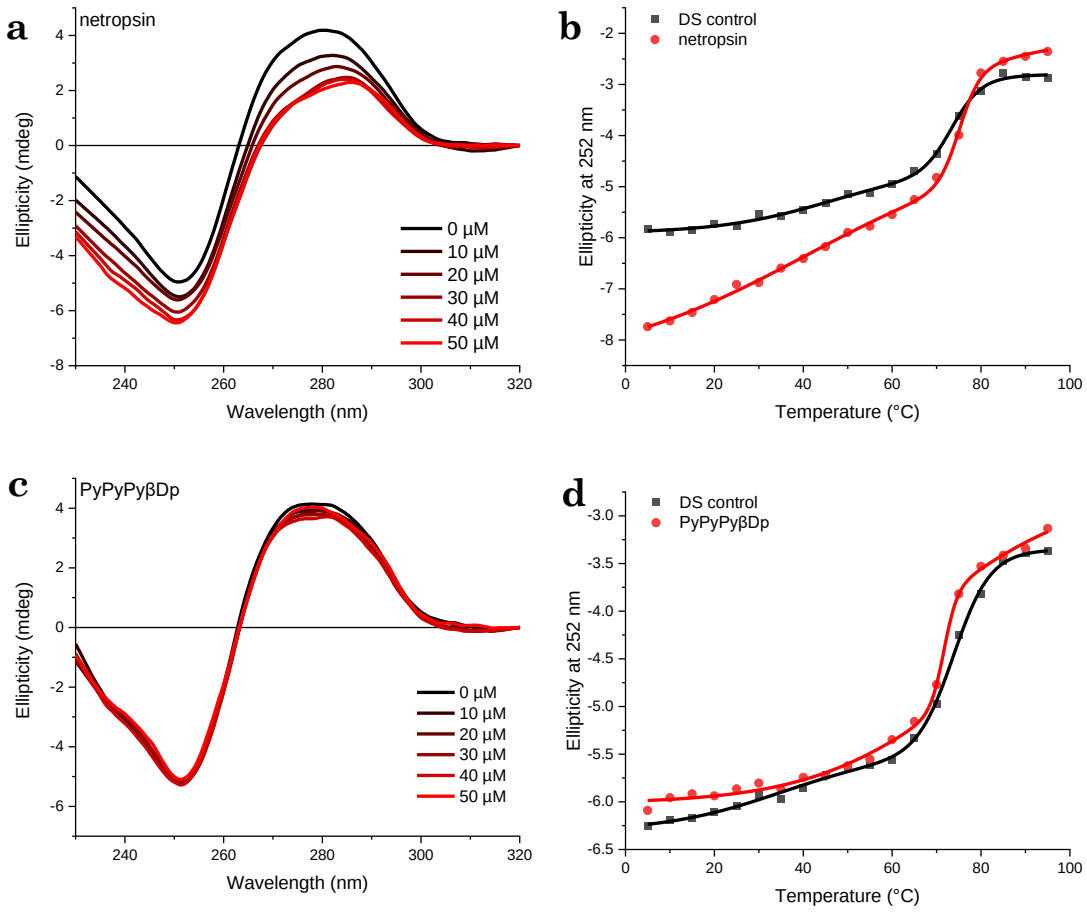


Figure 2.10 *a). CD titration of 0 – 50 μM netropsin to double-stranded DNA, CD signal of ligand was subtracted, b). CD melting of 50 μM netropsin to double-stranded DNA, c). CD titration of 0 – 50 μM PyPyPyβDp to double-stranded DNA, CD signal of ligand was subtracted, d). CD melting of 50 μM PyPyPyβDp to double-stranded DNA. All experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 7.0.*

The CD signal of *hTelo*G in the absence of ligand showed positive peaks between 250 – 260 nm and ~290 nm, indicating both parallel and antiparallel topologies formed, and the shape of the spectra is consistent with the literature.[209,210] Upon the addition of **netropsin**, there was a clear decrease between 250 – 260 nm, but no significant change at ~290 nm (***Figure 2.11 a***), suggesting **netropsin** alters the structure towards the antiparallel form. CD melting showed no effect on the thermal stability of *hTelo*G (***Figure 2.11 b***), which is consistent with UV melting. When *hTelo*G was titrated with **PyPyPyβDp**, no significant change for parallel structure, while a decrease of CD signal at 295 nm was observed (***Figure 2.11 c***),

CD melting indicated **PyPyPyβDp** slightly destabilizes *hTelo*G structure with ΔT_m of -1.0°C (***Figure 2.11 d***), however, ΔT_m calculated from the UV melting suggested no significant stabilization effect (ΔT_m = +0.9°C), however, the errors showed both CD melting and UV melting results are in the moderate range, indicating no significant effect on melting temperature of G-quadruplex structure.

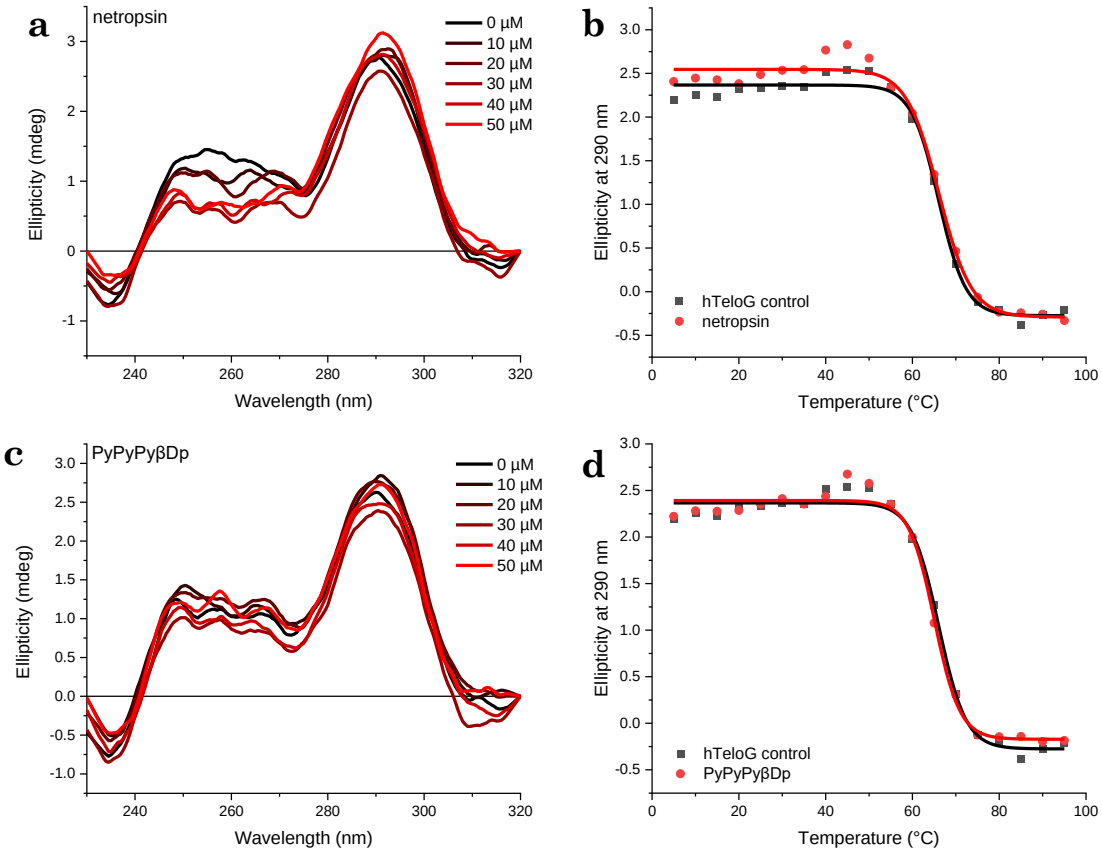


Figure 2.11 *a). CD titration of 0 – 50 μM netropsin to hTeloG, CD signal of ligand was subtracted, b). CD melting of 50 μM netropsin to hTeloG, c). CD titration of 0 – 50 μM PyPyPyβDp to hTeloG, CD signal of ligand was subtracted, d). CD melting of 50 μM PyPyPyβDp to hTeloG. All experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 7.0.*

CD signals of the i-motif forming sequences from *Hif1a*, *DAP*, *Nrf2*, and *c-MYC* showed a positive peak at 288 nm and a negative peak at 260 nm, indicating all these sequences folded into i-motif structures.[211] With the titration of **netropsin** to *Hif1a* (***Figure 2.12 a***), there were both hypochromic and hyperchromic shifts

at 288 nm and 255 nm observed, respectively. Also, an iso-elliptic point at ~273 nm was detected, suggesting the interaction between DNA and ligand. CD melting then (***Figure 2.12*** b) showed netropsin slightly stabilize *Hif1a* (ΔT_m = +1.2°C). Compared to **netropsin**, CD titration of **PyPyPyβDp** showed comparatively less change in both positive and negative ellipticity (***Figure 2.12 c***), at ~275 nm, there was an iso-elliptic point that can be observed, which is indicative of ligand-DNA interaction. From CD melting results, **PyPyPyβDp** presented a stabilization effect on DNA melting temperature (ΔT_m = +1.5°C).

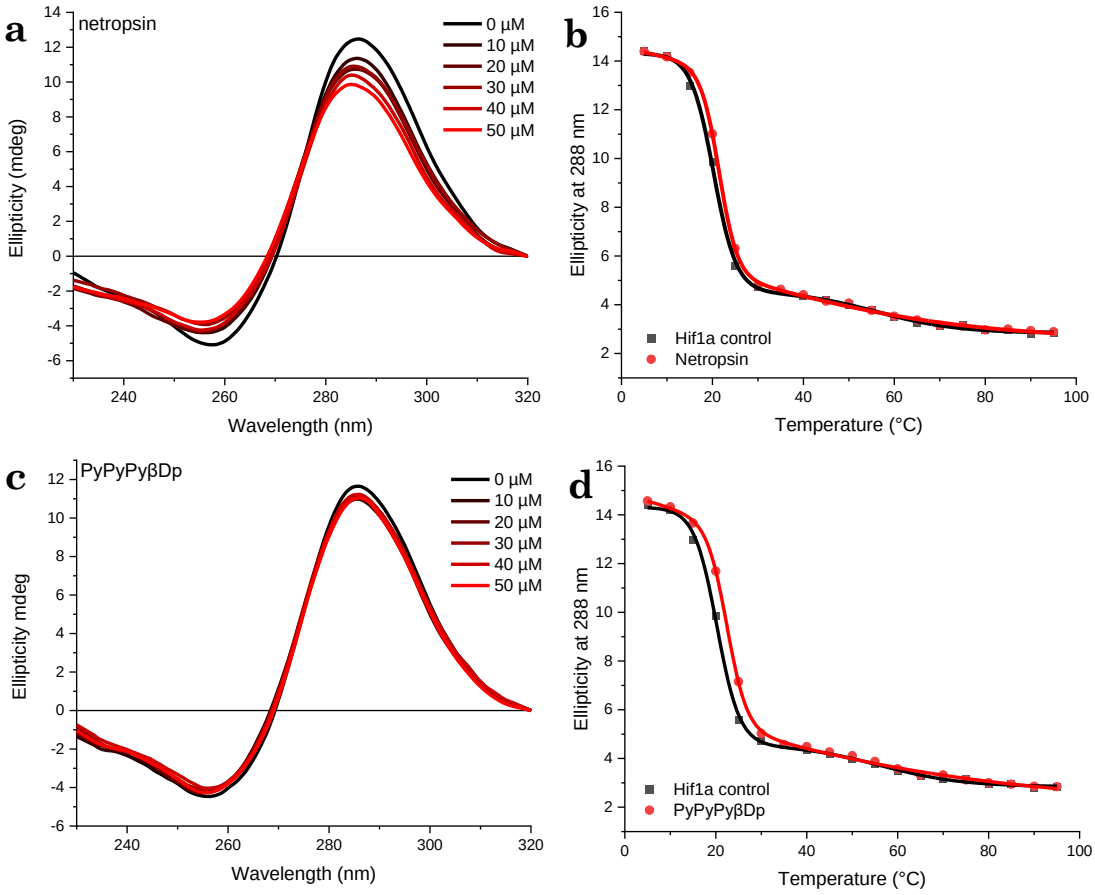


Figure 2.12 *a). CD titration of 0 – 50 µM netropsin to Hif1a, CD signal of ligand was subtracted b). CD melting of 50 µM netropsin to Hif1a, c). CD titration of 0 – 50 µM PyPyPyβDp to Hif1a, CD signal of ligand was subtracted, d). CD melting of 50 µM PyPyPyβDp to Hif1a. All experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 7.0.*

Upon the titration of netropsin to the *DAP* i-motif, a clear increasing CD signal at 288 nm was observed with the first titration, then slightly decreased, also an iso-elliptic point at ~275 nm can be detected (***Figure 2.13 a***), which suggests the interaction of DNA and ligand. CD melting experiments were then performed, which indicated the destabilization effect of netropsin on the DAP i-motif with ΔT_m -1.5°C (***Figure 2.13 b***). Then CD titration with **PyPyPyβDp** showed only a slight change in CD signal at 288 nm, and no iso-elliptic point was observed (***Figure 2.13 c***), therefore, this compound may not affect the *DAP* i-motif, this was consistent with the CD melting experiments that no change in melting temperature of DAP (***Figure 2.13 d***).

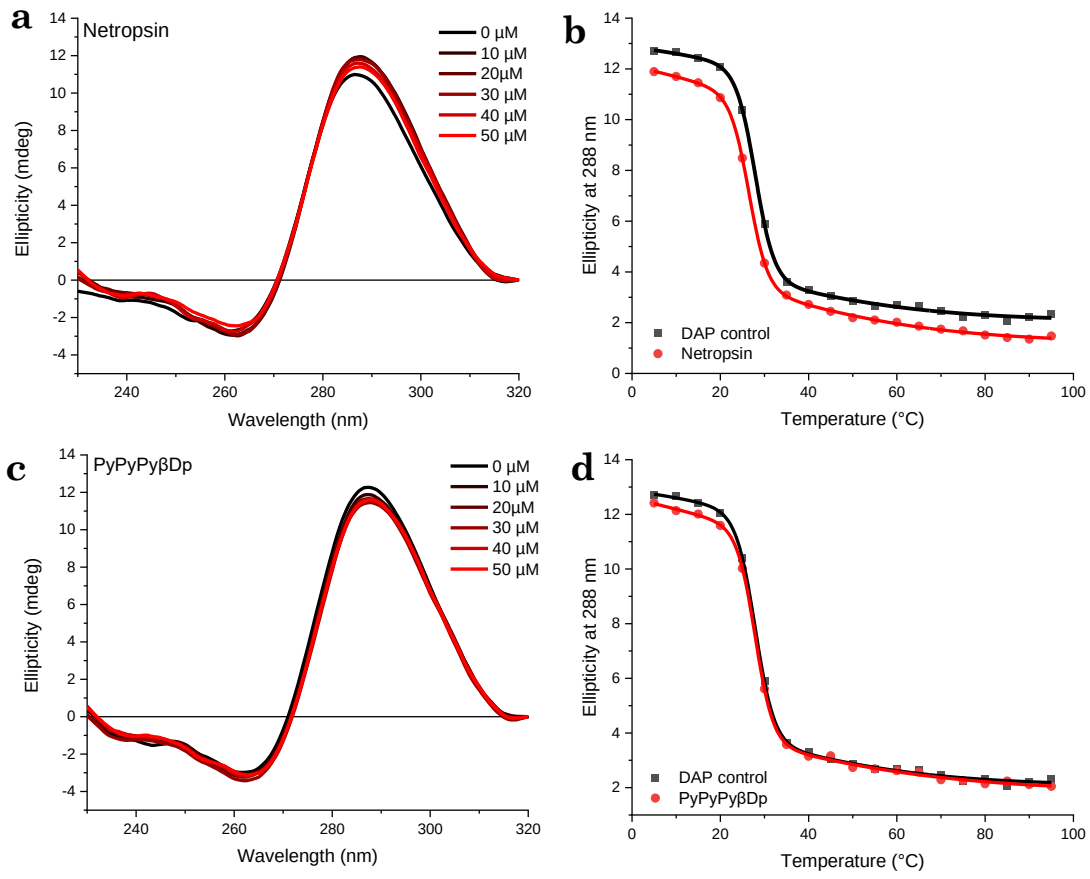


Figure 2.13 *a). CD titration of 0 – 50 μM netropsin to DAP, CD signal of ligand was subtracted b). CD melting of 50 μM netropsin to DAP, c). CD titration of 0 – 50 μM PyPyPyβDp to DAP, CD signal of ligand was subtracted, d). CD melting of 50 μM PyPyPyβDp to DAP. All experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 7.0.*

CD titrations of both netropsin and **PyPyPyβDp** to *Nrf2* i-motif did not show a significant change in CD signals (***Figure 2.14 a,c***), indicating they may not have an effect on the topology of *Nrf2* i-motif, but from the CD melting results, both minor groove binders showed a small stabilization effect on the i-motif with ΔT_m +1.3°C and +2.4°C, respectively.

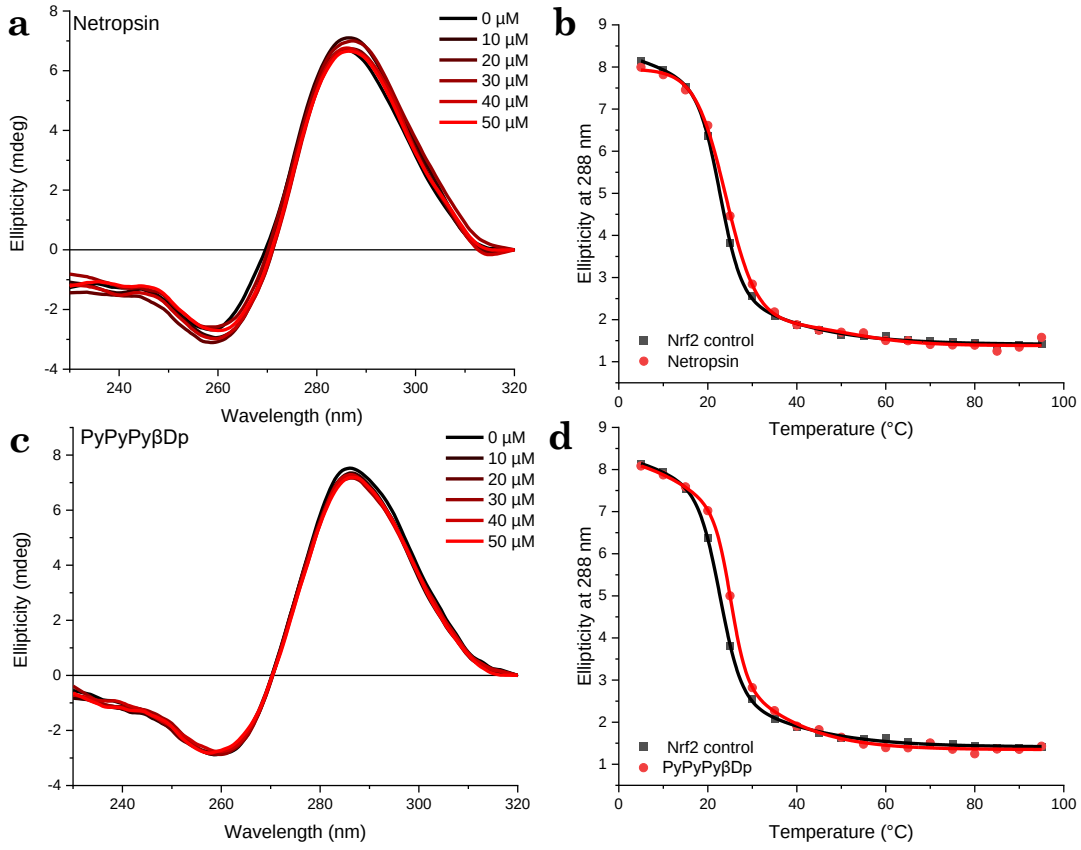


Figure 2.14 *a). CD titration of 0 – 50 μM netropsin to Nrf2, CD signal of ligand was subtracted b). CD melting of 50 μM netropsin to Nrf2, c). CD titration of 0 – 50 μM PyPyPyβDp to Nrf2, CD signal of ligand was subtracted, d). CD melting of 50 μM PyPyPyβDp to Nrf2. All experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 6.5.*

Upon the titration of **netropsin** to *c-MYC* i-motif, there was a clear decrease and increase in CD signal at 288 nm and 255 nm observed, respectively, and two iso-elliptic points at ~240 and 275 nm were detected (***Figure 2.15 a***), furthermore, at 50 μM of **netropsin**, the DNA melting curve was changed compared to the one in

the absence of ligand (***Figure 2.15 b***), this was indicative of some interaction between **netropsin** and *c-MYC* i-motif structure, with a small destabilizing effect (ΔT_m = -1.3°C). On the contrary, with the addition of **PyPyPyβDp**, there was no change observed in the CD signal, however, the CD melting showed that **PyPyPyβDp** produced a small effect on the *c-MYC* i-motif (ΔT_m = +1.8°C).

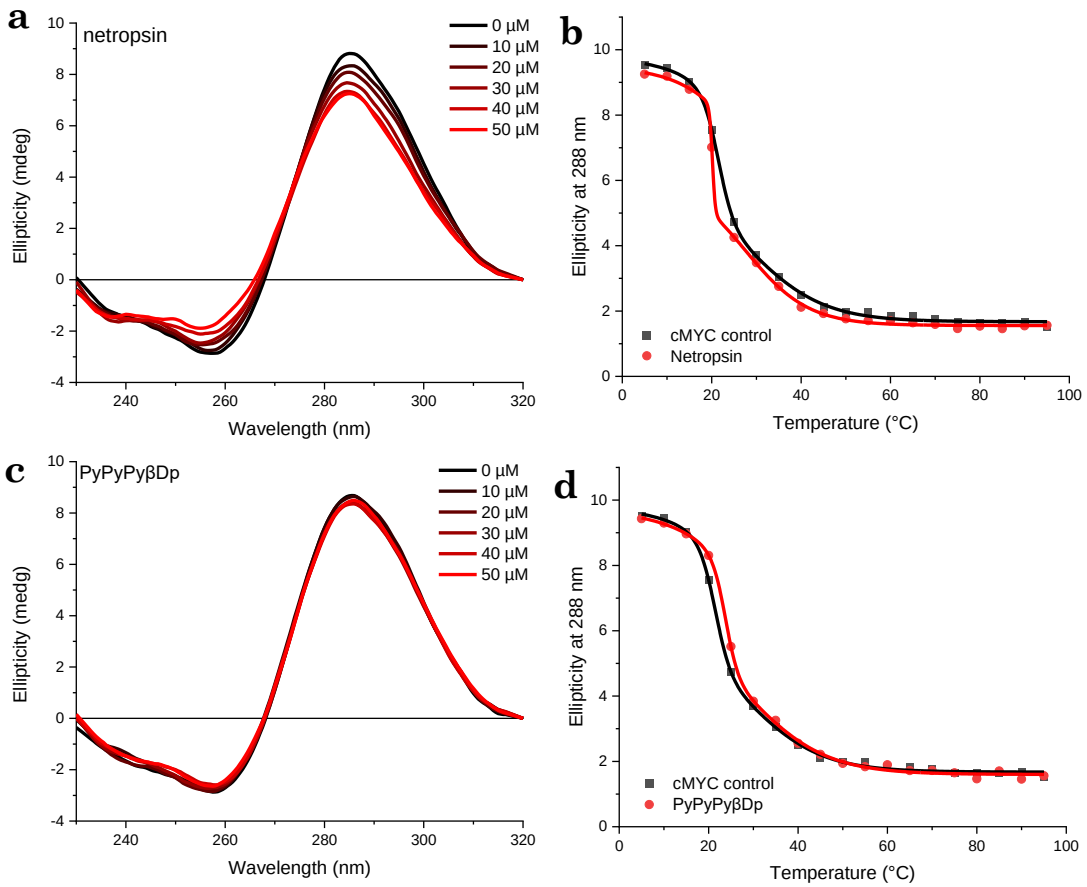


Figure 2.15 *a). CD titration of 0 – 50 µM netropsin to c-MYC, CD signal of ligand was subtracted b). CD melting of 50 µM netropsin to c-MYC, c). CD titration of 0 – 50 µM PyPyPyβDp to c-MYC, CD signal of ligand was subtracted, d). CD melting of 50 µM PyPyPyβDp to c-MYC. All experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 6.5.*

The results of the change in melting temperature of examined DNA sequences with minor groove binders are presented in ***Table 2.4***, determined by both CD and UV melting together with the ΔT_m values. The UV melting graphs of tested DNA sequences are presented in **Appendix 5** and **Appendix 6**.

Table 2.4 *ΔT_ms of examined DNA sequences measured by CD and UV in the absence and presence of 50 μM of netropsin and PyPyPyβDp. ΔT_ms were calculated from the melting temperature which was derived from the R-square values from sigmoidal curve fitting, errors show the standard derivative across two repeats.*

			Netropsin		PyPyPyβDp	
Name	T_m control (°C) CD	T_m control (°C) UV	ΔT_m (°C) CD	ΔT_m (°C) UV	ΔT_m (°C) CD	ΔT_m (°C) UV
DS	71.1±1.2	72.2±0.3	1.8 ± 0.6	2.1 ± 0.5	-1.0 ± 0.6/-14.0 ± 1.1	-0.7 ± 0.8
*hTelo*G	66.5 ± 0.5	64.1±1.2	-0.1 ± 0.3	1.6 ± 1.1	-1.0 ± 0.4	0.9 ± 0.5
Hif1a	20.3±0.2	20.5±0.6	1.2 ± 0.8	2.1 ± 0.3	1.5 ± 0.3	1.2 ± 0.4
DAP	27.9 ±0.1	27.4 ±0.1	-1.5 ± 0.7	-2.3 ± 0.2	0.0 ± 0.1	-0.6 ± 0.3
Nrf2	22.8±0.1	22.8±0.3	1.3 ± 0.4	0.4 ± 0.3	2.4 ± 0.1	1.8 ± 0.1
c-MYC	21.5±0.3	22.1±0.5	-1.3 ± 0.5/6.1±0.8	0.1 ± 0.3	1.8 ± 0.1	1.8 ± 0.3

2.5 Discussion and Conclusion

Previous work showed that **netropsin** and **PyPyPyβDp** interact with double-stranded DNA,[186,212] however, in examination with other types of DNAs, both minor groove binders showed an effect on G-quadruplex and i-motif structures either on topology or thermal stability. Ligands that target B-form duplex or triplex DNA will often intercalate or interact with a groove, while most G-quadruplex binding ligands can specifically stack onto terminal G-tetrads.[213–216] Nguyen and co-workers identified a dual-specific ligand that can interact at two binding sites of a duplex-quadruplex hybrid.[217] They achieved this with the addition of netropsin and Phen-DC_3 to *QDH1*, this complex showed netropsin binding to the AT-rich region of duplex and Phen-DC_3 binding to the terminal G-tetrad of *QDH1* simultaneously.[217] Interestingly, both **Netropsin** and **PyPyPyβDp** do not affect the stability of double-stranded or G-quadruplex DNA significantly, but they were found to interact with several i-motif structures: *Hif1α*, *Nrf2*, and *c-MYC*. Netropsin is a known minor-groove binding compound; the reason for the binding data for double helical DNA might not be as good as expected is that the duplex sequence (DS/DS_{comp}) has only has $(AT)_3$ tracts and also contains TpA steps. Although the duplex sequence used in the experiments was not ideal for these specific compounds, both **netropsin** and **PyPyPyβDp** have already been studied as minor groove binders. The sequence we used allows comparison between ligands against a general double helical sequence, to compare without favouring to one specific type of sequence. The main purpose for this project was to investigate whether minor groove binders interact with i-motifs or G-quadruplexes. Although the CD melting results do not show significant changes in melting temperature (1-2°C), the fitted data to the points were performed in duplicate and the differences were repeatable. This indicates that it may be worth examining these further, by performing-by-performing further repeats, and/or further concentrations. In addition to the study of G-quadruplex and i-motif with **netropsin** and **PyPyPyβDp**, some other classical minor groove binders have been previously examined with quadruplex structures. The Mayol group investigated interaction of netropsin and distamycin A with G-quadruplex and

duplex DNAs.[218] From the NOESY NMR results, it was suggested distamycin A interacts with medium width grooves of G-quadruplexes, and netropsin showed low binding affinity with G-quadruplex structures compared with duplex DNA. Souvik and co-workers identified that the minor groove binder Hoechst 33258 binds the G-quadruplex from the promoter region of *c-MYC*, [219] they also proposed the likelihood that binding with the loop formed by AAGGT. Lijun and co-workers screened several fluorescent dyes including Hoechst 33258 for developing an i-motif probe.[220] The ligand was added to crystal violet/i-motif system at pH 6.0, the fluorescence at 490 nm was significantly enhanced, indicating Hoechst 33258 interacts with i-motif structure, and it may bind to the narrow groove of i-motif.

The small molecule ellipticine is a known intercalator but has been shown to stabilize i-motif DNA preferentially over G-quadruplex or double helical DNA.[63] This indicates that **PyPyPyβDp** could also be used to stabilize some i-motifs generically. Though it would be better for the stabilization potential to be larger, this could be increased with further derivatives. Another interesting point to note is that **PyPyPyβDp** did not affect the stability of the i-motif from the promoter region of DAP. This sequence has only single nucleotide loops, whereas the other i-motifs have longer loops. This indicates that **PyPyPyβDp** may bind the loop regions of the i-motif structures. However, to understand this, further experiments need to be performed, e.g., computational modelling experiments. Job plot or ITC could also be used to determine the stoichiometry for the ratio of ligand-DNA binding.

Nevertheless, looking back at our ligand design **strategy i**, the first step was trying to find a good minor groove binder, which meant the candidate should bind to double-stranded DNA specifically. Here we have shown that the minor groove binding agents also interact with other types of DNA structures, and perhaps even as a proof of concept before targeting using the Dervan-style polyamides, these compounds would not be a good start for our intended purpose. **PyPyPyβDp** shows potential as a specific i-motif stabilizing agent (ΔT_m of 1.5 to 2.4, but does not stabilize DS or G-quadruplex), but only provides weak stabilisation.

This part of the project was performed in parallel with the discovery of an i-motif stabilizing compound (Chapter 3), to conjugate to the minor groove binder. The outcome of this part of the project indicates that further work is required, or the synthesis of a Dervan-style hairpin polyamide is necessary, even as a proof of concept. Regardless, the discovery that **PyPyPyβDp** binds i-motif is new and interesting and this could be useful as a probe for biological experiments.

Chapter 3

Studies on the Effects of Small Molecules

3.1 Discovery of i-Motif Ligand

In parallel to **Chapter 2** where minor groove binders were studied, at the same time, we also focused on i-motif binding ligands. To design a specific i-motif ligand, if conjugating to a double-helical sequence targeting moiety there also needs to be an i-motif component. The reality of the i-motif field is that there are not yet enough compounds to target i-motif well and before developing a conjugate for specific targeting a good general i-motif stabilizing compound is also required. Although there is a great deal of published literature describing ligands that can bind i-motif, most of them do not stabilize the i-motif structure and many are also not specific to i-motif,[124,221] therefore, there is the need to identify more and better small molecules that can interact and stabilize the i-motif. It is significant to search for a small molecule that can bind specifically to the i-motif structure, because: firstly, as discussed in **Chapter 2**, we are trying to design an i-motif-specific ligand. The have been many studies of ligands that bind to the i-motif, but i-motif-specific binding ligands are rare. Secondly, i-motif binding ligands can be used as tools to understand the biological functions of i-motif structures.

This chapter will discuss the binding of different sources of potential i-motif ligands including compounds from the National Cancer Institute (NCI) Diversity Set VI library and synthetic compounds. These compounds were tested with a variety of i-motif sequences and double-stranded and G-quadruplex DNAs.

There are three biophysical techniques used in this chapter: FRET melting assay, FID assay, and CD spectroscopy, similar to how they were used in **Chapter 2**. High throughput FID assays were used to investigate the relative binding affinity. Medium throughput FRET melting experiments were used to investigate the stabilization potential of small molecules with DNA secondary structures. Although FID and FRET melting experiments can test a large number of compounds at the same time, there are some main drawbacks to each of these methods. The FID assay can only establish whether small molecules bind to DNA, with no information about their stabilizing ability. In FRET melting experiments, two fluorophores attached to DNA were used, so the ligands could bind to the fluorophores, not to the DNA structure. Moreover, both experiments are

fluorescent experiments, if direct fluorescence quenching occurs, this can interfere with the results. Therefore, circular dichroism (CD) was one of the main methods to confirm the effects of ligands on the topology and stability of DNA sequences.

3.2.1 Targeting the Human Telomeric i-Motif with Small Molecules from the National Cancer Institute

In this section, our work focused on investigating i-motif binding ligands from the NCI Set VI library. Firstly, computational work was used to mimic small molecules binding to i-motif, and *in silico* screening was performed by Dr. Philip Spence.[222] The DNA sequence used in this section is the human telomeric i-motif (*hTelo*C, 5′-[TAA-CCC]$_4$-3′).[206, 222] The NMR-derived i-motif structure (PDB: 1EL2) was used to model the i-motif sequence.[223] The structure was edited to remove the non-canonical bases used for NMR structure determination and this refined model,[57] representative of the native sequence, was used for *in silico* screening (***Figure 3.1***). The molecular structure model of *hTelo*C was chosen to perform a computational docking experiment with the 1584 small molecules from the NCI library.[222]

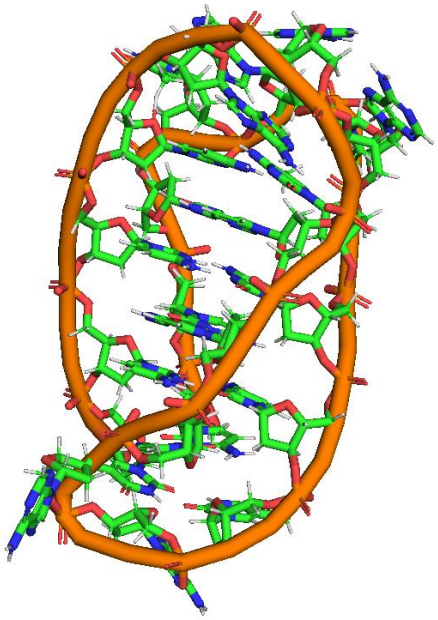
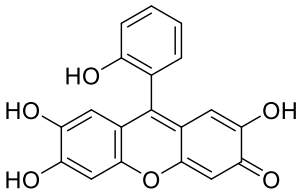
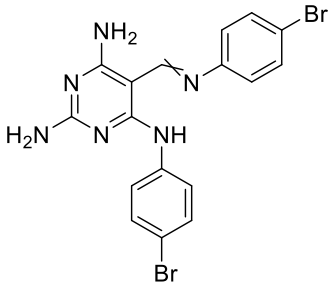


Figure 3.1 *NMR structure of modified human telomeric i-motif (PDB id: 1EL2).*

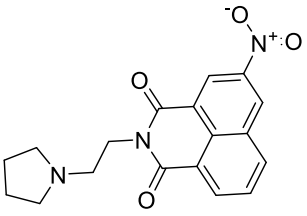
The NCI Diversity set VI was selected using defined pharmacophoric centres including hydrogen bond donor, acceptor, positive charge, aromatic, acid, base, and hydrophobic.[224] Over one million molecules were used for the Diversity selection, the selection principle is based on the pharmacophores of each molecule and also each of its isomers. Then the diversity was set up by only accepting the molecules with more than five pharmacophores, all the molecules contain five or fewer rotatable bonds, one or fewer chiral centres, or contained undesirable features, e,g. obvious leaving groups, weakly bonded heteroatoms or polycyclic aromatic hydrocarbons, were also removed, then it was given the final set of 1584 molecules.[224] The NCI molecules were screened against the i-motif structure in both the loops and wide grooves,[222] using AutoDock Vina.[222] The docking methods sifted compounds for further biophysical experiments, the top 28 molecules were examined using CD titration and melting experiments, and the structures of these compounds are presented in ***Figure 3.2***.



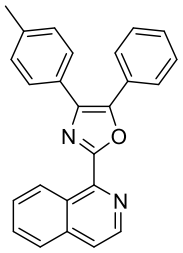
NSC 9037



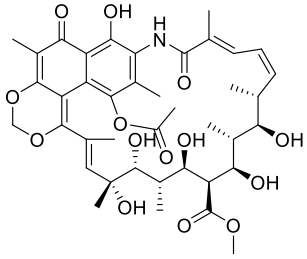
NSC 103520



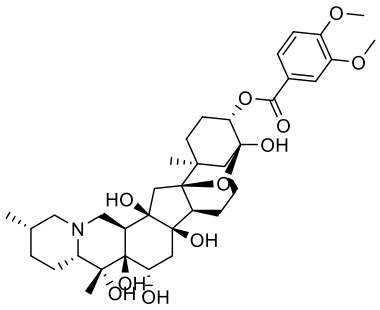
NSC 300289



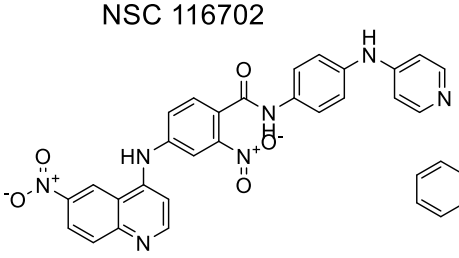
NSC 116702



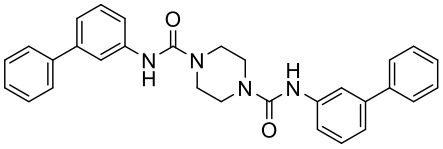
NSC 19990



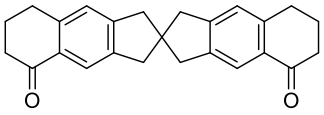
NSC 7524



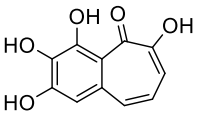
NSC 202386



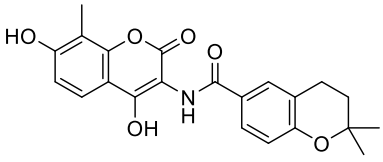
NSC 37553



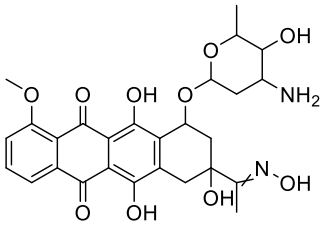
NSC 670283



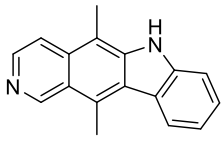
NSC 35676



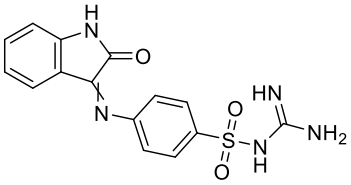
NSC 5157



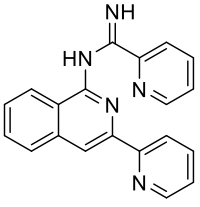
NSC 143491



NSC 71795



NSC 133075



NSC 637578

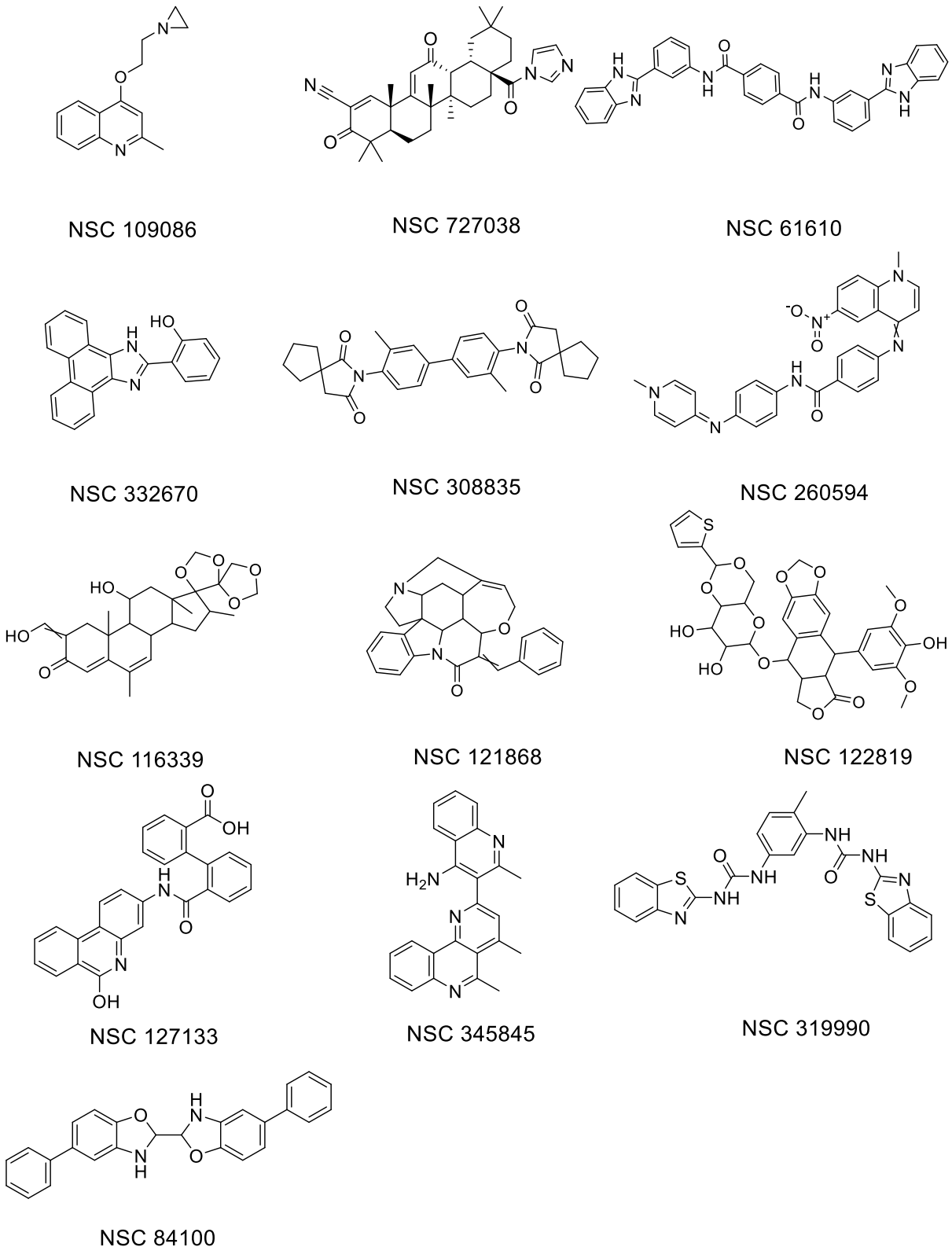


Figure 3.2 *Structures of 28 examined NCI compounds using CD titration and melting experiments.*

3.2.2 CD Titration and Melting Experiments with the i-Motif Forming Sequence from the Human Telomere and Ligands from the NCI Diversity Set VI.

According to the docking results, the potential binding sites could be three loop regions and two wide grooves. We envisaged that 50 μM would provide a decent opportunity the caputure potential moderate binding ligands, without the overall ligands:DNA ratio becoming too high. All the experiments were performed in 50 mM sodium cacodylate at pH 6.0 for *hTelo*C i-motif with the ligands. To determine the effects of the compounds on the *hTelo*C i-motif structure, CD titrations were performed to determine any ligand-induced changes in the secondary structure. In the absence of a ligand, the CD signal for each sample showed a characteristic positive peak at 288 nm and a negative peak at 255 nm, indicating an i-motif structure formed. Then with the addition of 1 equivalent of ligand to the DNA each time, up to 50 μM. CD melting experiments examined the *hTelo*C i-motif in the presence of 50 μM of ligand, and also measured the DNA in the absence of ligand as a control. Monitoring the ellipticity at 288 nm from CD against temperature change can directly report the melting curve, melting temperature can be determined using sigmoidal curve fitting. Herein, the compounds that showed a clear change in titration or melting temperature will be discussed. Out of the 28 compounds, two compounds showed potential to stabilize the i-motif from the human telomere, whereas twelve compounds were found to destabilize the structure. seven compounds did not affect the i-motif structure at all. Full results are shown in ***Table 3.1***.

Table 3.1 *Effects of tested NCI ligands on topology and thermal stability of hTeloC i-motif structure. ΔT_m of 10 μM hTeloC with 50 μM NSC ligands was measured by CD, and experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride pH 6.0 buffer. T_ms were calculated using R-square values from the statistics on the data fit.*

Ligand (NSC)	Effect on CD Spectrum	Effect on shape of melting curve	Effect on T_m	ΔT_m (°C)
19990	+	+	+	+0.2 / +33.0
71795	+	–	+	-2.5 / +24.2
61610	–	+	+	-2.8
116702	–	+	+	-2.7
727038	–	+	+	-2.4
319990	+	–	+	-2.4
127133	–	+	+	-2.3
332670	+	–	+	-2.3
308835	+	–	+	-2.2
109086	+	–	+	-2.0
84100	+	–	+	-1.6
345845	+	–	+	-1.5
143491	+	–	+	-1.4
35676	–	+	–	-1.2
5157	–	–	–	-0.9
122819	–	+	–	0.6
9037	–	–	–	-0.1
103520	+	–	–	-0.5
300289	–	–	–	0.3
7524	–	–	–	-0.2
202386	+	–	–	0.1
37553	–	–	–	-0.3
670283	+	–	–	-0.4
133075	–	–	–	-0.5
637578	–	+	–	-0.4
260594	–	–	–	-0.1
116339	–	–	–	0.1
121868	–	+	–	0.2

Six NSC ligands for further discussion as those compounds showed potential interaction with hTeloC i-motif forming sequence. The CD titration and melting graphs of rest 22 compounds are presented in **Appendix 7**. A similar decrease in the CD signal intensity was found in **NSC 670283** (***Figure 3.3 a***) and **NSC 143491** (***Figure 3.3c***) with titration up to 50 µM, and one iso-elliptic point observed at 268 nm, indicating some change to the i-motif structure, from the CD titration experiments, **NSC 670283** showed no change in melting temperature (***Figure 3.3 b***), while **NSC 143491** presented a slight decrease in melting temperature of *hTelo*C with ΔT_m –1.4°C (***Figure 3.3 d***), indicating a potential weak destabilizing effect, but this would need to be repeated many times to validate as the magnitude is small.

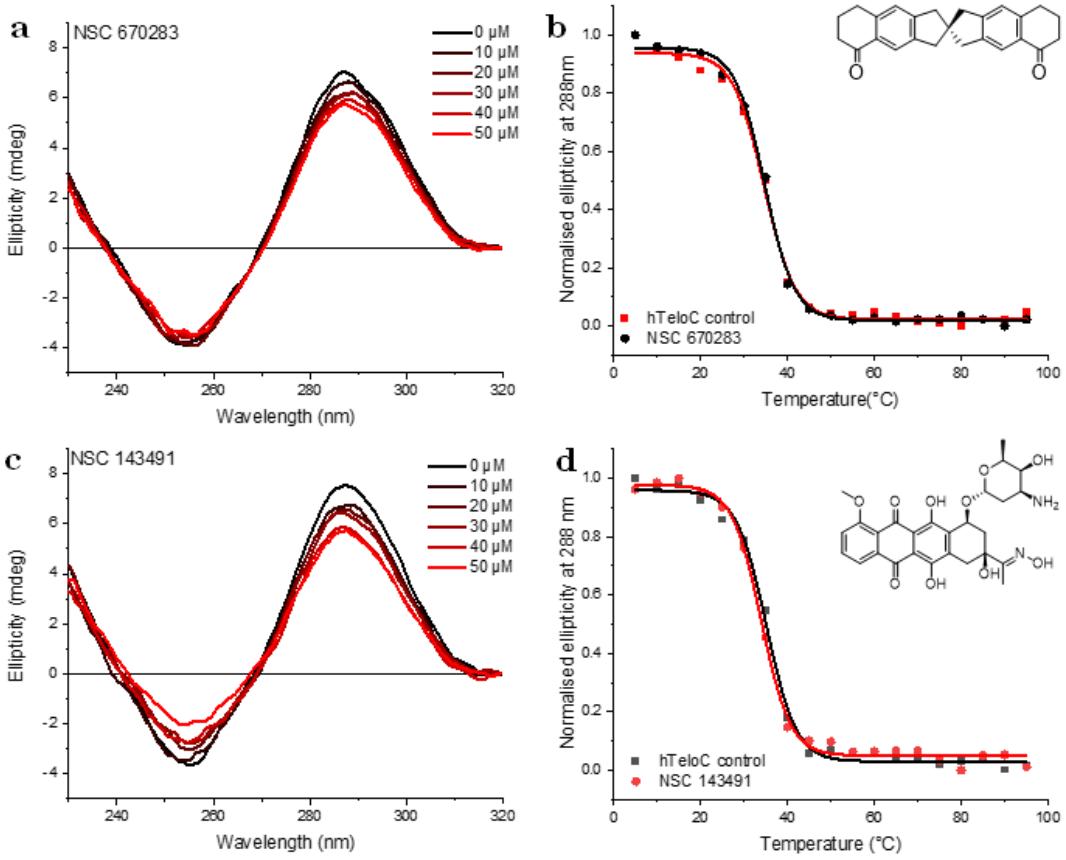


Figure 3.3 *a). CD spectra of titration experiments of 0 – 50 µM of NSC 670283, CD signal of ligand was subtracted, b). CD melting experiment of 50 µM ligand, c).CD spectra of titration experiments of 0 – 50 µM of NSC 143491, CD signal of ligand was subtracted, d). CD melting experiment of 50 µM ligand, with 10 µM hTeloC in buffer containing 50 mM sodium cacodylate at pH 6.0.*

In contrast, **NSC 19990** showed a larger change in the CD spectra intensity of *hTelo*C (***Figure 3.4 a***), the CD spectra looked relatively unchanged in the presence of 10 µM of ligand, then the large decrease of CD signal was obtained from the addition of 20 µM of ligand up to 50 µM. An additional signal starts to appear at 245 nm ***(Figure 3.4a)*** but this may be due to an artefact from subtraction of the effects of the chiral ligand. The CD melting experiment also showed a large effect on the CD melting curve. The CD melting experiment also showed that NSC 19990 affected the CD melting curve, for which an additional transition which melted with a ΔTm of +33°C is seen in the presence of 50 µM ligand (***Figure 3.4 b***). It is not clear why this only affects a fraction of the melting. It might indicate a strong stabilizing effect on a sub-population of DNA structures, though since this is such a large effect it is unusual that it does not affect the entire melting curve. We must therefore consider that this is an experimental artefact.

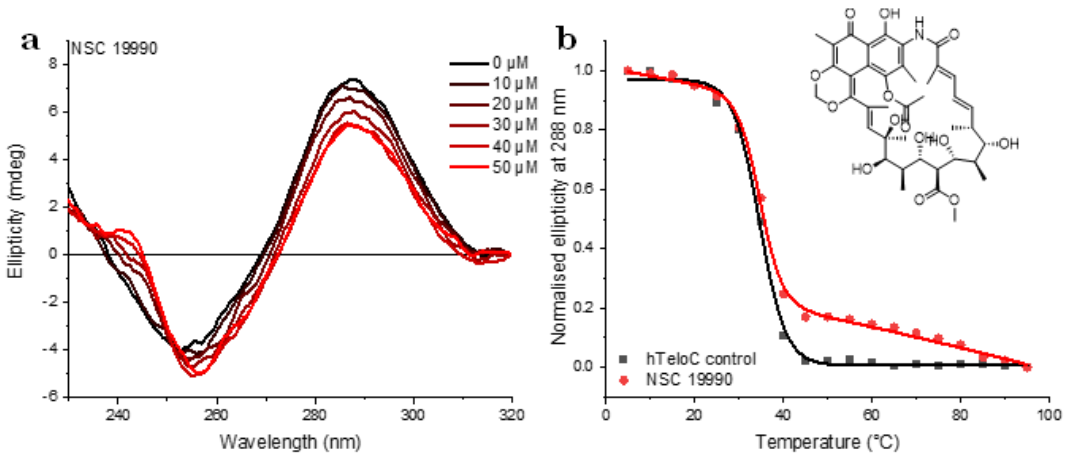


Figure 3.4 *a). CD spectra of titration experiments of 0 – 50 µM of NSC 19990, CD signal of ligand was subtracted, b). CD melting experiment of 50 µM ligand, with 10 µM hTeloC in buffer containing 50 mM sodium cacodylate at pH 6.0.*

The *hTelo*C CD spectrum was also affected by the addition of 10 µM of **NSC 71795** (***Figure 3.5 a***) and the CD signal decreased above 20 µM of ligand, indicating a possible change in i-motif structure. The CD melting with 50 µM of NSC 71795 (**Figure 3.5 b**) showed as small decrease in melting temperature (ΔTm = -2.5°C), for the majority of the melting transition. However, as seem the NSC 19990, a

small proportion melted at a higher temperature (ΔTm = +24.2°C). This may be consistent with work by Smith and coworkers,[225] where they observed a second transition on addition of ellipticine with the i-motif forming sequence from c-MYC. The presence of the second transition may suggest that NSC 71795 is able to stabilise a minor sub-population of i-motif in solution. However, the lack of effect on most of the melting transition means that this may be an experimental artefact.

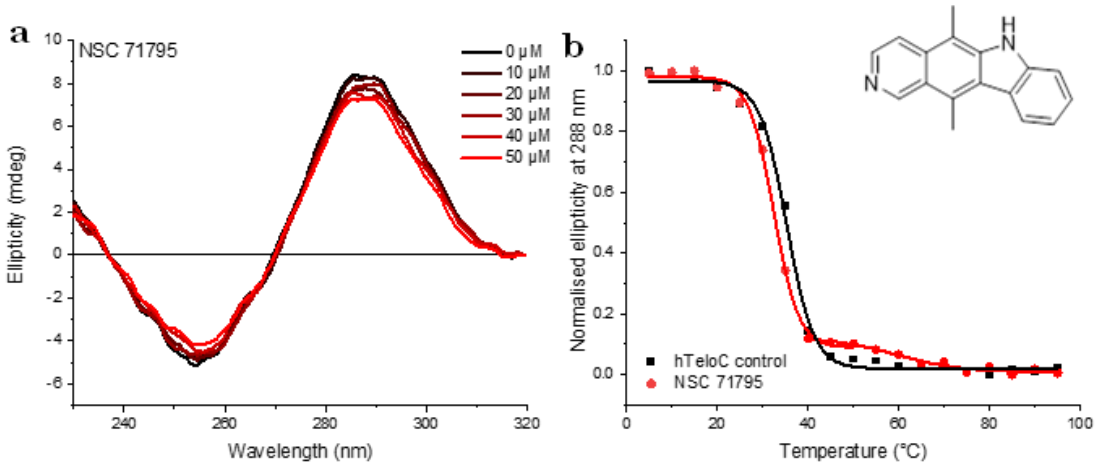


Figure 3.5 *a). CD spectra of titration experiments of 0 – 50 µM of NSC 71795, CD signal of ligand was subtracted, b). CD melting experiment of 50 µM ligand, with 10 µM hTeloC in buffer containing 50 mM sodium cacodylate at pH 6.0*

Upon titration of up to 50 µM **NSC 116702** into *hTelo*C (***Figure 3.6 a***), there was no significant change in the CD signal intensity, however, interestingly, the CD melting showed decreasing in melting temperature (ΔT_m = -2.7°C), showing a destabilizing effect on i-motif structure.

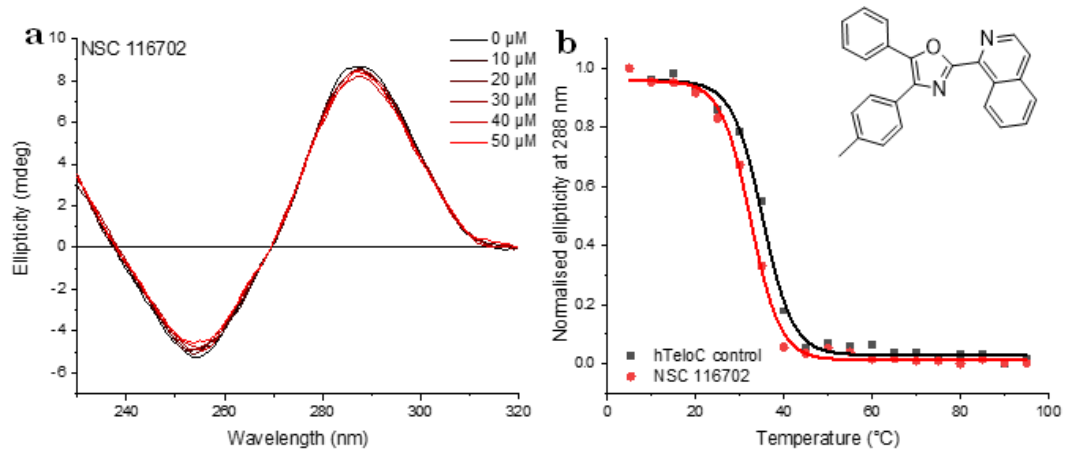


Figure 3.6 *a). CD spectra of titration experiments of 0 – 50 µM of NSC 116702, CD signal of ligand was subtracted, b). CD melting experiment of 50 µM ligand, with 10 µM hTeloC in buffer containing 50 mM sodium cacodylate at pH 6.0*

The ligand **NSC 61610** did not show a big change in CD signal with titration experiments up to 50 µM (***Figure 3.7 a***). While the CD melting experiment (***Figure 3.7 b***), showed a change in melting temperature (ΔT_m = -2.8°C), which suggests that this ligand destabilizes the *hTelo*C i-motif structure.

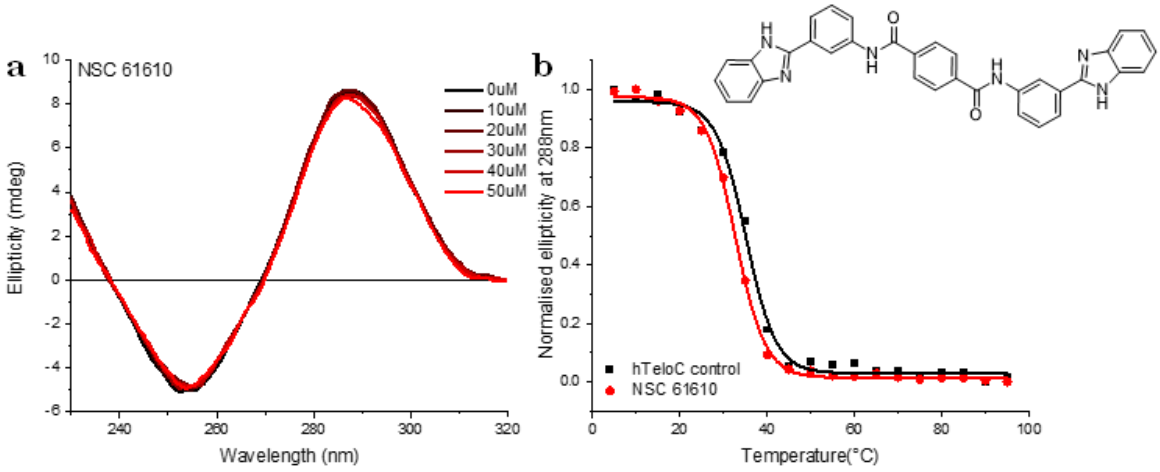


Figure 3.7 *a). CD spectra of titration experiments of 0 – 50 µM of NSC 61610, CD signal of ligand was subtracted. b). CD melting experiment of 50 µM ligand, with 10 µM hTeloC in buffer containing 50 mM sodium cacodylate at pH 6.0*

In addition to the example ligands discussed above, some other compounds also showed effects on the structure or stability of the *hTelo*C i-motif. CD titration and melting graphs for the rest of the NCI ligands are presented in ***Appendix 8***.

3.2.3 Discussion

Overall, the correlation between the *in silico* screening and CD spectra was shown to be moderate, the 28 NCI ligands from the top one hundred *in silico* assays were examined, and 20 of them showed a small interaction with *hTelo* i-motif structure, suggesting around 71% of success rate, with results matched the modelling experiments. From both CD titration and melting results, those ligands (**NSC 103520, 202386** and **670283**) showed a slight change in the intensity of the bands in the CD titration and no significant effect on the melting temperature of the hTeloC structure, indicating that there might be a weak interaction with i-motif structure. For the ligands (**NSC 116702, 35676, 727038, 61610,** and **127133**) that showed no change in titration experiments but have effects on the stability of the i-motif structure, all these seven ligands destabilize the i-motif. The ligands (**NSC 143491, 19990**, **71795**, **109086, 332670, 308835, 345845, 319990,** and **84100**) that showed both interactions on i-motif topology and a change in melting temperature appear to stabilize or destabilize the i-motif structure. The way the experiments were performed means that small changes in stability are difficult to determine with confidence. The melting experiments were performed using a whole spectrum scanning method. This means that the whole spectrum can be monitored across the temperature range, providing lots of information of potential changes in topology or induced CD effects. For these experiments, each spectrum is scanned from 320 nm to 230 nm and the average of four scans for each spectrum are taken. This method was used because this provides the opportunity to observe changes in topology over the temperature range, but also because CD melting which are followed at a single wavelength have high signal to noise, compared to UV-based melting experiments. Full scans were taken every five degrees, providing a data set of 19 points across the 5-95°C range. The data interval is too large to determine the melting temperatures by the first derivative method and these data were instead fitted to either a sigmoidal or bi-sigmoidal curves to determine the melting temperatures. What we observe from this is exceptional detail in what is happening to the topology over the temperature range, but does mean that the melting data relies on the quality of data fitting. It also means that

small transitions, of smaller populations within a larger bulk solution can be missed. For example in ***Figure 3.5*** we observed a second transition on addition of **NSC 71795**, which could be put down to an artefact, but this secondary transition has also been observed by others. For example, **NSC 71795** has already been an identified i-motif stabilizer that showed stabilizing effect for i-motif (ΔT_m = +23°C) at pH 7.4 by Smith and co-workers.[63] Importantly, this compound was shown in separate experiments to stabilise i-motifs and induce i-motif formation at neutral pH.[63] Although in the paper, for *hTelo*C at pH 6.0, there was only one transition measured on the melting curve for hTeloC at 50 µM ligand, however, when looking closer at the melting curve in the presence of **NSC 71795**, it is possible to see that there is a slight second transition, indicating a potential stabilizing effect[63](***Figure 3.8***), however, due to the data interval in these experiments this had been missed. In this study, as we used different buffer conditions, the second transition is more clear, and the stabilizing effect was strong (ΔT_m = +24.2°C).

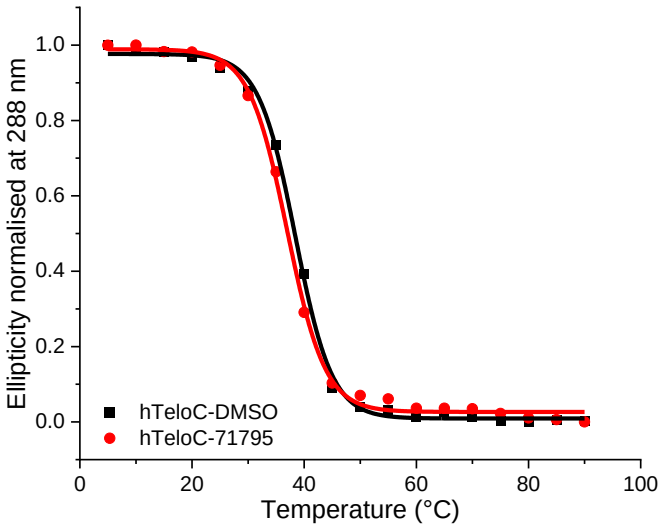


Figure 3.8 *CD melting of hTeloC in the presence of 50µM NSC 71795, in 10 mM sodium cacodylate buffer at pH 6.0. These experiments were performed by Dr. Rupesh Chikhale.*

NSC 19990 also showed two transitions in the melting curve (***Figure 3.4***), and it is the largest change in the melting curve compared to other ligands (ΔT_m = +33°C). The second melting temperature showed the biggest stabilization effect on the stability of the i-motif. As this compound is chiral this does mean that the ligand

itself has a CD signal. Although these signals were subtracted off appropriately, it might be that some effects observed in the spectrum could be put down to artefacts due to differences in subtraction. Regardless of this **19990** shows the same additional population as **71795**, which has no CD signal of its own. Therefore, both **NSC 29990** and **71795** could be used as potential candidates for i-motif stabilizers for the ligand design project, but further binding studies and repeats would be needed to ensure the compounds are not false positives caused by the data processing pipeline.

In addition to this, there was a concern that some of these ligands may only work for the human telomeric i-motif, rather than i-motifs in general, which is what we wanted for the design of a ligand-conjugate. Also, some of the ligands could have potential effects on other i-motifs. For example, **NSC 9037** did not show an effect on *hTelo*C, but it does have a stabilization effect on *ILPR* i-motif (5′-TGTCCCCACACCCCTGTCCCCACACCCCTGT-3′) with ΔT_m = +2.0°C (***Figure 3.9***), which has been reported by Dr. Dilek Guneri, and it has biological effects on insulin-secreting cell lines.

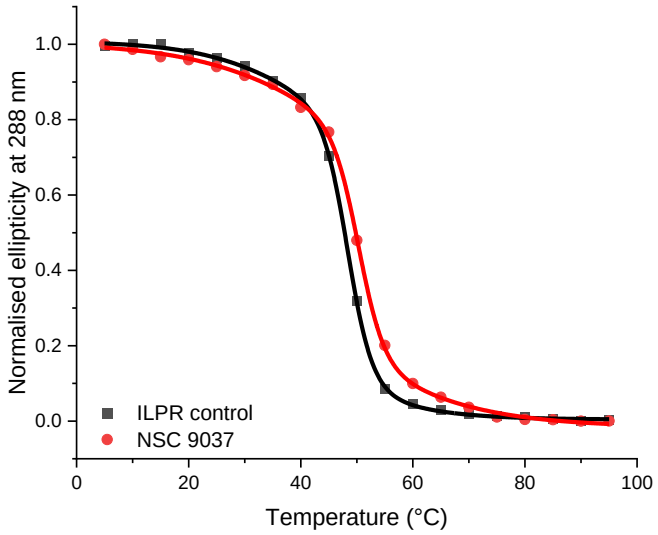


Figure 3.9 *CD melting experiment of 10 μM ILPR with 50 μM NSC 9037, this experiment was performed by Dr. Dilek Guneri.*

The outcomes from this study indicated a few compounds that could work by stabilizing the i-motif structure. Ligand **71795** showed interesting properties that were dependent on the buffer conditions, indicating that the environment where these compounds are being examined is critical to whether a hit is identified. Indeed in many cases where there were compounds that showed very small effects in the CD experiments and stabilized a secondary population of i-motif (in the case of **71795** and **19990**). What species is this? Is it the difference between the 5′E and 3′E? Further work using NMR spectroscopy or X-ray crystallography would be required to determine this. Regardless, this work showed a number of compounds that could potentially be used to stabilize i-motif for further studies in this area.

3.3.1 The Study of Metal Complexes of Bis(pyridyl)allenes

Parts of this section have been published and are presented in the appendix. Natural products or synthetic compounds with allene functional groups have been suggested to have biological activity, therefore, some attempts have been made in recent years to combine the allenic moiety with pharmacologically active molecules to increase their interaction with biological targets.[226] Additionally, in comparison with traditional organic compounds, metal complexes have some advantages.[227] From recent studies, it has been suggested that metal complexes (27%) are more likely to be active antifungal and antibacterial agents than just organic compounds (2%).[228] Herein, a series of Pd(II), Pt(IV), and Au(III) bis(pyridyl)allene-containing compounds were synthesized by Dr, Hanna Maliszewska from the Munoz group at UEA[229] (***Figure 3.10***), and these compounds were explored in their potential metallodrugs in three characterizations, as anticancer drugs, as antibacterial agents and their interaction with different DNAs.

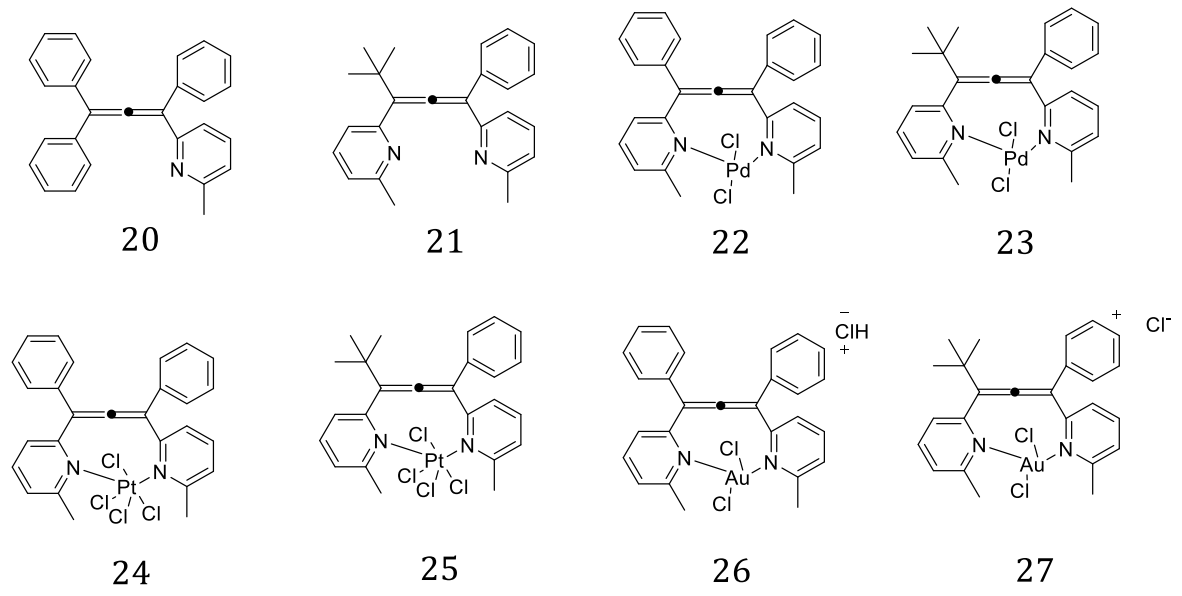


Figure 3.10 *Eight allene compounds, two free ligands (20 and 21), and six complexes used to be studied in interaction with different DNAs.*

As for many metal-based drugs, the mechanism of drug action is potentially around their interaction with DNA as an intracellular target, similar to oxaliplatin, cisplatin, and carboplatin.[230–234] The mechanism of Pt complexes bind to DNA would be they could covalently react with N7 of guanine,[235] which would be different to the examined NCI ligands showed above. This project intended to investigate the possible source of activity of allenic metal compounds in anticancer assays by studying interactions with different DNA structures, as well as the common double-stranded structure, other DNA secondary structures including i-motif and G-quadruplex was also chosen.[236]

3.3.2 Biophysical Results

FID assays

The FID assay can quickly determine whether ligands bind to DNA, so herein, the novel synthetic compounds **22**, **23**, **24**, **25**, **26**, and **27**, and the free ligands **20** and **21** were treated with different types of DNA: double-stranded DNA (DS), the i-motif (*hTelo*C) and G-quadruplex (*hTelo*G) sequences from the human telomere and the i-motif sequences from promoter regions *Hif1a* and *DAP*. Thiazole orange (TO) was used as a fluorescent probe in assays. The ratio of ligand to DNA was 5:1

(0.5 μM DNA, 2.5 μM ligand), and mitoxantrone was used as a positive control as it has previously been shown to displace thiazole orange effectively from i-motif DNA.[113,137] In the assay, the free ligands **20** and **21** did not show any significant change in TO displacement with all DNA sequences, which was consistent with the lack of activity in anticancer and antimicrobial assays carried by the Munoz group.[237] However, these two complexes showed slight negative displacement to *hTelo*C (**20** = -0.9%, **21** = -8.9%) and DAP (**20** =-15%, **21** =-7%) which is indicative the fluorescence was enhanced. The reason could be these complexes switch the equilibrium from any remaining single stranded DNA to folded i-motif species, therefore, more folded i-motif species in the solution and enhanced fluorescence. The other possible reasons could be these complexes are fluorescent or they could become fluorescent when they bind to DNA. The Pd (II)-complexes **22** and **23** showed displacement of TO from *hTelo*G, with moderate affinity for G-quadruplex structure. Moreover, these two compounds were also the most active binders for all three selected i-motif sequences, with 87% of displacement for **23** with *hTelo*C. The percentage of TO displacement was displayed in ***Table 3.2***. The graph of TO displacement of allene ligands with different DNA sequences is presented in **Appendix 9**.

Table 3.2 *FID results for allene ligands and allene-containing complexes. Experiments were performed in 10 mM sodium cacodylate and 100 mM KCl at pH 7.0 for dsDNA and hTeloG, pH 6.8 for Hif1a and DAP, and pH 5.5 for hTeloC, [DNA] was 0.5 μM, [TO] was 1 μM, and [ligands] was 2.5 μM, errors show the standard deviation across three repeats.*

	TO displacement (%)				
Compound	dsDNA	*hTelo*C	*Hif1a*	*DAP*	*hTeloG*
20	9.8 ± 2.8	-0.9 ± 8.8	4.9 ± 0.8	-8.9 ± 3.7	-1.7 ± 3.3
21	8.1 ± 0.9	-14.8 ± 10.0	2.8 ± 3.5	-7.4 ± 2.2	-1.1 ± 4.8
22	13.0 ± 3.7	80.8 ± 6.8	45.6 ± 6.5	62.5 ± 2.5	34.5 ± 3.9
23	24.7 ± 0.5	87.6 ± 1.1	49.2 ± 2.1	62.1 ± 3.3	47.9 ± 4.8
24	8.0 ± 2.6	0.8 ± 12.2	1.6 ± 3.0	-8.9 ± 8.0	-1.6 ± 7.4
25	11.5 ± 3.8	9.2 ± 8.1	5.9 ± 3.7	-3.6 ± 1.7	0.7 ± 1.4
26	10.6 ± 1.1	17.6 ± 6.0	9.8 ± 8.2	30.0 ± 3.6	-10.0 ± 4.1
27	9.3 ± 2.2	16.5 ± 11.7	6.2 ± 5.1	20.6 ± 3.0	-2.1 ± 2.9

Pt(IV) complexes **24** and **25** presented inactivity on the same level as the free allene compounds **20** and **21**, and Au(III) complexes **26** and **27**, also showed displacement with DAP i-motif with 30% and 20%, respectively. However, Au(III) ligands **26** and **27** showed much less interaction with DNA compared to ligands **22** and **23**. This would be consistent with other Au(III) complexes that are more likely to interact with protein targets.[238–242]

FRET melting experiments

FRET melting experiments were used to measure complex-induced stabilization or destabilization of DNA.[113] This was performed with the ratio of DNA and ligands of 1:5 (0.2 μM of DNA, 1 μM of ligands), and the change in melting temperature (ΔT_m) was calculated for each DNA-ligand pair. Through the FRET melting assays, as the start temperature was 25°C, for some unstable i-motifs, for example, the *DAP* i-motif, the melting process has already been proceeding before running the experiments, so the melting temperature of *DAP* was not determined, and the ΔT_ms of the rest DNA are shown in ***Table 3.3***.

Table 3.3 *ΔT_m of bis(pyridyl)allenes complexes with different DNA sequences measured by FRET melting experiments. 0.2 μM DNA and 1 μM ligands in 10 mM sodium cacodylate, 100 mM potassium chloride at pH 5.5 (hTeloC), 6.8 (Hif1a) and pH 7.0 (hTeloG, DS) Errors show the standard deviation from three repeats.*

	ΔT_m(°C) ± error(°C)			
Compounds	*hTelo*C	*Hif1a*	*hTelo*G	DS
20	0.1 ± 0.7	-0.1 ± 0.5	-0.7 ±1.3	0.0 ± 0.4
21	-0.3 ± 0.8	0.0 ± 0.3	-1.2 ±1.0	-0.4 ± 0.0
22	0.0 ± 0.5	0.2 ± 0.2	1.9 ± 0.2	-0.1 ± 0.3
23	-0.2 ± 0.6	0.1 ± 0.2	-1.0 ± 1.3	0.1 ± 0.5
24	-0.2 ± 0.6	-0.4 ± 0.5	-0.7 ± 0.2	-0.4 ± 0.0
25	-0.1 ± 0.7	0.0 ± 0.3	-0.8 ± 0.2	-0.2 ± 0.2
26	0.1 ± 0.6	0.0 ± 0.3	-0.6 ±1.0	-0.2 ± 0.2
27	-0.2 ± 0.6	0.1 ± 0.4	-0.9 ± 0.9	-0.2 ± 0.2

Based on the change in melting temperature values, it looked like none of these compounds stabilize or destabilize i-motif and double-stranded DNA, compounds **21** and **23** seemed to destabilize *hTelo*G (ΔT_m = -1.2°C and ΔT_m = -1.0°C), while **22** stabilizes *hTelo*G with ΔT_m = +1.9°C. In these experiments, *hTelo*C and *Hif1a* were performed at different pHs, because compared to *Hif1a*, *hTelo*C is more stable under acidic condition, but this did not give a fair comparison of effects of the ligands. At the higher pH, it would be easier disrupt the structure, however, this is the drawback of working with i-motifs where each structure has different pH properties. Also, in FRET melting experiments, the ligand concentrations are much too low (1 μM). These are probably a long way below the K_d, this would be another drawback in using ratios rather than ligand concentrations for

experimental planning. Most FRET melting temperatures for DNA in the literature are typically quoted at 1 µM,[243] however, this is for strong ligands. The state of the art for targeting i-motif DNA structures is significantly behind that of the G-quadruplex field, so a higher concentration should be used as a screening standard, to enable measurement of weaker ligands prior to development. FRET is a preliminary technique to study the interaction of DNA and ligands as it only presents the fluorescence, thus more detail needed to be investigated by other methods.

CD titration and melting experiments.

Combining the FRET melting and FID results from ***Table 3.2*** and ***Table 3.3***, it is suggested that Pd complexes **22**, and **23** had some interactions with DNA. From FID assay, both ligands show high TO displacement from 45% - 87% of i-motif forming sequences, they also show 13%-48% TO displacement of double stranded and G-quadruplex. FRET melting results indicate ligand **22** slightly stabilizes *hTelo*G (ΔT_m = 1.9°C) and ligand **23** shows destabilization effect of *hTelo*G (ΔT_m = -1.0°C). Therefore, both complexes were further studied using CD spectroscopy.[7] There have been a few previous reports of Pd compounds that bind to G-quadruplexes.[244,245,246] To the best of our knowledge, there is no evidence that Pd complexes have been studied with i-motif DNA.

CD titrations were performed with the complexes **22** and **23** against the i-motif forming sequences from the human telomere (*hTelo*C), and the promoter regions from *DAP* and *Hif1a* as well as the G-quadruplex forming sequence from the human telomere (*hTelo*G). The ratio of ligand to DNA was kept constant at 5:1. After the titrations, CD melting experiments were performed. All the changes in melting temperature in the presence of complexes **22** and **23** are presented in ***Table 3.4***.

Table 3.4 *Change in melting temperature (ΔT_m) of hTeloC, DAP, Hif1a, and hTeloG measured by CD melting experiments for Pd complexes **22** and **23**. Errors indicated standard deviations from duplicate experiments.*

	ΔT_m / °C			
Compound	*hTelo*C	*DAP*	*Hif1a*	*hTelo*G
22	-8.1 ± 0.8°C	-1.0 ± 0.1°C	-0.1 ± 0.3°C	-0.3 ± 0.8°C
23	-2.1 ± 0.1°C	-0.7 ± 0.6°C	-0.3 ± 0.9°C	-0.7 ± 0.6°C

The positive peak for characterizing i-motif was observed at 288 nm, while the negative peak was at 255 nm.[211] With the addition of **22** into the *hTelo*C solution, a visible hypochromic shift on both peaks happened, which could be attributed to precipitation, however, no signals for scattering at 320 nm were observed, indicating there was no precipitation. Following titration with the compound, the DNA-ligand complex was melted. The ellipticity at 288 nm was plotted against temperature and the melting temperatures were calculated by fitting the data to a sigmoidal single dose-response curve using Origin (***Figure 3.11***). For compound **22** with *hTelo*C, the change in melting temperature (ΔT_m) was found to be -8.1°C, indicating destabilization of the i-motif from the human telomere by Pd complex **22**.

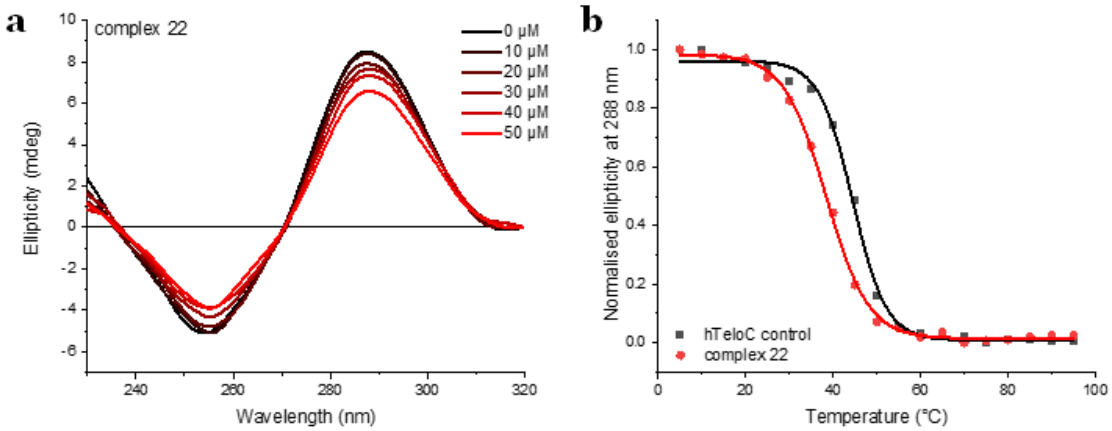


Figure 3.11 *Interaction of compound* ***22*** *with i-motif DNA. (a) CD titration of hTeloC (10 μM) in 10 mM sodium cacodylate buffer, 100 mM potassium chloride, pH 5.5. 0–50 μM of compound* ***22****, CD signal of ligand was subtracted. (b) CD melting experiment with hTeloC in the absence and the presence of 5 equiv. of compound* ***22****.*

With the addition of complex **23**, only a small decrease at 288 nm was observed compared to **22**, and an iso-elliptic point at ~273 nm was found, this suggested that complex **23** showed less interaction with *hTelo*C i-motif structure compared to complex **22**. This was also consistent with the CD melting, complex **23** showed destabilization of *hTelo*C with ΔT_m = -2.1°C, less than complex **22** (ΔT_m = -8.1°C) (***Figure 3.12***).

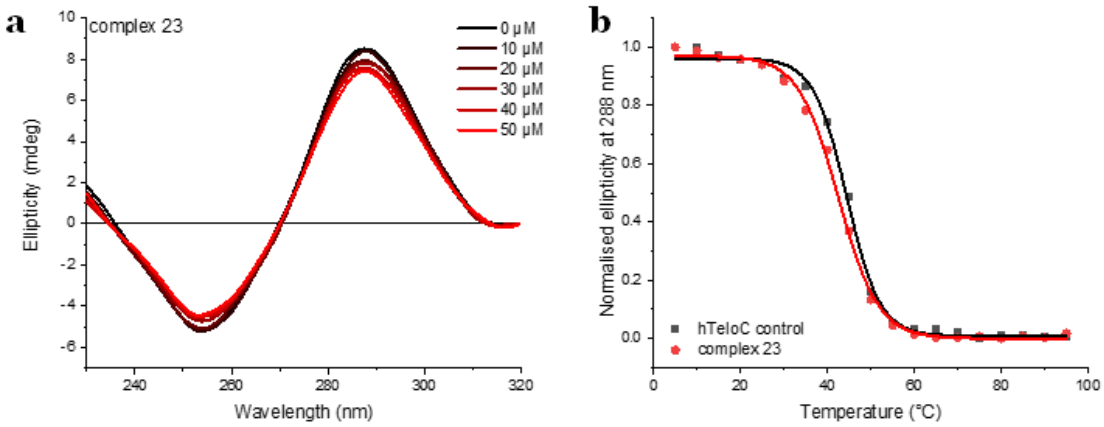


Figure 3.12 *Interaction of compound* ***23*** *with i-motif DNA. (a) CD titration of hTeloC (10 μM) in 10 mM sodium cacodylate buffer, 100 mM potassium chloride, pH 5.5. 0–50 μM of compound* ***23****, CD signal of ligand was subtracted. (b) CD melting experiment with hTeloC in the absence and the presence of 5 equiv. of compound* ***23****.*

Upon the titration of **22** and **23** to the i-motif from the promoter region of DAP, there was also a large decrease in the signals at both 288 and 255 nm observed, moreover, iso-elliptic points at ~270 nm were detected for both complexes, respectively (***Figure 3.13***), indicating interactions with the DNA. CD melting experiments were then performed, and the melting curve was changed in the presence of 50 μM complex **22**, however, compared with *hTelo*C i-motif, the change in melting temperature was quite small (ΔT_m = -1.0°C). Complex **23** also showed a change in melting temperature, but again was relatively modest (ΔT_m = -0.9°C).

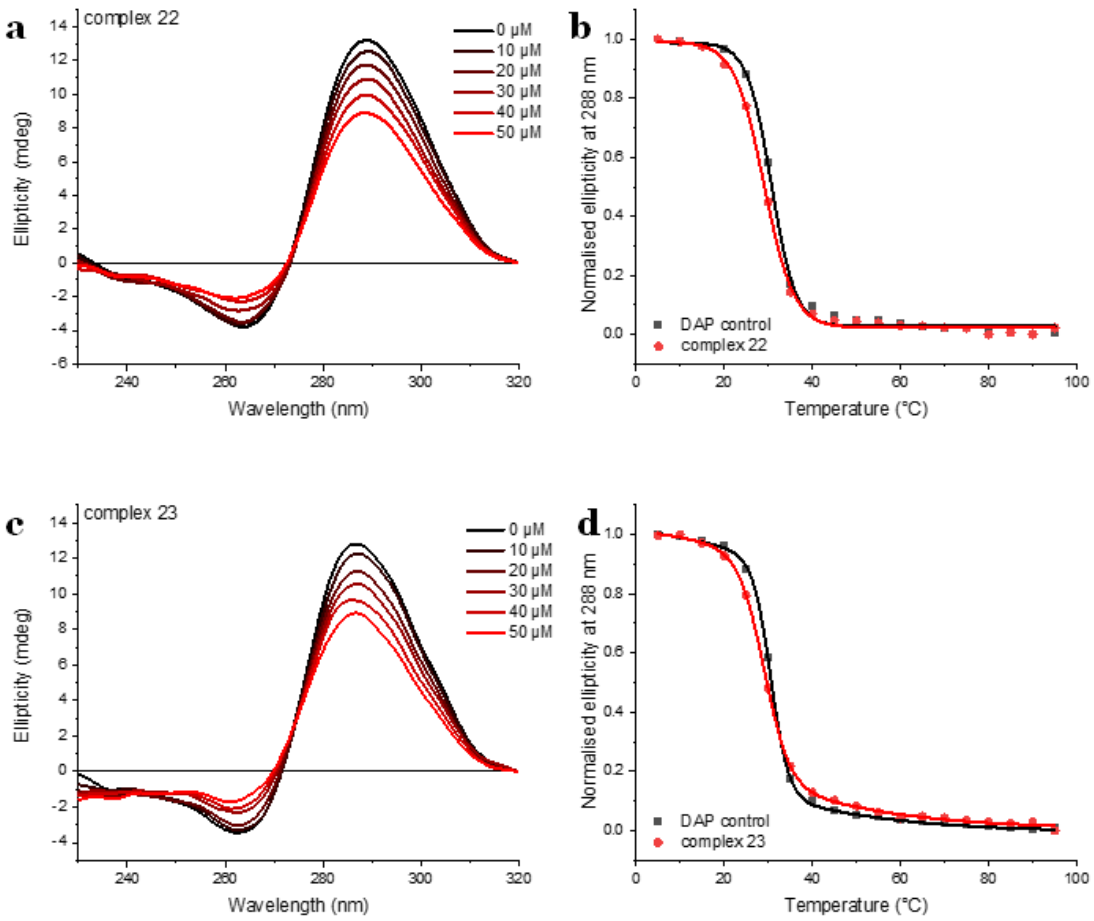


Figure 3.13 *Interaction of compounds* ***22*** *and* ***23*** *with i-motif DNA. (a) CD titration of DAP (10 μM) 0–50 μM of compound* ***22****, CD signal of ligand was subtracted. (b) CD melting experiment with DAP in the absence and the presence of 50 μM of compound* ***22****. c). CD titration of DAP (10 μM) with 0–50 μM of compound* ***23****. d). CD melting experiment with DAP in the absence and the presence of 50 μM of compound* ***23****. All experiments were performed in 10 mM sodium cacodylate buffer, 100 mM potassium chloride, pH 6.8.*

Both complexes were then also examined with *Hif1a* i-motif. From the titration experiments, there was a very clear decrease in the signals at 288 and 255 nm, respectively. However, these two complexes did not affect the thermal stability of the *Hif1a* i-motif structure. In the presence of complex **23**, the shape of the melting curve of *Hif1a* was significantly changed, although there was no change in melting temperature, this indicates that the complex still had an effect on *Hif1a* i-motif (***Figure 3.14***).

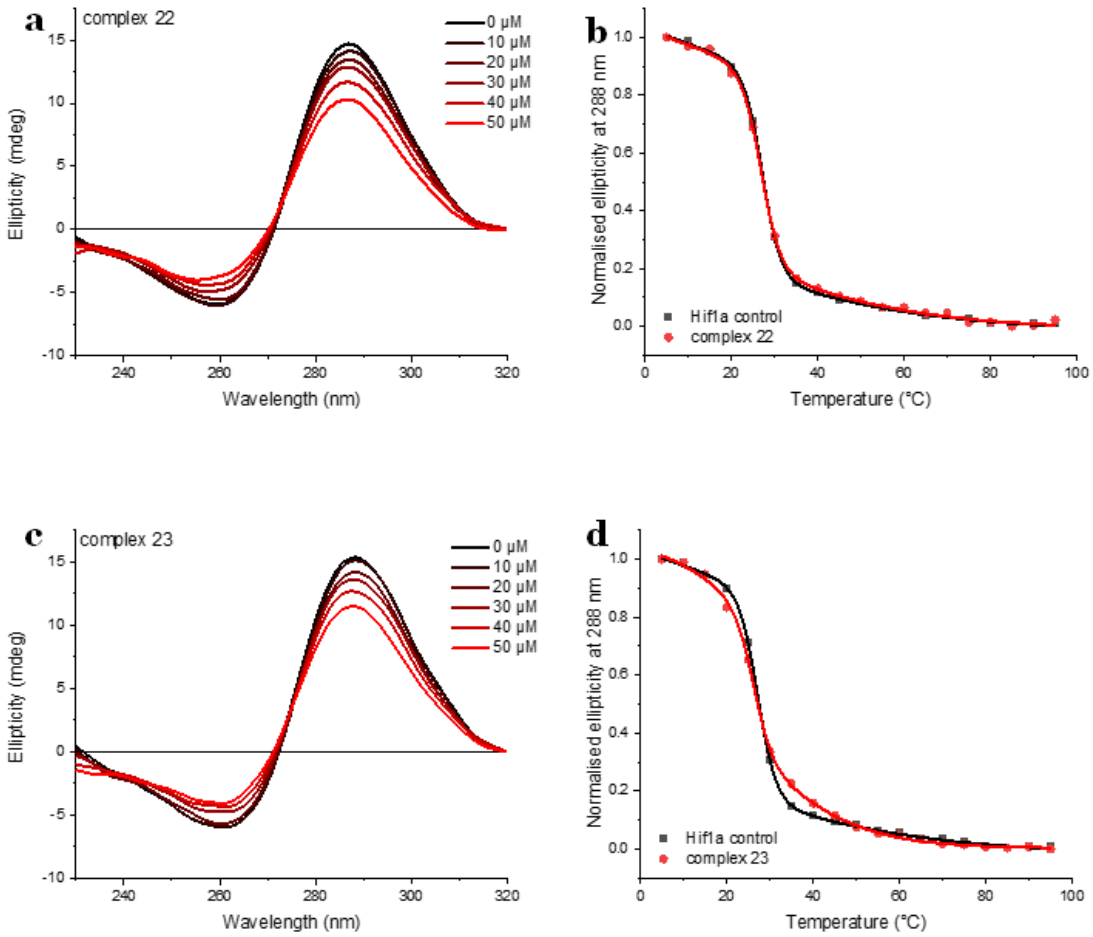


Figure 3.14 *Interaction of compounds* ***22*** *and* ***23*** *with i-motif DNA. (a) CD titration of Hif1α (10 μM) 0–50 μM of compound* ***22****, CD signal of ligand was subtracted. (b) CD melting experiment with Hif1α in the absence and the presence of 50 μM of compound* ***22****. c). CD titration of Hif1α (10 μM) with 0–50 μM of compound* ***23****. d). CD melting experiment with Hif1α in the absence and the presence of 50 μM of compound* ***23****. All experiments were performed in 10 mM sodium cacodylate buffer, 100 mM potassium chloride, pH 6.8.*

Then complexes **22** and **23** were tested with the G-quadruplex structure from the human telomere (***Figure 3.15***). The CD spectra of *hTelo*G showed positive peaks at both 245-260 nm and 290 nm, indicating parallel and antiparallel G-quadruplex species formed. Upon the titration of **22**, a large decrease at 245-260 nm and a little increase at 290 nm was detected, which suggested this ligand unfolds the parallel G-quadruplex topology and induces antiparallel structure formed. With the titration of complex **23**, a decrease in CD signal for both parallel and

antiparallel structures was observed. However, the quality of data in the presence of ligands are quite noisy, the reason might be subtracting the CD data of ligand. Furthermore, there is no significant change at 295 nm might also be the lost in the noise during the subtract of ligand CD signal. Although CD melting showed both ligands have no change in melting temperature of *hTelo*G structure, **22** and **23** altered the melting curves, which suggested some sort of interaction happened, but the stability of the DNA remained the same.

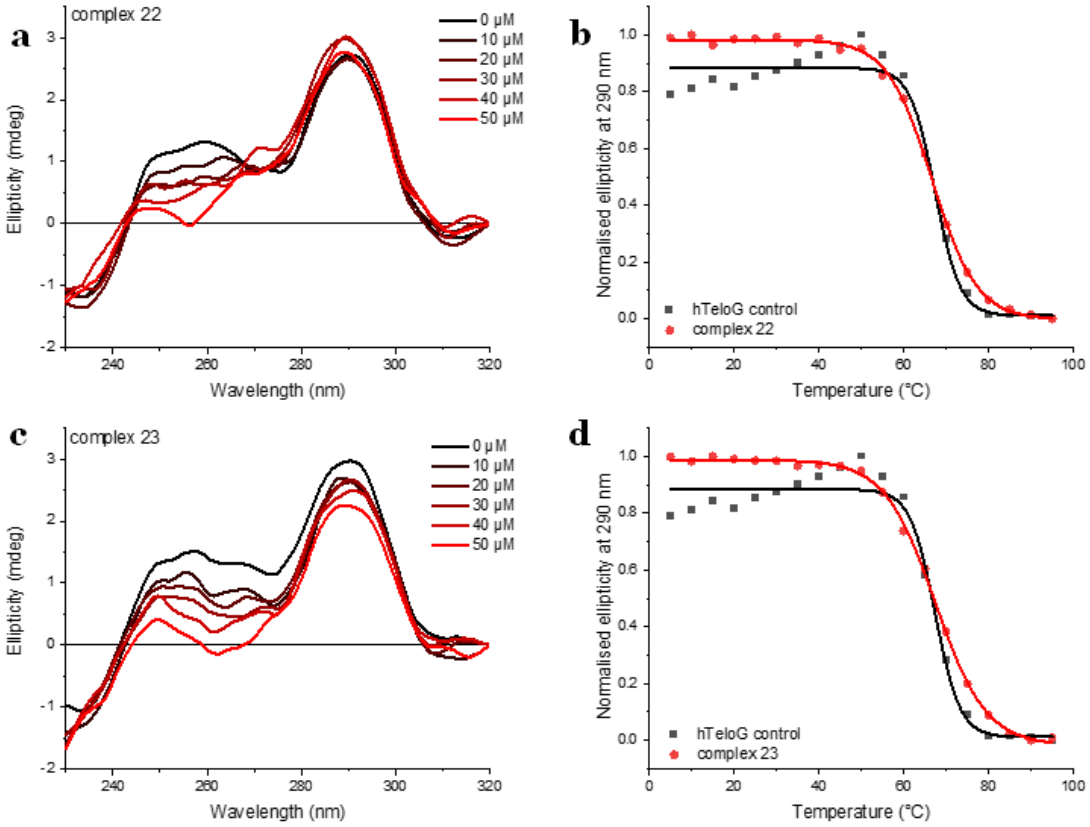


Figure 3.15 *Interaction of compounds* ***22*** *and* ***23*** *with G-quadruplex DNA. (a) CD titration of hTeloG (10 μM) 0–50 μM of compound* ***22****, CD signal of ligand was subtracted. (b) CD melting experiment with hTeloG in the absence and the presence of 50 μM of compound* ***22****. c). CD titration of hTeloG (10 μM) with 0–50 μM of compound* ***23****. d). CD melting experiment with hTeloG in the absence and the presence of 50 μM of compound* ***23****. All experiments were performed in 10 mM sodium cacodylate buffer, 100 mM potassium chloride, pH 7.0.*

3.3.3 Discussion

This was the first investigation, to the best of our knowledge, of the interaction of Pd complexes and i-motif structures.[237] Following biophysical experiments testing these compounds with DNA sequences, the results indicated that the two Pd analogues **22** and **23** interact with different DNA structures. FID assays showed much less TO displacement for double-stranded and *hTelo*G DNA, suggesting **22** and **23** have a preference to bind i-motif structures. CD titration experiments suggested both Pd complexes affect the topologies of both i-motif and G-quadruplex DNAs. In FID assays, DNA structures were carefully annealed and left sufficient time for equilibrating. However, when ligands are added, there is still an equilibrium that will exist between the ligand and oligonucleotide. For this experiment, 96-well plates were used, therefore, the equilibration time may not be exactly the same for each ligand. In addition, as the measurement was performed immediately after adding ligand, any slow structural changes would not be captured in this experiment. This might cause inaccurate fluorescence results.

From CD melting experiments, with three i-motifs (*hTelo*C, *DAP*, and *Hif1α*), **22** showed destabilizing effects on *hTelo*C. Ligand **23** only showed destabilization of *hTelo*C, indicating it has a preference for stabilization of the i-motif from the human telomere. Furthermore, the Pd complexes **22**, and **23** presented no change in melting temperature for *hTelo*G, suggesting these two complexes might have selectivity to i-motif over other DNA structure types. As described in **Section 3.2**, we have found many ligands from the NCI that have destabilizing effects on *hTelo*C i-motif, but the destabilization effects were quite less (ΔT_m between -1 to -3°C) compared to complex **22** (ΔT_m = -8.1°C). Also, in the literature, the Waller group identified some G-quadruplex ligands that destabilize i-motif structures, e.g., berberine destabilizes the i-motif forming sequences from the promoter *ATXN2L* and *DAP* with ΔT_m = -1.2°C and -3.4°C, respectively, BRACO-19 showed a quite strong destabilizing effect on DAP i-motif (ΔT_m = -11.6°C). RHPS4 destabilizes *ATXN2L* and *DAP* relatively strongly (ΔT_m = -7.5°C and ΔT_m = -8.7°C, respectively). And TMPyP4 strongly destabilizes *DAP* i-motif with ΔT_m of -16.0°C.[97,211] The study of i-motif destabilizers could be used in aiding the

understanding of the mechanism of the DNA-ligand binding process. In anticancer studies, it has been suggested that DNA replication might be influenced by drug-induced destabilization, as DNA opening needs less energy when the double helical structure is destabilized by the ligand.[247] In the line with i-motif, could also play an essential role in gene expression. Hurley's work from 2014 showed that destabilization of the i-motif forming sequence from the promoter region of *BCL2* resulted in a reduction in the expression of the gene.[248] Given that up to 69% of all oncogenes contain a putative quadruplex forming sequence in the promoter region, the i-motif destabilizers described may be used to downregulate these genes and could be used to develop potential antitumour agents.

3.4.1 The Study of an Iron Cylinder Binding to Quadruplex DNA Structures

As discussed in **Chapter 1**, i-motif forming sequences could be a potential target for genetic diseases treatment, and small molecules that bind to i-motif can be potential drugs to treat cancer or other diseases. The synthetic metallo-supramolecular cylinders were designed by the Hannon group, and each of them has a similar size and shape to the zinc finger motifs in DNA-recognition proteins.[249–252] These metal cylinders are thought to bind B-DNA in the major groove as they are predicted to be too large to fit into the minor groove. The cylinders can be designed with different metal ions, but herein, the focus is on the interaction between DNA and di-iron (II) cylinder $[Fe_cL_3]^{4+}$, $[Fe_2(C_{25}H_{20}N_4)_3]Cl_4$ (***Figure 3.16***), which has a deep purple colour. From the Hannon group's study, they investigated the effect of an iron cylinder on proliferation and survival in tumour and normal healthy cell lines.[253] The iron complex was found to reduce the mitochondrial activity of cell cultures and inhibit the cell cycles, furthermore, promoting apoptosis and leading to increased cell death. In addition, compared to the existing anticancer drug, the iron cylinder is not genotoxic, therefore, the iron cylinder could have potential anticancer properties with no genotoxicity for chemotherapy.[253]

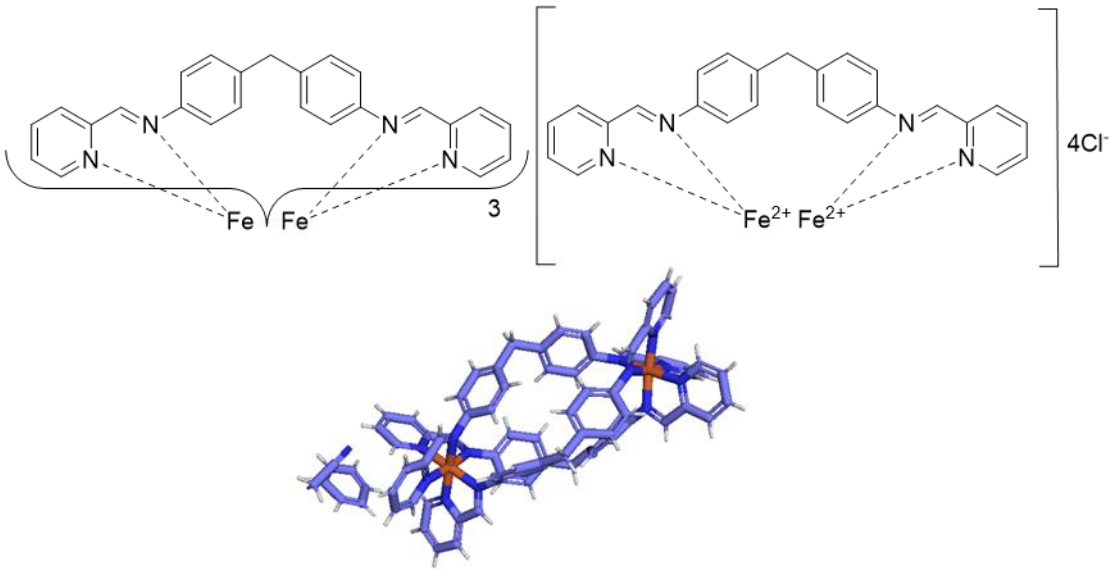


Figure 3.16 *Chemical structure of* ***Fe-cylinder*** *(top) and the crystal structure*[254] *(bottom).*

Hannon and co-workers identified that **Fe-cylinder** has biological effects that increase cell death and inhibits the cell cycle,[253] they also suggested that it cannot only bind to the major groove of B-form DNA, but it can fit into three-way junctions.[249,251,255] however, we were curious its ability to interact with other types of DNA structure, such as GC-rich quadruplexes, therefore, in this project, we studied the interactions between **Fe-cylinder** and different DNA sequences, including i-motif forming sequences from human telomere (*hTelo*C) and promoter region of *Hif1a* and *DAP*, double-stranded and G-quadruplex from the human telomere (*hTelo*G). Several biophysical methods were performed to investigate the effects of **Fe-cylinder** on DNA, including CD titration, CD melting, and Job Plot.

3.4.2 CD Titration

The first experiments performed were CD titration to determine the effect of the small molecule on DNA structure, the buffers used contained 10 mM sodium cacodylate and 100 mM potassium chloride buffer at desired the pH for each DNA sequence (pH 5.5 for *hTelo*C, pH 6.8 for *Hif1a* and *DAP*, pH 7.0 for *hTelo*G and DS).

The CD spectra of double-stranded DNA (pH 7.0) showed a characteristic positive peak at ~280 nm and a negative peak at 251 nm,[256] suggesting a double helical structure formed. With the addition of **Fe-cylinder** (***Figure 3.17***), there is a decrease in CD signal intensity when titrated up to 30 μM, after the titration up to 40 μM, there was a hypsochromic shift from ~280 nm to ~275 nm, also an iso-elliptic point was observed ~263 nm, suggesting the interaction between **Fe-cylinder** and double-stranded DNA, which is consistent with the previous work that the cylinder binds to double helical DNA.

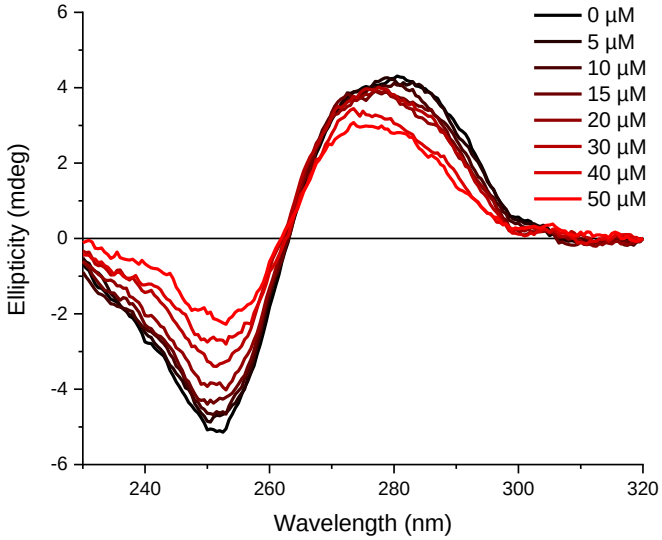


Figure 3.17 *CD spectra of titration experiments of 0 – 50 μM of* ***Fe-cylinder*** *with 10 μM DS in buffer containing 10 mM sodium cacodylate and 100 mM potassium chloride at pH 7.0. CD signal of ligand was subtracted.*

After the titration of **Fe-cylinder** with double-stranded DNA, this ligand was also examined with *hTelo*G (pH 7.0). In the CD signal (***Figure 3.18***), *hTelo*G showed a positive peak at 250 – 260 nm (parallel G-quadruplex form) and 290 nm (anti-parallel G-quadruplex form),[209,210] indicating the mixed species existed in the solution. Upon titration of up to 50 μM of **Fe-cylinder**, there was a slight decrease of CD signal at 290 nm, but a clear decrease at 250 nm was observed that indicated a possible unfolding parallel G-quadruplex structure, however, compared to the loss of the parallel form, there is no significant change in antiparallel species.

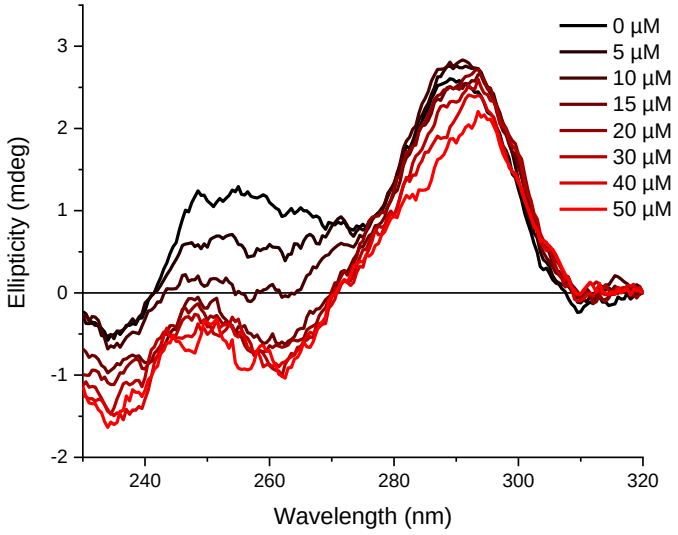


Figure 3.18 *CD spectra of titration experiments of 0 – 50 µM of* ***Fe-cylinder*** *with 10 µM hTeloG in buffer containing 10 mM sodium cacodylate and 100 mM potassium chloride at pH 7.0. CD signal of ligand was subtracted.*

The **Fe-cylinder** was then examined with i-motif forming sequences from the human telomere and *Hif1a* i-motif up to 50 µM of ligand. In *Hif1a*, there was a large decrease in CD signal at 288 nm, two iso-elliptic points at 235 and 275 nm were observed, and a hypsochromic shift from 288 nm to 280 nm, which indicated the ligand significantly unfolds Hif1α structure. While in the titration with the *hTelo*C sequence, only a slight decrease was found in the CD signal at 288 nm. Although compared to *Hif1a* and *hTelo*C had big difference in titration, they were performed at different pHs, as they have different pH properties, therefore, if *Hif1a* was repeated at low pH where could form quite stable i-motif structure, the **Fe-cylinder** may not have such huge effect on the highly stable i-motif structure. The additional CD titrations were also performed, we would like to add more **Fe-cylinder** until the CD signal reaches to plateau or no further change (***Figure 3.19***).

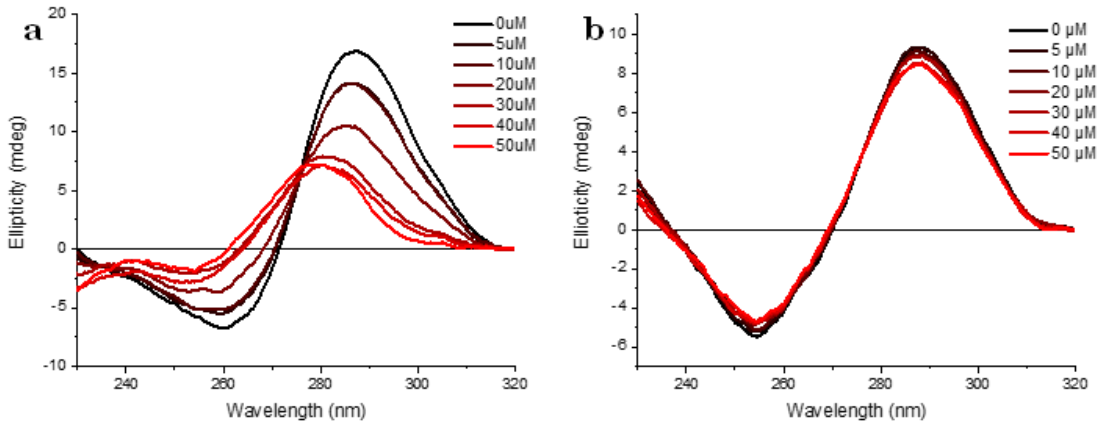


Figure 3.19 *CD spectra of titration experiments of 0 – 50 µM of* ***Fe-cylinder*** *with 10 µM i-motif DNA a). Hif1a at pH 6.8, b). hTeloC at pH 5.5, all experiments were performed in a buffer containing 10 mM sodium cacodylate and 100 mM potassium chloride. CD signal of ligand was subtracted.*

A further CD titration of *Hif1a* showed a positive peak around 288 nm and a negative one at 260 nm, indicating an i-motif structure formed. Upon titration of up to 5 µM of **Fe-cylinder** into *Hif1a* solution (pH 6.8), there was a decrease in CD signal at 288 nm, then upon titration up to a higher concentration of **Fe-cylinder**, the CD signal intensity at 288 nm kept decreasing in a dose-dependent fashion until ~ 100 µM in which no further reduction in ellipticity was observed. At this point, it was observed that some purple precipitation was visible within the cuvette, so this was the maximum concentration this was performed. During the titration, there was a hypsochromic shift of around 10 nm from 288 nm to 278 nm observed (***Figure 3.20***), both decreases in the CD signal and hypsochromic shift suggested ligand-dependent disruption or unfolding of the i-motif structure,[257] the precipitation suggests DNA-complex aggregation could form.

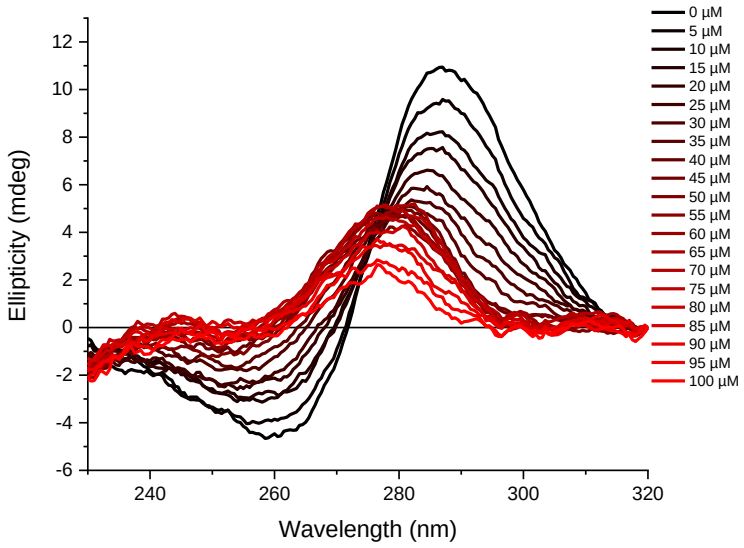


Figure 3.20 *CD spectra of titration experiments of 0 – 100 µM of* ***Fe-cylinder*** *with 10 µM Hif1a in buffer containing 10 mM sodium cacodylate and 100 mM potassium chloride at pH 6.8. CD signal of ligand was subtracted.*

In contrast, the **Fe-cylinder** had no significant effect on the CD spectrum of *hTelo*C (***Figure 3.21***), and the spectrum as unchanged in the presence of 100 µM **Fe-cylinder.** For example, the hypochromic shift of ~5.8 mdeg in *Hif1a* at 35 µM was observed, while in the *hTelo*C, only ~1.7 mdeg at the concentration of 100 µM was detected. Moreover, in this case, there was no significant hypsochromic or bathochromic shift, indicating **Fe-cylinder** had little interaction with *hTelo*C i-motif structure. It is, however, considered that *hTelo*C was performed at pH 5.5, so further experimental controls could be performed to determine the general effect of pH on the interaction.

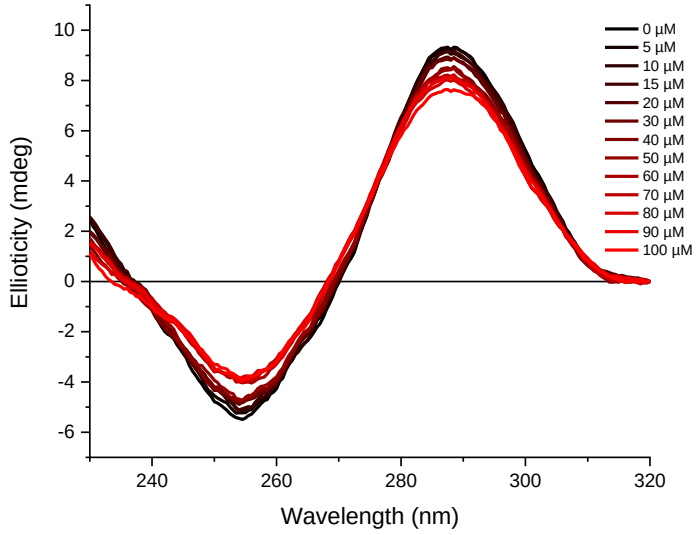


Figure 3.21 *CD spectra of titration experiments of 0 – 100 µM of* ***Fe-cylinder*** *with 10 µM hTeloC in buffer containing 10 mM sodium cacodylate and 100 mM potassium chloride at pH 5.5. CD signal of ligand was subtracted.*

The i-motif forming sequence from the promoter *DAP* was also studied with the **Fe-cylinder**. This time **Fe-cylinder** was titrated to DNA directly up to 100 µM, and no precipitation was observed. Although there was no significant hypsochromic shift at 288 nm (1 nm) compared to *Hif1α* (10 nm), a large loss in the CD signal at both 288 and 265 nm, and two iso-elliptic points at ~235 and ~276 nm were detected (***Figure 3.22***). As in the solution, an equilibrium between single-stranded and i-motif species exists at the same time, such a large loss in signal (8 mdeg) indicates upon the addition of **Fe-cylinder**, the equilibrium was shifted towards single-stranded species, which suggested the **Fe-cylinder** also showed some of the unfolding effects on *DAP* i-motif structure. However, at pH 7.0, the DNA i-motif is not completely unfolded, in order to observe the spectrum of an totally unfolded strand, the experiments could be performed at higher pH conditions such as 8.0 or higher.

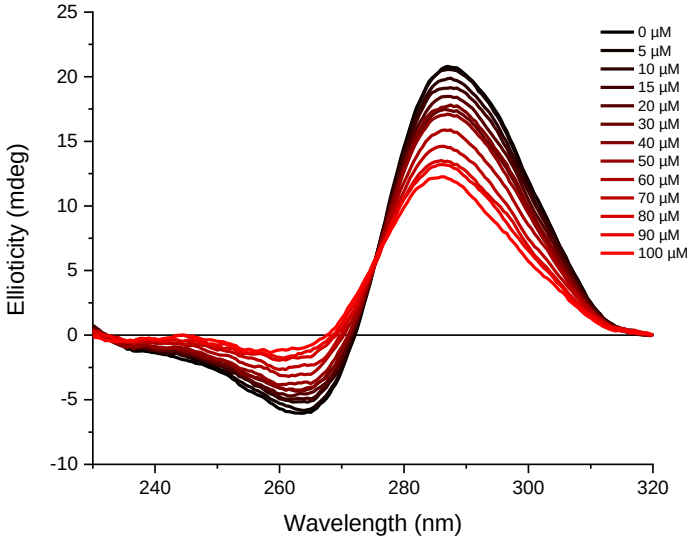


Figure 3.22 *CD spectra of titration experiments of 0 – 100 µM of* ***Fe-cylinder*** *with 10 µM DAP in buffer containing 10 mM sodium cacodylate and 100 mM potassium chloride at pH 6.8. CD signal of ligand was subtracted.*

3.4.3 Unfolding and Destabilization of i-Motif Structure from Hif1α

According to the CD titration experiments, it was suggested that the **Fe-cylinder** seems to unfold the *Hif1a* i-motif structure but has less interaction with the *hTelo*C sequence. After plotting the molar ellipticity at 288 nm of *Hif1a, hTelo*C, and *DAP* against the concentration of **Fe-cylinder** (***Figure 3.23***), there was a clear decrease from the point at 70 µM, indicating precipitation starting aggregated, therefore, the further fitting used the curve from 0 – 70 µM.

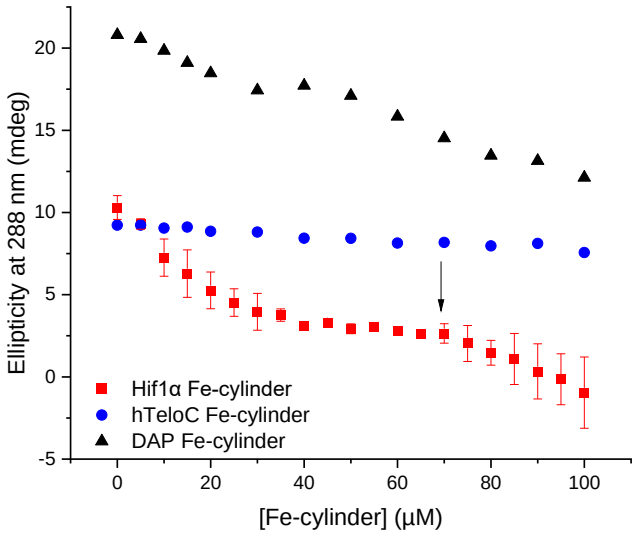


Figure 3.23 *CD plot of molar ellipticity at 288 nm of Hif1α (■ red squares), hTeloC (● blue circle), and DAP (▲ black triangle) against Fe-cylinder. The arrow indicates precipitation formed in Hif1α. Error bars showed the standard deviation across two repeats.*

The plot of the molar ellipticity of *Hif1α* against the concentration of the **Fe-cylinder** showed a sigmoidal curve, indicating a cooperative effect[258] that the **Fe-cylinder** at one binding site changes the binding affinity at other sites, which suggested *Hif1α* i-motif has potentially more than one binding site for **Fe-cylinder**, therefore, when the increased concentration of ligand, **Fe-cylinder** would possibly bind to multiple sites thus to unfold the i-motif structure (***Figure 3.24***).

The sigmoidal curve showed a cooperative relationship between the ligand and DNA. This suggests that as the ligand binds DNA it creates more binding sites and induces unfolding. These data were therefore fitted using the Hill1 equation (***Equation 8***), where θ is the fraction of ligand binding sites filled, n is the Hill coefficient and [L] is ligand concentration, K_A is ligand concentration occupying half of the binding sites.

***Equation** 8 Hill Fit.*

$$\theta = \frac{[L]^n}{[K_A]^n + [L]^n}$$

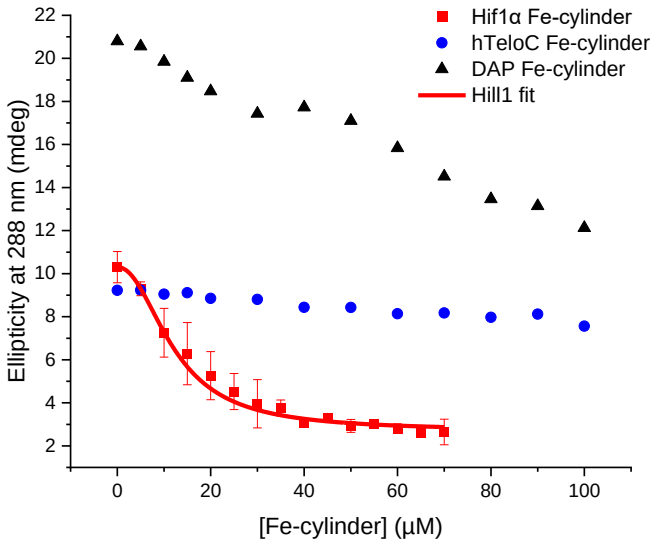


***Figure 3.24** CD plot of molar ellipticity at 288 nm of Hif1α (■ red squares), hTeloC (● blue circle) and DAP (▲ black triangle) against the concentration of **Fe-cylinder**. Ellipticity at 288 nm against ligand concentration of Hif1α was fitted by the Hill equation, errors bars showed the standard deviation across two repeats.*

In general, cooperative binding may be positive (n > 1) or negative (n < 1) depending on the way ligand binding increases or decreases the binding affinity of other binding sites. From the Hill fitting, the Hill coefficient of 2.1 was determined for **Fe-cylinder** to *Hif1α*, which has identified **Fe-cylinder** has positive cooperativity of *Hif1α*. In addition, $[DC]_{50}$ value, the concentration of **Fe-cylinder** required to get a 50% of reduction in molar ellipticity, was also calculated to be 17.2 ± 2.2 μM for Hif1α, indicating a relatively strong unfolding effect on *Hif1α*. In contrast, the plot for **Fe-cylinder** and *hTelo*C (***Figure 3.24***), did not present strong evidence of cooperative effect, because the corresponding binding curve gave a comparative linear loss of signal. For the plot of *DAP* and the **Fe-cylinder**, although it seems a binding curve, the ligand for these experiments were not from

the exact same stock, to allow for the difference in concentration range, therefore, it could be that these differences may cause a discrepancy here in the results.

From the CD titration and an analysis of the correlation between molar ellipticity and the concentration of ligand, *Hif1α* was selected for further study. This sequence was identified as forming a stable structure at neutral pH and thus may be more physiologically relevant.[62] It also had the largest effect observed from titration experiments. These results suggest that the **Fe-cylinder** can unfold the Hif1α i-motif, so the next step was to explore its effects on the thermal stability of the i-motif. As precipitation was observed at 70 μM in the CD titrations CD melting experiments were performed and melting temperatures were calculated in the absence and presence of 50 μM of the compound. The plot of normalized molar ellipticity at 288 nm (fraction folded) gave the sigmoidal curves for *Hif1α* with and without **Fe-cylinder** (***Figure 3.25***), and they were fitted using the Dose-Response sigmoidal fitting in OriginLab to reveal the melting temperatures. The ΔT_m was found to be -17.3°C, indicating a strong destabilizing effect on the thermal stability of Hif1α i-motif.

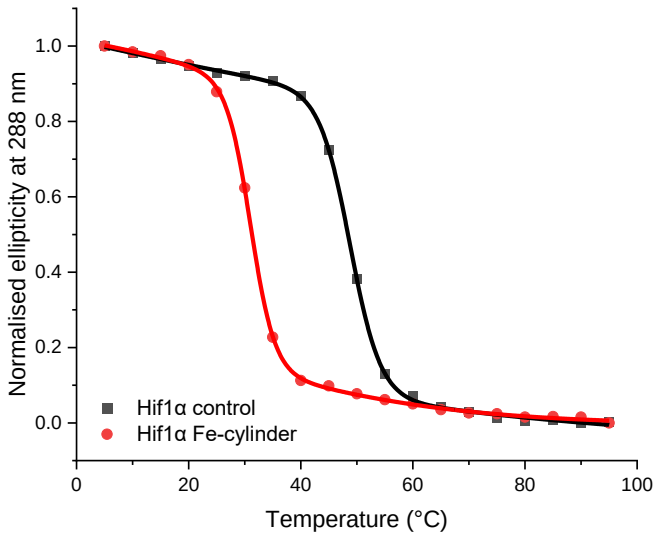


Figure 3.25 *CD melting of Hif1α in the absence (black square) and the presence (red circle) of 50 μM* ***Fe-cylinder****, with 10 μM Hif1α in the buffer containing 10 mM sodium cacodylate and 100 mM potassium chloride at pH 6.8.*

3.4.4 Stoichiometry

To examine the potential binding sites on *Hif1α* for the **Fe-cylinder**, a continuous variation binding analysis (Job Plot) was performed.[259260] A group of samples was prepared in which the ratio of **Fe-cylinder**:Hif1α-i-motif varied between 1:9 and 9:1, but the whole concentration of ligand + DNA in each sample was kept constant at 20 μM, and then the CD spectra were measured for each sample. Also, a series of samples with the same concentration of DNA but in the absence of **Fe-cylinder** were examined, for the analysis, this data was subtracted from the first set to get the change in molar ellipticity. The plot of the change in molar ellipticity at 288 nm against the ratio of DNA gave a graph (***Figure 3.26***) that would correspond to the stoichiometry.

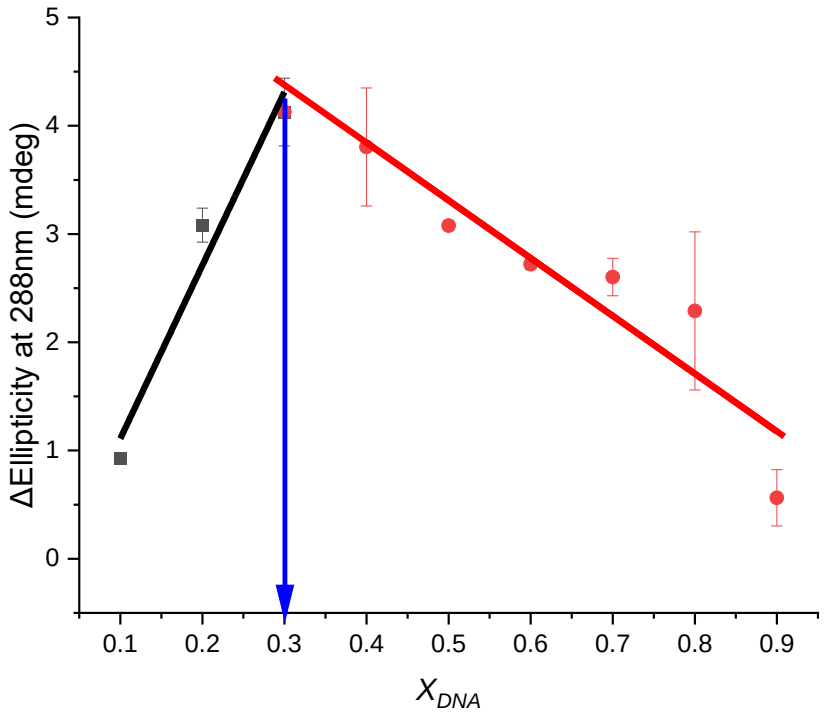


Figure 3.26 *Job Plot presenting the stoichiometry of the interaction between Hif1α and* ***Fe-cylinder*** *in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 6.8. The error bars show the standard deviation across 2 repeats.*

To determine stoichiometry, the plotted points were fitted by two linear lines that intersect each other. The intersection point reflects the X-axis giving the molar ratio of DNA, which can be converted into stoichiometry. From the analysis, the graph gave an average stoichiometry of 2.3:1, which equals to roughly 2:1 of **Fe-cylinder**:DNA. We considered that there might be the possibility that the stoichiometry may be higher than 2:1 but analysis of the two intersecting linear standard curves indicates that the stoichiometry of 2.3:1 is within error of 2:1. The 2:1 stoichiometry also seems reasonable, given the size of the **Fe-cylinder**, it is likely that it would target the larger potential binding sites in the structure, and these might be the major grooves. Other options could be end-stacking or interaction with the loops, however, as **Fe-cylinder** destabilises *Hif1α* at pH 7.0, this is indicative that the ligand might bind other structures or single-stranded DNA rather than the i-motif structure.

3.4.5 Discussion

From the CD titration experiments, the **Fe-cylinder** showed interaction with double-stranded DNA. This result was not surprising as the previous reports in the literature showed that the **Fe-cylinder** binds to the major groove of duplex DNA.[250]The CD titration indicated that it might unfold the parallel G-quadruplex structure of *hTelo*G. Although this effect was modest, it could be worth investigating the unfolding ability further in the future. However, it was clear that there was a much stronger interaction and effect with i-motif DNA. In the examination with i-motif structure, among the tested three sequences, Hif1α showed a much stronger effect in the presence of **Fe-cylinder** compared with *DAP* and *hTelo*C, suggesting it has selectivity to *Hif1α* i-motif. CD melting showed that the **Fe-cylinder** gave the largest decrease in melting temperature (-17.3°C) of all ligands that have been presented in this thesis. From the data, the unfolding of the i-motif seemed relatively easy compared to the unfolding G-quadruplex. In fact, in general, within this thesis and the literature there seem to be more destabilizing compounds for i-motif compared to G-quadruplex. The destabilization of the G-quadruplex by comparison is more unusual. The classic ligand TMPyP4 has also been identified to unfold both G-quadruplex and i-motif

structures,[97] the next chapter, will discuss some of TMPyP4's properties and unfolding effects.

3.5 Conclusions

Overall, we have identified many ligands that interact with i-motif structures. The outcomes from this project were two NCI compounds **71795** and **19990** could be used as potential i-motif stabilizers for i-motif conjugate design. While from the last chapter, we have not investigated a double-stranded selective minor groove binder, therefore, currently, we cannot proceed with the proof of concept. In the study of metal complexes, Pt(I) complexes **22** and **23** showed selectivity to human telomeric i-motif rather than other types of DNA structures, and **22** showed a significant destabilization effect on the stability of *hTelo*C rather than **23**. Both Pb complexes also showed interactions with other i-motif and G-quadruplex structures, but no moderated effect on their thermal stability. The **Fe-cylinder** was examined with several i-motif sequences, it shows strong destabilizing effect on Hif1α while no significant effect on *hTelo*C. The reason could be different pHs were used for *Hif1a* and *hTelo*C experiments, and it is easier for **Fe-cylinder** to destabilise the i-motif structure at higher pH, therefore, if the tested i-motif sequences were repeated at pH 5.5, **Fe-cylinder** may not have such huge destabilising effect on *Hif1a*. In the the future work, it will be worth to investigate it. For the many ligands that showed destabilizing effects on i-motif structure, although they are not suitable for i-motif conjugate synthesis, they all contributed to the study of the interaction between ligands and i-motif DNA structures.

Chapter 4

Unfolding RNA G-Quadruplex by TMPyP4

4.1 Introduction

This chapter is as part of the paper "Mechanistic Insights into the Ligand-Induced Unfolding of an RNA G-Quadruplex" in JACS.[145]

In the previous chapter, an abundance of i-motif unfolders has been identified, further supporting other previous work from the literature.[97] This work identified some small molecules that could slightly destabilize and unfold G-quadruplex DNA. There is not much known about G-quadruplex unfolding compounds, despite there being thousands of well-characterized G-quadruplex ligands.[261] Prof. Shozeb Haider led work to investigate the unfolding effects of **TMPyP4** on RNA G-quadruplex using well-tempered metadynamics (WT-MetaD). The Waller group has already previously characterized the binding properties of **TMPyP4**, which had been shown to destabilize i-motif and could significantly unfold i-motif from the promoter regions of *DAP* and *ATXN2L* compared to other compounds at pH 7.0.[97] Also previous work on ligand-DNA unfolding (**Chapter 3**) gave me the necessary experience to study the unfolding of G-quadruplexes. Thus, we were able to collaborate and support Prof. Haider's computational study of **TMPyP4** binding RNA G-quadruplex with the necessary biophysical characterization.

As discussed in **Chapter 1**, G-quadruplex forming sequences play an important role in regulating gene expression. Wide human genomic analyses have identified >500,000 DNA and >6,000 RNA putative G-quadruplex forming sequences (*PQFS*).[32,262–264] Their existence has potential important functions in regions involved in telomere maintenance and DNA damage response in aging.[265] For instance, both human telomeric DNA d(TTAGGG) and telomeric repeat-containing RNA r(UUAGGG) can form into G-quadruplex structures and are significant to telomere biology and epigenetics regulation.[266] Thorough mapping of the DNA replication origins has indicated that almost 90% of start sites contain RNA G-quadruplex structures.[267] With a high population of *PQFS* in gene transcription start sites, indicates a role of the G-quadruplex in regulating the transcription process.[268,26] Recent research indicated that an RNA G-quadruplex plays a role in the regulation of RNA expression in mitochondria,[269] and leads to

the formation of epitranscriptomics[270] and membrane-less organelles.[271,272] It has been reported that a G-quadruplex may be involved in the regulation of translation because of the high frequency of *PQFS* in the regions of 5′ untranslated regions (5′-UTR) of encoded mRNA.[273,274] Furthermore, it has been shown that RNA G-quadruplex in cells is involved in translation, replication, and transcription.[275] In addition to the human genome, *PQFS* also widely exists in bacteria[276] and viruses.[277]

The studies of G-quadruplex binding ligands are mostly focused on the stabilization of the G-quadruplex structure.[278–280] However, the importance of unfolding the G-quadruplex structure is not fully apprehended, the study has shown unfolding of the G-quadruplex structure can enhance the efficiency of translation in the *FMR1* gene and also the 5′UTR of FMR1 mRNA in Fragile X syndrome.[281,282] In addition, age-related changes or genetic loss in the proteins with G-quadruplex modulation, have been suggested to promote neurological disorders and brain aging.[283] Therefore, destabilization of the G-quadruplex structure can be significant in controlling gene expression and in developing applications of treatment for age-related neuro-disorders. A number of small molecules have been reported to destabilize the G-quadruplex structure, for example, **TMPyP4** is identified to disrupt the G-quadruplex structures in Fragile X FMR1 gene[282] and the MT3 endopeptidase mRNA sequence.[284] These results of G-quadruplex destabilising by **TMPyP4** are unusual, as it is normally considered to be a quadruplex stabiliser. [285,286]

In comparison with other classical G-quadruplex binding ligands, **TMPyP4** (***Figure 4.1***) is quite special as it can either stabilize or destabilize G-quadruplex structures, depending on the sequences studied. It has been shown that **TMPyP4** stabilizes the G-quadruplex structure and also has antitumor[285,286] and antivirus activity.[287] **TMPyP4** was reported to stabilize *c-MYC Pu24I* sequence stacking on the top of G-tetrad using NMR (PDB 2A5R).[288] The mechanism of binding mode between human telomeric G-quadruplex and **TMPyP4** is stacking with the loop TTA instead of G-quartet (PDB 2HRI).[289] In contrast, there is also research suggesting that **TMPyP4** can destabilize and unfold G-quadruplex structures, for

example, the Basu group identified that **TMPyP4** unfolds the extreme stable *M3Q* G-quadruplex structure of *MT3-MMP* mRNA, and results in enhanced translation of mRNA.[284]

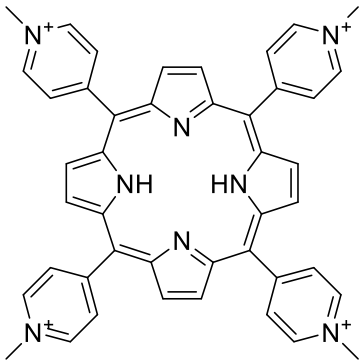


Figure 4.1 *Structure of TMPyP4.*

In this study, we investigated the interaction between **TMPyP4** and one RNA G-quadruplex forming sequence *PQS18-1* r(GGCUCGGCGGCGGA) from the non-coding region of the pseudorabies virus (*PRV*).[290] It's a parallel G-quadruplex structure consisting of four-stacked tetrads stabilized by three potassium cations on the central axis. The crystal structure of the complex of **TMPyP4** and G-quadruplex *PQS18-1* has previously been published (PDB id 6JJH, 6JJI).[291] From the structure, there are two molecules of **TMPyP4** bound to the G-quadruplex, one intercalated in the G-tetrad and one external to the RNA G-quadruplex (***Figure 4.2***).

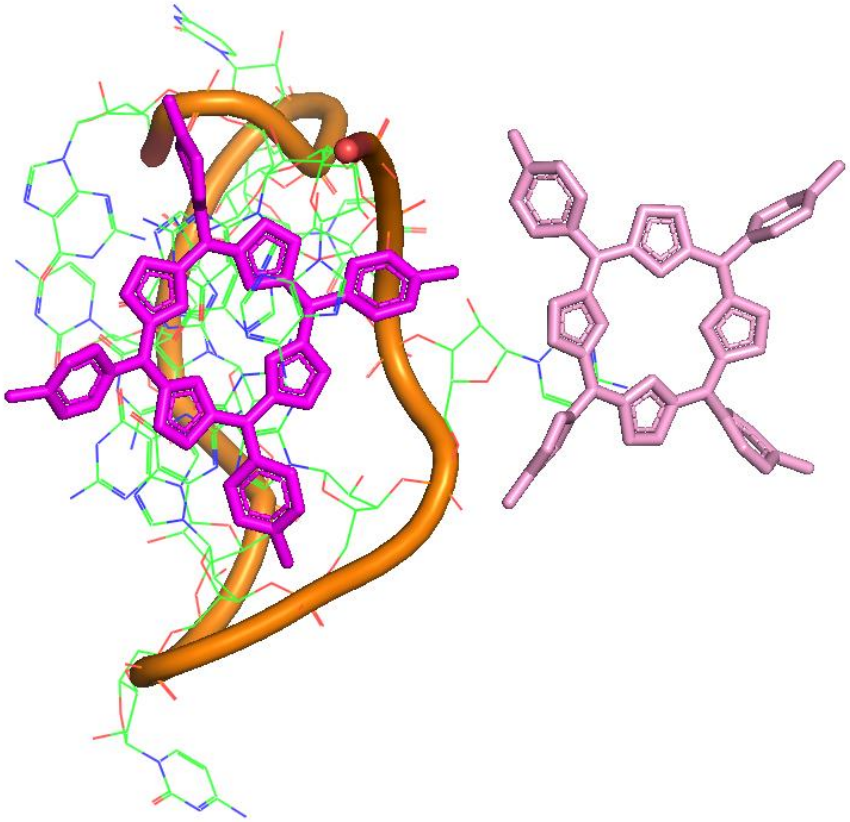


Figure 4.2 *Crystal structure of the PQS18-1 RNA-TMPyP4 complex (PDB ID 6JJH).*

Herein, this project was part of a collaboration with the Haider group, and the mechanism of the unwinding of the G-quadruplex RNA structure by a porphyrin compound **TMPyP4** was studied using WT-MetaD enhanced sampling simulations.[145] The work built from the crystal structure of **TMPyP4** bound to the *PQS18-1* RNA and **TMPyP4** was simulated to bind on the top of the face and in one of the grooves of G-quadruplex RNA.[291] To support and access the reliability of the theoretical results, several biophysical experiments were performed, and CD, UV absorbance, and FRET titrations were used.

4.2 CD Spectra Results

CD spectra were used to investigate nucleic acid structure and folding.[159] The RNA *PQS18-1* samples were prepared in 10 mM lithium cacodylate and 100 mM potassium chloride buffer, as *PQS18-1* RNA was identified to form parallel G-

quadruplex in the presence of K^+.[291] Lithium-based cacodylate buffer, rather than sodium cacodylate, was used to prevent the formation of antiparallel topology,[41] and avoid competition between Na^+ and K^+ in coordinating with the G-tetrad. As discussed in **Chapter 1**, the type of cations plays a key role in G-quadruplex topology. The size of Li^+ is very small, compared to other monovalent cations from group I,[41] therefore, using a lithium-based buffer will ensure only K^+ could coordinate with G-tetrad and avoids the unnecessary complication of mixtures of G-quadruplex topologies.

CD was used to measure the spectrum of *PQS18-1* RNA between 200 and 650 nm. Typically, CD spectra on nucleic acids would be measured between 200 and 320 nm. An extended range of wavelengths was used as previous studies of **TMPyP4** with a G-quadruplex forming sequence from Bcl2, showed an induced CD signal at 436 nm.[292] With the additions of **TMPyP4** to RNA *PQS18-1*, a continually concentration-dependent decrease in CD signal at 264 nm was observed (***Figure 4.3***), it decreased down to a plateau at around 60 – 70 μM (6 – 7 eq), and the purple precipitation was observed, consistent with unfolding effect. A bathochromic shift of the positive band at 264 nm was observed, also the iso-elliptic point at around 248 nm was observed. Thes data suggested that **TMPyP4** was interacting with the *PQS18-1* RNA G-quadruplex structure. The hypochromic change suggested that the ligand was unfolding the G-quadruplex structure, furthermore, **TMPyP4** induced a bathochromic shift at low concentration, which suggests that other structures were adopted. On the other hand, there was no CD signal observed at 295 nm in the absence of **TMPyP4**, after the addition of 10 μM (1 equivalent), a small peak appears at this wavelength, suggesting an antiparallel G-quadruplex formed. With the increased concentration of **TMPyP4**, this signal reaches a plateau at 15 μM (1.5 equivalents) of **TMPyP4** added.

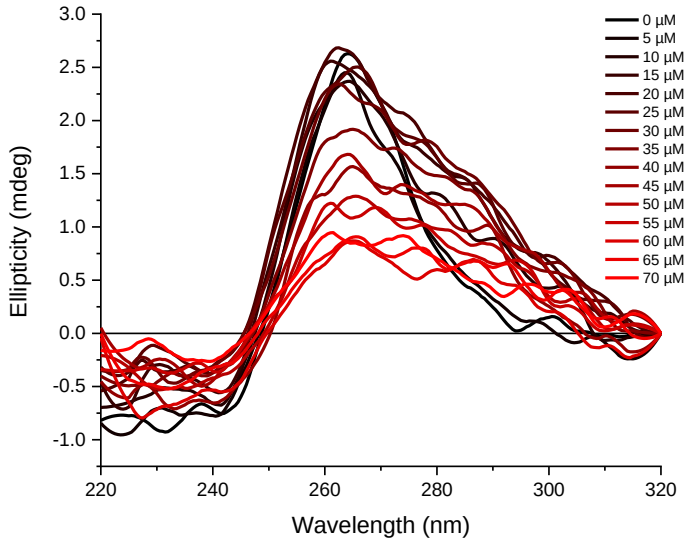


Figure 4.3 *CD spectra of titration experiments of 0 – 70 µM of TMPyP4 with 10 µM PQS18-1 RNA in buffer containing 10 mM lithium cacodylate and 100 mM potassium chloride at pH 7.0. CD signal of ligand was subtracted.*

We considered that this may be an important indicator for ligand binding, and we wanted to see whether **TMPyP4** had the same effect on RNA G-quadruplex. We performed a CD titration where we titrated small aliquots of **TMPyP4** into the RNA (***Figure 4.4 a***). The CD titration of the *PQS18-1* DNA G-quadruplex sequence (***Figure 4.4 b***) with **TMPyP4** was also performed to directly compare it to RNA G-quadruplex. Typically, the CD would usually be performed with 10 µM of nucleic acids, but we found the signal for the DNA d(GGCTCGGCGGCGGA) was very poor. This is due to the fact that RNA G-quadruplexes are much more stable than DNA G-quadruplexes. In this case, we had to use double the quantity of DNA, so the CD titration of DNA was performed at 20 µM.

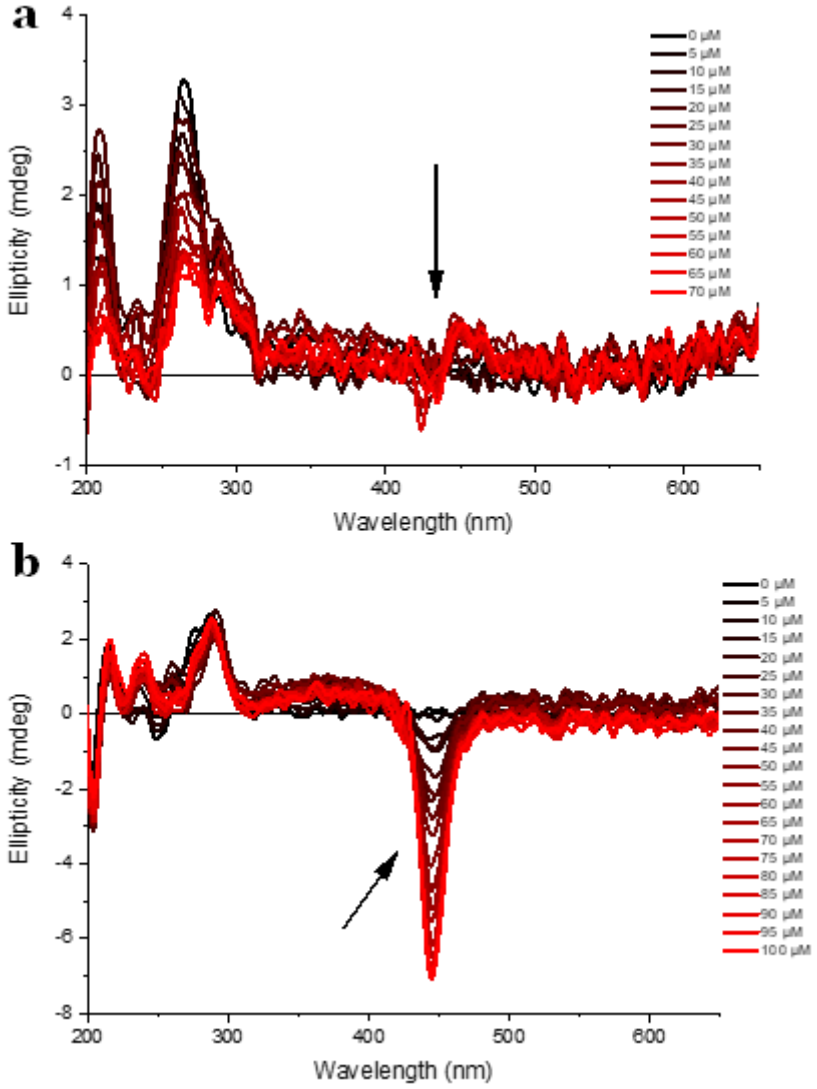


Figure 4.4 *a) CD titration of 10 µM PQS18-1 RNA in the presence of 0–70 µM (0 -7 equivalents) TMPyP4. b) CD titration of 20 µM PQS18-1 DNA in the presence of 0–100 µM (0 – 5 equivalents) TMPyP4. Both experiments were performed in 10 mM lithium cacodylate and 100 mM KCl buffer, pH 7.0. CD signal of ligand was subtracted.*

During the whole scanning wavelength, for **TMPyP4** with *PQS18-1* RNA, there was no significant CD signal observed between 320 – 650 nm. In the CD titration of **TMPyP4** with DNA, the DNA in the absence of **TMPyP4**, showed a positive peak at 295 nm and a negative one at ~240 nm, which had the same features as anti-parallel G-quadruplex structure. No unfolding was observed after adding **TMPyP4**, though, a significant signal appeared at 445 nm that increased in CD intensity with the addition of the compound. This was indicative of binding

between DNA G-quadruplex and **TMPyP4** and it altered the DNA conformation, which is consistent with previous studies of DNA G-quadruplex with **TMPyP4**.[293–295] On the contrary, there is only weak induced exciton splitting in RNA G-quadruplex titration. These results suggest the formation of dimers or high-order complexes, through groove binding or external stacking-binding mode, which is consistent with molecular dynamics studies.[144,296,297] The CD titrations between PQS18-1 DNA and RNA show that **TMPyP4** can unfold *PQS18-1* RNA G-quadruplex structure and interact with *PQS18-1* DNA G-quadruplex, however, **TMPyP4** seemed not to unfold *PQS18-1* DNA G-quadruplex structure.

From the CD titrations, it can be observed that **TMPyP4** unfolds parallel G-quadruplex RNA. The plot of the ellipticity at 264 nm against the concentration of **TMPyP4** (0 – 70 μM) showed a sigmoidal-shaped curve (***Figure 4.5***), indicating a cooperativity binding process,[258] this sigmoidal curve was fitted using the Hill equation (***Equation 8***), a Hill coefficient (n) of 4.1 which suggests the positive cooperativity ($n > 1$) of the binding of **TMPyP4** to *PQS18-1* RNA G-quadruplex. The half-denaturing concentration $[DC]_{50}$ was also identified to be 42 ± 0.5 μM.

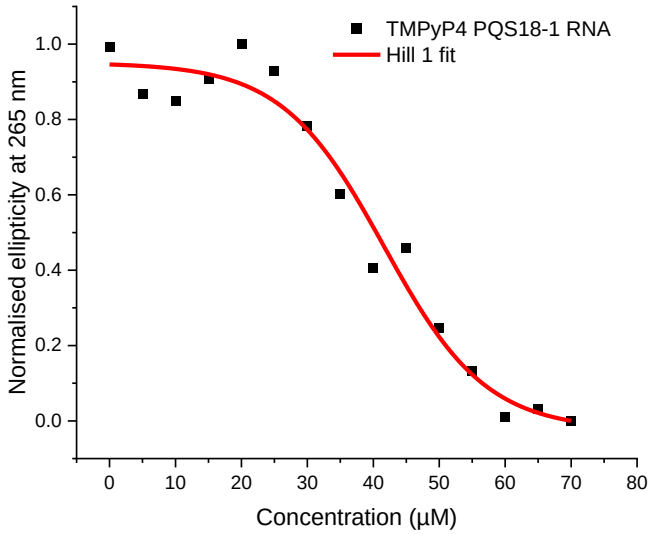


Figure 4.5 *Plot of molar ellipticity at 264 nm against the concentration of TMPyP4 and corresponding Hill fit.*

Job Plot experiments were performed to determine the stoichiometry of **TMPyP4** to RNA. A set of samples were prepared where the ratio of **TMPyP4** to RNA was between 1:9 and 9:1 the whole concentration of **TMPyP4** + RNA in each sample was 20 μM, then the CD spectra were measured for each sample between the wavelength 200 – 320 nm, and the set of samples of RNA but without ligand was measured, then the change in molar ellipticity at 264 nm was plotted against the ratio of RNA, two linear lines intersected to get the stoichiometry (***Figure 4.6***). The stoichiometry of **TMPyP4** to RNA is around 1.5:1, which is between 1:1 and 2:1, indicating there may be more than one binding site for **TMPyP4**. The Haider group identified the stoichiometry to be 2:1 for **TMPyP4** and RNA using ITC,[145] and the modelling indicated 2:1 binding so this result is broadly in-line with these. We also reanalyzed the linear curves by removing the point with a large error (***Appendix 10***), a range of stoichiometry was calculated between 2:1 – 1.4:1, and 2:1 was within the error.

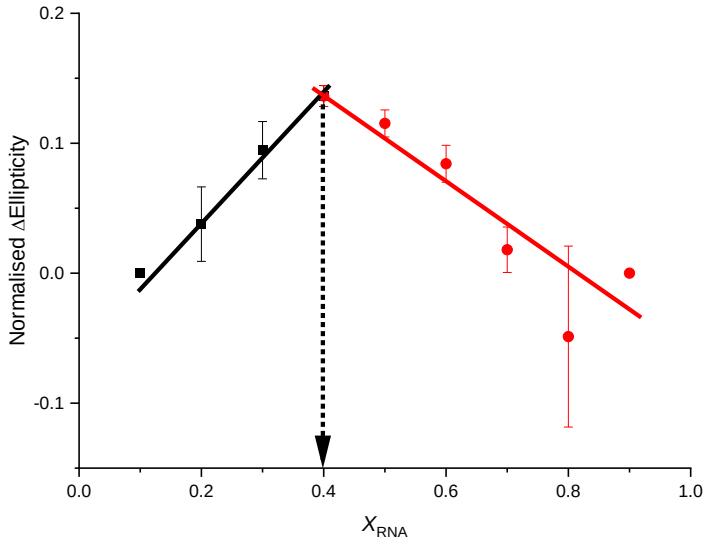


Figure 4.6 a), *Job plot of PQS18-1 RNA G-quadruplex and TMPyP4 in 10 mM lithium cacodylate and 100mM potassium chloride buffer at pH 7.0. b), reanalyzed Job Plot linear curves. The black and red symbols represent the points used for fitting the respective linear best fits to determine the intercept.*

4.3 UV Spectra of TMPyP4 and PQS18-1 RNA G-Quadruplex

During the CD titration, when increasing the concentration of **TMPyP4** up to 70 μM, aggregation and precipitation were observed, so we considered some CD signal losses might be due to aggregation, rather than unfolding and disruption of RNA G-quadruplex structure, therefore, during the CD titration, the UV absorbance spectra were also recorded (***Figure 4.7***) to investigate the behaviour of **TMPyP4** in solution.

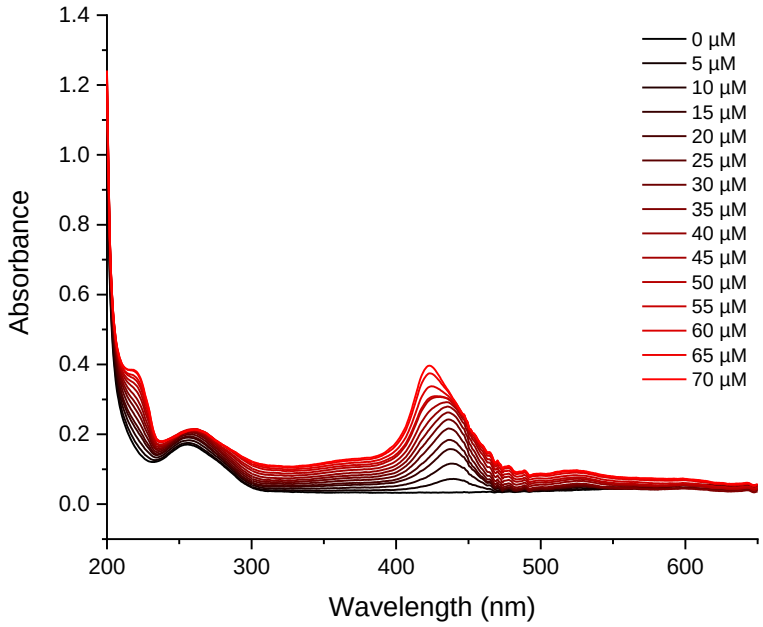


Figure 4.7 *Example UV titration spectra RNA PQS18-1 in the presence of TMPyP4. Experiments were performed at 10 μM RNA in 10 mM lithium cacodylate and 100 mM of potassium chloride buffer at pH 7.0, and 0 to 70 μM of TMPyP4.*

In a plot of the absorbance at 217 nm against increased concentration up to 70 μM (***Figure 4.8 a***), as can be seen, after 50 μM the data starts to plateau, this suggests that above this concentration, aggregation occurs. The plot of absorbance with a concentration of **TMPyP4** (0 - 50 μM) gave a linear correlation (***Figure 4.8 b***), which is consistent with the Beer-Lambert Law, thus suggesting there was no aggregation at a such concentration under the conditions examined.

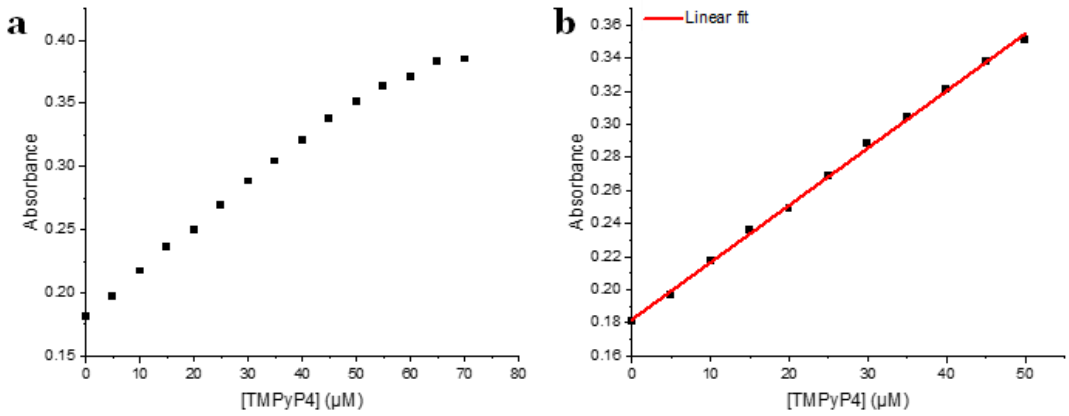


Figure 4.8 *Plot of UV absorbance of the addition of different concentrations of the ligand to the RNA G-quadruplex complex. a) A plot of UV absorbance at 217 nm against the concentration (0 – 70 μM) of TMPyP4. b) Linear plot of UV absorbance at 217 nm against the concentration (0 – 50 μM) of TMPyP4.*

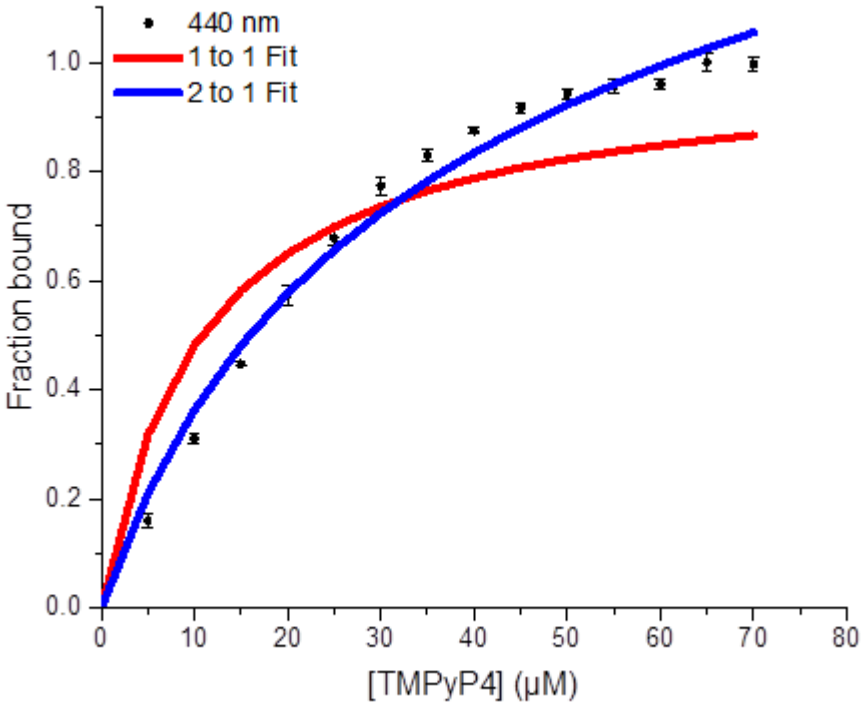
From the UV data, at 440 nm, hyperchromic and hypsochromic shifts were observed, indicating a strong effect could be detected at such a wavelength. The change in absorbance at 440 nm was converted into a fraction bound between 0 to 1, then a plot of the fraction bound of UV absorbance at 440 nm against concentration gave a binding curve. One to one and two to one binding models (Equation 9 and 10) were used to fit the curve (***Figure 4.9***). K_a is the equilibrium association constant, K_1 and K_2 are the association constants for the first and second binding sites, respectively. K_d is the reciprocal value of K_a.

Equation 9*: one to one binding mode*

$$\theta = \frac{\mathrm{Ka[ligand]}}{1 + \mathrm{Ka[ligand]}}$$

Equation 10*: two to one binding mode*

$$\theta = \frac{\mathrm{K1[ligand] + 2K1K2[ligand]\text{^}2}}{1 + \mathrm{K1[ligand] + 2K1K2[ligand]\text{^}2}}$$



.

Figure 4.9 *Graph shows the fraction bound with increasing concentration of TMPyP4 from the change in UV absorbance at 440 nm. Experiments were performed at 10 μM RNA in 10 mM lithium cacodylate and 100 mM of potassium chloride buffer at pH 7.0, and 0 to 70 μM of TMPyP4. Data is fitted with 1 to 1 and 2 to 1 binding models using Slover. Error bars show the standard deviation across two repeats.*

The dissociation constants were then determined. From the 1 to 1 binding model, K_d is 10 ± 1.2 μM, 2 to 1 binding model gives K_{d1} = 19 ± 0.7 μM and K_{d2} = 188 ± 19 μM. The 2 to 1 binding model may give a much better fit. This indicates that **TMPyP4** has a moderate binding affinity to the RNA G-quadruplex binding sites, potentially with a higher binding affinity for the first binding site and a weaker affinity for the second binding site. This is consistent with the stoichiometry and computational work by the Haider group.

4.4 CD melting

CD melting experiments were performed to assess the effect of **TMPyP4** on the stability of the *PQS18-1* RNA G-quadruplex structure. The melting curve of RNA

PQS18-1 showed two transitions, indicating there were two melting temperatures, one at 45°C and another at 73°C, this suggested that the *PQS18-1* RNA structure may have more than one conformation. When 10 μM of **TMPyP4** was added to RNA, the shape of the curve slightly changed while the change in melting temperature did not change (***Figure 4.10***).

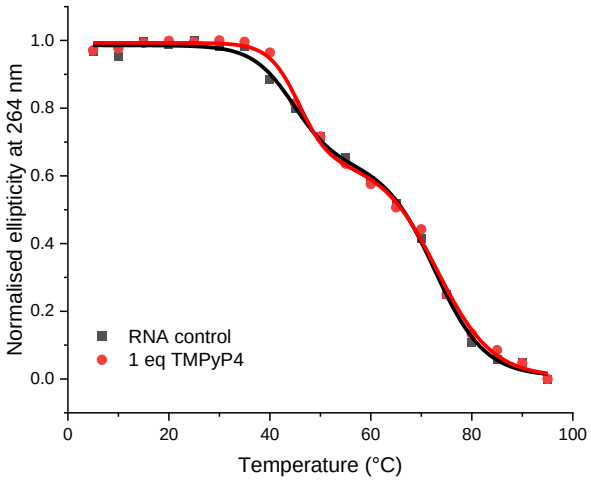


Figure 4.10 *CD melting of PQS18-1 RNA (10 μM) in the presence of 0 μM (black), 10 μM of TMPyP4 (red) in 10 mM lithium cacodylate and 100 mM KCl, pH 7.0. The plot was fitted using bisigmoidal curve fitting.*

Then we did the CD melting of RNA in the presence of 20 μM of **TMPyP4**, and the shape of the curve changed again, tending towards a single transition instead of two separate ones (***Figure 4.11)***, suggesting **TMPyP4** altered the structure of RNA G-quadruplex. The melting temperature was around 71°C, indicating the change in melting temperature of -2°C, which showed a slight destabilization effect on RNA structure. The shape of the graph could be interpreted to give a transition at the midpoint between the folded and unfolded, but the overall change in melting temperature across the two melting experiments is between -2 and -5°C.

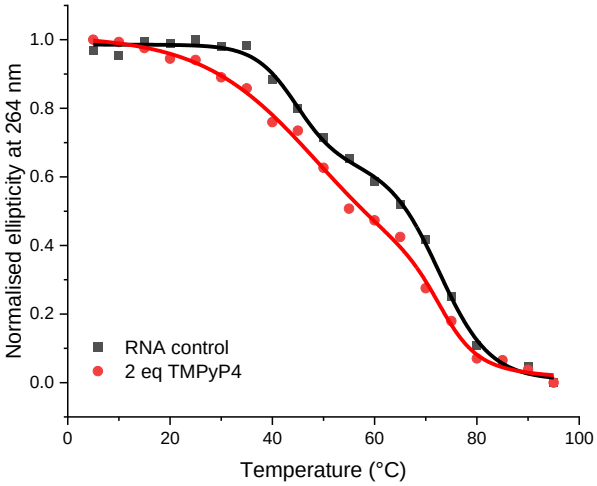


Figure 4.11 *CD melting of PQS18-1 RNA (10 μM) in the presence of 0 μM (black), 20 μM of TMPyP4 (red) in 10 mM lithium cacodylate and 100 mM potassium, pH 7.0. The plot was fitted using bisigmoidal curve fitting.*

When the concentration of **TMPyP4** was increased further, up to 50 μM, the change in the shape of the melting curve also increased again (***Figure 4.12***), also there were again two clear transitions at 37 and 70°C, which indicates ΔT_m values of -8 and -3°C.

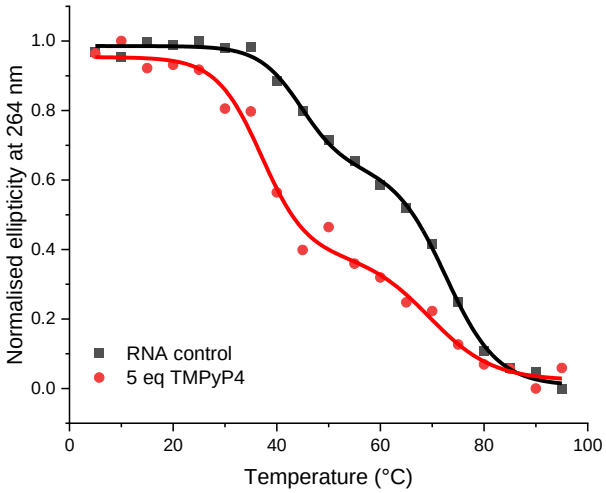


Figure 4.12 *CD melting of PQS18-1 RNA (10 μM) in the presence of 0 μM (black), 50 μM of TMPyP4 (red) in 10 mM lithium cacodylate and 100 mM potassium, pH 7.0. The plot was fitted using bisigmoidal curve fitting.*

The CD melting results showed a decrease in the melting temperature of RNA G-quadruplex, indicating **TMPyP4** causes the destabilization of *PQS18-1* RNA G-quadruplex structure. All the melting curves are biphasic which indicates more than one complex is formed. The possible reason could be there are potentially two species (parallel and antiparallel species). As in the CD titration experiments, an initial shift from parallel towards antiparallel can be observed before unfolding. **TMPyP4** changed the shape of the RNA melting curve, consistent with **TMPyP4** altering the structure of RNA G-quadruplex, and the RNA-ligand complex being less stable than the free-RNA.

Meanwhile, we also did a study on the corresponding annealing analysis in the absence and presence of **TMPyP4**. ***Figure 4.13*** shows the CD melting and annealing for RNA *PQS18-1* in the absence of **TMPyP4.**The thermal CD spectrum, the melting and annealing process only showed a positive peak at ~264 nm, which was consistent with parallel G-quadruplex structure. However, in the plots of ellipticity at 264 nm against temperature, both thermal curves showed two clear transitions, indicating more than one species formed during heating and annealing processes.

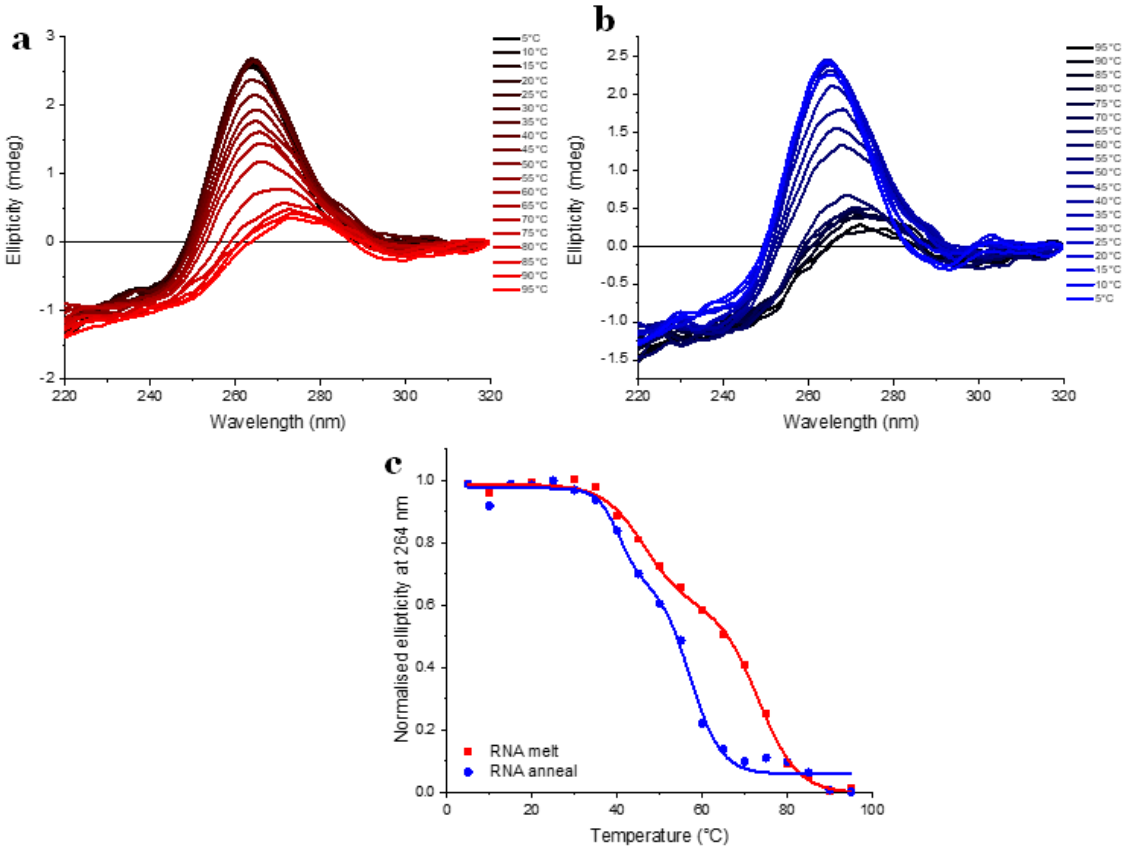


Figure 4.13 *Example CD melting (a) and annealing (b) of RNA PQS18-1 and thermal curves (c). Experiments were performed at 10 μM RNA in 10 mM lithium cacodylate and 100 mM potassium chloride buffer, pH 7.0.*

The CD melting and annealing of RNA *PQS18-1* in the presence of 20 μM **TMPyP4** (***Figure 4.14***), during the melting process, showed two transitions that were similar to the melting curve in absence of **TMPyP4** shown in ***Figure 4.14*** c, however, during the cooling process, the annealing curve changed and only one transition was observed, indicating **TMPyP4** interacted with RNA and altered the RNA structure, furthermore, at the end of annealing, dark purple precipitation was observed, suggesting ligand-RNA aggregation formed.

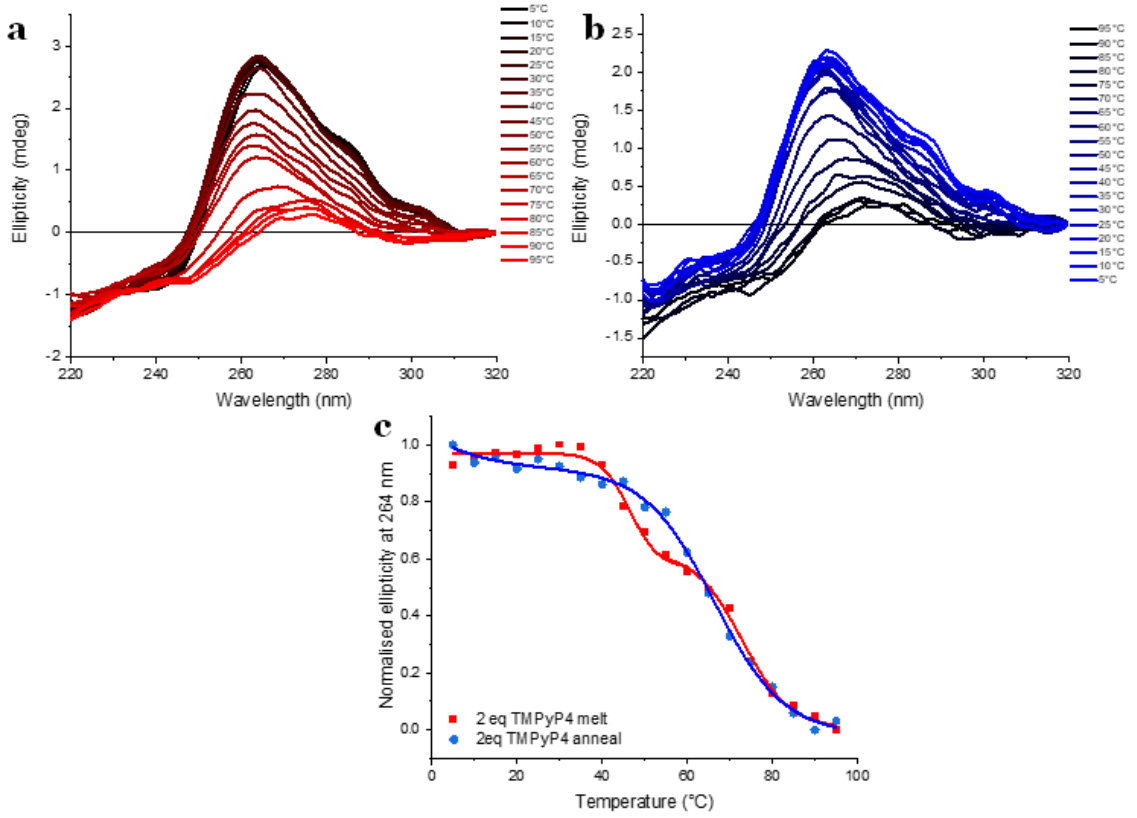


Figure 4.14 *Example CD melting (a) and annealing (b) of RNA PQS18-1 and thermal curves (c). Experiments were performed at 10 μM RNA in the presence of 20 μM TMPyP4 in 10 mM lithium cacodylate and 100 mM potassium chloride buffer, pH 7.0.*

Then the CD melting and annealing of RNA with 50 μM **TMPyP4** were performed (***Figure 4.15***). In the annealing spectra, there was a decrease in CD signal at 295 nm, and unlike the other two annealing of RNA control and RNA with 20 μM 2 **TMPyP4** at which a clear decrease was observed at 264 nm, the annealing of RNA with 50 μM ligand did not show a significant decrease at this wavelength. Then in the thermal curves, the melting curve showed clear two transitions, but the annealing curve still showed one transition, and significantly changed the shape, there was a significant right shift in annealing. Interestingly there is a clear increase in the annealing temperature of the RNA with each progressive addition of **TMPyP4**. For example, the *PQS18-1* RNA in the absence of **TMPyP4**, the annealing temperature (T_{a1} = 40°C, T_{a2} = 57°C) was lower than the melting temperature (T_{m1} = 46°C, T_{m2} = 73°C); then *PQS18-1* RNA with 20 μM **TMPyP4**,

the T_a and T_m were roughly the same. However, in the experiment where we examined *PQS18-1* RNA with 50 μM of **TMPyP4**, the annealing temperature was significantly more than the melting temperature (T_{m1} = 34°C, T_{m2} = 69°C, T_a = 90°C). During the annealing process, re-folding occurred when the temperature decreased to around 58°C, but the signal then started to decrease again (***Figure 4.15 d***). There was also precipitation observed at the end of the annealing, this indicates the annealing experiments were complicated by precipitation and ligand-RNA aggregation processes. The CD melting and annealing curves do not look very clean, because there is high signal to noise. The *PQS18-1* RNA G-quadruplex is a 2-tetrad G-quadruplex structure so it has not as much signal, we had tried to use higher concentration of RNA for the experiments, but the solubility of the **TMPyP4**-RNA complex is quite low, therefore, we proceeded to continue using 10 μM RNA, in spite of the signal to noise issues.

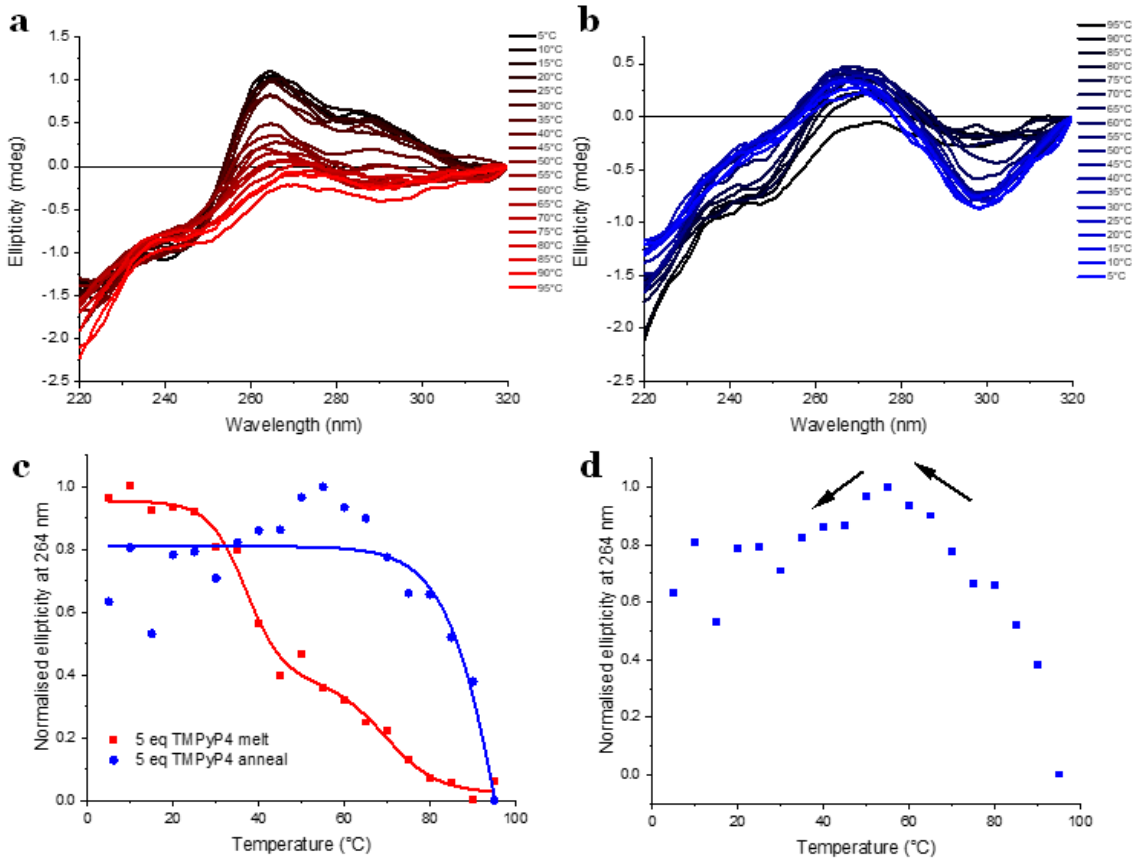


Figure 4.15 *Example (a) CD melting and (b) annealing of RNA PQS18-1, (c) thermal curves, and (d) annealing curve. Experiments were performed at 10 μM RNA in the presence of 50 μM TMPyP4 in 10 mM lithium cacodylate and 100 mM potassium chloride buffer, pH 7.0.*

4.5 FRET Titration Results of TMPyP4 with *PQS18-1* RNA G-Quadruplex

To further investigate the change in structure observed in CD experiments, FRET titration was performed to explore the apparent unfolding, using FRET labelled RNA *PQS18-1* with FAM at 5′-end and TAMRA at 3′-end. In this design, the fluorophores would be close together when the G-quadruplex folded, and energy will be transferred from FAM to TAMRA, then the FRET would occur. The FRET labelled RNA *PQS18-1* was excited at 490 nm and the emission spectrum was recorded from 500 to 650 nm, the fluorescence of FAM and TAMRA can be detected at 515 and 585 nm respectively. Unlike the FRET melting experiments, FRET

titration allows us to observe the folding in real-time using a fluorimeter. In this experiment with (***Figure 4.16***) when the RNA G-quadruplex was folded as expected the signals from TAMRA were high (at 585 nm) while FAM (515 nm) was low, consistent with the fluorophores being in close proximity and FRET occurring. Upon the addition of 0 – 0.5 μM (0 – 5 equivalents) of **TMPyP4**, there was no significant decrease in the emission of FAM at 515 nm. On the contrary, the emission of TAMRA was observed to decrease at 585 nm to the folded *PQS18-1* RNA (0.1 μM), indicating TAMRA emission is decreasing and could be indicative of less FRET occurring and thus also the unfolding of RNA G-quadruplex structure. However, compared to the conditions of CD experiments, the conditions of FRET titrations are quite different, as much lower concentration of RNA and ligand were used, the concentration differences are important, therefore, the fraction bound could be very different.

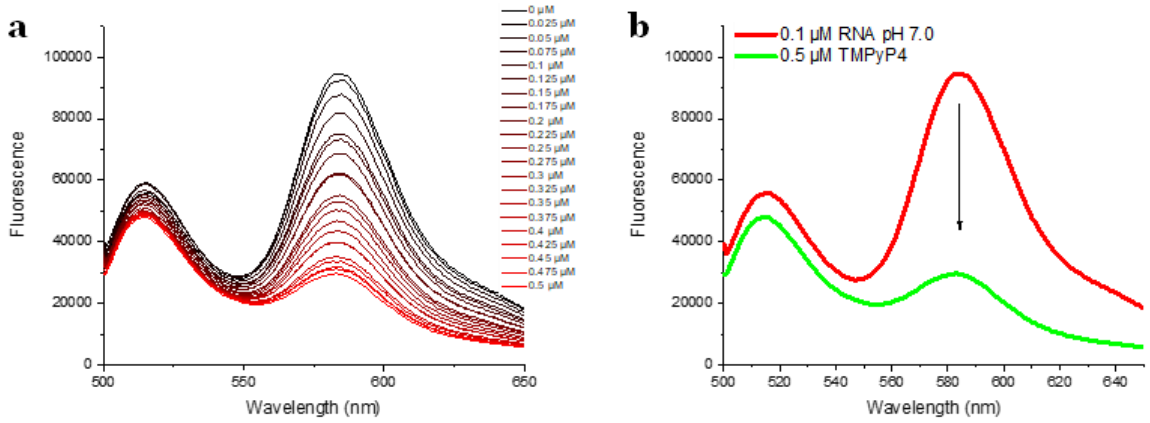


Figure 4.16 *a) Example fluorescence spectra of 0.1 μM RNA PQS18-1 labelled with FAM and TAMRA in the presence of TMPyP4 RNA in 10 mM lithium cacodylate and 100 mM of potassium chloride buffer, pH 7.0. b) fluorescence of RNA in the absence of TMPyP4 (red) and the presence of 0.5 μM of TMPyP4 (green).*

The change in fluorescence can be used to calculate the relative FRET efficiency using ***Equation 3***, where I_A is the fluorescence intensity of acceptor fluorophore TAMRA and I_D is the fluorescence intensity of donor fluorophore FAM, when RNA is fully unfolded, the plot of $E_{rel.}$ gives a decreasing curve indicating the unfolding process.

$$\textbf{\textit{Equation 3}}\ E_{rel.} = \frac{I_A}{I_D + I_A}$$

The FRET efficiency was fitted using the Hill equation (***Equation* 8**) to determine $[DC]_{50}$ of **TMPyP4** at which half RNA is unfolded and the cooperativity of the interaction (***Figure 4.17***). The error bars are larger at higher concentrations of ligand. This could be as the concentration of both RNA and ligand were quite low during FRET titration experiments, thus the error of pipetting would be greater than normal, or the RNA might have denatured, this should be checked in future studies.

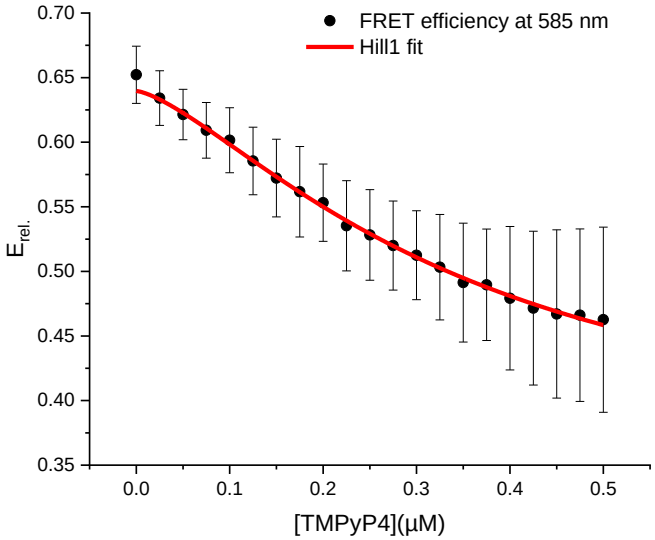


Figure 4.17 *FRET efficiency showing unfolding of 0.1 μM RNA PQS18-1 G-quadruplex with increasing concentration of TMPyP4 (0 – 0.5 μM) in 10 mM lithium cacodylate and 100 mM potassium chloride at pH 7.0. Also corresponding to Hill fit. Error bars show the standard deviation across 3 repeats.*

Using the fitting method, a value for $[DC]_{50}$ of 0.35 μM was determined, and low positive cooperativity (n >1) of n = 1.42 was calculated. These values are roughly in-line with those observed in the CD titration. Thus, the concentration-dependent decrease in E_{FRET} also happened with the addition of **TMPyP4**, which is consistent with the unfolding of the *PQS18-1* RNA G-quadruplex structure with **TMPyP4** but monitored by a complementary method.

4.6 Conclusions

According to the biophysical data presented in this project, **TMPyP4** showed the ability to unfold RNA *PQS18-1* G-quadruplex structure. The computational work performed by the Haider group showed that, in addition to having stabilizing properties for some G-quadruplexes, the ligand could also act as an unfolding agent on G-quadruplex, depending on the type of G-quadruplex structure.[145] As there is not much information on the dynamic structure of the ligand binding mode, biophysical characterization was important. The Job Plot indicated binding to more than one site, which is consistent with computational experiments showing that there are high- and low-affinity binding sites at the top-face and major groove-bound sites, respectively.[145] This "two independent sites" model also showed in ITC and UV titration experiments performed by the Haider group.[145] The concentration-dependent biophysical difference indicates the formation of a ligand-RNA intermediate during the titration process. The CD melting and annealing results supported the titration that **TMPyP4** interacts with RNA G-quadruplex and could form a ligand-RNA complex, and with the addition of **TMPyP4**, annealing temperatures were significantly increased. All the biophysical experiments were performed at a large range of concentrations of *PQS18-1* RNA from 0.1 μM (in the FRET titration) to 10 μM (in the CD titration and melting), indicating the folding effects can occur at a wide range of RNA concentrations and there is no limit with high concentrations of RNA.

Although from the previous study, the binding mechanism of **TMPyP4** to G-quadruplex is controversial, some studies showed it stabilizes G-quadruplex structure under high potassium concentration,[286,288] while some others suggest that **TMPyP4** destabilizes G-quadruplex,[281,284,298] and a few reports that **TMPyP4** can alter the topology of G-quadruplex.[299] From our biophysical experiments, both destabilizing and alteration G-quadruplex structure exist, which depends on the concentration ratio and temperature. According to the results, it is difficult to predict the interaction between G-quadruplex and ligands, therefore, the approaches for ligands design cannot only be based on the structure but dynamic interaction should also be considered. This understanding of the

unfolding effect of **TMPyP4** on RNA G-quadruplex will be helpful to further study the i-motif or G-quadruplex ligands binding process.

Chapter 5

Study of Repetitive GC-rich Sequences from Promoter Region of *ALOX5*

5.1 Introduction

As discussed in **Chapter 1**, DNA can adopt multiple topologies, for example, G-quadruplexes and i-motif structures, and they also have potential biological functions. G-quadruplexes have been identified as involved in the regulation of gene expression.[300–304] The structure of complementary C-rich sequences can be more complicated than G-quadruplex forming sequences. It has been suggested the equilibrium in the solution of C-rich sequence exists between single strand, i-motif, duplex, and hairpin, this has indicated that i-motif structure could be relevant to the regulation of transcription.[69,248,305] Many ligands bind to the i-motif and then affect gene expression, for example, the Hurley group identified that the ligands IMC-48 and IMC-76 affect the *BCL2* expression,[248] suggesting the i-motif is not just a special structure and can only be used in biotechnological applications,[306,307] but can also alter biological effects by targeting molecules or proteins. The Christ group also showed the relevance of this structure through their experiments using an i-motif specific antibody iMab. They used this antibody to visualize the presence of i-motifs *in celluo* in the nuclei of human cells and showed that the formation of i-motifs in cells is both pH and cell-type dependent.[308] In addition, the Feng group identified the existence of i-motif structures in the testis cell of silkworms,[309] it was the first time that the *in celluo* i-motif structures were observed in invertebrate cells. The *in celluo* i-motifs found exist in both vertebrate and invertebrate cells, suggesting the potential significance of i-motif as a modulatory mechanism in gene transcription across different organisms.

I have a great interest in leukaemia; therefore, I was interested in studying a specific i-motif sequence from the promoter region related to leukaemia. In a previous study on leukaemia, other DNA structures were used as a target, for example, the G-quadruplex structure is widely used and well-studied in studying as a target for leukaemia chemotherapy.[310] Many G-quadruplex ligands (e.g. TMPyP4,[98] ATTP-253,[311] telomestatin[312] and SYUIQ-5[313]) were used to examine their effects on leukaemia cell growth, and most of the tests identified that leukaemia cell growth was inhibited when G-quadruplex structure stabilized by

ligands *in vitro* and *in vivo*.[311,312,314] From the Ohyashiki group, G-quadruplex stabilized by telomestatin can cause apoptosis in leukaemia cells in four acute myeloid leukaemia (AML) patients.[315] Furthermore, other DNA secondary structures were also studied as a target for leukaemia, the work from the Yoshikawa group, for example, selected an actual DNA sequence from the promoter region of *Stat5* at which murine leukaemia retrovirus (MLV) commonly integrates,[316] and from the promoter region of *c-MYC* that is one of the common integration sites[317] for the study. In *in vitro* assays, both sequences adopted cruciform DNA secondary structures and successfully integrated MLV into the target DNA.[318] As discussed in **Chapter 1**, the i-motif also plays a role in gene expression, thus i-motif sequence could also be a potential target for cancer therapy. Herein, the project aims to identify potential i-motif targets that could be characterized by their potential ability for developing chemotherapy and diagnostics for leukaemia.

In this project, we selected a series of i-motif sequences from the human genome, investigating the stability and folding potential of different DNA structures. Several biophysical methods were used to characterize the oligonucleotides, including CD,[160] UV,[319] and TDS.[167] This study would enable us to observe which sequences can fold under neutral conditions and are more likely to have potential biological functions.

5.2 Exploring GC-Rich Sequences from the Human Genome Involved in Leukaemia using Bioinformatics

The Waller group previously found 637 predicted i-motif sequences that overlap gene promoters.[54] To explore leukaemia-related i-motif sequences, I used this known list of i-motifs that have already been found and narrowed these down based on whether an i-motif sequence is a cancer or leukaemia related. Basic Local Alignment Search Tool (BLAST) authorized by National Centre for Biotechnology

Information (NCBI) was the main search engine and SnapGene was the analysis software used in the bioinformatics study.

Firstly, we investigated how many sequences in the list can overlap or are partially similar to the Sp1 (specificity protein 1) transcription factor, because Sp1 is one of the best-characterized transcription factors of a family including Sp2, SP3, and SP4.[320] It binds GC-rich sequences which could be involved in expression and regulation in numbers of genes.[321] Sp1 and other transcription factors of this family have been identified as over-expressed in several cancers In addition, Sp1 binding to gene promoters could be a target for DNA binding ligands that show binding preference to GC-rich region in DNA.[322] Therefore, when a sequence from the i-motif list has a similar sequence to the Sp1 binding site (which has the consensus sequence of 5'-(G/T)GGGCGG(G/A)(G/A)(C/T)-3')), it could be a potential target for cancer chemotherapy or diagnostics. To find the similarity between i-motifs and the DNA sequence which Sp1 binds, NCBI and SnapGene were used to compare i-motif sequences and the DNA sequence which Sp1 binds sequence identity was also calculated,[323] which refers to the percentage of two DNA sequences that have the same residues at the same position in the alignment. All the 637 sequences were checked to determine any overlap with the Sp1 gene, the result is presented in ***Table 5.1***. The alignment identity results of all 637 i-motif sequences with Sp1 transcription factor binding sites are presented in ***Appendix 25***.

Table 5.1 *Sequence alignment results of 637 i-motifs overlapping promoter regions against Sp1 transcription factor binding site.*

Identity	>80%	70 – 80%	65 – 70%	<65%
Number of sequences	35	119	137	346
percentage	5.94%	18.68%	21.51%	54.32%

From the results, we selected the sequences with higher than 65% identity for the next study as these sequences showed a moderate similarity to the target Sp1 gene. Then using NCBI search engine to find the main biological functions of these genes corresponding to the GC-rich sequences. 39 genes have been identified related to

cancer, and among these genes, nine were found involved in leukaemia (***Table 5.2***).

Table 5.2 *Leukaemia-related GC-rich sequences from the i-motif list that overlap Sp1 transcription factor from bioinformatics study.*

Gene	Chromosome	Transcription start site	Sequence 5′ → 3′	Identity (%)
ALOX5	10	45374175	GGG-GGC-GGG-GGC-GGG-GGC-GGG-GGC-GGG-GGC-GGG-GG	84
HOXB9	17	48627356	CCC-CCC-CAG-CCC-CCC-ACC-CTC-CCA-CCC-CCA-CCC-CC	81
CCND1	11	69641086	CCC-CCT-CCC-CCT-GCG-CCC-GCC-CCC-GCC-CCC-C	80
SATB1-AS1	3	18445023	GGG-GGG-GGG-AGA-AGG-GGG-AGG-GGG-CGG-CGG-CGG-GGG-CGG-GAG-GGG-G	68
SH3GL1	19	4401547	CCC-CCG-TCC-CTG-CCC-CCG-TCC-CTG-CCC-CCG-TCC-CCG-CCC-CGT-CCC-CGC-CCC-CGC-CCC-CGC-CCC-CGT-CCC-CGC-CCC-C	68
PEG10	7	94656324	CCC-CCA-CCC-CCA-TCC-CCC-ATC-CCC-GCA-CCC-CCC-TCC-ACC-CCC	66
PUS1	12	131929199	CCC-CCA-CCC-CCT-AGT-CTC-CGC-CCC-CAT-CAC-CCC-C	66
TLE1	9	81690305	GGG-GGG-CGC-GAG-CGC-GGA-GCC-GGC-GGG-GGA-CGT-GAG-CCC-GGG-AGG-TTG-GGG-GGT-GGG-GGG-GGG-G	65
RSPO3	6	127118603	CCC-CCG-AGC-TCC-CCC-CCG-CCC-CCC-ACA-GCC-ATC-CCC-CCT-CCT-CTC-CCC-CTC-CCT-GCC-CTC-CTC-TCC-CTC-CCC-CC	65

Among these nine leukemia-related sequences, the sequence of *ALOX5* only contains GC bases, which is special, as this tandem repeat sequence is similar to the already identified i-motif sequences *DAP* and *MSMO1*, therefore currently we are focused on *ALOX5*. ***Figure 5.1*** shows a flow chart overview of the pathway toward this target.

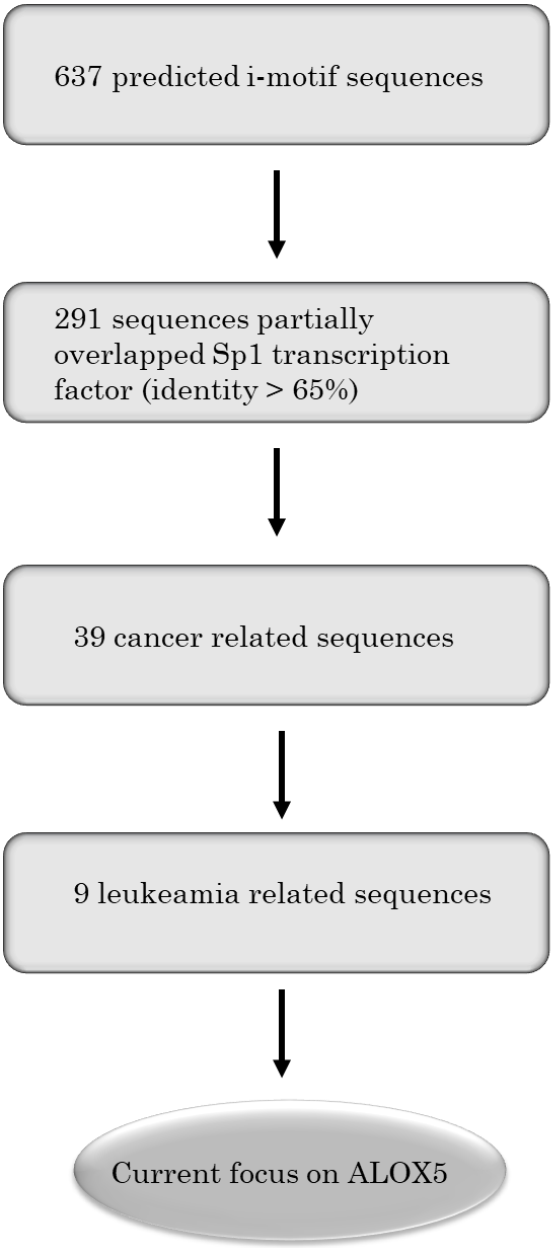


Figure 5.1 *The flow chart shows the process of selecting our interested i-motif ALOX5.*

The reason that *ALOX5* was firstly selected as it showed the highest identity (84.38%) with Sp1 transcription factor binding site among the top nine leukaemia-related i-motif sequences, it also has a similar sequence to the previously described

i-motif forming sequences *DAP* and *MSMO1* (***Table 5.3***), they shared the same repeat motif sequence. It was interesting because these sequences are common across the genome, aside from *DAP* and *MSMO1*, they are present in a lot of places. This confirms that they overlap with the sequence for a master regulator protein or transcription factor, such as Sp1. We hypothesize that these are therefore critical regions where the i-motif structure may play an important role in transcription factors binding the DNA and thus biological function. From the i-motifs examined previously by the Waller group,[54] we already knew about *DAP* and *MSMO1* that have tandem repeat sequences of CCCCCG, but we were curious how prevalent sequences of this type are across the genome and whether the length of these is important for their properties, especially if they are at points of regulation by Sp1. Overall, we looked at sequence lengths for this repeat between three and seven tracts of cytosines (***Table 5.3***). The sequence from *MSMO1* has four stretches of cytosines, *DAP* has five, *ALOX5* has six. We also included sequences with three C-tracts (found in the promoter region of *MADD*, *MAP* kinase-activating death domain protein) and also seven C-tracts (*ZNF480*, zinc finger protein 480), which were not previously characterized by the Waller group. BLAST searches of the human genome did not find sequences longer than seven repeats in length.

Table 5.3 *Tandem repeats C-rich sequences and number of cytosine tract.*

Sequence name	Sequence 5′ – 3′	Number of C-tract
MADD-C	$(CCCCCG)_2CCCCC$	3
MSMO1-C	$(CCCCCG)_3CCCCC$	4
DAP-C	$(CCCCCG)_4CCCCC$	5
ALOX5-C	$(CCCCCG)_5CCCCC$	6
ZNF480-C	$(CCCCCG)_6CCCCC$	7

Mutations in the 5-lipoxygenase (*ALOX5*) promoter region contribute to a reduced response to antileukotriene drugs used in the treatment of asthma and may also be involved in atherosclerosis and cancers including leukaemia.[324] The Li group found the *ALOX5* gene to be a regulator for leukaemia stem cells in BCR-ABL-induced chronic myeloid leukaemia (CML). They demonstrated, using a mouse

model, that in the absence of *ALOX5*, *BCR-ABL* failed to induce CML in mice.[325] ALOX5 protein is also a key enzyme in the biosynthesis of allergic reaction responses, mediators in inflammation, and leukotrienes.[326] *ALOX5* has also been implicated to contribute towards atherosclerosis.[327] The *ALOX5* gene promoter contains 8 GC boxes as the major transcription initiation site, five of these GC boxes are tandem repeats.[328] The Radmark group showed that gene expression of *ALOX5* was reduced by the small molecule mithramycin, which inhibits the binding of Sp1 to GC boxes.[329] This indicates that the GC boxes are important for the transcription of the *ALOX5* gene. In the *ALOX5* promoter, natural mutations are related to the number of GC boxes in tandem. For example, in HeLa cells addition and deletion of GC boxes in tandem resulted in reduced reporter gene activity, while in Drosophila SL2 cells, addition and deletion of GC boxes resulted in decreased or increased activity respectively.[330] A novel Sp1-binding site was observed close to the major transcription start site, it was also indicated that the GC-rich core region including Sp1 and Egr-1 sites is significant for *ALOX5* promoter activity.[328] Sp1 binds the G-rich region and is well known that Sp1 binds G-quadruplex forming sequences.[329] Egr-1 and Sp1 interact functionally with the 5-Lipoxygenase promoter and its naturally occurring mutants. Although the transcription factor Egr-1 binds the C-rich region,[329,331,332] Egr-1 protein may be binding i-motif structure as opposed to the C-rich sequence alone. Or it may be possible to find ligands that disrupt this binding process, similar to mithramycin. To investigate this fully, we must first characterize the repeat sequences to understand the potential structural consequences of deletion events on the number of C-tracts.

5.3 Study of the C-Rich Repeat Sequences $(CCCCCG)_n$

5.3.1 Circular Dichroism of C-Rich Sequences

To give a broad overview of the stability of the different length C-rich sequences $(CCCCCG)_n$ we first used CD spectroscopy to determine their folded state and their transitional pH. For all experiments, oligonucleotides were diluted in a buffer containing 10 mM sodium cacodylate and 100 mM sodium chloride at the

pH as detailed. These conditions were to enable us to compare directly with the sequences which had been previously studied.[54] Then the work continued to characterize the different numbers of C-tracts. *MADD* with three C-tracts (n = **3**), *MSMO1* has four C-tracts (n = **4**) and *DAP* has five C-tracts (n = **5**), *ALOX5* has six C-tracts (n = **6**) and a longer sequence *ZNF480* with seven C-tracts (n = **7**). pH_T values were measured by CD spectra, pH range was from pH 4.0 – 9.0 because, from the previous study, the pH stability of i-motif depends on the number of C-tracts, n = **6** showed a positive peak at 288 nm and a negative peak at 264 nm at pH 4.0, indicating it formed i-motif structure. With increased pH, the folded n = **6** structure started to unfold, and it reached a plateau at pH 8.5. The plot of ellipticity at 288 nm against pH gave a bisigmoidal curve (***Figure 5.2***), there were two transitions observed, suggesting n = **6** has two transitional pHs due to potential different i-motif populations, therefore, in this case, the curve was fitted with bi-sigmoidal fitting. In the literature, most pHs changes have been single transitions, but the Smith group also used the bisigmoidal fitting method to analyze the CD data of an i-motif forming sequence from *c-MYC* in the presence of a ligand.[225]

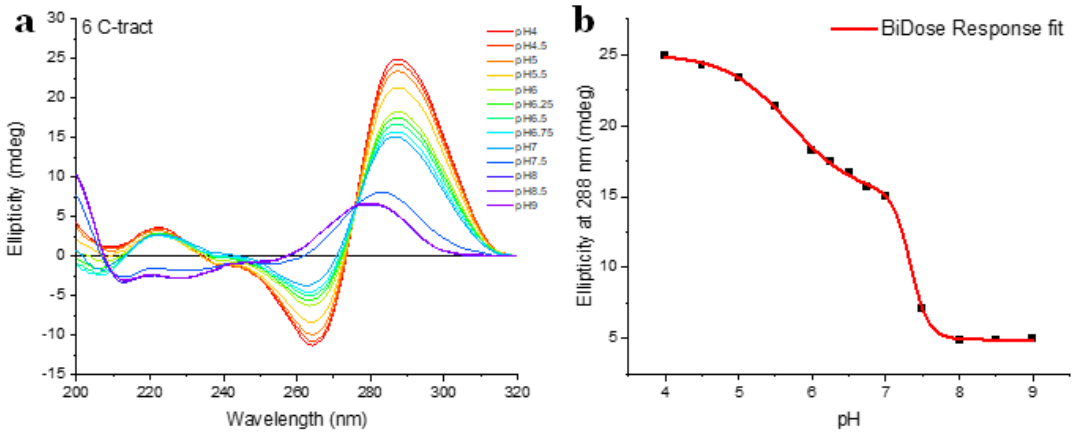


Figure 5.2 *a) CD spectrum of n = 6 (ALOX5) at pH range 4.0 to 9.0. b) Plot of molar ellipticity at 288 nm with pH for n = 6 C-rich sequence and corresponding Bisigmoidal fitting. [DNA] was 10 μM, and the sample was prepared in 10 mM sodium cacodylate and 100 mM sodium chloride buffer.*

Similar to n = **6**, two transitions were also observed in n = **3** and **7** (***Figure 5.3 a, d***), all these three CCCCCG repeat sequences showed two pH_Ts, however, the previous study of *DAP* and *MSMO1* showed only one pH_T using single sigmoidal curve fitting, so we were curious whether the curves could be bisigmoidal fitted as they have similar sequences, therefore, the CD data of n = **4** and n = **5** were reanalyzed (***Figure 5.3 b, c***), using bisigmoidal fitted the data set better than Dose Response which can be seen both sequences had two transitions, indicating their sequences can adopt more than one i-motif topology. All CD spectra of the selected C-rich sequences in different pH conditions and bisigmoidal fitted data are presented in **Appendix 11**.

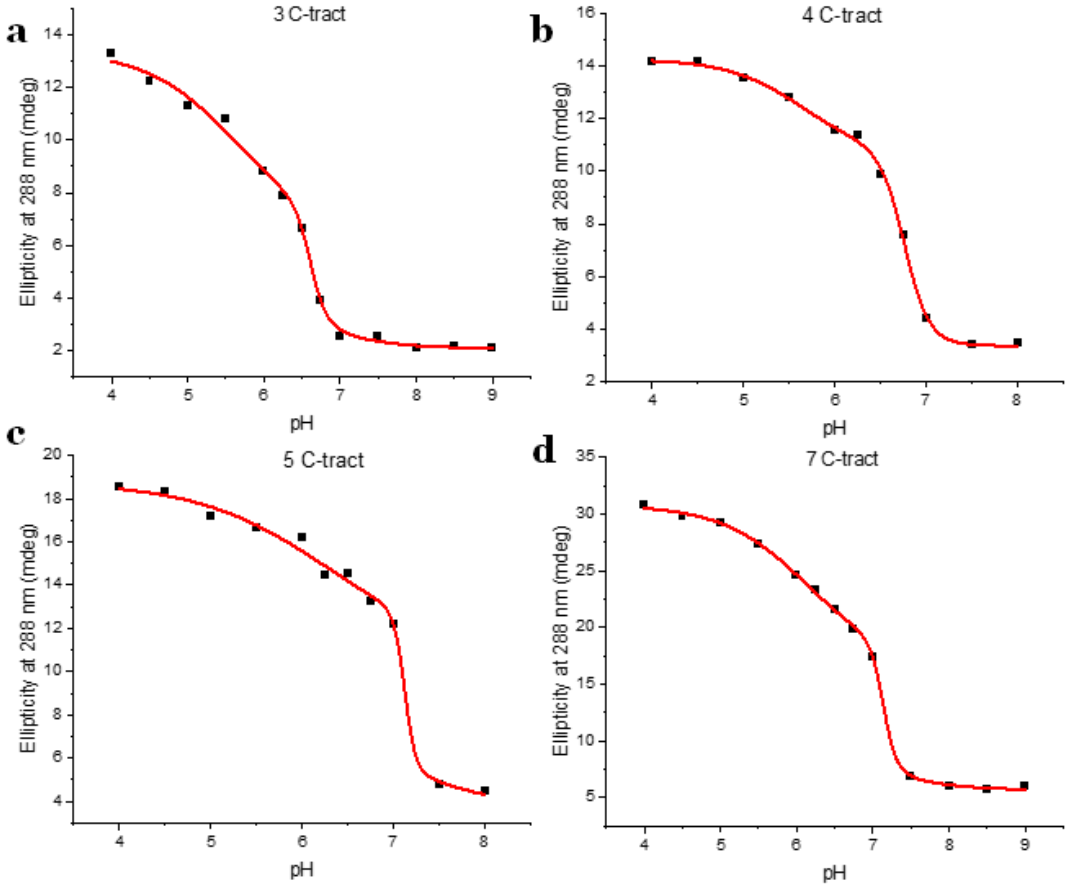


Figure 5.3 *Plot of molar ellipticity at 288 nm with pH for C-rich sequence and corresponding Bisigmoidal fitting.*

From the results, all these C-rich sequences have two transitional pHs (***Table 5.4***), suggesting in some environments they could fold multi-conformation DNA structures. For example, n = **3** is a very short C-rich sequence that may be difficult to self-fold intramolecular i-motif and possibly be more likely to form bimolecular i-motif, therefore, the DNA sample could be mixed with both species; in contrast, the number of C-tract of **6** and **7** have a long repetitive sequence, so it may form a huge i-motif with long loops or two small i-motif "beads on a string", therefore, different i-motif structures may have different features and result in more than one transitional pH. However, from CD, we cannot identify what the different topologies are, therefore, further experiments need to be performed, such as using X-ray crystallography to get crystal structures of these sequences. Then we plotted the relationship between transitional pH and number of C-tract (***Figure 5.4***).

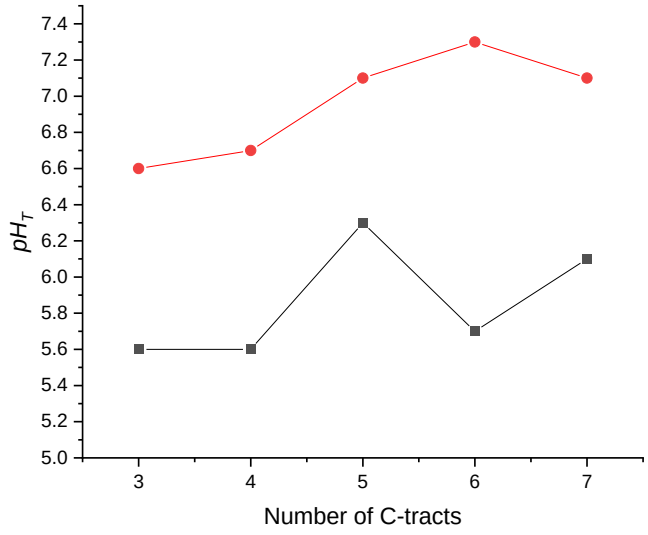


Figure 5.4 *Relationship between transitional pHs and number of C-tract. The black square presents the first* pH_T*, red one shows the second* pH_T*.*

From the result, it suggests with increasing numbers of C-tracts, the transitional pH also increased, indicating long GC-repeat i-motifs are more stable at higher pHs. However, the maximum pH_T reaches a maximum around pH 7.0 after n = **6** a plateau in pH_T was observed. This is supported by the literature from the Waller group that increasing the number of C-tracts raises the transitional pH up to pH 7.4, but not much above this pH.[54] The Burrows group has identified that the

number of cytosines in a single tract (i.e. poly dC) increases the stability of the i-motif, with increasing length of sequence, up to 30 cytosine bases, however beyond 30 was found to cause destabilization.[333] This study found dCn strands of length 15, 19, 23, and 27 nucleotides (i.e., 4n–1) to have pH_T values >7.2 and thermal stabilities >37 °C at pH 7.0.). Nevertheless, many of these C-rich sequences were shown to form i-motif structures at neutral conditions and even the less stable ones still had pH_T values above pH 6.5.[333]

CD melting experiments were then performed. Two sets of DNA samples were prepared, one set was at pH 5.5 (***Figure 5.5 a***) at which all the selected sequences should form an i-motif, and another was at pH 7.0 (***Figure 5.5 b***) to observe whether these sequences could be stable at neutral pH,[54] the melting temperature results are shown in ***Table 5.4***. The CD melting graphs of c-rich sequences at pH 5.5 and pH 7.0 are presented in **Appendix 12** and **Appendix 13**, respectively.

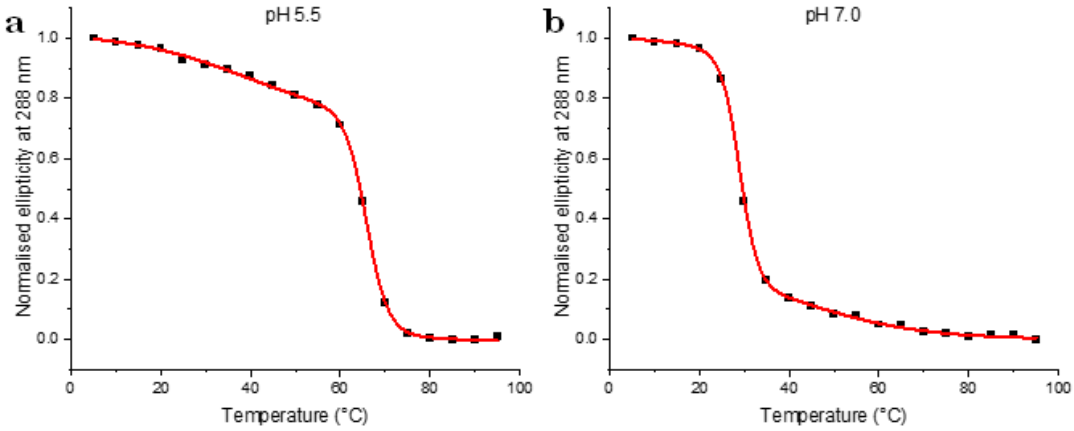


Figure 5.5 *Example of CD melting of C-rich sequence. a), CD melting of n= 6 sequence at pH 5.5 and corresponding bisigmoidal curve fitting. b), CD melting of n= 6 sequence at pH 7.0 and corresponding sigmoidal curve fitting.*

Table 5.4 *pH and melting temperature of GC-repeat i-motif sequences measured by CD, all experiments were performed in 10 mM sodium cacodylate and 100 mM sodium chloride buffer, pH range 4.0 – 9.0.* T_m*,* pH_T*, and error values were calculated using R-square values from Sigmoidal fitting.*

Name	Bases	C-tract	T_m(°C) pH 5.5	T_m(°C) pH 7.0	pH_{T1}	pH_{T2}
MADD-C	18	3	55.1±0.4	52.1±0.8	5.6±0.3	6.6±0.1
MSMO1-C	23	4	53.6±7.6/74.5±0.2	15.7±0.2	5.6±0.4	6.7±0.0
DAP-C	29	5	66.4±8.7/79.7±0.2	25.7±0.1	6.3±0.1	7.1±0.2
ALOX5-C	35	6	38.8±2.9/65.8±0.1	29.6±0.4	5.7±0.1	7.3±0.0
SNF480-C	42	7	61.8±4.6/76.3±0.2	33.1±1.7/60.5±0.6	6.1±0.2	7.1±0.1

From the results, T_m increases with increasing number of repeats of C-tracts starting from **n= 4**, which is consistent with previous work on CCCT repeat sequences.[334] However, n= **3** showed unusual behavior in that it has the shortest number of C-tracts, but it showed T_m (52.1°C) much higher than n= **4**, **5,** and **6**. This may indicate other DNA structures might have formed, the CD still appears like an i-motif, but there are not enough C-tracts to form the same type of intramolecular i-motif as the analogues with more C-tracts. It might be there is a mixture of populations in this sample, both intramolecular and intermolecular. n = **7** showed two melting temperatures of 33.1°C and 60.5°C, potentially indicating more than one DNA structure formed. Intermolecular quadruplex structure formation is very slow, taking often days for the four strands to come together.[152] Therefore, in this case, it is likely that most of the species observed are intramolecular. Further experiments using gel electrophoresis may help with this,[335] which would indicate whether there are any intermolecular species formed in solution and provide an estimate of the proportions of each type of species in solution.

5.3.2 Ultraviolet Melting and Annealing of C-Rich Sequences

UV was used to perform DNA melting and annealing experiments. Compared with CD, UV is higher throughput as the instrumentation in the lab has a 6-cell changer and the sensitivity is much higher, allowing a range of concentrations to be measured. DNA absorbs light at 260 nm when it is irradiated, but folded DNA structure absorbs differently to unfolded DNA due to different π-π stacking interactions, therefore, UV absorbance can be used to monitor the DNA folding and unfolding process. Both melting and annealing temperatures can be determined for each i-motif sequence. All the sequences were thermally melted and annealed at concentrations between 0.5-10 µM, to allow assessment for any concentration-dependent effects. We used two different pHs, pH 5.5, where most i-motifs are stably folded, and pH 7.0 (neutral pH). At both pHs, most of the selected C-rich sequences showed both melting and annealing curves (***Figure 5.6***) indicating they all have folded and unfolded states, the melting (T_m) and annealing (T_a) temperatures were determined using the first derivative method. The results are shown in ***Table 5.5*** and the melting and annealing curves are presented in ***Appendix 14 – 18***.

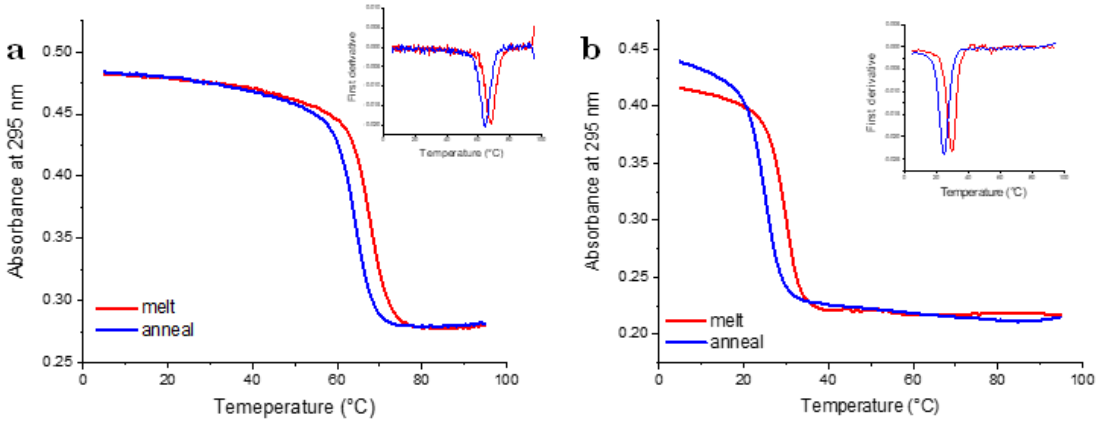


Figure 5.6 *Example of UV melting (red line) and annealing (blue line) process for 2.5 µM ALOX5 in buffer containing 10 mM sodium cacodylate and 100 mM sodium chloride at pH 5.5. The small graphs show the plot of the first derivative of melting (red) and annealing (blue) against temperature.*

Table 5.5 *GC-repeat i-motif sequences and data for their melting temperature (T_m) and annealing temperature (T_a) at 2.5 μM. Hysteresis shows the difference between melting and annealing temperatures. All UV experiments were performed in 10 mM sodium cacodylate and 100 mM sodium chloride at pH 5.5 and 7.0. Errors show the standard deviation across three repeats.*

		2.5 μM pH 5.5			2.5 μM pH 7.0		
Sequence name	Number of C-tract	T_m (°C)	T_a (°C)	Hysteresis (°C)	T_m (°C)	T_a (°C)	Hysteresis (°C)
MADD-C	3	51.3 ± 2.3	42.3 ± 2.5	9.0 ± 0.7	16.2±0.2	<5	ND
MSMO1-C	4	62.3 ± 2.5	59.3 ± 2.5	3.0 ± 0.5	16.7±0.5	14.9±0.6	1.8±0.3
DAP-C	5	63.3 ± 3.5	59.7 ± 2.1	3.7 ± 0.9	25.6±0.3	21.8±0.1	3.8±0.1
ALOX5-C	6	65.3 ± 2.5	61.3 ± 1.5	4.0 ± 0.9	30.0±0.2	25.5±0.2	4.5±0.2
ZNF480-C	7	65.0 ± 3	62.0 ± 2.6	3.0 ± 0.5	30.1±0.2	24.2±0.1	5.9±0.2

The melting and annealing temperatures for sequences at 2.5 μM, pH 7.0 are shown in ***Table 5.6***. Under both pH 5.5 and 7.0 experimental conditions, the T_m was found to increase with an increasing number of C-tracts (***Figure 5.7 a***). In the cases at pH 5.5, the increase in T_m from n= **3** to **4** was quite large (11°C), but from lengths from n= **4** through to **7**, the increase in T_m was marginal with each incremental C-tract added. A similar pattern was observed for the T_a. At pH 7.0, the T_m from n= **3** to **4** C-tracts only increased slightly (0.5°C), whereas the increase in T_m from four to five C-tracts was a lot larger (8.9°C). The increase in T_m from n= **5** to n=**6** was 4.4°C then after this, no further significant changes were observed. The same pattern of change in annealing temperature was also observed (***Figure 5.7 b***), which is consistent with the results from the Waller group published in 2022.[334] Interestingly, the plot of T_m and T_a against the number of C-tracts at pH 7.0 can be fitted perfectly using a sigmoidal curve, indicating that five tracts are an inflection point in behavior of these types of sequences.

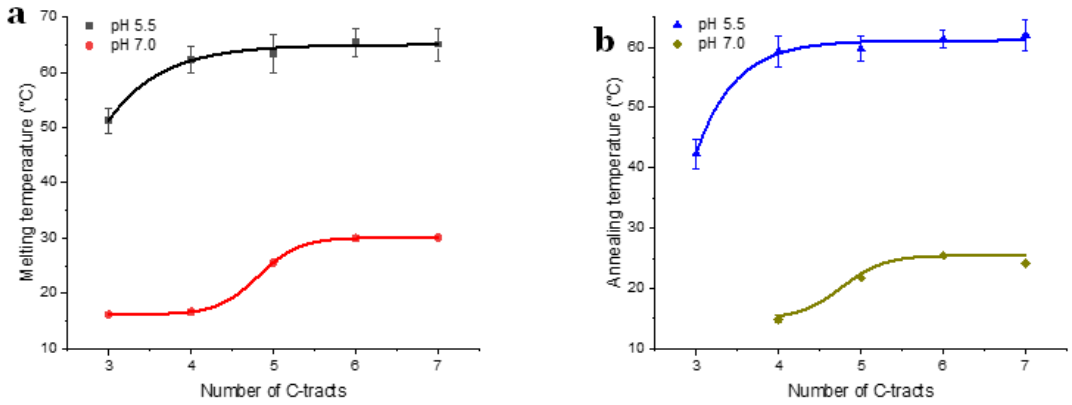


Figure 5.7 *Thermal stability of C-rich sequences with an increasing number of C-tracts. The experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at the indicated pH. a) T_m pH 5.5 (black), pH 7.0 (red). b) T_a pH 5.5 (blue), pH 7.0 (dark yellow). Error bars show the standard deviation across three repeats.*

From the determination of the transitional pH, all selected sequences presented two transitions, that could form more than one subset of structures, these could be intermolecular or intramolecular. From the previous G-quadruplex studies, in the case of intramolecular equilibrium, the melting temperature is concentration-dependent[336] as multimolecular reactions are concentration dependent, while

unimolecular (intramolecular) are not. Intramolecular quadruplex (both G-quadruplex or i-motif) structures may have different stabilities from intermolecular ones. In comparison with the T_ms derived from CD melting were observed to be slightly different to those derived from UV melting experiments. For example, for sequences with C-tract length n=**3** the T_m was found to be 52.1°C by CD whereas the analogous experiment in the UV was found to be 16.2°C. Similarly, the analogous sequence where there are seven repeats, n=**7**, there are two melting transitions observed in the CD (T_{m1} 33.1°C, T_{m2} 60.5°C) whereas there is only one in the UV experiments (30.1°C). The reason for this might arise from the fact that CD and UV measured different things in DNA. For example, in the CD, the signal at 288 nm might signify a different aspect of i-motif structure than the signal at 295 nm in the UV. Another possible reason might be the difference in concentration of DNA, as 10 μM sample used in CD while 2.5 μM in UV, in higher concentration solution, a more single strand would have more chance to form different DNA structures, for example, intermolecular or GC duplex structure, therefore, the next experiments performed UV melting and annealing for C-rich sequences at different concentration, including 0.5, 1.0 and 10 μM The melting and annealing temperatures of different concentration are presented in ***Table 5.6***.

Table 5.6 *GC-repeat i-motif sequences and data for their melting temperature (T_m) and annealing temperature (T_a) determined by UV melting experiments at 0.5 μM, 1 μM, 2.5 μM, and 10 μM. T_m and T_a values are derived from the first derivative. Hysteresis shows the difference between melting and annealing temperatures, all UV experiments were performed at pH 7.0. Errors show the standard deviation across three repeats.*

	0.5 μM			1 μM			2.5 μM			10 μM		
DNA	T_m	T_a	hysteresis	T_m	T_a	hysteresis	T_m	T_a	hysteresis	T_m	T_a	hysteresis
MADD-C	<5	<5	ND	<5	<5	ND	16.2±0.2	<5	ND	17.8 ± 0.1	<5	ND
MSMO1-C	19.0±0.7	15.1±0.6	3.9±0.3	19.0±0.6	16.1±0.2	2.9±0.2	16.7±0.5	14.9±0.6	1.8±0.3	18.6 ± 0.1	14.5 ± 0.1	4.1±0.1
DAP-C	26.6±1	22.9±0.6	4.2±0.5	27.3±0.8	21.9±0.1	5.4±0.1	25.6±0.3	21.8±0.1	3.8±0.1	25.7 ± 0.2	21.7 ± 0	4.0±0
ALOX5-C	30.1±0.01	24.6±0.4	5.5±0.3	29.8±0.1	24.8±0.2	5±0.1	30.0±0.2	25.5±0.2	4.5±0.2	29.8 ± 0	25.8 ± 0	4.0±0
ZNF480-C	30.6±1	24.5±0.2	6.1±0.1	30.1±0.5	24.7±0.2	5.4±0.3	30.1±0.2	24.2±0.1	5.9±0.2	30.0 ± 0.1	25.5 ± 0.1	4.5±0.1

All the selected C-rich sequences were studied using UV melting and annealing in the same buffer at pH 7.0. According to the results, at each concentration of DNA, the T_m values increased with an increasing number of C-tracts. The T_m and T_a values of different DNA concentrations, in general, are roughly the same.

The relationship between the number of C-tracts and hysteresis[337] was also measured (***Figure 5.8***). Hysteresis happened when a melting curve is not overlapped ideally on its corresponding annealing curve, typically caused when the rate folding/unfolding is slower than the temperature ramp in the experiment. The difference between the two thermal curves is hysteresis. The presence of hysteresis shows the kinetics of association and dissociation are slower than the temperature ramp of the experiment. If the hysteresis is concentration independent, this indicates that formation of structures is a slow intramolecular process.[338,337]

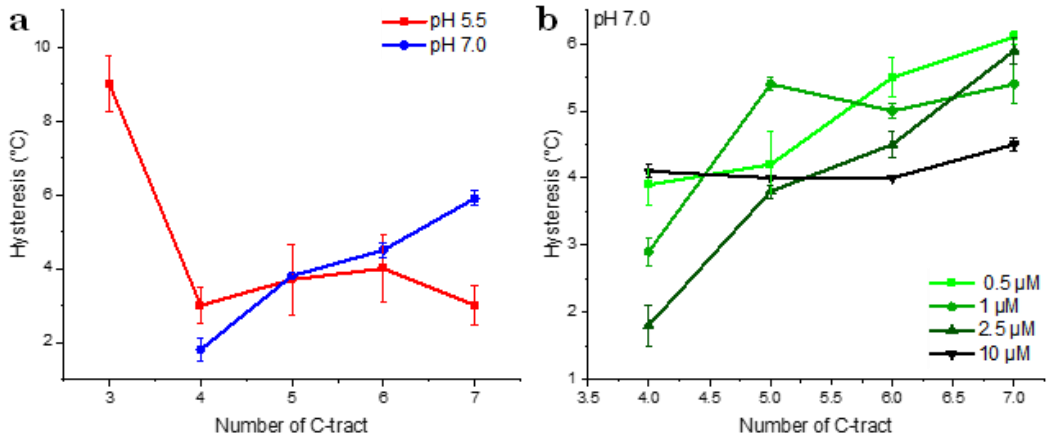


Figure 5.8 *The effect of the number of cytosine tracts on hysteresis. a) Hysteresis of 2.5 µM DNA at pH 5.5 (red) and pH 7.0 (blue) b) Hysteresis of different DNA concentration at pH 7.0. All samples were prepared in 10 mM sodium cacodylate and 100 mM sodium chloride.*

At pH 5.5, the size of hysteresis decreased from cytosine tracts to **n= 4**, and then from **4** to **7**, there was no further change, this indicates that **n= 3** has a different type of population compared to the other lengths, possibly evidence of intermolecular structures. At pH 7.0, we examined the hysteresis at different

concentrations for each number of C-tract repeats. The hysteresis was found to vary with concentration of DNA, for all the different sequences examined but the overall differences in hysteresis were only up to 2.5°C, which although is minimal, is significant. This indicates that the folding of DNA structures at neutral pH is complex, and the association kinetics are slower than the experimental conditions. In a biological context, the formation with slow kinetics might inhibit any regulatory functions in genomic C-rich sequences but could be influenced by any chemical interaction or protein binding.[54]

5.3.3 Thermal Difference Spectra of C-Rich Sequences

The thermal difference spectrum was used to determine the dominant species of each folded DNA sequence. TDS were measured at both pH 5.5 and 7.0 at 2.5 μM DNA concentration. TDS at pH 7.0 were also measured for DNA concentrations of 0.5, 1.0, and 10 μM. UV absorbance was measured for each sequence at 5°C where DNA was completely folded and at 95°C in which DNA was unfolded. The difference between these two UV spectra can be calculated and the plot of the difference in absorbance against wavelength gives an indicative spectrum that can be used for characterizing DNA structure.[167] At pH 5.5, the TDS (***Figure 5.9***) of all the selected C-rich sequences showed a characterized positive peak at around 240 nm and a negative one at around 295 nm, the result was consistent with the previous TDS of characterized folded i-motif structures[167,54], indicating all these C-rich sequences were folded into i-motif structures at pH 5.5.

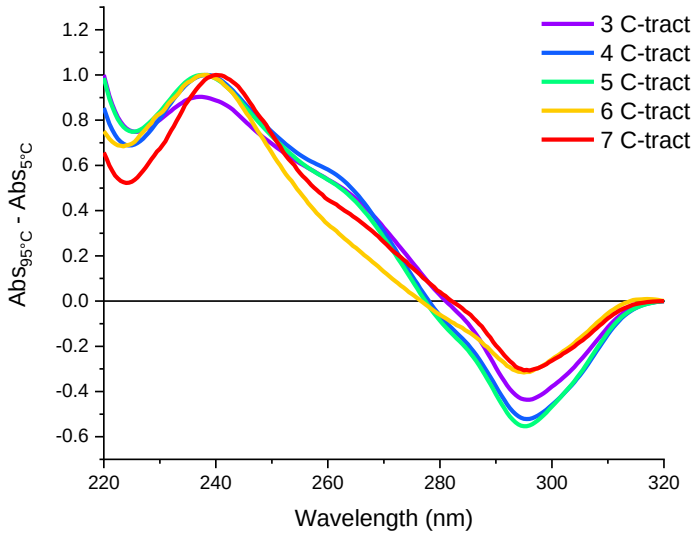


Figure 5.9 *Thermal difference spectra of 2.5 μM C-rich sequences in 10 mM sodium cacodylate and 100 mM sodium chloride at pH 5.5.*

At pH 7.0, the absorbance of 10 μM DNA samples was high and exceeded the maximum value of the machine, so the TDS could not be determined for this concentration. In the case of 0.5 μM of DNA (***Figure 5.10***), all the sequences showed featured negative peaks at 295 nm, for the sequence with C-tract from **4** to **7**, the positive peaks at ~240 nm[167] were also observed, indicating these four sequences can form i-motif structures. In the case of n= **7**, another positive peak at ~270 nm was observed, which looks different to the other i-motif species. According to the TDS spectra of the i-motif from the Mergny group, the shape of n = **7** TDS is within the error of an i-motif TDS, which is indicative of the predominated species for n= **7** were i-motif structures. The TDS for n= **3** seemed different from others, as it showed a positive peak at ~280 nm and a negative one at ~262 nm, which was similar to the duplex TDS.[167]

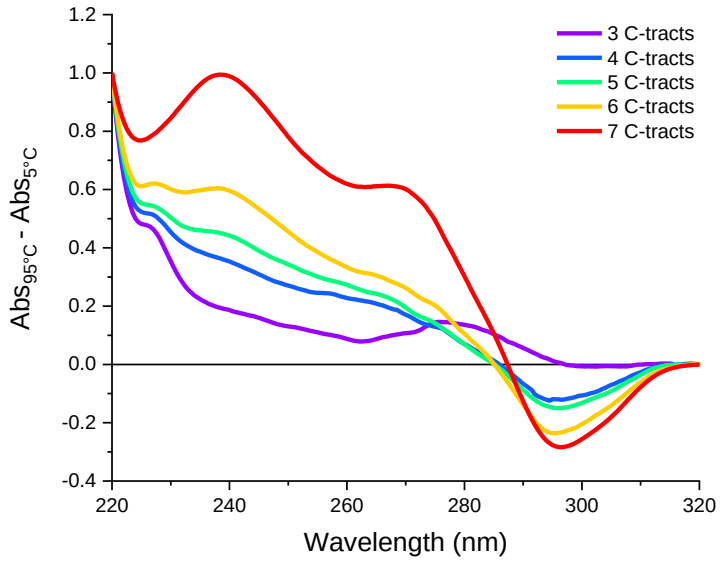


Figure 5.10 *Thermal difference spectra of 0.5 μM C-rich sequences in 10 mM sodium cacodylate and 100 mM sodium chloride at pH 7.0.*

In the case of 1.0 μM of DNA (***Figure 5.11***), as the concentration was increased, so the spectrum signal was clearer. It was observed that n= **4** to **7** C-tract sequences had clear positive peaks at ~240 nm and negative peaks at ~295 nm, suggesting they all can fold into an i-motif. However, in n= **4** and **7** C-tract sequences, a clear positive peak at ~270 nm was detected, and **6** C-tracts also had a weak positive peak at this wavelength, however, the CD spectra for these three sequences showed the wavelength for i-motif, indicating the dominated species were i-motif. For n= **3**, at this concentration, it showed more strong peaks at both ~262 and 280 nm, this was consistent with a duplex TDS.

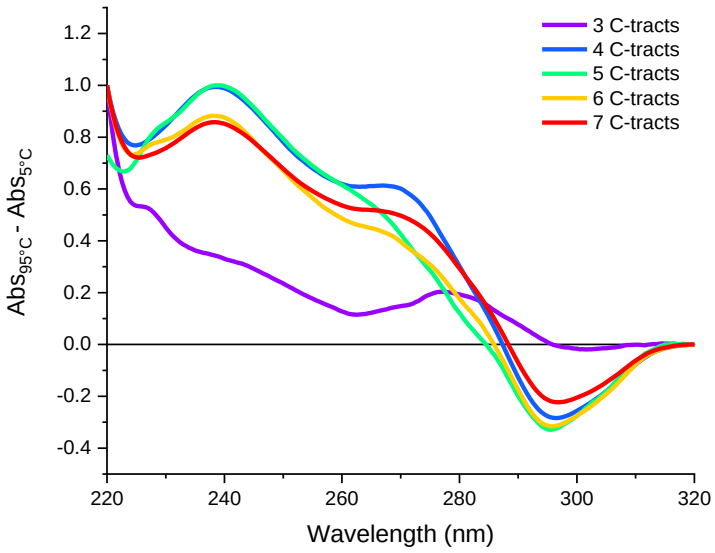


Figure 5.11 *Thermal difference spectra of 1.0 μM C-rich sequences in 10 mM sodium cacodylate and 100 mM sodium chloride at pH 7.0.*

From the TDS of 2.5 μM DNA (***Figure 5.12***), all the sequences showed positive peaks at ~240 nm and negative ones at ~295 nm, indicating that i-motif structures formed. For n= **3** another positive peak was presented at ~280 nm and a negative one at ~262 nm, suggesting it had mixed species in a solution of i-motif and duplex. ***Table 5.7*** presents the wavelength for each C-rich sequence at different DNA concentrations.

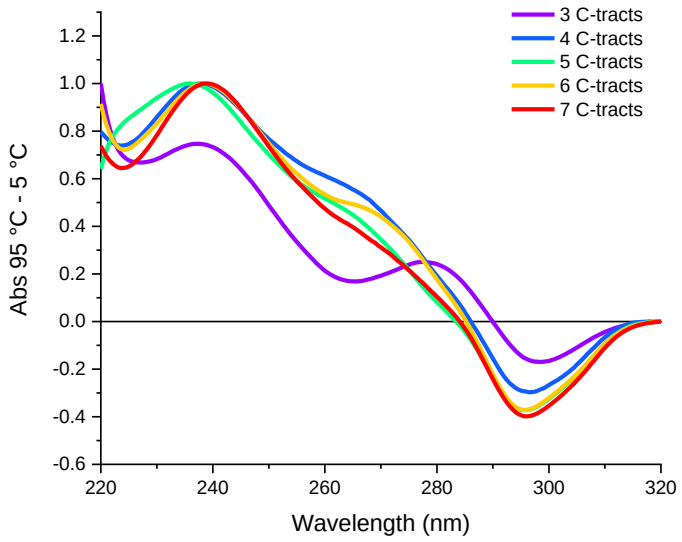


Figure 5.12 *Thermal difference spectra of 2.5 μM C-rich sequences in 10 mM sodium cacodylate and 100 mM sodium chloride at pH 7.0.*

Table 5.7 *Wavelength and possible species for each C-rich sequence according to TDS.*

Sequence	pH	[DNA](μM)	Positive peak (nm)	Negative peak (nm)	Dominant Species
MADD-C 3 C-tracts	5.5	2.5	240	295	i-motif
	7.0	0.5	280	262 / 297	duplex
		1.0	279 nm	261 / 295	duplex
		2.5	239 / 279	264 / 298	i-motif/duplex
	5.5	2.5	237	294	i-motif
		0.5	239	295	i-motif

MSMO1-C 4 C-tracts	7.0	1.0	239 / 269	260 / 295	i-motif
		2.5	239	295	i-motif
DAP-C 5 C-tracts	5.5	2.5	239	295	i-motif
	7.0	0.5	239	295	i-motif
		1.0	239	295	i-motif
		2.5	239	295	i-motif
ALOX5-C 6 C-tracts	5.5	2.5	239	295	i-motif
	7.0	0.5	239	295	i-motif
		1.0	239 / 267	260 / 295	i-motif
		2.5	239	295	i-motif
ZNF480-C 7 C-tracts	5.5	2.5	240	295	i-motif
	7.0	0.5	239 /270	260 / 295	i-motif
		1.0	239 / 267	260 / 295	i-motif
		2.5	239	295	i-motif

From the CD data, all the sequences with C-tracts n=**3** to **7** showed two transitional pHs, suggesting an equilibrium of different i-motif structures such as intermolecular and intramolecular i-motif, hairpin, and duplex. UV melting and

annealing experiments showed no correlation between melting temperature and different DNA concentrations for n= **4**–**7**, which suggested these sequences are more likely to form intramolecular structures at neutral pH. However, the n= **3** sequence is more likely to form an intermolecular i-motif structure at high DNA concentration. The TDS were consistent with i-motif structures at pH 5.5. At pH 7.0, n= **3** showed different properties, as its spectrum was consistent with GC duplex, which indicated the dominated species of n= **3** was potentially not i-motif.

5.4 Study of G-Rich Repeat Sequences $(GGGGGC)_n$

5.4.1 Circular Dichroism Spectra of G-Rich Sequences

The mechanisms of i-motif and G-quadruplex folding have been used to study the propensity of their formation across the genome and representation in regulatory regions.[54,339] The i-motif and G-quadruplexes have been widely used as targets for the design of ligands that stabilize i-motif and G-quadruplex structures and thus regulate gene expression.[236,303,340] Most G-quadruplexes have been identified as inhibitors of transcription,[302] while, more and more evidence has shown that i-motifs may play a role in activating gene expression.[248,340] Therefore, understanding the i-motif and G-quadruplex dynamics relative to each other is important for developing targeted therapeutics in this field. For example, work in 2020 indicated in cell-based experiments that i-motif and G-quadruplex are highly dependent on the formation of each other.[225] This study has implications indicating that the study of both i-motif and G-quadruplex at the same time is important. Thus, following on from the study of the C-rich repeat sequences, having established the characterization for i-motif sequences, the complementary sequences to C-rich ones were also studied (***Table 5.8***).

Table 5.8 *Tandem repeats G-rich sequences and number of guanine tract.*

Sequence name	Sequence 5′ – 3′	Number of G-tracts
MADD-G	$(GGGGGC)_2GGGGG$	3
MSMO1-G	$(GGGGGC)_3GGGGG$	4
DAP-G	$(GGGGGC)_4GGGGG$	5
ALOX5-G	$(GGGGGC)_5GGGGG$	6
ZNF480-G	$(GGGGGC)_6GGGGG$	7

G-quadruplex structures can adopt a wide range of topologies broadly defined as parallel, antiparallel or mixed, and these are dependent on the sequence and length of the strand and loops.[341] As discussed in **Section 1.2.1**, the formation of G-quadruplex has a typical cation dependency,[342] so within this study, different salt conditions were used to characterize the sequences and identify whether these G-rich sequences had typical G-quadruplex forming behaviours. Potassium and sodium ions are physiologically relevant monovalent ions involved in cell function[343] and both have a stabilizing influence ($K^+ > Na^+$) on the stability of the G-quadruplex. Lithium ions, within the same group in the periodic table, are too small to coordinate the G-tetrad,[344,345] therefore, all these G-rich sequences were prepared in 10 mM lithium cacodylate, and 100 mM of additional monovalent cations in the form of either potassium, sodium or lithium chloride at pH 7.0. The Fox group previously studied some short oligonucleotides sequences with short loops ($d[G_3T]_4$, $d[G_3T_2]_4$ and $d[G_4T]_4$). In the presence of potassium and sodium cations they all adopt parallel structures,[346] as the short loops are not sufficient to fold into antiparallel structures. For some longer G-rich sequences, a folded G-quadruplex therefore need to have G residues in the loops, as when you increase the stack of G-tetrads the distance to form parallel G-quadruplex obviously also increases with each additional stack. For example, where the G-tracts were longer e.g. $d[G_5T]_4$, $d[G_6T]_4$ and $d[G_7T]_4$, it was found that a single T was not sufficient to bridge the gap between the top and bottom of seven G-quartets.[346] For comparison with the data for the i-motif, experiments were also performed in sodium

cacodylate. The CD spectra showed the G-tract repeat sequences can adopt different topologies in K^+, Na^+, and Li^+.

Using the same buffer conditions that were used for the the i-motif forming sequences (including 100 mM potassium cations, ***Figure 5.13***), all the G-rich sequences showed positive peaks at 260 nm, which is indicative of a parallel G-quadruplex.[42] Sequence repeat n = **3** only showed a parallel G-quadruplex structure with a positive peak at 260 nm and a negative minimum at 240 nm,[159] **4** and **5** showed positive peaks at 260 and 295 nm, indicating they are mixed types with roughly equal parallel and antiparallel, **6** and **7** are mainly parallel but do have some antiparallel species.

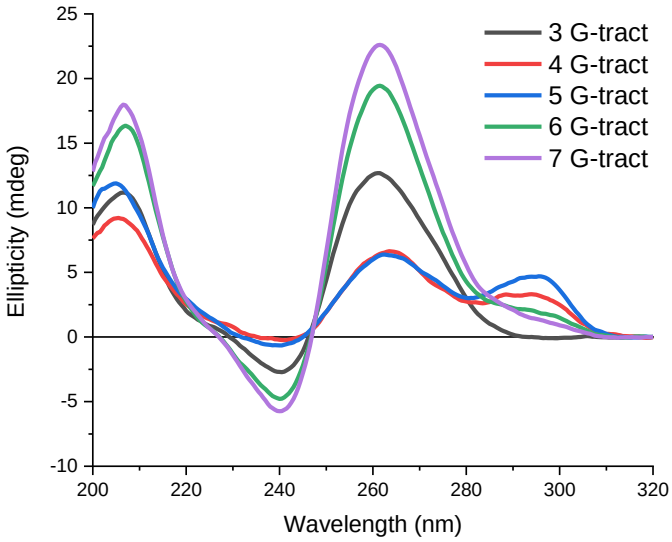


Figure 5.13 *CD spectrum for each 10 μM G-tract sequence in buffer containing 10 mM sodium cacodylate and 100 mM potassium chloride at pH 7.0*

In buffers containing 100 mM NaCl (***Figure 5.14***), there were also a wide variety of different DNA topologies formed. The sequence with **3** G-tracts mainly showed a positive peak at 260 nm and a negative one at 240 nm, also only a very slight peak at 295 nm was observed, suggesting it mainly formed a parallel G-quadruplex structure and a small amount of antiparallel. Note that the sequence with **3** G-tracts, although it is not possible to fold this sequence into an

intramolecular structure with a stack of five G-tetrads, the sequence is long enough still to fold into other intramolecular G-quadruplexes. For example, there is the potential to fold a G-quadruplex with two G-tetrads (i.e. GGGGGCGGGGGCGGGGG or another analogous combination of G2-tracts as the sequence is long enough for there to be a few different potential combinations), or a G-quadruplex with three tetrads, with bulges[347] (i.e. GGGGG**C**GGGGGG**C**GGGGG, where the two cytosines indicated in bold would be bulges). So this sequence is expected to behave differently compared to others in the series. The sequences with **5** and **6** G-tracts had a single positive peak at 295 nm, respectively, indicating the predominated formation of antiparallel G-quadruplex folded, this is consistent with the previous work. The first intramolecular G-quadruplex was identified by the Patel group in 1993, they found the sequence AGGG(TTAGGG)$_3$ could form antiparallel structures in the presence of Na^+.[348] **4** and **7** displayed positive peaks at both 260 nm and 295 nm, and a negative one at 240 nm, this is indicative of the mixture of parallel and antiparallel G-quadruplex topologies.

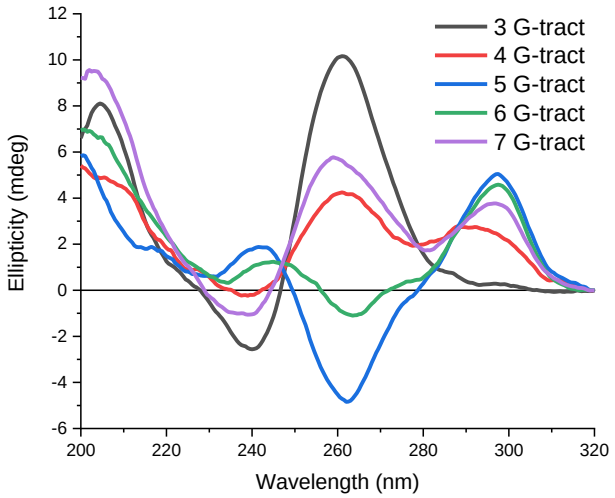


Figure 5.14 *CD spectrum for each 10 μM G-tract sequence in buffer containing 10 mM sodium cacodylate and 100 mM sodium chloride at pH 7.0*

In buffers containing 100 mM LiCl (***Figure 5.15***), **3**, **6,** and **7** clearly showed a positive peak at 260 and a negative peak at 240, indicating they only formed a

parallel G-quadruplex structure. The other two **4** and **5** showed a positive peak at 260 nm and a negative one at 240 nm, but they also displayed a slight positive peak at 295 nm, indicating a small population of antiparallel species existed in Li^+ conditions. It is important to note that in this buffer there is still 10 mM Na^+ cations, from the cacodylate, so any observed amount of antiparallel species may be due to the presence of sodium in the conditions.

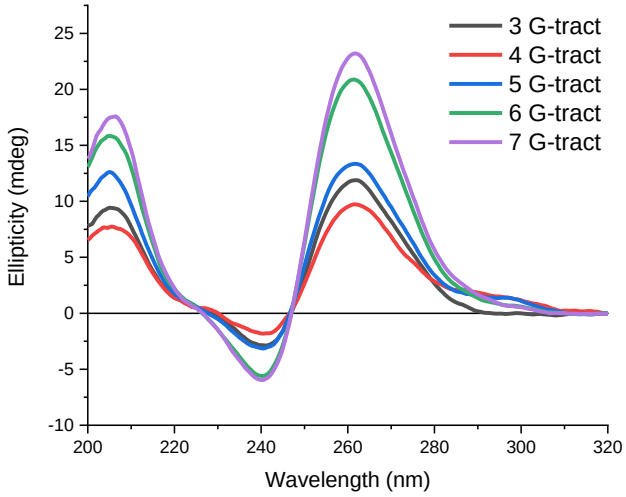


Figure 5.15 *CD spectrum for each 10 μM G-tract sequence in buffer containing 10 mM sodium cacodylate and 100 mM lithium chloride at pH 7.0*

The CD spectrum of each sequence was put with three cations respectively, shown in ***Figure 5.16***.

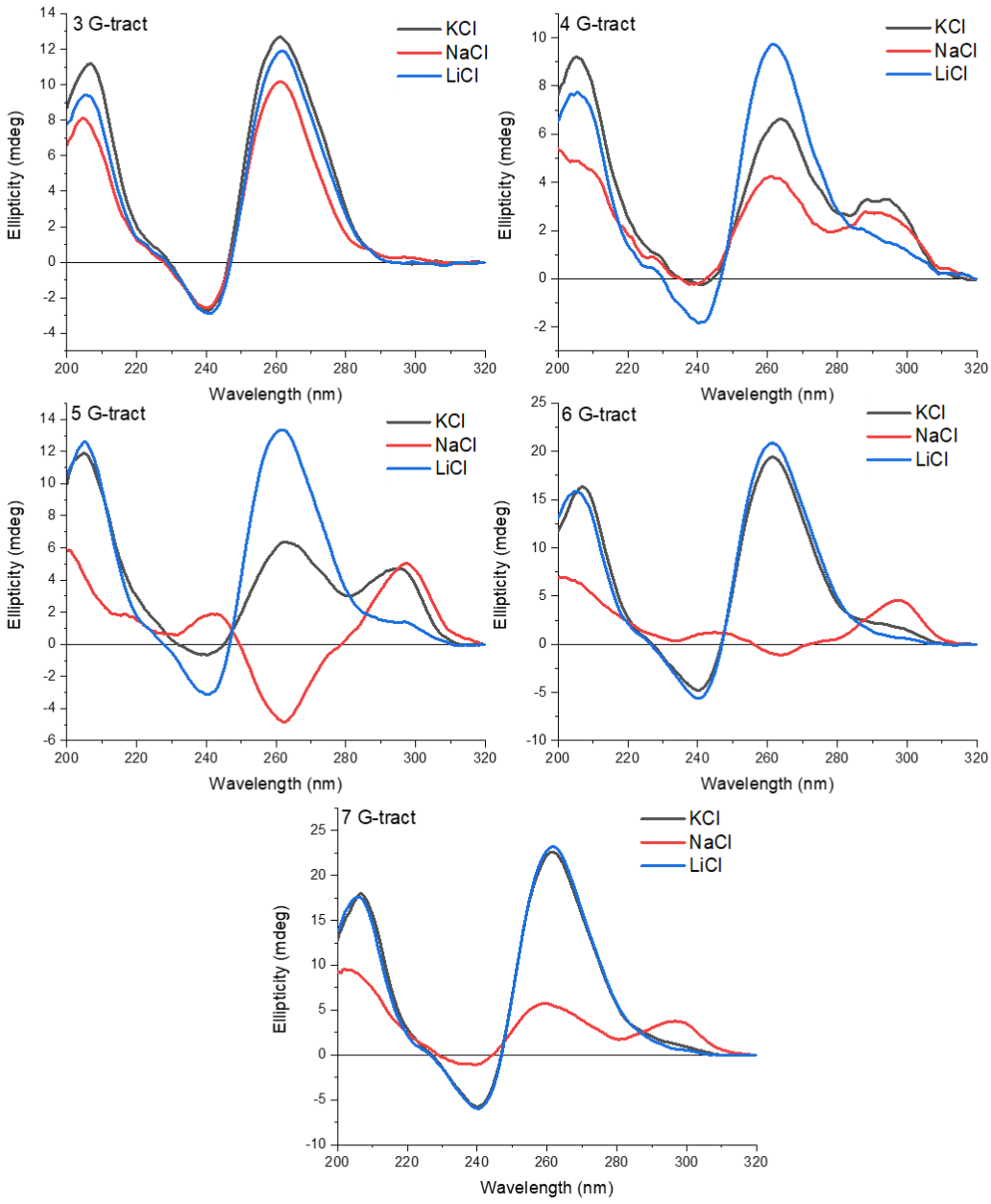


Figure 5.16 *CD spectra of selected G-rich sequences. Each sequence was performed in 10 mM sodium cacodylate and 100 mM potassium chloride / sodium chloride / lithium chloride buffer at pH 7.0 with concentration of DNA 10 μM.*

From the graphs, **4**, **5**, **6**, and **7** seemed special, as they all showed mixed species of parallel (260 nm) and antiparallel (295 nm) topologies in the presence of K^+ and Na^+, even **7** also had a very small peak at 295 nm in K^+. The reason could be these sequences are quite long, so they were able to adopt different topologies. Marin and co-workers had studied long G-rich sequences $G_5H_1G_5H_1G_5H_1G_5$ and

$G_6H_1G_6H_1G_6H_1G_6$ (H is any nucleotide except G) with single nucleotide loops. According to their CD spectra experiments, in the presence of K^+, five and six stacked quartets showed two peaks at 264 and 296 nm, which means 5G and 6G sequences could fold into both parallel and antiparallel quadruplex structures.[349] The other study by the Panyutin group, used ^{125}I-radioprobing to probe the human telomeric sequence and showed that it formed an antiparallel G-quadruplex structure in the presence of K^+ and Na^+.[350] Herein, our G-rich sequences all contained five guanines that could stack to form quadruplex $(G_5C)_nG_5$, there is no surprise that they can also adopt mixed species in K^+. Interestingly, however, the sequences showed the highest ellipticity for parallel G-quadruplex in Li^+ solution, which is quite unusual.

As the effects of cations on the topology of G-quadruplex structures were discussed in **Section 1.2.1**, the topology of G-quadruplex affected by these three cations would be $K^+ > Na^+ > Li^+$, depending on their size of ion radius,[351] so the CD signal intensity should show the same spectra, as strongest spectra are K^+, then Na^+ and the less strong is Li^+. The study in G-quadruplex topology indicated that Na^+ induces the formation of an antiparallel structure rather than K^+, however, this was strongly dependent on the DNA sequence.[352,353] In this case, it was observed that there were mixed species in the 100 mM KCl buffer, and in some cases only antiparallel in the 100 mM NaCl buffer. The reason for the mixed species in the KCl buffer, and the appearance of quadruplex CD signatures in the presence of 100 mM Li, might be due to the 10 mM Na from the cacodylate. Therefore, these were repeated using a fresh stock solution of lithium cacodylate: the samples were prepared in 10 mM lithium cacodylate and 100 mM potassium/sodium/lithium buffer at pH 7.0 (***Figure 5.17***).

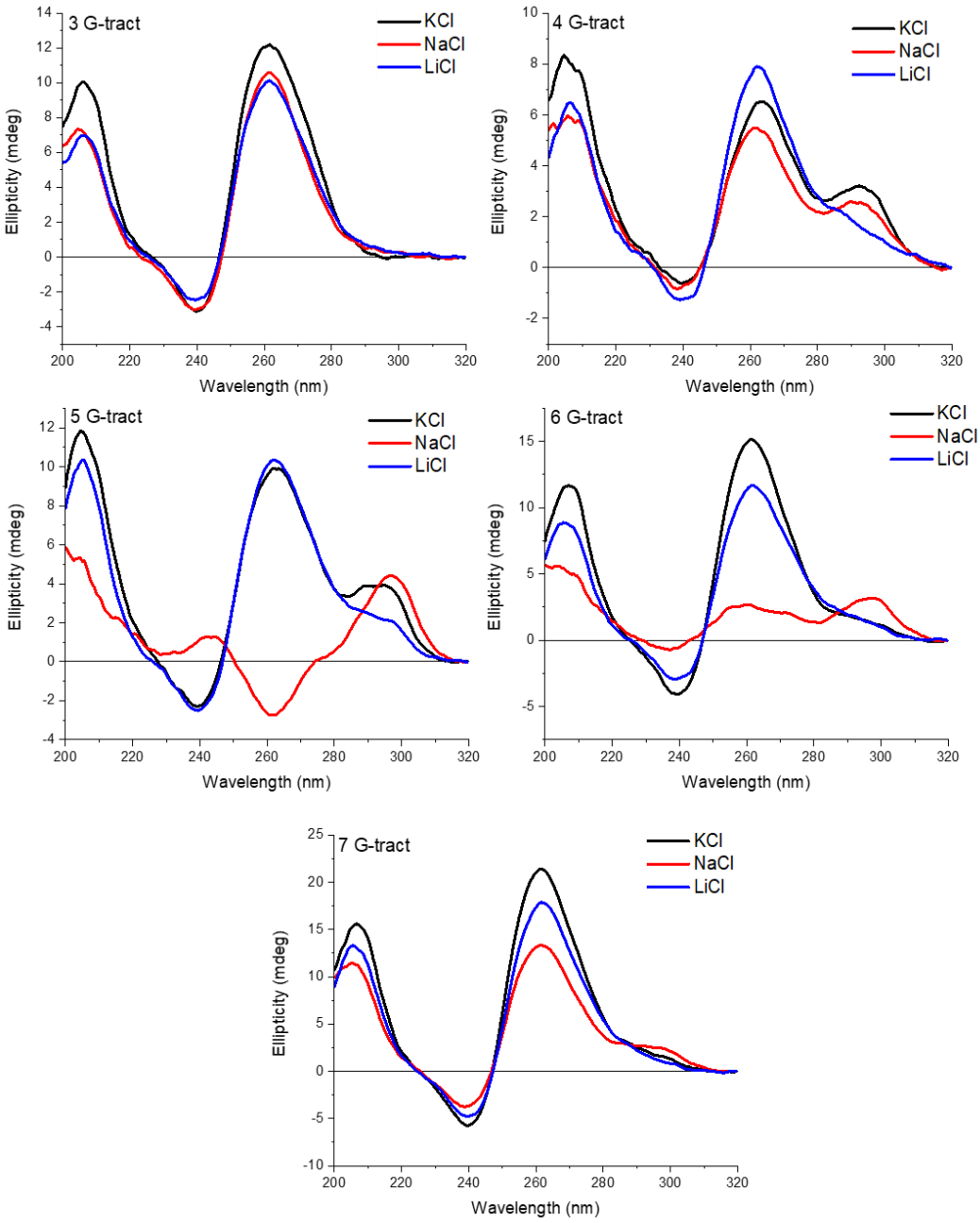


Figure 5.17 *CD spectra of selected G-rich sequences. Each sequence was performed in 10 mM lithium cacodylate and 100 mM potassium chloride / sodium chloride / lithium chloride buffer at pH 7.0 with 10 μM DNA.*

In the new buffer conditions, with the exception of n= **4**, the sequences showed the highest CD signal in the K^+ solution, which is consistent with the previous G-quadruplex study.[38] n= **3** displayed a positive peak at 260 nm and a negative one at 240 nm in all three cations, indicating the formation of a parallel G-quadruplex.

n= **4** exhibited two positive peaks at 260 nm and 295 nm in K^+ and Na^+, which resulted from the mixed conformations of parallel and antiparallel structures. In the presence of Li^+ there was only a positive peak at 260 nm and a negative peak at 240 nm, indicating parallel quadruplex topology formed. For most of the sequences, the strongest CD signal at 260 nm was in the presence of K^+ cations. However, it is worth noting that this was not the case for the sequence where n= **4**, where the strongest signal was actually observed in Li^+ cations. In Na^+, there was only a positive peak at 295 nm and a negative one at 260 nm, indicating only antiparallel structure formed. n= **6** and **7** showed a positive peak and negative peak at 260 nm and 240 nm respectively in three buffers, indicating the formation of a parallel structure, though in Na^+, another positive peak appeared at 295 nm for both sequences, which suggested the mixed species existed in Na^+ condition.

One interesting observation is that when the number of G-tract increases, the amount of parallel topology also increases from **n=4-7**. This means that the stacks of G-tetrads in these species cannot be 4-7 stacks high respectively. One could easily imagine the sequence $(GGGGGC)_7$ to exist has a G-quadruplex with a stack of seven tetrads, but as Fox and co-workers have shown,[346] this is not possible as the theoretical distance between the top and bottom tetrads is too far to connect with a single nucleotide loop, the same can be said for anything higher than five tetrads. At a tract length of seven, we observe mostly parallel folding, but the loops will have to be longer and incorporate guanines, so the stack of G-tetrads will be less than seven.

5.4.2 Thermal Stability of G-Rich Sequences

After investigating the topologies of the G-rich sequences in different cation conditions, then the thermal stability of these was examined in the presence of 10 mM sodium cacodylate with 100 mM K^+, Na^+ and Li^+ using CD and UV melting experiments.

In the CD melting experiments (***Figure 5.18***), it is clear that these sequences have not finished melting under the conditions of the experiment **Appendix 19–21**.

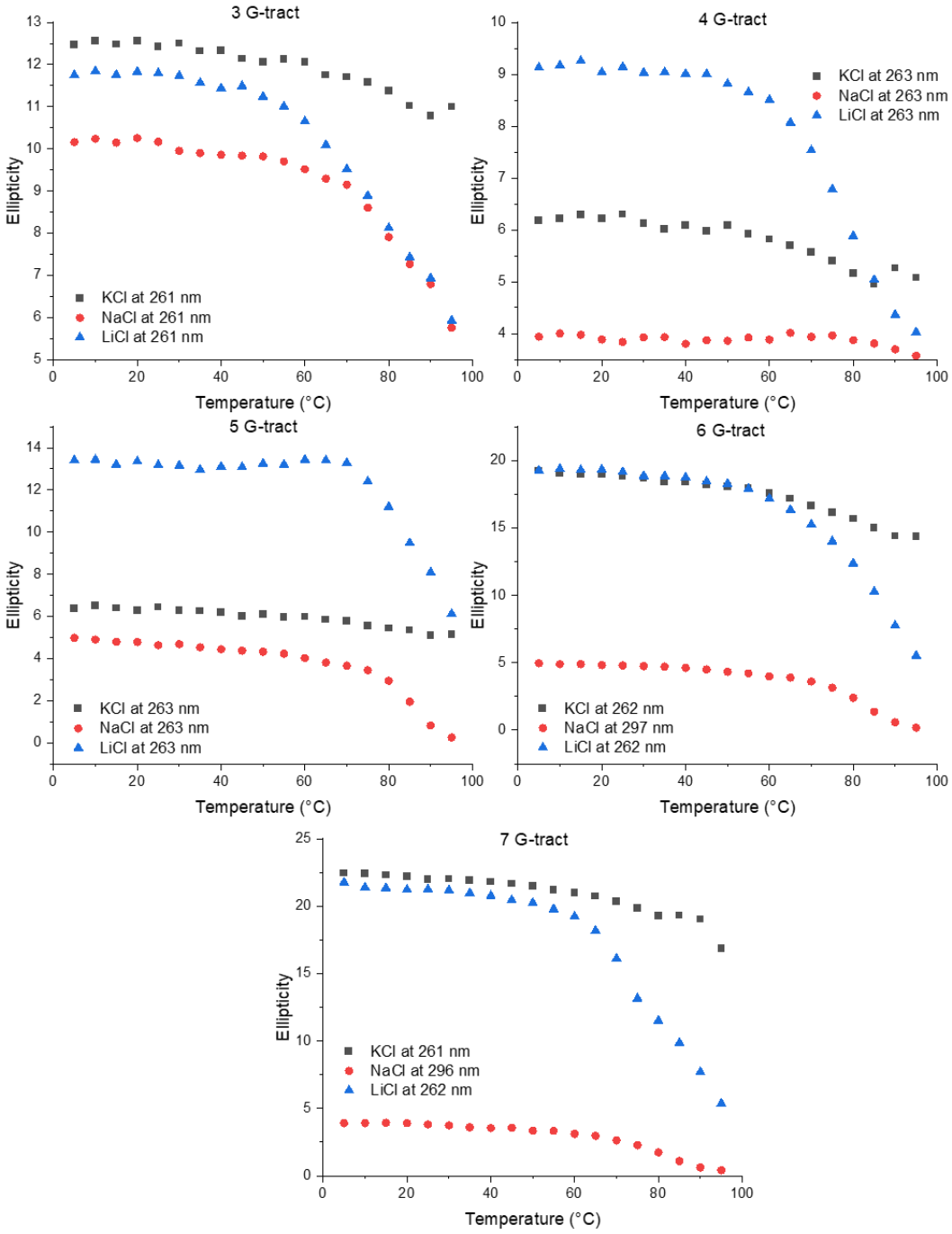


Figure 5.18 *CD melting of G-rich sequences in 10 mM sodium cacodylate and 100 mM KCl, NaCl, and LiCl buffer at pH 7.0.*

Then UV melting experiments of G-rich sequences were also performed with three different cations, the melting curves are presented in ***Figure 5.19***. Examples of both UV melting and annealing processes of G-rich sequences in K^+, Na^+ and Li^+ buffers are presented in **Appendix 22 – 24**.

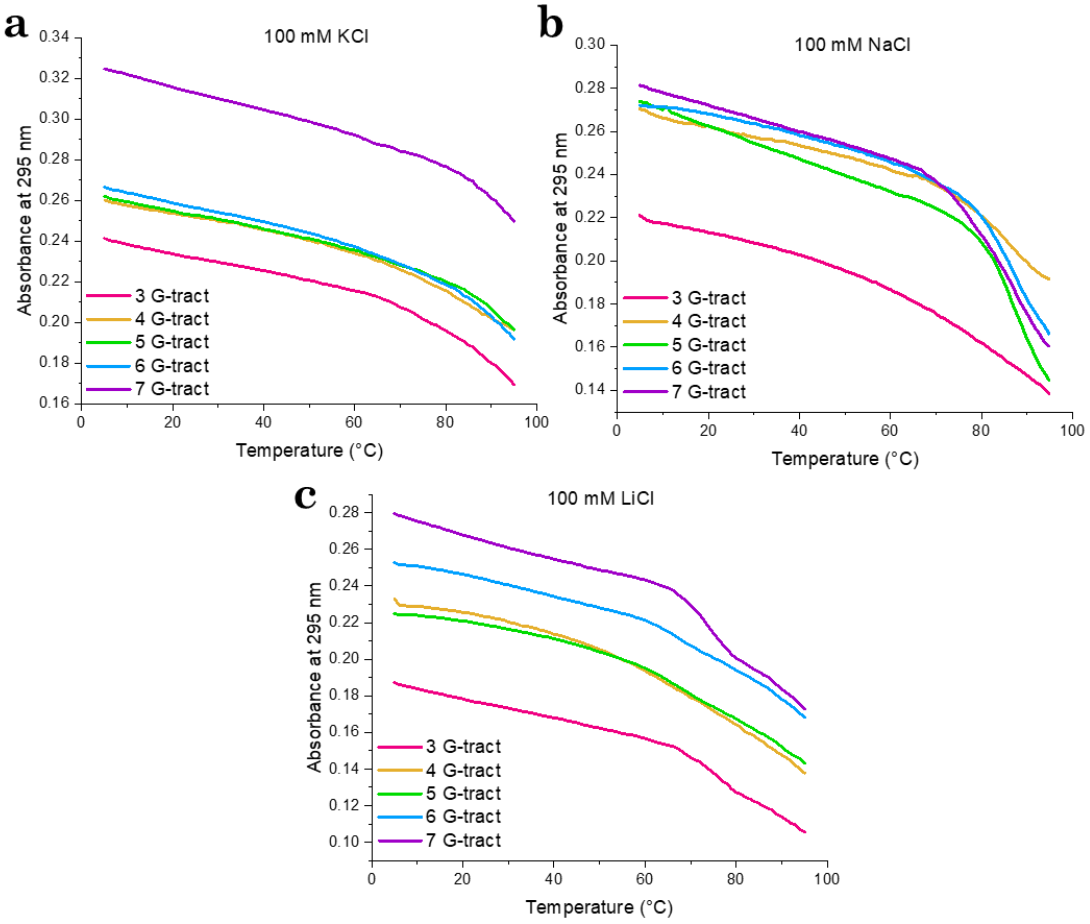


Figure 5.19 *UV melting curves and UV absorbance at 295 nm of G-rich sequences in 10 mM sodium cacodylate and 100 mM a). KCl, b). NaCl and c). LiCl at pH 7.0*

These data show that all the G-rich sequences were very stable, regardless of the conditions in which they were examined. Even in Lithium cations, the transitions are not complete at the highest temperature of the experiments (i.e. the T_m >95°C). The melting and annealing curves did not overlap (examples of UV melting and annealing curves are presented in ***Figure 5.20***), thus the hysteresis occurs between melting and annealing temperatures of G-tract sequences in different cations, demonstrating the slow kinetics of folding and/or unfolding. It is likely

that these sequences still fold into intramolecular species as it has been shown that intermolecular G-quadruplex formation is extremely slow.[152]

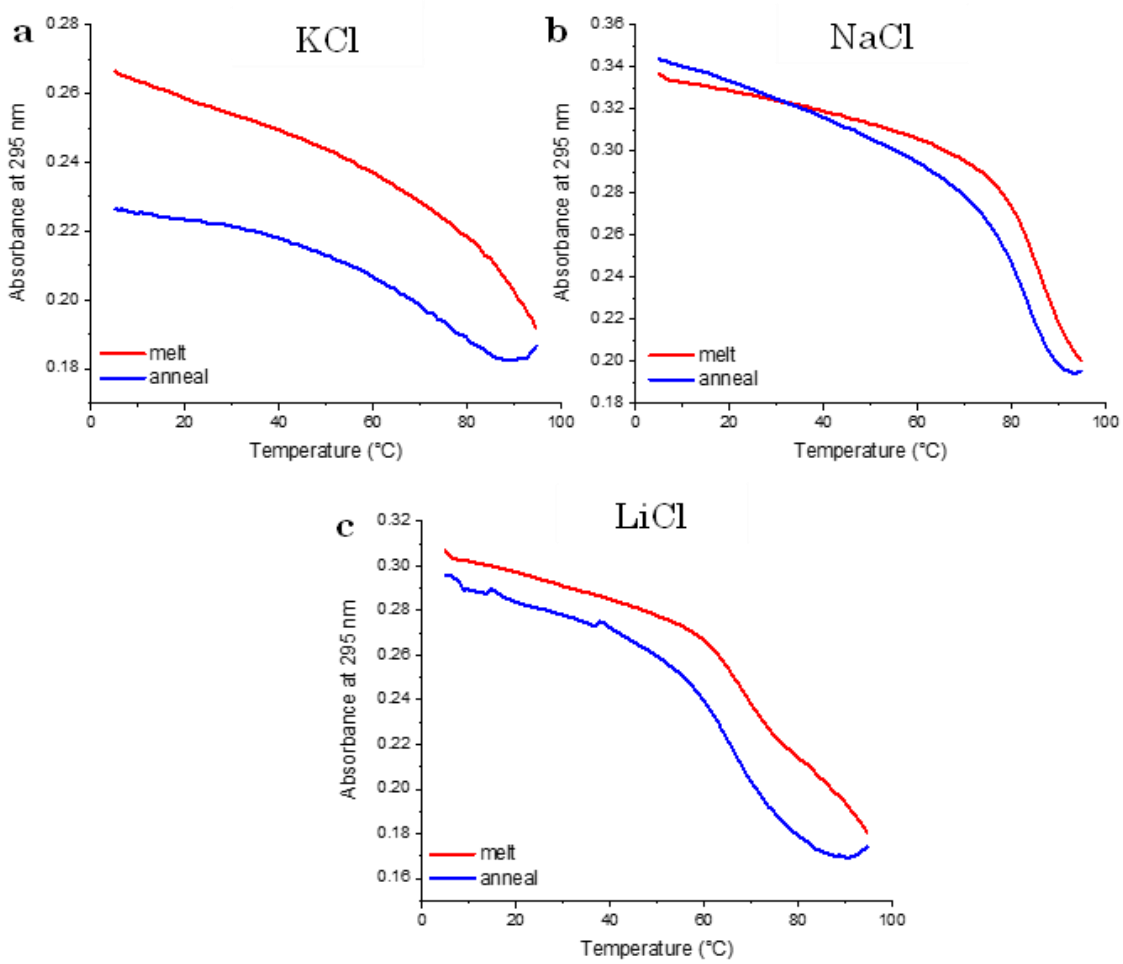


Figure 5.20 *Examples of UV melting and annealing curves of n= **6**. [DNA] was 2.5 μM, and experiments were performed in 10 mM sodium cacodylate and 100 mM a). KCl, b). NaCl and c). LiCl at pH 7.0.*

We calculated the estimated hysteresis for the G-rich sequences in three different cations conditions, and the results are presented in ***Table 5.9***. The relationship between hysteresis and the number of G-tract was also plotted (***Figure 5.21***).

Table 5.9 *Estimated hysteresis between T_m and T_a of G-tract sequences performed in 10 mM sodium cacodylate and 100 mM KCl, NaCl, and LiCl at pH 7.0.*

				Hysteresis (°C)		
Notation	Bases	n	Sequence 5′ → 3′	KCl	NaCl	LiCl
MADD-G	18	3	$(G_5C)_2G_5$	>28.6	>29.6	>25.0
MSMO-G	23	4	$(G_5C)_3G_5$	>28.9	>19.7	>25.4
DAP-G	29	5	$(G_5C)_4G_5$	>21.3	>8.3	>28.2
ALOX5-G	35	6	$(G_5C)_5G_5$	>24.6	>11.4	>29.3
ZNF480-G	42	7	$(G_5C)_6G_6$	>20.1	>14.1	>24.2

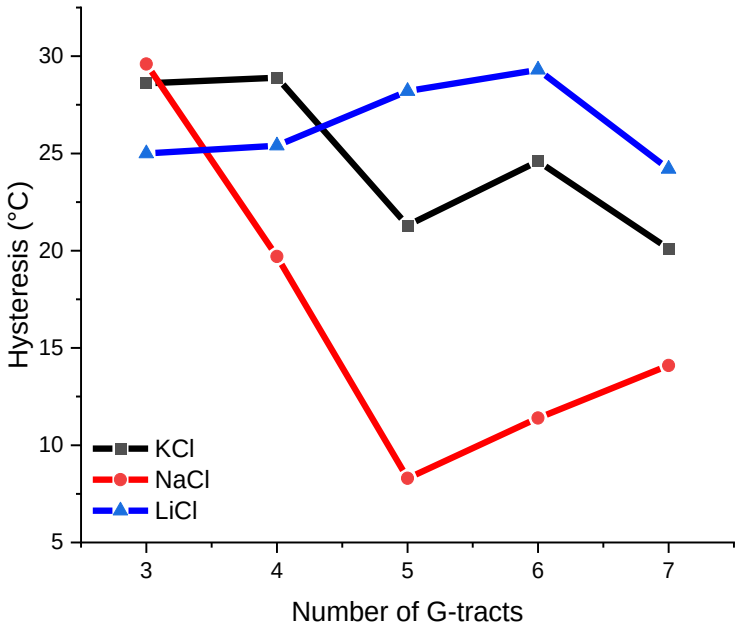


Figure 5.21 *The effect of G-tract number on estimated hysteresis in the presence of K^+(black), Na^+ (red), and Li^+ (blue).*

The hysteresis shows a slow association and/or slow dissociation process, The kinetics of intermolecular structure association will also depend on the DNA concentration. All these G-quadruplex structures are very stable, which is unsurprising as in previous studies with G-rich sequences with smaller tract lengths, have been shown to be exceptionally stable. For example, the Fox group identified the sequence $(G_3T)_4$ is exceptionally stable even in 1 mM potassium solution and doesn't form a double stranded structure with its complementary

strand.[65] From the hysteresis data, in lithium cations the hysteresis remains approximately constant, when varying the number of repeats. Whereas in sodium and potassium cations, the, hysteresis decreases from C-tract repeats n= **3** to **5**, then plateaus. The hysteresis showed that T_m from the melting curve is always higher than T_a, this is indicative that the DNA molecules are not at thermodynamic equilibrium due to the slow folding and unfolding processes.

5.4.3 Thermal Difference Spectra of G-Rich Sequences

After investigating the thermal stability of G-rich sequences in different cations, then TDS was performed to further characterize what structures they fold into. UV absorbance at 5°C and 95°C were measured and UV difference was calculated. TDS of G-rich sequences in K^+ are shown in ***Figure 5.22***.

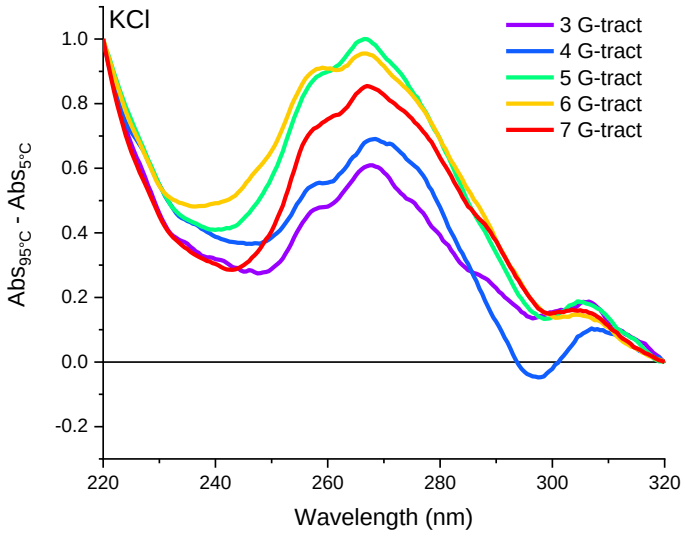


Figure 5.22 *Thermal difference spectra (TDS) calculated between 95 °C and 5 °C for each G-rich DNA sequence at pH 7.0 in 10 mM sodium cacodylate and 100 mM potassium chloride buffer, the concentration of oligonucleotide is 2.5 μM.*

From the graph, although all the sequences had positive and negative peaks at ~270 and 295 nm respectively, the spectrum shapes were quite different from the literature, this might be because from the UV melting experiments, these G-rich sequences were not completely melted, which means when taking TDS at 95°C,

the sequences were not unfolded completely (refer to ***Figure 5.19 a***). The reason that G-tract sequences did not show clear G-quadruplex TDS could be too much K^+ in the solution, so the thermal stability of G-quadruplex is beyond what can be measured under these conditions. Therefore, an additional TDS was performed, this time, all the sequences were prepared in 10 mM sodium cacodylate and 25 mM potassium chloride at pH 7.0 (***Figure 5.23***).

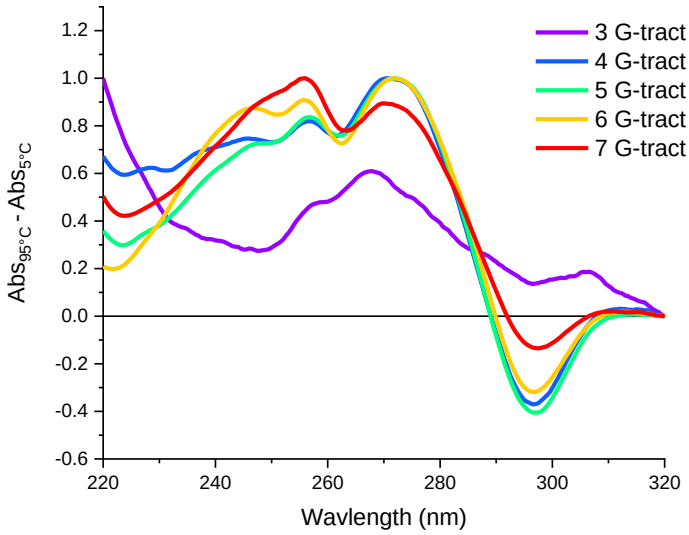


Figure 5.23 *Thermal difference spectra (TDS) calculated between 95 °C and 5 °C for each G-rich DNA sequence at pH 7.0 in 10 mM sodium cacodylate and 25 mM potassium chloride buffer, the concentration of oligonucleotide is 2.5 μM.*

The samples in the lower concentration of K^+, **n=4-7** G-rich sequences showed a featured positive peak at ∽273 nm and a negative one at 295 nm,[167] indicating the formation of G-quadruplex. n = **3** presented a different TDS shape compared to others, suggesting it may form other DNA species. However, the CD spectra of n= **3** showed it formed a parallel G-quadruplex structure, but these experiments were performed at a different concentration of DNA and higher cation concentration (100 mM).

Then TDS were performed in the presence of Na^+ and Li^+ buffers (***Figure 5.24***), and all the G-rich sequences exhibited a positive peak ~273 nm and a negative one

~295 nm,[167] this is indicative of G-quadruplex structures formed in both cation conditions.

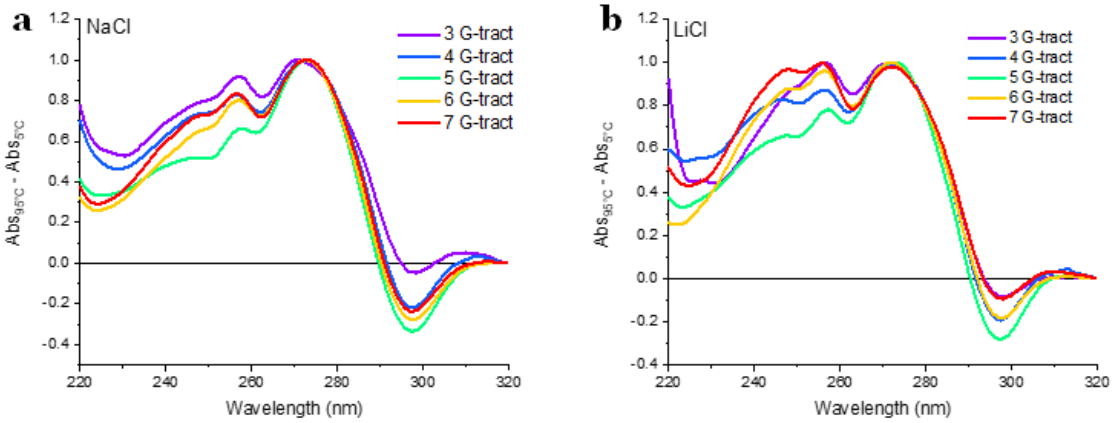


Figure 5.24 *Thermal difference spectra (TDS) calculated between 95 °C and 5 °C for each G-rich DNA sequence at pH 7.0 in 10 mM sodium cacodylate and 100 mM a). sodium chloride b). lithium chloride buffer, the concentration of oligonucleotide is 2.5 μM.*

Overall based on the current TDS results, in the presence of 100 mM Na^+ and Li^+, all the G-rich sequences can form G-quadruplex structures, in 100 mM K^+, n= **4–7** can adopt G-quadruplex structures, but n= **3** may adopt different DNA secondary structures depending on the DNA concentration. However, to determine the exact structure they form, additional experiments using X-ray crystallography or NMR[266] are required.

5.5 Exploring *ALOX5* Binding Ligands

After finishing the characterization of selected GC-repeat sequences, the next step was investigating potential ligands that bind to the i-motif sequence. As the *ALOX5* i-motif was the first time it was identified, there are no published ligands bound to this i-motif. To get a suitable starting point, we decided to screen the NCI Set VI library.

FID assays were used for this project, as a high throughput screening technique was that it enabled the screening of a huge number of ligands. The conditions for the screening experiments were optimized by myself and the screening was performed by Robert Yuan, under my supervision. The FID screening experiments were performed with the whole library of 1584 small molecules (***Figure 5.25***).

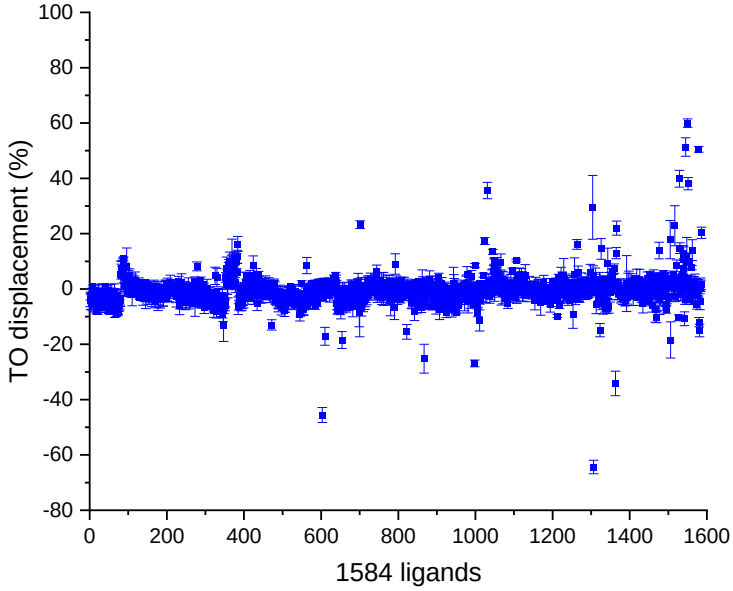


Figure 5.25 *Results of TO displacement of ALOX5 i-motif against NCI Set VI library. All the experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride buffer at pH 7.0, the concentration of DNA was 0.5 μM, the concentration of TO was 1 μM, and the concentration of ligands was 2.5 μM (5 equivalents).*

From the results it can be observed that seven compounds gave a displacement of 50% or above, 1 compound gave a displacement between 40% to 50%, four compounds showed a displacement between 30% to 40%, six compounds had 20% to 30% displacement and thirty-one compounds showed 10% to 20% displacement. Given the number of ligands in the whole library, and the fact that the most potent compounds may not be the most specific[115] it was decided that hits would be defined as anything that gave a percentage displacement of 10% or above. However, there were also some ligands with high *negative* displacement values. This was observed also with the minor groove binding ligands in **Chapter 2**. This could indicate either ligands show fluorescence themselves. However, it is

important to consider that folding of DNA structures is an equilibrium, and depending on the conditions, the equilibrium will be a mixture of folded, partially folded and unfolded species. For i-motifs in particular there is often some unfolded species, due to the variability between individual i-motifs, which have different transitional pHs. A ligand could shift the equilibrium between single-strand DNA and folded i-motif structure, therefore increasing the amount of TO capable of binding DNA and therefore increasing the fluorescence. Therefore, an additional control experiment was also performed to determine whether ligands are fluorescent, by examining their fluorescence in buffer only. However, it is also possible that ligands could act as fluorescent probes similar to TO, and become fluorescent on binding DNA. Then 34 compounds that displayed more than 10% displacement were finally selected as potential ALOX5 i-motif binding hits (***Figure 5.26***). ***Table 5.10*** presents the structure of hits and the percentage of TO displacement.

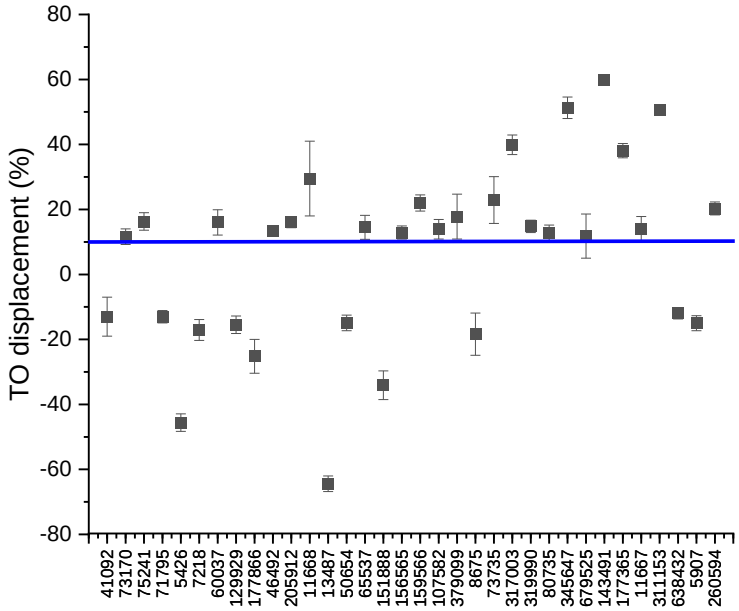
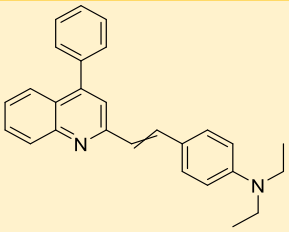
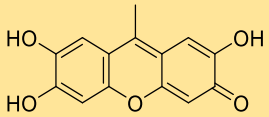
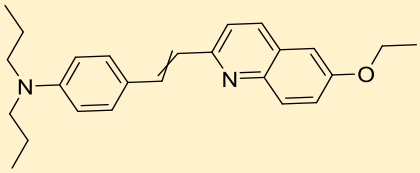
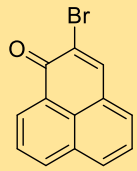
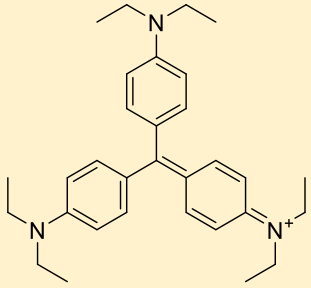
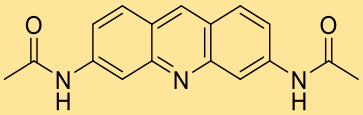
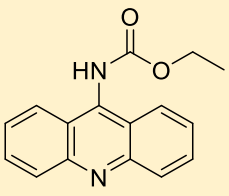
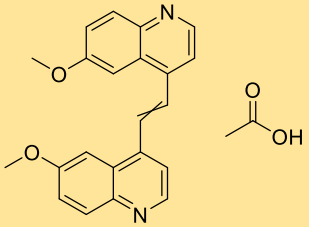
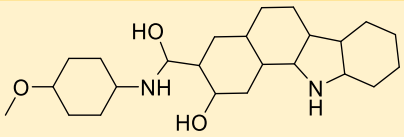
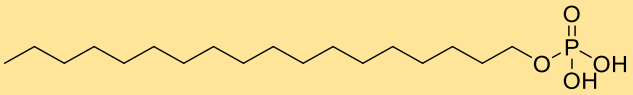
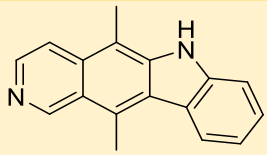
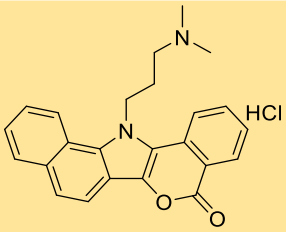
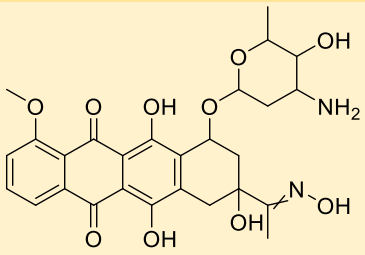
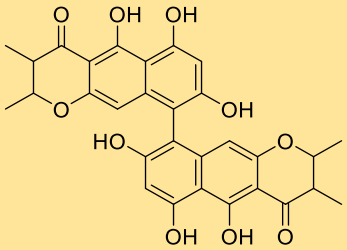
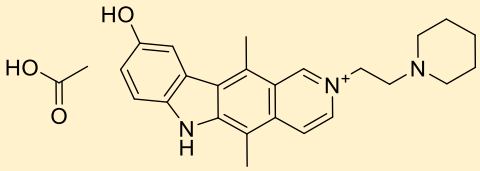
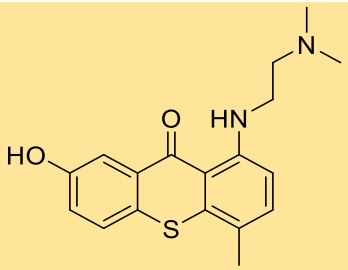
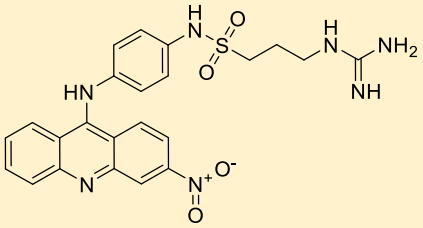
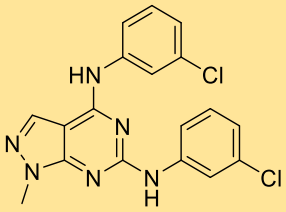
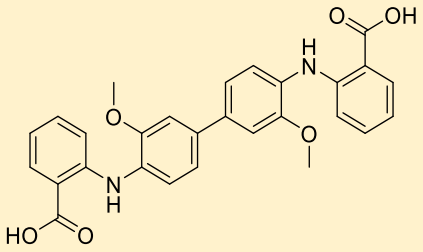
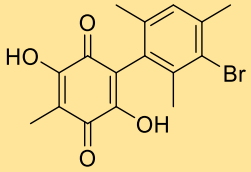
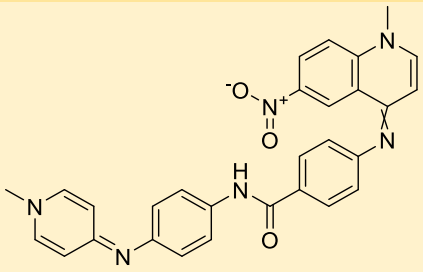
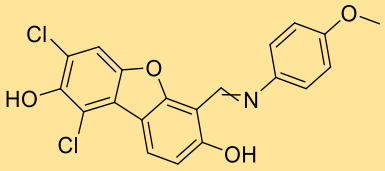


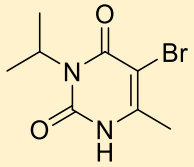
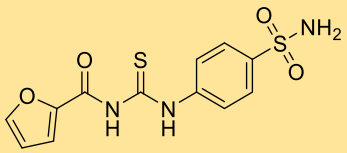
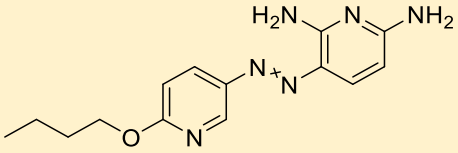
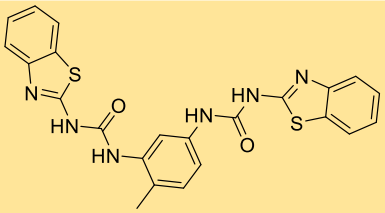
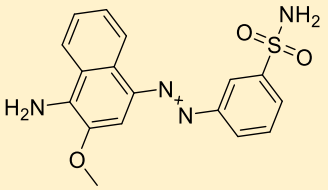
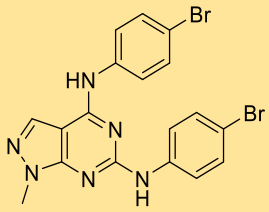
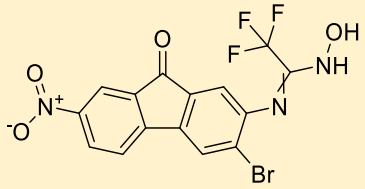
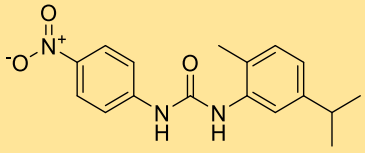
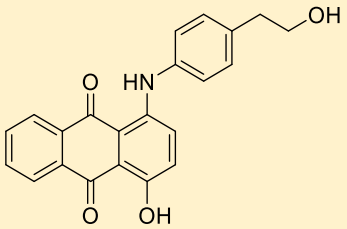
Figure 5.26 *Hits from FID screening, all the compounds showed over either positive or negative 10% TO displacement.*

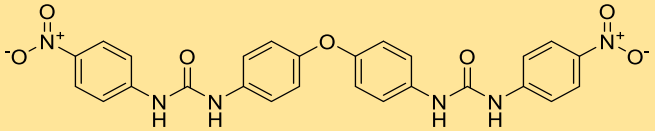
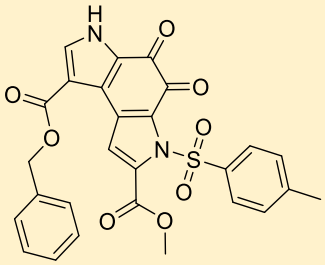
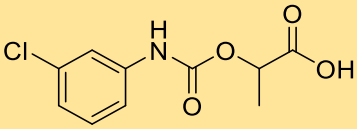
Table 5.10 *Structures of hits from NCI library and percentage of TO displacement against the i-motif forming sequence from ALOX5. All experiments were performed in 10 mM sodium cacodylate and 100 mM KCl at pH 7.0, [ALOX5-C] was 0.5 μM, [TO] was 1 μM, and [ligands] was 2.5 μM.* Positive displacement values represent displacement of TO from the DNA. Negative displacement represents enhanced fluorescence on addition of the ligand.

NSC	Structure	TO Displacement (%)
13487		-64.4 ± 2.4
5426		-45.6 ± 2.7
151888		-34.1 ± 4.4
177866		-25.2 ± 5.2
8675		-18.4 ± 6.5
7218		-17.1 ± 3.2
129929		-15.5 ± 2.7

5907		-15.0 ± 2.3
50654		-14.9 ± 2.4
41092		-13.0 ± 6.0
71795		-13.0 ± 1.9
638432		-12.0 ± 1.7
143491		60.0 ± 1.5
345647		51.3 ± 3.3
311153		50.5 ± 1.0

317003		39.9 ± 3.0
177365		38.1 ± 2.2
11668		29.5 ± 11.5
73735		22.9 ± 7.2
159566		22.0 ± 2.5
260594		20.3 ± 2.0
379099		17.8 ± 6.9

75241		16.3 ± 2.7
205912		16.1 ± 1.7
60037		16.0 ± 3.9
319990		14.8 ± 1.9
65537		14.5 ± 3.7
11667		14.0 ± 3.8
107582		13.9 ± 3.0
46492		13.4 ±0.9
156565		12.9 ± 2.0

80735		12.7 ± 2.5
679525		11.8 ± 6.8
73170		11.6 ± 2.4

From the ranking of the hits, 12 ligands were found to show negative displacement of TO, which is indicative of those ligands may either fluorescent or can shift the equilibrium between single strand and folded i-motif structure, so that when the ligands were added to *ALOX5-C* solution, more i-motif species could form. **NSC 12487** was found to be the highest negative TO displacement ligand for *ALOX5* with -64% TO displacement, compared to other ligands which showed negative TO displacement. **NSC 12487** seems significantly induce more i-motif species formed. **NSC 143491** (daunomycin) showed the highest TO displacement with 60%, suggesting high binding affinity to *ALOX5* i-motif, interestingly, in **Chapter 3**, this compound has been found to destabilize i-motif forming sequence from the human telomere (ΔT_m = -2.4°C). **NSC 143491** is an anticancer medication, especially used for AML,[354] thus it could be a good example for the further study of interaction with *ALOX5* quadruplex DNAs. Ligand **143491** is a known and well characterised duplex binding drug, daunomycin, which binds to the minor groove by intercalation.[355] It has also been shown to interact with G-quadruplex DNA.[356] Ligand **319990** (ΔT_m = -2.4°C) was previously identified to have a destabilizing effect on *hTelo*C i-motif, as discussed in **Chapter 3**, also demonstrated more than 10% TO displacement (15%) against the *ALOX5* i-motif. Compound **71795** showed 13% displacement of TO; this compound has already been identified as interacting with the i-motif forming sequence from the human telomere,[63] and was also

studied in **Chapter 3**, where it was shown to have a stabilizing effect on the i-motif *hTelo*C. **NSC 71795** has already been shown to have a potential preference for i-motif structures, this shows that it could potentially be a useful molecule in studying the biology of *ALOX5*. **NSC 260594** showed 20% TO displacement to *ALOX5* i-motif, while in **Chapter 3**, from both CD titration and melting experiments, **260594** did not affect *hTeloC* i-motif, indicating this ligand may show some selectivity to the *ALOX5* i-motif structure, but these experiments were also performed at different pHs.

5.6 Conclusions

In the first part of this chapter, we used bioinformatics to investigate interesting i-motif forming sequences from the human genome. Eight leukaemia-related genes were selected, as they all partially overlap Sp1 transcription factor binding sites, which suggested they may be possible sites for modulation of gene expression through altering the DNA structures. The sequence from the promoter region of *ALOX5* was the first to be studied, and other analogous C-rich repeat sequences with different number of repeats were also investigated. CD and UV were used to investigate their pH and thermal stability. The study showed all these C-rich sequences have two transitional pHs, indicating they all potentially exist in at least one different population in solution. UV melting experiments at different concentrations suggested the DNA C-rich sequences with n= **4** to **7** C-tracts are more likely to adopt intramolecular species, whereas n= **3** does not but can form intermolecular i-motif structures. TDS showed all these sequences can fold into i-motif structures at acidic conditions, however, at neutral pH, n = **3** was not able to form an i-motif structure and may form alternative DNA conformations, such as a hairpin.

The complementary G-rich sequences were also studied. CD spectra were performed in different cations, and aside from the G-rich sequence with n = **3** G-tracts, the rest of the G-rich sequences were shown to adopt both parallel and

antiparallel structures in the presence of K^+, Na^+, and Li^+. As the buffer conditions were performed in sodium cacodylate to compare with the i-motif sequences, this resulted in the use of mixed cationic conditions in some cases. To remove the potential competition from the sodium within the cacodylate, all the CD spectra were repeated in lithium cacodylate buffer. In these conditions we observed that when increased the number of G-tracts, the proportion of parallel G-quadruplex species also increased. The TDS results clearly showed that these sequences form G-quadruplex topologies in the presence of 100 mM Na^+ and Li^+. The TDS shapes in 100 mM K^+ seemed abnormal, but this is because the G-rich sequences are thermally stable in K^+ even at high temperatures so at 95°C, the structures are not completely unfolded. However, when the reduced concentration of K^+ to 25 mM, n= **4–7** showed G-quadruplex features, although even at this concentration of potassium cations, the G-quadruplexes may still not fully melt. Regardless, it suggests all the G-rich sequences adopt G-quadruplex species.

As both the G-rich and C-rich sequences fold into alternative DNA secondary structures this does pose some questions with regard to what Sp1 and Egr1 transcription factors bind to. It is well known that Sp1 can bind G-quadruplex DNA.[355] The results here indicate that above three tracts, the C-rich sequences adopt stable i-motif structures at neutral pH. Further work to determine the binding properties of Egr1 against these sequences is now required to determine whether this transcription factor binds the i-motif itself, rather than the double-helical DNA sequence.

The initial ligands that have been screened showed many potential hits for *ALOX5*. We do not yet know whether they are specific for the *ALOX5* i-motif, and whether they also bind to other types of DNAs, thus other sequences need to be examined and complementary techniques need to be performed. Among these hits, **NSC 71795**, **319990**, and **143491** also showed interaction with human telomeric i-motif as discussed in **Chapter 3**, which suggests these four might have some sort of preference to bind i-motif structures. The next steps would be to investigate their effect on the thermal stability of the i-motif structure. Thus, FRET or CD melting experiments should be carried out. These ligands could potentially be used for

disrupting protein-DNA interactions between Egr1 and *ALOX5* and they could be used for modulation of *ALOX5* gene expression. This work is now being continued by a different Ph.D. student within the group, who is now characterizing the ligands further, expressing Egr1 for binding experiments, and making a reporter-gene construct to study the effects of GC-tract length on gene expression as well as the effects of the hit ligands.

Aside from *ALOX5*, from the original bioinformatics search, there were other eight sequences involved in leukaemia that have been identified (refer to ***Table 5.***2) and these could also be studied further.

Chapter 6

Conclusion and Future Work

6.1 Conclusions

Initially, my Ph.D. study aimed to investigate and design a novel specific, and selective i-motif ligand, and to study interactions between ligands and DNA secondary structures. At the current stage, although we only made one minor groove binder **PyPyPyβDp**, it was originally hoped to be the part that would bind a double-helical part of DNA but actually in the end we found it could bind i-motif. This did mean it was unsuitable for the intended purpose, but it was the first time that a minor-groove binder was identified to interact with i-motif. Future work to make i-motif-ligand conjugate should start with hairpin polyamides for the proof of concept as they would be more specific to target the minor groove. However, it may be that even these compounds have i-motif binding capabilities, only testing them would reveal whether this is the case. Regardless, it was clear that more i-motif interacting compounds were required before this was able to work effectively as a strategy.

Then my work moved to the study of interactions between ligands and DNA structures. Plenty of small molecules were examined with different DNA and RNA structures: double helical, G-quadruplex, and i-motifs. A variety of biophysical techniques was used to explore the interaction between ligands and oligonucleotides including FRET, FID, CD, and UV. After examining the top 28 NCI compounds from a bioinformatics study using CD titration and CD melting experiments, two ligands **NSC 19990** and **71795** showed a strong stabilization effect on *hTelo*C i-motif. These ligands could be used as potential agents for i-motif-conjugate designing. Many of the others had a destabilizing effect on the *hTelo*C i-motif. Some other ligands from collaborating groups were found to have interesting interactions with i-motif forming sequences, such as Pd complex **22** showed a preference for *hTelo*C i-motif, also the **Fe-cylinder** strongly unfolds and destabilizes the i-motif forming sequence from Hif1α. The fact that we had studied a lot of destabilizing compounds gave rise to the opportunity to study the unfolding effects of **TMPyP4** with RNA G-quadruplex. The biophysical evidence of unfolding *PQS18-1* RNA G-quadruplex structure was able to support the computational results performed by the Haider group. TMPyP4 is an interesting molecule that

continues to be used in the literature as a G-quadruplex stabilizing agent, though these results raise further questions as to whether this is a good ligand to use for such experiments. It does not seem that one can make assumptions as to what TMPyP4 should do with nucleic acids and should be used with caution or supported with significant biophysical data.

The project performed in **Chapter 5** was the study of GC-rich repetitive sequences present across the human genome. Bioinformatics methods were used to align Sp1 transcription factor and GC-sequences from 637 i-motif-sequences lists, from the alignment results, nine sequences related to leukaemia were selected and the sequence from the promoter region of *ALOX5* was selected for characterization. Some previous experiments were reanalyzed and got more accurate transitional pHs and it was possible to observe that once a range of sequences was characterized, the previously published sequences from *DAP* and *MSMO* showed single transitions for their transitional pHs. For all the C-rich sequences, two transitional pHs were observed, indicating they all can adopt more than one DNA secondary structure population in solution. Thermal stability experiments suggested melting temperature increased with an increasing number of C-tracts. UV melting performed with different oligonucleotide concentrations showed no variability in T_m and T_a, indicating the dominant structures of sequences n = **4** to **7** are likely to be intramolecular i-motif whereas n= **3** was found to form intermolecular i-motif structures. Characterization for G-rich sequences showed an interesting observation that with increased G-tract, parallel topology also increased. During the FID screening experiments, a number of NCI compounds were found to bind to *ALOX5* i-motif, among the hits, three ligands (**NSC 71795**, **319990,** and **143491**) were also found to interact with *hTelo*C in **Chapter 3**, thus they could have potential selectivity to i-motif structures. Further work to characterize these ligands further is necessary to explore their utility in altering *ALOX5* gene expression.

6.2 Future Work

Firstly, in the study of designing a conjugate that specifically binds to i-motif, as our two minor groove binders interact with i-motif structures, thus future work should focus on Dervan-style hairpin polyamides minor groove binders. Also, as discussed in **Chapter 2**, there are other two strategies, which will be worth investigating in the future: ligand-ligand conjugates to target i-motifs that exist in tandem and ligand-nucleic acid conjugates to combine strong i-motif stabilizers with sequence-specific DNA recognition. These types of conjugates are unexplored in the i-motif field so would represent a significant advance. Our work here has shown that the design may be straightforward, but the reality is much more complicated.

Secondly, in the ligands study, ligands **NSC 71795** and **19990** were proved to be potential stabilizers to *hTelo*C i-motif, however, we have not examined them with other DNA secondary structures, thus this should be performed. Other techniques could be used to study their effect on DNA structures, such as ITC and SPR can be used to directly study the binding affinity of **71795** and **19990** with oligonucleotides. **Fe-cylinder** showed unfolding and destabilizing effects on the i-motif forming sequence from *Hif1a*, but we have not yet investigated its effect on other DNA structures, thus CD melting experiments should be performed with other DNAs. We already know the unfolding mechanism of **TMPyP4** to *PQS18-1* RNA G-quadruplex, and also that it interacts very differently with *PQS18-1* DNA G-quadruplex. Work is required to determine exactly why the effects are different between RNA and DNA structures. As this has not been deeply studied, future work could focus on the differences between DNA and RNA, computational work should be performed to assist the biophysical experiments.

As the C-rich repeated sequences were found to be stable at neutral pH, it may be worth using X-ray crystallography to determine their respective structures. This may reveal insights into the different conformations and how they vary with the number of C-tracts and whether these n= **3**–**7** sequences fold into intermolecular or intramolecular i-motif structures. Additional experiments using gel-

electrophoresis will also help in this aspect. Then NMR and X-ray crystallography[266] should be performed to study the precise structure of G-rich sequences. this greater understanding of the GC-rich structures, their dynamics, and their binding to Sp1 and Egr1 transcription factors are important in the development of ligands to target *ALOX5* for future cancer treatments.

Chapter 7

Experimental

7.1 General Experimental

Chemical reagents were general-purpose grade and unless otherwise indicated, purchased from Sigma-Aldrich or Thermo Fisher, and used without further purification.

Oligonucleotides were purchased from Eurogentec, and each of them is dried and purified using reverse phase HPLC. The dry DNA was dissolved in ultrapure water to prepare a stock solution, the labelled DNA samples were prepared at approximately 100 μM and the unlabelled DNA was made at 1 mM. The concentration of the stock solutions was identified from their UV absorbance at 260 nm with a NanoDrop by using extinction coefficients provided by Eurogentec.

DNA annealing: The DNA samples were prepared at final concentration and annealed in the heating block for all experiments. The DNA samples were always prepared in the final buffer. When using the heating block, the temperature was increased to 95°C (1°C/min) and then DNAs were put into the block for 5 minutes then it was switched off, and the samples were left in it cooling down to room temperature overnight with a cooling rate of around 1°C/min.

Table 7.1 *Custom oligonucleotide sequences used throughout this research.*

Name	Sequence 5′ → 3′
*hTelo*C	TAA-CCC-TAA-CCC-TAA-CCC-TAA-CCC
*hTelo*G	GGG-TTA-GGG-TTA-GGG-TTA-GGG
DS	GGC-ATA-GTG-CGT-GGG-CGT-TAG-C
DS_{comp}	GCT-AAC-GCC-CAC-GCA-CTA-TGC-C
Hif1a	CGC-GCT-CCC-GCC-CCC-TCT-CCC-CTC-CCC-GCG-C
DAP	CCC-CCG-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CC
ATXN2L	CCC-CCC-CCC-CCC-CCC-CCC-CCC-CCC
Nrf2	CCC-TCC-CGC-CCT-TGC-TCC-CTT-CCC
c-MYC	CCT-TCC-CCA-CCC-TCC-CCA-CCC-TCC-CCA
$PQS18\text{-}1_{RNA}$	GGC-UCG-GCG-GCG-GA
$PQS18\text{-}1_{DNA}$	AGG-CTC-GGC-GGC-GGA
$PQS18\text{-}1_{RNAFRET}$	FAM-GGC-UCG-GCG-GCG-GA-TAMRA
$hTeloC_{FRET}$	FAM-TAA-CCC-TAA-CCC-TAA-CCC-TAA-CCC-TAMRA
$hTeloG_{FRET}$	FAM-GGG-TTA-GGG-TTA-GGG-TTA-GGG-TAMRA
DS_{FRET}	FAM-TAT-AGC-TAT-A-HEG(18)-TAT-AGC-TAT-A-TAMRA
$Hif1a_{FRET}$	FAM-CGC-GCT-CCC-GCC-CCC-TCT-CCC-CTC-CCC-GCG-C-TAMRA
DAP_{FRET}	FAM-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CC-TAMRA
$ATXN2L_{FRET}$	FAM-CCC-CCC-CCC-CCC-CCC-CCC-CCC-CCC-TAMRA
$Nrf2_{FRET}$	FAM-CCC-TCC-CGC-CCT-TGC-TCC-CTT-CCC-TAMRA
$c\text{-}MYC_{FRET}$	FAM-CCT-TCC-CCA-CCC-TCC-CCA-CCC-TCC-CCA-TAMRA
MADD-C	CCC-CCG-CCC-CCG-CCC-CC
MSMO1-C	CCC-CCG-CCC-CCG-CCC-CCG-CCC-CC
ALOX5-C	CCC-CCG-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CC
ZNF480-C	CCC-CCG-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CC

MADD-G	GGG-GGC-GGG-GGC-GGG-GG
MSMO1-G	GGG-GGC-GGG-GGC-GGG-GGC-GGG-GG
DAP-G	GGG-GGC-GGG-GGC-GGG-GGC-GGG-GGC-GGG-GG
ALOX5-G	GGG-GGC-GGG-GGC-GGG-GGC-GGG-GGC-GGG-GGC-GGG-GG
ZNF480-G	GGG-GGC-GGG-GGC-GGG-GGC-GGG-GGC-GGG-GGC-GGG-GGC-GGG-GG

Mass spectroscopy

MALDI was performed on Kratos Analytical Axima MALDI-TOF. The matrix, α-cyano-4- hydroxycinnamic acid (>98% Sigma), was prepared in ethanol at a concentration of 10 mg/mL. Compounds to be analyzed were prepared at a concentration of 1 mg/mL in methanol. Samples were prepared by mixing equal volumes (5 μL) of matrix and compound. The MALDI experiment was performed in reflectron mode under positive ionization conditions. The following parameters were used: 1000 profiles per sample, shots off, ion gate off, pulsed extraction on, and max laser rep rate 10.

Chromatographic Techniques

Preparative RP-HPLC was performed on an Agilent 1260 infinity using an Agilent Eclipse XDB-C18 column, 21.2 x 150mm, 5μM, and a flow rate of 20 mL/min. Solvent A = 5% MeOH in Water + 0.05% TFA and Solvent B = 5% Water in MeOH + 0.05% TFA. Gradient 0% B → 100% B over 20 minutes. Detection wavelengths: 214 and 254 nm.

FRET melting

Experiments were carried out on a Qiagen Roto-gene Q-PCR machine using Qiagen PCR tubes and a 7500 Fast Real-Time PCR system, 18 μL of 0.2 μM of oligonucleotide sample was prepared in MicroAmp Optical 96-well Reaction Plate, then 2 μL of 10 μM ligand was added into each well, covered by clear polyester sealing film. All experiments were repeated in triplicate. The samples were then heated in the PCR machine from 25 to 95 °C and the fluorescence of the FAM group was measured at 510 nm using a green filter at 1 °C interval. This was repeated for each ligand and with G-quadruplex and duplex DNA sequences. The raw data gives the fluorescence melting curves, and the melting temperatures were calculated by first the derivative.

FID assay

The FID assays were performed on a BMG CLARIOstar Plate Reader using Corning 96-Well Solid Black Flat Bottom plates. A 10 mM stock solution of TO was prepared in DMSO and dilution to 1 / 2 μM (depending on the concentration of DNA) in the appropriate buffer, a ratio of TO with DNA is 2:1. 90 μL of the 2 μM TO solution was added to each well and fluorescence emission at 450 nm was measured with excitation at 430 nm; this was normalized to 0% representing background fluorescence. 1 μL of 45 / 90 μM DNA was then added, shaken using 114 double orbitals shaking at 700 rpm in the plate reader for 30 seconds, and allowed to equilibrate for 15 - 20 minutes (depending on oligonucleotide sequence), then fluorescence emission was measured again, and values are normalized to 100% of the maximum fluorescence from the thiazole orange binding to DNA structures. Then 2.25 μL of 100 μM ligands were titrated into each well and the fluorescence was measured immediately as mentioned above, each experiment was performed in triplicate. The fluorescence after the addition of ligands was normalized between 0 and 100% for each independent well. The percentage of TO displacement was the difference calculated between normalized 100% fluorescence and normalized fluorescence measured after the addition of ligand.

CD spectroscopy

CD titration and CD melting experiments were performed on Jasco J-810 or Jasco J-1500 spectropolarimeter, oligonucleotide samples were performed in a 1 mm quartz cuvette with a silicone lid. CD scans were performed with a scanning speed of 100 (**Chapter 3 Section 3.2**) or 200 nm/min (the rest of the chapters), response time of 1 s, 0.5 nm pitch, and 2 nm bandwidth. The spectra are an average from 4 scans, and all are zero corrected at the longest wavelength of the measurement. CD melting experiments were carried out with the same parameters, each sample was heated from 5°C to 95°C at a rate of 1°C/min and measured at 5°C intervals. The melting temperature (T_m) of half unfolded was calculated using OriginPro software, and normalized ellipticity at 288 nm against temperature was plotted and fitted using sigmoidal curve fitting methods.

UV spectroscopy

Unless otherwise indicated, all UV spectroscopy experiments were performed on Jasco V-750 Spectrophotometer equipped by PAC 743R Temperature Controller, using a quartz cuvette with a 1 cm path length. Oligonucleotide samples were diluted to 2.5 μM in indicated buffer at the desired pH, then thermally annealed. 250 μL of the sample was transferred to a cuvette with a silicone lid. The UV absorbance was measured at 295 nm, the sample was held at 5°C for 10 min and then heated to 95°C at a rate of 0.5°C/min, then held at 95°C for 10 min and the process was reversed, each melting and annealing process was repeated three times. Data was measured every 1°C during both the heating and cooling state, each data point was an accumulation of three scans. Melting and annealing temperatures were determined using the first derivative method. Thermal difference spectra were the difference calculated by subtracting the spectrum of folded structure at 5°C from the spectrum of unfolded structure at 95°C, the spectrum wavelength was 220 to 320 nm at the scanning speed of 200 nm/min, and the data was normalized, and the maximum change of spectrum was set to 1.

FRET titration

FRET titration experiments were performed in an Edinburgh Instruments FS5 spectrofluorometer, using a quartz cuvette with a 10 mm path length. The sample volume was 250 μL. The FRET-labelled *PQS18-1*RNA (FAM-GGC-UCG-GCG-GCG-GA-TAMRA) was diluted in pH 7.0 buffer to 0.1 μM. The sample was excited at 490 nm, and the fluorescence emission was measured from 500 to 650 nm. TMPyP4 was diluted in buffer and added into a sample in 0.025 μM increments from 0.025 to 0.5 μM. The relative FRET efficiency ($E_{rel.}$) was calculated using the equation 3:

$$\textbf{\textit{Equation 3}}: \mathrm{E_{rel.}} = \frac{I\mathrm{a}}{I\mathrm{d} + I\mathrm{a}}$$

where I_d is the fluorescence intensity of the donor and I_a is the fluorescence intensity of the acceptor. The experiment was performed in triplicate, and the

error bars represent the standard deviation. The data were analyzed by OriginPro 2020.

7.2 Experimental Section for Chapter 2

Synthesis: The building blocks chlorotrityl chloride resin preloaded with ß alanine (**1**), Fmoc 1*H*-Pyrrole-2-carboxylic acid (**2**), 1-Methyl-1*H*-pyrrole-2-carboxylic acid (**7**) and 3-Dimethylaminopropylamine (**9**) were purchased from Merck.

FRET melting experiments were performed as described above, all experiments were performed in a 7500 Fast Real-Time PCR system. Labelled DNA samples were diluted in 10 mM sodium cacodylate and 100 mM potassium chloride buffer, the pH used for *hTelo*C is 6, *Nrf2* was in pH 6.5, *dsDNA*, *hTelo*G, *DAP*, *ATXN2L*, and *Hif1a* were prepared in pH 7.0. 1 mM ligands stock was diluted into 10 µM with distilled water. For control experiments, 2 µL of distilled water was added to DNA.

The **FID assays** were performed on a BMG CLARIOstar Plate Reader and measured as discussed above, buffers were used the same as FRET melting, and ligands were prepared in each well as 1, 2.5, 5, 10, 20, 50, and 100 µM.

CD experiments were carried out as described above. 10 µM of *hTelo*C and *DAP* in 10 mM sodium cacodylate and 100 mM potassium chloride buffer at pH 7.0, 1 µL of 1 mM ligand was titrated into DNA until 50 µM, CD scans measured between 230 to 320 nm, then melting experiments were performed at such concentration and measured the same wavelength. All titration data were ligand subtracted, and both titration and melting data were zero corrected at 320 nm.

7.3 Experimental Section for Chapter 3

Compounds: All the NCI compounds were purchased from NCI SET VI. All the bis(pyridyl)allene metal complexes were synthesized by Dr. Hanna K. Maliszewska. The iron cylinder was a Kind gift from Prof Mike Hannon.

FRET melting experiments were performed as described above, all the experiments were performed on a Qiagen Rotor-Gene Q-series PCR machine. The final analysis of the data was carried out using Qiagen Rotor-Gene Q-series software and Origin or Excel.

FID assays were performed on a BMG CLARIOstar Plate Reader as described above.

CD experiments were carried out on a Jasco J-810 spectropolarimeter (**Section 3.2**) and a J-1500 spectropolarimeter (**Section 3.3** and **3.4**) as described above.

7.4 Experimental Section for Chapter 4

Oligonucleotides purchased from Eurogentec (*PQS18-1* RNA, *PQS18-1* DNA, and *PQS18-1* FRET). The dry RNA/DNA was dissolved in nuclease-free water to prepare stock solutions. Further dilutions were carried out in a buffer containing 10 mM lithium cacodylate and 100 mM potassium chloride, pH 7.0.

CD experiments were performed with the JASCO J-1500 spectropolarimeter as described above. The scanning range for CD titration was 200–650 nm, the other parameters were the same as described above. Unfolding data were fitted using the Hill equation $\theta = C^n/(DC_{50}^n + C^n)$ where θ = fraction of bound TMPyP4, C = concentration of TMPyP4, n = Hill coefficient, and DC_{50} = the half denaturing concentration.

Corresponding UV–vis absorption data were also collected at the same time as the CD. UV–vis binding data were fitted to a two-inequivalent-sites binding model: $\theta = (K_1C + 2K_1K_2C^2)/(1 + K_1C + 2K_1K_2C^2)$ where θ = fraction of bound TMPyP4, C = concentration of TMPyP4, and K_1 and K_2 are the association constants for the first and second sites.

CD/UV–vis melting/annealing profiles were recorded by monitoring the CD at 264 nm, as a function of temperature using a sealed cuvette. RNA samples were cooled to 5 °C and then heated to 95 °C with a heating rate of 1 °C/min, held at 95 °C for 5 min, and then cooled to 5 °C at the same rate. The T_m values of the complexes were calculated by normalizing the experimental curves to give fraction folded, and the data fitted to sigmoidal fittings were used to determine the T_m values. All experiments were repeated twice, and the error represents the standard deviation from the mean. Final data were analyzed in OriginPro 2020.

FRET titration experiments were performed in an Edinburgh Instruments FS5 spectrofluorometer, using a quartz cuvette with a 10 mm path length. The sample volume was 250 µL. The FRET-labelled RNA was diluted in pH 7.0 buffer to 0.1 µM. The sample was excited at 490 nm, and the fluorescence emission was measured from 500 to 650 nm. TMPyP4 was diluted in buffer and added into a sample in 0.025 µM increments from 0.025 to 0.5 µM. The relative FRET efficiency (E_{FRET}) was calculated using $E_{FRET} = (I_a/I_d + I_a)$ where I_d is the fluorescence intensity of the donor and I_a is the fluorescence intensity of the acceptor. The experiment was performed in triplicate, and the error bars represent the standard deviation. The data were analyzed by OriginPro 2020.

7.5 Experimental Section for Chapter 5

CD spectroscopy experiments were performed on a Jasco J-810 spectropolarimeter. DNA samples were diluted to 10 µM (total volume: 100 µl) in buffer at pH increments of 0.5 pH unit from 4.0 to 9.0 (including pH 6.25 and pH 6.75). The scans were recorded at room temperature (20°C) between 200 and 320 nm. The parameters were used the same as described above.

UV spectroscopy experiments were performed on Jasco V-750 Spectrophotometer as described above, both melting and annealing temperatures were determined by the first derivative method

References

1. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).

2. Chargaff, E., Zamenhof, S. & Green, C. Composition of human desoxypentose nucleic acid. *Nature* **165**, 756–757 (1950).

3. Franklin, R. E. & Gosling, R. G. Molecular configuration in sodium thymonucleate. *Nature* **171**, 740–741 (1953).

4. Wilkins, M. H. F., Stokes, A. R. & Wilson, H. R. Molecular structure of deoxypentose nucleic acids. *Nature* **171**, 738–740 (1953).

5. Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet* **13**, 770–780 (2012).

6. Amato, J., Iaccarino, N., Randazzo, A., Novellino, E. & Pagano, B. Noncanonical DNA secondary structures as drug targets: The prospect of the i-motif. *ChemMedChem* **9**, 2026–2030 (2014).

7. Vorlíčková, M., Kejnovská, I., Bednářová, K., Renčiuk, D. & Kypr, J. Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. *Chirality* **24**, 691–698 (2012).

8. Pray, L. A. & Aa, A. A. Discovery of DNA Structure and Function: Watson and Crick The First Piece of the Puzzle: Miescher Discovers DNA Citation: Pray, L. (2008) Discovery of DNA structure and function: Watson and Crick. Nature Education 1(1):100. (2008).

9. Peck, L. J. & Wang, J. C. Energetics of B-to-Z transition in DNA. *Proc Natl Acad Sci U S A* **80**, 6206–6210 (1983).

10. Ho, P. S., Ellison, M. J., Quigley, G. J. & Rich, A. A computer aided thermodynamic approach for predicting the formation of Z-DNA in naturally occurring sequences. *EMBO J* **5**, 2737–2744 (1986).

11. Rich, A. & Zhang, S. Z-DNA: The long road to biological function. *Nat Rev Genet* **4**, 566–572 (2003).

12. Catasti, P. *et al.* Cystosine-rich strands of the insulin minisatellite adopt hairpins with intercalated Cytosine+ ·Cytosine pairs. *J Mol Biol* **272**, 369–382 (1997).

13. Lee, J. S., Johnson, D. A. & Morgan, A. R. Complexes formed by (pyrimidine)n . (purine)n DNAs on lowering the pH are three-stranded. *Nucleic Acids Res* **6**, 3073 (1979).

14. Panayotatos, N. & Wells, R. D. Cruciform structures in supercoiled DNA. *Nature* **289**, 466–470 (1981).

15. Wu, B. *et al.* Global structure of a DNA three-way junction by solution NMR: towards prediction of 3H fold. *Nucleic Acids Res* **32**, 3228–3239 (2004).

16. DNA: Alternative Conformations and Biology - Madame Curie Bioscience Database - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK6545/.

17. Ma, Y., Iida, K. & Nagasawa, K. Topologies of G-quadruplex: Biological functions and regulation by ligands. *Biochem Biophys Res Commun* **531**, 3–17 (2020).

18. Aboul-Ela, F., Murchie, A. I. H. & Lilley, D. M. J. NMR study of parallel-stranded tetraplex formation by the hexadeoxynucleotide d(TG4T). *Nature* **360**, 280–282 (1992).

19. Smith, F. W. & Feigon, J. Quadruplex structure of Oxytricha telomeric DNA oligonucleotides. *Nature* **356**, 164–168 (1992).

20. Greene, K. L., Wang, Y. & Live, D. Influence of the glycosidic torsion angle on 13C and 15N shifts in guanosine nucleotides: Investigations of G-tetrad models with alternating syn and anti bases. *Journal of Biomolecular NMR* **5**, 333–338 (1995).

21. Catasti, P., Chen, X., Moyzis, R. K., Bradbury, E. M. & Gupta, G. Structure–Function Correlations of the Insulin-linked Polymorphic Region. *J Mol Biol* **264**, 534–545 (1996).

22. Webba Da Silva, M. Geometric Formalism for DNA Quadruplex Folding. *Chemistry – A European Journal* **13**, 9738–9745 (2007).

23. Smargiasso, N. *et al.* G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. *J Am Chem Soc* **130**, 10208–10216 (2008).

24. Gellert, M., Lipsett, M.N. & Davies, D.R. Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences* **48**, 2013–2018 (1962).

25. Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. Inhibition of telomerase by G-quartet DMA structures. *Nature* **350**, 718–720 (1991).

26. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res* **43**, 8627–8637 (2015).

27. Ekundayo, B. & Bleichert, F. Origins of DNA replication. *PLoS Genet* **15**, (2019).

28. Dai, X. & Rothman-Denes, L. B. DNA structure and transcription. *Curr Opin Microbiol* **2**, 126–130 (1999).

29. Taanman, J. W. The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* **1410**, 103–123 (1999).

30. Todd, A. K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res* **33**, 2901–2907 (2005).

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

32. Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nature Biotechnology* **33**, 877–881 (2015).

33. Tan, Z. J. & Chen, S. J. Nucleic Acid Helix Stability: Effects of Salt Concentration, Cation Valence and Size, and Chain Length. *Biophys J* **90**, 1175 (2006).

34. Clever, G. H. & Shionoya, M. Metal–base pairing in DNA. *Coord Chem Rev* **254**, 2391–2402 (2010).

35. Müller, J. Functional metal ions in nucleic acids. *Metallomics* **2**, 318–327 (2010).

36. Waller, Z. A. E., Howell, L. A., Macdonald, C. J., O'Connell, M. A. & Searcey, M. Identification and characterisation of a G-quadruplex forming sequence in the promoter region of nuclear factor (erythroid-derived 2)-like 2 (Nrf2). *Biochem Biophys Res Commun* **447**, 128–132 (2014).

37. Howell, L. A., Waller, Z. A. E., Bowater, R., O'Connell, M. & Searcey, M. A small molecule that induces assembly of a four way DNA junction at low temperature. *Chemical Communications* **47**, 8262–8264 (2011).

38. Hardin, C. C., Watson, T., Corregan, M. & Bailey, C. Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG). *Biochemistry* **31**, 833–841 (1992).

39. Venczel, E. A. & Sen, D. Parallel and Antiparallel G-DNA Structures from a Complex Telomeric Sequence*. *Biochemistry* **32**, 6220–6228 (1993).

40. Riou J F, Gomez D, Morjani H, et al. Quadruplex ligand recognition: biological aspects[J]. Quadruplex Nucleic Acids, 154-179 (2006).

41. Bhattacharyya, D., Arachchilage, G. M. & Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front Chem* **4**, (2016).

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

43. Williamson, J. R., Raghuraman, M. K. & Cech, T. R. Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell* **59**, 871–880 (1989).

44. Lee, M. P. H., Parkinson, G. N., Hazel, P. & Neidle, S. Observation of the coexistence of sodium and calcium ions in a DNA G-quadruplex ion channel. *J Am Chem Soc* **129**, 10106–10107 (2007).

45. Miyoshi, D., Nakao, A. & Sugimoto, N. Structural transition of d(G4T4G4) from antiparallel to parallel G-quartet induced by divalent cations. *Nucleic Acids Symp Ser* **1**, 259–260 (2001).

46. Chen, F. M. Sr2+ facilitates intermolecular G-quadruplex formation of telomeric sequences. *Biochemistry* **31**, 3769–3776 (1992).

47. Gehring, K., Leroy, J.-L. & Guéron, M. A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* **363**, 561–565 (1993).

48. Marsh, R. E., Bierstedt, R. & Eichhorn, E. L. The crystal structure of cytosine-5-acetic acid. *Acta Crystallogr* **15**, 310–316 (2002).

49. Kang, C. H. *et al.* Crystal structure of intercalated four-stranded d(C3T) at 1.4 A resolution. *Proc Natl Acad Sci U S A* **91**, 11636–40 (1994).

50. Assi, H. A. *et al.* Stabilization of i-motif structures by 2'-β-fluorination of DNA. *Nucleic Acids Res* **44**, 4998–5009 (2016).

51. Berger, I., Egli, M. & Rich, A. Inter-strand C-H...O hydrogen bonds stabilizing four-stranded intercalated molecules: stereoelectronic effects of O4' in cytosine-rich DNA. *Proc Natl Acad Sci U S A* **93**, 12116–21 (1996).

52. Phan, A. T. & Leroy, J.-L. Intramolecular i-Motif Structures of Telomeric DNA. *J Biomol Struct Dyn* **17**, 245–251 (2000).

53. Brooks, T. A., Kendrick, S. & Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS Journal* **277**, 3459–3469 (2010).

54. Wright, E. P., Huppert, J. L. & Waller, Z. A. E. Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res* **45**, 2951–2959 (2017).

55. Cui, J., Waltman, P., Le, V. H. & Lewis, E. A. The effect of molecular crowding on the stability of human c-MYC promoter sequence I-motif at neutral pH. *Molecules* **18**, 12751–12767 (2013).

56. Zhou, J. *et al.* Formation of i-motif structure at neutral and slightly alkaline pH. *Mol Biosyst* **6**, 580–586 (2010).

57. Abdelhamid, M. A. *et al.* Redox-dependent control of i-Motif DNA structure using copper cations. *Nucleic Acids Res* **46**, 5886–5893 (2018).

58. Day, H. A., Huguin, C. & Waller, Z. A. E. Silver cations fold i-motif at neutral pH. *Chemical Communications* **49**, 7696 (2013).

59. 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* **52**, 2863–2874 (2009).

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

61. Mergny, J. L., Lacroix, L., Hélène, C., Han, X. & Leroy, J. L. Intramolecular Folding of Pyrimidine Oligodeoxynucleotides into an i-DNA Motif. *J Am Chem Soc* **117**, 8887–8898 (1995).

62. Brazier, J. A., Shah, A. & Brown, G. D. I-Motif formation in gene promoters: Unusually stable formation in sequences complementary to known G-quadruplexes. *Chemical Communications* **48**, 10739–10741 (2012).

63. King, J. J. *et al.* DNA G-Quadruplex and i-Motif Structure Formation Is Interdependent in Human Cells. *J Am Chem Soc* **142**, 20600–20604 (2020).

64. Mendoza, O., Elezgaray, J. & Mergny, J. L. Kinetics of quadruplex to duplex conversion. *Biochimie* **118**, 225–233 (2015).

65. Risitano, A. & Fox, K. R. Stability of intramolecular DNA quadruplexes: Comparison with DNA duplexes. *Biochemistry* **42**, 6507–6513 (2003).

66. Chalikian, T. V., Liu, L. & Macgregor, R. B. Duplex-tetraplex equilibria in guanine- and cytosine-rich DNA. *Biophys Chem* **267**, (2020).

67. Garavís, M., González, C. & Villasante, A. On the Origin of the Eukaryotic Chromosome: The Role of Noncanonical DNA Structures in Telomere Evolution. *Genome Biol Evol* **5**, 1142–1150 (2013).

68. Školáková, P. *et al.* Composite 5-methylations of cytosines modulate i-motif stability in a sequence-specific manner: Implications for DNA nanotechnology and epigenetic regulation of plant telomeric DNA. *Biochim Biophys Acta Gen Subj* **1864**, (2020).

69. Kendrick, S. *et al.* The Dynamic Character of the *BCL2* Promoter i-Motif Provides a Mechanism for Modulation of Gene Expression by Compounds That Bind Selectively to the Alternative DNA Hairpin Structure. *J Am Chem Soc* **136**, 4161–4171 (2014).

70. Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* **35**, 406–413 (2007).

71. Day, H. A., Pavlou, P. & Waller, Z. A. E. I-Motif DNA: Structure, stability and targeting with ligands. *Bioorg Med Chem* **22**, 4407–4418 (2014).

72. Zikich, D., Liu, K., Sagiv, L., Porath, D. & Kotlyar, A. I-Motif Nanospheres: Unusual Self-Assembly of Long Cytosine Strands. *Small* **7**, 1029–1034 (2011).

73. Guo, K., Gokhale, V., Hurley, L. H. & Sun, D. Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene. *Nucleic Acids Res* **36**, 4598–4608 (2008).

74. Dai, J., Hatzakis, E., Hurley, L. H. & Yang, D. I-motif structures formed in the human c-MYC promoter are highly dynamic-insights into sequence redundancy and I-motif stability. *PLoS One* **5**, 1–8 (2010).

75. Xu, Y. & Sugiyama, H. Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res* **34**, 949–954 (2006).

76. Kaiser, C. E. *et al.* Insight into the Complexity of the i-Motif and G-Quadruplex DNA Structures Formed in the KRAS Promoter and

Subsequent Drug-Induced Gene Repression. *J Am Chem Soc* **139**, 8522 (2017).

77. Dhakal, S., Lafontaine, J. L., Yu, Z., Koirala, D. & Mao, H. Intramolecular folding in human ILPR fragment with three C-rich repeats. *PLoS One* **7**, (2012).

78. Leroy, J. L., Guéron, M., Mergny, J. L. & Hélène, C. Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif. *Nucleic Acids Res* **22**, 1600–6 (1994).

79. Phan, A. T. & Mergny, J. Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res* **30**, 4618–4625 (2002).

80. Wellinger, R. J. In the end, what's the problem? *Mol Cell* **53**, 855–856 (2014).

81. Watson, J. D. Origin of concatemeric T7 DNA. *Nat New Biol* **239**, 197–201 (1972).

82. Stroik, S. & Hendrickson, E. A. Telomere replication-When the going gets tough. *DNA Repair (Amst)* **94**, (2020).

83. Phan, A. T. & Mergny, J.-L. Human telomeric DNA: G-quadruplex, i-motif and Watson-Phan, A. T., & Mergny, J.-L. (2002). Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res* **30**, 4618–25 (2002).

84. Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011–5 (1994).

85. Chen, Y. *et al.* Insights into the biomedical effects of carboxylated single-wall carbon nanotubes on telomerase and telomeres. *Nat Commun* **3**, 1074 (2012).

86. Simonsson, T., Pribylova, M. & Vorlickova, M. A Nuclease Hypersensitive Element in the Human c-myc Promoter Adopts Several Distinct i-Tetraplex Structures. *Biochem Biophys Res Commun* **278**, 158–166 (2000).

87. S J Berberich, E H Postel. PuF / NM23-H2 / NDPK-B transactivates a human c-myc promoter-CAT gene via a functional nuclease hypersensitive element . *PubMed Commons*. 1–2 (2014).

88. González, V. & Hurley, L. H. The c- *MYC* NHE III $_1$: Function and Regulation. *Annu Rev Pharmacol Toxicol* **50**, 111–129 (2010).

89. Brooks, T. A., Kendrick, S. & Hurley, L. Oncogene Promoters. *FEBS Jopurnal* **277**, 3459–3469 (2010).

90. 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* **52**, 2863–2874 (2009).

91. Simonsson, T., Pribylova, M. & Vorlickova, M. A Nuclease Hypersensitive Element in the Human c-myc Promoter Adopts Several Distinct i-Tetraplex Structures. *Biochem Biophys Res Commun* **278**, 158–166 (2000).

92. Shu, B. *et al.* Syntheses and evaluation of new acridone derivatives for selective binding of oncogene c- *myc* promoter i-motifs in gene transcriptional regulation. *Chemical Communications* **54**, 2036–2039 (2018).

93. Kuang, G. *et al.* Syntheses and Evaluation of New Bisacridine Derivatives for Dual Binding of G-Quadruplex and i-Motif in Regulating Oncogene c-myc Expression. *J Med Chem* **63**, 9136–9153 (2020).

94. Safran, M. & Kaelin, W. G. HIF hydroxylation and the mammalian oxygen-sensing pathway. *J Clin Invest* **111**, 779–783 (2003).

95. H Zhong, A M De Marzo, E Laughner, M Lim, D A Hilton, D Zagzag, P Buechler, W B Isaacs, G L Semenza, J W Simons. Overexpression of hypoxia-inducible factor 1alpha in common human cancers and their metastases. *Oncogene* **10**, 2343-2347 (1995).

96. Semenza, G. L. *et al.* Structural and functional analysis of hypoxia-inducible factor 1. *Kidney Int* **51**, 553–555 (1997).

97. Abdelhamid, M. A. S., Gates, A. J. & Waller, Z. A. E. Destabilization of i-Motif DNA at Neutral pH by G-Quadruplex Ligands. *Biochemistry* **58**, 245–249 (2019).

98. Fedoroff, O. Y., Rangan, A., Chemeris, V. V. & Hurley, L. H. Cationic porphyrins promote the formation of i-motif DNA and bind peripherally by a nonintercalative mechanism. *Biochemistry* **39**, 15083–15090 (2000).

99. Nagasawa, K. *et al.* Design and Synthesis of Telomestatin Derivatives and Their Inhibitory Activity of Telomerase. *Heterocycles* **69**, 505 (2006).

100. Sedghi Masoud, S. & Nagasawa, K. i-Motif-Binding Ligands and Their Effects on the Structure and Biological Functions of i-Motif. *Chem Pharm Bull (Tokyo)* **66**, 1091–1103 (2018).

101. Rosu, F., Gabelica, V., Shin-Ya, K. & de Pauw, E. Telomestatin-induced stabilization of the human telomeric DNA quadruplex monitored by electrospray mass spectrometry. *Chemical Communications* 2702–2703 (2003).

102. Kim, M. Y., Vankayalapati, H., Shin-Ya, K., Wierzba, K. & Hurley, L. H. Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular G-quadruplex. *J Am Chem Soc* **124**, 2098–2099 (2002).

103. Shin-ya, K. *et al.* Telomestatin, a novel telomerase inhibitor from Streptomyces anulatus [17]. *J Am Chem Soc* **123**, 1262–1263 (2001).

104. Nagasawa, K., Majima, S., Tera, M., Iida, K. & Shin-ya, K. Design and Synthesis of Telomestatin Derivatives Containing Methyl Oxazole and Their G-Quadruplex Stabilizing Activities. *Heterocycles* **82**, 1345 (2010).

105. Sedghi Masoud, S. *et al.* Analysis of Interactions between Telomeric i-Motif DNA and a Cyclic Tetraoxazole Compound. *ChemBioChem* **19**, 2268–2272 (2018).

106. Alberti P, Ren J, Teulade-Fichou MP, Guittat L, Riou JF, Chaires J, Hélène C, Vigneron JP, Lehn JM, Mergny JL. Interaction of an acridine dimer with DNA quadruplex structures. *J Biomol Struct Dyn* **19**, 505-513 (2001).

107. Albert Day, H. Investigating the Effect of Small Molecule Ligands and Cations on i-Motif DNA. PhD thesis (2015).

108. Li, X., Peng, Y. & Qu, X. Carbon nanotubes selective destabilization of duplex and triplex DNA and inducing B-A transition in solution. *Nucleic Acids Res* **34**, 3670–3676 (2006).

109. Zhao, C., Ren, J. & Qu, X. Single-walled carbon nanotubes binding to human telomeric i-motif DNA under molecular-crowding conditions: More water molecules released. *Chemistry - A European Journal* **14**, 5435–5439 (2008).

110. Li, X., Peng, Y., Ren, J. & Qu, X. Carboxyl-modified single-walled carbon nanotubes selectively induce human telomeric i-motif formation. *Proceedings of the National Academy of Sciences* **103**, 19658–19663 (2006).

111. Ren, J. *et al.* Insights into the biomedical effects of carboxylated single-wall carbon nanotubes on telomerase and telomeres. *Nat Commun* **3**, (2012).

112. Xue, L., Ranjan, N. & Arya, D. P. Synthesis and spectroscopic studies of the aminoglycoside (Neomycin)-perylene conjugate binding to human telomeric DNA. *Biochemistry* **50**, 2838–2849 (2011).

113. Wright, E. P. *et al.* Mitoxantrone and Analogues Bind and Stabilize i-Motif Forming DNA Sequences. *Sci Rep* **6**, 39456 (2016).

114. Pagano, A. *et al.* Common G-quadruplex binding agents found to interact with i-motif-forming DNA: Unexpected multi-target-directed compounds. *Front Chem* **6**, 281 (2018).

115. Sheng, Q. *et al.* Identification of new DNA i-motif binding ligands through a fluorescent intercalator displacement assay. *Org Biomol Chem* **15**, 5669–5673 (2017).

116. Kimira, M., Arai, Y., Shimoi, K. & Watanabe, S. Japanese intake of flavonoids and isoflavonoids from foods. *J Epidemiol* **8**, 168–175 (1998).

117. Arai, Y. *et al.* Dietary Intakes of Flavonols, Flavones and Isoflavones by Japanese Women and the Inverse Correlation between Quercetin Intake and Plasma LDL Cholesterol Concentration. *J Nutr* **130**, 2243–2250 (2000).

118. Prasath, G. S. & Subramanian, S. P. Modulatory effects of fisetin, a bioflavonoid, on hyperglycemia by attenuating the key enzymes of carbohydrate metabolism in hepatic and renal tissues in streptozotocin-induced diabetic rats. *Eur J Pharmacol* **668**, 492–496 (2011).

119. Syed, D. N., Adhami, V. M., Khan, N., Khan, M. I. & Mukhtar, H. Exploring the molecular targets of dietary flavonoid fisetin in cancer. *Seminars in Cancer Biology* vol. **40_41** 130–140 (2016).

120. Fazel Nabavi, S. *et al.* Neuroprotective Effects of Fisetin in Alzheimer's and Parkinson's Diseases: From Chemistry to Medicine. *Curr Top Med Chem* **16**, 1910–1915 (2016).

121. Bhattacharjee, S., Chakraborty, S., Sengupta, P. K. & Bhowmik, S. Exploring the Interactions of the Dietary Plant Flavonoids Fisetin and Naringenin with G-Quadruplex and Duplex DNA, Showing Contrasting Binding Behavior: Spectroscopic and Molecular Modeling Approaches. *Journal of Physical Chemistry B* **120**, 8942–8952 (2016).

122. Takahashi, S., Bhattacharjee, S., Ghosh, S., Sugimoto, N. & Bhowmik, S. Preferential targeting cancer-related i-motif DNAs by the plant flavonol fisetin for theranostics applications. *Sci Rep* **10**, 1–13 (2020).

123. Stephen Neidle. DNA and RNA Quadruplex Structures: Therapeutic Applications of Quadruplex Nucleic Acids. *Academic Press,* Pages 21-42 (2012).

124. Wang, L., Wu, Y., Chen, T. & Wei, C. The interactions of phenanthroline compounds with DNAs: Preferential binding to telomeric quadruplex over duplex. *Int J Biol Macromol* **52**, 1–8 (2013).

125. Gao, N., Wang, Y. & Wei, C. Interactions of phenanthroline compounds with i-motif DNA. *Chemical Research in Chinese Universities 2014 30:3* **30**, 495–499 (2014).

126. Brown, R. V. *et al.* The Consequences of Overlapping G-Quadruplexes and i-Motifs in the Platelet-Derived Growth Factor Receptor ß Core Promoter Nuclease Hypersensitive Element Can Explain the Unexpected Effects of Mutations and Provide Opportunities for Selective Targeting of Both Structures by Small Molecules To Downregulate Gene Expression. *J Am Chem Soc* **139**, 7456–7475 (2017).

127. Debnath, M. *et al.* Preferential targeting of i-motifs and G-quadruplexes by small molecules. *Chem. Sci.* **8**, 7448–7456 (2017).

128. Gunaratnam, M. *et al.* G-quadruplex compounds and cis-platin act synergistically to inhibit cancer cell growth in vitro and in vivo. *Biochem Pharmacol* **78**, 115–122 (2009).

129. Pizarro, A. M. & Sadler, P. J. Unusual DNA binding modes for metal anticancer complexes. *Biochimie* **91**, 1198–1211 (2009).

130. Bertrand, B. *et al.* Cytotoxicity of Pyrazine-Based Cyclometalated (C^Npz^C)Au(III) Carbene Complexes: Impact of the Nature of the Ancillary Ligand on the Biological Properties. *Inorg Chem* **56**, 5728–5740 (2017).

131. Ma, D. L. *et al.* Crystal violet as a fluorescent switch-on probe for i-motif: Label-free DNA-based logic gate. *Analyst* **136**, 2692–2696 (2011).

132. Erkkila, K. E., Odom, D. T. & Barton, J. K. Recognition and reaction of metallointercalators with DNA. *Chem Rev* **99**, 2777–2795 (1999).

133. Yu, H. J., Yu, L., Hao, Z. F. & Zhao, Y. Interactions of ruthenium complexes containing indoloquinoline moiety with human telomeric G-quadruplex DNA. *Spectrochim Acta A Mol Biomol Spectrosc* **124**, 187–193 (2014).

134. Kostova, I. Ruthenium complexes as anticancer agents. *Curr Med Chem* **13**, 1085–1107 (2006).

135. Shi, S. *et al.* Molecular “light switch” for G-quadruplexes and i-motif of human telomeric DNA: [Ru(phen)2(dppz)]2+. *Dalton Transactions* **39**, 2490–2493 (2010).

136. Ji, L. *et al.* Interaction of [Ru(bpy)2(dppz)]2+ with human telomeric DNA: Preferential binding to G-quadruplexes over i-motif. *Biochimie* **92**, 370–377 (2010).

137. Sheng, Q. *et al.* Identification of new DNA i-motif binding ligands through a fluorescent intercalator displacement assay. *Org Biomol Chem* **15**, 5669–5673 (2017).

138. Abrahamsson, M. *et al.* A 3.0 μs room temperature excited state lifetime of a bistridentate RuII-polypyridine complex for rod-like molecular arrays. *J Am Chem Soc* **128**, 12616–12617 (2006).

139. Abrahamsson, M., Becker, H. C. & Hammarström, L. Microsecond 3MLCT excited state lifetimes in bis-tridentate Ru(II)-complexes: significant reductions of non-radiative rate constants. *Dalton Transactions* **46**, 13314–13321 (2017).

140. Schlotthauer, T. *et al.* Aryl-Decorated RuII Polypyridyl-type Photosensitizer Approaching NIR Emission with Microsecond Excited State Lifetimes. *Inorg Chem* **55**, 5405–5416 (2016).

141. Jäger, M., Kumar, R. J., Görls, H., Bergquist, J. & Johansson, O. Facile synthesis of bistridentate Ru II complexes based on 2,6-Di(quinolin-8-yl)pyridyl Ligands: Sensitizers with microsecond 3MLCT excited state lifetimes. *Inorg Chem* **48**, 3228–3238 (2009).

142. Spence, P., Fielden, J. & Waller, Z. A. E. Beyond Solvent Exclusion: I-Motif Detecting Capability and an Alternative DNA Light-Switching Mechanism in a Ruthenium(II) Polypyridyl Complex. *J Am Chem Soc* **142**, 13856–13866 (2020).

143. Dash, J., Waller, Z. A. E., Pantoş, G. D. & Balasubramanian, S. Synthesis and binding studies of novel diethynyl-pyridine amides with genomic promoter DNA G-quadruplexes. *Chemistry* **17**, 4571–4581 (2011).

144. Garbett, N. C., Ragazzon, P. A. & Chaires, J. O. B. Circular dichroism to determine binding mode and affinity of ligand–DNA interactions. *Nature Protocols* **2**, 3166–3172 (2007).

145. Haldar, S. *et al.* Mechanistic Insights into the Ligand-Induced Unfolding of an RNA G-Quadruplex. *J Am Chem Soc* **144**, 935–950 (2022).

146. Chirio-Lebrun, M.-C. & Prats, M. *Biochemical Fluorescence resonance energy transfer (FRET): theory and experiments*. *Biochcfnical Education* vol. 26 (1998).

147. Calculate Resonance Energy Transfer (FRET) Efficiencies - The fluorescence laboratory. http://www.fluortools.com/software/ae/documentation/tools/FRET.

148. Kendrick, S. *et al.* The dynamic character of the BCL2 promoter i-motif provides a mechanism for modulation of gene expression by compounds that bind selectively to the alternative DNA hairpin structure. *J Am Chem Soc* **136**, 4161–4171 (2014).

149. Renčiuk, D. *et al.* A FRET-based screening assay for nucleic acid ligands. *Methods* **57**, 122–128 (2012).

150. Mergny, J. L. Fluorescence energy transfer as a probe for tetraplex formation: The i- motif. *Biochemistry* **38**, 1573–1581 (1999).

151. Mergny, J. L. & Lacroix, L. Analysis of Thermal Melting Curves. *Oligonucleotides* **vol. 13** 515–537 (2003).

152. Merkina, E. E. & Fox, K. R. Kinetic stability of intermolecular DNA quadruplexes. *Biophys J* **89**, 365–373 (2005).

153. Brown, N. M., Rachwal, P. A., Brown, T. & Fox, K. R. Exceptionally slow kinetics of the intramolecular quadruplex formed by the Oxytricha telomeric repeat. *Org Biomol Chem* **3**, 4153–4157 (2005).

154. Mergny, J. L., de Cian, A., Ghelab, A., Saccà, B. & Lacroix, L. Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res* **33**, 81–94 (2005).

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

156. Lewis, M. A. & Long, E. C. Fluorescent intercalator displacement analyses of DNA binding by the peptide-derived natural products netropsin, actinomycin, and bleomycin. *Bioorg Med Chem* **14**, 3481–3490 (2006).

157. Ham, Y. W., Tse, W. C. & Boger, D. L. High-Resolution assessment of protein DNA binding affinity and selectivity utilizing a fluorescent intercalator displacement (FID) assay. *Bioorg Med Chem Lett* **13**, 3805–3807 (2003).

158. Vorlíčková, M. *et al.* Circular dichroism and guanine quadruplexes. *Methods* **57**, 64–75 (2012).

159. del Villar-Guerra, R., Trent, J. O. & Chaires, J. B. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angewandte Chemie - International Edition* **57**, 7171–7175 (2018).

160. Bishop, G. R. & Chaires, J. B. Characterization of DNA Structures by Circular Dichroism. in *Current Protocols in Nucleic Acid Chemistry* vol. Chapter 7 7.11.1-7.11.8 (John Wiley & Sons, Inc., 2002).

161. Kumar, N., Nielsen, J. T., Maiti, S. & Petersen, M. i-Motif formation with locked nucleic acid (LNA). *Angewandte Chemie - International Edition* **46**, 9220–9222 (2007).

162. Kypr, J., Kejnovská, I., Renciuk, D. & Vorlícková, M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res* **37**, 1713–25 (2009).

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

164. Gray, D. M. *et al.* Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality* **20**, 431–440 (2008).

165. Ultraviolet Radiation. https://ehs.lbl.gov/resource/documents/radiation-protection/non-ionizing-radiation/ultraviolet-radiation/.

166. Mergny, J.-L. & Lacroix, L. Analysis of Thermal Melting Curves. *Oligonucleotides* **13**, 515–537 (2003).

167. Mergny, J. L., Li, J., Lacroix, L., Amrane, S. & Chaires, J. B. Thermal difference spectra: A specific signature for nucleic acid structures. *Nucleic Acids Res* **33**, 1–6 (2005).

168. Gralla, J. & Crothers, D. M. Free energy of imperfect nucleic acid helices: III. Small internal loops resulting from mismatches. *J Mol Biol* **78**, 301–319 (1973).

169. Assi, H. A., Garavís, M., González, C. & Damha, M. J. i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res* **46**, 8038–8056 (2018).

170. Cai, L. *et al.* Intercalated cytosine motif and novel adenine clusters in the crystal structure of the Tetrahymena telomere. *Nucleic Acids Res* **26**, 4696–4705 (1998).

171. Qin, T., Liu, K., Song, D., Yang, C. & Su, H. Porphyrin Bound to i-Motifs: Intercalation versus External Groove Binding. *Chem Asian J* **12**, 1578–1586 (2017).

172. Chen, S. Bin *et al.* Visualization of NRAS RNA G-Quadruplex Structures in Cells with an Engineered Fluorogenic Hybridization Probe. *J Am Chem Soc* **138**, 10382–10385 (2016).

173. Angelbello, A. J. *et al.* Using Genome Sequence to Enable the Design of Medicines and Chemical Probes. *Chem Rev* **118**, 1599–1663 (2018).

174. Blackledge, M. S. & Melander, C. Programmable DNA-binding Small Molecules. *Bioorg Med Chem* **21**, 6101 (2013).

175. Lauria A, Montalbano A, Barraja P, Dattolo G, Almerico AM. DNA minor groove binders: an overview on molecular modeling and QSAR approaches. *Curr Med Chem* **14**, 2136–2160 (2007).

176. Wemmer, D. E. & Dervan, P. B. Targeting the minor groove of DNA. *Curr Opin Struct Biol* **7**, 355–361 (1997).

177. Dervan, P. B. Molecular recognition of DNA by small molecules. *Bioorg Med Chem* **9**, 2215–2235 (2001).

178. Abu-daya, A., Brown, P. M. & Fox, K. R. DNA sequence preferences of several AT-selective minor groove binding ligands. *Nucleic Acids Res* **23**, 3385–3392 (1995).

179. Zhang, X. *et al.* New Insight into the Molecular Mechanisms of the Biological Effects of DNA Minor Groove Binders. *PLoS One* **6**, e25822 (2011).

180. Cai, X., Gray, P. J. & Von Hoff, D. D. DNA minor groove binders: Back in the groove. *Cancer Treat Rev* **35**, 437–450 (2009).

181. Arcamone,F., Penco,S., Orezzi,P., Nicolella,V &Pirelli,A. Structure and Synthesis of Distamycin A. *Nature* **203**, 1064–1065 (1964).

182. Pelton, J. G. & Wemmer, D. E. Structural characterization of a 2:1 distamycin A d(CGCAAATI'GGC) complex by two-dimensional NMR (drug-DNA interactions/drug binding/DNA minor groove/intermolecular contacts). *Library Services* **86**, 5723–5727 (1989).

183. Patel, D. J. Antibiotic-DNA interactions: Intermolecular nuclear Overhauser effects in the netropsin-d(C-G-C-G-A-A-T-T-C-G-C-G) complex in solution. *Proc Natl Acad Sci U S A* **79**, 6424–6428 (1982).

184. Freyer, M. W., Buscaglia, R., Nguyen, B., David Wilson, W. & Lewis, E. A. Binding of netropsin and 4,6-diamidino-2-phenylindole to an A2T2 DNA hairpin: a comparison of biophysical techniques. *Anal Biochem* **355**, 259–266 (2006).

185. Coll, M., Frederick, C. A., Wang, A. H. & Rich, A. A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc Natl Acad Sci U S A* **84**, 8385–8389 (1987).

186. Huijuan, Z. *et al.* Fluorescence dynamics of interactions between polyamide PyPyPyβDp and DNA. *Sci China B Chem* **49**, 75 (2006).

187. Lewis, E. A. *et al.* Complexity in the binding of minor groove agents: netropsin has two thermodynamically different DNA binding modes at a single site. *Nucleic Acids Res* **39**, 9649 (2011).

188. Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E. & Dervan, P. B. Regulation of gene expression by small molecules. *Nature 1997 387:6629* **387**, 202–205 (1997).

189. Turner, P. R. & Denny, W. A. The mutagenic properties of DNA minor-groove binding ligands. *Mutat Res* **355**, 141–169 (1996).

190. Loontiens, F. G., Dumortier, L., Regenfuss, P., Zechel, A. & Clegg, R. M. Binding characteristics of Hoechst 33258 with calf thymus DNA, poly[d(A-T)], and d(CCGGAATTCCGG): multiple stoichiometries and determination of tight binding with a wide spectrum of site affinities. *Biochemistry* **29**, 9029–9039 (1990).

191. Drobyshev, A. L., Zasedatelev, A. S., Yershov, G. M. & Mirzabekov, A. D. Massive parallel analysis of DNA-Hoechst 33258 binding specificity with a generic oligodeoxyribonucleotide microchip. *Nucleic Acids Res* **27**, 4100–4105 (1999).

192. Ramakers, L. A. I. *et al.* 2D-IR Spectroscopy Shows that Optimized DNA Minor Groove Binding of Hoechst33258 Follows an Induced Fit Model. *J Phys Chem B* **121**, 1295–1303 (2017).

193. Embrey, K. J., Searle, M. S. & Craik, D. J. Interaction of Hoechst 33258 with the minor groove of the A + T-rich DNA duplex d(GGTAATTACC)2 studied in solution by NMR spectroscopy. *Eur J Biochem* **211**, 437–447 (1993).

194. Carter, M. L. J., Rusling, D. A., Gurr, S., Brown, T. & Fox, K. R. Stability of the different arms of a DNA tetrahedron and its interaction with a minor groove ligand. *Biophys Chem* **256**, (2020).

195. Spink, N., Brown, D. G., Skelly, J. V. & Neidle, S. Sequence-dependent effects in drug-DNA interaction: the crystal structure of Hoechst 33258 bound to the d(CGCAAATTTGCG)2 duplex. *Nucleic Acids Res* **22**, 1607–1612 (1994).

196. Chandra, P., Mildner, B., Dann, O. & Metz, A. Influence of 4'-6'-diamidino-2-phenylindole on the secondary structure and template activities of DNA and polydeoxynucleotides. *Mol Cell Biochem* **18**, 81–86 (1977).

197. Mildner, B., Metz, A. & Chandra, P. Interaction of 4'-6-diamidino-2-phenylindole to nucleic acids, and its implication to their template activity in RNA-polymerase reaction of E. coli bacteria and of Friend-virus infected mouse spleen. *Cancer Lett* **4**, 89–98 (1978).

198. Fox, K. R., Sansom, C. E. & Stevens, M. F. G. Footprinting studies on the sequence-selective binding of pentamidine to DNA. *FEBS Lett* **266**, 150–154 (1990).

199. White, S., Baird, E. E. & Dervan, P. B. On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem Biol* **4**, 569–578 (1997).

200. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E. & Dervan, P. B. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **391**, 468–471 (1998).

201. Wang, C. C. C., Ellervik, U. & Dervan, P. B. Expanding the recognition of the minor groove of DNA by incorporation of beta-alanine in hairpin polyamides. *Bioorg Med Chem* **9**, 653–657 (2001).

202. Turner, J. M., Swalley, S. E., Baird, E. E. & Dervan, P. B. Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *J Am Chem Soc* **120**, 6219–6226 (1998).

203. Tang, F., Wang, J., Liu, D., Li, J. & Yuan, G. Systematical synthesis of distamycin analogues and their interaction with Herring Sperm DNA. *Journal of the Chinese Chemical Society* **49**, 1073–1077 (2002).

204. Chan, W. C. & White, P. D. Fmoc solid phase peptide synthesis : a practical approach. (*Oxford University Press*, 2000).

205. Melnikova AF, Zasedatelev AS, Kolchinsky AM, Gursky GV, Zhuze AL, Grochovsky SL, Mirzabekov AD. Accessibility of the minor groove of DNA in chromatin to the binding of antibiotics netropsin and distamycin A. *Mol Biol Rep* **2** 135-142 (1975).

206. Mahmoud AS Abdelhamid. Synthesis of novel biological Nanostructures. PhD thesis, (2019).

207. Waller, Z. A. E. Targeting G-Quadruplex DNA with Triarylpyridines and Nile Red Derivatives. PhD Thesis. (2009).

208. Gray, D. M., Ratliff, R. L. & Vaughan, M. R. Circular dichroism spectroscopy of DNA. *Methods Enzymol* **211**, 389–406 (1992).

209. Viglasky, V., Bauer, L., Tluckova, K. & Javorsky, P. Evaluation of Human Telomeric G-Quadruplexes: The Influence of Overhanging Sequences on Quadruplex Stability and Folding. *J Nucleic Acids* **2010**, (2010).

210. Demkovičová, E. *et al.* Telomeric G-Quadruplexes: From Human to Tetrahymena Repeats. *J Nucleic Acids* **2017**, (2017).

211. Abdelhamid, M. A. S. & Waller, Z. A. E. Tricky Topology: Persistence of Folded Human Telomeric i-Motif DNA at Ambient Temperature and Neutral pH. *Front Chem* **8**, 40 (2020).

212. Dervan, P. B. & Edelson, B. S. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr Opin Struct Biol* **13**, 284–299 (2003).

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

214. 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* **133**, 17673–17680 (2011).

215. Collie, G. W. *et al.* Structural basis for telomeric G-quadruplex targeting by naphthalene diimide ligands. *J Am Chem Soc* **134**, 2723–2731 (2012).

216. Nicoludis, J. M. *et al.* Optimized end-stacking provides specificity of N-methyl mesoporphyrin IX for human telomeric G-quadruplex DNA. *J Am Chem Soc* **134**, 20446–20456 (2012).

217. Nguyen, T. Q. N., Lim, K. W. & Phan, A. T. A Dual-Specific Targeting Approach Based on the Simultaneous Recognition of Duplex and Quadruplex Motifs. *Scientific Reports* **7**, 1–7 (2017).

218. Randazzo A, Galeone A, Esposito V, Varra M, Mayol L. Interaction of distamycin A and netropsin with quadruplex and duplex structures: a comparative 1H-NMR study. *Nucleosides Nucleotides Nucleic Acids* **21** 535-545 (2002).

219. Maiti, S., Chaudhury, N. K. & Chowdhury, S. Hoechst 33258 binds to G-quadruplex in the promoter region of human c-myc. *Biochem Biophys Res Commun* **310**, 505–512 (2003).

220. Xu, L. *et al.* Thiazole Orange as a Fluorescent Light-Up Probe for the i-Motif and its Application to the Development of a Molecular Logic System. *Asian J Org Chem* **4**, 1375–1378 (2015).

221. Lin, S. *et al.* A G-quadruplex-selective luminescent probe with an anchor tail for the switch-on detection of thymine DNA glycosylase activity. *Biosens Bioelectron* **86**, 849–857 (2016).

222. Spence, P. Revealing DNA's secret structures : development of an i-Motif specific ruthenium-based probe. PhD Thesis, (2020).

223. Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform* **4**, (2012).

224. Available Plates | Drug Synthesis and Chemistry Branch (DSCB) | Developmental Therapeutics Program (DTP). https://dtp.cancer.gov/organization/dscb/obtaining/available_plates.htm.

225. King, J. J. *et al.* DNA G-Quadruplex and i-Motif Structure Formation Is Interdependent in Human Cells. *J Am Chem Soc* **142**, 20600–20604 (2020).

226. Hoffmann-Röder, A. & Krause, N. Synthesis and Properties of Allenic Natural Products and Pharmaceuticals. *Angewandte Chemie International Edition* **43**, 1196–1216 (2004).

227. Sierra, M. A., Casarrubios, L. & de la Torre, M. C. Bio-Organometallic Derivatives of Antibacterial Drugs. *Chemistry – A European Journal* **25**, 7232–7242 (2019).

228. Frei, A. *et al.* Metal complexes as a promising source for new antibiotics. *Chem Sci* **11**, 2627–2639 (2020).

229. Maliszewska, H. K. Study of allene-based ligands for transition metal complexes – synthesis and applications in catalysis and as metallodrugs. PhD thesis, (2021).

230. Novakova O, Kasparkova J, Malina J, Natile G, Brabec V. DNA-protein cross-linking by trans-[PtCl(2)(E-iminoether)(2)]. A concept for activation of the trans geometry in platinum antitumor complexes. *Nucleic Acids Res* **31**, 6450-6460 (2003).

231. Byabartta, P. Gold-dppm-Arylazoimidazole Complexes: Synthesis, Spectra, and Redox Study. *SRX Chemistry* **2010**, 1–6 (2010).

232. Saris, C. P., Van de Vaart, P. J. M., Rietbroek, R. C. & Blommaert, F. A. In vitro formation of DNA adducts by cisplatin, lobaplatin and oxaliplatin in calf thymus DNA in solution and in cultured human cells. *Carcinogenesis* **17**, 2763–2769 (1996).

233. Messori, L., Orioli, P., Tempi, C. & Marcon, G. Interactions of selected gold(III) complexes with calf thymus DNA. *Biochem Biophys Res Commun* **281**, 352–360 (2001).

234. Chaney, S. G. *et al.* Protein interactions with platinum-DNA adducts: from structure to function. *J Inorg Biochem* **98**, 1551–1559 (2004).

235. Scoditti, S., Dabbish, E., Russo, N., Mazzone, G. & Sicilia, E. Anticancer Activity, DNA Binding, and Photodynamic Properties of a N∧C∧N-Coordinated Pt(II) Complex. *Inorg Chem* **60**, 10350–10360 (2021).

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

237. Maliszewska, H. K. *et al.* Precious metal complexes of bis(pyridyl)allenes: synthesis and catalytic and medicinal applications. *Dalton Transactions* **50**, 16739–16750 (2021).

238. Bindoli, A. *et al.* Thioredoxin reductase: A target for gold compounds acting as potential anticancer drugs. *Coord Chem Rev* **253**, 1692–1707 (2009).

239. Milacic, V. & Dou, Q. P. The tumor proteasome as a novel target for gold(III) complexes: implications for breast cancer therapy. *Coord Chem Rev* **253**, 1649–1660 (2009).

240. Meggers, E. Targeting proteins with metal complexes. *Chemical Communications* **9** 1001–1010 (2009).

241. Casini, A. & Messori, L. Molecular Mechanisms and Proposed Targets for Selected Anticancer Gold Compounds. *Curr Top Med Chem* **11**, 2647–2660 (2011).

242. Dörr, M. & Meggers, E. Metal complexes as structural templates for targeting proteins. *Curr Opin Chem Biol* **19**, 76–81 (2014).

243. Decian, A. *et al.* Fluorescence-based melting assays for studying quadruplex ligands. *Methods* **42**, 183–195 (2007).

244. Gao, E. *et al.* Impact of the Carbon Chain Length of Novel Palladium(II) Complexes on Interaction with DNA and Cytotoxic Activity. *Inorg Chem* **49**, 3261–3270 (2010).

245. Bork, M. A. *et al.* Accessibility and external versus intercalative binding to DNA as assessed by oxygen-induced quenching of the palladium(II)-containing cationic porphyrins Pd(T4) and Pd(tD4). *Biochemistry* **53**, 714–724 (2014).

246. Onar, G. *et al.* Palladium(II) and ruthenium(II) complexes of benzotriazole functionalized N-heterocyclic carbenes: Cytotoxicity, antimicrobial, and DNA interaction studies. *J Organomet Chem* **886**, 48–56 (2019).

247. Chow, C. S., Whitehead, J. P. & Lippard, S. J. HMG Domain Proteins Induce Sharp Bends in Cisplatin-Modified DNA. *Biochemistry* **33**, 15124–15130 (1994).

248. Kang, H. J., Kendrick, S., Hecht, S. M. & Hurley, L. H. The transcriptional complex between the BCL2 i-motif and hnRNP LL is a molecular switch for control of gene expression that can be modulated by small molecules. *J Am Chem Soc* **136**, 4172–4185 (2014).

249. Hannon, M. J. *et al.* Intramolecular DNA Coiling Mediated by a Metallo-Supramolecular Cylinder**. *Angew. Chem. Int. Ed* **40**, (2001).

250. Meistermann, I. *et al.* Intramolecular DNA coiling mediated by metallo-supramolecular cylinders: Differential binding of P and M helical enantiomers. *Proc Natl Acad Sci U S A* **99**, 5069–5074 (2002).

251. Moldrheim, E., Hannon, M. J., Meistermann, I., Rodger, A. & Sletten, E. Interaction between a DNA oligonucleotide and a dinuclear iron(II) supramolecular cylinder; an NMR and molecular dynamics study. *J Biol Inorg Chem* **7**, 770–780 (2002).

252. Khalid, S., Hannon, M. J., Rodger, A. & Mark Rodger, P. Simulations of DNA coiling around a synthetic supramolecular cylinder that binds in the DNA major groove. *Chemistry* **12**, 3493–3506 (2006).

253. Hotze, A. C. G. *et al.* Supramolecular Iron Cylinder with Unprecedented DNA Binding Is a Potent Cytostatic and Apoptotic Agent without Exhibiting Genotoxicity. *Chem Biol* **15**, 1258–1267 (2008).

254. Kerckhoffs, J. M. C. A. *et al.* Enantiomeric resolution of supramolecular helicates with different surface topographies. *Dalton Transactions* 734–742 (2007).

255. Khalid, S., Hannon, M. J., Rodger, A. & Mark Rodger, P. Simulations of DNA Coiling around a Synthetic Supramolecular Cylinder That Binds in the DNA Major Groove. *Chemistry – A European Journal* **12**, 3493–3506 (2006).

256. Zabost, E., Nowicka, A. M., Donten, M. & Stojek, Z. Substantial difference between temperature dependencies of dsDNA predenaturation process obtained by voltammetry and spectroscopy. *Physical Chemistry Chemical Physics* **11**, 8933–8938 (2009).

257. Fan, J. H., Bochkareva, E., Bochkarev, A. & Gray, D. M. Circular dichroism spectra and electrophoretic mobility shift assays show that human replication protein A binds and melts intramolecular G-quadruplex structures. *Biochemistry* **48**, 1099–1111 (2009).

258. Van Holde, K. E., Curtis Johnson, W. & Shing Ho, P. Principles of Physical Biochemistry (van Holde, Kersal E.; Johnson, W. Curtis; Ho, P. Shing). *J Chem Educ* **76**, 474 (1999).

259. Job, P. Formation and Stability of Inorganic Complexes in Solution. Annales de Chimie (France), **9**, 113-203 (1928).

260. Jenkins, T. C. Optical absorbance and fluorescence techniques for measuring DNA-drug interactions. *Methods Mol Biol* **90**, 195–218 (1997).

261. Santos, T., Salgado, G. F., Cabrita, E. J. & Cruz, C. G-Quadruplexes and Their Ligands: Biophysical Methods to Unravel G-Quadruplex/Ligand Interactions. *Pharmaceuticals* **14**, (2021).

262. Kwok, C. K. & Merrick, C. J. G-Quadruplexes: Prediction, Characterization, and Biological Application. *Trends Biotechnol* **35**, 997–1013 (2017).

263. Yang, S. Y. *et al.* Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nature Communications* **9**, 1–11 (2018).

264. Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S. & Balasubramanian, S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nature Methods* **13**, 841–844 (2016).

265. Hewitt, G. *et al.* Telomeres are favoured targets of a persistent DNA damage response in ageing and stress-induced senescence. *Nature Communications* **3**, 1–9 (2012).

266. 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* **38**, 5569–5580 (2010).

267. Besnard, E. *et al.* Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature Structural & Molecular Biology* **19**, 837–844 (2012).

268. Tian, T., Chen, Y. Q., Wang, S. R. & Zhou, X. G-Quadruplex: A Regulator of Gene Expression and Its Chemical Targeting. *Chem* **4**, 1314–1344 (2018).

269. Pietras, Z. *et al.* Dedicated surveillance mechanism controls G-quadruplex forming non-coding RNAs in human mitochondria. *Nature Communications* **9**, 1–15 (2018).

270. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **169**, 1187–1200 (2017).

271. Fay, M. M., Anderson, P. J. & Ivanov, P. ALS/FTD-Associated C9ORF72 Repeat RNA Promotes Phase Transitions In Vitro and in Cells. *Cell Rep* **21**, 3573–3584 (2017).

272. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).

273. Kharel, P., Becker, G., Tsvetkov, V. & Ivanov, P. Properties and biological impact of RNA G-quadruplexes: from order to turmoil and back. *Nucleic Acids Res* **48**, 12534–12555 (2020).

274. Bugaut, A. & Balasubramanian, S. 5′-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res* **40**, 4727–4741 (2012).

275. Monchaud, D. Quadruplex detection in human cells. *Annu Rep Med Chem* **54**, 133–160 (2020).

276. Beaume, N. *et al.* Genome-wide study predicts promoter-G4 DNA motifs regulate selective functions in bacteria: radioresistance of D. radiodurans involves G4 DNA-mediated regulation. *Nucleic Acids Res* **41**, 76–89 (2013).

277. Abiri, A. *et al.* Unlocking G-Quadruplexes as Antiviral Targets. *Pharmacol Rev* **73**, 897–923 (2021).

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

279. Neidle, S. Quadruplex Nucleic Acids as Novel Therapeutic Targets. *J Med Chem* **59**, 5987–6011 (2016).

280. Neidle, S. Quadruplex nucleic acids as targets for anticancer therapeutics. *Nature Reviews Chemistry* **1**, 1–10 (2017).

281. Ofer, N., Weisman-Shomer, P., Shklover, J. & Fry, M. The quadruplex r(CGG) n destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs

to increase the translation efficiency of fragile X premutation mRNA. *Nucleic Acids Res* **37**, 2712–2722 (2009).

282. Weisman-Shomer, P. *et al.* The cationic porphyrin TMPyP4 destabilizes the tetraplex form of the fragile X syndrome expanded sequence d(CGG)n. *Nucleic Acids Res* **31**, 3963–3970 (2003).

283. Moruno-Manchon, J. F. *et al.* Small-molecule G-quadruplex stabilizers reveal a novel pathway of autophagy regulation in neurons. *Elife* **9**, (2020).

284. Morris, M. J., Wingate, K. L., Silwal, J., Leeper, T. C. & Basu, S. The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Res* **40**, 4137–4145 (2012).

285. Rha SY, Izbicka E, Lawrence R, Davidson K, Sun D, Moyer MP, Roodman GD, Hurley L, Von Hoff D. Effect of telomere and telomerase interactive agents on human tumor and normal cell lines. *Clin Cancer Res.* **3** 987-993 (2000).

286. Grand CL, Han H, Muñoz RM, Weitman S, Von Hoff DD, Hurley LH, Bearss DJ. The Cationic Porphyrin TMPyP4 Down-Regulates c-MYC and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth in vivo. *Mol Cancer Ther.* **8**, 565-73 2002

287. Artusi, S. *et al.* Antiviral Activity of the G-Quadruplex Ligand TMPyP4 against Herpes Simplex Virus-1. *Viruses* **13**, 196 (2021).

288. Phan, A. T., Kuryavyi, V., Gaw, H. Y. & Patel, D. J. Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nature Chemical Biology* **1**, 167–173 (2005).

289. Parkinson, G. N., Ghosh, R. & Neidle, S. Structural Basis for Binding of Porphyrin to Human Telomeres†,‡. *Biochemistry* **46**, 2390–2397 (2007).

290. Zhang, Y. *et al.* G2-quadruplex in the 3'UTR of IE180 regulates Pseudorabies virus replication by enhancing gene expression. *RNA Biol* **17**, 816–827 (2020).

291. Zhang, Y. *et al.* Native de novo structural determinations of non-canonical nucleic acid motifs by X-ray crystallography at long wavelengths. *Nucleic Acids Res* **48**, 9886–9898 (2020).

292. Nagesh, N. & Ganesh Kumar, A. Interaction of TMPyP4, TMPyP3, and TMPyP2 with Intramolecular G-Quadruplex Formed by Promoter Region of Bcl2 and KRAS NHPPE. *ISRN Biophysics* **2012**, 1–12 (2012).

293. 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* **8**, e72462 (2013).

294. Perrone, R. *et al.* Formation of a Unique Cluster of G-Quadruplex Structures in the HIV-1 nef Coding Region: Implications for Antiviral Activity. *PLoS One* **8**, e73121 (2013).

295. Chilakamarthi, U. *et al.* Novel Amphiphilic G-Quadruplex Binding Synthetic Derivative of TMPyP4 and Its Effect on Cancer Cell Proliferation and Apoptosis Induction. *Biochemistry* **57**, 6514–6527 (2018).

296. Pradeep, T. P., Tripathi, S. & Barthwal, R. Molecular recognition of parallel quadruplex [d-(TTGGGGT)]4 by mitoxantrone: binding with 1 : 4 stoichiometry leads to telomerase inhibition. *RSC Adv* **6**, 71652–71661 (2016).

297. White, E. W. *et al.* Structure-specific recognition of quadruplex DNA by organic cations: Influence of shape, substituents and charge. *Biophys Chem* **126**, 140–153 (2007).

298. Zamiri, B., Reddy, K., Macgregor, R. B. & Pearson, C. E. TMPyP4 Porphyrin Distorts RNA G-quadruplex Structures of the Disease-associated r(GGGGCC)n Repeat of the C9orf72 Gene and Blocks Interaction of RNA-binding Proteins. *Journal of Biological Chemistry* **289**, 4653–4659 (2014).

299. Martino, L., Pagano, B., Fotticchia, I., Neidle, S. & Giancola, C. Shedding Light on the Interaction between TMPyP4 and Human Telomeric Quadruplexes. *J Phys Chem B* **113**, 14779–14786 (2009).

300. Huppert, J. L., Bugaut, A., Kumari, S. & Balasubramanian, S. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res* **36**, 6260–6268 (2008).

301. Cogoi, S. & Xodo, L. E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res* **34**, 2536–2549 (2006).

302. 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. *Proceedings of the National Academy of Sciences* **99**, 11593–11598 (2002).

303. Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* **35**, 406–413 (2007).

304. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* **5**, 182–186 (2013).

305. Kang, H. *et al.* Molecular population dynamics of DNA structures in a bcl-2 promoter sequence is regulated by small molecules and the transcription factor hnRNP LL. *Nucleic Acids Res* **42**, 5755–5764 (2014).

306. Liu, D. *et al.* A Proton-Fuelled DNA Nanomachine. *Angewandte Chemie International Edition* **42**, 5734–5736 (2003).

307. Li, T. & Famulok, M. I - Motif-Programmed Functionalization of DNA Nanocircles. *J Am Chem Soc,* **4**, 1593-1599 (2013).

308. Zeraati, M. *et al.* I-motif DNA structures are formed in the nuclei of human cells. *Nature Chemistry* **10**, 631–637 (2018).

309. Tang, W. *et al.* In vivo visualization of the i-motif DNA secondary structure in the Bombyx mori testis. *Epigenetics Chromatin* **13**, 12 (2020).

310. Kosiol, N., Juranek, S., Brossart, P., Heine, A. & Paeschke, K. G-quadruplexes: a promising target for cancer therapy. *Molecular Cancer* **20**, 1–18 (2021).

311. Local, A. *et al.* APTO-253 Stabilizes G-quadruplex DNA, Inhibits MYC Expression, and Induces DNA Damage in Acute Myeloid Leukemia Cells. *Mol Cancer Ther* **17**, 1177–1186 (2018).

312. Tauchi, T. *et al.* Activity of a novel G-quadruplex-interactive telomerase inhibitor, telomestatin (SOT-095), against human leukemia cells: involvement of ATM-dependent DNA damage response pathways. *Oncogene* **22**, 5338–5347 (2003).

313. Liu, J. N. *et al.* Inhibition of myc promoter and telomerase activity and induction of delayed apoptosis by SYUIQ-5, a novel G-quadruplex interactive agent in leukemia cells. *Leukemia* **21**, 1300–1302 (2007).

314. Zidanloo, S. G., Hosseinzadeh Colagar, A., Ayatollahi, H. & Raoof, J. B. Downregulation of the WT1 gene expression via TMPyP4 stabilization of promoter G-quadruplexes in leukemia cells. *Tumour Biol* **37**, 9967–9977 (2016).

315. Nakajima, A. *et al.* Telomerase inhibition enhances apoptosis in human acute leukemia cells: possibility of antitelomerase therapy. *Leukemia* **17**, 560–567 (2003).

316. Tsuruyama, T. *et al.* Constitutive activation of Stat5a by retrovirus integration in early pre-B lymphomas of SL/Kh strain mice. *Proceedings of the National Academy of Sciences* **99**, 8253–8258 (2002).

317. Nielsen, A. A., Sørensen, A. B., Schmidt, J. & Pedersen, F. S. Analysis of Wild-Type and Mutant SL3-3 Murine Leukemia Virus Insertions in the c-myc Promoter during Lymphomagenesis Reveals Target Site Hot Spots, Virus-Dependent Patterns, and Frequent Error-Prone Gap Repair . *J Virol* **79**, 67–78 (2005).

318. Tsuruyama, T., Liu, W. & Yoshikawa, K. In Vitro Murine Leukemia Retroviral Integration and Structure Fluctuation of Target DNA. *PLoS One* **7**, e31533 (2012).

319. Markovitsi, D., Sage, E., Lewis, F. D. & Davies, J. Interaction of UV radiation with DNA. *Photochemical and Photobiological Sciences* **12** 1256–1258 (2013).

320. Vizcaíno, C., Mansilla, S. & Portugal, J. Sp1 transcription factor: A long-standing target in cancer chemotherapy. *Pharmacol Ther* **152**, 111–124 (2015).

321. Suske, G. The Sp-family of transcription factors. *Gene* **238**, 291–300 (1999).

322. Hurley, L. H. DNA and its associated processes as targets for cancer therapy. *Nature Reviews Cancer* **2**, 188–200 (2002).

323. Girgis, H. Z., James, B. T. & Luczak, B. B. Identity: rapid alignment-free prediction of sequence alignment identity scores using self-supervised general linear models. *NAR Genom Bioinform* **3**, (2021).

324. Romano, M. & Clària, J. Cyclooxygenase-2 and 5-lipoxygenase converging functions on cell proliferation and tumor angiogenesis: implications for cancer therapy. *The FASEB Journal* **17**, 1986–1995 (2003).

325. Chen, Y., Hu, Y., Zhang, H., Peng, C. & Li, S. Loss of the Alox5 gene impairs leukemia stem cells and prevents chronic myeloid leukemia. *Nature Genetics* **41**, 783–792 (2009).

326. Samuelsson, B., Dahlén, S. E., Lindgren, J. Å., Rouzer, C. A. & Serhan, C. N. Leukotrienes and lipoxins: Structures, biosynthesis, and biological effects. *Science (1979)* **237**, 1171–1176 (1987).

327. Funk, C. D. Leukotriene modifiers as potential therapeutics for cardiovascular disease. *Nature Reviews Drug Discovery,* **4** 664–672 (2005).

328. Funk, C. D., Hoshiko, S., Matsumoto, T., Radmark, O. & Samuelsson, B. Characterization of the human 5-lipoxygenase gene. *Proc Natl Acad Sci U S A* **86**, 2587–2591 (1989).

329. Silverman, E. S. *et al.* Egr-1 and Sp1 interact functionally with the 5-Lipoxygenase promoter and its naturally occurring mutants. *Am J Respir Cell Mol Biol* **19**, 316–323 (1998).

330. Dishart, D. *et al.* GC-rich sequences in the 5-lipoxygenase gene promoter are required for expression in Mono Mac 6 cells, characterization of a novel Sp1 binding site. *Biochim Biophys Acta Mol Cell Biol Lipids* **1738**, 37–47 (2005).

331. Hoshiko, S., Rådmark, O. & Samuelsson, B. Characterization of the human 5-lipoxygenase gene promoter. *Proc Natl Acad Sci U S A* **87**, 9073–9077 (1990).

332. In, K. H. *et al.* Naturally occurring mutations in the human 5-lipoxygenase gene promoter that modify transcription factor binding and reporter gene transcription. *Journal of Clinical Investigation* **99**, 1130–1137 (1997).

333. Fleming, A. M. *et al.* 4n-1 Is a 'Sweet Spot' in DNA i-Motif Folding of 2'-Deoxycytidine Homopolymers. *J Am Chem Soc* **139**, 4682–4689 (2017).

334. Martella, M. *et al.* i-Motif formation and spontaneous deletions in human cells. *Nucleic Acids Res* **50**, 3445–3455 (2022).

335. Víglaský, V., Bauer, L. & Tlučková, K. Structural features of intra- and intermolecular G-quadruplexes derived from telomeric repeats. *Biochemistry* **49**, 2110–2120 (2010).

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

337. Mergny, J. L. & Lacroix, L. Kinetics and thermodynamics of i-DNA formation: Phosphodiester versus modified oligodeoxynucleotides. *Nucleic Acids Res* **26**, 4797–4803 (1998).

338. Jean-Louis Mergny, Laurent Lacroix, Xiaogang Han, Jean-Louis Leroy, and Claude Helene. Intramolecular folding of pyrimidine oligodeoxynucleotides into an i-DNA motif. *Journal of the American Chemical Society* **35**, 8887-8898 (1995).

339. Hänsel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory chromatin. *Nature Genetics* **48**, 1267–1272 (2016).

340. Sutherland, C., Cui, Y., Mao, H. & Hurley, L. H. A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the *MYC* Promoter NHE III $_1$. *J Am Chem Soc* **138**, 14138–14151 (2016).

341. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* **34**, 5402–5415 (2006).

342. Largy, E., Marchand, A., Amrane, S., Gabelica, V. & Mergny, J. L. Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches. *J Am Chem Soc* **138**, 2780–2792 (2016).

343. Bischof, H. *et al.* Live-Cell Imaging of Physiologically Relevant Metal Ions Using Genetically Encoded FRET-Based Probes. *Cells* **8**, 492 (2019).

344. Saintomé, C., Amrane, S., Mergny, J. L. & Alberti, P. The exception that confirms the rule: A higher-order telomeric G-quadruplex structure more stable in sodium than in potassium. *Nucleic Acids Res* **44**, 2926–2935 (2016).

345. Liao, J. Y., Anosova, I., Bala, S., Van Horn, W. D. & Chaput, J. C. A parallel stranded G-quadruplex composed of threose nucleic acid (TNA). *Biopolymers* **107**, (2017).

346. Rachwal, P. A., Brown, T. & Fox, K. R. Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes. *Biochemistry* **46**, 3036–3044 (2007).

347. Mukundan, V. T. & Phan, A. T. Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J Am Chem Soc* **135**, 5017–5028 (2013).

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

349. Bartas, M., Brázda, V., Karlický, V., Červeň, J. & Pečinka, P. Bioinformatics analyses and in vitro evidence for five and six stacked G-quadruplex forming sequences. *Biochimie* **150**, 70–75 (2018).

350. He, Y., Neumann, R. D. & Panyutin, I. G. Intramolecular quadruplex conformation of human telomeric DNA assessed with 125I-radioprobing. *Nucleic Acids Res* **32**, 5359–5367 (2004).

351. Venczel, E. A. & Sen, D. Parallel and Antiparallel G-DNA Structures from a Complex Telomeric Sequence*. *Biochemistry* **32**, 6220–6228 (1993).

352. Kim, B. G., Evans, H. M., Dubins, D. N. & Chalikian, T. V. Effects of Salt on the Stability of a G-Quadruplex from the Human c-MYC Promoter. *Biochemistry* **54**, 3420–3430 (2015).

353. Kim, B. G., Shek, Y. L. & Chalikian, T. V. Polyelectrolyte effects in G-quadruplexes. *Biophys Chem* **184**, 95–100 (2013).

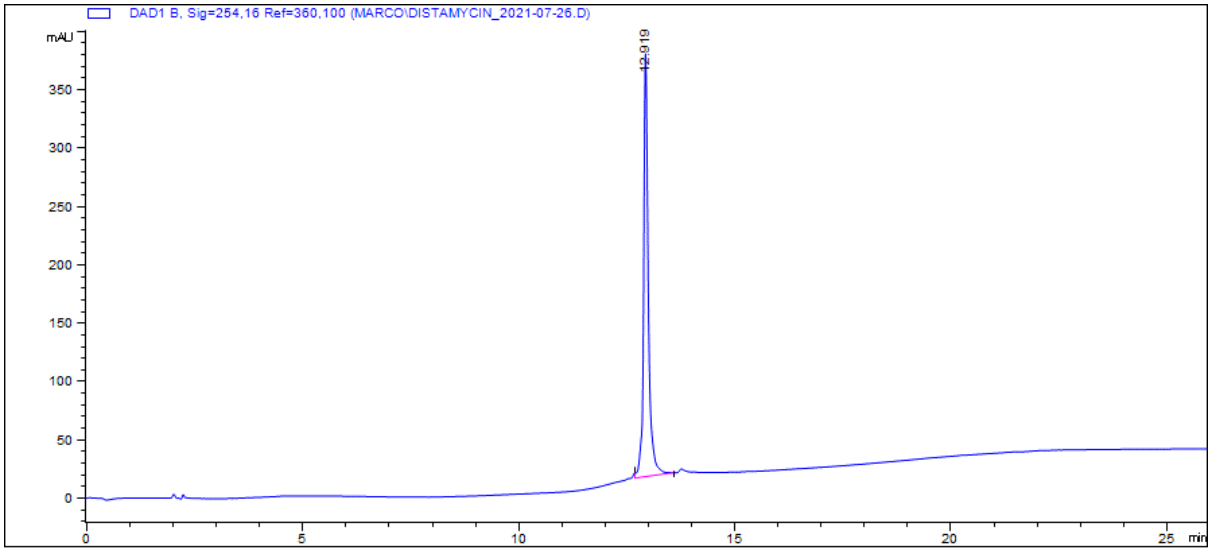
354. Von Hoff, D. D., Rozencweig, M. & Slavik, M. Daunomycin: an anthracycline antibiotic effective in acute leukemia. *Adv Pharmacol Chemother* **15**, 1–50 (1978).

355. Moore, M. H., Hunter, W. N., d'Estaintot, B. L. & Kennard, O. DNA-drug interactions. The crystal structure of d(CGATCG) complexed with daunomycin. *J Mol Biol* **206**, 693–705 (1989).

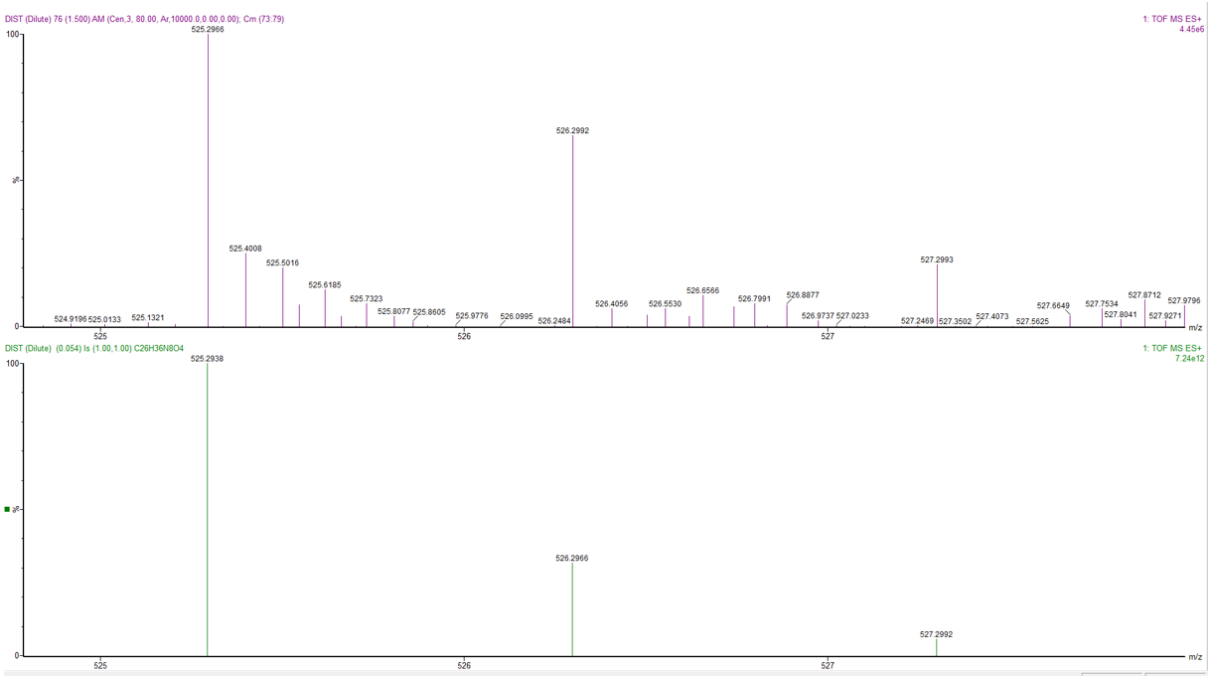
356. Clark, G. R., Pytel, P. D., Squire, C. J. & Neidle, S. Structure of the first parallel DNA quadruplex-drug complex. *J Am Chem Soc* **125**, 4066–4067 (2003).

Appendix

Appendix 1 *HPLC trace of PyPyPyβDp at 214 nm, using 5% MeOH: H2O, 0.05% TFA*



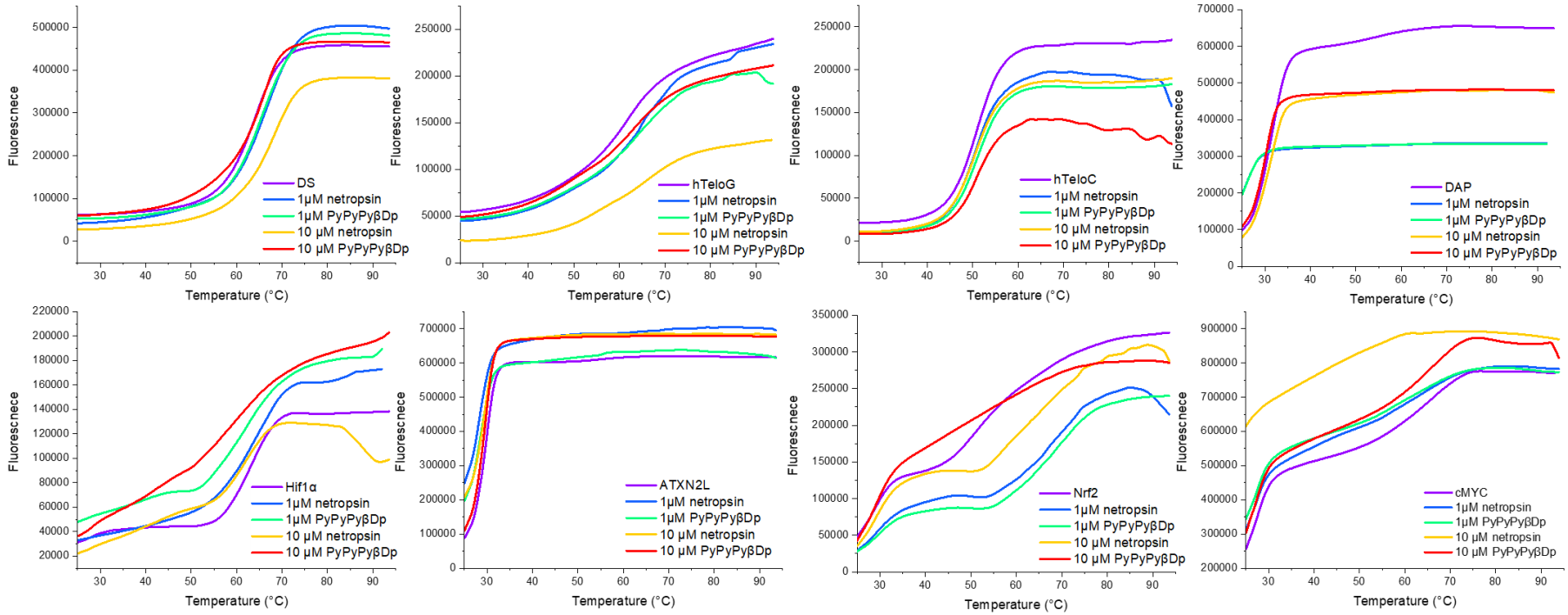
Appendix 2 *MALDI mass spectroscopy of PyPyPyβDp, samples were prepared at 1 mg/mL in MeOH.*



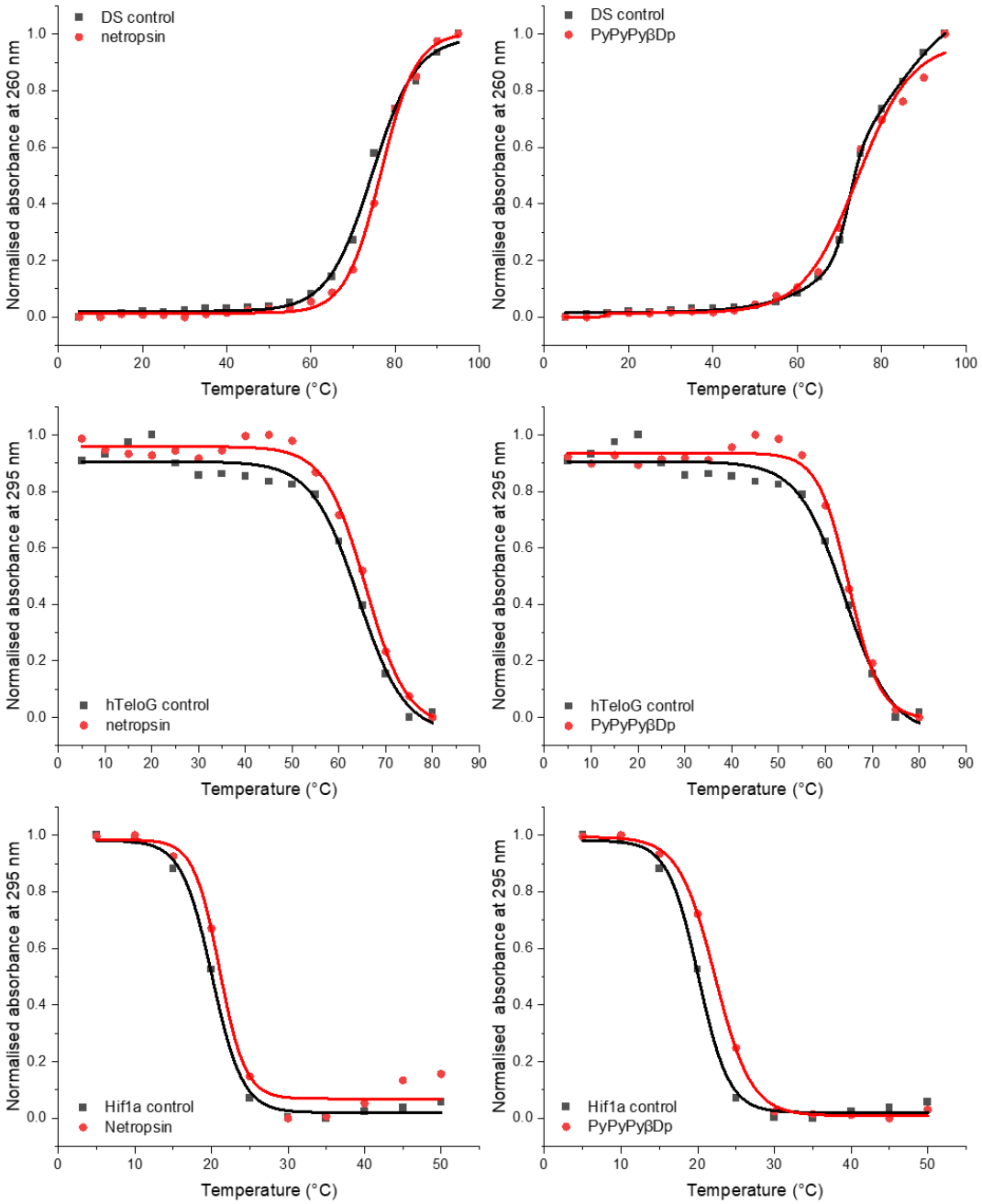
Appendix 3 *Percentage of TO displacement of minor groove binders with tested DNA sequences, the concentration of DNA was 0.5 μM, the buffer is 10 mM sodium cacodylate and 100 mM potassium chloride at desired pH (pH 7.0 for DS, hTeloG, DAP, and Hif1a, pH 6.5 for Nrf2 and c-MYC, pH 5.5 for hTeloC). Errors show the standard deviation across three repeats.*

TO displacement (%)							
	[PyPyPyβDp]						
DNA	1 μM	2.5 μM	5 μM	10 μM	20 μM	50 μM	100 μM
DS	-22.5 ± 8.6	-11.4 ± 3.5	-13.2 ± 5.1	-3.6 ± 5.5	1.3 ± 6.5	-11.8 ± 6.1	-18.8 ± 7.8
*hTelo*G	-4.1 ± 1.3	-6.6 ± 2.1	-7.3 ± 3.0	-6.7 ± 3.2	-10.8 ± 3.4	-9.2 ± 6.8	2.1 ± 2.1
*hTelo*C	-14.1 ± 5.7	-7.1 ± 1.1	-8.0 ± 1.1	-12.2 ± 10.7	2.9 ± 6.1	-18.9 ± 4.4	-46.3 ± 6.0
Nrf2	-6.4 ± 2.6	-6.2 ± 4.0	-4.6 ± 2.6	-1.4 ± 4.1	3.0 ± 0.8	1.4 ± 3.4	4.6 ± 1.2
DAP	7.7 ± 1.4	6.8 ± 0.5	6.1 ± 0.6	7.0 ± 1.1	6.8 ± 0.8	10.6 ± 0.4	7.9 ± 2.5
Hif1a	-0.2 ± 1.8	-0.9 ± 4.0	-8.2 ± 5.9	-7.0 ± 0.2	-5.2 ± 1.3	-3.1 ± 3.6	-1.9 ± 1.0
c-MYC	-34.7 ± 15.6	-9.7 ± 12.8	-7.7 ± 4.6	-25.1 ± 13.4	-6.7 ± 16.8	-122.7 ± 13.3	-122.9 ± 12.9
	[netropsin]						
DS	-4.4 ± 1.3	-3.9 ± 4.9	15.6 ± 2.1	22.3 ± 2.9	30.1 ± 2.5	40.2 ± 6.4	53.3 ± 10.0
*hTelo*G	-5.8 ± 1.8	0.2 ± 2.8	1.6 ± 2.9	3.6 ± 1.5	7.4 ± 0.6	15.1 ± 1.1	18.4 ± 4.3
*hTelo*C	-6.7 ± 5.5	-14.5 ± 5.2	0.2 ± 0.9	2.2 ± 6.8	6.3 ± 3.5	16.2 ± 7.1	30.1 ± 7.4
Nrf2	-1.4 ± 1.2	-6.6 ± 5.5	-1.7 ± 1.2	0.8 ± 1.1	3.2 ± 0.6	12.9 ± 0.5	26.4 ± 2.0
DAP	1.8 ± 1.7	10.3 ± 1.1	0.7 ± 0.9	1.4 ± 3.2	3.8 ± 0.7	7.8 ± 1.1	9.8 ± 3.4
Hif1a	-4.7 ± 1.4	5.5 ± 4.1	14.5 ± 1.5	23.5 ± 1.3	32.9 ± 1.2	48.3 ± 1.3	65.8 ± 1.9
c-MYC	14.4 ± 10.3	-1.4 ± 5.2	5.2 ± 10.9	-20.4 ± 12.4	1.1 ± 9.6	-10.2 ± 12.5	0.9 ± 7.8

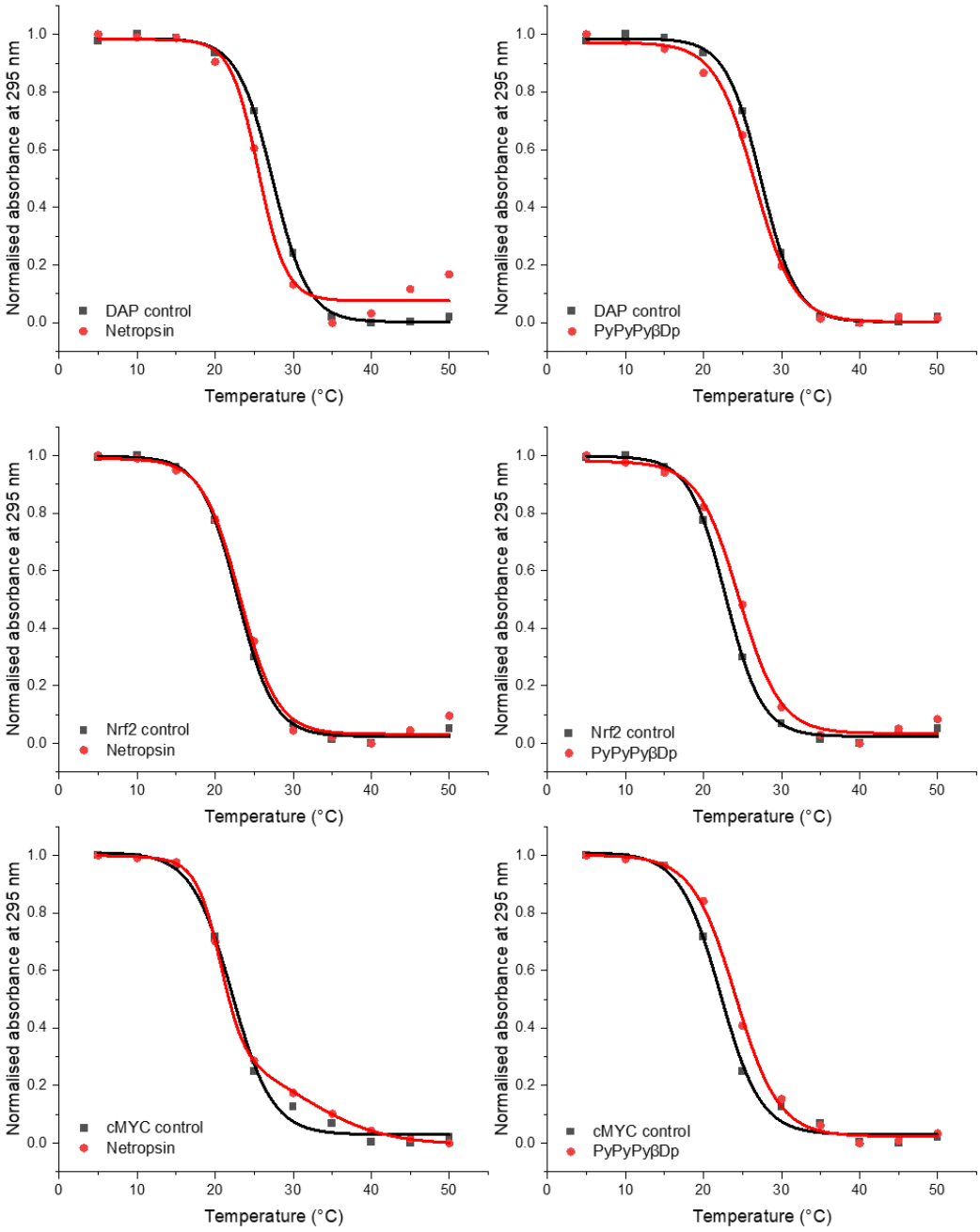
Appendix 4 *FRET melting curves for minor groove binders with DNA sequences, the concentration of DNA was 0.2 μM, the buffer is 10 mM sodium cacodylate and 100 mM potassium chloride at desired pH (pH 7.0 for DS, hTeloG, DAP, and Hif1a, pH 6.5 for Nrf2 and c-MYC, ph 5.5 for hTeloC), concentrations of minor groove binders were 1 μM and 10 μM.*



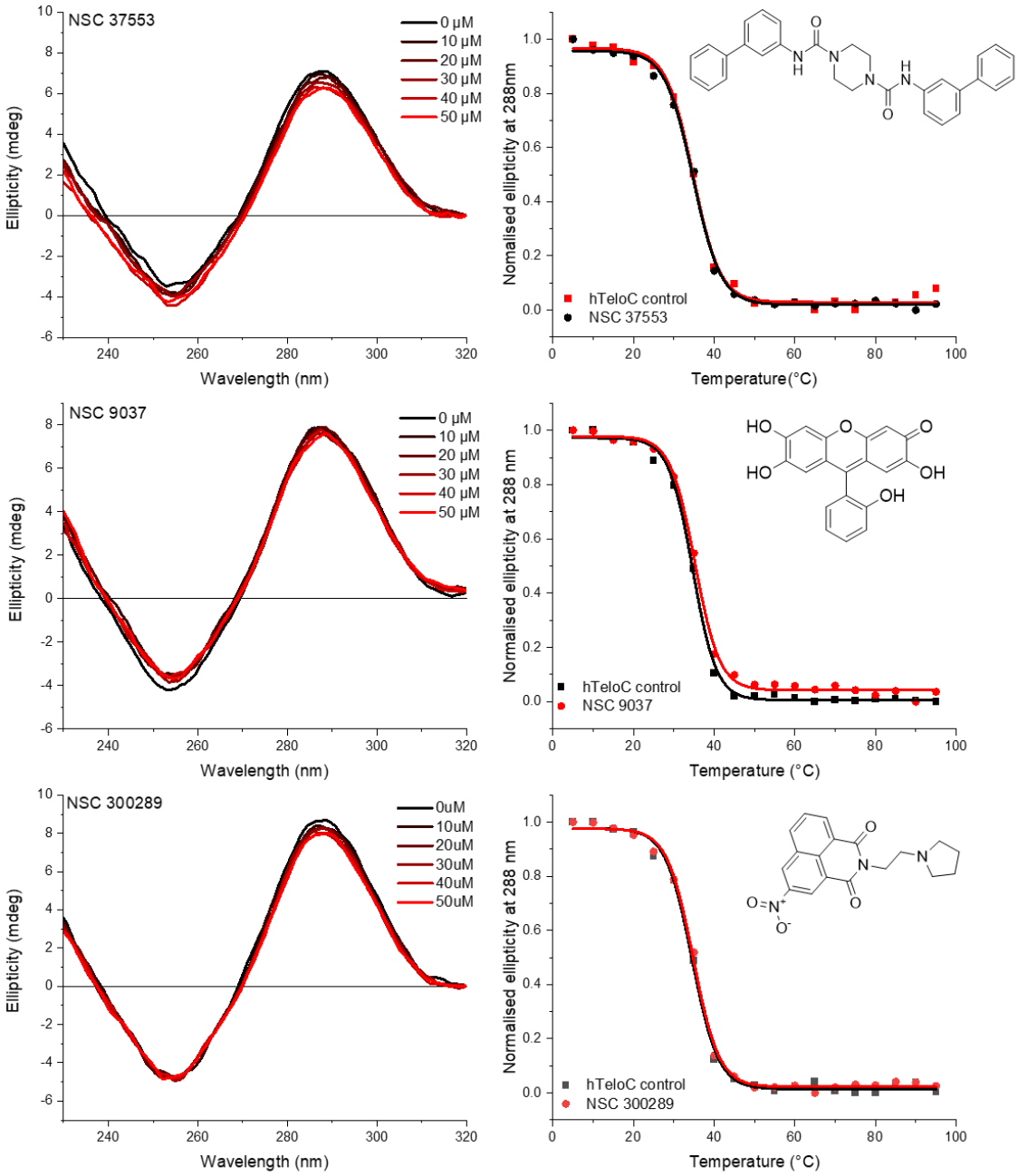
Appendix 5 *UV melting of DS, hTeloG, and Hif1a in the absence and presence of 50 μM of minor groove binders. [DNA] is 10 μM, and all experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 7.0. The data were fitted by sigmoidal curve using Origin 2022.*

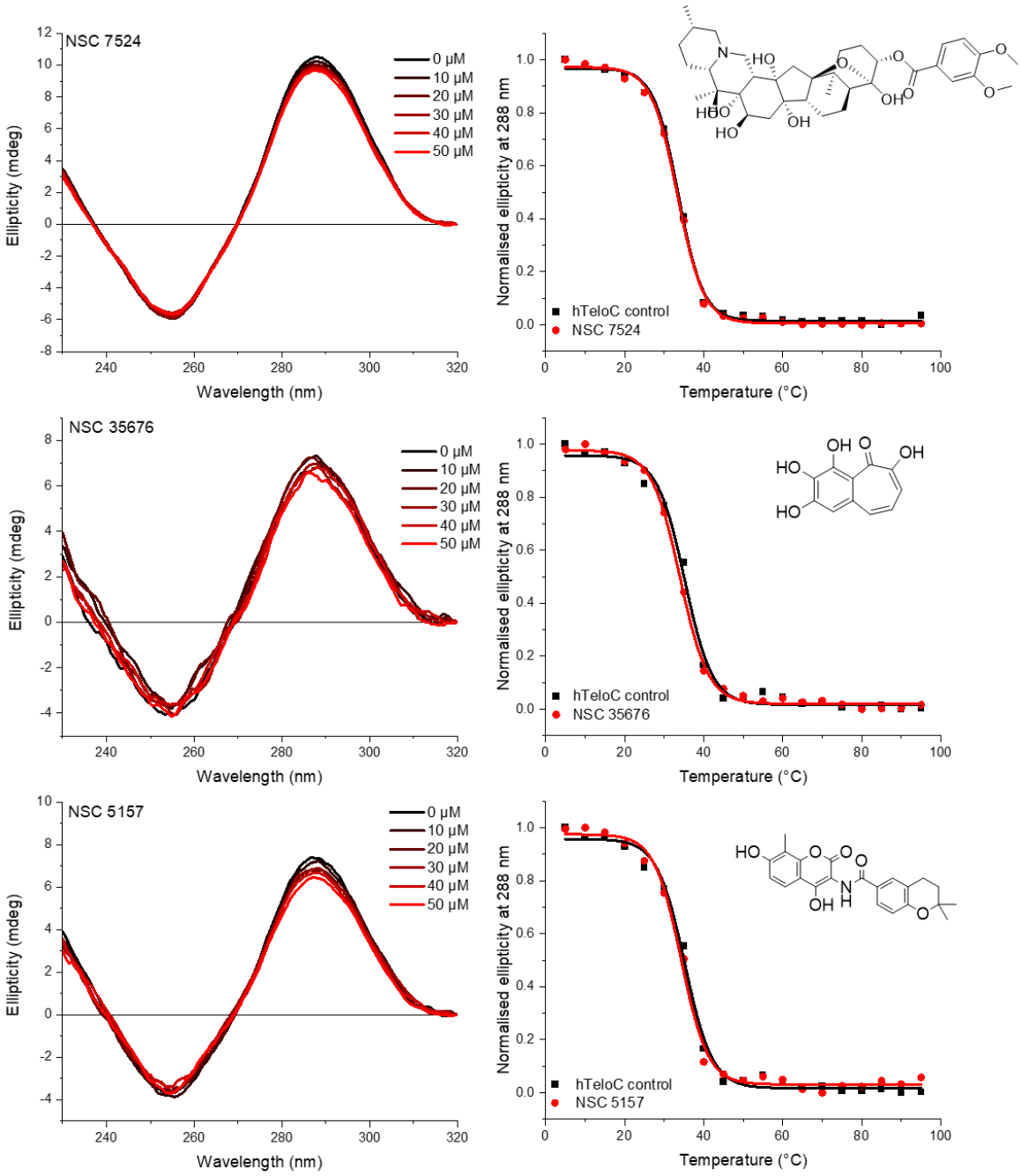


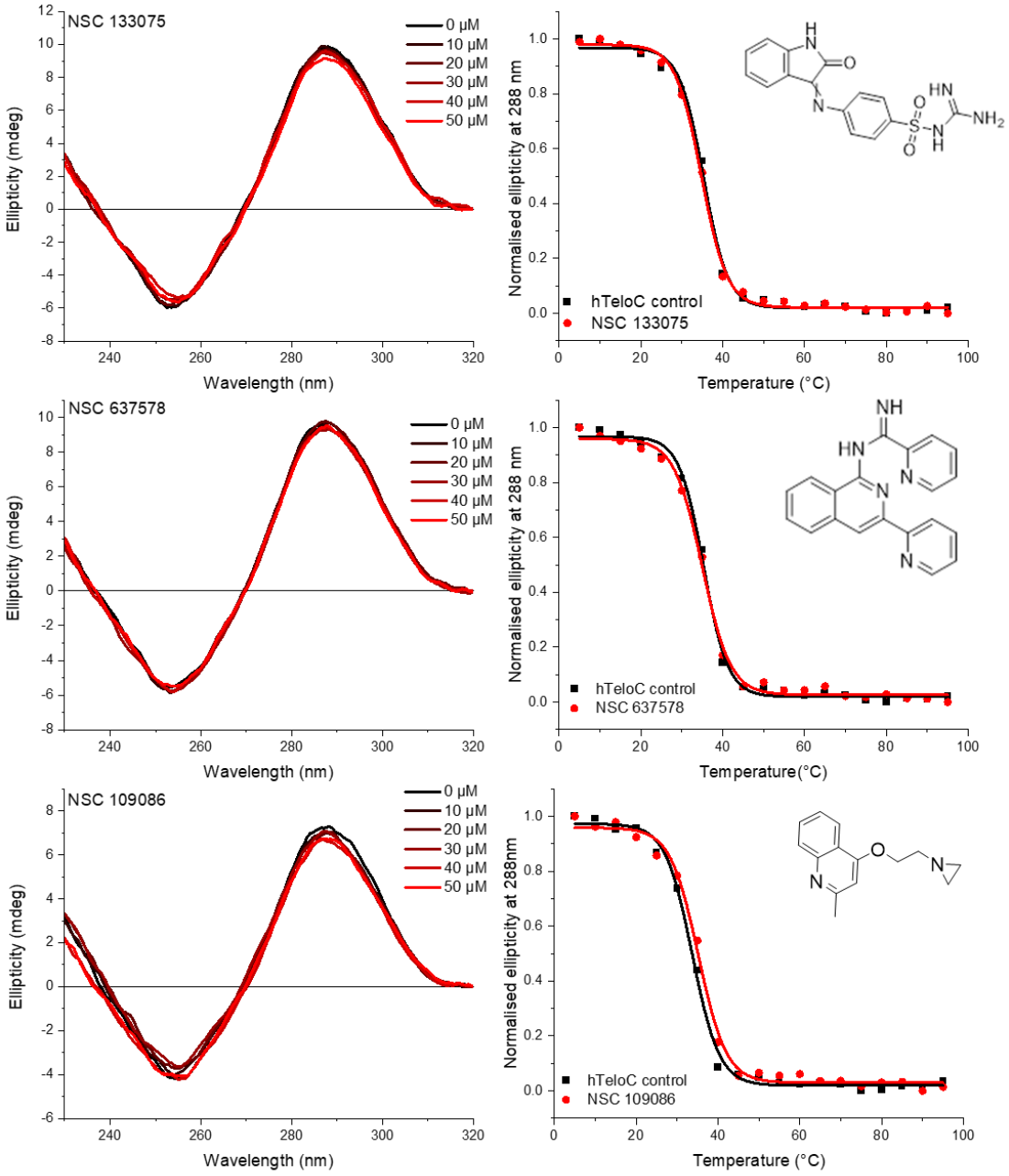
Appendix 6 *UV melting of DAP, Nrf2, and c-MYC i-motif in the absence and presence of 50 μM of minor groove binders. [DNA] is 10 μM, and all experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride at pH 7.0 for DAP, pH 6.5 for Nrf2, and c-MYC. The data were fitted by sigmoidal curve using Origin 2022.*

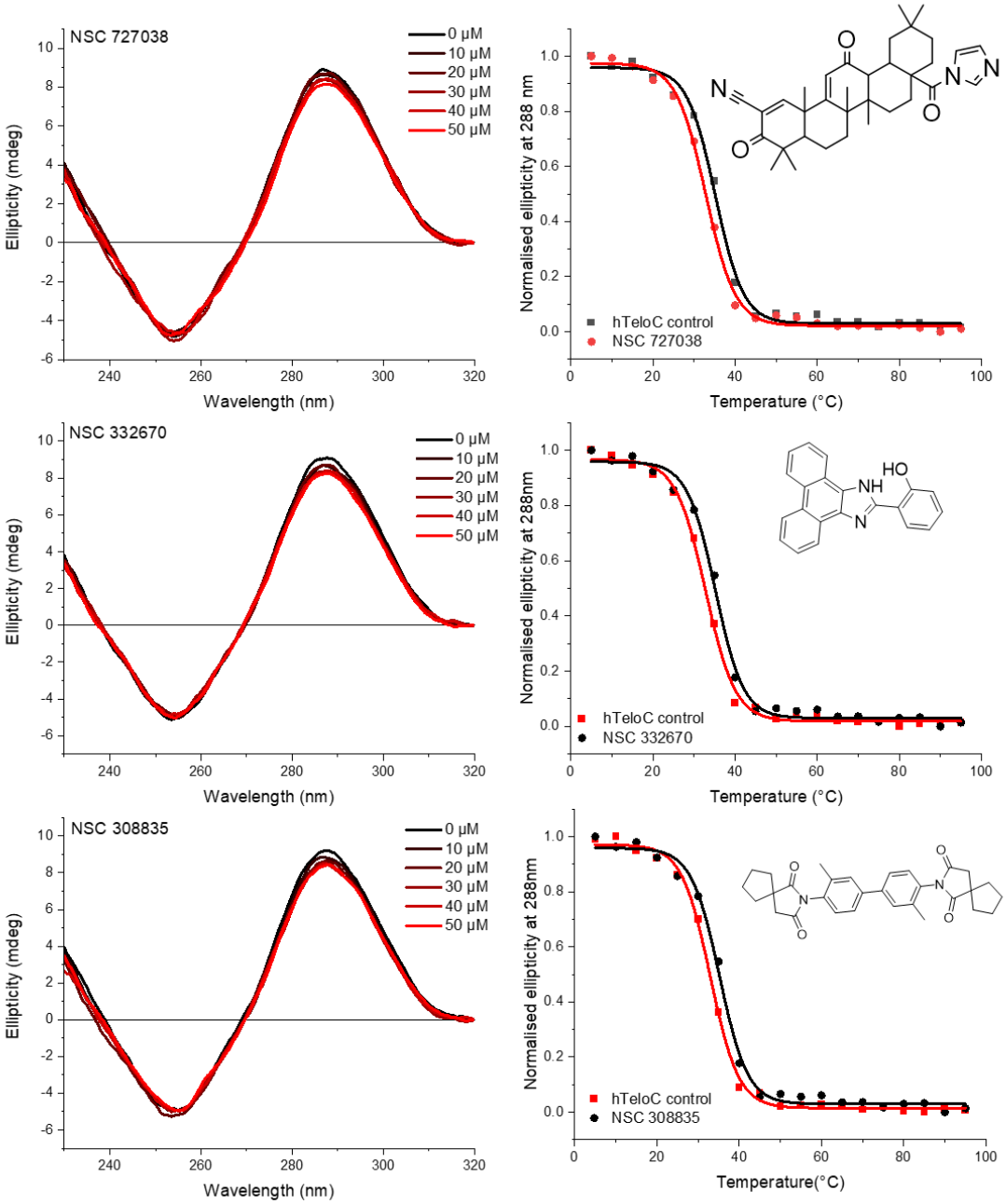


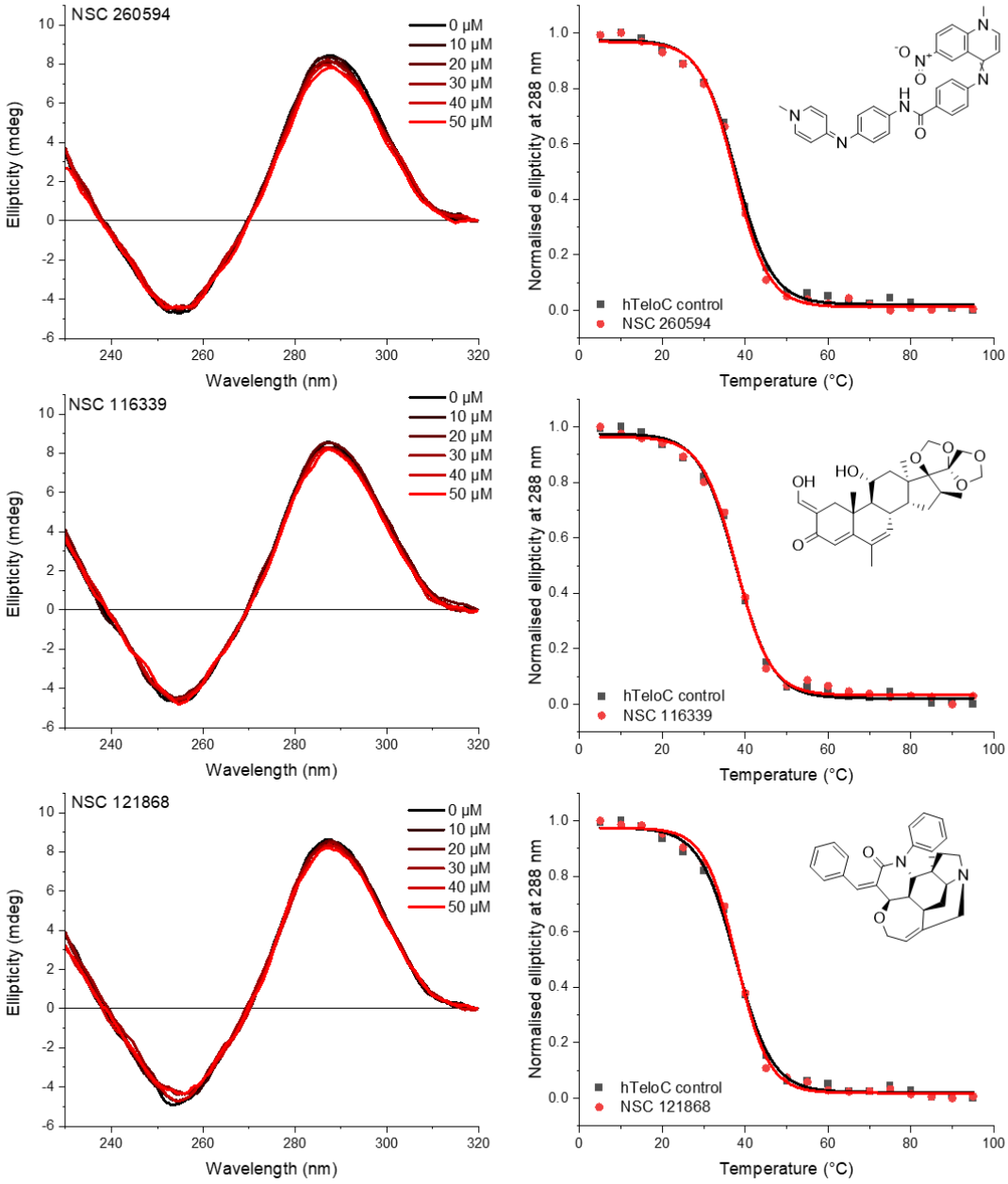
Appendix 7 *CD titration and melting of NCI ligands with structures, all experiments were performed in 50 mM sodium cacodylate at pH 6.0, [hTeloC] is 10μM. The plot of normalized ellipticity at 288 nm against temperature gave sigmoidal curves.*

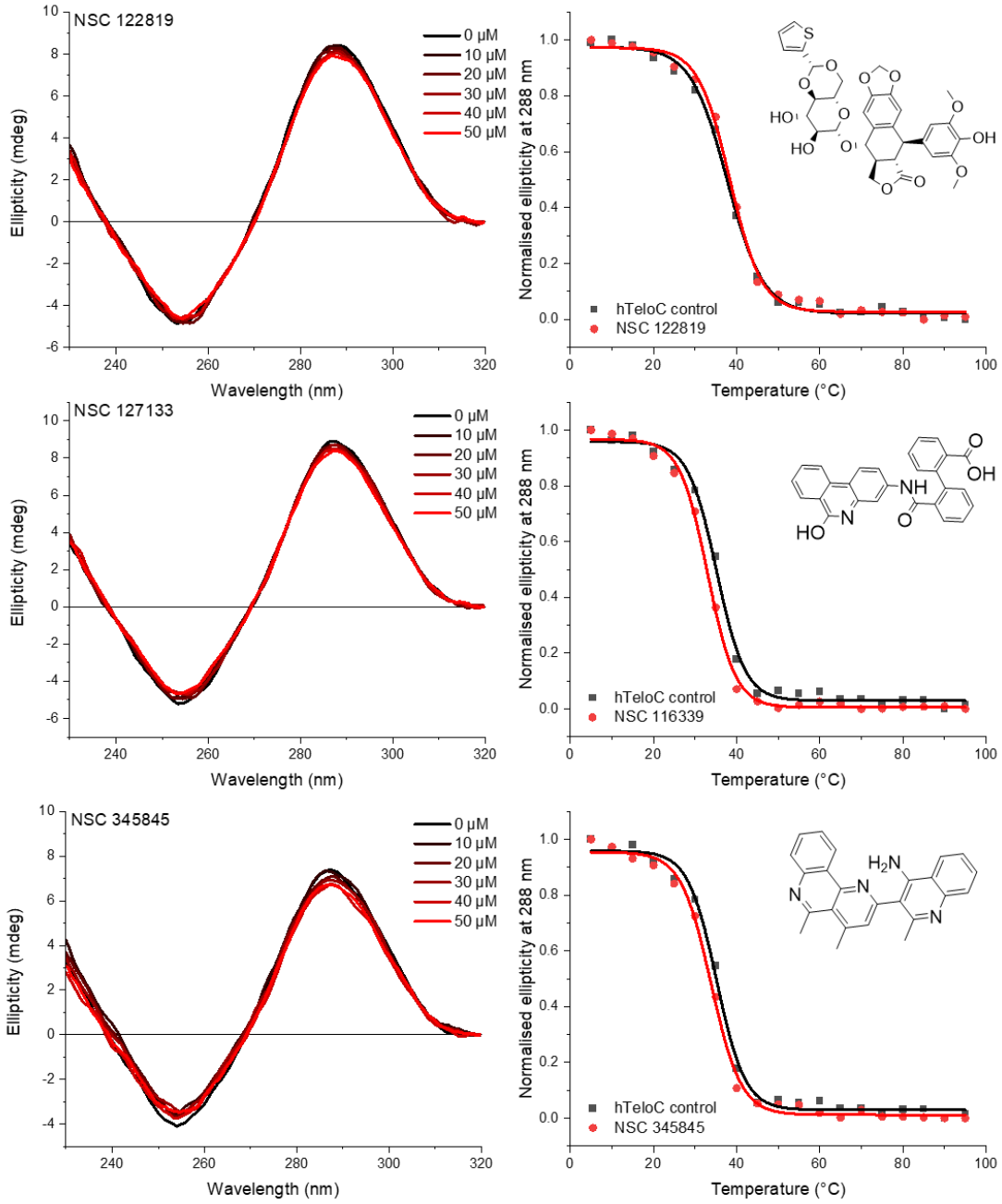


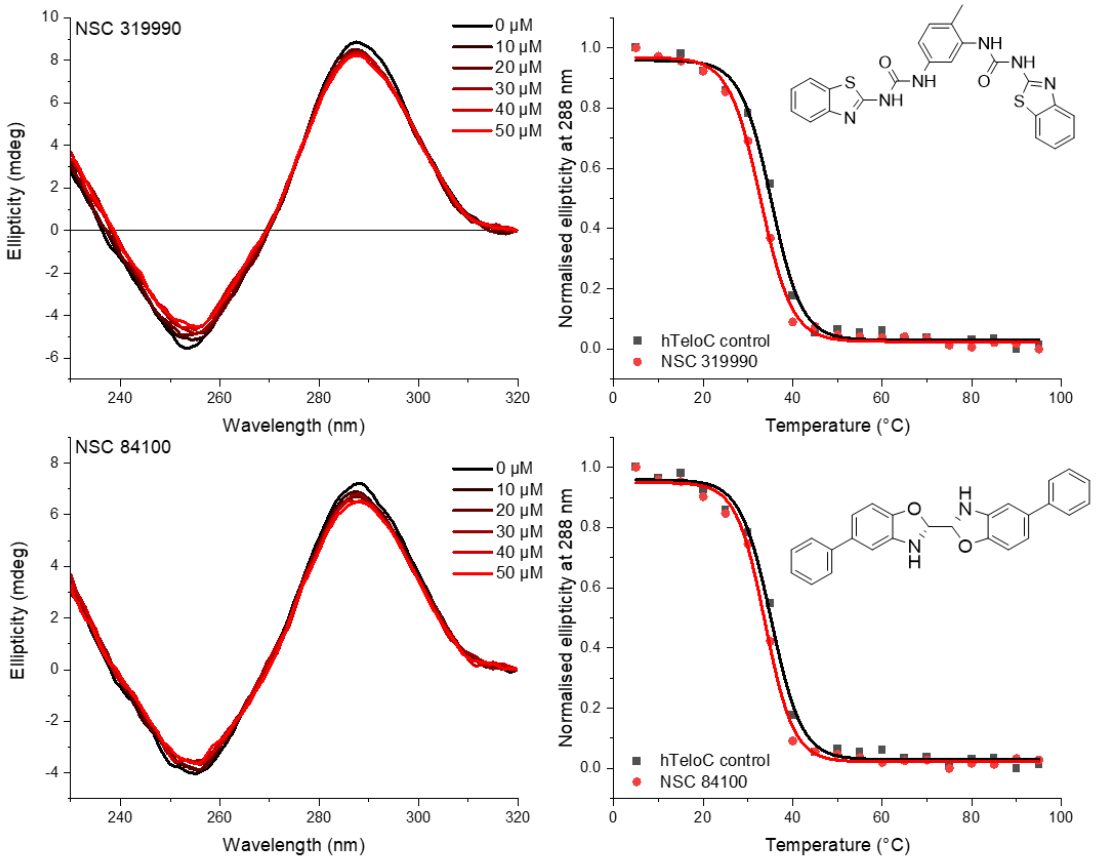

NSC 7524
0 µM
10 µM
20 µM
30 µM
40 µM
50 µM
Ellipticity (mdeg)
Wavelength (nm)
Normalised ellipticity at 288 nm
Temperature (°C)
hTeloC control
NSC 7524
NSC 35676
hTeloC control
NSC 35676
NSC 5157
hTeloC control
NSC 5157

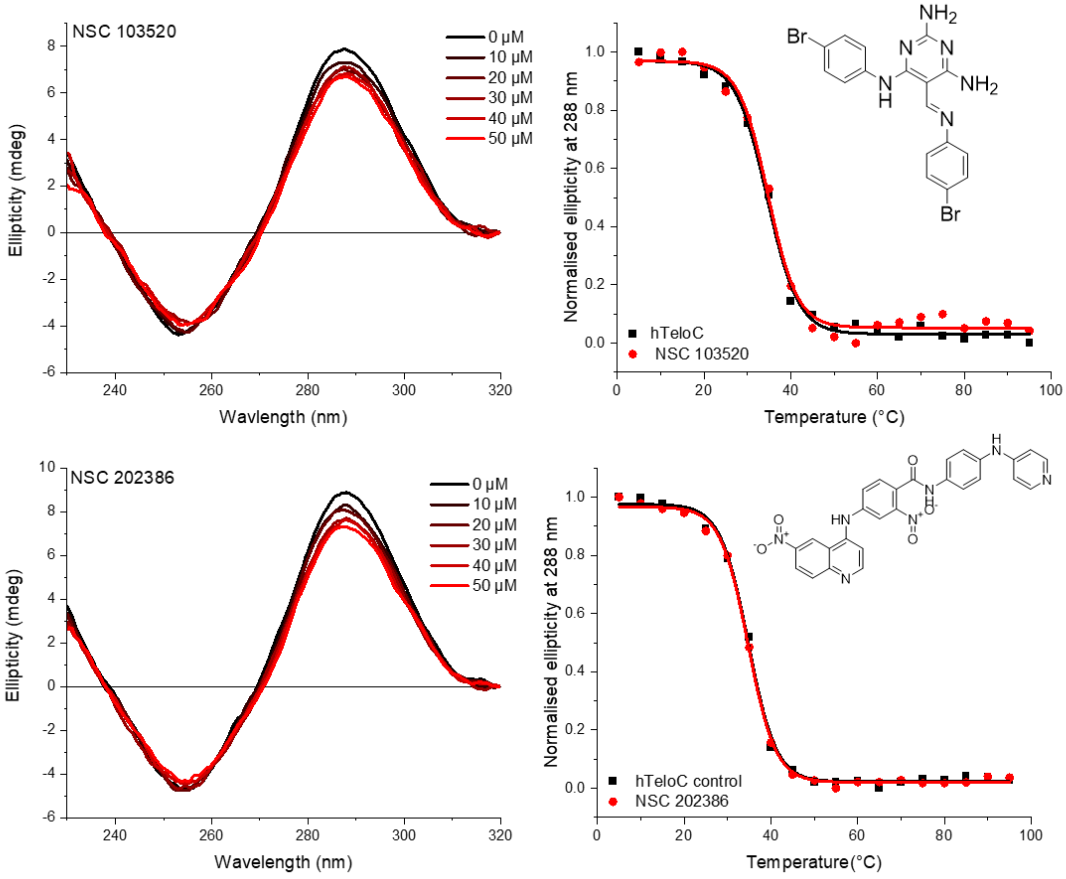

NSC 133075
0 µM
10 µM
20 µM
30 µM
40 µM
50 µM
Ellipticity (mdeg)
Wavelength (nm)
Normalised ellipticity at 288 nm
Temperature (°C)
hTeloC control
NSC 133075
NSC 637578
hTeloC control
NSC 637578
NSC 109086
Normalised ellipticity at 288nm
hTeloC control
NSC 109086


NSC 727038
0 μM
10 μM
20 μM
30 μM
40 μM
50 μM
Ellipticity (mdeg)
Wavelength (nm)
Normalised ellipticity at 288 nm
Temperature (°C)
hTeloC control
NSC 727038
NSC 332670
Normalised ellipticity at 288nm
hTeloC control
NSC 332670
NSC 308835
Temperature(°C)
hTeloC control
NSC 308835

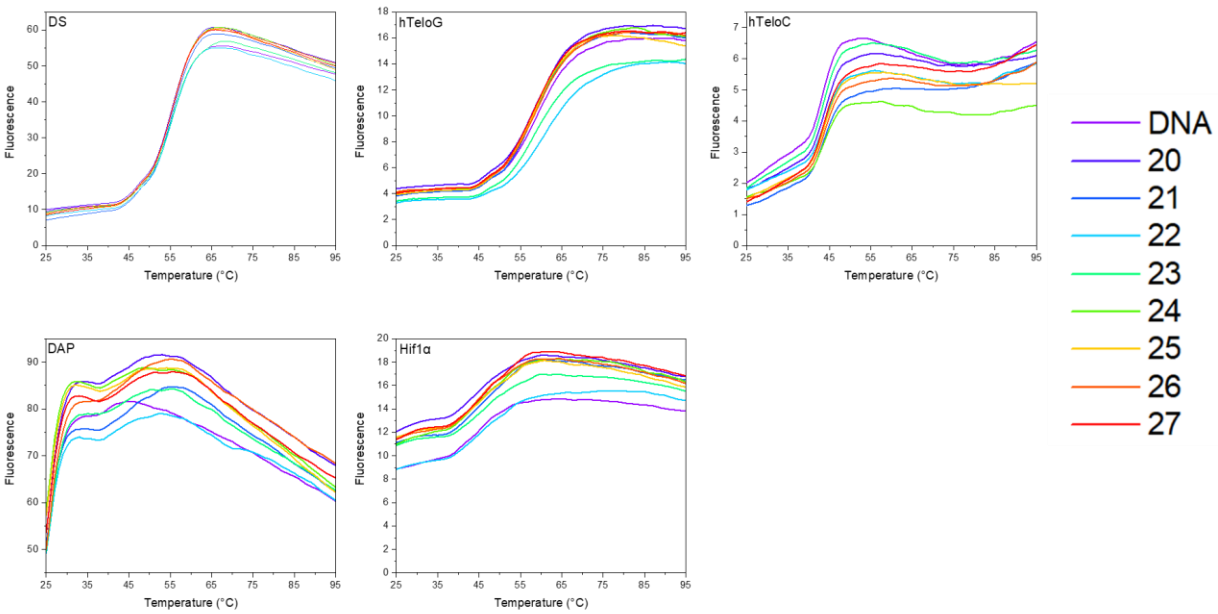

NSC 260594
0 µM
10 µM
20 µM
30 µM
40 µM
50 µM
Ellipticity (mdeg)
Wavelength (nm)
Normalised ellipticity at 288 nm
Temperature (°C)
hTeloC control
NSC 260594
NSC 116339
hTeloC control
NSC 116339
NSC 121868
hTeloC control
NSC 121868


NSC 122819
0 µM
10 µM
20 µM
30 µM
40 µM
50 µM
Ellipticity (mdeg)
Wavelength (nm)
Normalised ellipticity at 288 nm
Temperature (°C)
hTeloC control
NSC 122819
NSC 127133
NSC 116339
NSC 345845

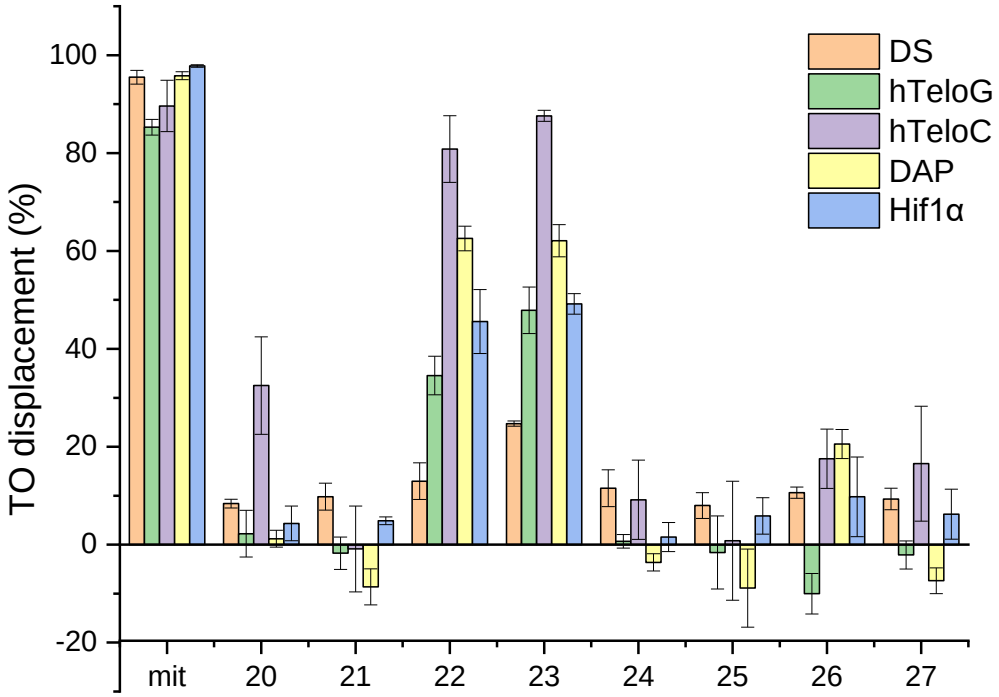

NSC 319990
0 µM
10 µM
20 µM
30 µM
40 µM
50 µM
Ellipticity (mdeg)
Wavelength (nm)
Normalised ellipticity at 288 nm
Temperature (°C)
hTeloC control
NSC 319990
NSC 84100
hTeloC control
NSC 84100


NSC 103520
0 µM
10 µM
20 µM
30 µM
40 µM
50 µM
Ellipticity (mdeg)
Wavlength (nm)
Normalised ellipticity at 288 nm
hTeloC
NSC 103520
Temperature (°C)
NSC 202386
0 µM
10 µM
20 µM
30 µM
40 µM
50 µM
Ellipticity (mdeg)
Wavelength (nm)
Normalised ellipticity at 288 nm
hTeloC control
NSC 202386
Temperature (°C)

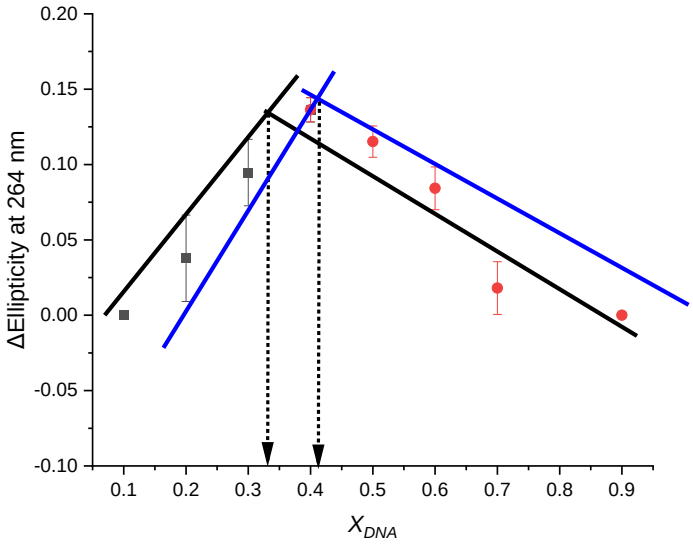
Appendix 8 *FRET melting curves of bis(pyridyl)allenes complexes, all experiments were performed in 10 mM sodium cacodylate and 100 mM potassium chloride buffer, [DNA] was 0.2 μM, [ligand] was 10 μM. hTeloC, pH 5.5; DAP, pH 6.8; Hif1a, pH 6.8; hTeloG, pH 7.0; DS, pH 7.0. DNAs were examined with complexes 20, 21, 22, 23, 24, 25, 26, and 27.*



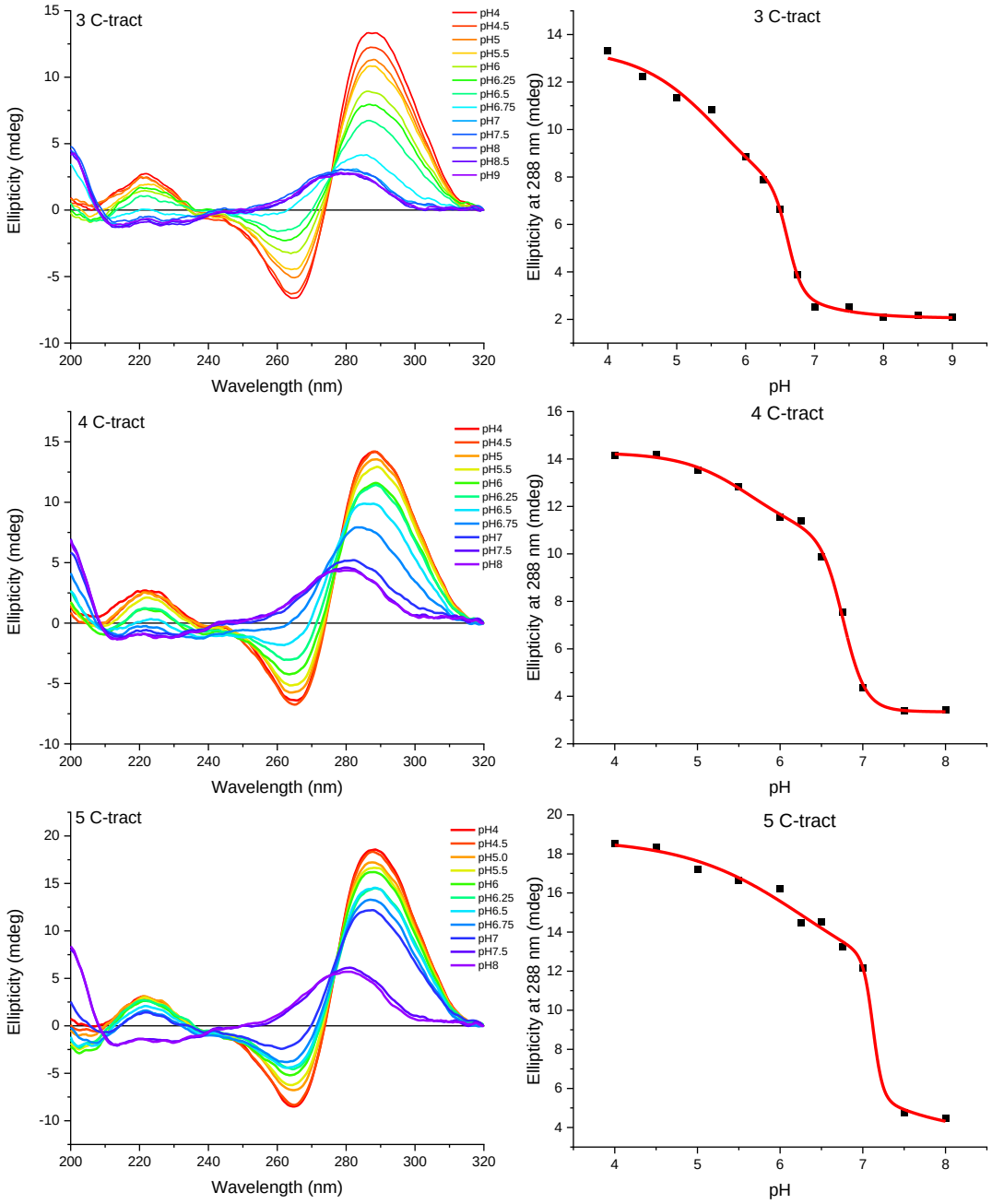
Appendix 9 *FID TO displacement assay for allene ligands 1a-b and allene-containing complexes at 2.5 μM in 10 mM sodium cacodylate and 100 mM KCl: 0.5 μM hTeloC, pH 5.5; 0.5 μM DAP, pH 6.8; 0.5 μM Hif1a, pH 6.8; 0.5 μM hTeloG, pH 7.0; 0.5 μM DS, pH 7.0.*

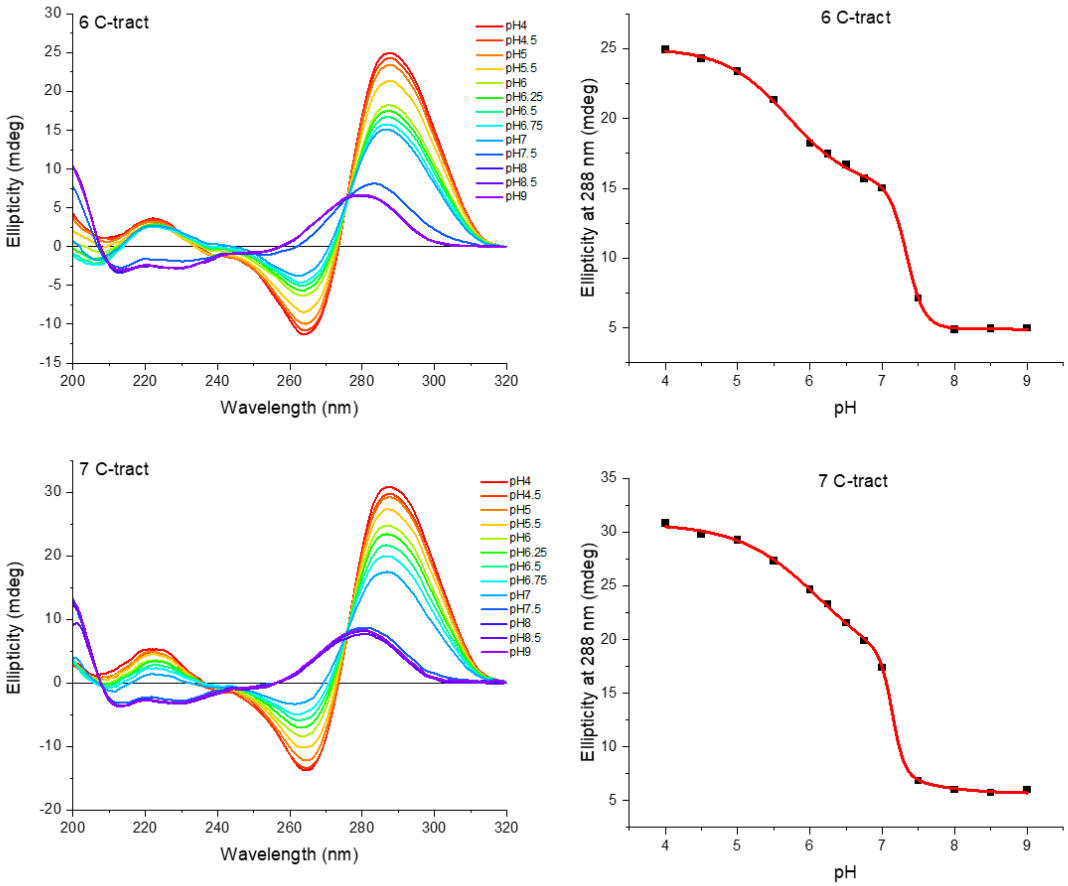


Appendix 10 *Calculation of error of Job Plot of TMPyP4 with PQS18-1 RNA G-quadruplex*

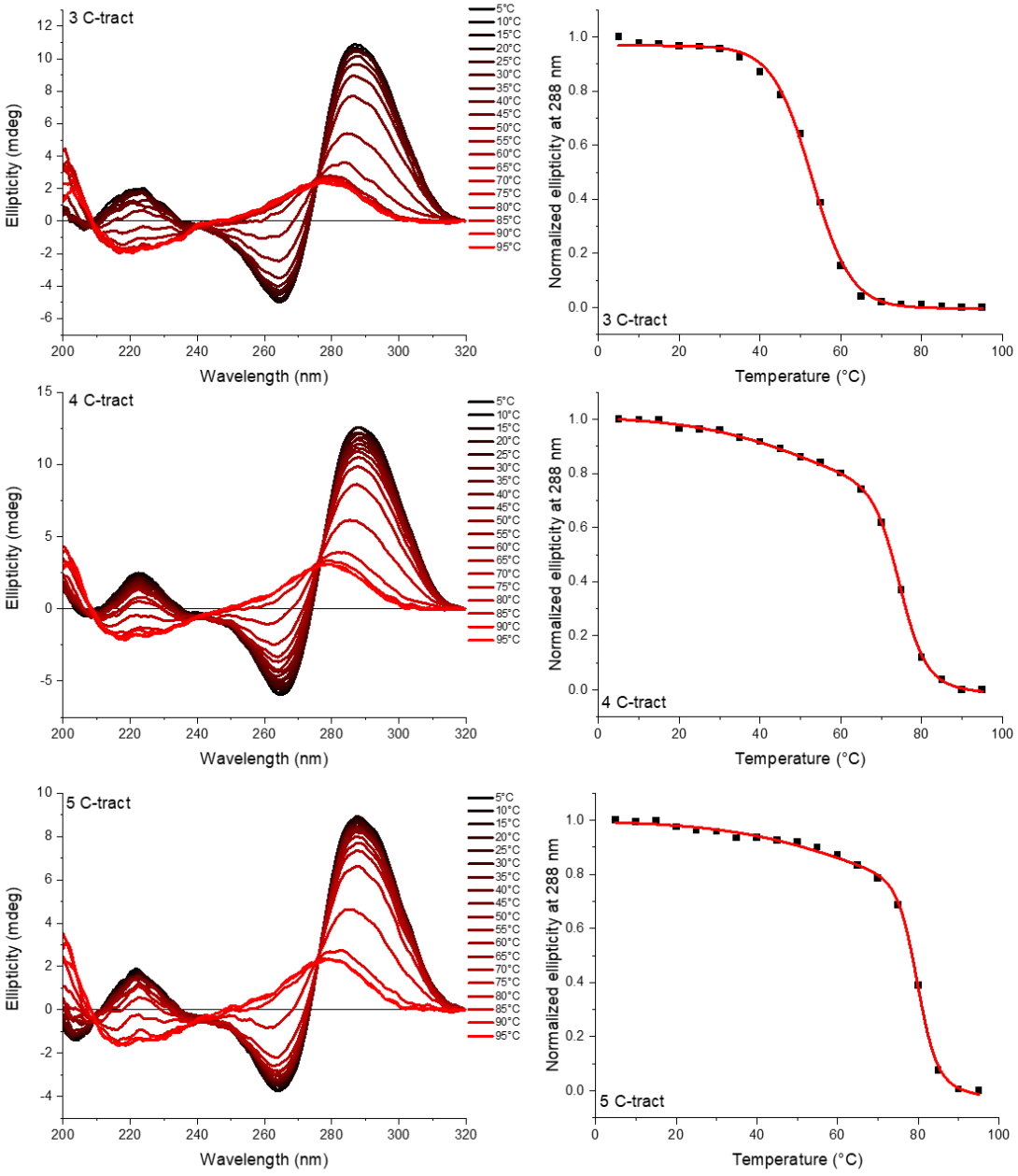


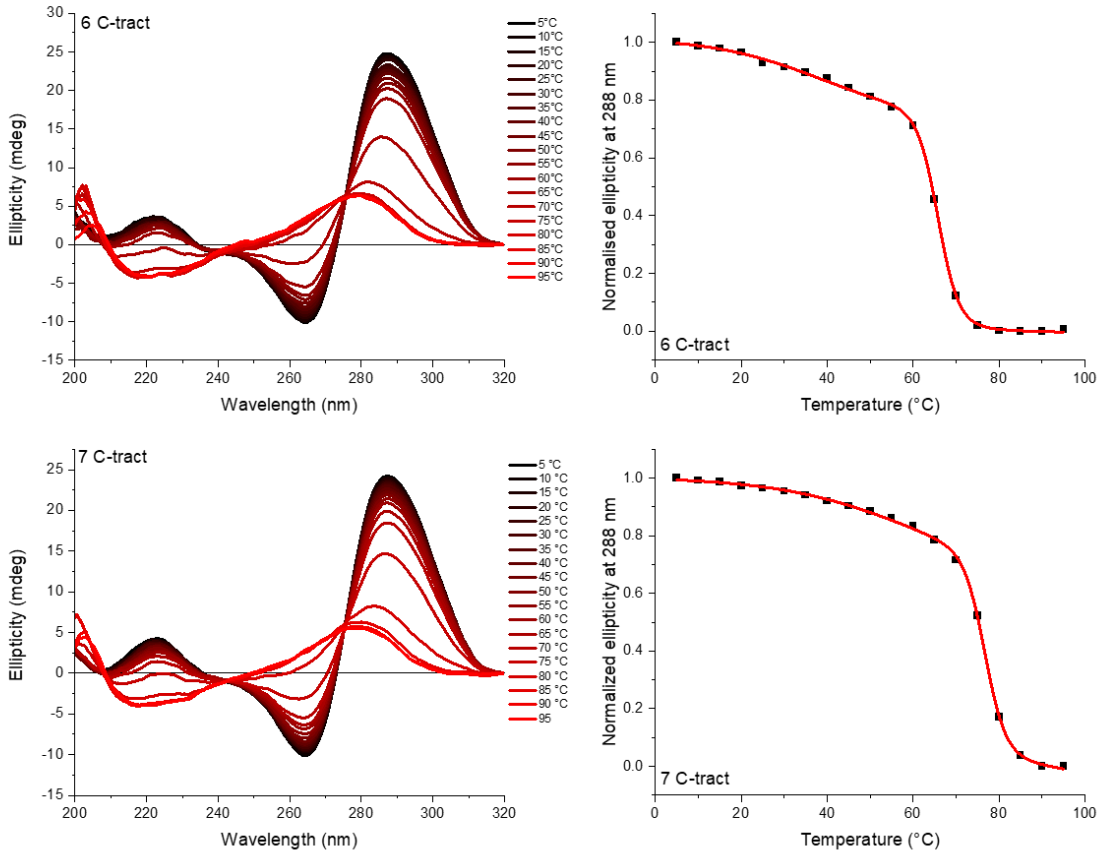
Appendix 11 *CD spectra of C-rich sequences with number of C-tract 3–7 at different pHs, and corresponding bisigmoidal curve fitting. [DNA] was 10 μM, and the buffer was containing 10 mM sodium cacodylate and 100 mM sodium chloride.*



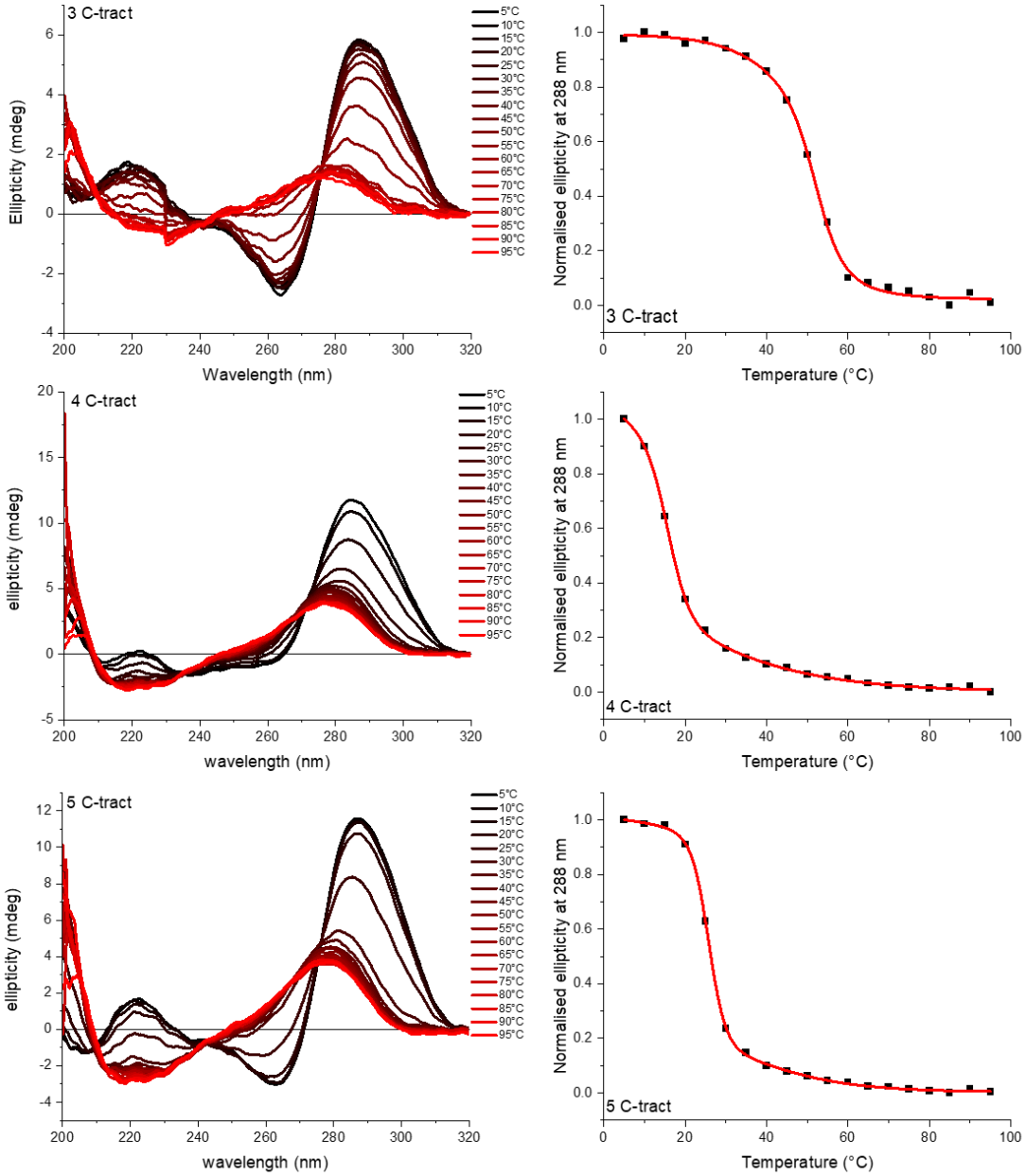

6 C-tract
pH4
pH4.5
pH5
pH5.5
pH6
pH6.25
pH6.5
pH6.75
pH7
pH7.5
pH8
pH8.5
pH9
Ellipticity (mdeg)
Wavelength (nm)
6 C-tract
Ellipticity at 288 nm (mdeg)
pH
7 C-tract
Ellipticity (mdeg)
Wavelength (nm)
7 C-tract
Ellipticity at 288 nm (mdeg)
pH

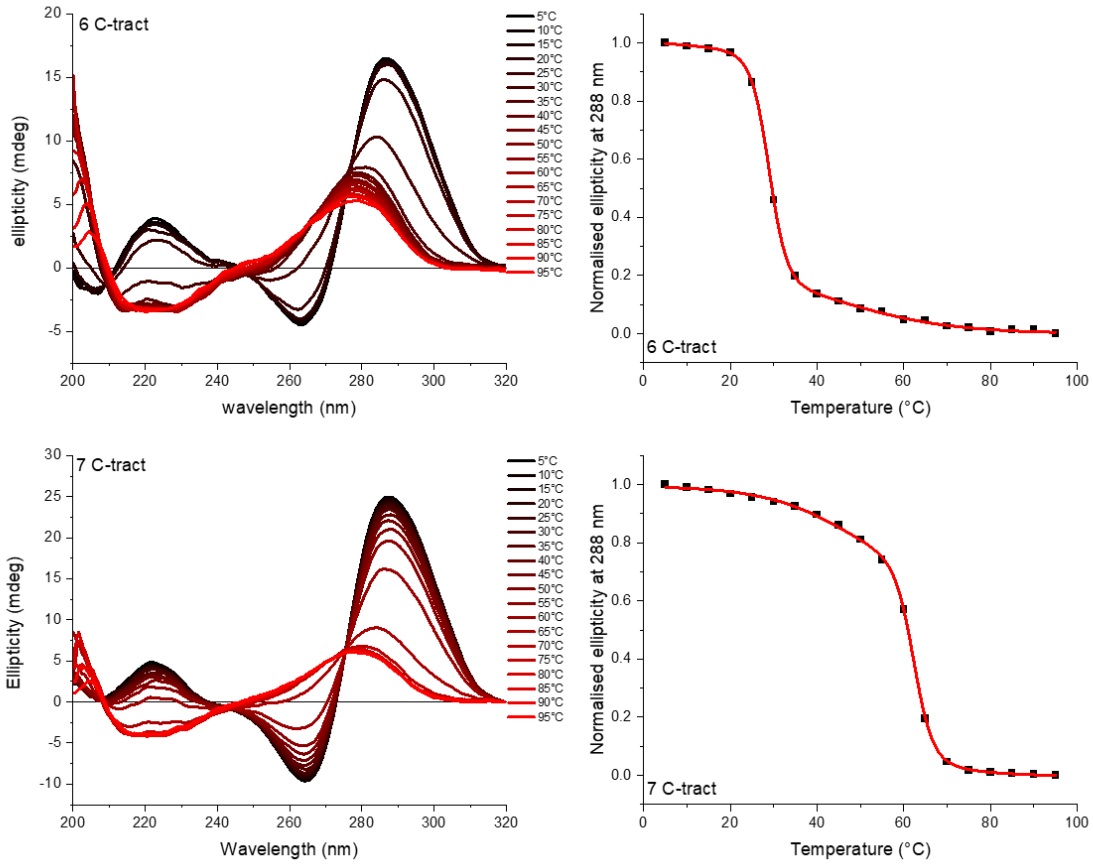
Appendix 12 *CD melting of C-tract 3 to 7 sequences (10 μM) measured in 10 mM sodium cacodylate with 100 mM sodium chloride at pH 5.5.*



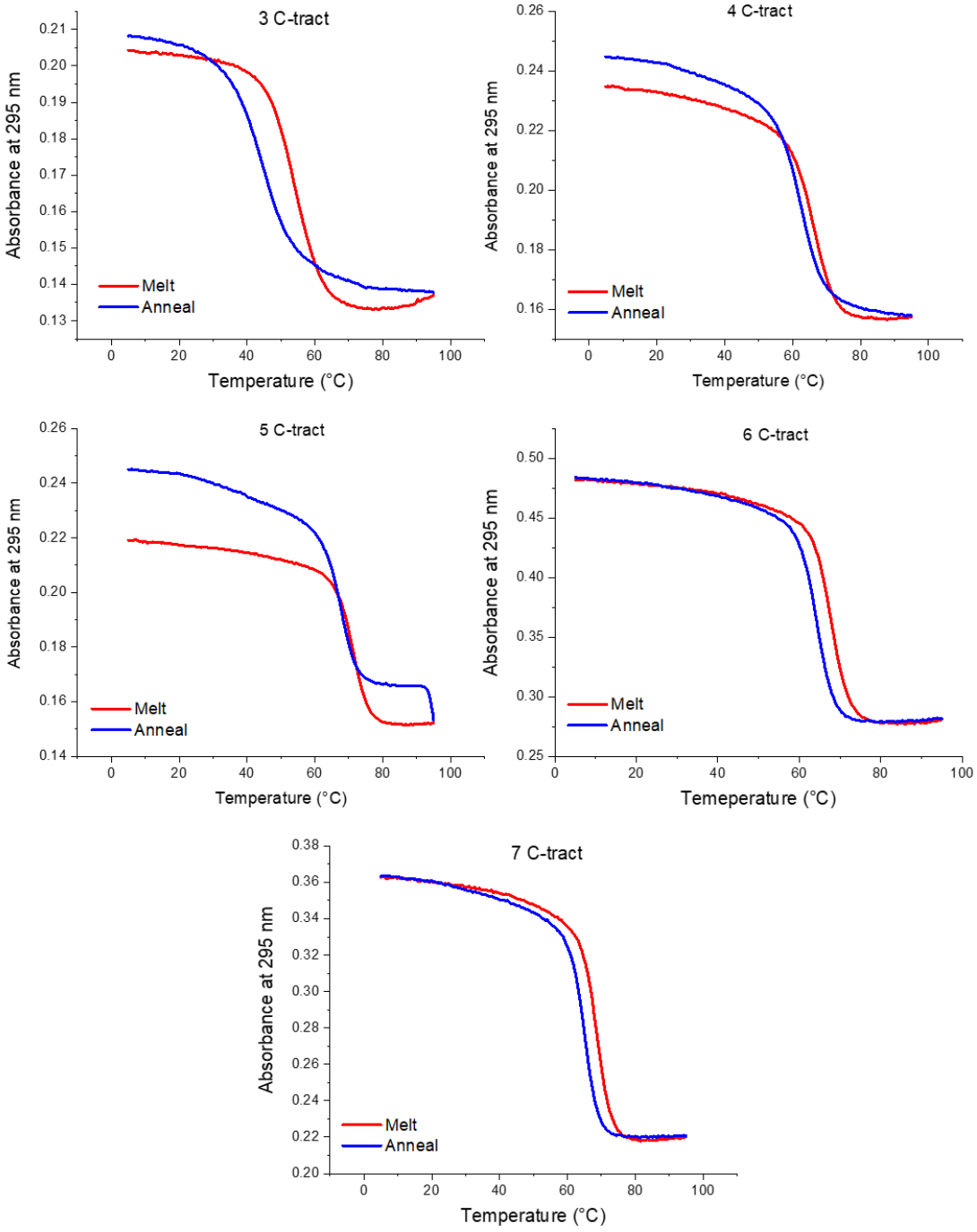

6 C-tract
Ellipticity (mdeg)
Wavelength (nm)
5°C
10°C
15°C
20°C
25°C
30°C
35°C
40°C
45°C
50°C
55°C
60°C
65°C
70°C
75°C
80°C
85°C
90°C
95°C
Normalised ellipticity at 288 nm
Temperature (°C)
6 C-tract
7 C-tract
Ellipticity (mdeg)
Wavelength (nm)
5 °C
10 °C
15 °C
20 °C
25 °C
30 °C
35 °C
40 °C
45 °C
50 °C
55 °C
60 °C
65 °C
70 °C
75 °C
80 °C
85 °C
90 °C
95
Normalized ellipticity at 288 nm
Temperature (°C)
7 C-tract

Appendix 13 *CD melting of C-tract 3 to 7 sequences (10 μM) measured in 10 mM sodium cacodylate with 100 mM sodium chloride at pH 7.0.*

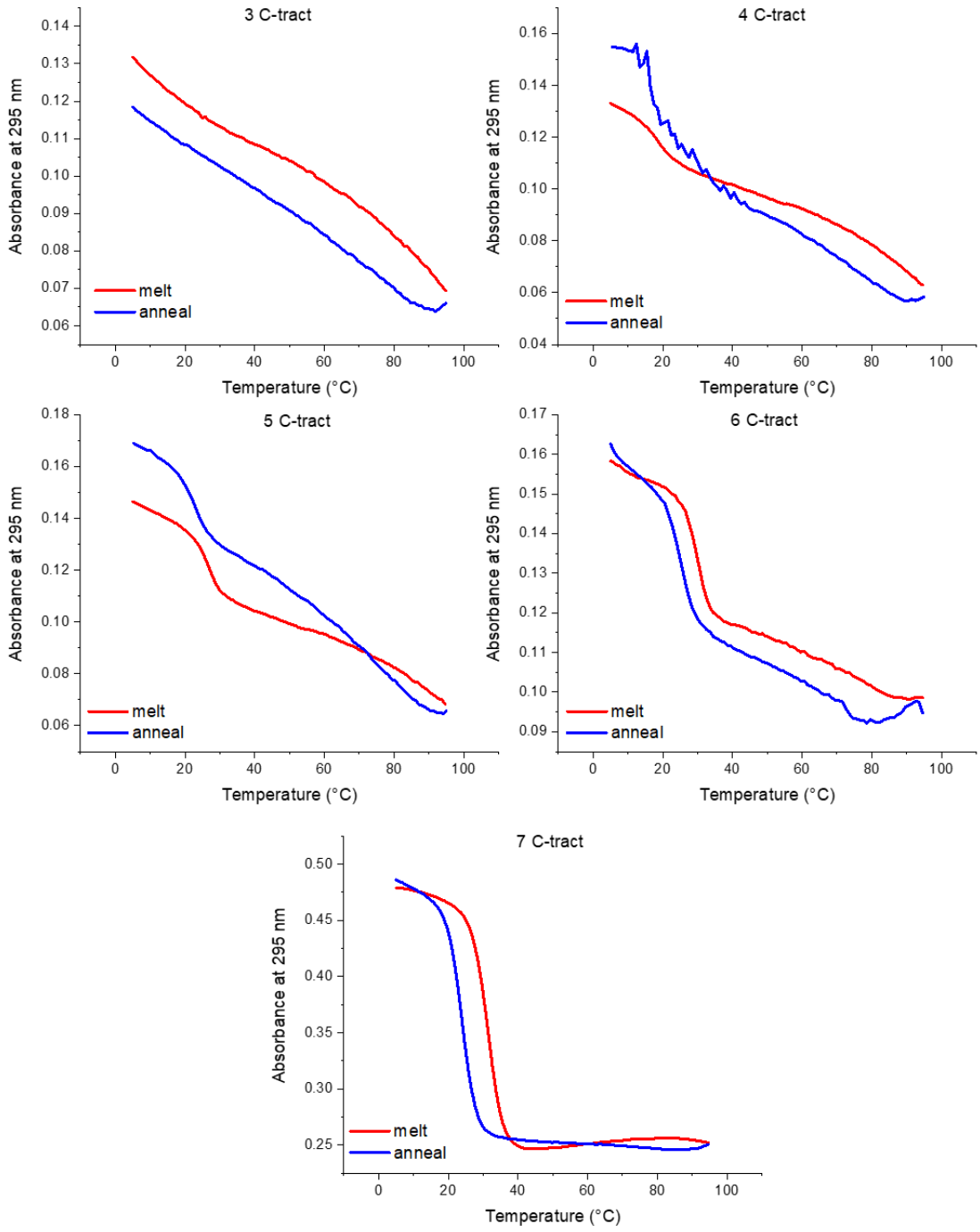



6 C-tract
ellipticity (mdeg)
wavelength (nm)
5°C
10°C
15°C
20°C
25°C
30°C
35°C
40°C
45°C
50°C
55°C
60°C
65°C
70°C
75°C
80°C
85°C
90°C
95°C
Normalised ellipticity at 288 nm
Temperature (°C)
7 C-tract
Ellipticity (mdeg)
Wavelength (nm)

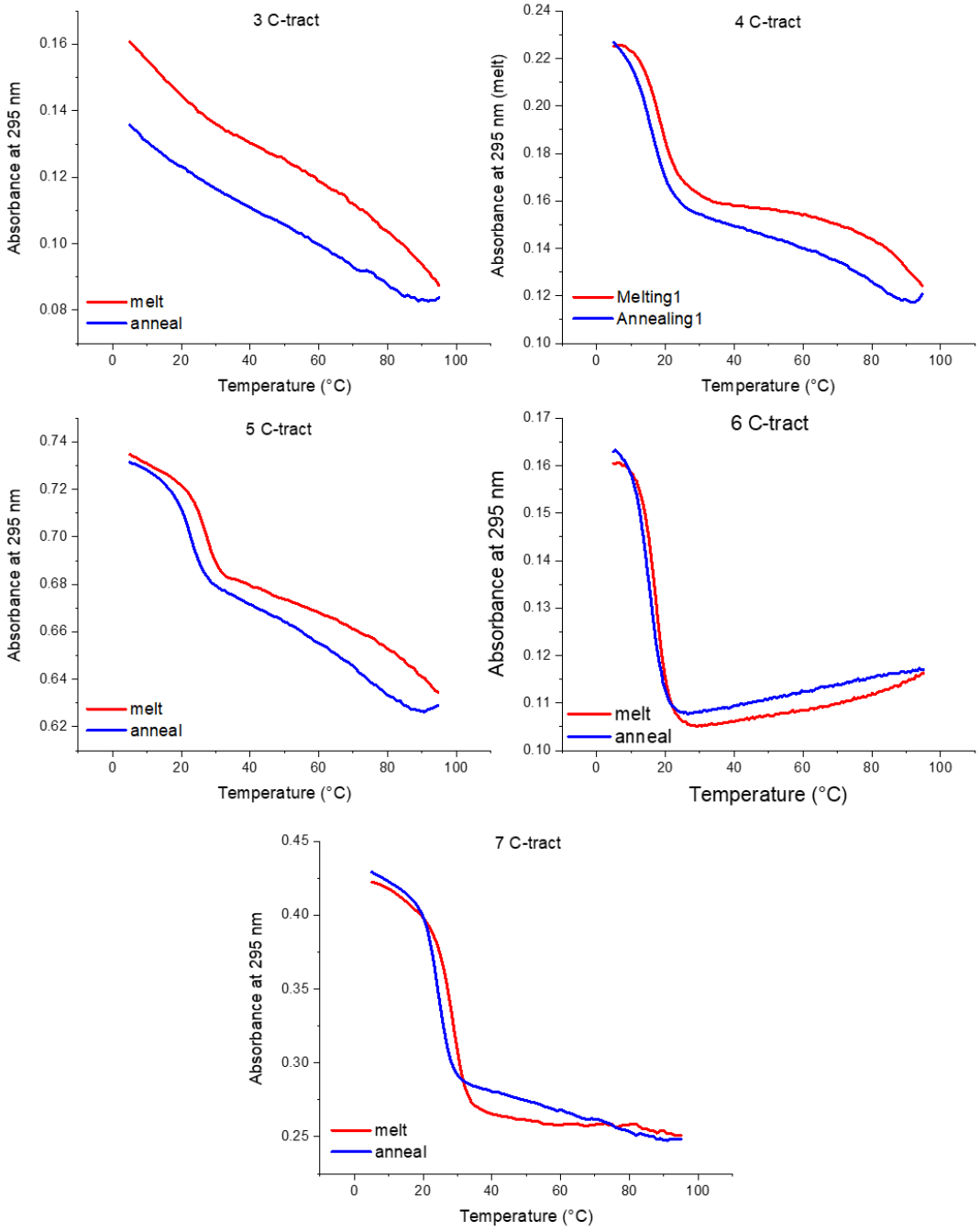
Appendix 14 *Example of UV melting and annealing process of each GC-repeat i-motif measured in 10 mM sodium cacodylate and 100 mM sodium chloride buffer at pH 5.5, the concentration of DNA is 2.5 μM.*



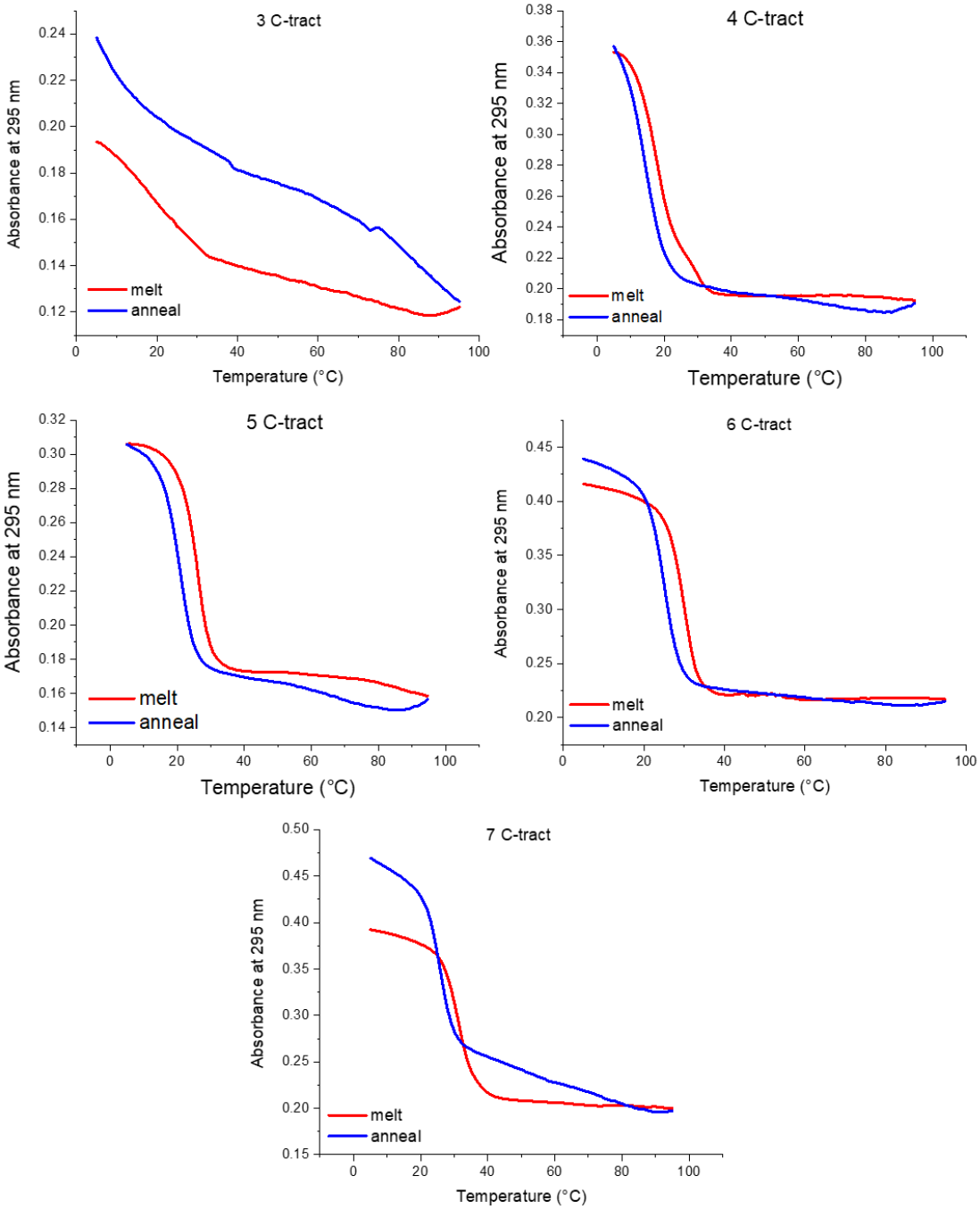
Appendix 15 *Example of UV melting and annealing process of each C-tract of n= 3 to 7 i-motif measured in 10 mM sodium cacodylate and 100 mM sodium chloride buffer at pH 7.0, the concentration of DNA is 0.5 μM.*



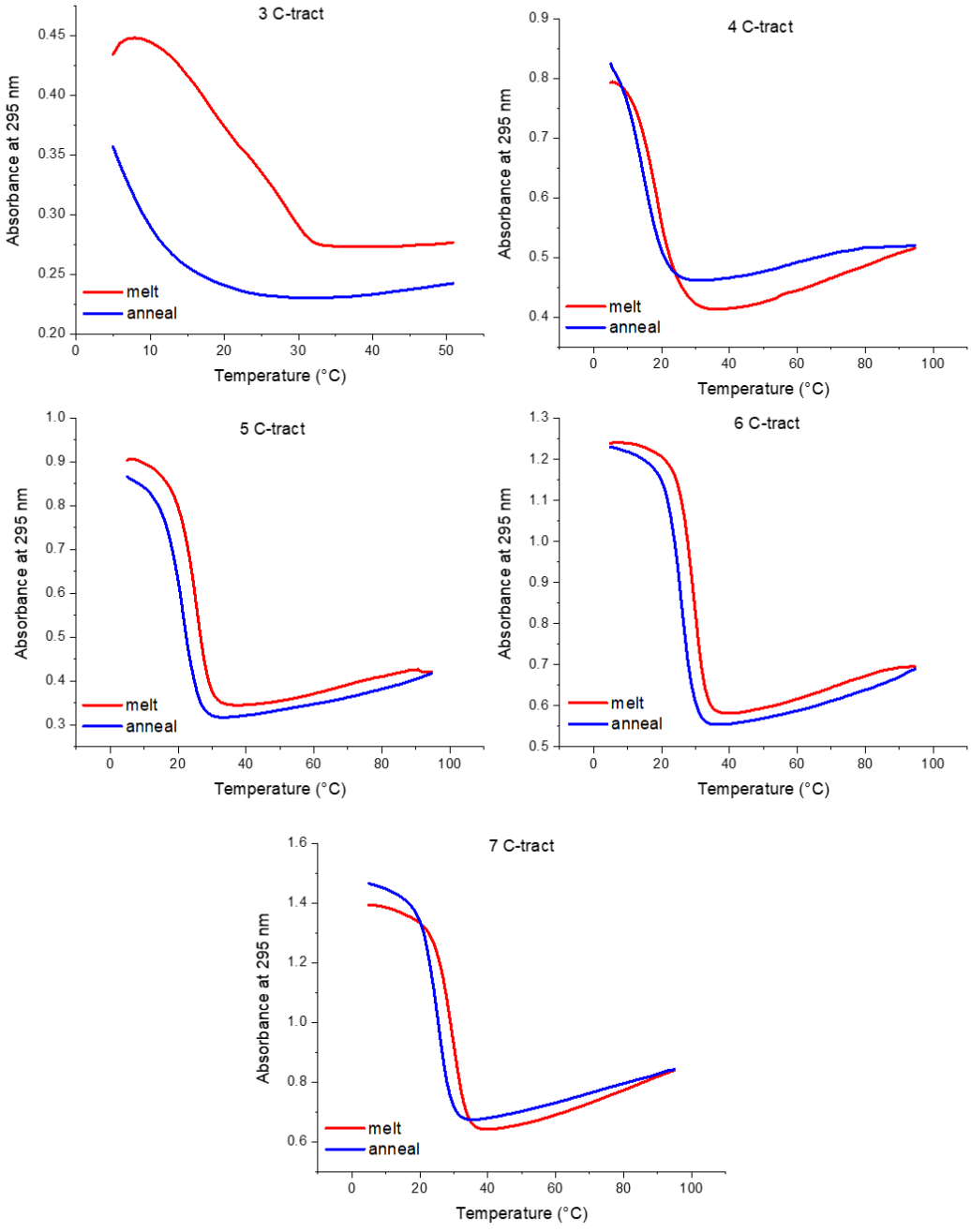
Appendix 16 *Example of UV melting and annealing process of each C-tract of n= 3 to 7 i-motif measured in 10 mM sodium cacodylate and 100 mM sodium chloride buffer at pH 7.0, the concentration of DNA is 1 μM.*



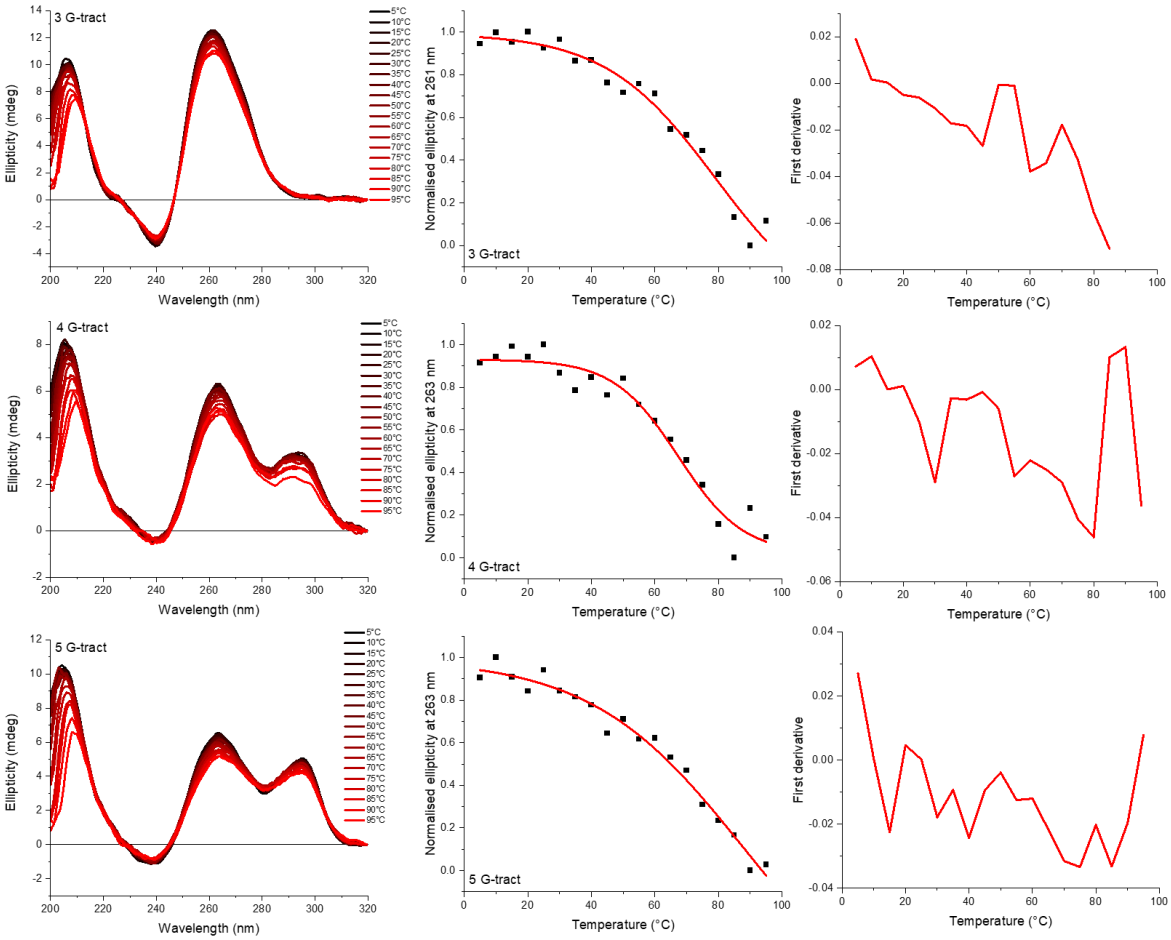
Appendix 17 *Example of UV melting and annealing process of each C-tract of n= 3 to 7 i-motif measured in 10 mM sodium cacodylate and 100 mM sodium chloride buffer at pH 7.0, the concentration of DNA is 2.5 μM.*

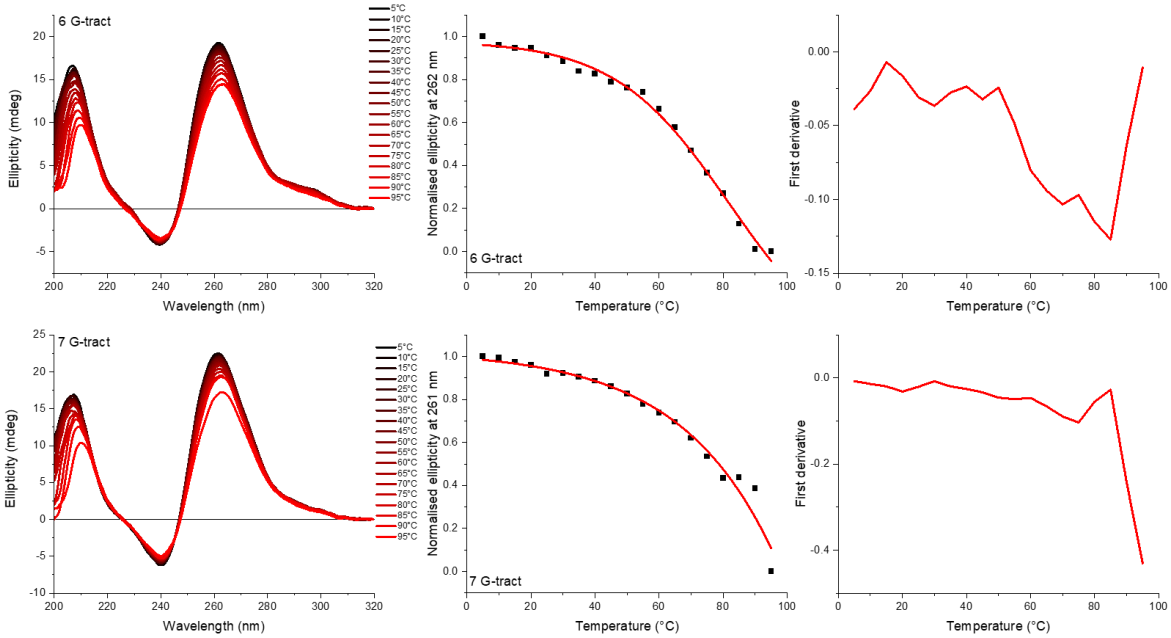


Appendix 18 *Example of UV melting and annealing process of each C-tract of n= 3 to 7 i-motif measured in 10 mM sodium cacodylate and 100 mM sodium chloride buffer at pH 7.0, the concentration of DNA is 10 μM.*

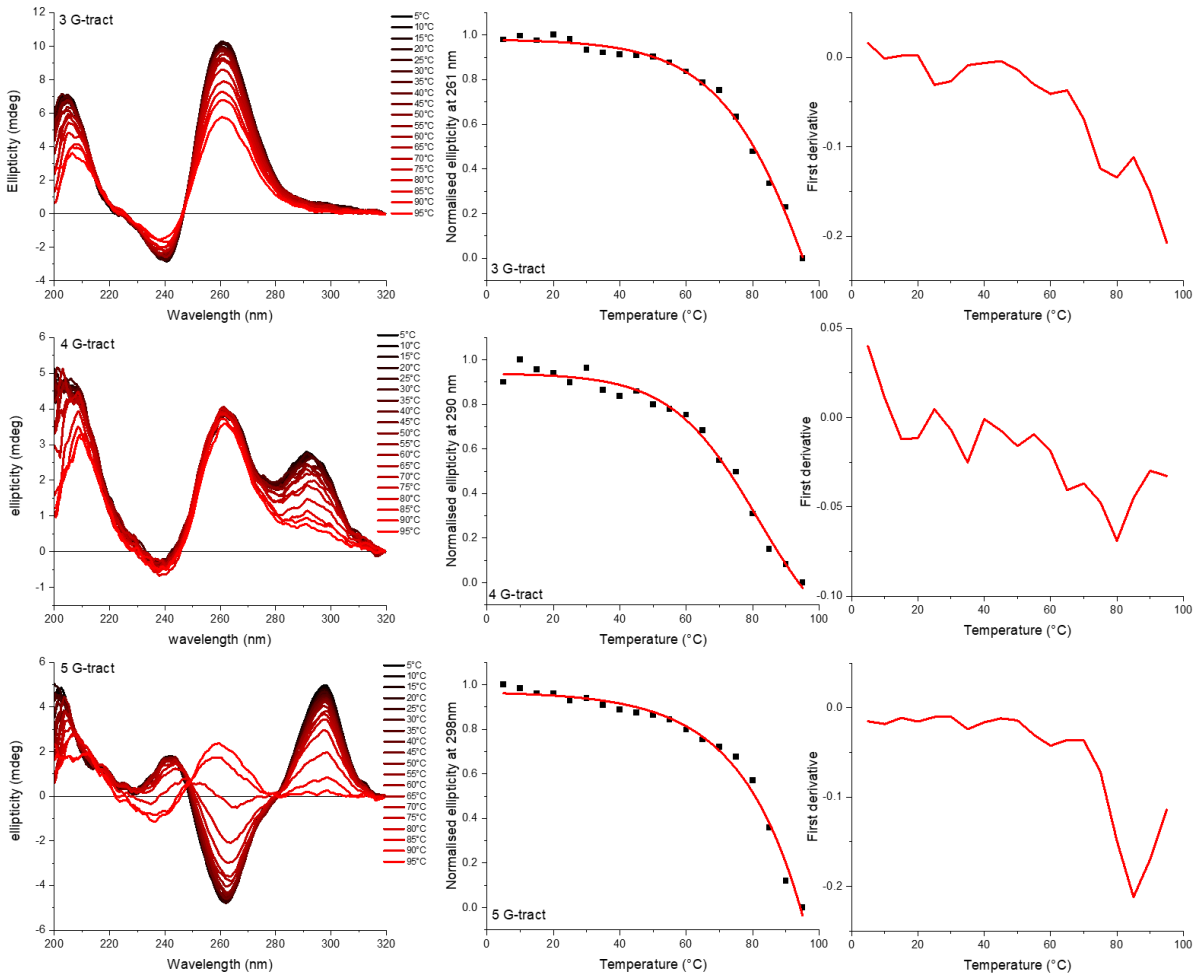


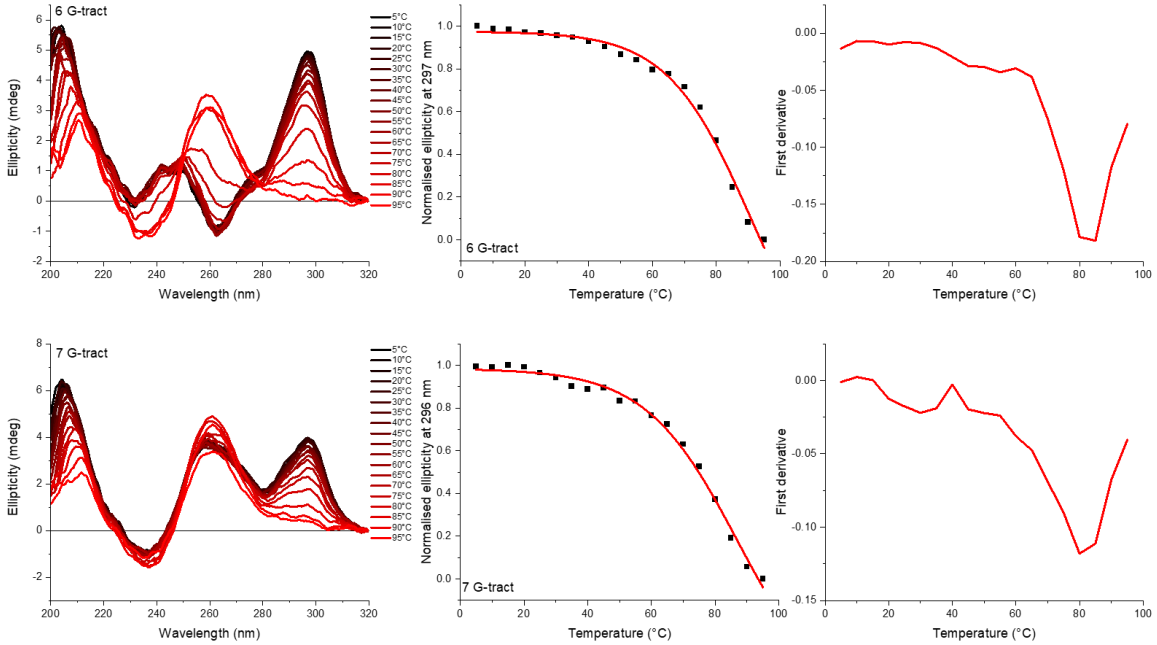
Appendix 19 *CD melting of G-rich sequences with the number of G-tract 3–7, and corresponding sigmoidal curve fitting and first derivative. [DNA] was 10 μM, and all experiments were performed in 10 mM sodium cacodylate and 100 mM KCl at pH 7.0.*



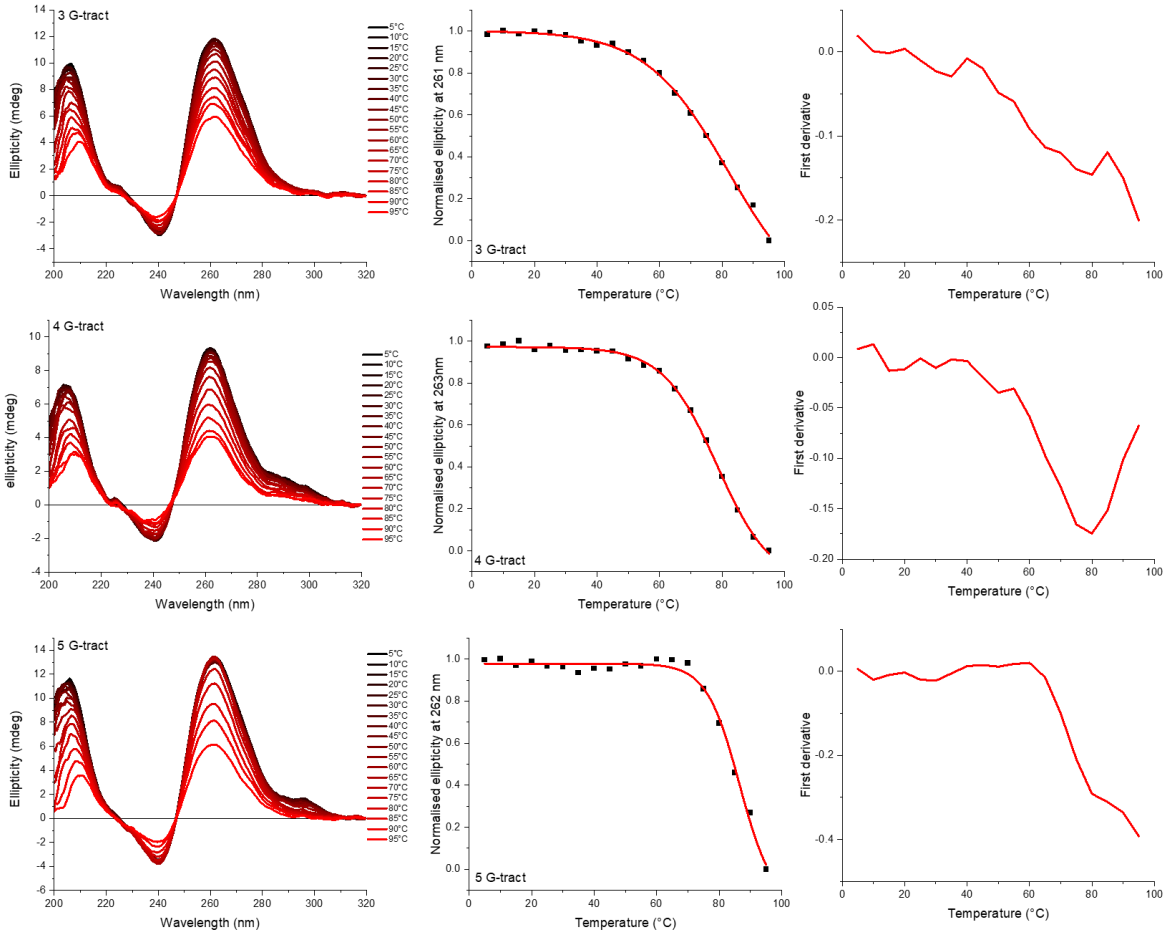

6 G-tract
Ellipticity (mdeg)
Wavelength (nm)
Normalised ellipticity at 262 nm
Temperature (°C)
First derivative
Temperature (°C)
7 G-tract
Ellipticity (mdeg)
Wavelength (nm)
Normalised ellipticity at 261 nm
Temperature (°C)
First derivative
Temperature (°C)

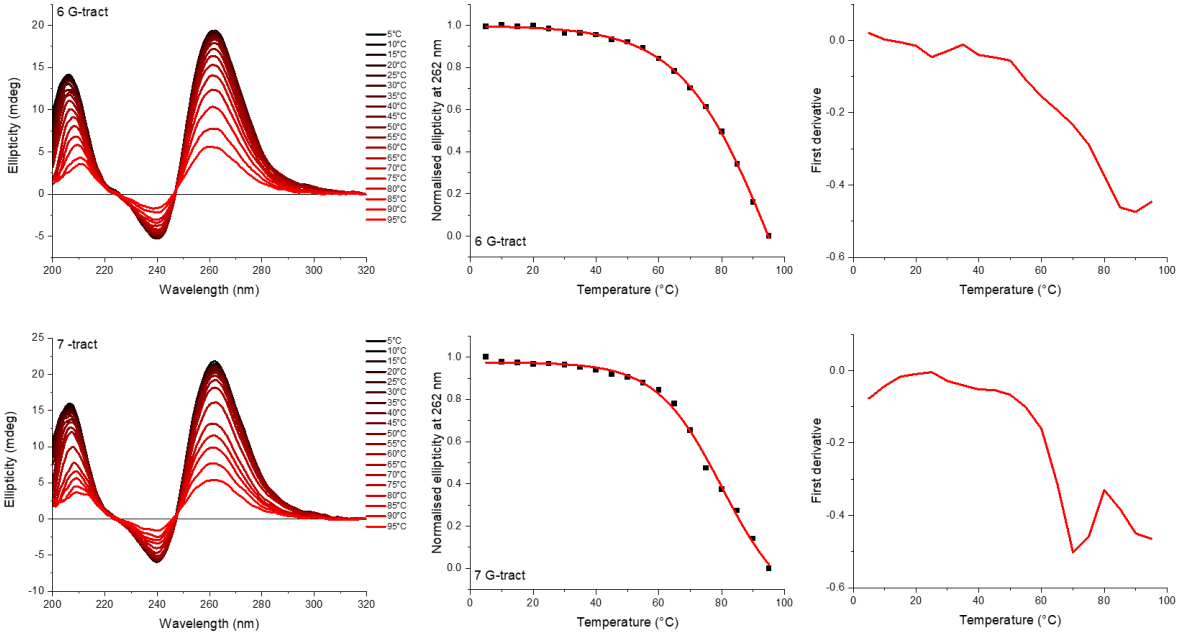
Appendix 20 *CD melting of G-rich sequences with the number of G-tract 3–7, and corresponding sigmoidal curve fitting and first derivative. [DNA] was 10 μM, and all experiments were performed in 10 mM sodium cacodylate and 100 mM NaCl at pH 7.0.*



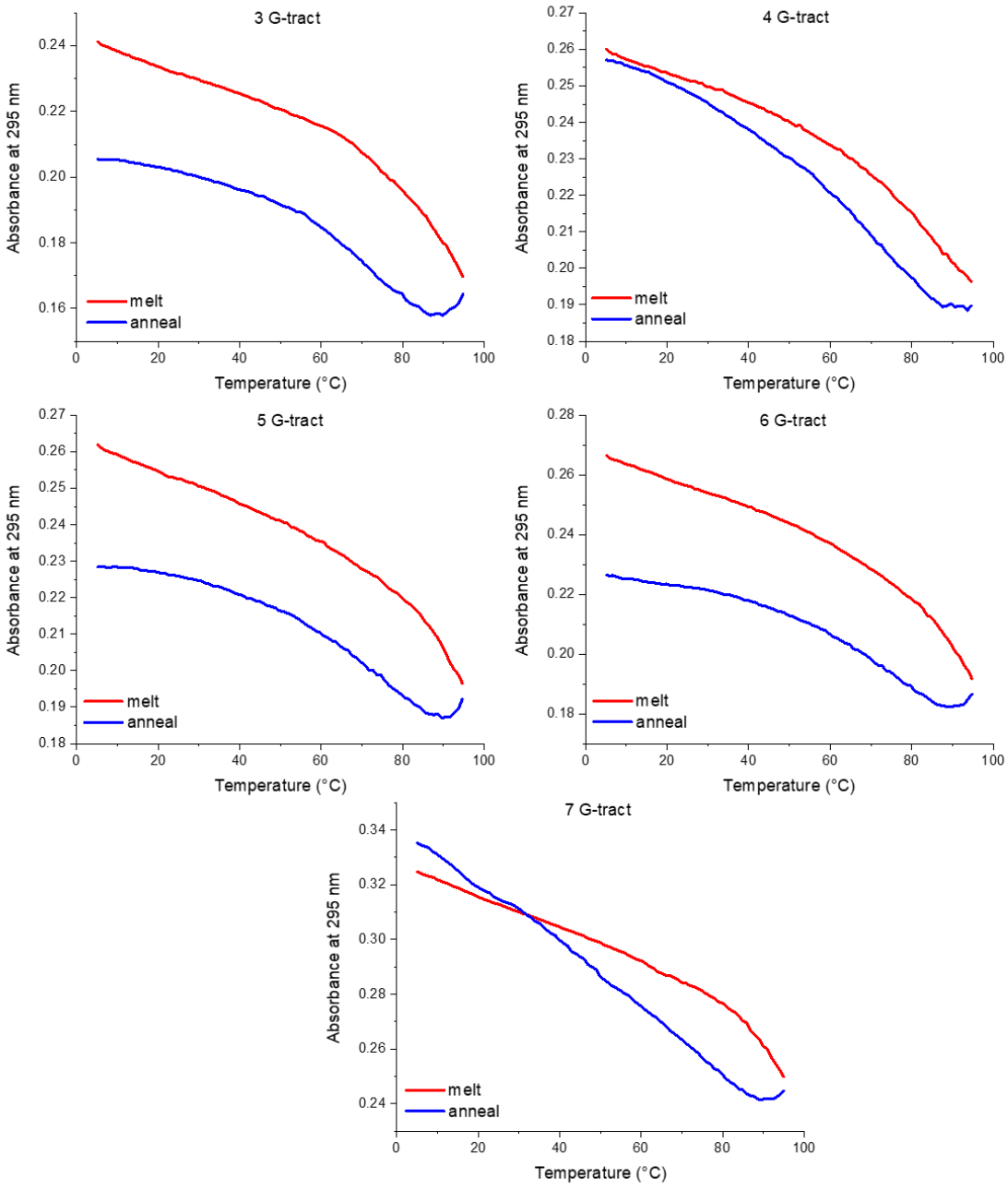

6 G-tract
Ellipticity (mdeg)
Wavelength (nm)
5°C
10°C
15°C
20°C
25°C
30°C
35°C
40°C
45°C
50°C
55°C
60°C
65°C
70°C
75°C
80°C
85°C
90°C
95°C
Normalised ellipticity at 297 nm
6 G-tract
Temperature (°C)
First derivative
Temperature (°C)
7 G-tract
Ellipticity (mdeg)
Wavelength (nm)
Normalised ellipticity at 296 nm
7 G-tract
Temperature (°C)
First derivative
Temperature (°C)

Appendix 21 *CD melting of G-rich sequences with the number of G-tract 3–7, and corresponding sigmoidal curve fitting and first derivative. [DNA] was 10 μM, and all experiments were performed in 10 mM sodium cacodylate and 100 mM LiCl at pH 7.0.*

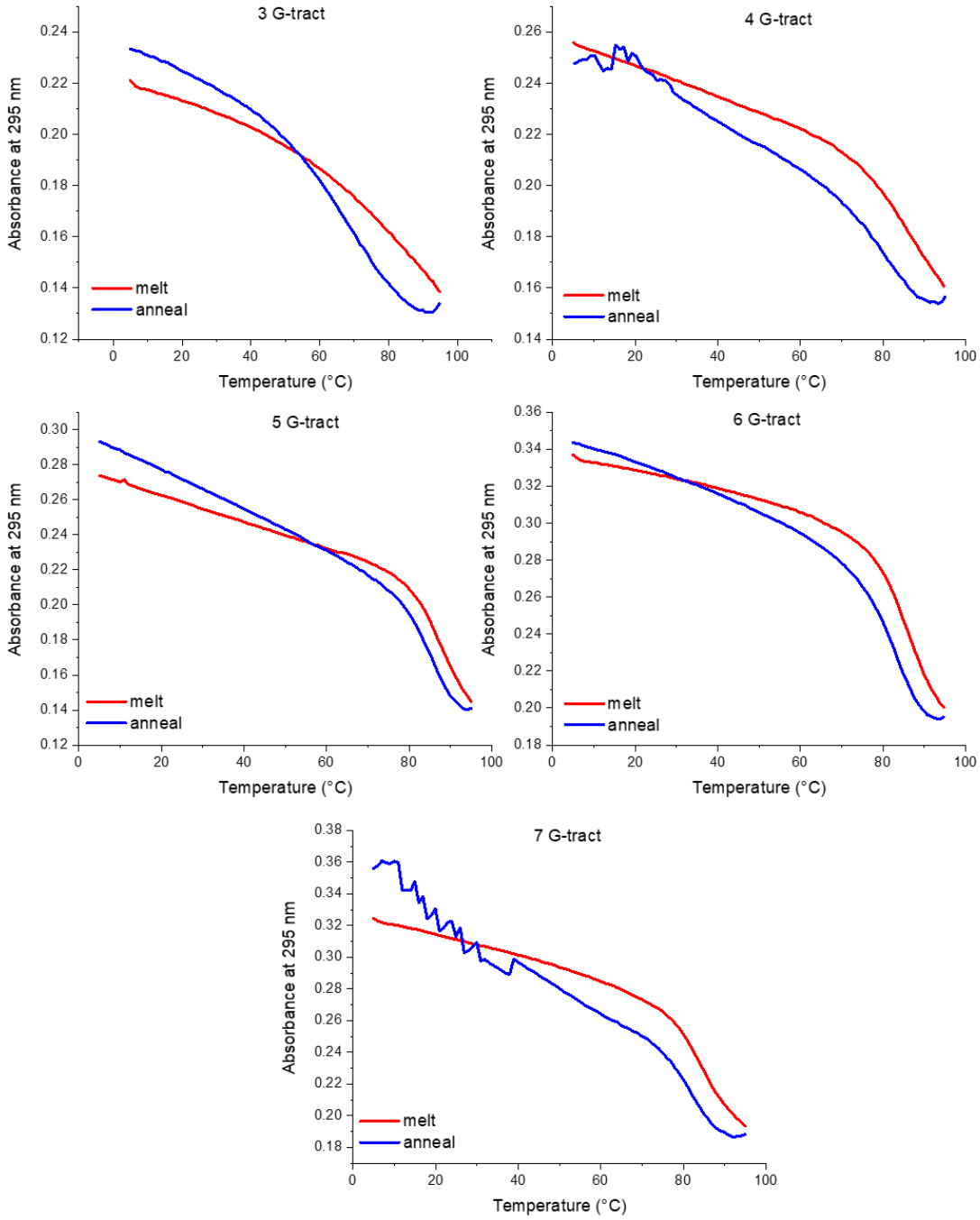



6 G-tract
Ellipticity (mdeg)
Wavelength (nm)
5°C
10°C
15°C
20°C
25°C
30°C
35°C
40°C
45°C
50°C
55°C
60°C
65°C
70°C
75°C
80°C
85°C
90°C
95°C
Normalised ellipticity at 262 nm
6 G-tract
Temperature (°C)
First derivative
Temperature (°C)
7 -tract
Ellipticity (mdeg)
Wavelength (nm)
Normalised ellipticity at 262 nm
7 G-tract
Temperature (°C)
First derivative
Temperature (°C)

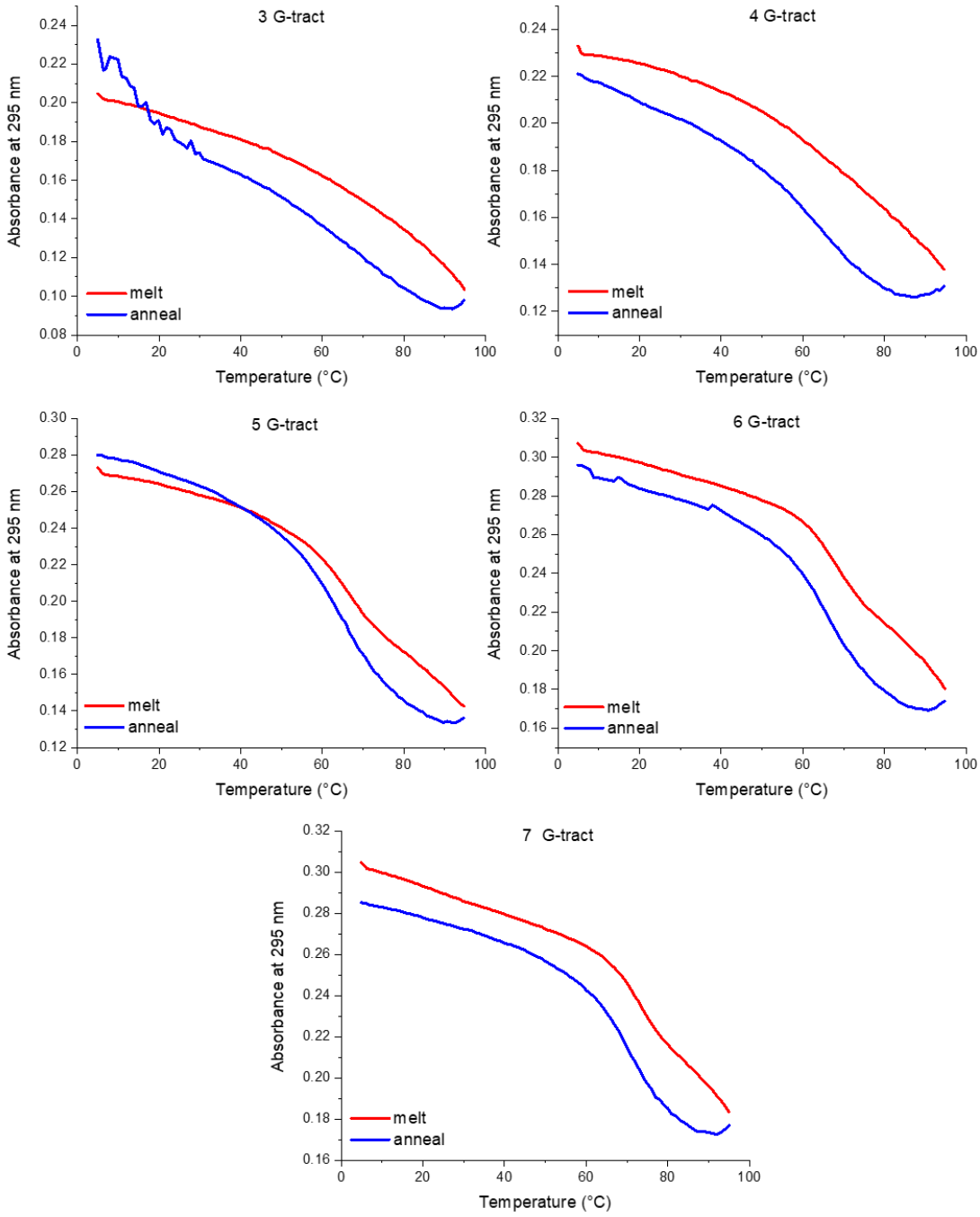
Appendix 22 *Example of UV melting and annealing process of each G-tract of n= 3 to 7 G-quadruplex measured in 10 mM sodium cacodylate and 100 mM KCl buffer at pH 7.0, the concentration of DNA is 2.5 μM.*



Appendix 23 *Example of UV melting and annealing process of each G-tract of n= 3 to 7 G-quadruplex measured in 10 mM sodium cacodylate and 100 mM NaCl buffer at pH 7.0, the concentration of DNA is 2.5 μM.*



Appendix 24 *Example of UV melting and annealing process of each G-tract of n= 3 to 7 G-quadruplex measured in 10 mM sodium cacodylate and 100 mM LiCl buffer at pH 7.0, the concentration of DNA is 2.5 μM.*



Appendix 25 *Alignment identities of 637 i-motif sequences with Sp1 transcription factor. BLAST and NCBI were used as searching tools.*

GeneCode	Chromosome	PrmStart	PrmEnd/TSS	iSeq	Sp1 overlapping identy	Rank
RP11-752G15.4	15	82692359	82693358	CCCCCCGCCCCCACCCCCCCGCCCCC	88	1
TMEM194A	12	57088064	57089063	CCCCCGCCCCCACCCCCGCCCCC	87.5	2
RPL17P46	18	50822577	50823576	CCCCCGCCCCCCGCCCCCCGCCCCCC	87.5	3
PAQR6	1	156248118	156249117	CCCCCTCCCCCCCCCCCCCCCCC	86.36	4
ALOX5	10	45373176	45374175	GGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGG	84.85	5
RSRC1	3	158104855	158105854	CCCCCGCCGCCCCCCCCCCCCGCCCCC	84.62	6
RNVU1-1	1	148362534	148363533	CCCCCCACCGCCCCCCACCCCCACCCCC	84.62	7
COL11A2	6	33192500	33193499	CCCCCACCTCGCCCCCGCCCCCGGCCCGGCCCCC	84.62	8
LOXL4	10	98268251	98269250	GGGGGGCGGGGGGAGCGGGGGGGCGGCGGGGGG	84.38	9
RP11-1079K10.2	17	49374380	49375379	CCCCCCGCCCCCCACCCCCCACCCCC	84	10
JAZF1	7	28180744	28181743	GGGGGGAGGGCGGGGGCGGGGGCGGGGGGGGG	83.87	11
SOX2	3	181710924	181711923	CCCCCTCCTCGCCTGCCCCCTCCTCCCCCGGCCTCCCCC	83.33	12
CAMK2N2	3	184261464	184262463	GGGGGCGGGGGCGGGGGGCGGGGG	83.33	13
ARHGAP4	X	153935000	153935999	GGGGGTGGGGGCGGGGGAGGGGGCGGGGG	82.76	14
NAA10	X	153935224	153936223	GGGGGTGGGGGCGGGGGAGGGGGCGGGGG	82.76	15
RP11-327F22.1	16	50742816	50743815	GGGGGGTGGGGGGCGAATGGGGGGCGGGGG	82.76	16
TRABD	22	50184915	50185914	GGGGGGGGGGGGGGGGGCGGGGG	81.82	17
RP4-647J21.1	7	44960910	44961909	GGGGGTTGAGGGGGCTGGGGGTCGGGGGG	81.82	18
FO538757.3	1	181393	182392	GGGGGCGGGGGGGGGGGGGGGG	81.82	19
RNVU1-18	1	143729571	143730570	CCCCCCCCCCCCCCCCCCCACCCCC	81.48	20
RUNDC3A	17	44307413	44308412	GGGGGCAGGGAGGGGACGGGGGTGGGGGGTGGGGG	81.25	21
RP3-388N13.5	8	93916683	93917682	GGGGGCCGTCGGGGGGCTGGGGGCGGGGGCGGGGG	81.25	22
HOXB9	17	48626357	48627356	CCCCCCCAGCCCCCCACCCTCCCACCCCCACCCCC	81.08	23
RP11-20I20.2	4	1112639	1113638	CCCCCTCCCCCTGCCCCCTCCCCC	80.95	24
SCML1	X	17736449	17737448	CCCCCCCCCCCCCCCCCCCCCCCCC	80.77	25
RP11-1348G14.5	16	28822034	28823033	CCCCCCCCCCCCCCCCCCCCCCCCC	80.77	26
ATXN2L	16	28822035	28823034	CCCCCCCCCCCCCCCCCCCCCCC	80.77	27
THEGL	4	56529609	56530608	CCCCCACACCCCCCGCCCCCCGCCCCC	80.77	28
JRKL-AS1	11	96506827	96507826	CCCCCCCCCCCCCCCCCCCCCCCCC	80.77	29
AL031601.5	10	42241721	42242720	CCCCCCCCCCCCCCCCCCCCCCC	80.77	30
RP11-455F5.3	16	30095430	30096429	GGGGGTGGGGGCGGGCGCCGGGGGAGGGGG	80.65	31
BACH2	6	90296909	90297908	GGGGGCTGGGGCGGGGGGCGGGGGACGGGGG	80.65	32
CTB-50L17.16	19	4471494	4472493	GGGGGGGGGGGGGGGGGCAGCGGGGG	80	33
CCND1	11	69640087	69641086	CCCCCTCCCCCTGCGCCCGCCCCCGCCCCCC	80	34
RP11-452L6.6	16	31472109	31473108	CCCCCGCCCCCCGCCCCCACCGGACACGGCCCCC	80	35
WRNIP1	6	2764414	2765413	CCCCCCCCCCCGCCCCTCCCCCGCCCCC	79.31	36
GDF15	19	18373731	18374730	GGGGGTGGGGGCGGGGGGGAAGAGGGGG	79.17	37
SNAP25	20	10217830	10218829	GGGGGTGGGGGTGGGGGTGGGGGGG	79.17	38
MSMO1	4	165326623	165327622	CCCCCGCCCCCGCCCCCGCCCCC	79.17	39
PRR35	16	559422	560421	GGGGGCGGCGGGGGGGCGCGCGGGCGGCGGGGGCGCGGGGTTCGGGCCGGGGG	78.85	40
AC011380.1	5	140286313	140287312	GGGGGGGGGGGGGGGGGGGGGGGCGGGGG	78.67	41
DDX11L5	9	11134	12133	GGGGGCGGGGGGGGGGGGGGGGGGGGGG	78.57	42
KIF13B	8	29263125	29264124	GGGGGCCGGGGGCGGAGCCGGGGGTGGGGACCGGGCTGGGGG	78.38	43
CCDC148	2	158456754	158457753	GGGGGTGGTGGGAGAGCCGCTCCGGGGGCGGGGGCCGGTGGGGG	78.31	44
GID8	20	62937119	62938118	CCCCCGCCCCCGCCCCCGCCTCCGCCGCGGCCCCC	78.13	45
DIDO1	20	62937953	62938952	CCCCCGCCCCCGCCCCCGCCTCCGCCGCGGCCCCC	78.13	46
U1	1	146375807	146376806	GGGGGTGGGGGGGGGTGGGGGGGGGGGG	77.78	47
RP11-140C5.3	10	49550308	49551307	GGGGGGGGGGGGGGGGGGGGGCGGGGGG	77.78	48
HOXD10	2	176107790	176108789	GGGGGCCGCGGGGGAGGGGGGGGGG	77.78	49
TMEM132C	12	128266403	128267402	GGGGGCGGGGCCCGGGCGGGGGCGGGGGCGGGGCCGGGGG	77.5	50
RP13-582L3.4	4	8860828	8861827	GGGGGCGGGGGCGAGGGCGGGGGAGGGGG	77.42	51
DAP	5	10761273	10762272	GGGGGCGGGGGCGGGGGCGGGGGCGGGGG	77.42	52
NR4A2	2	156342349	156343348	CCCCCCTCCCCCGGCCCCCGGCCCCCGGCCCCC	77.42	53
UGGT1	2	128090200	128091199	GGGGGGTTGGGGGGTGGGGGGCGCGGGCGGGGG	77.14	54
RP11-586D19.1	4	6199733	6200732	GGGGGCTGGGGGCGGGACCGGGAGGGGGTGTTGGGGG	76.67	55
RP4-647J21.1	7	44960910	44961909	GGGGGTTGAGGGGGCTGGGGGCTGGGGGGCTGAGGGCTGAGGGGGG	76.47	56
AARS	16	70289544	70290543	CCCCCCCCCGCCCCCCCCGCCCCCCCCCGCCCCC	76.32	57
GDF11	12	55742280	55743279	CCCCCATCCTGATTTCTCTCCCCCCCACCCCCCCCCCC	76.19	58
ARHGAP22	10	48656266	48657265	GGGGGGGGGGGGGCGGGGGGGGGG	76	59

HOXC10	12	53984065	53985064	CCCCCACCCCCACCCCCACCCCCC	75.86	60
RP11-834C11.12	12	53984845	53985844	CCCCCACCCCCACCCCCACCCCCC	75.86	61
CNN2	19	1025581	1026580	GGGGGTGCTGTGCCCAGGCGGGGGTCCGG GGGCGACCGAGGGGG	75.76	62
MAFG-AS1	17	81926829	81927828	CCCCCTCGGCCCCCGCCGCCCCCGCCGCGC CGCCCCC	75.76	63
CAMSAP2	1	200738558	200739557	GGGGGCGGGGGCGGGGGCGGGGGTTGGGG GGGG	75.76	64
RP11-334C17.3	17	80146250	80147249	CCCCCGGCTCGCCCCCGACTCGCCCCCGGC TGGCCCCC	75.68	65
ABHD6	3	58236506	58237505	GGGGGCGGGGGCGGGGGCGGGGG	75	66
RNVU1-15	1	144411576	144412575	GGGGGCGGGGGGGGTGGGGGGGGGGGG	75	67
STON1-GTF2A1L	2	48528925	48529924	GGGGGGTGGGGGGTGGGGGTGGGGTGAGG CGGGGGG	75	68
C14orf93	14	23010167	23011166	CCCCCACCCCCACCCCCACCCCC	75	69
HYAL1	3	50312382	50313381	GGGGGAGGGGGAGGGGGAGGGGAGGGGG AGGGGGAGGGGG	75	70
CTB-147C22.8	19	50949185	50950184	CCCCCACCCCCACTTCCCCACCCCCACCCCC	75	71
FER1L6-AS2	8	124171523	124172522	GGGGGGGGGGGGGGGGGGGGGGGGGGGG	75	72
RP5-1160K1.6	1	109630306	109631305	GGGGGGTTGGGGGGATGGGGGGGCGGGGG	75	73
ZNF213-AS1	16	3134883	3135882	GGGGGCGGGGGCGGGGGCGGGGG	75	74
VPS36	13	52450629	52451628	CCCCCTCCACCCCAACTCCCCCACGCCACTC CACTCCCCCACCCCC	74.74	75
ALS2CL	3	46693705	46694704	CCCCCACCCTGCCCCCTGCCAGTCCCGCCC CCTGCCAGCCCCGCCCCC	74.51	76
RP11-355B11.2	2	61470188	61471187	GGGGGGGAAGGACGGGGGGAGGGGAGAGG GGGGGAGGGGG	74.42	77
USP34	2	61470770	61471769	GGGGGGGAAGGACGGGGGGAGGGGAGAGG GGGGGAGGGGG	74.42	78
MICU3	8	17026238	17027237	CCCCCCCCCCCGCCCCCGCCCCTCCGTTCT CTGCCCCC	74.36	79
MAGEA1	X	153178285	153179284	CCCCCACCCCCACCCCCACGCCCACTCCCA CCCCC	74.36	80
ACTN1-AS1	14	68978682	68979681	GGGGGCTGGGGGCTGGGGGCTGCAGCGGG CGGGGG	74.29	81
RP11-24M17.4	15	75736820	75737819	GGGGGCGGGGGCGGGGGCGGGGTGGGGG	74.19	82
DES	2	219417377	219418376	GGGGGATGAATCAGGGAGGGGGCGGGGGA CCCAGGGGG	74.19	83
USH1G	17	74923257	74924256	GGGGGCGGGGGCGGAGGGGGCGGGGG	74.19	84
OTOP2	17	74923275	74924274	GGGGGCGGGGGCGGAGGGGGCGGGGG	74.19	85
PPP1R15B	1	204411792	204412791	GGGGGGGGGGGGGGGGGGGGGGGGGGG	74.07	86
CRIM1	2	36354926	36355925	GGGGGAGGGGGAGGCGGAGGGGGCCGGCC GGGCCGGGGAGGGGGGG	74	87
RP11-490M8.1	2	36355115	36356114	GGGGGAGGGGGAGGCGGAGGGGGCCGGCC GGGCCGGGGAGGGGGGG	74	88
RP11-5N11.4	5	31742328	31743327	CCCCCTCCCCCAACTCCCCCCCTCCCCACCC CATTCCCCC	73.81	89
AP000936.4	11	117097987	117098986	GGGGGAGGGGGCAAGGGGGGGCGGGAAAG GGGGGG	73.68	90
UNC5C	4	95549207	95550206	GGGGGTGGGGGTGGGGGTGGGGGTTGAAG TGGGGGG	73.68	91
SIK3	11	117098438	117099437	GGGGGAGGGGGCAAGGGGGGGCGGGAAAG GGGGGG	73.68	92
STX17	9	99905633	99906632	GGGGGCCCTGCGGGGCGGGGGCGGGGGCG GGGG	73.53	93
STX17-AS1	9	99906602	99907601	GGGGGCCCTGCGGGGCGGGGGCGGGGGCG GGGG	73.53	94
RP13-516M14.4	17	82292815	82293814	GGGGGGAGGGAGGGGACGGGGGGAGGGA GGGGAGGGTGGGGGAAGGAGGGGAGGGGA GGGTGGGGG	73.44	95
SH3BP4	2	234950973	234951972	CCCCCCGCGCCCCCGCCCCCGCCCCCGCCC CCGCCCCC	73.33	96
MILR1	17	64448037	64449036	GGGGGTGTGGGGGGGTGGGGGGGGCGGGG G	73.33	97
NFATC1	18	79394856	79395855	CCCCCGTTTCCCCCGCCAGCCCCAGCGCCC CCCTGCCCGGCCCCC	73.33	98
TMEM216	11	61390687	61391686	GGGGGGGGGGGGGGGGGGGGGGGGGGGG GG	73.33	99
SSR1	6	7347447	7348446	GGGGGGGGGGGTTGGGGGGGCGGTGGGG G	73.33	100
DUX4L18	Y	11320557	11321556	CCCCCAAAACGCGCCCCCTTCAGCACTTCC CCCTCGCCCCC	72.73	101
FLJ44511	7	519391	520390	CCCCCGCCCCCGCCCCCGCCCCCCCCCCCC GCCTCCCCC	72.73	102
TRIM8	10	102643496	102644495	GGGGGCCGGGGGGGCGAAGGGGGAGGGCC GGGGG	72.73	103
MIR3197	21	41166557	41167556	GGGGGCGAAGGGGGCGGGGGTGGGGGG	72.41	104
BACE2	21	41166801	41167800	GGGGGCGAAGGGGGCGGGGGTGGGGGG	72.41	105
BHLHE40	3	4978116	4979115	CCCCCTCCCCCAGCACCCCCGTCCCTTCCA GCTTCCGCGCCCCCC	72.34	106
ADGRA1-AS1	10	133088492	133089491	GGGGGTCCCGGGGGTCCTGCAGCTGGGGG CTAGGCCGCTGCGGGGGGG	72.09	107
SIM2	21	36698133	36699132	CCCCCGCCCCCGCCGCGGTCCCTCCCCCCT CCCGCTCCCCCC	71.74	108
ITGA9-AS1	3	37861781	37862780	CCCCCCCGCGCCGCGCCCCCGCGCCCCCCG CGCCGCGCCCCC	71.74	109
COX6B2	19	55354815	55355814	CCCCCCCGCCCCCCTCGTCCTCCCCCCTCC GTTACCCGCCCCCCCCACCCTCCTCCCTCGC CCCCCTCCCCCCGCCCCC	71.59	110
ADCY7	16	50245137	50246136	CCCCCCAGCCCCCAGCCCCCCGCCCCGGCT CCCGCCCCC	71.43	111
RP1-172N19.5	X	72776608	72777607	CCCCCGCCGCCCCCCGCCCCCGCAGTACCC CTCCGGCCCCC	71.43	112

SNORD112	14	100896920	100897919	GGGGGGGGGGGTGGGGGGGTGGGGGGCG GGGGGGGG	71.43	113
ENTPD8	9	137441817	137442816	GGGGGTGTTGGGGGGCGGGGAATAGGGTG GGGGAATGCGGGGGGG	71.43	114
RP5-942I16.1	7	69597449	69598448	CCCCCTCCACGCCCCCGCCCCCGCCCCCCT CACAGCTCCCCACCGCCCCC	71.19	115
AUTS2	7	69597919	69598918	CCCCCTCCACGCCCCCGCCCCCGCCCCCCT CACAGCTCCCCACCGCCCCC	71.19	116
SYT15	10	46577217	46578216	CCCCCCAAAACACGCACTCCCCCACAACCA CCCCCTCCCCCCCCCC	71.11	117
IGF2BP3	7	23470468	23471467	GGGGGAGTGGGTGAGGGGGGCAGGGGGTG GGGG	71.05	118
QSOX1	1	180153834	180154833	GGGGGCGGGGGCTCGGGGGCGGGGG	70.97	119
AC010525.4	19	55694559	55695558	CCCCCCATGATGCCCCCGGGCCCCCCGGCC CCC	70.97	120
CXXC4	4	104494902	104495901	GGGGGCGGGGGCGCGGGGGTGGGGG	70.97	121
SHANK1	19	50719451	50720450	CCCCCAAACCTCGCCCATCCCCCACCACCG CCGGCTCCGCCCCCTACCCTTGCCCTCCCC C	70.91	122
NOP14-AS1	4	2933899	2934898	CCCCCAACCCGAGCGCCCCCTACGCCCCCC ACCCCGGGCGCCCCC	70.83	123
MLLT4-AS1	6	167826710	167827709	CCCCCTCCCCCCTCCGCCCCTTCTCTCCCCT CCCCCTCCCCC	70.83	124
ZNRF3-IT1	22	28991721	28992720	CCCCCCAACCCCCCGCCCCGGCCCCGCACA CCACCCCCCACCCCCC	70.83	125
B4GALNT4	11	368796	369795	CCCCCCCGCCCCCCGGCCTCTCTCGCCCGC CCCCATTCCCCGCCCCCC	70.83	126
PREX1	20	48827884	48828883	GGGGGCCGGGCGGCTGGGCCGGGGGCGGA GGCAGCGCGCGGGGGCCGGGCGAGGGGCG GGAGGGGG	70.76	127
GUSBP1	5	21340833	21341832	CCCCCCGCCCCCGTCCCCTGCCCCCCGCCC CCC	70.73	128
DUX4L16	Y	11305918	11306917	CCCCCGAATCGTGCCCCCCTCACCCCTACC CCCTCGCCCCC	70.73	129
DUX4L8	4	190066935	190067934	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	70.73	130
DUX4L7	4	190070232	190071231	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	70.73	131
DUX4L6	4	190073525	190074524	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	70.73	132
DUX4L5	4	190076818	190077817	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	70.73	133
DUX4L3	4	190086705	190087704	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	70.73	134
DUX4L4	4	190080118	190081117	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	70.73	135
DUX4L1	4	190083412	190084411	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	70.73	136
DUX4L2	4	190090005	190091004	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	70.73	137
DUX4	4	190172774	190173773	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	70.73	138
DUX4L24	10	133669450	133670449	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCC	70.73	139
DUX4L23	10	133672749	133673748	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCC	70.73	140
DUX4L20	10	133682668	133683667	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCC	70.73	141
DUX4L13	10	133752250	133753249	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCC	70.73	142
DUX4L15	10	133758859	133759858	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCC	70.73	143
RP11-374A4.1	5	21341376	21342375	CCCCCCGCCCCCGTCCCCTGCCCCCCGCCC CCC	70.73	144
PTPRR	12	70920844	70921843	GGGGGTGGGGGTGGGGGTGGGGAAGGGGG AAGTGCGGGGG	70.73	145
DKKL1	19	49360783	49361782	GGGGGACAGAGACCAGAGGGAGGGGGGGA CAGAGACCAGAGGGAGGGGGGGACAGAGA CCAGAGAGGGGG	70.59	146
RAC3	17	82030624	82031623	CCCCCCGCCCGCCCCCGCCCCCGCCCCCCG CCCCCC	70.45	147
KB-1562D12.1	8	101492500	101493499	CCCCCCTCCCCCGCCTCCCTCCCCCGGGTC CGCCCCC	70.45	148
AC108004.2	17	191588	192587	CCCCCGCCCCCGCGACAGTCACCCGTTCCC CCGCCCCCGCGACAGTCGCCCGTTCCCCC	70.31	149
AEBP1	7	44103361	44104360	CCCCCCGCCCCCCGCCCCCCGCCCCCCC	70.27	150
RP4-663N10.3	1	115478428	115479427	GGGGGGAAAGGGGGGAAAGGGGGGTGAAA GGGGGG	70.27	151
MIR4487	11	47399970	47400969	CCCCCCAACTCCAGCCCCCGCCGCAGGCCC GGCCCCCGGCCCCGCCCCCCTAGGGCCCCT CCCCCGCGCCGCCCGCCCCC	70.13	152
AC012531.25	12	54018910	54019909	CCCCCTCCCCCACCCCCCCACCCCC	70	153
CHD5	1	6180124	6181123	CCCCCCCACCCCCCCAAACCTCCACCCCCC CGTCTCGACCCCCC	70	154
GGN	19	38388083	38389082	CCCCCAAGCCGCCCCTTCCCCCCCCACCCG TCTTCCAGGTTCCCCCTCCCTCCCCCC	69.84	155
VGLL2	6	117264558	117265557	CCCCCCGGTCCCCGCCCCCCACCGCCCCCT CCCGCTCCGCAGCCCCC	69,81	156
AC116614.1	2	950275	951274	CCCCCATCCACCCCCGCGGGAGCGCCCCTG TCCCCCCATCCTGCACTTCCCCC	63,79	157
AC005391.2	19	928818	929817	CCCCCCGATCCCCCTGCTGCACCCCCCGGC CGGGCCCGGGCTGCCCCC	69.81	158
GRID1	10	86366494	86367493	CCCCCTCCCCTCCCCCTCCCCCTCAGCTCCG CTCCCCC	69.77	159
AC108004.2	17	191588	192587	CCCCCGCCCCCGCGACAGTCGCCCGTTCCC CCGCCCCCGCGACAGTCGCCCGTTCCCCC	69.7	160

RP11-513G11.2	3	194275682	194276681	GGGGGTAGTGTGGGGGTGGGGGTGGGGTG GGGGTAGGGTGGGGG	69.57	161
PPAP2C	19	291505	292504	GGGGGGAGGGGGAGGGAAGGGAAGAAGGG GGGGAGGGGAGGAGGGAGGGGGG	69.49	162
IGLON5	19	51310848	51311847	CCCCCTCCCCCTCCCCCCTCTCCCCCC	69.44	163
PPA1	10	70233912	70234911	CCCCCGCGCCCCCCGCCCCCTCCATCCCCC C	69.44	164
SLC30A3	2	27275818	27276817	CCCCCGCCCCCGCCCCCGCCCCCGCCCCC	69.44	165
WDR45B	17	82648554	82649553	CCCCCTTCAGGAGACAGAGACCCCCAGGAC ACAGAGGGCCCCCAAGTCACTCAGCCCCCC C	69.39	166
RP4-545L17.12	20	1185092	1186091	CCCCCACTCTGCCACTTCCCCTTCCCCCCAC CCCCACCCCGCACGCAGCCCCC	69.23	167
TMEM74B	20	1185416	1186415	CCCCCACTCTGCCACTTCCCCTTCCCCCCAC CCCCACCCCGCACGCAGCCCCC	69.23	168
LINC01451	9	136617182	136618181	GGGGGTGATGGGGGTGTGGGATTGGGGGC TGGGGG	69.23	169
ATP11B	3	182792500	182793499	GGGGGCGGGCCGGGGCGCCGGGGGCGGTC CCGGGTGGAGCCGGGGGGCCGGGCGGGGC GGGGCGGGGGTGGGGTTGCGGGGG	69.07	170
RP11-225N10.1	3	182793395	182794394	GGGGGCGGGCCGGGGCGCCGGGGGCGGTC CCGGGTGGAGCCGGGGGGCCGGGCGGGGC GGGGCGGGGGTGGGGTTGCGGGGG	69.07	171
HOXB8	17	48614940	48615939	GGGGGAGAGGGGGAGAGAGAGGGGGAGAG GGAGAGGGGG	69.05	172
ZNF480	19	52296177	52297176	CCCCCGCCCCCGCCCCCGCCCCCGCCCCCG CCCCCGCCCCCGCCCCCGCCCCCGCCCCCG CCCCCGCCCCCGCCCCCGCCCCC	68.89	173
MIR200B	1	1166092	1167091	CCCCCACCCACACCCCCACCCCCACCCCCG CCTGCACCCCC	68.89	174
DKKL1	19	49360783	49361782	GGGGGAGACAGAGACCAGAAAGAGGGGGA CAGAGACCGGGGGGGGGGG	68.89	175
AZIN1-AS1	8	102863300	102864299	CCCCCGGCGGCGAGACCCCCGCGCGCTCCC CCACTGGCACACGCCCCC	68.89	176
AC108004.2	17	191588	192587	CCCCCGCCCCCGCGACAGTCACCCGTTCCC CCGCCCCCGCGACAGTCACCCGTTCCCCC	68.75	177
PKD1P6	16	15131602	15132601	GGGGGAGGGGGAGGAGAATGGGAATTGGG GGAGGGGGATAAGGATGGGAAGGGGGAGG GGAGGAAGGGGG	68.66	178
PPP1R26P3	22	18404207	18405206	GGGGGAGGCGGCTGCCGGGGGTGGTGGGG GAGGGGGAGGGGG	68.63	179
ASCL1	12	102956686	102957685	GGGGGTGGCGGTAGGGGGCGATGGTACAG GGGGCCAGAGGTCATCCTAGGGGG	68.52	180
RP11-385H1.1	15	24847357	24848356	GGGGGGCGGGGGCGGCGGTGGGGGCGGG GGCGGCGGCGGGGGCGGGGGCGGGGCG GGGG	68.49	181
FXYD7	19	35142250	35143249	GGGGGCGGGGGCGGGGGTGGCCTGCACGG GGG	68.42	182
SATB1-AS1	3	18444024	18445023	GGGGGGGGGAGAAGGGGGAGGGGGCGGC GGCGGGGGCGGGAGGGGG	68.42	183
LGI4	19	35142452	35143451	GGGGGCGGGGGCGGGGGTGGCCTGCACGG GGG	68.42	184
RP11-560J1.2	6	15245001	15246000	GGGGGCGGGGCGCGCGGGGAGGGGGCTGG CGGCCGCCAGGGGAGGGGGGCCGGGGCCC GGGGGG	68.42	185
TBC1D5	3	18444818	18445817	GGGGGGGGGAGAAGGGGGAGGGGGCGGC GGCGGGGGCGGGAGGGGG	68.42	186
FTHL17	X	31072054	31073053	GGGGGGAGGGGCGGCGGGGGACAGAGTTG CGGGGCAAGGGGGTGGGGG	68.42	187
SH3GL1	19	4400548	4401547	CCCCCGCGCGCGCCCCCGGCCCCCGCGCGC GCCCCC	68.42	188
AC079776.1	2	129877554	129878553	GGGGGGGCGGGCGCGCGTGTCGGGGGCGG GCGCGCGTGTGGGGGGGCGGGCGCGCGTG TCGGGGGGCGGGCGCGGGGG	68.29	189
AC108004.2	17	191588	192587	CCCCCGCCCCCGCGACAGTCGCCCGTTCCC CCCCCCGCGACAGTCGCCCGTTCCCCC	68.18	190
ADNP	20	50931241	50932240	GGGGGAGGGGGCACAAGATGGCGGCGGCC GGGGGGGGGGGGG	68.18	191
CTD-2537I9.18	19	55613098	55614097	CCCCCGGCCCCCCGCCGCAGCCCCCGCCGC AGCCCCCTCCCCC	68.09	192
RP11-432J24.3	10	132441485	132442484	GGGGGGGCCGGAATGGGGGGGGCAGATGG GGGATTGGAATGGGGGGG	68	193
AKIRIN1	1	38990223	38991222	GGGGGGGATGGGACCCTGGGAGGGGCGGG GGTGGGGGTGGGGG	68	194
TBCD	17	82751064	82752063	CCCCCGGTCCCACCCGCCCCCGCCCCGCCC CCGCCCCGCCCCCC	68	195
DSTN	20	17568863	17569862	CCCCCACCTCCCCCACTAAGTGCCCCCTCG CCCCTCCCCCCACATCCCCC	67.86	196
BFSP1	20	17569221	17570220	CCCCCACCTCCCCCACTAAGTGCCCCCTCG CCCCTCCCCCCACATCCCCC	67.86	197
MIA3	1	222617086	222618085	CCCCCGTCCCTGCCCCCGTCCCTGCCCCCG TCCCCGCCCCGTCCCCGCCCCCGCCCCCGC CCCCGTCCCCGCCCCC	67.82	198
RP3-395M20.12	1	2545465	2546464	GGGGGCGGGGGGAGCGGGGAGTGGGGGGT GTGGGGGGTGAGGGGGGTGGGGGTGGGGG TGGCCCGGCCCGGAGCTAGGGGG	67.82	199
WNT7A	3	13880122	13881121	CCCCCGCCCCTCCCTCCTTTCCCCCGTCCCT CCCCCGCCCCCTCCCCC	67.8	200
CHRNG	2	232538727	232539726	GGGGGATGCCGAGTGAGGGGGAGGCCGAG TGAGGGGGATGCCCAGTGAGGGGG	67.69	201
AC007391.2	2	37324340	37325339	GGGGGGTTGGGGGTTTGGGGGATGGGGG	67.65	202
CDK6	7	92836595	92837594	GGGGGCTAGAAGGGCGCGGTGGGGGTGGG GGTGGGGGGG	67.57	203
SHANK1	19	50719451	50720450	CCCCCCTCCCCCCACCCCCCACCCCCCC	67.57	204

GRK5-IT1	10	119207531	119208530	GGGGGGTGGGGGGGAGCGTGTTGAGGGAG GGGGGAGGGGGG	67.5	205
HEXIM1	17	45147502	45148501	CCCCCGCTTCCCCCTCCACTCCCCCGCCCG CCCCC	67.5	206
SLC6A8	X	153687099	153688098	GGGGGAGGGGGAGGGGGGCCGGCCGGCCG TGGGGG	67.44	207
SYNDIG1L	14	74426103	74427102	CCCCCTGCCCCCCTCCCCCTGTCCCCCTTCC CCC	67.44	208
RP11-159D12.3	17	57987180	57988179	CCCCCCTGTCCCCCCCGTGCCCCCCCGGGG GGGCCCCC	67.4	209
KLHL25	15	85795031	85796030	GGGGGGTGGGGGGCCGGGGGCGGGGAGG GGGGGGGACCGGGGG	67.39	210
SLC24A3	20	19211646	19212645	CCCCCTCCGCCCCCGCCTCCCGGCCGCCCC CGCCGCCCCC	67.35	211
LPPR3	19	821978	822977	GGGGGGCGCCATCGAGACGGGGGCGGAGG CGGGGGCGGGGG	67.35	212
MIR4497	12	109832348	109833347	CCCCCACCCCCACCCCCGCTACCCACCCCC AGGAGCCACCCGCCCCC	67.31	213
RP11-65J21.3	16	14301288	14302287	GGGGGTTGTGTGTGTTTGTGCGGCGGGGGA AGCGGGGGGTGCTGAAAATTGGGGG	67.31	214
RP11-326A19.5	15	89201769	89202768	GGGGGCGGCTTGCTCGCTGGGGGCGGGGG ACGATGGCGAGAGGGGAGGGGG	67.27	215
ZFP41	8	143245821	143246820	GGGGGCTGGGGGTCGGCGGAGGCGGGAGC TGGGGGTCGGCGGGGGCTGGGGG	67.24	216
ZFP41	8	143246110	143247109	GGGGGCTGGGGGTCGGCGGAGGCGGGAGC TGGGGGTCGGCGGGGGCTGGGGG	67.24	217
POU3F3	2	104854511	104855510	GGGGGCCCGGGGCGGGGGCGGGGAAGGAG GGGGGGAGGAGGCGGGAGGCGGGGGG	67.21	218
LINC01521	22	31345777	31346776	CCCCCCTTCCCTATTTATACACCCCCCCCCC GCAGTCCGCTCCCCCCGCTTCCTGCCCCCC C	67.12	219
PATZ1	22	31346233	31347232	CCCCCCTTCCCTATTTATACACCCCCCCCCC GCAGTCCGCTCCCCCCGCTTCCTGCCCCCC C	67.12	220
AGPAT3	21	43864186	43865185	CCCCCCGTTCAGCCCCGACCCCCCTGATCC CCGCGCCCCCCGCGGTCCCGGCCCGCCCCC TCCCCC	67.11	221
ARHGEF7-AS2	13	111115679	111116678	GGGGGGCGCCATTGTGTGCGGGGCAGGGG GAGGGGGCCGGGGAGCGGGGTCGGGGGGA GGGGCGGCGCGGGCGGCGGGGG	67.06	222
AKAP11	13	42271153	42272152	GGGGGCGGGCCGCGGGGCTGGGAGGGGGA GGGCTGGGGGTCAGCGCGGGAGGGGG	66.67	223
PEG10	7	94655325	94656324	CCCCCACCCCCATCCCCCATCCCCGCACCC CCCTCCACCCCC	66.67	224
GAREML	2	26172091	26173090	GGGGGCGGGGGTCCGGGAGGAGGGGGCCG GGGTCCGGGAGGAGGGGGCGGGGCCCGGG GG	66.67	225
ABAT	16	8673565	8674564	CCCCCGTCTCCCCCGCCCCCGGAGGCTGGA CTCTGCCCCC	66.67	226
TCN2	22	30605838	30606837	CCCCCCCCCCACCGAACTGCCCAACTCCCC CTCCCCAACTGCCCAACTCCCCCACCCCC	66.67	227
THEG	19	376014	377013	CCCCCCCCGTGCCTCCACCCAGCCTCTCCC CCTCTGTGCCCCCACCCAGCCTCTCCCCC	66.67	228
TRIM6-TRIM34	11	5595725	5596724	CCCCCCTTCCCCCGTTCTCCCCTTCCCCCCT TCCCCCTTCCCCCTTCCCCCTTCCCCCTTCC CCTCCCCC	66.67	229
C20orf166-AS1	20	62551527	62552526	CCCCCACCCCCCGCCCCCACCCACCCCC	66.67	230
KCNAB3	17	7929804	7930803	GGGGGTGGGGGTGCGGGGGCTTACAGTGG GGG	66.67	231
GALNT18	11	11622006	11623005	CCCCCCCCCCACCCCCCACCCCCCACCCCC	66.67	232
AC010907.5	2	3704154	3705153	GGGGGAGGGGGGAGGGGCGTGGGGGGGA GGGGCGTGGGGGG	66.67	233
SLC16A2	X	74420461	74421460	GGGGGCAGGGGGAGACTCCCGGGGGTGCG CAGGACCGCGGGGGG	66.67	234
AP006547.3	8	142619373	142620372	CCCCCCACCCCCACCATCCCCCCCCCCCC	66.67	235
DNMT3L	21	44262217	44263216	GGGGGATTAGGGAGTGGGGGTGGGGGTGG ATGCTGAGTTGGAGATGGGGG	66.67	236
TMEM253	14	21097937	21098936	GGGGGGAGGGGGCCGGGGGCGGGGGCCG GGTGGCGGACGGGGGGGCCCGGGAGTCGG GGG	66.67	237
GSC	14	94770231	94771230	GGGGGAAGAGGGTTGGGGGGAGGGGGAGA GGTGCCAAGGGGAGGGGG	66.67	238
SMARCA1	X	129523501	129524500	CCCCCCGTCACTCTCCCCCCTTCCTCCACCC CCCCCCCC	66.67	239
GATA2	3	128493186	128494185	GGGGGGGCAGGGGGTGGCGACGGGGGCGG GAGCGCCGCAGGGGG	66.67	240
SLC7A10	19	33225851	33226850	CCCCCTCCCCCACCTCCCTTAAAGGGAACG CCCCCTCCCCC	66.67	241
SGCE	7	94656210	94657209	CCCCCACCCCCATCCCCCATCCCCGCACCC CCCTCCACCCCC	66.67	242
AC010761.14	17	28726258	28727257	CCCCCGCCCCCAGCCGCCCGCGCCCCCGCC CCGCCCGCCTGCCCCCGCCCCGTCCGCCTG CCCCC	66.67	243
RASA4B	7	102517782	102518781	GGGGGGGCGGGCCGGGGAGGAGGCGGGG GCGGGGGCGGGGCGGCGGGGG	66.67	244
SYNE2	14	63851983	63852982	GGGGGGGGGGGGGGGGGGAATGAAATGGA TGAAATGGGGGGGGGGGG	66.67	245
GSDMD	8	143552207	143553206	GGGGGTGGGGGGATAGGGAGGTGGGGTGG GGGGTGGGTTGTGTGTAGAGCAGGGGG	66.1	246
SLC12A7	5	1112036	1113035	CCCCCCCACCCTTCCTGCACCCCCTCCTGCT GGGCCCCCCTGCCCCTATGAACCCCC	66.1	247

RP3-441A12.1	6	37819219	37820218	GGGGGTAAGTGCCCGGCCGGGTGGGGGCG GGGGGCGGGGGCCGGGGG	66.04	248
PPP1R26P2	22	18653618	18654617	GGGGGAGGCGGCTGCCGGGGGTGGTGGGG GAGGGAGAGGGGG	66.04	249
TEAD3	6	35497077	35498076	GGGGGTGGGGGCTGCGGGAAGGAGGGGGC GGGGCGCGCAGGGGG	66.04	250
PPFIA3	19	49118389	49119388	GGGGGGGGAGGGGGAGGGAGGAGGAATTT TTTTGGGGGGAAGGTGGGATTTGGGGGG	66	251
PPP1R26P5	22	21340595	21341594	CCCCCTCCCCCTCCCCCACCACCCCCGGCA GCCGCCTCCCCC	66	252
PPP1R26P4	22	18714815	18715814	CCCCCTCCCCCTCCCCCACCACCCCCGGCA GCCGCCTCCCCC	66	253
C19orf73	19	49119141	49120140	GGGGGGGGAGGGGGAGGGAGGAGGAATTT TTTTGGGGGGAAGGTGGGATTTGGGGGG	66	254
SLC22A23	6	3457023	3458022	CCCCCTCCTCCCTGCCCCCCGGCCCGCCCC CTCTCCCCCC	66	255
HIST4H4	12	14771132	14772131	CCCCCACCTACACTCCACCCCCACCTCCCCC CGACCCCC	65.96	256
CA11	19	48646313	48647312	GGGGGAGGAGGTGAAGGAGGAGGGGGAGG AGGGAGAGGAGGGGGAGGGAGAGGAGGGG GAGGAGGGAGAGGAGGAGGGGG	65.93	257
PUS1	12	131928200	131929199	CCCCCACCCCCTAGTCTCCGCCCCCATCACC CCC	65.91	258
PAXIP1-AS1	7	155002448	155003447	CCCCCCGCCCCCGGCCCCCGCGGCGCCCGG CGCCCCC	65.91	259
C1orf233	1	1600097	1601096	GGGGGATCGGGGGTGGGGGGGAGGAGGGG G	65.85	260
CNOT6LP1	15	56007177	56008176	GGGGGAGGCGGCGGCGGCGGGGGAGGCGG CGGCGGCGGGGGAGGCGGCGGCGGCGGGG GAGGCGGCGGCGGCGGGGGAGGCGGCGGC GGCGGGGG	65.81	261
RP11-436F23.1	4	46389255	46390254	CCCCCCCTCCCCCCCCTCCCCCCCCCCC	65.79	262
HERC5	4	88456117	88457116	GGGGGGTGGGGGTGGGGGTAGGGGTGGGG GGG	65.79	263
RP11-20I20.2	4	1112639	1113638	CCCCCCATCTCCTCGGTATCCCCCTTCCCCC CCAAGCCTTGACCCCCCACATCCCCCACCTT GCCCCCC	65.79	264
AC018878.3	2	106178542	106179541	CCCCCACCCCCAGCCCCCTTTCCCCC	65.71	265
LMNA	1	156081573	156082572	GGGGGCGGTGGGGGGCTGGGGCGTCTGCG CTGCGGGGGCCGCCGTCGGGGCGGCCGGG GG	65.67	266
PTK7	6	43075268	43076267	GGGGGCGTCGAGGGGGCGGGGCGCCGTGC GGGGGCCGGAGGGGAGGGGG	65.63	267
ULK1	12	131893651	131894650	GGGGGCGGGAGGGGAGCGGGGGGGAAGGC ATGGGAGGGGGCGGAAGGGAGGGGG	65.63	268
MIR431	14	100880007	100881006	GGGGGATGGGGGTTGGGGGGCGGGGG	65.63	269
RP1-169K13.3	X	118116801	118117800	GGGGGCAGTGGGGGGGGGAGCGCGCGGGC GGGGGAAACATGGCGGCGGGGG	65.57	270
DUSP4	8	29350669	29351668	CCCCCGCCCCCCGCGCTGTCCCCCGCCCCC CGCGCTGACTCCCCGCCCCCTCCGGGCCCC TCCTCCCACCCCC	65.48	271
MIR3665	13	77698117	77699116	GGGGGCTTCGGGCACCGGGGAGCGGGGGC GGGGGGCGGACGGGGG	65.45	272
POLR2J	7	102478908	102479907	CCCCCCCCCCCGCCCCCGCCACCCCGCACC CCGCCCCCC	65.38	273
AC093668.2	7	102479080	102480079	CCCCCCCCCCCGCCCCCGCCACCCCGCACC CCGCCCCCC	65.38	274
RP4-725G10.4	7	56174634	56175633	GGGGGTGGGGGGGGGCGGGGGGGCGTCCT CTCCAGGGGG	65.31	275
CTD-2132N18.2	17	42043118	42044117	CCCCCAACCCCCACCTGAATGCTCTTGCCCC CCTGCCTCCCCC	65.31	276
RP11-154D17.1	9	81688713	81689712	GGGGGGCGCGAGCGCGGAGCCGGCGGGGG ACGTGAGCCCGGGAGGTTGGGGGGTGGGG GGGGGG	65.28	277
TLE1	9	81689306	81690305	GGGGGGCGCGAGCGCGGAGCCGGCGGGGG ACGTGAGCCCGGGAGGTTGGGGGGTGGGG GGGGGG	65.28	278
DGCR5	22	18969514	18970513	CCCCCCAGCCCCCGCCCAGGCCCGCGCCCC CCAGCCCCC	65.22	279
PRRC2B	9	131393093	131394092	CCCCCGTCACTGCCCCCTCCATCCCTTCCCC CTCTTTCCGGGCGCCCTTCCCCC	65.15	280
FAM86FP	12	8242949	8243948	GGGGGCGGGTCCGGGGCAGTGCGCGGGGG CAGGGAGGGGGCGGGGCCTGGGGGG	65.15	281
NRG1	8	31638386	31639385	GGGGGTAGGGGGACAGATGGGGGTGGAGT TGGGGG	65.12	282
HOMER3	19	18941262	18942261	GGGGGTTAGGGGCGGTGGAGGGGGAGTGG GGGGGGGGG	65.12	283
PASK	2	241150265	241151264	GGGGGAGCGGGGGAGGCGCTGGACCTCGG GGGAGCCCCGGCCCGGGGG	65.08	284
PDE9A	21	42652636	42653635	GGGGGAGGGAGGGGGGAGGGGGGAGGGG G	65	285
GABRD	1	2018324	2019323	GGGGGCGCGGCTGGCGGAGGGGGGCGGGA TGGGGCGGGGGGCGGGGGG	65	286
STK32A	5	147233963	147234962	CCCCCCTTCTCCCCCCCCCCCCCCC	65	287
PID1	2	229271286	229272285	GGGGGGAAGGGCCTGGAGCCGGGGGTGGT GGGGGCAGCGGGGATGTCGGGGGTGGCGG GGGTGGGGG	65	288
MN1	22	27801499	27802498	GGGGGGAGGGGGGCGCGGGGGCAGCTCTG GGGGG	65	289
KCNQ2	20	63472678	63473677	GGGGGGAGACAGGAGTGGAGGGGGAGGGG AGGGGGGAGAGGAGGGGGAGTGGGGGAGG GGGAGGGGAGGGCAGGGGGAGAGGAGGGG G	64.89	290
MIR6751	11	65129979	65130978	GGGGGGCTGAGGCAAGAGTGGTGGGGGGT GGGGGGCAGGGGG	64.81	291
RP11-65J21.3	16	14301288	14302287	CCCCCCCACCCCGGCCCCCGCCCCCACCCC CACCCCCACCCCCACCCCCACCCCC	64.79	292

PNPLA3	22	43922739	43923738	CCCCCGCCCCCAATCCCCCTCCCAGACCCG GTCCCCGCCCCCATCCCCCGCCAGACGTCG TCCCCACCCCCATCCCCC	64.77	293
RSPO3	6	127117604	127118603	CCCCCGAGCTCCCCCCCGCCCCCCACAGCC ATCCCCCCTCCTCTCCCCCTCCCTGCCCTCC TCTCCCTCCCCCC	64.71	294
SOCS5P4	X	71045471	71046470	CCCCCACCCATTCCCCAGCCATTTCCCCCTC TTCACCCATTTCCCCTCCCCCCACAATCCCC C	64.62	295
GDF11	12	55742280	55743279	CCCCCGGCCCCCCGGGTCCCCCCTCGGCCG GGCAGCCCCCAATCCCGCGCCGCCCGGACC CCC	64.47	296
FXYD7	19	35142250	35143249	CCCCCATTGGCTGCAGGCCCCCCCACGCCA GGGCTCCCCCCAGCTCCCCCCC	64.41	297
LGI4	19	35142452	35143451	CCCCCATTGGCTGCAGGCCCCCCCACGCCA GGGCTCCCCCCAGCTCCCCCCC	64.41	298
SCT	11	627144	628143	GGGGGCGGAGGGGCCGGAGGGCTTGGGGG AGGGTTTGGGGGGGGACTGGGGGG	64.41	299
CAMK1D	10	12348482	12349481	GGGGGAGGAGGGGGAGGAGGGGGAGGAG GCGCGAGGTGGGGGGG	64.29	300
FBXL19-AS1	16	30923270	30924269	GGGGGAGGGGGAGAGGTCGGGGGTGCGCG CGGGCTGGGGGGGG	64.29	301
SYT15	10	47763593	47764592	GGGGGGGGGAGGGGGTGGTTGTGGGGGAG TGCGTGTTTTGGGGGG	64.29	302
BRAF	7	140924765	140925764	GGGGGAAGGACGCTAGGCGGGGGTGCGGG GGGGAGCGGGGGAAGGGGG	64.29	303
RP11-444D3.1	12	24562591	24563590	GGGGGGTGCGGCGGGGGCGGGGGGGGGG CGCGGGGGCGGTGCCGGGGG	64.29	304
CSAG1	X	152733736	152734735	CCCCCCGCACCCCCACCACCATTCCCATCCC CCAACACCAACCCCACCCCCATCCCCC	64.18	305
EPN2	17	19214615	19215614	CCCCCGCCCCCAACCCTCCCGCCACCCCCG CCGCCCCC	64.15	306
B3GAT1	11	134411919	134412918	GGGGGGCGGGGGGCGGGGAGGGGGAGCG GGGAGGGGGAGCGGGGAGCGGGCGCGGGG GCGAGAGGGGCGAGGGGGG	64.13	307
CRLF2	X	1212751	1213750	CCCCCCTCCCCCCACCCCCTCCTCCCTCCCC TCTCCCTCCCCCTTCCGTCTCCTCCCCTCCC CCTCCCCTCTCCTCCCCTCCCCC	64	308
DUX4L25	10	133666151	133667150	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	64	309
DUX4L22	10	133676059	133677058	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	64	310
DUX4L21	10	133679369	133680368	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	64	311
DUX4L11	10	133745641	133746640	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	64	312
DUX4L12	10	133748950	133749949	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	64	313
DUX4L14	10	133755559	133756558	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	64	314
DUX4L10	10	133742332	133743331	CCCCCGAAACGCGCCCCCCTCCCCCCTCCC CCCTCTCCCCC	64	315
MIR4298	11	1859537	1860536	CCCCCAGCCCCCCAGCCCCCAGCCCCC	63.89	316
RP11-48B3.5	8	80485620	80486619	CCCCCCTCCCCCAGCCCGGCCCCCACCCTT TCCCCC	63.83	317
FAM132B	2	238157982	238158981	CCCCCTCCCCCACTCCCCCTCCCCCGCCCG GACACTCCCCCC	63.79	318
SNTG2	2	949868	950867	CCCCCATCCACCCCCGCGGGAGCGCCCCTG TCCCCCCATCCTGCACTTCCCCC	63.79	319
HMGCL	1	23838621	23839620	CCCCCGTCGCACGCCACACCCCCCTCGCAC GCCACACCCCCGTCGCACGCCACACCCCCC	63.79	320
BGN	X	153493939	153494938	GGGGGAGGGGGGGGGAGGTGGTGGGGGCG AGGGGGCGGGGG	63.64	321
RP11-452L6.8	16	31427180	31428179	CCCCCCCCCCGCAGCACCCCCCCCCGCCCC CGCAGCACCCCCCCCCGCCCCCGCAGCACC CCC	63.64	322
TAF4	20	62065811	62066810	GGGGGGCGGCGCGCGGCCGGCGGGGGGCG GGGGCGCGCGGCCGGGGGGCGGGGGAGCG GGGG	63.64	323
FBXO36	2	229921302	229922301	CCCCCCGCCCCCCCAACCCCTCAGCTGGCC CCCTCCCCCAAGGCTTGCCCCC	63.49	324
AKIP1	11	8910139	8911138	CCCCCAGCCCCCGACCCCCGACCCCCGGCC CCCGACCCCCGACCCCCGGCCCCC	63.38	325
ST5	11	8910952	8911951	CCCCCAGCCCCCGACCCCCGACCCCCGGCC CCCGACCCCCGACCCCCGGCCCCC	63.38	326
NDUFAF2	5	60944129	60945128	CCCCCCCCCCCACCCCCGGGGAACCCCCCCC ACCCCCGCCGCCCCCC	63.33	327
CTC-492K19.7	19	40425115	40426114	CCCCCACCCCCGACTCCGCCCCCGGGTCCC GGCGCCCCC	63.27	328
MEIOB	16	1872179	1873178	CCCCCACCGCTCACTGACTCCGCCCCCACC GCTCACTGGCTCCGCCCCCACCGCTCACTG GCTCCGCCCCC	63.1	329
THEG	19	376014	377013	CCCCCCTGTGCCCCACCCAGCCTCTCCCCCT CTGTGCCCCCACCCAGCCTCTCCCCC	63.08	330
XKR6	8	11201367	11202366	GGGGGGGAAACGAATGGAGAGGAAGGGGG GCGGGGAGGAAGCGGGGGAGCAAACGAAC GAGGGGGG	63.01	331
MEF2BNB-MEF2B	19	19192159	19193158	GGGGGGCGGGGGTGCGCAAGCGCAGTGGG GGAGCTCTGGGGTGGGGG	62.96	332
MEF2BNB	19	19192592	19193591	GGGGGGCGGGGGTGCGCAAGCGCAGTGGG GGAGCTCTGGGGTGGGGG	62.96	333
STX1B	16	31010629	31011628	GGGGGCATGCGCAGCACCCGGGCCGGGGG CGCCTGGGGGTTGTAGTCACGGGCTTGGGG GACACCCGAGGAGGGGG	62.96	334
TMEM177	2	119678167	119679166	CCCCCCACGGACATCCCCTTGAGGCCCCCC CAGCCACACCCCCAAGAGTCCCGCCCCC	62.9	335
AC007405.6	2	170770113	170771112	CCCCCGCCCCCGCCCCCGCCACCTCCCCCG CCGTCGCCCAATTCCCCCAGCCCCC	62.86	336

AC007405.4	2	170770769	170771768	CCCCCGCCCCCGCCCCCGCCACCTCCCCCG CCGTCGCCCAATTCCCCCAGCCCCC	62.86	337
C8orf34-AS1	8	68331492	68332491	CCCCCGCCTTGCCCCCGCCTTGCCCCCGCC CCCC	62.79	338
RP11-432J22.2	9	135906812	135907811	GGGGGCGGGGGCGGGCGCGGGGGCGGGA GCGGGCCGGGGG	62.75	339
RBPMS2	15	64775588	64776587	GGGGGACCGAGTTCTGGGGGTGATGGCCG GGGGGAGGGGG	62.75	340
BCORL1	X	129980107	129981106	GGGGGCTCTGCCCAGGTAAAGGGGGCGCC GTGAGGCGGGGGTGGTGGGGGAACGGTGC GGCCCTGCGGCGGGGGCGGGGGG	62.75	341
CAMSAP1	9	135907229	135908228	GGGGGCGGGGGCGGGCGCGGGGGCGGGA GCGGGCCGGGGG	62.75	342
MIR670HG	11	43555758	43556757	CCCCCCACCCTCTTTTTCTCTCTCCCCCTCC CTCTCACCCCCTCTCTCTCTCCCCC	62.69	343
NOC2L	1	959310	960309	GGGGGTCGGGGGAGCGGGGATGGCTTCCG GGGGAGGAGGAGGGCGAGGCCTGGGGGG	62.69	344
KLHL17	1	959587	960586	GGGGGTCGGGGGAGCGGGGATGGCTTCCG GGGGAGGAGGAGGGCGAGGCCTGGGGGG	62.69	345
LINC00567	13	110813085	110814084	CCCCCTCACGACGCCCCCGTCACGACGCGC CCCCTTCGTCACGTCCCCCTCACGACCTCCC CC	62.69	346
RP11-395B7.4	7	101017609	101018608	CCCCCGGGACTCCCTTCTCCCCCTGGGACT CCCTTCTCCCCCTGGGACTCCCTCCCTTCCC CC	62.67	347
NPIPB15	16	74376878	74377877	CCCCCCTTCCCCTCCCCTCCCCCTCCCCCTC CCCTCCCCCTCCCCTCCCCCTCCCCTCCCCC TCCCCTCCCCCTCCCCTCCCCC	62.62	348
PALM	19	708093	709092	CCCCCGCCAGGCCGCGTCCCCCTCCCCTCC CCTCCCCCGCGCGCCACCCGCGCCCGCCCC C	62.5	349
IKZF4	12	56006659	56007658	GGGGGAGGGGGACGGCTGTGATGGGGGGG GATATTTGGGGGGG	62.5	350
RP13-122B23.8	9	137295722	137296721	CCCCCACCCCCACATGCCCCCCACGGCCCG GCTCCCTCCCCCCGCACCCCGGCCCCAGCC CCCC	62.5	351
PRDM16	1	3068168	3069167	GGGGGCGAAGGGGAGGAGGCGGGGGCCGG GAGGCGCGAGGGGGGCGGTGGGGCCCGGA GGGGG	62.35	352
AL158069.1	X	69503705	69504704	GGGGGAGCCAAAAAGCTAGGACGGGGGCT GGTGCTTGGGTGGGGGGCGGGCGGGTGGG GGG	62.32	353
ANKRD20A11P	21	13980438	13981437	CCCCCAACAAGTCCCCCGCCGCGGGCGGTG CAGCCCCCAAACCACCCCCCC	62.3	354
TRERF1	6	42452052	42453051	GGGGGGGGGGAGCAGGAGCAAAAAGAGGG GGAGGGCGGGGGAAGGGAGGGGG	62.3	355
THTPA	14	23554988	23555987	CCCCCGCCCCCTTCCCTCCCCCCACCCCAAC CCCCC	62.26	356
AL591893.1	1	151993531	151994530	CCCCCAGCCCTGTCTCCTCCCCCTCTTCCTG CCCCCGACTCCCCC	62.26	357
RHOXF2	X	120157561	120158560	CCCCCCCGCCCTACTCCCCCCCTCCACCCTA CGCCCCCCCGCCCTACTCCCTCCCCCGCCC TACTCCCCCCACCCCGTCCTACTGCCCCC	62.16	358
RP11-126K1.2	1	151281930	151282929	CCCCCGCCCCTCCCCCACTTGGTCAACTACC CCCTTCTCCGCGCCGCCCCC	62.07	359
CTD-2589H19.6	5	674826	675825	GGGGGTGCAGTGTGGCTGGGGGTGCAGTG TGGGGGTGTGCAGTGTGGCCGGGGG	62	360
FP671120.7	21	8216953	8217952	CCCCCCGCCCTACCCCCCCGGCCCCGTCCG CCCCCCGTTCCCCCC	62	361
FP236383.12	21	8444223	8445222	CCCCCCGCCCTACCCCCCCGGCCCCGTCCG CCCCCCGTTCCCCCC	62	362
FP236383.4	21	8399991	8400990	CCCCCCGCCCTACCCCCCCGGCCCCGTCCG CCCCCCGTTCCCCCC	62	363
DUSP16	12	12562384	12563383	GGGGGAAAGGCGGGGGGGTGGGGTGGGGG GTTGGGGG	62	364
RNF212	4	1113563	1114562	GGGGGCGTGAGGCGGTGGGGGTGGGGTGT GATGAGGCGGTGGGGGTGGGGGGGGCAGT TGGGGGTGGGGG	61.9	365
PROSER2-AS1	10	11894701	11895700	GGGGGAGGGGGCCCACCGGAAGTGGGGGG AGCTTGTCCTCCGGAAAGCGGGGGAGGAGG CCCGCTGGAAATGGGGG	61.9	366
FP671120.6	21	8257914	8258913	GGGGGCGGGCTCCGGCGGGTGCGGGGGTG GGCGGGCGGGGCCGGGGGTGGGGTCGGCG GGGG	61.85	367
FP236383.9	21	8440949	8441948	GGGGGCGGGCTCCGGCGGGTGCGGGGGTG GGCGGGCGGGGCCGGGGGTGGGGTCGGCG GGGG	61.84	368
LIN28A	1	26409778	26410777	GGGGGAAGAGAGGGGGAGGGGAGCTCAAG GGCTGGGGGAGGGGG	61.82	369
JAZF1-AS1	7	28179322	28180321	CCCCCTCCCCCACCCGAGCCCGCTCCCCGC CCCCCGGCACCTCGCGCCCCCGCATCCCGC CACCCCC	61.8	370
DYNLL1	12	120468850	120469849	GGGGGGAGGGGAGGCCGGGGGGAGGGGA GCCCTGGGGGAGGGGACAGCAGGGAAGGC GGGGG	61.76	371
PI4KAP1	22	18577969	18578968	CCCCCGGGACAGGGACCCTGGCCCCCCCCG ACAGGCTGACGCCCACCCCCTCAAACTCTG GTGGACTTACCCCC	61.73	372
PI4KAP2	22	21517534	21518533	CCCCCGGGACAGGGACCCTGGCCCCCCCCG ACAGGCTGACGCCCACCCCCTCAAACTCTG GTGGACTTACCCCC	61.73	373
ZNF618	9	113875282	113876281	CCCCCCCCGCGCGCCCCCAGCGCCCCCGCC CCCCC	61.7	374
CRLF2	X	1212751	1213750	CCCCCACCCCCTCCTCCCTCCCCTCCCCCTC CCCC	61.7	375

RP11-365N19.2	14	102932574	102933573	GGGGGCTCCATGCGGGGGCAGGACTGGTG GGGGGGGGGG	61.7	376
CBX2	17	79777132	79778131	CCCCCAGCCAGGGGCCCGCCCGCGCCCCCA CCCCCGCCCCCGCCCCC	61.67	377
UTF1	10	133229274	133230273	GGGGGCGGCGGGGGCGGGGGCAGCACTGG GGGCCGGACAAGGGGG	61.67	378
AP000946.2	21	19899756	19900755	CCCCCACCACCGCCCCCCTTCCCCCCGCCG GCCTCCCTGCACCCTCCCCCC	61.64	379
RP11-482M8.3	16	87602191	87603190	GGGGGCGGGGGCGGGGGCGGGGGCGGGG GGCGGGGGGCGGGGGGCGGGCGGGGG	61.64	380
RHOXF1	X	120115938	120116937	GGGGGGATGGGAAGTTGGGGGGTGCGGGT GGGGAGTGGGGGTGTTGCAGGTGGGAGTG GGGGG	61.64	381
EEFSEC	3	128152454	128153453	CCCCCTCCCCCACGTCCCCCCGCCCTCCCC CGCCCCCCGCCGTCCCCCCGCCCTCCCCCC CGCCCCCGTCTTCCCGCCCCCCCCGCCCCC GTCTTCCCCCCGCCCACCCCGCCCCCCCCG CCCCCCTTCGCCCCCCC	61.58	382
DAGLA	11	61679433	61680432	GGGGGTAGGGGTGGGGGGTGGGGGGGTGG TAGGGAGGAGGGGG	61.54	383
CTD-2105E13.15	19	55378126	55379125	GGGGGCTGGAGGGGGCTTGGGGATGGAGG CTTAGGGGGCGCCAGGCTGTGGTCGTGGGG GTGTGAGCTGAGGGGG	61.54	384
TVP23C-CDRT4	17	15563562	15564561	GGGGGCGGGGCGAGCGGGGGCGGGGCGAG CGGGGGCGGGGCGAGCGGGGGCGGGGCGA GCGGGGGCGGGGCGAGCGGGGGCGGGGCG AGCGGGGG	61.42	385
TVP23C	17	15563596	15564595	GGGGGCGGGGCGAGCGGGGGCGGGGCGAG CGGGGGCGGGGCGAGCGGGGGCGGGGCGA GCGGGGGCGGGGCGAGCGGGGGCGGGGCG AGCGGGGG	61.42	386
QSER1	11	32892178	32893177	CCCCCTCTTCCTCCCCCGCCCCCTCTCCCTG CCCGCCCCC	61.4	387
MMD2	7	4959214	4960213	GGGGGTCTTGGAGGGAGTGGGGGACCTGG GGGTGGGGGGAGTGGGAAGTGGGGGTGTT GGAGGGAGCGGGGGACCTGGGGG	61.36	388
MIR3613	13	49996502	49997501	GGGGGAGGGGGAGGGGAAAGGGAGGGGGA GGGGGAGGGGAAAGGGAGGGGGAGGGGG GAAGGGACGGGAGGGGAGGGGGAAGGGAC GGGAGGGGAGGGGG	61.36	389
RHOXF1	X	120115938	120116937	GGGGGTTGCCGGTGGGGGTACATGGTGGG GGTGGGGG	61.36	390
AL662797.1	6	30683796	30684795	GGGGGGCAGTTGGGGCTGGGGGTGGGGGA GACAGAGGGGCTGGTGGTGGGGG	61.29	391
CTB-161C1.1	7	5555731	5556730	GGGGGCGCCCTCGGAGCCATGGGGGGAAG GGGGTATCGGCCGGGTGGGGG	61.29	392
MEIS3P1	17	15785618	15786617	CCCCCGAGCCGCCGCCGGACCCCCGCCGCC CCGGGCCGCCCCCCGCCCCCC	61.29	393
MEIS3P2	17	20588293	20589292	CCCCCGAGCCGCCGCCGGACCCCCGCCGCC CCGGGCCGCCCCCCGCCCCCC	61.29	394
SH3PXD2B	5	172454524	172455523	CCCCCGGCCCGCCCCCCGCACCGCCCCCCG CACCGCCCCCCGCACCGCCCCCCGCCGCC CGGCCCGCCCCCCTTAAAGGGCCCGGCCCC GACCCCC	61.29	395
RP11-467J12.4	16	53052874	53053873	GGGGGAGCAGAGTGGGGGCCGGGGGGAGG GGG	61.22	396
RNF212	4	1113563	1114562	GGGGGAGGGGGAGGGGGAGCGCCGAGGGC GGGGG	61.22	397
PASD1	X	151562622	151563621	GGGGGCAGCGGGGGGGGCGGGGGGTGGG GGGGCGGGGGGCGGGGGGTCGGGGGTCGG GGGGGCGGGGGGCGGGGGGGGCGGGGGG GCGGGGGGTGTCGGTGGTCGGGGGGGGGC GGGGGGGG	61.22	398
PLCB2	15	40307936	40308935	CCCCCGCCTCTTCTGGAGGCCCCCGCCCCC ACCCCC	61.22	399
MIR5004	6	33437331	33438330	GGGGGATCTGGGGGCATGGGTTCGGGAGG GGGAGGGGGCTCGGGGGGTGGCTCAGGGG G	61.19	400
ZXDC	3	126475920	126476919	GGGGGGCGGGGCGCGGGGTCGGGGGGCGG GGCGCGGGGTCGGGGGGCGGGGCGCGGGG TCGGGGGGCGGGGCGCGGGGTCGGGGGGC GGGGCGCGGGGTCGGGGGG	61.19	401
LINC00982	1	3068438	3069437	GGGGGCGAAGGGGAGGAGGCGGGGGCCGG GAGGCGCGAGGGGGGCGGTGGGGCCCGGA GGGGG	61.18	402
ADGRB1	8	142448430	142449429	CCCCCCCTTATCTCGCTCCCCTCCCCCCACA CCCCCGCCCCCCGCTAGCGCTCCCCC	61.11	403
CRIM1	2	36354926	36355925	GGGGGTGGGGGGAAGGTCTGGGGGAGGAG GTAGGATGGGGG	61.11	404
RP11-490M8.1	2	36355115	36356114	GGGGGTGGGGGGAAGGTCTGGGGGAGGAG GTAGGATGGGGG	61.11	405
BRD1	22	49827513	49828512	GGGGGAGGGGGCGCCGCGAGGAGCGGGGG CCAAAGGGGG	61.11	406
RP11-17M16.2	18	76490652	76491651	CCCCCGCCGGCTCCCCCCGCCCGCCCCGCC CGCCCCCGCCCCCCGGCCCGCACAGACCCC C	60.98	407
RAE1	20	57350010	57351009	GGGGGGACGGGGGGGCGGGGGGGCGGGG GGACGGGGGGCGGGGGGGGCGGGGGGGG CGGGGGGACGGGGGCGGGGGGCGGGGGG GGGCGGCGCGGGGGGGCGGGGGGGGCGG GGGG	60.96	408
RP11-462L8.1	10	32957845	32958844	CCCCCAACCGCTGGCGGCCGCCCCCGCCCC CGCCCCCGCCCCCACCCCC	60.94	409
C8orf58	8	22598601	22599600	CCCCCTAGGCCACGCCCCCCACCGCCCCGC CCCCCGCCCCCCCAGCCCCGCTCCCCGCCC CCCC	60.92	410
RP1-118J21.25	1	39787976	39788975	GGGGGGGCGGGGGCGGGGCCTGGGAGCGG GGGCCACGCCTGCGAGGGGG	60.87	411

RP11-896J10.3	14	36519522	36520521	CCCCCGCTAACACGCCTCCCCCAAGGTCTTCCCCCCCCCCCCCCACCCCC	60.87	412
MIR4733	17	31094426	31095425	CCCCCTCCCCCCAGCCTCCTTGCCAACGCCCCCTTTCCCTCTCCCCCTCCCGCTCGGCGCTGACCCCCCATCCCCACCCCC	60.87	413
TRPM2	21	44349163	44350162	GGGGGAGGGGGCTGCAGAGGTCTTCGAGAAGGGGGTCCTGGGTCTCGTTGTGGGGGGGG	60.87	414
BMP8B	1	39788862	39789861	GGGGGGGCGGGGGCGGGGCCTGGGAGCGGGGGCCACGCCTGCGAGGGGG	60.87	415
NOG	17	56592699	56593698	GGGGGTGGAGAAGTGGGGGGTGGGGTTGTTGTTTGGGGGGAAGAAGGGGGAGGGGG	60.61	416
RANBP3	19	5978143	5979142	GGGGGGTAAATCAGAGCTCATCGGCGGGGGCGGGGGTCGTGGTCAGGCTCTAGGAGGGGG	60.61	417
AC114730.2	2	241843380	241844379	GGGGGCGGCGCTGCAGGGTCGGGGGCGCGGGGGCGCGGGGGCGCGGGGG	60.56	418
RN7SL546P	14	102583178	102584177	GGGGGTGGGGGTGGGGGGTGGTGGGGG	60.53	419
AC017019.1	Y	9664229	9665228	GGGGGGAGGGGGGAGGGGGGGAGGGGGG	60.53	420
FAM150B	2	288852	289851	CCCCCCGTCTCCCGCTCCCCCTCAACTCCGCCCCCGCCCCCGCCCGGTCCCCCC	60.53	421
AC104534.3	19	38840179	38841178	GGGGGAGGGGGTGGGGGGTGCCCATACTGGGGG	60.47	422
ALX3	1	110070701	110071700	GGGGGGCGGGGGGCGGGGGGCAGGGGAGGGGG	60.47	423
TSPYL2	X	53081367	53082366	GGGGGAAGGGGCAGGACCCGGGATGGGGGAAGGGGCAGAAGCGGGGGTGGGGGAAGGAGCTGAAGGAAATGGGGG	60.44	424
AGPAT3	21	43864186	43865185	CCCCCGTGTCCCCTCCGCGTCCCCCCCCGCGCCGTCCCCCCGCGTCCCCCCCCGCGTCCTTCCCCCGTGTCCCCCC	60.42	425
AC073343.13	7	6662974	6663973	GGGGGTGATGGGGGAGGGGGTGGAGATTTAGAGGTGGGGGG	60.42	426
COTL1	16	84618078	84619077	CCCCCACTGCTTGCTTCCCCCACTTCCCTGACCTCTTCCCCCTCCCTTCCCTCCCCCTCTTTCTCCCTTCCCCC	60.42	427
SP9	2	174333946	174334945	CCCCCCCCCCGCAAGTATCCCCCACTACTCCCCCCCACCCCC	60.38	428
CPTP	1	1323756	1324755	GGGGGCTGGGGGACTGGGGGGCTGAGGGGCTGGGGGGCTGAGGGTCTGGGGGTCTGCTGGGCTGAGGGGCTGGGGGACTGAGGGGCTGGGGG	60.38	429
DCHS1	11	6655855	6656854	CCCCCGCCCCTCCGCCCGCGGAGCCCCCGCCCCCGCCCCC	60.38	430
ZBTB7B	1	155001630	155002629	GGGGGGGGGGAGTTTGGGGGGTGGCAGGGGCGGGGGGGAGGGAGGGGAGGGATGGGGGG	60.27	431
ELF2	4	139177219	139178218	CCCCCGCCCCGCTCCGCTCCCGGCCCCCTCCCTCCACCCCCCGCCTCGGCTCCTCCCCC	60.24	432
RND2	17	43024241	43025240	GGGGGTGGGAGGAGGGGCCGGGGGGGCGCGGCCGCCTGGCTGGGGGCGGGGCGGAGGGGGGGCCGCGGACCCGGGGCGGGGG	60.17	433
VAT1	17	43025124	43026123	GGGGGTGGGAGGAGGGGCCGGGGGGGCGCGGCCGCCTGGCTGGGGGCGGGGCGGAGGGGGGGCCGCGGACCCGGGGCGGGGG	60.17	434
RP11-17M16.2	18	76490652	76491651	GGGGGCGGGGGGCGGGGGGCGGGCGCCGGGGGGCGGGGG	60	435
WDR45B	17	82648554	82649553	CCCCCCCGGGACACAGAAACCCCCCCAGGACACAGCCCCCCTAAGACACACAGCCTCCCCC	60	436
CCDC64	12	119988869	119989868	GGGGGCGGGGGAGGGGGCGCGTGCCGCCGGCGCGGGGG	60	437
KCNJ5	11	128890356	128891355	GGGGGCAGGGGGAGCCGGGGGAGGGCGCAGGGGG	60	438
Y_RNA	20	49920695	49921694	CCCCCTCCCCCTCCCCCTCCCCCTCCCCC	60	439
ZBTB18	1	244047939	244048938	CCCCCTCCCCTCCCCCTCCCCCGCGCCCGCCCCCCCCCCCGCCCCCGCCCCCCC	60	440
RP11-38L15.2	10	46596918	46597917	CCCCCCCCTTCCCCCTCCCCCCAACCCCCCCCTTCCCCCCTCCCCCCCAACCCCCCCCC	60	441
HRASLS	3	193240125	193241124	GGGGGCGGAGGGGGGCCGGGGGAGGGGGG	60	442
UBTD2	5	172284072	172285071	GGGGGAGGGGACGCAGGGGGCTAGGTGGGGGGGAGGACTGCAGGCGGGGGG	60	443
MRPL14	6	44127458	44128457	CCCCCTCCCCCTCCCTCCCCCCGCTCCCCCTCCCCC	60	444
AL022328.1	22	50274674	50275673	CCCCCAGCACCCCTCCCAGCCCCCTCCACTTCCCCCCAGCACCCCTCCCAGCCCCCTCCACCTCCCCCCAGCACCCCTCCCAGCCCCCTCCACCTCCCCC	59.84	445
IGF2-AS	11	2139501	2140500	GGGGGGTGCGGGCTGGGGGCCCTGGGGCCCCGGGGGGGGGGCGGCGGTCGCAGGGGCTGGGGGGTGCGCGGGGGCGAGGCTGGGTGGGGGG	59.83	446
PXN	12	120265772	120266771	CCCCCGGCCGCCCCCTGTGCACGCGCGCCCCGCCCCCGGACACCCCCGCGAGCTTGCTGGCCCCGCCCCCTGCGCACGCTGGTCCCGCCCCC	59.8	447
RP11-464F9.20	10	73624996	73625995	CCCCCTCCCCAGGCGCTACCCCCTGCACCCCGTTACCCCCGGCCCCGCCCCCC	59.72	448
RP3-400B16.1	6	4135072	4136071	GGGGGCTCGGGGGTGGGGCCGGGTCTCGGGGGTGGAGCCGGGACTCGGGGG	59.68	449
ECI2	6	4135598	4136597	GGGGGCTCGGGGGTGGGGCCGGGTCTCGGGGGTGGAGCCGGGACTCGGGGG	59.68	450

SRRM3	7	76200900	76201899	GGGGGGGAGGGGGGCGAAGGGGGCGGGGAGGAGGGGG	59.65	451
RP11-154H23.3	3	71583943	71584942	GGGGGGGGGGGTTGGGGGGAGGGATCCAGGGAGGGGG	59.57	452
FOXP1	3	71583990	71584989	GGGGGGGGGGGTTGGGGGGAGGGATCCAGGGAGGGGG	59.57	453
CTD-2589H19.6	5	674826	675825	GGGGGTGCAGTGTGGCTGGGGGTGCAGTGCAGGGGGTACAGTGTGGCCGGGGGTGCAGCCCAGGGGG	59.49	454
U1	1	148521601	148522600	GGGGGGGGGGGGTGGGGGGGGGGGG	59.46	455
CAMK2G	10	73874592	73875591	CCCCCAGGCCCCGCCAGTCCCCCCCCCCGCCCGGCCCCCGGCCCGCCCCC	59.42	456
CASKIN1	16	2196526	2197525	GGGGGCGGGGAGCGCGTCACGGGCGGGGGGCGGCCCGGGGGCGGGGGCGCGCGGCGGGCGCGGGGCGGGGG	59.41	457
HRK	12	116881442	116882441	GGGGGGCGGGGAGGGGGGTTACATCATCCGGCCCCCGGGGGTTGGGGGGGGG	59.38	458
RP13-516M14.1	17	82292716	82293715	GGGGGGAGGGAGGGGACGGGGGGAGGGAGGGGAGGGTGGGGGAAGGAGGGGAGGGGAGGGTGGGGG	59.34	459
VENTX	10	133236404	133237403	CCCCCCACCCCTCCTGGCTTTACCCCCGAGGCCACGTCCAAAGCCCCCAGCGCCCCACCCCCGCCAGCCTCCCCC	59.3	460
AC034243.1	5	138753310	138754309	GGGGGCCGCGGGCGGGGGGCGGGCCGGGGGCGGGGG	59.26	461
CTD-2270P14.1	16	2673445	2674444	CCCCCCGCGATGCCCCCGCCCCCCGCGACGCCCCGCCCCC	59.26	462
DPYSL4	10	132185900	132186899	GGGGGCGGGGCCTGCCGGGGGCGGAGCCGGGGTGGGGGCGGGGCCCAGCGGGGG	59.21	463
RP11-140A10.3	10	132185963	132186962	GGGGGCGGGGCCTGCCGGGGGCGGAGCCGGGGTGGGGGCGGGGCCCAGCGGGGG	59.21	464
RP11-110G21.1	8	51898325	51899324	CCCCCGACCTGCCCGCCCTTAAGTCCCCCTGCCCCCGACTCTTCGGCTGCCGTCCCCC	59.21	465
LINC00685	X	319990	320989	GGGGGCTTGAGTCCAGCCCTGGGGGGCCCGGATGCCGCCCGGGGGTTCTCCTCGTCTGCCTTGGGGG	59.21	466
RP11-174M13.2	12	129698228	129699227	GGGGGAGGGGGAGGGGAGGGGAGGGGGAGGAGATGGGGAGGGGAGGGGG	59.15	467
OR7E156P	13	63741341	63742340	GGGGGGTGCCTCCCCACCCTGCGATGGGGGTTCCAAGTGTCAGGGGGGAAGAGGGGG	59.09	468
AC008278.3	2	15680485	15681484	GGGGGGAGAGGGAGGGAGGGGGAGGGGGAGGGGGAGGGGAGGGGGGAGGGGGAGGGGAGGGGGGAGGGGGGAGGGGGAGGGGGAGGGGGAGGGGGGGGAGGGGGGGAGGGGAGGGGGAGGGGGGAGGGGGAGGGGGTGGGGGAGGGGGGAGGGGAGGGGGTGGGGGAGGGGG	59	469
AC069368.3	15	64840883	64841882	CCCCCGGCCCCACCGCCCCCTCCCCGGCCCCGCCCCGGCCCCCGGAAGCCTCACCCCGCCCCCGGCCCCGCCCCCCCGAAGTTTCTTGGGCCCCC	58.88	470
PLEKHO2	15	64840883	64841882	CCCCCGGCCCCACCGCCCCCTCCCCGGCCCCGCCCCGGCCCCCGGAAGCCTCACCCCGCCCCCGGCCCCGCCCCCCCGAAGTTTCTTGGGCCCCC	58.88	471
MRPL41	9	137550199	137551198	GGGGGTACCGCTAGGGAGGGGGTGCAGCTGGGGAGGGGGTGGATCTGGGGTGTGGGGGTGCCGCTGGGGAGGGGGTGCAGCTGGGGAGGGGGTGGATCTGGGGCGTGGGGGTGCCGCTGGGGAGGGGGTGCAGCTGGGGAGGGGGTGGATCTGGGGCGTGGGGGTGCCGCTGGGGAGGGGG	58.87	472
PNRC1	6	89079751	89080750	CCCCCGTTCTTCCTGCCGCCCTCCCCCTGCCCCCTCCCTTCCCCACCCCC	58.82	473
TBL1XR1	3	177197397	177198396	GGGGGGAGGGGCGGGGGCGCACGCGGCCGGCGGCGGGGGGAGGCGGCGCGCGGGGG	58.75	474
RP11-435M3.2	10	25933711	25934710	GGGGGAAACGGGGGCGGGGGTGAACTCGAGGGGG	58.7	475
PKDCC	2	42047020	42048019	GGGGGAGGGGGCGGGGGAGGGGGCGGGGGCGGGGGCGGGGGG	58.62	476
RP11-272P10.2	9	13279188	13280187	CCCCCACCCCCACCCCCACCCCCATCCCCGCCCCCACCCCCACCCCC	58.57	477
SHANK3	22	50673415	50674414	CCCCCGCCTCCGGCGCAGCCCCCTCGCCACCCCCGCTTCCCTCCCGTCTCAGGCCCCCTCCCCCCGCCGCCCCCGCCCCC	58.49	478
FARS2	6	5260044	5261043	CCCCCGGTCCCCGGCCCCTGGCCCCCCGCCCCCGGCCCCC	58.49	479
FAM160A1	4	151408216	151409215	CCCCCGCCCCTGCCCCGCCCCGGCCCCCGCCCCCGCCCCCGCCCCCCCC	58.33	480
MCF2L	13	112893378	112894377	CCCCCTCCGCTCTGGGCTCTGGCCCCCGCCCCGCCCCCGCCCCCGGGCCCCTCTCCGCCCCTCCCCCTCCGGCCCGCCCCC	58.33	481
CCDC40	17	80035632	80036631	CCCCCGGCCCGCCCCCGGGCCCGCCCCCGGCCCGGCCCCTGCCTGCCCCC	58.33	482
RP11-610P16.1	4	151408836	151409835	CCCCCGCCCCTGCCCCGCCCCGGCCCCCGCCCCCGCCCCCGCCCCCCCC	58.33	483
TBC1D16	17	80035849	80036848	CCCCCGGCCCGCCCCCGGGCCCGCCCCCGGCCCGGCCCCTGCCTGCCCCC	58.33	484
CTD-3193O13.11	19	7869561	7870560	GGGGGCGTGGCCGAAGGAGACACTGGGGGTGAGGCAGGGGGAGAGAAAGGGGCCTGCACTGGGGGTGAGGGAGGGGG	58.25	485
CTD-3193O13.11	19	7869561	7870560	GGGGGTGGGGCCTGTGGAGGGGGCGTGGCTGAAGGAGACATTGGGGGAGAGGCAGGGGG	58.23	486
PLEKHA7	11	17014444	17015443	GGGGGGGAGGGGGCGGGGGCGCGGGCAGGGGCGGAGGCGGGGGCGGTGGCGGGGG	58.23	487

FST	5	53479409	53480408	CCCCCGACGCGTAGCCCCCAACCCCCGCCC CGGTCGCCTTCCCCC	58.18	488
PHF2	9	93575407	93576406	CCCCCCACCCGCCTTCTCCCCCGGCAGGGT CCCCCCGGCCGTCTTCTCCCCCGGCAGGGT CCCCCAACCCACCTTCTCCCCC	58.16	489
RP11-10K17.6	16	5097739	5098738	GGGGGCGGGCCCAGGGCAGTGAGCGGGGG CAGAGAGGGGGCGGGGCCCGGGGGTCCCC CGGGGGTCAGGAGGAACGTCGTGGGGG	58.12	490
EEF2KMT	16	5097809	5098808	GGGGGCGGGCCCAGGGCAGTGAGCGGGGG CAGAGAGGGGGCGGGGCCCGGGGGTCCCC CGGGGGTCAGGAGGAACGTCGTGGGGG	58.12	491
TRABD2B	1	47996896	47997895	GGGGGCTGGGGCCGCTGGGGCGGGGGAGG GGAGGACCTGGCGCGGGGGGCGGGGGGG	58.02	492
C1orf159	1	1116362	1117361	CCCCCCAAAACCCCCAGATCCCCCAACCCC CCATATACCCCCAACCCCTCAGACCCCCCAA CCCCCCAGACCCCCCAACCTCCCAGGCCCC CAGACCCCCC	57.98	493
RNASEH2C	11	65720948	65721947	CCCCCTCTCCTTCCCCCTCCCCCCTCCTTCC CCCTCCCCCCTCCCCC	57.97	494
VN1R91P	19	23289849	23290848	GGGGGAGGGGGAGGGGGAGGGGG	57.89	495
RNU6-96P	7	66394191	66395190	GGGGGAGGGGGAGGGGGAGGGGG	57.89	496
SHANK3	22	50673415	50674414	CCCCCCGCACCGAGGCCTAGGACTCCCCCC CCCAACCCCGTCACAGCCCCCCAGACCCCC GCCCCGTGGCTCGGCCCCC	57.89	497
RP11-114H24.6	15	77992405	77993404	CCCCCCCCGCCCCCCCACCCCCCCCAATCC CCCGCCCCCCCACTCCCCCCACTCCCCCCC ACTCCCCCCCCACTCCACTCCCCCCCACTCC CCCCCCACTCCCCCCCACTCCCCCCACTCCC CCCACTCCCCCC	57.8	498
RP11-816J6.3	3	169768649	169769648	CCCCCCGACAACACCCCCAGTCCCCCCACC CCCAGCCACCCCAGCCCCACCCCCC	57.75	499
AL022328.1	22	50274674	50275673	CCCCCTCCACCTCCCCCCAGCACCCCTCCCA GCCCCCTCCACCTCCCCC	57.75	500
C20orf194	20	3407626	3408625	GGGGGCGCGGCGCGCGGGCCCGGCGGGGG CGGGGGCGGGGGCGTGGAGGGGG	57.69	501
CHD3	17	7883806	7884805	GGGGGTTTCGGAGGGGGGTGGCCCGGGGG TTTGGGGGGCTGGGGG	57.63	502
RHOXF2B	X	120077706	120078705	GGGGGCAGTAGGACGGGGTGGGGGGAGTA GGGCGGGGGAGGGAGTAGGGCGGGGGGGC GTAGGGTGGAGGGGGGGAGTAGGGCGGGG GGG	57.63	503
NOVA2	19	45973547	45974546	GGGGGGGGGCGGGGGGCGGGGGGCGGGG GAGGGGGAGGGGAGGGGAGGTGGGAGGGG GAGGGGGCGGTGGAGGGATGGGGGAGGGG GAGCCCCGGAGAGGGAAGGCGGGGGTAGG GGGGAGCCGGTTCTCGGGGGG	57.61	504
RP11-100M12.3	15	27541992	27542991	GGGGGCTGCACTTACACAAGGGGGTGGGG GTGGGGGG	57.45	505
TMSB10	2	84904625	84905624	CCCCCCTCCCCCTACACCCCCCCGCCCCCC	57.45	506
RP4-647J21.1	7	44960910	44961909	GGGGGCTGAGCACTGAGGGGGTTGAGGGG GCTGGGGGGCTGAGGGCTGAGGGGG	57.35	507
TTLL12	22	43187134	43188133	CCCCCGCCCCGTCCTCCTCCCAGCCCCCCG CCCACCGCCCCGTCCTCCCCCCAGCCCCCC GCCCCTGCCGCGCCCCCTGCCCCCGCCCCC TGCCCCCC	57.35	508
RP11-168O16.1	1	201022949	201023948	GGGGGTAGGGGGCGGGATTGAAGGGTGGG GGTGGGGGTAGGGGTGGGAGTAGGGGTGG GGG	57.32	509
KIF21B	1	201023701	201024700	GGGGGTAGGGGGCGGGATTGAAGGGTGGG GGTGGGGGTAGGGGTGGGAGTAGGGGTGG GGG	57.32	510
AGRN	1	1019123	1020122	GGGGGAGGGAGGAGGGAGGGGGCGGGAG GGGGAAGGAAGGAGGAGGGGGAGGAGGAG GGCGCGGGGAGGGGGTAGGGGGGCGGGGA GGAGAGAGGGGG	57.25	511
LAMB2	3	49133119	49134118	GGGGGCGGCGGGGGTGGGGGTTGGGTGAG GGTTGGGTGGGGG	57.14	512
NHLH1	1	160366067	160367066	CCCCCGCTCCCCGCCCCCACCCAGGGCCAG CCCCCGCCCCC	57.14	513
BSG	19	570309	571308	GGGGGTGGGTTGGGAGGGGCGGAGGGGGT GGGTTGGGAGGGGCGGAGGGGGTGGGTTG GGAGGGGCGGAGGGGG	57.14	514
ABTB2	11	34358009	34359008	GGGGGTGGGGATGGGGGGGGATGGGGGGG ATGGGGGGGTGGGGG	57.14	515
TBCD	17	82751064	82752063	CCCCCCGTCCCCGTCCCCCGTCCCCGTCCC CCCTGTCCCTGTCCCCCTTCCCCC	56.96	516
MYL6	12	56157161	56158160	CCCCCCACGACCCCCGCCCCCCCACCTCCC CCCACGACCCCCCCC	56.94	517
PANX2	22	50169731	50170730	CCCCCCGCCCCGCCCGCGCCCCCCGCCGGG CCGCGCCCCCCGCCCCCC	56.94	518
PNPLA7	9	137550535	137551534	GGGGGTACCGCTAGGGAGGGGGTGCAGCT GGGGAGGGGGTGGATCTGGGGTGTGGGGG TGCCGCTGGGGAGGGGGTGCAGCTGGGGA GGGGGTGGATCTGGGGCGTGGGGGTGCCG CTGGGGAGGGGGTGCAGCTGGGGAGGGGG TGGATCTGGGGCGTGGGGGTGCCGCTGGG GAGGGGG	56.87	519
SLC12A7	5	1112036	1113035	CCCCCTCCCGCACCCCCTCCTGTTTGCCCCC CATTTCCTGCTCCAGCCCCCTACCCCTCCTT CTTCCCCCTCCTGTTCCCCCC	56.76	520
RPL7P14	2	203039108	203040107	CCCCCTGGCCCTCCCCTCCCCCCCTCCCCTC CCCCTCCCCCTCCTTCCCCC	56.76	521

AC073343.13	7	6662974	6663973	GGGGGGTGAAGGGGAGGGGGAGGCGATGT GGAGGTAGGGGGTGAAGGGAAGGGGGAGA GGGAAGGGATGTGGGGG	56.73	522
AC108004.2	17	191588	192587	CCCCCGCGACAGTCGCCCGTTCCCCCGCCC CCGCGACAGTCGCCCGTTCCCCCGCCCCCG CGACAGTCGCCCGTTCCCCCGCCCCCGCGA CAGTCGCCCGTTCCCCCCCCCCGCGACAGT CGCCCGTTCCCCCGCCCCCGTGACAGTCGC CCGTTCCCCC	56.73	523
CSH2	17	63881864	63882863	GGGGGCGGGGGCGGGGCGTGGGGATGGGG GTGGGGGTGGGGG	56.67	524
GH2	17	63881936	63882935	GGGGGCGGGGGCGGGGCGTGGGGATGGGG GTGGGGGTGGGGG	56.67	525
AP1G1	16	71809202	71810201	GGGGGTAGGGGGGTGGGGGGTAGGGGGGT GGGGGGGTAGGGGGGG	56.67	526
RP1-63G5.8	22	37370684	37371683	GGGGGTGTGAGGAGGGGTTGGGGGGCAGG TTTTGGGGGCCGGGGG	56.45	527
SOX1	13	112066647	112067646	CCCCCTGCAGGCCCCCCTGCGCCTCCCCCC CCCCGCCACTGGCGCCTGGCTTCCCCC	56.41	528
PKD1P6	16	15131602	15132601	GGGGGCAGGCAAAAAGGGGGAGCCGGAGG GTGGGGGCTGGGAGAAAGGGGGAACCTGA GGGGG	56.32	529
RNU7-140P	19	10551093	10552092	CCCCCTCCTCACTCCTCCCACCCCCCACCGC ACCCCCACTCACACCTGCACCCCC	56.25	530
MIR1238	19	10551122	10552121	CCCCCTCCTCACTCCTCCCACCCCCCACCGC ACCCCCACTCACACCTGCACCCCC	56.25	531
PAX1	20	21704659	21705658	GGGGGGGTGGTGGAAGAGGGCCGAGGGGG GAGGATAGAAGGAGGGGGTAGAGTTTCAGG GCGGGGAGGGGGG	56.25	532
SUN1	7	815615	816614	CCCCCAGGTGCAGACCCCTCCCCCAGGTGC AGACCCCTCCCCCAGGTGCAGATCCCTCCC CCAGATGCGGGCCCCTCCCCCAGATGCAAA CCCCCCCC	56.19	533
RP11-358D14.3	20	63421269	63422268	GGGGGAGCCCGAACAGGCACGGGGGAGCC CGCACAGGCATGAGGGGGCAGCCGGTGTTG GGGGCTCCTCCTCCAGCACGGGGG	56.19	534
RP11-20I20.2	4	1112639	1113638	CCCCCCCATGCCCCCCCCGCTCCATGCGCC GTTTCTCCCCCTCGGCCCCCCACGCGCCCC C	56.18	535
UNC119B	12	120709435	120710434	CCCCCCGGGCGCCCGCCCCCGCCCCCGCCC CGCCCCC	56.14	536
MGAT3	22	39456344	39457343	GGGGGAGGGGAGGGGGTTGCGGAGGGGGC GGGGTGGGGG	56.14	537
KDM1A	1	23018448	23019447	GGGGGACATGAAGGATGGGGGACATGGGG GATGGGGGACATGGGGG	56.14	538
RP11-40H20.4	1	27234353	27235352	CCCCCCAGCCCCCTCCCCGGGATCCGCGCC CCCCTTCCCGGGAGAGGGGCCGCCCCCCCC	56.1	539
FAM86C2P	11	67805337	67806336	GGGGGCGGGCCCAGGGCAGTGGGCGGGGG CAGAGAGGGGGCGGGGCCTGGGGG	56.1	540
AC004540.5	7	26397593	26398592	GGGGGTGAGGGAGGGGGTGAGGGAGGGGG TGAGGGAGGGGGTGAGGGTGGGGGCGGGG GG	56.1	541
MMD2	7	4959214	4960213	GGGGGCGTTGGTGGGGAGTGGGGGACTTG GGGGTGGGGGAGCGGAAAGTGGGGGGTTG GGGGGAGTGGGGGACCTGGGGGTGGAGGG AGTGGAAAGTGGGGGAAGTTGGGGGGAGT GGGGGACCTGGGGGTGGAGGGAGTGGGAA ATGGGGGTGTTGGGGGGTGGAGGGAGTGG AAAGTGGGGGGTTGGGGGGAGTGGGGGAC CTGGGGGTGGAGGAAGTGGGGGG	56.05	542
GPR173	X	53048091	53049090	GGGGGGGGGTGGGGGGTGGGGGGGGTAGG GGG	56	543
H6PD	1	9233775	9234774	CCCCCGCCCCGCCCCCGCCCCCGCCGCGCG CGCCCCCGCCGCGCGCGCCCCC	56	544
CHRD	3	184379073	184380072	CCCCCTCTTCGCGCCCCCTGGTCCCTCCCCT CCCCCTCCCCCGCTCCCCC	56	545
AC093642.5	2	242087633	242088632	CCCCCCCGCCCCCGCCCCCCCGCGCCCCCG CGCCCCCGCGCCCCCGCGCCCCC	55.95	546
RP11-341N2.1	2	242088458	242089457	CCCCCCCGCCCCCGCCCCCCCGCGCCCCCG CGCCCCCGCGCCCCCGCGCCCCC	55.95	547
GAS6-AS1	13	113814630	113815629	GGGGGTGCTGGGGGGTTTGGGGGTGCTGT GGCCCCGGGGCAGGGGGTGCTGGGGGG	55.94	548
CTCFL	20	57525653	57526652	GGGGGCGCTGGGATACCTTGTGGGGGAATG GATAAGGAGGGGAGGGGGTCCTGCAGGGG G	55.84	549
EIF4BP4	3	42339588	42340587	GGGGGGGAGGGGGAGTAGGAGGAGGGGGA GGAGGAGGAGGAGGGGGAGGAGGGGAAGG GGGAGGAGGGGGGAGGAGGAGCAGGGAGA GGGGG	55.81	550
DRP2	X	101218769	101219768	CCCCCTCTTCCCCTCTCCCCCTCTCCCCCTC TCTCCCTCTTCCCCCTCTCCTTGTCTCCTTC TCTCCCCC	55.67	551
TCL1B	14	95685417	95686416	CCCCCTTCCTCCTCCTCCCCCTCCTCCCCCG ACTGGCCCCGCCCCCACTGCCGGCCCCGCC CCC	55.67	552
TLK2P1	17	34039873	34040872	GGGGGGGTCCTGGGGGGTGGGGGAAGCGC CGACAGTGGGGGAGGAGGGGGAGGGGACG AGGAAGGTGGGGGGGG	55.66	553

RP11-19N8.3	16	33058936	33059935	CCCCCCCGTCGCCGCGGCTTTTTGCCCCCC CTGCCTCCGCGGCTTTTTGCCCCCCCTGCCT CCGCGGCTTTTTGCCCCCCCTGCCTCCGCG GCTTTTTGCCCCCCCCTGCCTCCGCGGCTTT TTGCCCCCCCCCTGCCTCCGCGGCTTTTTGCC CCCCCCTGCCTCCGCGGCTTTTTGCCCCCCCC CTGCCTCCGCGGCTTTTTGCCCCCCCCCTGCC TCCGCGGCTTTTTGCCCCCCCCTGCCGCCG CGGCTTTTTCCCCCC	55.59	554
TTLL12	22	43187134	43188133	CCCCCCCCCCCGCCCCGCCCTCCTCCCAGC CCCCCCCCGCCCCGCCCTCCTCCCAGCCCC CCCCCGCCCCGCCCTCCTCCCAGCCCCCCC CGCCTCCGCCCCCC	55.56	555
TYW3	1	74732152	74733151	GGGGGGGGGGGGGGGGGGTGCAGCGTGGG GCTGGGGGACAGGGAGTGCGGGGGG	55.41	556
PVRL3-AS1	3	111071554	111072553	GGGGGAGCCGGGGGGCGGGCGGGCGAGCG GGCCGGGGGGAGGGTGGGGG	55.41	557
IRF2BP2	1	234609526	234610525	GGGGGGCGGGGAGGCCGGGGGGGCAGGG GGCGGGGGG	55.36	558
CPLX1	4	826199	827198	GGGGGCGGGGGTGGACCCGGGGGCGGGGG CGGAGGCGGACCCGGGGGCGGGGGGCGGG GGGTGGACCCGGGGCGGGGGCGGGGGGTG GACCCGGGGGCGGGGGCGGGGGGCGGGGG CGGGGGGCGGACCCGGGGCGGGGGCGGGG GGTGGACCCGGGGCGGGGGCGGGGGTGGA CCCGGGGGCGGGGGCGGGGGGCGGGGGCG GGGGCGGGGGGTGGACCCGGGGCGGGGGC GGGGGTGGACCCGGGGGCGGGGGCGGGGG	55.29	559
SLC45A4	8	141308306	141309305	CCCCCGAACGGCCGCCCCGCCCCCGCCGCG CCCCTGCCCCGCCCCCGCGCGCACCGTGGC TCCCGCCCCCGTCTCGCCATGGCCCCGCCC CCAGCCATAATCCAGCCCCGCCCCCGCCCT CCTCCCGCTACGCCCCC	55.26	560
EPN1	19	55674226	55675225	GGGGGGCGGGGGTTGGGGGGTTGGGGGG	55.26	561
PIM1	6	37169203	37170202	CCCCCGACGCGCCCCCCAACACACAAACCC CCAGAATCCGCCCCC	55.17	562
NFATC1	18	79394856	79395855	CCCCCCTCCCCGTGCGCCCCCCCTCCCCGT GCGCCCCCCTCCCCGTGCGCCCCCCCTCCC CGGGCGCCCCCCTCCCCGGGCGCCCCCCCT CCCCGTGCGCCCCCCCCTCCCCGTGCGCCC CCCC	55.15	563
CDKN1C	11	2885882	2886881	CCCCCTCGCCCCATCCCCCCTGCAACACTCC CCCACCCCCCCGGCTCCCGCCCCCC	55.13	564
TKTL1	X	154294671	154295670	CCCCCTGACACGTGCGGCACCACCCCCTGA CACGTGCGGCACCACCCCCTGACACGTGCG GCACCACCCCCTGACACGTGCGGCACCACC CCC	55.12	565
RP11-83B20.4	12	124541503	124542502	GGGGGATGGGGGTGGATATGGAGGGGGAT GGGGGTGGATATGGAGGGGGATGGGGAGT GGAGATGGAGGGGGATGGGGAGTGGAGAT GGAGGGGGTGGGGAGTGGAGATGGAGGGG GTGGGGAGTGGAGAGCTGGAGGGGGAGGG GGTGGAGAGTTGGAGGGGG	55.12	566
TRAM2-AS1	6	52576307	52577306	CCCCCTTCCTCCGCCCGCCCCCGGCCGCCC GTAACCGCCCCCGGCCGCAGCCCCCGCCCA GCCCCC	55.06	567
CTD-2589H19.6	5	674826	675825	GGGGGAGCAGTGAGGGGGGCACAGTGTGG CCGGGGGAGCAGTGAGGGGGGTACAGTAT GGCTGGGGG	55.06	568
TRAM2	6	52576916	52577915	CCCCCTTCCTCCGCCCGCCCCCGGCCGCCC GTAACCGCCCCCGGCCGCAGCCCCCGCCCA GCCCCC	55.06	569
CBX7	22	39152675	39153674	CCCCCGAGCCCCGCCCCCATCGGAACCCCG CCCCCGGGGCCCGCCCCC	54.93	570
DLGAP3	1	34929586	34930585	GGGGGGGAGGGGGCCACGGTGGGGAAGCC GCAGGGGGAGGGGCTGAAGACGTGGGGG	54.88	571
FAM86GP	11	3422497	3423496	GGGGGCGGGCCCAGGGCAGTGCGCGGGGG CAGAGAGGGGGCGGGGCCTGGGGG	54.88	572
IGLON5	19	51310848	51311847	CCCCCCCAGCCCCCTCCCCCCGCCCCCGCC GCCCCCC	54.84	573
MYDGF	19	4670371	4671370	GGGGGCGGGACTTTGTCGAGGGGGCGGGA CTTTTGCCGAGGGGGCGGGGCCGCCGAGG GGGCGGGGCCGCCGAGGGGG	54.78	574
UBE2R2	9	33816567	33817566	CCCCCCGCGCAGCCCCCGCCGCCACCGCCG CGAGACCCCCGCACGCCGCTCTCCCCCC	54.76	575
TMEM253	14	21097937	21098936	CCCCCCCGCCCCCGGCCCGGCCCCCGCCCC CTCCCCGGTCCCCCGCCCCC	54.76	576
TSPYL2	X	53081367	53082366	GGGGGGGTGAGGGAAGGGGCAGAAGCGGG GGAGGAAGGGGCAGAAGCGGGGGGTGAAG GGGCAGAAGCGGGGGTGGGGTGGGGGAAG GGGCCGAAGCGGGGGATGGGGGAAGGGGC AGAAACGGGGGGTGGGGG	54.75	577
ENAH	1	225653143	225654142	GGGGGCGAGGGGGCGGTGCGGGGGGAGGG GAGGAAAGGGGAGGCGGGGGG	54.67	578
ASB1	2	238425742	238426741	CCCCCACTCCCGCCTCCAACCCCCACAGGG CCTGCCAGGTCTGCCCCCAGTGGGCCCCC	54.55	579
AC108004.2	17	191588	192587	CCCCCGCCCCCGCGACAGTCACCCGTTCCC CCCCCGCGACAGTCACCCGTTCCCCCCCCG CGACAGTCACCCGTTCCCCC	54.39	580
SPATA31A6	9	42182659	42183658	CCCCCTCCCCTGCCCCTCTCCCTCCCCCTCC CCCTCCTGCTGCACCTCCCCCTCCCTCTGCC CCTCCCCC	54.37	581

RBPMS2	15	64775588	64776587	GGGGGGCGGGCGCCGGGGGAGGGGGCGG GCGCCGGCGGCGGGGGG	54.17	582
LL22NC03-63E9.3	22	22558343	22559342	GGGGGAGGGGTGGGGATGGGGGAAGGGAG ACAGAAGGGGGGGAGGGGTGGGGATGGGG G	54.12	583
FAM201A	9	38619474	38620473	CCCCCGCAACCCGCGATCCCCCCCGCAACC CGCGATCCACCCCCAAATCCAAGATCCACC CCC	54.02	584
LRRC27	10	132331154	132332153	CCCCCTGCCACCCCCGCGCAGGCGCACCAC AACCCCCCCACCGCAAGCGCAACCCCCC	53.95	585
STK32C	10	132331848	132332847	CCCCCTGCCACCCCCGCGCAGGCGCACCAC AACCCCCCCACCGCAAGCGCAACCCCCC	53.95	586
CTD-2561B21.5	17	80971214	80972213	CCCCCTCCCCCGCCAGCCTGCACATCCTCC CCCATCGTCCATGCAGCCCCCC	53.95	587
URAHP	16	90047774	90048773	GGGGGGCGGGACCGCTGTGAGAGGGGGCG GGGGGACGGCGGGACGGGGGGATGGGGGG ATGGAGGGGATGGGGG	53.85	588
C11orf86	11	66974277	66975276	GGGGGAGGGGAGGGGAGAAGGAGGGGGGGA GGATGGGGGAGGAGGGAGGGGGGAGAGGA GGGGAGGGGGAGGGGAGGGGGAGGGGGG	53.79	589
HIC2	22	21416404	21417403	GGGGGTAAGTCCACCAGAGTTTGAGGGGGT GGGCGTCAGCCTGTCGGGGGGGGCCAGGG TCCCTGTCCCGGGGG	53.77	590
RP11-501C14.6	17	49013726	49014725	CCCCCTCAGTCTCAGCCCCCAGTCGCCGT CCCCCTCTGTCTCCCCCTCAGTCTCAGCCCC CCAGTCGCCGTCCCCC	53.7	591
RPUSD1	16	788398	789397	CCCCCCCCCCGGCCCGCCCCCGGCCCGCC CCC	53.7	592
TNFAIP8L1	19	4638518	4639517	CCCCCAGAGCCGCCTGACCCCGCCCCCAGG CTCCGCCCCTTCCCCCCACTTCCCCCCC	53.68	593
CTD-2545M3.8	19	50487639	50488638	GGGGGAGAGGGTGTTGGGGGAGAAGAGGG TGGGGGAGAGGGGG	53.62	594
RP11-395B7.4	7	101017609	101018608	CCCCCTGGGACTCCCTTCCTCCCCCTGGGA CCCCTTCCTTTCCCCCTGGGACTCCCTCCCT CCCCCTGGGACCCCTTCCCTTCCCCCTGGG ACTCCCTCCCTCCCCCTGGGACCCCTTCCCT TCCCCCTGGGATTCCCTCCCTCCCCCTGGG ACCCCTTCCCTTCCCCCTGGGACTCCCTCCC TCCCCCTGGGACCCCTTCCCTTCCCCCTGG GACTCCCTCCCCTCCCCCTGGGACCCCTTCC CTTCCCCCTGGGACTCCCTCCCTCCCCCTGG GACCCCTTCCCTTCCCCCGGGACTCCCTCCC TTCCCCCGGGACTCCCTCCTCCCCC	53.5	595
KDM2A	11	67118269	67119268	GGGGGGGGTGATGTGCTTTGTGGGGGGTC GGGGGTCACCACCCTCCCAGGGGGGTGGG GG	53.42	596
RP11-564A8.4	1	206906619	206907618	GGGGGGTGGGGGGGTGGGGGGGTGGGGG GGGGGTGGGGG	53.33	597
AL031721.1	16	1379424	1380423	CCCCCGTCGCACTGGCCACGCCCCCCGCGC TGCCCTCCCCCAACCTTCCCCCC	53.16	598
KLF6	10	3785282	3786281	CCCCCTCCGCCCCCGCCTGGCTCCCCCCAC TCCCCCTCCCCCC	52.94	599
RP11-561N12.5	7	64570438	64571437	GGGGGGAGAGGGGCAAGGGGGAGGGGGAA GGGGGAGGGGG	52.86	600
BHLHA15	7	98210427	98211426	GGGGGGAAGCTATCTCGGGGCGGTTGGGG GGCTGTCGGGGGAGGATGGGGAGAAGCGT GTGGGGGAGGCTGTCTGGGGGAACGAGGG TCGGGGGCTGTCGGGGGATGGTGGGGAGA GGCGGGCGGGGGTGGCTGTCTTGGGGGAC GAGGGTGGGGGGCAGTAGGGGGAGGATGG GGAGAGGCGGGCGGGGGAGGTTTTCTATCT TGGGGGCGGGGCGGAAGCTGGGGG	52.85	601
RP11-74E22.4	17	2749421	2750420	GGGGGGCTCGGGGGCGCCGAAGGGGGGTC GGGGG	52.73	602
ESX1	X	104254934	104255933	GGGGGCGAGGAGACGAGGGGGCGAGGGGG CGAGGGGGCGAGGGGG	52.7	603
URAHP	16	90047774	90048773	GGGGGCGGGGCCGCGGTGAGAGGGGGCGG GACCGCGGTGAGAGGGGGCGGGACCGCGG TGAGAGGGGGCGGGACCGCGGTGAGAGGG GG	52.63	604
CYB5R1	1	202967281	202968280	GGGGGGGCAGGAGGCGGGGCCTGGCGGGG GGGGGGGGGGGTGTCTGTCGGGGG	52.44	605
Y_RNA	10	68502152	68503151	GGGGGAGGGGAGGGGAAAGGAGGGGGAGG GGGAAGGGAAAGGAGGGGGAGGGGAGGGG AAAGGAGGGGGAGGGGAGGGGAAAGGAGG GGGAGAGGAGGGGAAAGGAGGGGGAGGGG AGGGGAAAGGAGGGGG	52.41	606
BRWD1-AS2	21	39312935	39313934	GGGGGGGGCGGGGGGCGGGGGGCCGGGG GCGGGCGGCGGGCGGCGGGCGGGGGG	52.22	607
RP11-20I20.2	4	1112639	1113638	CCCCCCACGTGCCCCCATGGCCCTCCCACC CAAGTCCCCCATCCCCCGGAGTTCCCTGAC CCCC	52.08	608

AC108004.2	17	191588	192587	CCCCCGCCCCCGCGACAGTCGCCCGTTCCC CCGCCCCCGCGACAGTCGCCCGTTCCCCCG CCCCCGCGACAGTCACCCGTTCCCCCGCCC CCGCGACAGTCACCCGTTCCCCCCCCACGA CAGTCACCCGTTCCCCCCCGCGACAGTCAC CCGTTCCCCCGCCCCCGCGACAGTCACCCG TTCCCCCGCCCCCGCGACAGTCACCCGTTC CCCCCCCCGCGACAGTCACCCGTTCCCCC	52	609
SAMD1	19	14091037	14092036	GGGGGCGGGGGAAAGCGGCGGGCGGGGGA GGGGGAGGGGAGTGCCGGGGGGAGTGGAG AGGGGG	51.89	610
AC005391.2	19	928818	929817	CCCCCGCGGCCCCCACGCTGCAGTGCGGCC GGGCCCCCTCCCCGCAGGGGCCGCCCCC	51.85	611
ANO9	11	442012	443011	CCCCCCACCCCGGGCCACGGCCCCCCACCC CGGGCCACGGCCTCCCCCACCCCGGGCCAC GGCCTCCCCCACCCCGGGCCACGGCCCCCC ACCCCGGGCCACGGCCCCCC	51,67	612
AL022341.3	16	649201	650200	CCCCCAGGAGAGCTGCCCCGCCCCCGAGGG CCCCCTCGCCCCC	51.56	613
CDKN2B-AS1	9	21993778	21994777	CCCCCCACCTTCACCCCCACCCCCACCCCAC CCCC	50.79	614
RP11-149I2.4	9	21994482	21995481	CCCCCCACCTTCACCCCCACCCCCACCCCAC CCCC	50.79	615
ADNP-AS1	20	50929984	50930983	GGGGGTGTGCGTGGGGGGGCTCCTTTTGGG GGTGGGGGG	50.77	616
HCN2	19	588893	589892	CCCCCGCCCCTCCCCCGCCCCTCCCCCGCC GCGAGGCTCCGCCCCCCCGCGGCTCCGCCCC GCCCCGCCCCCGCCTCCCCCC	50.72	617
ADD2	2	70768226	70769225	CCCCCACTACCGCCCCGCAGCGCCCCCCTC CTCCCCGCAGCGCCCCCGCCCGCGCCGCCG CAGACCCCCCTAGCCAGGCTTCCAGCCCCC TTCCACCCCC	50.64	618
LINC01436	21	36004338	36005337	GGGGGCGGGAGGGGGACGGGGGAGGCAGT GGGGGAGGCAGCAGTGGGGGAGGAGGTGG GGGGGGAGGAGGCGGGGGGGGAAGAGGCG GGGGGCAGGCGGGGGGGGGAGGAGGCGG GGGGAGGCGGGGGGGAGGAGGCGGGGGG AGGCGGGGGGGAGGCAGGGGGAGGCAGGA GAAAGGGGGGGAGGGGAGGCAGGGGGAG GCAGGGGAAAGGGGGGAGGGGAGGCAGGG GGAAGCAGGGGAAAGGGGGGAGGGGGAGA CAGGGGGAGGCAGGGGAAGGGAGGAGGGG GAGGCAGGGGGGGCAGGGGAAAGGGGGGA GGGAGAGGCAGGGGGAGGGGGGAGTCGGG GGGGCCGGGGGAAGGGGTGCGGGGGGGG	50.52	619
RUNX1	21	36004668	36005667	GGGGGCGGGAGGGGGACGGGGGAGGCAGT GGGGGAGGCAGCAGTGGGGGAGGAGGTGG GGGGGGAGGAGGCGGGGGGGGAAGAGGCG GGGGGCAGGCGGGGGGGGGAGGAGGCGG GGGGAGGCGGGGGGGAGGAGGCGGGGGG AGGCGGGGGGGAGGCAGGGGGAGGCAGGA GAAAGGGGGGGAGGGGAGGCAGGGGGAG GCAGGGGAAAGGGGGGAGGGGAGGCAGGG GGAAGCAGGGGAAAGGGGGGAGGGGGAGA CAGGGGGAGGCAGGGGAAGGGAGGAGGGG GAGGCAGGGGGGGCAGGGGAAAGGGGGGA GGGAGAGGCAGGGGGAGGGGGGAGTCGGG GGGGCCGGGGGAAGGGGTGCGGGGGGGG	50.52	620
PLXNB2	22	50307628	50308627	GGGGGAGAAGGAGGGGGAGGGGGAGGGG GGAGAGGGGGAGGGGGGAGAGGGGGAGG GGGAAGAGGGGGAGGGGGAGAGGGGGAGG GGGAAGAGGGGGAGGGGGAGGGGAGGGG G	49.48	621
MACF1	1	39080316	39081315	GGGGGGAGCGCGCATCGGGGGGCGGGGGG CGGCGTCGGCGCTGGGAGTGGGGG	49.33	622
APOC3	11	116828706	116829705	GGGGGAGGCAGCGGGGGGCACACAGGGTG GGGGCGGGTGGGGGG	49.33	623
PEA15	1	160204337	160205336	CCCCCGCCCCCTCCTGCCTCTCCATCCCTCC CCCTCCTCCCGCCATTCCCCGCCCCCC	49.04	624
AL121987.1	1	160204377	160205376	CCCCCGCCCCCTCCTGCCTCTCCATCCCTCC CCCTCCTCCCGCCATTCCCCGCCCCCC	49.04	625
UNKL	16	1414752	1415751	CCCCCGCCCCCGGATATCCCCGCGCGCCCC CGCGGCCACGCCCCCCGCGGCCCCGCCCCC	48.98	626
KIF25-AS1	6	167997078	167998077	GGGGGCGGGGGAATCTGCAGAGATGGGGG CGGGGGAATCTGCAGAGATGGGGGCGGGG GAATCTGCAGAGATGGGGGGCGGGGGAAT CTGCAGAGATGGGGGGTCGGGGGAATCTGC AGAGATGGGGGGCGGGGGAATCTGCAGAG ATGGGGGGTCGGGGGAATCTGCAGAGGTG GGGGCGGGGGAATCTGCAGAGATGGGGGC GGGGGAATCTGCAGAGATGGGGGCGGGGG AATCTGCAGAGATGGGGGGCGGGGGAATCT GCAGAGATGGGGGGCGGGGGAATCTGCAG AGATGGGGGGCGGGGGAATCTGCAGAGAT GGGGGGTCGGGGGAATCTGCAGAGATGGG GGGCGGGGGAATCTGCAGAGATGGGGGGC GGGGG	48.78	627
VANGL2	1	160399586	160400585		48.64	628
CMIP	16	81444170	81445169	CCCCCCGCGGGCAGCGCCCCCTCCCCTGCG GGCGCCCCCAGCCCCACACCCCCCC	48.28	629

BRD1	22	49827513	49828512	GGGGGGCCCCGAGGCAGGATGGGGGAGCA GCCGCCAGTCGGGGGCCCTGACGGGGGAA CGGCCAGGAGTGGGGGTCCCTGAGGCGGG GGGG	48.03	630
VDAC1P7	3	77317599	77318598	GGGGGCTGCTGGTTGCTGCTGGGGGGCTGC TGGGGGGGCTGCTGGGGGGCTACTAGGGG GCTGCTAGGGGTGCTGCTGGGGGGCTGCTA GGGGGGCTGCTGGGGGGGCTGCTGGGGGG CTGCTGGGGGGCTGCTAGGGGCGCTGCTGG GGGGCTGCTAGGGGGGCTGCTGGGGGGGC TGCTGGGGGGCTGCTGGGGGGCTGCTAGG GGGCTGCTGGGGGGCTGCTGGGGGGGCTG CTGGGGGGCTGCTGGGGGGCTGCTAGGGG CGCTGCTGGGGGGCTGCTAGGGGGG	46.95	631
ADGRA1	10	133069929	133070928	GGGGGCCATAGGGTGCAGAGTCAGGGGGC CCGGGGGTGCAGAGTCAGGGGGCGTAGGG TGCAGAGTCAGGGGGCCCTGGGGTGCAGTC AGGGGGCATAGGGTGCAGAGTCAGGGGGC CACAGGGTGCAGAGTCAGGGGGCCCTGGG GTGCAGAGTCAGGGGGCCATGGGGTACAGA GTCGGGGGGGACCTGGGTTGCAGAGTCGGG GGGCCCTGGGTTGCAGAGTCAGGGGG	46.43	632
AP001187.1	11	64925418	64926417	CCCCCTGCAAGCACCCCCACTAGCCCCTCC TCCCCCGGGCTGTCGCCTGTGCCCCC	46.07	633
AL356693.1	1	5815960	5816959	GGGGGATTACTGGTGTGTGGACATGGGGGA TTACCGGTGTGTGGACATGGGGGATTACCA GTGTGTGGATACGGGGGATTCCTGGTGTGT GAATACGGGGG	46.06	634
LA16c-OS12.2	16	186295	187294	GGGGGGAACAGCGACACGGGGGGAACAGC GACACGGGGGGAACAGCGACACGGGGGGA ACAGCGACACGGGGGGAACAGCGACACGG GGGGAACAGCGACACGGGGGGAACAGCGA CACGGGGGGGAACAGCGACACGGGGGGAAC AGCGACACGGGGGGAACAGCGACACGGGG GGAACAGCGACACGGGGGGAACAGCGACA CGGGGGGAACAGCGACACGGGGGGAACAG CGACACGGGGGGAACAGCGACACGGGGGG AACAGCGACACGGGGGGAACAGCGACACG GGGGGAACAGCGACACGGGGGGAACAGCG ACACGGGGGGAACAGCGACACGGGGGGAA CAGCGACACGGGGGGAACAGCGACACGGG GGGAACAGCGACACGGGGGGAACAGCGAC ACGGGGGGAACAGCGACACGGGGGGAACA GCGACACGGGGGGAACAGCGACACGGGGG GAACAGCGACACGGGGGGAACAGCGACAC GGGGGGAACAGCGACACGGGGGGAACAGC GACACGGGGGGAACAGCGACACGGGGGGA ACAGCGACACGGGGGGAACAGCGACACGG	43.44	635
GRTP1	13	113364149	113365148	GGGGGCGGGGGGGGGCAGGGCGGGGGGA GGGGCGGCTGGGGG	43	636
AC116609.1	2	731225	732224	GGGGGAGTCACGGCACGGGGGGAGTCACG GCACGGGGGGGAGTCACGGCACGGGGGGG GGGTCACGGCACAGGGGGAATCACGGCACA GGGGGAGTCACGGCACGGGGGGAGTCACG GTACGGGGGGAGTCACGGCACGGGGGGAG TCACGGCACGGGGGGAGTCACGGCACGGG GGGAGTCACGGCACGGGGGGGTCACGGCA CGGGGGAGTCACGGCACGGGGGGAGTCAC GGCACGGGGGGAATCACGGCACGGGGGGA GTCACGGCACAGGGGGAGTCACGGCCCGG GGGGGTCACGGCCCGGGGGGAGTCACGGC ACAGGGGGAGTCACGGCACAGGGGGGGTC ACGGCACGGGGGGAGTCACGGCACGGGGG GAGTCACGGCACGGGGGGGTCACGGCACG GGGGAGTCACGGCACGGGGGGAGTCACGG TACGGGGGGAGTCACGGCACGGGGGGAAT CACGGCACGGGGGGAGTCACGGCACAGGG GGAGTCACGGCCCGGGGGGGTCACGGCAC GGGGGGAGTCACGGCACGGGGGGGTCACG GCACGGGGGAGTCACGGCACGGGGGGAGT CACGGCACGGGGGGAATCACGGCACGGGG GGAGTCACGGCACAGGGGGAGTCACGGCC CGGGGGGGTCACGGCCCGGGGGGAGTCAC GGCACGGGGGGGAGTCACGGCACGGGGGGA GTCACGGCACGGGGGGAGTCACGGCACGG GGGGGTCACGGCACGGGGGAGTCACGGCA CGGGGGGAGTCACGGCACGGGGGGGGTCA CGGCACGGGGGGGAGTCACGGCACAGGGGG AGTCACGGCCCGGGGGGGTCACGGCCCGG GGGGGTCACGGCCCGGGGGGAGTCACGGC ACAGGGGGAGTCACGGCACAGGGGGGGTC	42.88	637
Out of a total of 5125 imotifs, there were 637 which overlapped gene promoters						

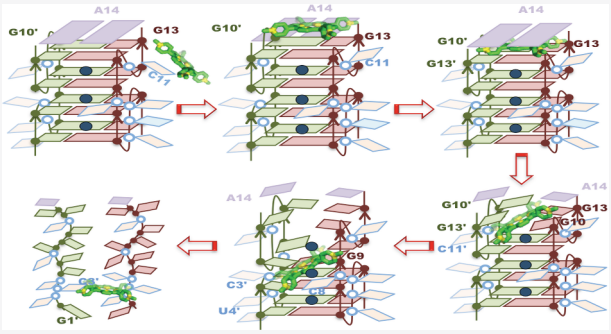
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Mechanistic Insights into the Ligand-Induced Unfolding of an RNA G-Quadruplex

Susanta Haldar,[#] Yashu Zhang,[#] Ying Xia,[#] Barira Islam, Sisi Liu, Francesco L. Gervasio, Adrian J. Mulholland, Zoë A. E. Waller, Dengguo Wei,* and Shozeb Haider*

Cite This: J. Am. Chem. Soc. 2022, 144, 935−950
Read Online
ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The cationic porphyrin TMPyP4 is a well-established DNA G-quadruplex (G4) binding ligand that can stabilize different topologies via multiple binding modes. However, TMPyP4 can have both a stabilizing and destabilizing effect on RNA G4 structures. The structural mechanisms that mediate RNA G4 unfolding remain unknown. Here, we report on the TMPyP4-induced RNA G4 unfolding mechanism studied by well-tempered metadynamics (WT-MetaD) with supporting biophysical experiments. The simulations predict a two-state mechanism of TMPyP4 interaction via a groove-bound and a top-face-bound conformation. The dynamics of TMPyP4 stacking on the top tetrad disrupts Hoogsteen H-bonds between guanine bases, resulting in the consecutive TMPyP4 intercalation from top-to-bottom G-tetrads. The results reveal a striking correlation between computational and experimental approaches and validate WT-MetaD simulations as a powerful tool for studying RNA G4−ligand interactions.

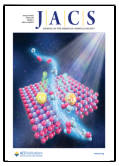


INTRODUCTION

Guanine-rich sequences in single-stranded DNA/RNA can self-associate, in a coplanar, cyclic arrangement called G-tetrads.[1,2] Each tetrad comprises four guanine bases, stabilized by eight Hoosteen hydrogen bonds reinforced by π and σ bonds.[3,4] The G-tetrads can stack on top of one another, held together by $\pi-\pi$ stacked nonbonded attractive interactions to form a G-quadruplex (G4).[5,6] The G4 structure is further stabilized by the presence of mono- or divalent cations, as they coordinate between two successive G-tetrads and shield the electrostatic repulsion between the carbonyl oxygens of guanines.[5] G4-DNA structures are polymorphic and can adopt varied topologies influenced by strand stoichiometry (one to four), polarity (parallel, antiparallel, hybrid), glycosidic conformation (syn/anti), intervening length of loops, and guanine stretches.[7] G4-RNA structures, however, are limited to predominantly parallel topology, primarily due to the anti-conformation of the glycosidic bonds and the presence of an additional 2′-OH group that contributes to enhanced hydrogen bond networks.[8,9]

G4s are widespread throughout the human genome. Extensive human genomic analyses have identified >500 000 DNA and >6000 RNA putative G-quadruplex forming sequences (PQFS).[10−13] They are found to colocalize with functional regions of the genome including specific sites such as telomeres and promoter regions of genes.[14−18] Their presence has important implications in regions involved in telomere maintenance and persistent DNA damage response in aging.[19] Human telomeric DNA d(TTAGGG) and the noncoding TERRA RNA r(UUAGGG) can both adopt characteristic G4 structures and are important participants in telomere biology and epigenetic regulation.[9,20] Exhaustive mapping of the genome-wide location of DNA replication origins have revealed that ∼90% of start sites contain the PQFS.[21] Such a high genomic distribution of PQFS in gene transcription start sites is indicative of a G4-mediated role in regulating the process of transcription.[22,23] Furthermore, significant enrichment of PQFS in the vicinity of somatic copy-number alteration breakpoints is an epigenetic determinant driving tissue-specific mutational landscapes in cancer.[24] New functional roles of G4-RNA have also been reported in the regulation of RNA expression in mitochondria,[25] in phase separation mechanisms leading to the formation of membraneless organelles,[26,27] and epitranscriptomics.[28] A high frequency of PQFS has also been reported in regions specifying the 5′-UTR of the encoded mRNAs, suggesting a G4-mediated role in regulating translation.[29,30] Moreover, the formation of G4-DNA in cells has been shown to be linked to replication and transcription and that of G4-RNA with translation.[31] Based on

Received: October 26, 2021
Published: January 6, 2022



https://doi.org/10.1021/jacs.1c11248
J. Am. Chem. Soc. 2022, 144, 935−950

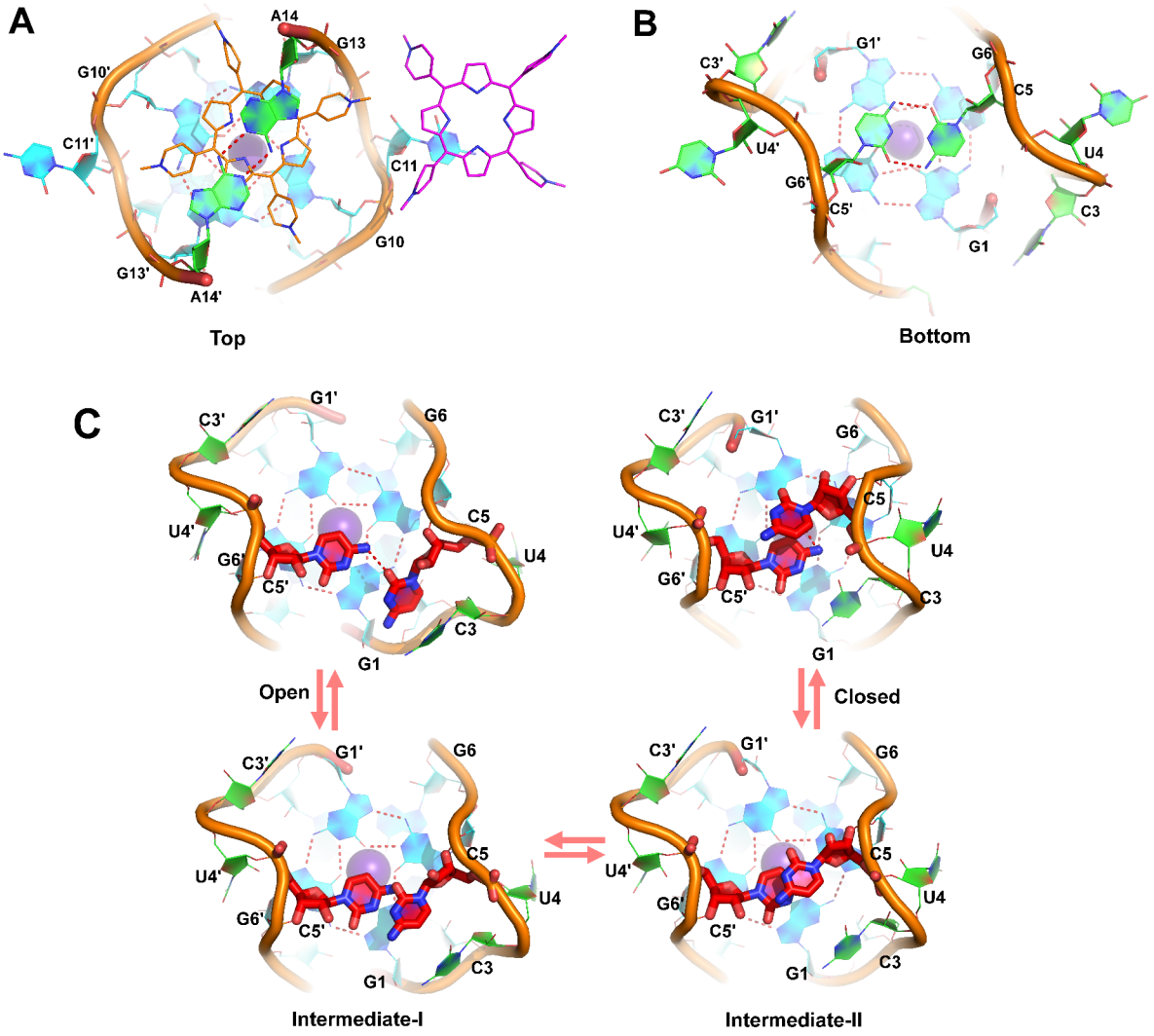


Figure 1. Crystal structure of the RNA-TMPyP4 complex (PDB id 6JJH). (a) Top-face view highlighting the two different bound states of TMPyP4. TMPyP4 shown in orange is sandwiched between A14, A14′, and guanine bases, G10, G10′, G13, and G13′, from the first G-tetrad. TMPyP4 shown in magenta is $\pi-\pi$ stacked with a C11 base. (b) Bottom-face view of the RNA−TMPyP4 complex highlighting a H-bond network between the capped bases C5 and C5′ in an open-like conformational arrangement. (c) Dynamic equilibrium between open and closed conformational transitions via two intermediate states (I and II). Both the C5 and C5′ bases are highlighted for better visuals with default colors of the atoms (carbon in red, oxygen in light red, and nitrogen in blue).

these observations, it has thus been proposed that G4 (un)folding *in vivo* could possibly represent another layer of non-genetic, but structural regulation of gene expression.[32] Besides the human genome, PQFS are also present extensively in bacteria[33] and viruses.[34]

The involvement of G4s in several biological processes has made them a potential target for therapeutic intervention. The structural features of G4s have been exploited to increase their thermodynamic stability via induced formation by small molecules.[15,35] G4s with high thermodynamic stability are an obstruction to processivity by the cellular replicative machinery. This was the central idea behind the design of G4 stabilizing agents targeting telomere maintenance[36] or associated suppression of transcriptional activation in proliferative cancer cells.[37] Most of the G4 interacting small molecules present in the literature focus on the stabilization of the G4 structure.[38−40] On the other hand, the importance of destabilizing G4 structures has been underexplored. Destabilization of G4 structures has been shown to enhance translational efficiency in the FMR1 gene and the 5′-UTR of the FMR1 mRNA in Fragile X syndrome.[41,42] Furthermore, a genetic loss or age-related changes in G4 modulating proteins, compounded by over-representation of G4s, have been shown to accelerate brain aging and foster neurological disorders.[43] Thus, destabilization of G4 structures can be another means of controlling gene expression or find applications in treating age-associated neurobiological disorders. However, it has not been trivial to reliably assess G4 destabilization by small molecules due to the lack of standard assays and protocols.[32] It is only recently (while this manuscript was under review) that Monchaud and co-workers have published a G4-helicase-based destabilization assay.[44] Nevertheless, there are small molecules that have been reported to destabilize G4s. For example, TMPyP4 disrupted the G4 structure in the Fragile X FMR1 gene[42] and in the MT3 endopeptidase mRNA sequence;[45] a triarylpyridine derivative disrupted G4 in the cKit-1 and 2 sequences;[46] an anthrathiophenedione[47] and a stiff-stilbene derivative were shown to unfold a sodium form of telomeric G4.[48]

https://doi.org/10.1021/jacs.1c11248
J. Am. Chem. Soc. 2022, 144, 935−950

The omnipresence of G4 structures in a cell-based setting presents a formidable challenge to stabilize or unwind a specific topology. The formation and dissolution of G4s have been studied by a wide variety of biophysical and chemical probe methods.[45,48−50] Several G4−ligand complexes have been characterized by crystallographic and NMR studies.[51−57] Recently, the DEAH/RHA helicase DHX36-G4 complex structure laid the structural foundation to explore the G4 unfolding mechanism by a helicase.[58] Besides, numerous helicases that unwind G4 have been identified and are being used as molecular tools to study G4 unwinding.[32]

In the absence of crystal structures, computational methods have been an indispensable tool to study such processes. Recently, Moraca et al. showed the binding mechanism of berberine (a polycyclic aromatic compound) to human telomeric G4-DNA and its stabilization through a combined effort using well-tempered metadynamics (WT-MetaD) simulation and steady-state fluorescence spectroscopy experiments.[59] In another study, O'Hagan et al. studied the reversible unfolding mechanism of G4-DNA by a photoresponsive stiff-stilbene ligand in the presence of sodium buffer using WT-MetaD simulations, circular dichroism, and NMR spectroscopy.[48] In particular, this study used human telomeric antiparallel G4-DNA to investigate unfolding. One common feature that was identified while looking at the ligand−complex structures is that the binders are mostly one or more polycyclic and planar aromatic chromophores, able to engage in $\pi-\pi$ stacking interactions with the terminal G-tetrads, and a positive charge that is necessary to interact with the DNA backbone phosphate groups.[40] Specifically, this study also showed that the unfolding of G4-DNA initiated from the groove via the formation of the $\pi-\pi$ stacking interaction between the stilbene moiety and G-tetrads.[48]

Among all classical G4 interacting ligands, TMPyP4 is a paradox that exhibits both G4 stabilizing and destabilizing properties. TMPyP4 has been shown to stabilize G4 structures and exert antitumor[60,61] and antiviral activity.[62] In the NMR structure (PDB id 2A5R), TMPyP4 was reported to stabilize the c-Myc Pu24I by stacking with the surface of the 5′-quartet.[57] Neidle and co-workers explored the interactions between TMPyP4 and biomolecular human telomeric quadruplex (PDB id 2HRI), in which TMPyP4 exhibits an alternative binding mode by stacking with the loop TTA nucleotides instead of the G-quartet.[63] However, although TMPyP4 has also been shown to stabilize an RNA G-quadruplex from the Ebola virus, there are also reports that TMPyP4 has a completely opposing effect on G4-RNA structures and unwinds them.[45,49,64] Until now, there is no structural insight on how TMPyP4 or any small molecule unwinds G4-RNA structures.

We approached this issue by studying the interactions between TMPyP4 and an RNA G4 forming sequence (PQS18-1; r(GGCUCGGCGGCGGA)) from the noncoding region of pseudorabies virus (PRV).[65] The structure of PQS18-1 (PDB id 6JJH, 6JJI) has been published recently.[66] It is a bimolecular, all-parallel stranded G4-RNA consisting of four-stacked tetrads with three K^+ ions positioned along the central axis. The structure contains two molecules of TMPyP4 ligand bound to the G4: one intercalated at the 3′ end between the top G-tetrad plane and an A-diad formed from A14:A14′ bases and one external to the G4-RNA in a stacked arrangement between a pair of uracils (U4:U4′) and a pair of cytosines (C11:C11′) (Figure 1).

Herein we report a computational study of the unfolding of an RNA G4 topology by a porphyrin TMPyP4 ligand from a well-tempered metadynamics (WT-MetaD) enhanced sampling simulation. Our simulation data (∼1.5 μs) showed that TMPyP4 binds on the top face as well as in the groove side of the RNA G4 mimicking the X-ray crystal structure binding poses.[66] Moreover, extending the simulation to a further 1.1 μs (a total of ∼2.6 μs) highlighted the complete unfolding of the RNA G4 topology. Further, to validate our theoretical results and assess the reliability of the method used, we performed several experiments. Through WT-MetaD simulations, circular dichroism (CD), ultraviolet−visible (UV−vis) absorbance, Förster resonance energy transfer (FRET) titrations , and isothermal titration calorimetry (ITC) experiments, we hypothesize that the unfolding process could be divided into three states: (1) an initial binding between the ligands and RNA G4, (2) dynamic movement of the intermediates or formation of the intermediate complex through opening of the G-tetrads via intercalation, and (3) unfolding. Finally, our results indicate that computational simulations can predict ligand-induced unfolding of G4s with precision. The WT-MetaD data complement experimental findings and provide extensive structural information for the TMPyP4-mediated RNA G4 unfolding process; therefore they can be used as a useful tool for investigating G4/ligand interactions.

RESULTS

Unbiased MD Simulations. To gain insights into TMPyP4 ligand binding to RNA G4, we carried out unbiased classical MD simulations of both the native RNA G4 and all available TMPyP4-PRV PQS18-1 RNA G4 complexes (PDB id 6JJH, 6JJI).[66] Crystal structures revealed a top-face-bound state of TMPyP4 where it is sandwiched in between the top fraying base A14 and A14′ and the bases from the first G-tetrad such as G10, G10′, G13, and G13′.[66] In the second bound state, TMPyP4 is mainly solvent exposed and stacked on the C11 base of the RNA G4 loop (Figure 1a). We performed 1 μs (×3) of classical MD simulation of both the native and TMPyP4 groove bound states to test the stability of these systems. Further, we carried out principal component analysis (PCA) of all the systems investigated and then compared the results to understand the dynamic behavior of the RNA G4 structure. PCA analysis on the native G4 shows that the RNA G4 is highly flexible in water and can rapidly change conformations between open and closed states. At the start, both the C5 and C5′ bases are positioned adjacent to one another on the top plane of the terminal G-tetrad (Figure 1b,c), resulting in the backbones of chain A and chain B spreading out. We consider this as the open-like conformation as found in the crystal structure (Figure 1b, bottom). In the closed state, both the C5 and C5′ bases are found to be stacked on top of each other, and the RNA G4 shrinks by bringing the backbone of chains A and B close to each other relative to the open state. Intermediate states (I and II, Figure 1) show how the C5 and C5′ bases transform their position from open to closed states (Figure 1c).

Further, MD simulation demonstrates that the top-face-bound state is stabilized in the open-like conformation, whereas the groove-bound state resembles the closed conformation. Figure 1 shows the sequence of transformation from open to closed conformation and *vice versa*. (See Supporting Information for more details on the discussion of PCA analysis of native RNA G4 and its complexes.)

https://doi.org/10.1021/jacs.1c11248
J. Am. Chem. Soc. 2022, 144, 935−950

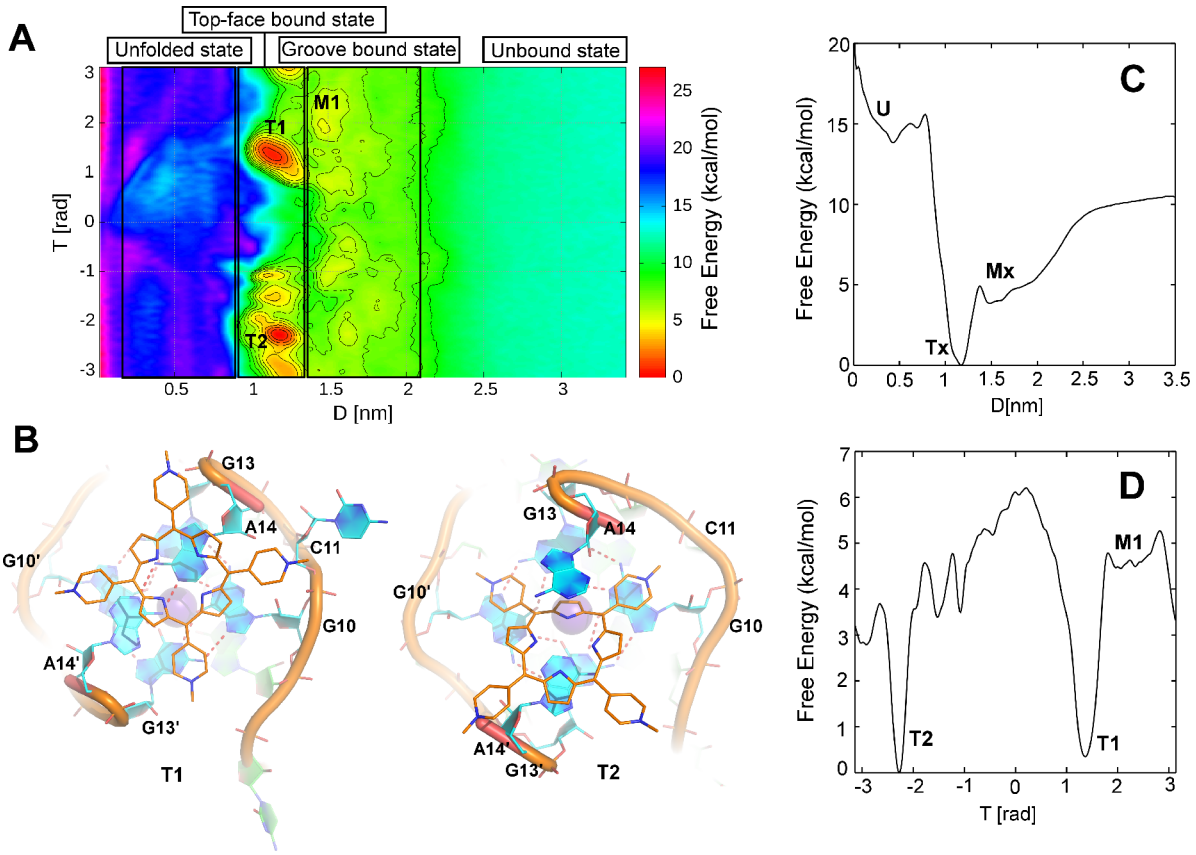


Figure 2. Free energy calculations of TMPyP4 binding to PQS18-1 RNA G4. (a) Free energy surface (FES) of binding of TMPyP4 to RNA G4 topology illustrated as a function of distance (D) and torsion (T) collective variables. Each box in the FES represents a different state of RNA such as TMPyP4 bound to the RNA top face ($0.9 \leq D \leq 1.4$), TMPyP4 bound to the RNA groove ($1.4 \leq D \leq 2.1$), unbound state (>3.0), and TMPyP4-mediated RNA unfolded state (U) ($0.0 \leq D \leq 0.9$). The two most pronounced basins are found in the top-face-bound state, T1 and T2. (b) The most populated clusters are shown for both the top-face-bound states. In T1, TMPyP4 binds on the extreme top of the RNA interacting with A14 and A14′ fraying bases, whereas, in T2, TMPyP4 is found to be slightly tilted relative to the T1 pose, interacting with first with the G-quartet bases, G10, G10′, and G13, and A14 and A14′ nucleotide bases and mimicking the native crystal TMPyP4-RNA G4 bound complex. (c) The one-dimensional potential of mean force (PMF) is plotted as a function of distance (D) in nm, showing $\Delta G_{cal} = -10.5$ kcal/mol. Free energy basins such as Tx, Mx, and U represent the corresponding top-face, groove-bound, and unfolded states. (d) One-dimensional PMF plotted as a function of torsion (T) CV in radians, illustrating the absolute free energy difference between T1 and T2 basins.

Free Energy Calculations Using Well-Tempered Metadynamics Simulations. To study the complete binding and unfolding of the TMPyP4 ligand to RNA G4, we performed WT-MetaD simulation. Metadynamics (MetaD) is an enhanced sampling simulation technique that allows simulating long-time-scale events considered as rare events such as protein−ligand binding[67] and the protein folding/unfolding mechanism[68] in a reasonable computational time cost.[69] At the end of the simulation, the free energy landscape of the simulated process of interest can be computed using the history-dependent biasing potential which was added during the simulation on a chosen degree of freedom called collective variables (CVs).[69] This technique has already been successfully used by us and also by a few other research groups to simulate biological processes such as the folding/unfolding mechanism of RNA tetraloops[70,71] and G4−ligand binding[48,72] and also in materials science such as binding of small ligands to surfaces.[73,74]

In the present study, MetaD simulation was used to (a) explore the available binding modes of the TMPyP4 ligand around RNA G4, (b) identify a possible unfolding mechanism of RNA G4, and (c) compare the observed aforementioned phenomena to the available experimental data. Two collective variables were used for WT-MetaD to study binding and unbinding of the TMPyP4 ligand around RNA G4. These are the distance (D) between the center of mass (COM) of the ligand to the COM of the middle G-tetrad in the RNA G4 and a torsion (T) angle between the ligand and the G4. The torsion angle between the ligand and G4 was measured as the two points from the ligand to two points from the RNA G4 (see Table S1 in the Supporting Information for details). The simulation took over 2.6 μs to converge with the following protocol: first, a single WT-MetaD simulation is performed to reach a semiquantitative convergence where we have sampled all the possible free energy minima for the ligand binding to RNA G4. Particularly, two separate binding conformations (basins) are sampled, and they are the top-face and groove-binding conformations. To further accelerate the sampling of these conformations, four consecutive walkers are placed along the path from successive basins using the multiple walkers technique for rigorous sampling (see below). A two-dimensional representation of the free energy surface (FES), as a function of D and T, is shown in Figure 2. Further, to have a quantitatively well-characterized free energy profile, various recrossing events between the different states, such as bound and unbound states, visited by the system should be seen. To

https://doi.org/10.1021/jacs.1c11248
J. Am. Chem. Soc. 2022, 144, 935−950

Table 1. Thermodynamic Parameters of ITC Experiments[a]

experiment	K_a (M^{-1})	ΔH (cal/mol)	ΔS (cal/mol/deg)	n	ΔG_{expt}* (kcal/mol)	ΔG_{cal} (kcal/mol)
RNA PQS18-1 with TMPyP4	$5.74(\pm0.46) \times 10^5$	$-2.17(\pm0.09) \times 10^4$	−46.5	1.24 ± 0.02	7.86	6.7
	$2.83(\pm2.67) \times 10^7$	$-6.87(\pm4.08) \times 10^2$	31.8	0.32 ± 0.04	10.17	10.5

[a]The ΔG_{expt} is calculated from the equation of $-RT \ln K_a$, where R, T, and K_a are the universal gas constant, standard temperature of the reaction, and binding constant, respectively.

provide a picture of the convergence of the binding free energy estimation, the free energy difference between the bound and unbound states was computed as a function of the simulation time (Figure S1). The estimate of free energy of binding of TMPyP4 to RNA G4 converges to −10.5 kcal/mol (ΔG_{cal}), which is close to the experimentally obtained binding free energy (ΔG_{expt}) of −10.2 kcal/mol (Table 1).

TMPyP4-RNA G4 Interactions: Top-Face and Major Groove-Binding Modes. The available crystal structure confirmed the top-face and groove-binding modes of TMPyP4 to RNA G4.[66] The two-dimensional FES (Figure 2a) can distinguish different states including top-face-bound state, groove-bound state, unbound state, and the unfolded state. Top-face-bound states are represented as Tx where x corresponds to the number of bound states based on their position and orientation (Figure 2). The top-face-bound states are assigned within the distance $0.9 \leq D \leq 1.4$ nm in the D versus T curve. In basin T1, a cluster analysis using an RMSD cutoff value of 0.17 nm indicated that TMPyP4 acquires a top-face binding conformation (with 93.4% population) where the major contribution comes from the stacking interaction with both the top A14 and A14′ nucleotide bases (Figure 2). In particular, all four pyrrole rings are involved in the $\pi-\pi$ stacking interaction with the aforementioned bases (see T1, Figure 2). We note that a few ionic interactions are also seen between the pyridinium cation and negatively charged oxygen atoms of the phosphate backbone of RNA G4. Basin T2 shows a rather tilted top-face binding conformation where TMPyP4 is intercalated in between the A14 and A14′ fraying bases. Further, TMPyP4 is also seen to partially stack with G10, G13, and G10′ bases with a population of 75.6%, as calculated from the cluster analysis. This pose is found to be very similar to the available top-face TMPyP4-RNA G4 binding pose in the X-ray crystal structure (Figure 1). Energetically, both the top-face-bound states (T1 and T2) are found to be very similar in their binding free energy with a difference of 0.3 kcal/mol (Figure 1). Although the present WT-MetaD simulation is not able to capture the exact crystal structure top-face binding conformation, it has been quite successful in sampling the near-native binding conformations (Figure 1). Further, PCA on the top-face binding pose revealed that TMPyP4 binding on the top-face allows the RNA to adopt an open-like conformation (Figure S3). Moreover, the PCA analysis is also performed on the native RNA itself to differentiate the open to closed conformational changes and compare it with the bound states (Figure S2).

Major groove binding modes are represented as Mx where x accounts for the number of TMPyP4 groove-bound states differing in their position interacting with different nucleotide bases around RNA G4. The groove-bound states are assigned within the distance $1.4 \leq D \leq 2.1$ nm in the D versus T curve. A cluster analysis on the most stable basin (M1, Figure 1 and Figure S5) in the groove-binding region revealed a high degree of heterogeneity on the ligand-binding position, i.e., toward different binding sites since RNA G4 has a top-face and four groove-binding sites. Three equally populated clusters from the basin M1 were obtained (Figure S5). TMPyP4 is either bound to the top face of RNA G4 interacting with fraying adenine (A14 and A14′) bases or in the groove site binding individually to C11 and C11′ bases (see the Supporting Information for more details). Thus, the groove-binding sites have not been well separated by the FES portrayed with D and T CVs in Figure 2.

To address this issue, we explored several variants of CVs. Following a previous study by O'Hagan et al.,[75] we have adopted two new CVs such as ligand position in the x- and y-axis (taken as the vector of the distance in the x and y direction, i.e., $d{\cdot}x$ and $d{\cdot}y$, respectively). This condition only applies when the RNA principal axis is aligned with the z-axis (Figure S6). The vectors are able to describe the possible groove-binding modes of TMPyP4 since the groove sides lie in the x and y plane (Figure 3). The major advantage of using these two CVs is to separate each groove-bound state in the 4-fold symmetry. These CVs are also able to separate the top-face-bound state as well. Looking at the FES, four additional basins are found along with T, the most stable one, which represents the top-face, TMPyP4 bound state (Figure 3). In basin M1, a cluster analysis revealed that TMPyP4 adopts a groove-binding conformation in which it is mostly interacting with the C11 base via a $\pi-\pi$ stacking interaction. This pose agrees well with our crystal structure (PDB id 6JJH).[66] In particular, one of the pyrrole rings from TMPyP4 is involved in the stacking interaction. We carried out a 1 μs (×3) unbiased MD simulation with the ligand bound in this pose. The TMPyP4 remained stacked with C11, and no significant deviation from the starting structure was observed, indicating that this pose was stable (Figure S7). Further, a PCA analysis on the MD data of RNA G4 with TMPyP4 in the groove-bound state suggests that the RNA G4 topology remains in a closed conformation for the M1 binding mode (Figure S4). This finding has great importance on the TMPyP4-mediated unfolding mechanism of the RNA G4 topology.

A cluster analysis was performed for basin M2, where TMPyP4 remains on the opposite side to C11−C11′ base stacking interactions. In particular, several electrostatic interactions are established between the pyridinium cation and backbone oxygen atoms of the G3′, G12′, G10′, G7′, and U4′ nucleotides. The stability of this pose is again validated with ∼1 μs (×3) unbiased MD simulation. As no significant deviation from the starting structure was observed, we concluded that it is a stable binding pose (Figure S7). This pose was similar to the available X-ray crystal structure[66] since the C11′ base is equivalent to C11 due to the observed symmetry between both RNA strands (chain A and chain B). Basin M3 is a groove-binding pose involving mainly electrostatic interactions and partially one of the ligand's pyridine rings interacting with the C3′ base via $\pi-\pi$ stacking interactions. Moreover, in this conformation, one of the pyrrole rings of TMPyP4 makes a stacking interaction with the G1′ sugar pucker ring. Furthermore, an electrostatic

https://doi.org/10.1021/jacs.1c11248
J. Am. Chem. Soc. 2022, 144, 935−950

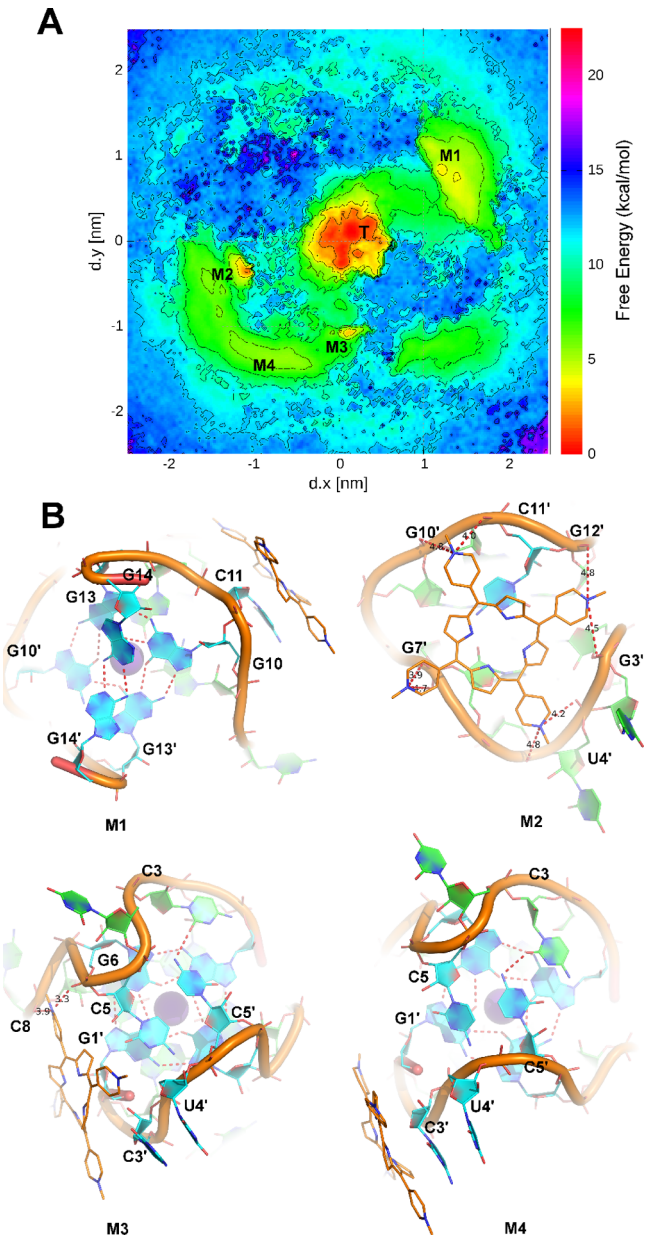


Figure 3. Free energy surface of groove binding states. (a) Free energy surface plot considering $d{\cdot}x$ and $d{\cdot}y$ as the two major collective variables (CVs). The CVs were able to capture the TMPyP4−RNA G4 groove-bound state. Basin T at the center represents the top-face binding conformation, whereas Mx corresponds to the groove-binding conformations. (b) Basins M1 and M2 represent the X-ray crystal-structure-like groove-bound states, where TMPyP4 interacts with C11 and C11′ bases, respectively. In basin M3, TMPyP4 partially interacts with the C3′ base, in addition to the negatively charged oxygen atoms from the RNA backbone with its positively charged pyridine nitrogens. Basin M4 is the least stable free energy basin, where TMPyP4 mainly interacts with the C3′ base.

interaction is established between the pyridinium cation and RNA backbone oxygen atoms of the C8 and G6 nucleotides. In basin M4, a solvent-exposed conformation of TMPyP4 is observed where it interacts with the C3′ base via a $\pi-\pi$ stacking interaction. In particular, one of the pyrrole rings from the porphyrin moiety and the pyridine ring is involved in the $\pi-\pi$ stacking interaction with the C3′ base. The only difference between basins M3 and M4 is that in the present conformation TMPyP4 has slightly shifted toward the C3′ base by breaking the noncovalent interaction between the G1′ sugar pucker and pyrrole ring and at the same time rotates slightly on the plane perpendicular to the G-RNA stabilizing the mentioned $\pi-\pi$ stacking interaction.

The available X-ray crystal structure[66] resolves two binding modes of TMPyP4: (a) top-face and (b) groove-site interactions with specific bases such as C11 and C11′ bases. The results from the WT-MetaD simulation complement the crystalline data and capture other binding possibilities of TMPyP4 to RNA G4 such as binding to the C3′ base and also a $\pi-\pi$ interaction between the pyrrole ring and sugar pucker ring, thereby highlighting the robustness of the present simulation. At the end of the simulation, basin M3 is found to be equally stable with the groove-bound mode such as M1 and M2. Furthermore, the free energy difference between the top-face binding mode and groove-binding mode is calculated to be ∼3.8 kcal/mol ($\Delta\Delta G_{cal}$), which is in good agreement with experimental ITC data ($\Delta\Delta G_{expt}$ = 2.3 kcal/mol) (Table 1) The FES is able to identify a suitable pathway of binding/unbinding events of TMPyP4. The possible binding/unbinding events generally occur via the interaction between TMPyP4 and the C11 nucleotide base, resulting in the formation of the M1 basin (Figure 3). This particular base plays a key role in bringing TMPyP4 back to the top face of RNA G4 from the bulk solvent. The rebinding mechanism of TMPyP4 to RNA G4 is shown in Movie S1 in the Supporting Information.

We then employed several orthogonal biophysical methods to further explore our computational findings. To study the conformational states of G4 structures formed from the RNA PQS18-1 sequence, we employed CD, a widely used analytical method that gives information about DNA structure and folding.[76] The RNA PQS18-1 was first annealed in a buffer containing 10 mM lithium cacodylate (pH 7.0) with a metal ion concentration of 100 mM KCl.

The CD spectra of RNA PQS18-1 under these buffer conditions gave rise to a positive peak at 264 nm and a negative peak at 240 nm, which is consistent with a parallel-stranded RNA G4 topology (Figure 4). CD titration methods were then utilized to investigate the formation of the G4/ligand complex using TMPyP4.[77] We observed that as TMPyP4 was added to RNA PQS18-1, a concentration-dependent decrease in the CD signal at 264 nm was observed across all the concentrations, to a plateau at ∼60−70 equiv (60−70 μM). Meanwhile, although initially there was no signal observed at 295 nm, after addition of 1 equiv of TMPyP4 (10 μM) a shoulder appears at this wavelength, consistent with remodeling of the RNA in an antiparallel formation during the binding process. As further equivalents of TMPyP4 are added, this signal increases up to 15 μM, reaches a plateau, and then also decreases in a similar fashion to the parallel signal at 264 nm. Plotting the ellipticity at 264 nm against concentration of TMPyP4 added gave a sigmoidal-shaped curve, indicating a cooperative process. We fitted this curve to the Hill equation, which identified a Hill coefficient (n) of 4.1 ± 1.3. This reveals that the binding of TMPyP4 to PQS18-1 exhibits positive cooperativity (n > 1). The half degrading concentration ($[DC]_{50}$) was determined to be 42 ± 0.5 μM. These results are consistent with the unfolding mechanism observed in our modeling experiments. As a direct comparison we also performed a similar titration with the analogous DNA sequence using CD (Figure 5). In the DNA sequence equivalent, on addition of TMPyP4, there was no unfolding

https://doi.org/10.1021/jacs.1c11248
J. Am. Chem. Soc. 2022, 144, 935−950

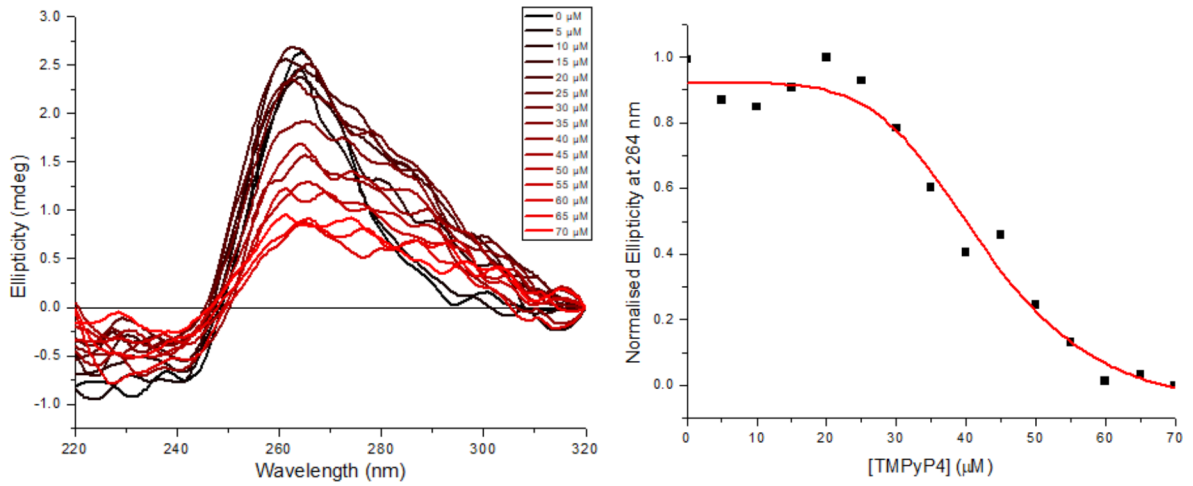


Figure 4. Left: CD titration of PQS18-1 RNA (10 μM) and TMPyP4 (0−70 μM) in 10 mM lithium cacodylate and 100 mM KCl, pH 7.0. Right: Plot of ellipticity at 264 nm against the concentration of TMPyP4 and corresponding Hill fitting.

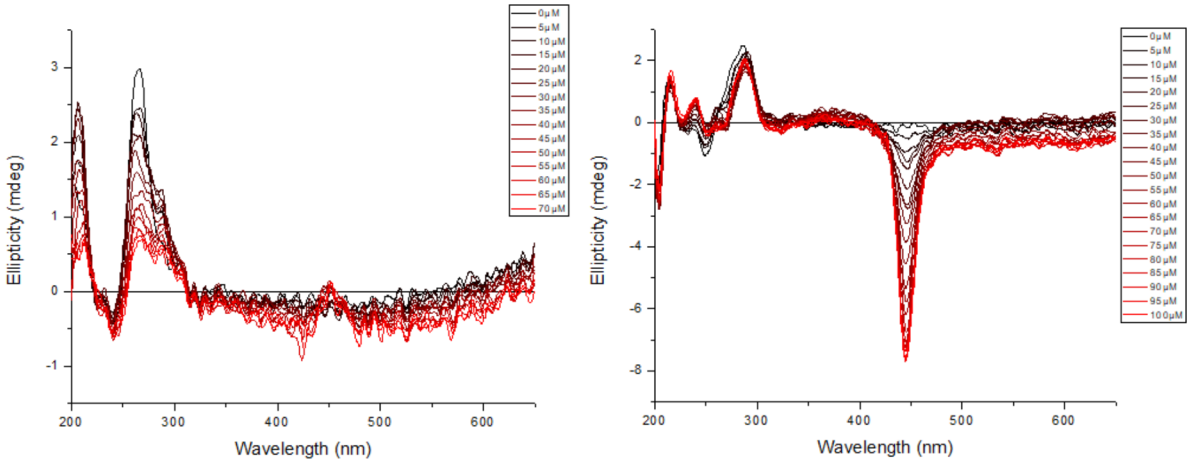


Figure 5. Left: CD titration of PQS18-1 RNA (10 μM) in the presence of 0−70 μM TMPyP4. Right: CD titration of PQS18-1 DNA (20 μM) in the presence of 0−100 μM TMPyP4. Both experiments were performed in 10 mM lithium cacodylate and 100 mM KCl, pH 7.0.

effect observed; however a prominent band at 445 nm appeared, which increased in intensity with further additions of the ligand. We attribute this to an induced circular dichroism (ICD), indicative of strong binding between the DNA G4 and the TMPyP4. This is consistent with other reports of TMPyP4 with DNA G4.[78−80] In contrast to this, in the equivalent titration with RNA G4 there is only weak induced exciton splitting (Figure 5, left panel). This suggests formation of dimers, or higher order complexes, either in a groove-binding or an external stack-binding mode, consistent with our molecular dynamics studies.[81−83] This direct comparison between both DNA and RNA highlights the differences between the manner TMPyP4 interacts with G4 structures, depending on its composite nucleic acid.

To further analyze the effect of TMPyP4 on RNA PQS18-1, we performed a Job's analysis using CD to indicate the stoichiometry is between 1:1 and 2:1, averaging 1.5:1 (Figure S8). We performed CD melting analysis to determine the ligand-induced effect of TMPyP4 on the stability of RNA PQS18-1. The melting of RNA PQS18-1 in the CD seemed to give rise to two melting events, one at 45 °C and another at 73 °C (Figure 6). On addition of 1 equiv of TMPyP4, the shape of the curve starts to change, but the overall T_m values did not change. Addition of another equivalent of TMPyP4 changed

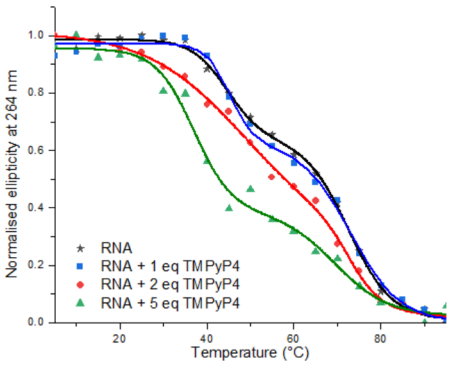


Figure 6. CD melting of PQS18-1 RNA (10 μM) in the presence of 0 equiv (black stars), 1 equiv (blue squares), 2 equiv (red circles), and 5 equiv (green triangles) of TMPyP4 in 10 mM lithium cacodylate and 100 mM KCl, pH 7.0.

the shape of the curve again, to give an overall T_m of 71 °C, which indicates a ΔT_m of −2 °C. This change in the shape of the melting curve is further increased when the number of equivalents of TMPyP4 is increased to 5 equiv, where again there are two clear transitions at 37 and 70 °C, indicative of ΔT_m values of −8 and −3 °C. These results indicate TMPyP4

https://doi.org/10.1021/jacs.1c11248
J. Am. Chem. Soc. 2022, 144, 935−950

has a destabilizing effect on the RNA G4 structure. After the melting experiments, we also studied the corresponding annealing experiments in the absence and presence of ligand. These indicated that the melting and annealing processes in the presence of TMPyP4 are not reversible, which is not unexpected for the liganded complexes (Figures S9–S11). However, we also observed precipitation of complex in the annealed samples, so the results from the annealing experiments are complicated by precipitation and aggregation processes. The corresponding UV–vis melting and annealing experiments were performed in parallel, but were also affected by TMPyP4 absorption in the same region as the DNA (Figures S12–S15).

We considered that some of the signal losses in the CD spectra might be through the effects of aggregation, rather than unfolding or disruption of the G4 structure, so during the CD titrations we also monitored the UV absorption spectra (Figure S16). Using this we were able to observe a linear dose–response on addition of TMPyP4 at 217 nm, consistent with the Beer–Lambert law, indicating no aggregation at the concentrations examined at room temperature. Beyond these concentrations we sometimes observed precipitation, easily recognized as TMPyP4 is colored. The apparent unfolding effects were observed at much lower concentrations than this, and here we present data only at concentrations where precipitation was not evident.

UV–vis titrations of TMPyP4 with RNA titrated in (Figure 7) gave rise to a reduction and shift in the visible absorption

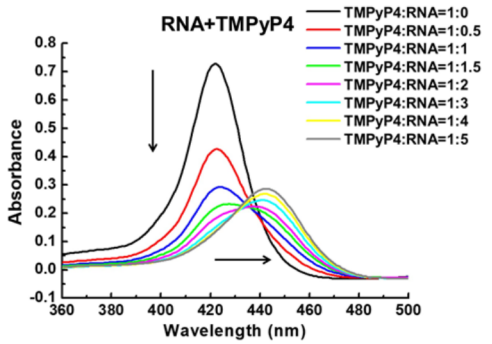


Figure 7. UV titration spectra. UV titration of TMPyP4 with RNA PQS18-1 in solution buffer (10 mM K_2HPO_4/KH_2PO_4, pH 7.0, 100 mM KCl). Conditions: [ligand] = 4 μM; titrant: [RNA] = 0–20 μM. The vertical arrow indicates the decrease in the absorbance of TMPyP4, and the horizontal arrow indicates the shifted Soret band.

spectra of TMPyP4 in the absence and presence of RNA PQS18-1. The strength and type of binding are indicated by the significant changes in wavelength maxima of absorption spectra upon addition of RNA G4, both the bathochromic shifts ($\Delta\lambda$ = 20 nm) and hypochromicity (70%) shift to hyperchromicity (13%). Here the high values of bathochromic shifts and hypochromicity indicate strong stacking interactions between TMPyP4 and G-tetrads.

We also recorded UV spectra during the CD titrations between RNA PQS18-1 and TMPyP4, i.e., RNA with TMPyP4 added. Plotting the absorption at 440 nm against concentration gave a binding curve (Figures S17), which we fitted to a 2:1 binding model to indicate K_ds of 19 ± 0.7 and 188 ± 19 μM.

As an alternative to UV and CD, we also explored the apparent unfolding using FRET titrations, using dual-labeled RNA PQS18-1, with FAM and TAMRA as a FRET pair. When in the folded G4 conformation, the fluorophores will be in close proximity and FRET will occur, and it is possible to follow unfolding of G4 using this method. Folded RNA PQS18-1, when excited at 490 nm, gives rise to an emission spectrum with maxima at 515 and 585 nm, from the fluorophores FAM and TAMRA respectively. Addition of 0–5 equiv of TMPyP4 to the folded RNA PQS18-1 caused a decrease in emission at 585 nm, indicative that TAMRA emission is decreasing and potentially unfolding (Figure S18). TMPyP4 was also observed to cause an overall decrease in fluorescence emission through quenching, so we determined the FRET efficiency (E_{FRET}) between the two fluorophores. A concentration-dependent decrease in E_{FRET} was observed on addition of TMPyP4, consistent with unfolding of the RNA PQS18-1 structure in the presence of TMPyP4.

Finally, we next investigated the interaction between TMPyP4 and RNA PQS18-1 G4 by calculating complete thermodynamic parameters, including the stoichiometry and the values of free energy (ΔG), enthalpy (ΔH), and entropy (ΔS) changes of the binding reaction using ITC. Correlating such thermodynamic data with a structural description increases our understanding of the molecular recognition process involved in complex formation and maintenance. The TMPyP4 (500 μM) was titrated into a RNA solution (20 μM, 200 μL) with 10 mM K_2HPO_4/KH_2PO_4, pH 7.0, and 100 mM KCl buffer at 25 °C. Figure 8 shows the integrated heat change data (after the correction of heat of dilution), and the corresponding heat changes are fit with the multiple-sites binding model. All the fitting parameters are summarized in Table 1. The two-site binding data were then fitted to a two-independent-site binding model to determine the affinity and binding enthalpy of each of the two sites. Hence, the ITC

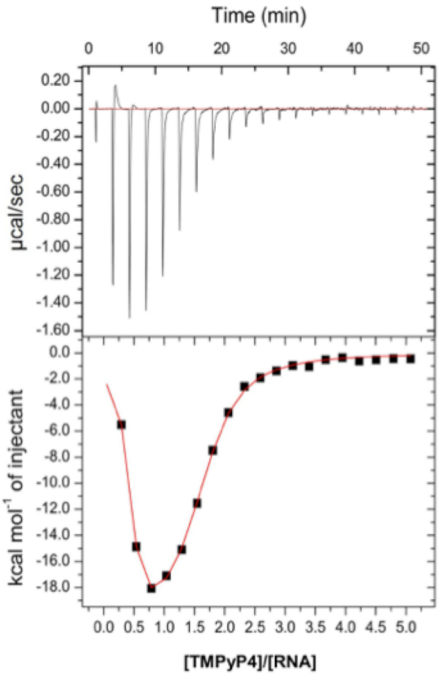


Figure 8. ITC of TMPyP4–RNA G4 complex ITC binding profile of titration of 500 μM TMPyP4 solution into 20 μM RNA PQS18-1 solution in 10 mM K_2HPO_4/KH_2PO_4, pH 7.0, and 100 mM KCl buffer at 25 °C. Raw and fitted isotherms are shown in the panel. It is corrected for the corresponding heat of dilution by blank titration of 500 μM TMPyP4 solution into buffer.

https://doi.org/10.1021/jacs.1c11248
J. Am. Chem. Soc. 2022, 144, 935–950

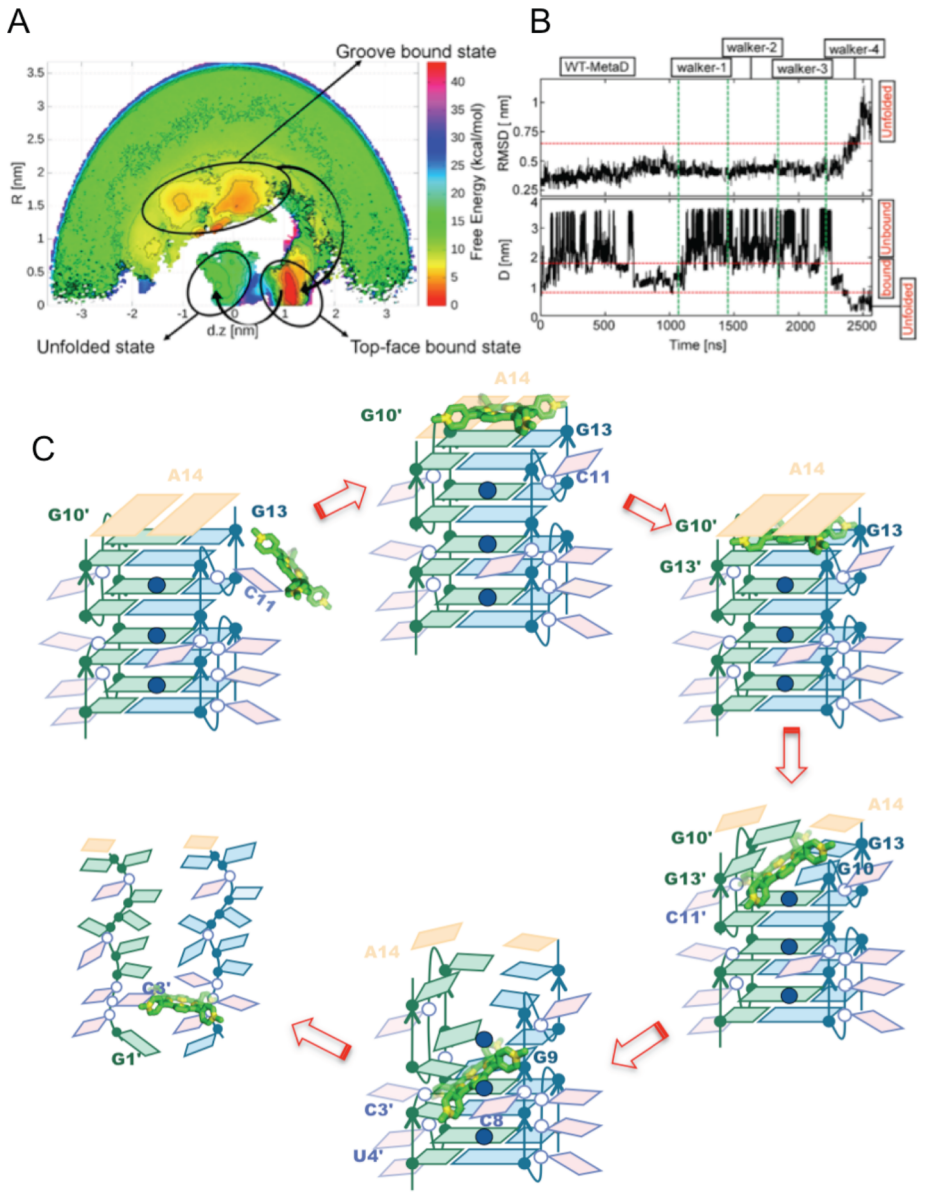


Figure 9. Unfolding mechanism. (a) Two-dimensional free energy surface plotted as a function of POA (position on the axis as $d{\cdot}z$ in nm) and DFA (distance from axis as R in nm), highlighting the pathway between various states. TMPyP4-mediated unfolding is initiated via stacking interactions identified in the top-face-bound state. (b) RMSD of the RNA G4 backbone throughout the 2.6 μs of simulation time (top). As observed in the last replica (walker-4), RMSD increased from an average value of ∼3.5 nm to >1.0 nm (thereby decreasing the CV distance from ∼1.0 nm, i.e., from the top-face-bound state to ∼0.2 nm), indicating unfolding via intercalation. The simulation statistics, illustrating the distance between TMPyP4 and RNA G4 where all the walkers are combined together (bottom). (c) Proposed model for ion-assisted recognition and unfolding mechanism of G4 by TMPyP4. Guanine nucleotides are in blue (chain A) and green (chain B). Loop nucleotides are in pink. Potassium ions are in blue. TMPyP4 is illustrated in green sticks with nitrogen atoms in yellow.

experiments resulted in K_{a1} value of $(5.74 \pm 0.46) \times 10^5$ M^{-1}, a ΔH_1 of -21.7 ± 0.9 kcal/mol, and a ΔS_1 of -46.5 cal/mol/deg (low affinity site), while the high-affinity site has a K_{a2} value of $(2.83 \pm 2.67) \times 10^7$ M^{-1}, a ΔH_2 of -0.68 ± 0.4 kcal/mol, and a ΔS_2 of 31.8 cal/mol/deg. The results indicated that the binding of TMPyP4 to two sites was different. The nonsigmoidal binding curve revealed that TMPyP4 could bind to the RNA at more than one site. The ITC data indicate that the stoichiometry of this multisite interaction is two-site binding as the thermogram saturates after the addition of four molar equivalents of TMPyP4. The dip at the start of the ITC thermogram indicates that the second (lower affinity) site is more exothermic than the first (higher affinity) site (Figure 7). The enthalpy of the system starts decreasing as the second site starts to be populated. The total enthalpy is negative when both sites are saturated with the ligand. These data are strikingly correlated with our WT-MetaD simulation results, where we found two stable binding sites as described above. The high- and low-affinity sites correspond to the top-face and groove-bound states, respectively. Since the groove-bound state is completely solvent exposed, as seen in the crystal data as well as from the WT-MetaD simulation, it is evident that while moving from a top-face-bound state to a groove-bound state, the enthalpic contribution to the free energy of binding decreases as the solvation entropy plays a vital role in the ligand binding. The binding mechanism predicts that TMPyP4

https://doi.org/10.1021/jacs.1c11248
J. Am. Chem. Soc. 2022, 144, 935−950

is being caught by the fraying base C11 through a $\pi-\pi$ stacking interaction, where it forms the groove-bound state, showing the importance of the C11 base on the G4-TMPyP4 association. While sampling this particular state, the two top-face fraying bases A14 and A14′ remain in a wide-open conformational state due to the loss of H-bond between themselves, exposure to the bulk solvent, and excessive flexible nature. Finally, TMPyP4 transfers from the groove to the top face of RNA and forms the top-face-bound state by stacking on the terminal G-tetrad. This is followed by the loop closure, and TMPyP4 is eventually sandwiched between the bases from the terminal G-tetrad and top fraying bases A14 and A14 (for a better understanding of the binding mechanism, see Movie S1 in the Supporting Information).

Unfolding Mechanism. To trace the unfolding mechanism of a biomacromolecule using the molecular simulation technique, such as MD simulation, one has to simulate the system sufficiently long since the sampling time for a typical unfolding remains on the scale of microseconds to milliseconds.[84] However, one can overcome this time scale problem using an enhanced sampling technique.[85,86] A combination of WT-MetaD and the multiple walkers method has enabled us, for the first time, to characterize a possible TMPyP4-mediated unfolding mechanism of an RNA G4 topology (Figure 9b).

In this report, we have identified two novel CVs, POA (position on the axis as $d{\cdot}z$ in nm) and DFA (distance from axis as R in nm), initially applied to study the DNA ligand association process by O'Hagan et al.[75] and adapted them to study RNA PQS18-1 G4 and its interaction with TMPyP4. These CVs can configure and describe all the possible bound/unbound states along with unfolded states. The FES highlights the connections between the described states (Figure 9a). The groove-bound, top-face, and unfolded states are assigned with the following distance in the $d{\cdot}z$ versus R curve: $-1.5 \leq d{\cdot}z \leq 0.7$ nm and $1.0 \leq R \leq 2.0$ nm; $0.8 \leq d{\cdot}z \leq 1.5$ nm and $0.0 \leq R \leq 0.5$ nm; and $-0.5 \leq d{\cdot}z \leq 0.2$ nm and $0.0 \leq R \leq 0.5$ nm, respectively. A cartoon representation of the possible mechanism of the TMPyP4-mediated unfolding is illustrated in Figure 9c. As observed in the FES, the unfolding region is close to ∼0.0 nm, suggests that the unfolding occurs via the intercalation process as the CV distance between the DNA and ligand decreases. In the first step, the C11 base pulls TMPyP4 back on the top face of the RNA; that is, the rapid transfer of ligand occurs from groove-bound state to the top face (see T1 and T2 in Figure 2 and Supporting Information, Movie S1, where the ligand-rebinding process is illustrated). Before TMPyP4 jumps toward the top-face-bound state, the groove-bound state remains in a closed-like conformation (Figure S3). As TMPyP4 moves to the top face, the RNA conformation shifts from a closed to an open state. This open conformation of the RNA is also observed in our unbiased MD simulation and resembles the top-face-bound state (Figure S3). In the second step, TMPyP4 slides down toward the groove, stays in a slightly tilted conformation, and interacts with the bases from the first G-tetrad such as G10 and G13′. In the next step, TMPyP4 intercalates through the first and second G-tetrad by breaking their associated Hoogsteen bonds. As a result, TMPyP4 finally reaches the third G-tetrad by breaking the RNA G4 topology completely to an unwound state (Supporting Information, Movie S2).

DISCUSSION

The present work reflects on the complexity of ligand interactions on the stability of the G4s. We show that the ligands could behave as both stabilizers and destabilizers of G4s based on how they interact with the G4s. TMPyP4 has been reported to both stabilize and unfold G4s.[42,45,49,87] The scarcity of the dynamic structural information about the ligand binding makes it necessary to rely heavily on biophysical characterization. The Job plot indicated that the stoichiometry of TMPyP4 binding to PQS18-1 RNA G4 is more than 1:1, indicating binding at two or more distinct sites or modes (Figure S9). The "two independent sites" model exhibited in the ITC assay showed that more than two states exist in the unfolding processes (Figure 8). When ligand:DNA = 3:4 (ITC experiment)/2:3 (UV experiment), both the UV absorption and ITC thermogram exhibited a big change. This concentration-dependent biophysical character variance suggests the existence of a structural intermediate in the titration process. Importantly, our biophysical experiments were performed at a range of concentrations of RNA, from 0.1 μM (in the FRET titrations) to 20 μM (in the ITC experiments). This indicates the unfolding effect is observed at a wide range of concentrations of RNA and is not limited to when the RNA is at higher concentrations.

The ITC experiments show that TMPyP4 could completely disrupt the RNA G4 structure. This extent of TMPyP4 disruption was plausibly due to the higher melting temperature of the RNA PQS18-1 G4 structure (T_m = 68.4 °C). A lower concentration of TMPyP4 was required for RNA PQS18-1 destabilization at higher temperatures. We found that the apparent initial rates of G4 unfolding decreased as a function of thermal stability imparted by ligand binding. While previous reports have indeed shown that high potassium concentrations have a stabilizing effect on the T_m of various G4s, it is not trivial to conclude that the unfolding kinetics positively correlates with the thermal stability.

On the basis of our simulation and experimental results, we propose that the observed unfolding could be broadly divided into three steps: (a) TMPyP4 approaches PQS18-1 RNA G4 and forms an initial complex structure; (b) the stability of the intermediate complex depends upon the dynamic interactions between the ligand and the G4 structure. The dynamic interactions determine whether small molecules stabilize or unfold a G4 structure and (c) the strength of the dynamic interactions between the ligand and G4 structure determines the fate of the interquartet cations. In the case of destabilizers, like TMPyP4, the interquartet cations are expected to be ejected, which accelerates the unfolding process.

The binding mode of TMPyP4 to G4 remains controversial: some reports say that it binds and stabilizes G4s,[57,61] a few suggest that TMPyP4 alters the G4 structure,[87] while others have demonstrated that it unfolds G4s.[41,45,49] Here we show that all types of binding coexist and that the distribution of final states depends on the ligand concentration ratio and the temperature. Finally, it is becoming increasingly clear that it is extremely difficult to predict the behavior of G4 interacting ligands on different G4 topologies and nucleic acids. This emphasizes that ligands cannot be designed based on the traditional approaches of exploiting the topology, without taking dynamic interactions into account.

https://doi.org/10.1021/jacs.1c11248
J. Am. Chem. Soc. 2022, 144, 935−950

METHODS

Standard (Unbiased) MD Simulations. We carried out standard (unbiased) MD simulations of all the available TMPyP4-bound poses to RNA G4. The initial crystal structure revealed a top-face-bound state of TMPyP4, where it is sandwiched in between the top fraying base A14 and A14′ and the bases G10, G10′, G13, and G13′ from the first G-tetrad. In the second bound state TMPyP4 stacks on the C11 base, and it is mainly found to be a solvent exposed (Figure 1).

All the unbiased simulations were performed in triplicates using the Gromacs-5.0 software package.[88,89] The bsc0χ_{OL3} force field was used for the RNA parametrization.[90−93] For the ligand, the General Amber Force Field (GAFF) was used to generate parameters.[94] The charges were calculated using the restrained electrostatic potential (RESP) fitting procedure.[95,96] The RESP fit was performed onto a grid of electrostatic potential points calculated at the HF/6-31G(d) level as recommended by many works.[97,98]

The K^+ ion in the central axis of the structure was treated as an integral part of the structure. The RNA−TMPyP4 complex was solvated in a cubic box with the dimensions of 7.8 × 7.8 × 7.8 nm^3 along with 15 364 TIP3P explicit water molecules.[99] The Joung and Cheatham cations optimized for TIP3P water were used to generate the 100 mM KCl concentration of the system.[100] The RNA−TMPyP4 complexes were minimized before the equilibration and production run as follows: the minimization of the solute hydrogen atoms on the RNA and TMPyP4 was followed by the minimization of the counterions and the water molecules within the box. In the next step, the RNA backbone and all the heavy atoms on TMPyP4 were restrained, and the solvent molecules with counterions were allowed to move during a short 50 ps MD run, therefore relaxing the density of the whole system. In the next step, the nucleobases were relaxed in several minimization runs with decreasing force constants applied to the RNA backbone atoms; however, only a few phosphate atoms were kept restrained with a force constant of 0.239 kcal/mol/nm^2. After the full relaxation, the system was slowly heated to 300 K using a velocity rescaling thermostat[101] with a coupling constant of 0.5 ps employing the NVT ensemble. As the system reached the temperature of interest (300 K), the equilibration simulation was performed for 10 ns using an NPT ensemble with a Berendsen thermostat and a Berendsen barostat,[102] and 0.5 ps was used again as the coupling constant for both temperature and pressure, respectively. Finally, the production run was set for 1 μs using a Nose-Hoover thermostat[103] and a Parrinello−Rahman barostat[104] with the same coupling constant as previously taken in the equilibration simulation in the NPT ensemble. All the simulations were carried out under the periodic boundary conditions (PBC). The particle-mesh Ewald (PME) method was used to calculate the electrostatic interactions within a cutoff of 10 Å.[105,106] The same cutoff was used for Lennard-Jones (LJ) interactions. All simulations were performed with a 1.0 fs integration time step.

Well-Tempered Metadynamics Simulations. We performed a well-tempered variant of the metadynamics simulation of TMPyP4 binding to the RNA G-quadruplex. The WT-MetaD helps to understand a complete binding/unbinding mechanism of TMPyP4. Further, a pathway for the TmPyP4-mediated unfolding of the RNA G4 is disclosed for the very first time. The WT-MetaD simulation was started with a well-equilibrated structure generated from the crystal structure of the RNA−TMPyP4 top-face-bound state. In particular, the starting structure for the WT-MetaD simulation was taken after 20 ns of unbiased MD simulation, which was found to be a rather stable ligand-binding conformation. We performed over 2.5 μs of WT-MetaD simulation[107] in order to obtain an accurate estimate of RNA−TMPyP4 binding free energies through sampling all the individual states such as bound, unbound, and unfolded states. We used a combination of the following scheme:

1. Well-tempered variant (WT) of the metadynamics
2. The multiple walker technique, placing four walkers based on the TMPyp4 binding sites:
   a. Top-face-bound state (one-replica)
   b. Groove-bound state (two-replica)
   c. Unbound state (one-replica)

The same systems and MD settings as described previously were used for the WT-MetaD simulation. The Plumed 2.3 plugin[108,109] was used to carry out the simulation with the Gromacs-5.0.7 code.[89] The bias potential was calculated according to the WT-MetaD scheme as follows:

$$V(s, t) = \sum_{t'=0,\tau_G,2\tau_G,\ldots}^{t'<t} \omega\tau_G e^{-V(s(q(t'),t'))/\Delta T} e^{-\sum_{i=1}^{2}[(s_i(q)-s_i(q(t')))^2/2\sigma_i^2]}$$

where the deposition rate, ω, and deposition stride, τ_G, of the Gaussian hills were set to 0.358 kcal/mol/ps (1.5 kJ/mol/ps) and 1.0 ps, respectively. The bias factor $(T + \Delta T)/T$ was set to 15, and the final FES was calculated as follows:

$$F(s, t) = -\frac{T + \Delta T}{\Delta T}(V(s, t) - C(T))$$

where $V(s, t)$ is the bias potential added to the collective variables used and T represents the simulation temperature. ΔT is the difference between the temperature of the CV and the simulation temperature. The bias potential is grown as the sum of the Gaussian hills deposited along the chosen CV space, and finally the sampling of a particular CV space can be controlled with the tuning of the ΔT parameter.[66] The torsions used in the metadynamics simulations are illustrated in Figure S19.

Analysis. As discussed in the previous section, the X-ray crystal structures[110] were downloaded from the PDB data bank and prepared for the MD simulation. The visualization and analysis of the MD statistics are performed through the VMD molecular visualization program.[111] Hydrogen bond analysis of the quartets (Figure S20) was carried out using analysis tools implemented in VMD. All the necessary graphs are first prepared with the Gnuplot program (http://gnuplot.info) and later modified with the Gimp 2.0 (https://www.gimp.org) program. The analysis of the metadynamics simulation is performed through the Plumed 2.3 package.[109] The high-resolution figures were prepared through the PyMOL program (www.schrodinger.com).

Materials. Experiments based in China (UV and ITC) were performed using oligonucleotide sequence RNA PQS18-1 5′-[r(GGCUCGGCGGCGGA)]-3′ purchased from Tsingke Biological Technology (Beijing, China), and TMPyP4 was purchased from Frontier Scientific (Logan, UT, USA) and Sigma-Aldrich. The concentration was determined using the Beer−Lambert law by measuring the absorbance at 260 nm using a Nanodrop photometer N60 (Implen, Germany). The extinction coefficients were obtained from the IDT Web site (https://sg.idtdna.com/calc/analyzer). Further dilutions were carried out in buffer containing 10 mM K_2HPO_4/KH_2PO_4, pH 7.0 and 100 mM KCl. The starting oligonucleotide solutions were annealed in the corresponding buffer by heating to 95 °C for 5 min, followed by gradual cooling to room temperature.

Experiments based in the UK (CD, UV, and FRET) were performed using oligonucleotides purchased from Eurogentec (PQS18-1 5′-[r(GGCUCGGCGGCGGA)]-3′ PQS18-1 DNA 5′-[d(GGCTCGGCGGCGGA)]-3′, and PQS18-1 FRET 5′-[FAM-r(GGCUCGGCGGCGGA)-TAMRA]-3′); each of them were purified using reverse-phase HPLC and supplied dry. The dry RNA/DNA was dissolved in nuclease-free water in order to prepare stock solutions, the FRET-labeled RNA sample was prepared at approximately 100 μM, and the unlabeled RNA was prepared at 1 mM. The concentration of the stock solutions was identified from their UV absorbance at 260 nm with a NanoDrop by using extinction coefficients provided by Eurogentec. Further dilutions were carried out in buffer containing 10 mM lithium cacodylate, pH 7.0, and 100 mM KCl. RNA and DNA samples were annealed using a heating block, and samples were put into the block for 95 °C for 5 min followed by gradual cooling to room temperature and left overnight.

Circular Dichroism and UV−Vis Spectroscopy. CD experiments were performed with the JASCO 1500 spectropolarimeter (JASCO, Japan) under a constant flow of nitrogen. All measurements were done at 25 °C with a 1 mm quartz cuvette and covering a

https://doi.org/10.1021/jacs.1c11248
J. Am. Chem. Soc. 2022, 144, 935−950

spectral range of 200−650 nm using a scan rate of 200 nm/min and with a 1 s response time and 1 nm bandwidth. Each spectrum was obtained as an average of four measurements and was buffer/ligand subtracted, zero-corrected at 320 nm, and smoothed using a Savitsky-Golay 5 point window. During titrations with TMPyP4 spectra were taken immediately after addition of TMPyP4. Unfolding data were fitted using the Hill equation $\theta = C^n/(DC_{50}{}^n + C^n)$ where θ = fraction of bound TMPyP4, C = concentration of TMPyP4, n = Hill coefficient, and DC_{50} = the half degrading concentration.

Corresponding UV−vis absorption data were also collected at the same time as the CD. UV−vis binding data were fitted to a two-inequivalent-sites binding model: $\theta = (K_1C + 2K_1K_2C^2)/(1 + K_1C + 2K_1K_2C^2)$ where θ = fraction of bound TMPyP4, C = concentration of TMPyP4, and K_1 and K_2 are the association constants for the first and second sites.

CD/UV−vis melting/annealing profiles were recorded by monitoring the CD at 264 nm, as a function of temperature using a sealed cuvette. RNA samples were cooled to 5 °C and then heated to 95 °C with a heating rate of 1 °C/min, held at 95 °C for 5 min, and then cooled to 5 °C at the same rate. The T_m values of the complexes were calculated by normalizing the experimental curves to give fraction folded, and the data fitted to sigmoidal fittings were used to determine the T_m values. All experiments were repeated twice, and the error represents the standard deviation from the mean. Final data were analyzed in OriginPro 2020.

UV−Vis Spectroscopy. Initially, 150 μL solutions of the blank buffer and the ligand sample (4 μM) were placed in the reference and sample cuvettes (1.0 cm path length), respectively, and then the first spectrum was recorded in the range of 200−600 nm. During the titration, an aliquot of buffered RNA solution was added to the cuvette. Complex solutions were incubated for 5 min before absorption spectra were recorded. The absorption spectra were recorded on a UV1800 spectrophotometer (Shimadzu Technologies, Japan) at 25 °C.

Förster Resonance Energy Transfer (FRET) Experiments. FRET titration experiments were performed in an Edinburgh Instruments FS5 spectrofluorometer, using a quartz cuvette with a 10 mm path length. The sample volume was 250 μL. The FRET-labeled RNA was diluted in pH 7.0 buffer to 0.1 μM. The sample was excited at 490 nm, and the fluorescence emission was measured from 500 to 650 nm. TMPyP4 was diluted in buffer and added into a sample in 0.025 μM increments from 0.025 to 0.5 μM. The relative FRET efficiency (E_{FRET}) was calculated using $E_{FRET} = (I_a/I_d + I_a)$ where I_d is the fluorescence intensity of the donor and I_a is the fluorescence intensity of the acceptor. The experiment was performed in triplicate, and the error bars represent the standard deviation. The data were analyzed by OriginPro 2020.

Isothermal Titration Calorimetry. ITC experiments were performed on an Auto-iTC100 titration calorimeter (MicroCal) at 25 °C. All solutions (buffer, RNA, and ligand) were degassed before carrying out the experiments. A 20 μM solution of RNA PQS18-1 was placed into the cell, and 500 μM TMPyP4 was taken in the rotating syringe (750 rpm). A total of 40 μL of TMPyP4 was added to the RNA in 20 injections, and the time gap for two injections was 150 s, while the first injection was 0.4 μL to account for diffusion from the syringe into the cell during equilibration. This initial injection was not used in fitting the data. Similarly, dilution experiments were also carried out taking buffer (10 mM K_2HPO_4/KH_2PO_4, pH 7.0, and 100 mM KCl) in the cell, and TMPyP4 was kept in the syringe. We have used Microcal ITC analysis software for the analysis of the ITC raw data using the two-site-binding model. We have subtracted the dilution data from the raw data of the interaction of TMPyP4 with RNA before analysis. After fitting these experimental data, the enthalpy change during the process was obtained.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.1c11248.

PC analysis of bound states, native RNA, top-face RNA−TMPyP4 bound state, and groove site RNA−TMPyP4 bound state; description of the free energy surface; circular dichroism melting; 1D PMF; FES of native RNA G4 topology, top-face TMPyP4−RNA, groove-bound TMPyP4−RNA complex, most populated clusters of the RNA−TMPyP4 complex; construction of CVs; MD simulation statistics; Job plot; CD melting/annealing; UV melting/annealing; UV−vis titration; fluorescence titrations; details of CVs (PDF)

Movie 1: TMPyP4 binding to the RNA G-Quadruplex (MP4)

Movie 2: Unfolding of the RNA G-Quadruplex (MP4)

■ AUTHOR INFORMATION

Corresponding Authors

Dengguo Wei − *State Key Laboratory of Agricultural Microbiology, College of Vetrinary Medicine, Huazhong Agricultural University, Wuhan 430070, China; National Reference Laboratory of Veterinary Drug Residues and MAO Key Laboratory for Detection of Veterinary Drug Residues, Hubei Hongshan Laboratory, and Interdisciplinary Sciences Institute, Huazhong Agricultural University, Wuhan 430070, China;* orcid.org/0000-0001-9923-789X; Email: dgwei@mail.hzau.edu.cn

Shozeb Haider − *UCL School of Pharmacy, University College London, London WC1N 1AX, U.K.; UCL Centre for Advanced Research Computing, University College London, London WC1H 9RN, U.K.;* orcid.org/0000-0003-2650-2925; Email: shozeb.haider@ucl.ac.uk

Authors

Susanta Haldar − *School of Chemistry, University of Bristol, Bristol BS8 1TS, U.K.; D.E. Shaw India Private Ltd., Hyderabad, Telangana 500096, India*

Yashu Zhang − *State Key Laboratory of Agricultural Microbiology, College of Vetrinary Medicine, Huazhong Agricultural University, Wuhan 430070, China; College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China*

Ying Xia − *UCL School of Pharmacy, University College London, London WC1N 1AX, U.K.*

Barira Islam − *Department of Pharmacy, School of Applied Sciences, University of Huddersfield, Huddersfield HD1 3DH, U.K.;* orcid.org/0000-0001-5882-6903

Sisi Liu − *College of Science, Huazhong Agricultural University, Wuhan 430070, China;* orcid.org/0000-0002-6814-134X

Francesco L. Gervasio − *Department of Chemistry, University College London, London WC1H 0AJ, U.K.; Pharmaceutical Sciences, University of Geneva, Geneva CH-1211, Switzerland; Institute of Pharmaceutical Sciences of Western Switzerland (ISPSO), Geneva CH-1211, Switzerland;* orcid.org/0000-0003-4831-5039

Adrian J. Mulholland − *School of Chemistry, University of Bristol, Bristol BS8 1TS, U.K.;* orcid.org/0000-0003-1015-4567

Zoë A. E. Waller − *UCL School of Pharmacy, University College London, London WC1N 1AX, U.K.;* orcid.org/0000-0001-8538-0484

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.1c11248

Author Contributions

[#]S.H., Y.Z., and Y.X. are joint first authors.

Notes

The authors declare no competing financial interest.

The unbiased simulation trajectories have been uploaded to zenodo (doi 10.5281/zenodo.5594466). Metadynamics simulation trajectories can be obtained from the authors upon request.

■ ACKNOWLEDGMENTS

D.G.W. is supported by the National Natural Science Foundation of China (21732002, 22077043, 31672558). A.J.M. and S.H. would like to acknowledge the EPSRC grants for this study (EP/N024117/1 and EP/M022609/1). S.H. would like to thank Dr. Gary N. Parkinson for critical reading of the manuscript. D.G.W would like to thank Chen Dong and Qi Zhou.

■ REFERENCES

(1) Gellert, M.; Lipsett, M. N.; Davies, D. R. Helix Formation by Guanylic Acid. *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48* (12), 2013−2018.

(2) Laughlan, G.; Murchie, A.; Norman, D.; Moore, M.; Moody, P.; Lilley, D.; Luisi, B. The High-Resolution Crystal Structure of a Parallel-Stranded Guanine Tetraplex. *Science* **1994**, *265* (5171), 520−524.

(3) Gilli, G.; Bellucci, F.; Ferretti, V.; Bertolasi, V. Evidence for Resonance-Assisted Hydrogen Bonding from Crystal-Structure Correlations on the Enol Form of the. Beta.-Diketone Fragment. *J. Am. Chem. Soc.* **1989**, *111* (3), 1023−1028.

(4) Fonseca Guerra, C.; Zijlstra, H.; Paragi, G.; Bickelhaupt, F. M. Telomere Structure and Stability: Covalency in Hydrogen Bonds, Not Resonance Assistance, Causes Cooperativity in Guanine Quartets. *Chem.−Eur. J.* **2011**, *17* (45), 12612−12622.

(5) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.* **2006**, *34* (19), 5402−5415.

(6) 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.

(7) Lightfoot, H. L.; Hagen, T.; Tatum, N. J.; Hall, J. The Diverse Structural Landscape of Quadruplexes. *FEBS Lett.* **2019**, *593* (16), 2083−2102.

(8) Lyu, K.; Chow, E. Y.-C.; Mou, X.; Chan, T.-F.; Kwok, C. K. RNA G-Quadruplexes (RG4s): Genomics and Biological Functions. *Nucleic Acids Res.* **2021**, *49* (10), 5426−5450.

(9) Collie, G. W.; Haider, S. M.; Neidle, S.; Parkinson, G. N. A Crystallographic and Modelling Study of a Human Telomeric RNA (TERRA) Quadruplex. *Nucleic Acids Res.* **2010**, *38* (16), 5569−5580.

(10) Kwok, C. K.; Merrick, C. J. G-Quadruplexes: Prediction, Characterization, and Biological Application. *Trends Biotechnol.* **2017**, *35* (10), 997−1013.

(11) 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**, *33* (8), 877−881.

(12) Yang, S. Y.; Lejault, P.; Chevrier, S.; Boidot, R.; Robertson, A. G.; Wong, J. M. Y.; Monchaud, D. Transcriptome-Wide Identification of Transient RNA G-Quadruplexes in Human Cells. *Nat. Commun.* **2018**, *9* (1), 4730−4741.

(13) Kwok, C. K.; Marsico, G.; Sahakyan, A. B.; Chambers, V. S.; Balasubramanian, S. RG4-Seq Reveals Widespread Formation of G-Quadruplex Structures in the Human Transcriptome. *Nat. Methods* **2016**, *13* (10), 841−844.

(14) Huppert, J. L.; Balasubramanian, S. G-Quadruplexes in Promoters throughout the Human Genome. *Nucleic Acids Res.* **2007**, *35* (2), 406−413.

(15) Bochman, M. L.; Paeschke, K.; Zakian, V. A. DNA Secondary Structures: Stability and Function of G-Quadruplex Structures. *Nat. Rev. Genet* **2012**, *13* (11), 770−780.

(16) Todd, A. K.; Johnston, M.; Neidle, S. Highly Prevalent Putative Quadruplex Sequence Motifs in Human DNA. *Nucleic Acids Res.* **2005**, *33* (9), 2901−2907.

(17) Lipps, H. J.; Rhodes, D. G-Quadruplex Structures: In Vivo Evidence and Function. *Trends in Cell Biology* **2009**, *19* (8), 414−422.

(18) Maizels, N.; Gray, L. T. The G4 Genome. *PLoS Genet* **2013**, *9* (4), 1−6.

(19) Hewitt, G.; Jurk, D.; Marques, F. D. M.; Correia-Melo, C.; Hardy, T.; Gackowska, A.; Anderson, R.; Taschuk, M.; Mann, J.; Passos, J. F. Telomeres Are Favoured Targets of a Persistent DNA Damage Response in Ageing and Stress-Induced Senescence. *Nat. Commun.* **2012**, *3* (1), 708−718.

(20) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal Structure of Parallel Quadruplexes from Human Telomeric DNA. *Nature* **2002**, *417* (6891), 876−880.

(21) Besnard, E.; Babled, A.; Lapasset, L.; Milhavet, O.; Parrinello, H.; Dantec, C.; Marin, J.-M.; Lemaitre, J.-M. Unraveling Cell Type−Specific and Reprogrammable Human Replication Origin Signatures Associated with G-Quadruplex Consensus Motifs. *Nat. Struct Mol. Biol.* **2012**, *19* (8), 837−844.

(22) Tian, T.; Chen, Y.-Q.; Wang, S.-R.; Zhou, X. G-Quadruplex: A Regulator of Gene Expression and Its Chemical Targeting. *Chem.* **2018**, *4* (6), 1314−1344.

(23) Rhodes, D.; Lipps, H. J. G-Quadruplexes and Their Regulatory Roles in Biology. *Nucleic Acids Res.* **2015**, *43* (18), 8627−8637.

(24) De, S.; Michor, F. DNA Secondary Structures and Epigenetic Determinants of Cancer Genome Evolution. *Nat. Struct Mol. Biol.* **2011**, *18* (8), 950−955.

(25) Pietras, Z.; Wojcik, M. A.; Borowski, L. S.; Szewczyk, M.; Kulinski, T. M.; Cysewski, D.; Stepien, P. P.; Dziembowski, A.; Szczesny, R. J. Dedicated Surveillance Mechanism Controls G-Quadruplex Forming Non-Coding RNAs in Human Mitochondria. *Nat. Commun.* **2018**, *9* (1), 2558−2573.

(26) Fay, M. M.; Anderson, P. J.; Ivanov, P. ALS/FTD-Associated C9ORF72 Repeat RNA Promotes Phase Transitions In Vitro and in Cells. *Cell Reports* **2017**, *21* (12), 3573−3584.

(27) Jain, A.; Vale, R. D. RNA Phase Transitions in Repeat Expansion Disorders. *Nature* **2017**, *546* (7657), 243−247.

(28) Roundtree, I. A.; Evans, M. E.; Pan, T.; He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **2017**, *169* (7), 1187−1200.

(29) Kharel, P.; Becker, G.; Tsvetkov, V.; Ivanov, P. Properties and Biological Impact of RNA G-Quadruplexes: From Order to Turmoil and Back. *Nucleic Acids Res.* **2020**, *48* (22), 12534−12555.

(30) Bugaut, A.; Balasubramanian, S. 5′-UTR RNA G-Quadruplexes: Translation Regulation and Targeting. *Nucleic Acids Res.* **2012**, *40* (11), 4727−4741.

(31) Monchaud, D. Quadruplex Detection in Human Cells. In *Quadruplex nucleic acids as targets for medicinal chemistry*; Annual Reports in Medicinal Chemistry; 2020; Vol. *54*, pp 133−161.

(32) Lejault, P.; Mitteaux, J.; Sperti, F. R.; Monchaud, D. How to Untie G-Quadruplex Knots and Why? *Cell Chemical Biology* **2021**, *28* (4), 436−455.

(33) Beaume, N.; Pathak, R.; Yadav, V. K.; Kota, S.; Misra, H. S.; Gautam, H. K.; Chowdhury, S. Genome-Wide Study Predicts Promoter-G4 DNA Motifs Regulate Selective Functions in Bacteria: Radioresistance of D. Radiodurans Involves G4 DNA-Mediated Regulation. *Nucleic Acids Res.* **2013**, *41* (1), 76−89.

(34) Abiri, A.; Lavigne, M.; Rezaei, M.; Nikzad, S.; Zare, P.; Mergny, J.-L.; Rahimi, H.-R. Unlocking G-Quadruplexes as Antiviral Targets. *Pharmacol Rev.* **2021**, *73* (3), 897−923.

(35) Ruggiero, E.; Richter, S. N. G-Quadruplexes and G-Quadruplex Ligands: Targets and Tools in Antiviral Therapy. *Nucleic Acids Res.* **2018**, *46* (7), 3270−3283.

(36) 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

https://doi.org/10.1021/jacs.1c11248
J. Am. Chem. Soc. 2022, 144, 935−950

Human Telomerase by a G-Quadruplex-Interactive Compound. *J. Med. Chem.* **1997**, *40* (14), 2113−2116.
(37) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (18), 11593−11598.
(38) Monchaud, D.; Teulade-Fichou, M.-P. A Hitchhiker's Guide to G-Quadruplex Ligands. *Org. Biomol. Chem.* **2008**, *6* (4), 627−636.
(39) Neidle, S. Quadruplex Nucleic Acids as Novel Therapeutic Targets. *J. Med. Chem.* **2016**, *59* (13), 5987−6011.
(40) Neidle, S. Quadruplex Nucleic Acids as Targets for Anticancer Therapeutics. *Nat. Rev. Chem.* **2017**, *1* (5), 1−10.
(41) Ofer, N.; Weisman-Shomer, P.; Shklover, J.; Fry, M. The Quadruplex r(CGG)n Destabilizing Cationic Porphyrin TMPyP4 Cooperates with HnRNPs to Increase the Translation Efficiency of Fragile X Premutation MRNA. *Nucleic Acids Res.* **2009**, *37* (8), 2712−2722.
(42) Weisman-Shomer, P. The Cationic Porphyrin TMPyP4 Destabilizes the Tetraplex Form of the Fragile X Syndrome Expanded Sequence d(CGG)n. *Nucleic Acids Res.* **2003**, *31* (14), 3963−3970.
(43) Moruno-Manchon, J. F.; Lejault, P.; Wang, Y.; McCauley, B.; Honarpisheh, P.; Morales Scheihing, D. A.; Singh, S.; Dang, W.; Kim, N.; Urayama, A.; Zhu, L.; Monchaud, D.; McCullough, L. D.; Tsvetkov, A. S. Small-Molecule G-Quadruplex Stabilizers Reveal a Novel Pathway of Autophagy Regulation in Neurons. *eLife* **2020**, *9*, 1−26.
(44) Mitteaux, J.; Lejault, P.; Wojciechowski, F.; Joubert, A.; Boudon, J.; Desbois, N.; Gros, C. P.; Hudson, R. H. E.; Boulé, J.-B.; Granzhan, A.; Monchaud, D. Identifying G-Quadruplex-DNA-Disrupting Small Molecules. *J. Am. Chem. Soc.* **2021**, *143* (32), 12567−12577.
(45) Morris, M. J.; Wingate, K. L.; Silwal, J.; Leeper, T. C.; Basu, S. The Porphyrin TmPyP4 Unfolds the Extremely Stable G-Quadruplex in MT3-MMP MRNA and Alleviates Its Repressive Effect to Enhance Translation in Eukaryotic Cells. *Nucleic Acids Res.* **2012**, *40* (9), 4137−4145.
(46) Waller, Z. A. E.; Sewitz, S. A.; Hsu, S.-T. D.; Balasubramanian, S. A Small Molecule That Disrupts G-Quadruplex DNA Structure and Enhances Gene Expression. *J. Am. Chem. Soc.* **2009**, *131* (35), 12628−12633.
(47) Kaluzhny, D.; Ilyinsky, N.; Shchekotikhin, A.; Sinkevich, Y.; Tsvetkov, P. O.; Tsvetkov, V.; Veselovsky, A.; Livshits, M.; Borisova, O.; Shtil, A.; Shchyolkina, A. Disordering of Human Telomeric G-Quadruplex with Novel Antiproliferative Anthrathiophenedione. *PLoS One* **2011**, *6* (11), 1−6.
(48) O'Hagan, M. P.; Haldar, S.; Duchi, M.; Oliver, T. A. A.; Mulholland, A. J.; Morales, J. C.; Galan, M. C. A Photoresponsive Stiff-Stilbene Ligand Fuels the Reversible Unfolding of G-Quadruplex DNA. *Angew. Chem., Int. Ed.* **2019**, *58* (13), 4334−4338.
(49) Zamiri, B.; Reddy, K.; Macgregor, R. B.; Pearson, C. E. TMPyP4 Porphyrin Distorts RNA G-Quadruplex Structures of the Disease-Associated r(GGGGCC)n Repeat of the C9orf72 Gene and Blocks Interaction of RNA-Binding Proteins. *J. Biol. Chem.* **2014**, *289* (8), 4653−4659.
(50) Liu, H.; Zheng, K.; He, Y.; Chen, Q.; Hao, Y.; Tan, Z. RNA G-Quadruplex Formation in Defined Sequence in Living Cells Detected by Bimolecular Fluorescence Complementation. *Chem. Sci.* **2016**, *7* (7), 4573−4581.
(51) Ohnmacht, S. A.; Marchetti, C.; Gunaratnam, M.; Besser, R. J.; Haider, S. M.; Di Vita, G.; Lowe, H. L.; Mellinas-Gomez, M.; Diocou, S.; Robson, M.; Šponer, J.; Islam, B.; Barbara Pedley, R.; Hartley, J. A.; Neidle, S. A G-Quadruplex-Binding Compound Showing Anti-Tumour Activity in an in Vivo Model for Pancreatic Cancer. *Sci. Rep* **2015**, *5* (1), 11385−11396.
(52) 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. Chem., Int. Ed.* **2017**, *56* (25), 7102−7106.
(53) Haider, S. M.; Parkinson, G. N.; Neidle, S. Structure of a G-Quadruplex−Ligand Complex. *J. Mol. Biol.* **2003**, *326* (1), 117−125.
(54) Collie, G. W.; Promontorio, R.; Hampel, S. M.; Micco, M.; Neidle, S.; Parkinson, G. N. Structural Basis for Telomeric G-Quadruplex Targeting by Naphthalene Diimide Ligands. *J. Am. Chem. Soc.* **2012**, *134* (5), 2723−2731.
(55) Liu, W.; Lin, C.; Wu, G.; Dai, J.; Chang, T.-C.; Yang, D. Structures of 1:1 and 2:1 Complexes of BMVC and MYC Promoter G-Quadruplex Reveal a Mechanism of Ligand Conformation Adjustment for G4-Recognition. *Nucleic Acids Res.* **2019**, *46* (22), 11931−11942.
(56) Chung, W. J.; Heddi, B.; Hamon, F.; Teulade-Fichou, M.-P.; Phan, A. T. Solution Structure of a G-Quadruplex Bound to the Bisquinolinium Compound Phen-DC $_3$. *Angew. Chem., Int. Ed.* **2014**, *53* (4), 999−1002.
(57) Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. Small-Molecule Interaction with a Five-Guanine-Tract G-Quadruplex Structure from the Human MYC Promoter. *Nat. Chem. Biol.* **2005**, *1* (3), 167−173.
(58) Chen, M. C.; Tippana, R.; Demeshkina, N. A.; Murat, P.; Balasubramanian, S.; Myong, S.; Ferré-D'Amaré, A. R. Structural Basis of G-Quadruplex Unfolding by the DEAH/RHA Helicase DHX36. *Nature* **2018**, *558* (7710), 465−469.
(59) Moraca, F.; Amato, J.; Ortuso, F.; Artese, A.; Pagano, B.; Novellino, E.; Alcaro, S.; Parrinello, M.; Limongelli, V. Ligand Binding to Telomeric G-Quadruplex DNA Investigated by Funnel-Metadynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (11), E2136−E2145.
(60) Rha, S. Y.; Izbicka, E.; Lawrence, R.; Davidson, K.; Sun, D.; Moyer, M. P.; Roodman, G. D.; Hurley, L.; Von Hoff, D. Effect of Telomere and Telomerase Interactive Agents on Human Tumor and Normal Cell Lines. *Clin. Cancer Res.* **2000**, *6* (3), 987−993.
(61) Grand, C. L.; Han, H.; Muñoz, R. M.; Weitman, S.; Von Hoff, D. D.; Hurley, L. H.; Bearss, D. J. The Cationic Porphyrin TMPyP4 Down-Regulates c-MYC and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth in Vivo. *Mol. Cancer Ther.* **2002**, *1* (8), 565−573.
(62) Artusi, S.; Ruggiero, E.; Nadai, M.; Tosoni, B.; Perrone, R.; Ferino, A.; Zanin, I.; Xodo, L.; Flamand, L.; Richter, S. N. Antiviral Activity of the G-Quadruplex Ligand TMPyP4 against Herpes Simplex Virus-1. *Viruses* **2021**, *13* (2), 196−215.
(63) Parkinson, G. N.; Ghosh, R.; Neidle, S. Structural Basis for Binding of Porphyrin to Human Telomeres. *Biochemistry* **2007**, *46* (9), 2390−2397.
(64) Wang, S.-R.; Zhang, Q.-Y.; Wang, J.-Q.; Ge, X.-Y.; Song, Y.-Y.; Wang, Y.-F.; Li, X.-D.; Fu, B.-S.; Xu, G.-H.; Shu, B.; Gong, P.; Zhang, B.; Tian, T.; Zhou, X. Chemical Targeting of a G-Quadruplex RNA in the Ebola Virus L Gene. *Cell Chemical Biology* **2016**, *23* (9), 1113−1122.
(65) Zhang, Y.; Liu, S.; Jiang, H.; Deng, H.; Dong, C.; Shen, W.; Chen, H.; Gao, C.; Xiao, S.; Liu, Z.-F.; Wei, D. G-Quadruplex in the 3′UTR of IE180 Regulates Pseudorabies Virus Replication by Enhancing Gene Expression. *RNA Biology* **2020**, *17* (6), 816−827.
(66) Zhang, Y.; El Omari, K.; Duman, R.; Liu, S.; Haider, S.; Wagner, A.; Parkinson, G. N.; Wei, D. Native de Novo Structural Determinations of Non-Canonical Nucleic Acid Motifs by X-Ray Crystallography at Long Wavelengths. *Nucleic Acids Res.* **2020**, *48* (17), 9886−9898.
(67) Saladino, G.; Gauthier, L.; Bianciotto, M.; Gervasio, F. L. Assessing the Performance of Metadynamics and Path Variables in Predicting the Binding Free Energies of P38 Inhibitors. *J. Chem. Theory Comput.* **2012**, *8* (4), 1165−1170.
(68) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations. *PLoS Comput. Biol.* **2009**, *5* (8), 1−18.
(69) Valsson, O.; Tiwary, P.; Parrinello, M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annu. Rev. Phys. Chem.* **2016**, *67* (1), 159−184.

(70) Kührová, P.; Best, R. B.; Bottaro, S.; Bussi, G.; Šponer, J.; Otyepka, M.; Banáš, P. Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *J. Chem. Theory Comput.* **2016**, *12* (9), 4534−4548.
(71) Haldar, S.; Kührová, P.; Banáš, P.; Spiwok, V.; Šponer, J.; Hobza, P.; Otyepka, M. Insights into Stability and Folding of GNRA and UNCG Tetraloops Revealed by Microsecond Molecular Dynamics and Well-Tempered Metadynamics. *J. Chem. Theory Comput.* **2015**, *11* (8), 3866−3877.
(72) Di Leva, F. S.; Novellino, E.; Cavalli, A.; Parrinello, M.; Limongelli, V. Mechanistic Insight into Ligand Binding to G-Quadruplex DNA. *Nucleic Acids Res.* **2014**, *42* (9), 5447−5455.
(73) Haldar, S.; Spiwok, V.; Hobza, P. On the Association of the Base Pairs on the Silica Surface Based on Free Energy Biased Molecular Dynamics Simulation and Quantum Mechanical Calculations. *J. Phys. Chem. C* **2013**, *117* (21), 11066−11075.
(74) Miriyala, V. M.; Lo, R.; Haldar, S.; Sarmah, A.; Hobza, P. Structure and Properties of Double-Sandwich Complexes at the Graphene Surface: A Theoretical Study. *J. Phys. Chem. C* **2019**, *123* (23), 14712−14724.
(75) O'Hagan, M. P.; Haldar, S.; Morales, J. C.; Mulholland, A. J.; Galan, M. C. Enhanced Sampling Molecular Dynamics Simulations Correctly Predict the Diverse Activities of a Series of Stiff-Stilbene G-Quadruplex DNA Ligands. *Chem. Sci.* **2021**, *12* (4), 1415−1426.
(76) del Villar-Guerra, R.; Trent, J. O.; Chaires, J. B. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew. Chem., Int. Ed.* **2018**, *57* (24), 7171−7175.
(77) Murat, P.; Singh, Y.; Defrancq, E. Methods for Investigating G-Quadruplex DNA/Ligand Interactions. *Chem. Soc. Rev.* **2011**, *40* (11), 5293−5307.
(78) 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**, *8* (8), 1−9.
(79) Perrone, R.; Nadai, M.; Poe, J. A.; Frasson, I.; Palumbo, M.; Palù, G.; Smithgall, T. E.; Richter, S. N. Formation of a Unique Cluster of G-Quadruplex Structures in the HIV-1 Nef Coding Region: Implications for Antiviral Activity. *PLoS One* **2013**, *8* (8), 1−14.
(80) Chilakamarthi, U.; Koteshwar, D.; Jinka, S.; Vamsi Krishna, N.; Sridharan, K.; Nagesh, N.; Giribabu, L. Novel Amphiphilic G-Quadruplex Binding Synthetic Derivative of TMPyP4 and Its Effect on Cancer Cell Proliferation and Apoptosis Induction. *Biochemistry* **2018**, *57* (46), 6514−6527.
(81) Garbett, N. C.; Ragazzon, P. A.; Chaires, J. B. Circular Dichroism to Determine Binding Mode and Affinity of Ligand−DNA Interactions. *Nat. Protoc* **2007**, *2* (12), 3166−3172.
(82) Pradeep, T. P.; Tripathi, S.; Barthwal, R. Molecular Recognition of Parallel Quadruplex [d-(TTGGGGT)] $_4$ by Mitoxantrone: Binding with 1:4 Stoichiometry Leads to Telomerase Inhibition. *RSC Adv.* **2016**, *6* (75), 71652−71661.
(83) White, E. W.; Tanious, F.; Ismail, M. A.; Reszka, A. P.; Neidle, S.; Boykin, D. W.; Wilson, W. D. Structure-Specific Recognition of Quadruplex DNA by Organic Cations: Influence of Shape, Substituents and Charge. *Biophys. Chem.* **2007**, *126* (1−3), 140−153.
(84) Gray, R. D.; Trent, J. O.; Chaires, J. B. Folding and Unfolding Pathways of the Human Telomeric G-Quadruplex. *J. Mol. Biol.* **2014**, *426* (8), 1629−1650.
(85) Zhang, J.; Chen, M. Unfolding Hidden Barriers by Active Enhanced Sampling. *Phys. Rev. Lett.* **2018**, *121* (1), 010601−010605.
(86) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562−12566.
(87) Martino, L.; Pagano, B.; Fotticchia, I.; Neidle, S.; Giancola, C. Shedding Light on the Interaction between TMPyP4 and Human Telomeric Quadruplexes. *J. Phys. Chem. B* **2009**, *113* (44), 14779−14786.
(88) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal Of Chemical Theory And Computation* **2008**, *4* (3), 435−447.
(89) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.
(90) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M. Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2016**, *13* (1), 55−58.
(91) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179−5197.
(92) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886−2902.
(93) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92* (11), 3817−3829.
(94) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157−1174.
(95) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J. Am. Chem. Soc.* **1993**, *115* (21), 9620−9631.
(96) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97* (40), 10269−10280.
(97) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *Journal of Molecular Graphics and Modelling* **2006**, *25* (2), 247−260.
(98) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97* (40), 10269−10280.
(99) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.
(100) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112* (30), 9020−9041.
(101) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101−014107.
(102) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684−3690.
(103) Evans, D. J.; Holian, B. L. The Nose−Hoover Thermostat. *J. Chem. Phys.* **1985**, *83* (8), 4069−4074.
(104) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method Polymorphic Transitions in Single Crystals. *J. Appl. Phys.* **1981**, *52* (1981), 7182−7190.
(105) Wells, B. A.; Chaffee, A. L. Ewald Summation for Molecular Simulations. *J. Chem. Theory Comput.* **2015**, *11* (8), 3684−3695.
(106) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.
(107) Bussi, G.; Laio, A. Using Metadynamics to Explore Complex Free-Energy Landscapes. *Nat. Rev. Phys.* **2020**, *2* (4), 200−212.

https://doi.org/10.1021/jacs.1c11248
J. Am. Chem. Soc. 2022, 144, 935−950

(108) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180* (10), 1961−1972.
(109) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604−613.
(110) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α-Ketoamide Inhibitors. *Science* **2020**, *368* (6489), 409−412.
(111) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph* **1996**, *14* (1), 33−38.

https://doi.org/10.1021/jacs.1c11248
J. Am. Chem. Soc. 2022, 144, 935−950

Dalton Transactions

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Dalton Trans.*, 2021, **50**, 16739

Precious metal complexes of bis(pyridyl)allenes: synthesis and catalytic and medicinal applications†

Hanna K. Maliszewska, [a] Carla Arnau del Valle, [a] Ying Xia,[b] María J. Marín, [a] Zoë A. E. Waller [b,c] and María Paz Muñoz *[a]

Received 31st August 2021,
Accepted 28th October 2021

DOI: 10.1039/d1dt02929k

rsc.li/dalton

The incorporation of donor-type substituents on the allene core opens up the possibility of coordination complexes in which the metal is bonded to the donor groups, with or without interaction with the double bond system. Despite the challenges in the synthesis of such allene-containing metal complexes, their unique 3D environments and dual functionality (allene and metal) could facilitate catalysis and interaction with chemical and biological systems. Bis(pyridyl)allenes are presented here as robust ligands for novel Pd(II), Pt(IV) and Au(III) complexes. Their synthesis, characterisation and first application as catalysts of benchmark reactions for Pd, Pt and Au are presented with interesting reactivity and selectivities. The complexes have also been probed as antimicrobial and anticancer agents with promising activities, and the first studies on their unusual interaction with several DNA structures will open new avenues for research in the area of metallodrugs with new mechanisms of action.

Introduction

The development of ligand architectures with novel 3D arrangements is an important challenge in organic synthesis[1–4] with implications not only in the field of catalysis (and asymmetric catalysis), but also in the discovery of new organometallic complexes with potential for biological applications, currently an area of increasing growth.[5–7]

In this context, organic molecules displaying axial chirality, like biaryls or spiranes, have proven to be privileged structures in catalysis and have revolutionised the field of asymmetric catalysis.[8–10] However, despite progress in the field, it is still necessary to develop new chiral ligands to overcome limitations of established catalytic systems in terms of substrate scope, catalytic loading, turnover or enantioselectivity. Similarly, in medicinal organometallic chemistry, irrelevant to the metal studied, the rational design of the ligand seems crucial to control physical properties and selectivity towards specific biological targets. In many examples, stabilisation of a metal complex is achieved by chelating the metal centre with multidentate ligands.[11]

In the context of novel ligand design, allenes[12] contain a backbone that goes beyond the tetrahedral chiral carbon or the chiral biaryl structures, creating a very attractive and unique 3D (axially chiral) environment that can be modulated and exploited in unprecedented applications.

Allene-containing natural products and synthetic molecules have been demonstrated to exhibit interesting biological activity, although little is known about their mode of action. Consequently, some efforts have been made in recent years to incorporate the allene moiety in the backbone of pharmacologically active compounds, including natural and non-natural products, to increase their interaction with biological targets, *e.g.* as enzymatic inhibitors.[13]

Besides, the use of organometallic and coordination metal complexes as drug candidates has some key advantages over classic organic compounds.[14] Recently, it has been reported that metal-containing compounds are statistically more likely to be active antimicrobial and antifungal agents than purely organic compounds (27% *vs.* 2% of compounds showing activity in each group, respectively).[15]

Moving away from the more documented class of η^2 allene–metal complexes where the metal is solely bonded to one of the allene's double bonds (Fig. 1a),[16–19] the incorporation of donor-type substituents on the allene core opens the possibility of coordination complexes in which the metal is bonded to the donor groups, with or without interaction with the double bond system (LB = Lewis base, Fig. 1b).[20–24]

[a]*School of Chemistry, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK. E-mail: m.munoz-herranz@uea.ac.uk*
[b]*School of Pharmacy, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK*
[c]*UCL School of Pharmacy, 29-39 Brunswick Square, London, WC1N 1AX, UK*

† Electronic supplementary information (ESI) available: Synthesis and characterisation of ligands and metal complexes, catalytic and electrochemical experiments, spectral data, X-ray details for complex **2b**, experiments of stability of **3a–b** in solution, experimental details of the antimicrobial, anticancer and interaction with DNA studies. CCDC 2094617. For ESI and crystallographic data in CIF or other electronic format see DOI: 10.1039/d1dt02929k

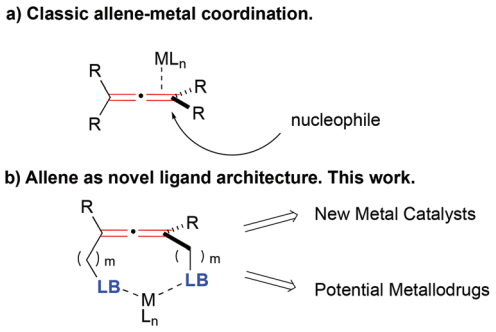


Fig. 1 Different types of interaction of allene frameworks with metals.

Creation of such allene containing-metal complexes presents significant synthetic challenges, *e.g.* competing coordination of the metal to the double bond and the donor atoms, and the control of nucleophilic attack of one donor atom to the allene activated by the metal that would destroy the allene backbone. Despite these inherent challenges, this design offers many advantages: the unique stereochemistry of the allene backbone creates novel 3D environments that could facilitate catalysis and asymmetric induction in chemical and biological systems; the multiple sites of coordination could increase the stability of the new metal species; and the presence of the metal in close proximity to the allene core can potentially activate the allene as a Michael acceptor to react with nucleophilic residues of enzymes, proteins or nucleic acids, giving these new metallodrugs alternative and/or synergistic modes of action, unique from current organometallic drug molecules. These characteristics could be key to the development of new catalysts and drugs. Therefore, allene-containing metal complexes have the potential to be exploited for their unique 3D shapes for studies in both catalytic and medicinal settings.

Despite some initial promising results using allene-containing phosphines as ligands in Rh(I) (Fig. 2a) and Au(I) (Fig. 2b, bottom) catalysis, reported by Ready[23] and Fensterbank[24] respectively, the Pd(II) and Pt(II) complexes of phosphine containing ligands (Fig. 2b, top)[24] have not been reported in any catalytic applications. Since these seminal works, there have not been further developments in the field, possibly due to the rapid oxidation of these systems when exposed to air.[25] Specifically, this makes phosphine analogues less attractive to use as ligands for complexes to be used in biological applications.[20,21,26]

In contrast, N-containing ligands such as bi- and ter-pyridines, have been widely used for the stabilisation of metal complexes by at least two chelating nitrogen donors, which lowers the reduction potential of the metal centre, making them more stable under physiological conditions.[11] This makes pyridyl groups stand out as especially attractive structures to incorporate in the allene ligand design.

In 2008, Krause and co-workers prepared bis(pyridyl)allene ligands and postulated the formation of their Cu(I) and Ag(I) complexes (Fig. 3a, top). However, not all the metallic structures were fully characterised and their use in any applications has not been reported since.[22] Later on, Fensterbank and co-workers observed that pyridyl-containing phosphine oxide (allenes) underwent nucleophilic intramolecular attack of the pyridine onto the allene to form cyclic metal species in the presence of Au(I) (Fig. 3a, bottom).[27] The difficulty in isolation and characterisation of pyridyl-based allene-containing metal complexes was further underpinned by our investigation into Au(I) and Au(III) coordination chemistry of bis(pyridyl)allenes, that also underwent cyclisation in the presence of these metal centres (Fig. 3a, bottom).[28]

To date applications of bis(pyridyl)allene-based metal systems remain totally unexplored, not only in the area of catalysis, but also in their use in a biological context. This latter application has also been completely unchartered in reported phosphine analogues, which gives this work additional signifi-

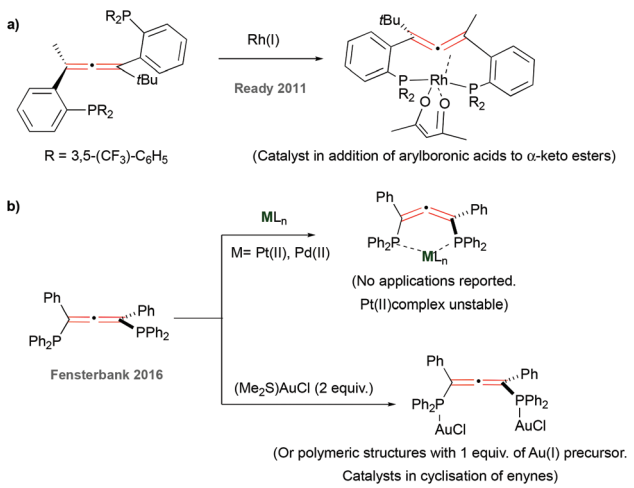


Fig. 2 Previous work with allene-containing phosphines as ligands.

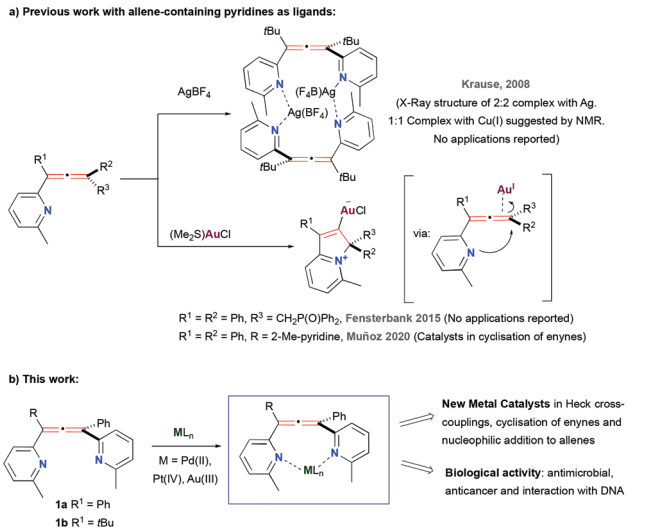


Fig. 3 Previous (a) and present (b) work with allene-containing pyridines as ligands.

cance. The many benefits and unexplored applications of these systems prompted us to explore their synthesis further. Here, we report the synthesis and characterisation of Pd(II), Pt (IV) and Au(III) bis(pyridyl)allene-containing complexes, and the explorative exploitation of their unique structures as catalysts and as potential metallodrugs in three settings, as antimicrobial agents, as anticancer drugs and their specific interaction with different structures of DNA. The preliminary data reported here will be crucial for further development of the complexes with applications in both areas (Fig. 3b, this work).

Results and discussion

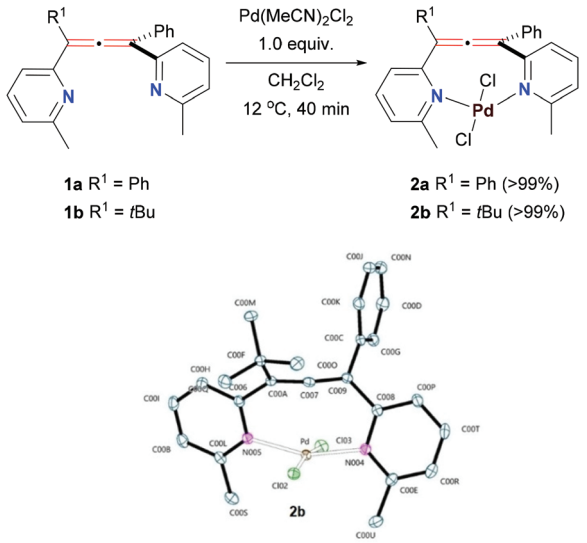
Synthesis of bis(pyridyl)allene metal complexes

We achieved the racemic synthesis of bis(pyridyl)allene ligands **1a–b** (Fig. 3b) according to modified literature procedures (see ESI,† section 2.1). These compounds are stable in solution (in chlorinated solvents, alcohols, DMSO, MeCN) and solid state, at room and elevated temperature, and under air, which highlights their robustness as promising ligands.

With the ligands in hand, we proceeded to expand the coordination properties of bis(pyridyl)allenes beyond the reported Cu(I) and Ag(I) complexes. The choice of metals for the complexation studies was two-fold. On the one hand, Pd, Pt and Au are versatile metals in important catalytic processes, so there is ample data to compare with the state-of-the art metal complexes in those transformations.[29–32] On the other hand, Pt and Au are one of the most common metals used in medicine,[33–35] and although Pd has been much less explored, there is emerging data that shows it might have good potential in interaction with biological systems.[36]

Starting with Pd(II) precursors, the ligands were reacted with 1 equiv. of $Pd(MeCN)_2Cl_2$ in dichloromethane. The reaction mixture changed colour from yellow to red over the course of 40 min, giving quantitative formation of new metal species (**2a–b**, Scheme 1, top).

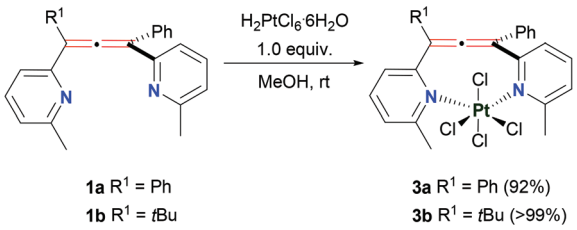
^{1}H and ^{13}C NMR analyses of **2a** revealed that the symmetry of the ligand was maintained in the complex, but the corresponding peaks were shifted. For instance, the methyl peak on the ^{1}H NMR spectrum moved from 2.58 ppm in **1a** to 3.17 ppm in **2a**, while the peak corresponding to the central allene carbon atom shifted up-field from 212.7 ppm to 204.1 ppm, respectively ($\Delta\delta$ = 8.6 ppm, see ESI,† section 2.2). These observations led us to propose the bidentate structure of **2a**, where Pd is captured by both pyridyl units and sits in the pocket created by the frame of the ligand, with the allene group itself not engaged in bonding to the metal centre. HRMS and EA analysis provided further evidence supporting the structure with 1 : 1 ligand to metal stoichiometry. Additionally, we obtained good quality crystals of the non-symmetric complex **2b**, and its structure was unambiguously confirmed by X-ray crystallography (Scheme 1, bottom). The structure **2b** clearly shows the undisturbed allene unit with the C(A)–C(7)–C(9) angle measured at 172.3°. Similarly, the typical double bond distances of the allene found at 1.321(2) Å and



Scheme 1 Synthesis of Pd(II)-complexes **2a–b** and ORTEP representation of the X-ray of complex **2b** (Hs omitted).

1.322(2) Å are indicative of a non-metal bonded allene group. The long distance of the Pd centre to the central carbon of the allene, 2.5038 (1) Å, confirms the Pd centre is not bonding to the central carbon of the allene. As we expected, the nitrogen atoms of both pyridines are facing inwards the ligand's pocket and are engaged in bonding to the Pd, with bond lengths of Pd–N(5) 2.0489(1) Å and Pd–N(4) 2.0269(1) Å. Overall, the perpendicular arrangement of the allene ligand is maintained and Pd is enclosed within. Complex **2b** features very slightly distorted square planar geometry typical for Pd(II) complexes. The coordination sphere is completed by two chloride ligands equally distanced from the metal centre (2.3300(4) Å and 2.3203(5) Å). Interestingly, **2b** is bonded to Pd in a *trans* arrangement. There are only few examples of such bidentate, *trans*-spanning ligands based on pyridyl groups reported to date.[37,38] Formation of *trans* isomers of complexes with these ligands might have beneficial consequences for their use in catalysis. For example, *trans*-spanning pyridyl ligands have been observed to be more active catalysts in Heck coupling reaction.[39] We could not obtain good enough crystals of the symmetric **2a** to perform the crystallographic analysis, so full assignment of the geometry (*cis* or *trans*) could not be confirmed. Although we initially assumed *trans* geometry of **2a**, based on the similarity of the remaining characterisation data, the differences in catalytic behaviour of both complexes in the Pd-catalysed cross-coupling reactions (*vide infra*) would suggest that **2a** could be predominantly *cis*.[40]

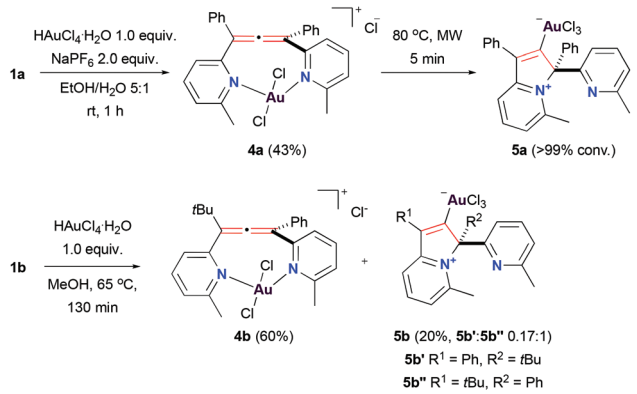
Interestingly, **1a–b** were completely unreactive towards Pt(II) salts when we tried to access the Pt(II) analogues of **2a–b** in a similar manner. Reactions with a variety of Pt precursors ($PtCl_2$, $Pt(MeCN)_2Cl_2$, K_2PtCl_4, *cis*-$Pt(DMSO)_2Cl_2$, *etc.*) under a range of conditions (*e.g.* solvents and temperature, see ESI,† section 2.2) usually resulted in the recovery of unreacted starting materials. We considered that the reaction with a smaller



1a R^1 = Ph
1b R^1 = *t*Bu

3a R^1 = Ph (92%)
3b R^1 = *t*Bu (>99%)

Scheme 2 Synthesis of Pt(IV)-complexes **3a–b**.



Scheme 3 Synthesis of Au(III)-complexes **4a–b** and **5a–b**.

Pt(IV) centre might be more feasible.[41] To test this, we reacted ligands **1a–b** with 1 equiv. of chloroplatinic(IV) acid (Scheme 2). We were happy to observe the immediate formation of orange powder products in both cases. These powders were the new Pt(IV) complexes **3a–b** with 1 : 1 ligand to metal stoichiometry, confirmed by ^{1}H and ^{13}C NMR, HRMS and EA analyses (see ESI,† section 2.2). Noteworthy, the NMR spectra of **3a** were still characterised by an increased symmetry indicative of a bidentate nature of the ligand. The data also suggested that the allene core did not participate in the metal complexation. Both complexes **3a–b** were only soluble in DMSO, with **3a** not being very stable in solution over long periods of time (see ESI,† section 5), which made obtaining good crystals for X-ray analysis difficult. In these cases, we proposed an octahedral geometry of the metal centre typical for Pt(IV) complexes and draw the structures as *cis* isomers. The *trans* isomer would require one of the chloride ligands to be placed directly in the space between the Pt atom and the centre of the allene or the distortion of the octahedral geometry. Although, the NMR data indicated that the allene environment did not substantially change in **3a–b** compared to the free ligands, we cannot completely rule out either isomer. In fact, the broadening of signals observed on the NMR spectra could be a reflection of the equilibrium between the *cis* and *trans* isomers in solution.[42]

The synthesis of the novel Pt(IV) bis(pyridyl)allene complexes brings a two-fold advantage. First of all, Pt(IV) complexes have been extensively studied as pro-drugs for anticancer therapy.[43–45] Pt(IV) compounds are considered as potential precursors to the active Pt(II) species and are designed to circumvent some of the undesired side effects of the latter. Hence, the cytotoxicity and redox chemistry of **3a–b** would be of much interest. Secondly, Pt(IV) compounds can be used as synthetic precursors to more challenging Pt(II) equivalents by means of chemical or electrochemical reduction. Unfortunately, our attempts to access direct Pt(II) analogues of **3a–b** *via* chemical reduction were unsuccessful (see ESI,† section 2.2). The more reactive **3b** underwent cycloplatination under reducing conditions losing the allene core in the process.[28] Inertness of **3a–b** under milder reduction conditions might be explained by their quite negative Pt(IV)/Pt(II) reduction potentials found at −1.0 V and −1.1 V, respectively for **3a** and **3b**, as measured by cyclic voltammetry of these complexes (see ESI,† section 2.2).

We explored the chemistry of bis(pyridyl)allenes with Au next. We previously reported that in the presence of Au(I) sources these ligands undergo cyclisation triggered by activation of the allene unit by the carbophilic Au(I) and attack of the pyridyl group onto the terminal carbon of the allene.[28] However, the different nature of the Au(III) centre allowed the access to unprecedented Au(III) complexes with the allene moiety intact (Scheme 3).

In this case, each of the ligands required slightly different reaction conditions to facilitate the formation of the allene-containing complex. Thus, **1a** reacted with 1 equiv. of chloroauric(III) acid in a mixture of ethanol and water at room temperature to give Au(III) complex **4a** in 43% yield. Attempts to improve the efficiency of the reaction, for example, by increasing the temperature, resulted in full conversion into the cyclic product **5a**, observed in our previous work.[28] On the other hand, reaction of **1b** in refluxing methanol yielded a mixture of allene-containing complex **4b** and mixture of isomers of the cyclised **5b** that could be easily separated afterwards, as previously reported.[28] Characterisation of **4a–b** confirmed the presence of the allene moiety and the *N*,*N*-chelating character of these complexes. Complexes **4a–b** are drawn as *trans* isomers in analogy to the square planar Pd(II) examples due to the similarities on the ^{13}C NMR shifts on the central carbon of the allene (see ESI,† section 2.2).

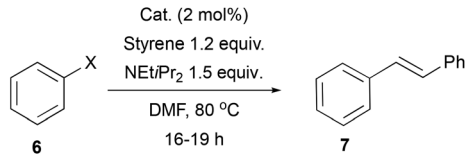
Catalytic applications

To explore the utility of our compounds as catalysts, we chose several benchmark reactions used in the evaluation of new catalysts depending on the metal used. We only tested the complexes in their racemic version, but success on this area will open avenues of research in the asymmetric variants (due to the axial chirality of the allene framework), making them worth exploring as catalysts.

Mizoroki–Heck cross-coupling

We began testing the Pd compounds **2a–b** in the Mizoroki–Heck cross-coupling reaction (Table 1).[46–48] A standard coupling between styrene and aryl halides (**6**, PhX, X = I or Br) was performed with 2 mol% catalyst loading. We observed formation of (*E*)-stilbene **7** in reaction with iodobenzene catalysed

Table 1 Catalytic activity of Pd(II)-complexes **2a–b** in Mizoroki–Heck couplings



6 7

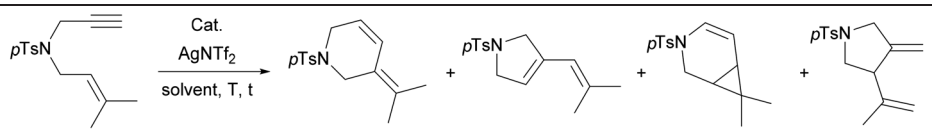
Entry	Cat.	X	Additives	Yield (%)
1	**2a**	I	HCO_2NH_4 (10 mol%)	87
2	**2a**	I	—	93
3	**2a**	Br	—	20
4	**2b**	I	—	36
5	**2b**	Br	—	—

by **2a**. The use of additives, such as HCO_2NH_4, was not necessary to obtain an excellent yield of the product (Table 1, entry 1 *vs*. entry 2). The activity of **2a** was substantially diminished in a reaction with bromobenzene (Table 1, entry 3). Non-symmetric complex **2b** was less active, giving a moderate yield in the reaction with PhI (36% yield, Table 1, entry 4) and showing no activity in the coupling with PhBr (Table 1, entry 5). The difference in activity between the two catalysts could be attributed, for example, to the steric bulk near the catalytic centre in **2b**, *e.g.* the access of starting materials might be more hindered for the *t*-butyl substituted complex. We could also explain the discrepancy if we consider complex **2a** to be in fact predominantly *cis* rather than *trans* like **2b**. Although further tuning of the catalyst design is needed to increase the activity with more challenging substrates, we present here the first example of an allene-derived catalytic system used in such an important process that could open further research and developments also in the asymmetric version.

Cyclisation of enynes

The catalytic activity of all complexes was next evaluated in reactions of 1,6-enynes (Table 2 and ESI,† sections 2.4 and 2.5).[30,49–52] We observed an initial lack of activity of Pd(II) and Pt(IV) compounds in the absence of $AgNTf_2$, either at room temperature or at 80 °C (Table 2, entries 1–4). These observations were consistent with a known heightened reactivity of cationic complexes in the catalysis involving enynes.[53,54] Thus, to ensure the formation of the cationic catalytic species and improve the activity, the synthesised catalysts were subsequently used in the presence of excess $AgNTf_2$ in toluene at 80 °C (Table 2, entries 5–8). One-hour long reactions with all the Pd(II) and Pt(IV) catalysts resulted in excellent conversions and yields. Interestingly, a switch between catalysts based on different metal centres allowed to achieve selectivity towards different isomeric products. Pd-Complexes **2a–b** favoured formation of a cyclopropane bicycle **9c**, while Pt-ones **3a–b** gave 6-membered skeletal rearrangement **9a** and Alder–ene **9d** as major reaction products. We restricted the use of Au(III) allene-containing complexes **4a–b** to a near room temperature conditions to avoid the cyclisation to carbene species during the catalytic reaction. Although these complexes showed to be active even without addition of $AgNTf_2$ providing further evidence of their cationic character (see ESI,† section 2.4), they were tested with the addition of $AgNTf_2$ here to boost the reactivity at room temperature (Table 2, entries 9 and 10). We observed only moderate conversions after 1.5 h, although this time the major product of the reaction was the 5-membered skeletal rearrangement product **9b**. Furthermore, Pt-complexes **3a–b** performed well as catalysts in the alkoxycyclisation of all enynes tested, when the reaction was carried out in MeOH. Interestingly, when N-tethered enynes (**8**) were investigated, Au-complexes **4a–b** seemed to give better results (see ESI,† section 2.5).[55–57]

Table 2 Catalytic activity of complexes **2a–b**, **3a–b** and **4a–b** in cyclisation of enynes



8 9a 9b 9c 9d

Entry	Cat. (mol%)	[$AgNTf_2$] (mol%)	Solvent	*T* (°C)	*t* (h)	**8**	**9a**	**9b**	**9c**	**9d**	Conv. (%)	Yield (%)
1	**2a** (3)	—	CH_2Cl_2	rt	1.5	1.0	—	—	—	—	0	—
2	**3a** (3)	—	CH_2Cl_2	rt	1.5	1.0	—	—	—	—	0	—
3	**2a** (3)	—[a]	PhMe	80	1	1.0	—	—	—	—	n. is.[b]	n. is[b]
4	**3a** (3)	—	PhMe	80	16	1.0	—	—	—	—	0	—
5	**2a** (3)	6.5	PhMe	80	1	—	0.6	—	1.0	0.06	>99	>99
6	**2b** (3)	6.5	PhMe	80	1	0.08	0.4	—	1.0	0.07	95	85
7	**3a** (3)	6.5	PhMe	80	1	—	0.8	0.3	0.4	1.0[c]	>99	88
8	**3b** (3)	6.5	PhMe	80	1	—	1.0	0.35	0.4	1.0[c]	>99	>99
9	**4a** (2)	2.5	CH_2Cl_2	rt	1.5	1.0	0.14	0.4	0.05	—	37	36
10	**4b** (2)	2.5	CH_2Cl_2	rt	1.5	1.0	0.08	0.25	0.03	—	26	26

[a] $[Ag]^+$ added after 1 h. [b] n. is. = not isolated. [c] Unidentified product also obtained (0.35 ratio).

Nucleophilic addition to allenes

To further test the robustness as catalysts of our new allene-derived complexes, we thought it would be an interesting challenge to use them in reactions involving allenes, like the well-explored metal-catalysed nucleophilic addition to allenes.[58,59] For this, we chose, as model, the reaction between allene–phthalimide **10** and benzyl alcohol. The Pt(IV)-complexes **3a–b** were the most active catalysts from the group (Table 3, entries 1 and 2). Both resulted in complete conversions and formation of the distal double addition product – acetal **11a** corresponding to the usual reactivity with Pt(II) catalysts.[60–63] Acetal **11a** was also isolated in its hydrolysed form of the aldehyde **12a**. Interestingly, a minor product of the reaction (**11b**) was identified as the double addition product to the central carbon atom of the allene that readily hydrolysed to ketone **12b** upon handling. This type of reactivity has been previously observed only with thiol nucleophiles under Au(III) catalysis.[64] However, our Au(III)-complex **4b** promoted the formation of the single distal addition alkene product **13**, albeit with low conversion (Table 3, entry 3), showing reactivity typical for Au(I) catalysis.[65,66] Pd(II)-Complex **2a** however, was unreactive under the reaction conditions (Table 3, entry 4).

Although the Au(III) and Pd(II) complexes yielded low reactivity, the Pt(IV) ones gave results in line with state-of-the-art reactivity of Pt complexes with the advantage of working at room temperature and being the first examples in which an allene-containing Pt catalyst is used to catalyse a reaction involving allene compounds.[67]

Bioactivity

In a more novel application, we explored the unique 3D geometries of the new organometallics containing allenes, as potential metallodrugs in three settings: as antimicrobial agents (in collaboration with Community for Open Antimicrobial Drug Discovery – CO-ADD), as anticancer drugs and their specific interaction with different structures of DNA (double helical, i-motif and G-quadruplex). Promising results in these areas will open future studies on the mechanism of action, speciation and the nature of the bioactive species in each of the biological settings, which are out of the scope of this work.

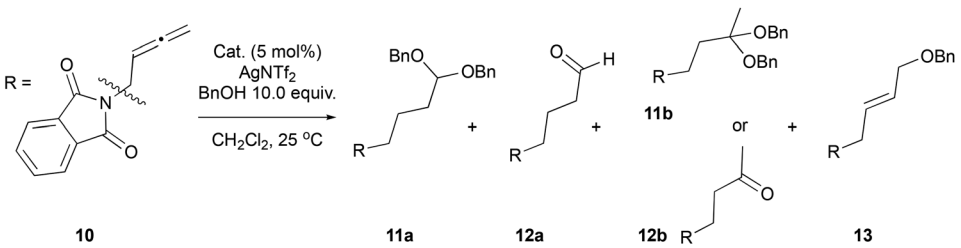
Antimicrobial studies

Complexes **2–4** and ligands **1a–b** were tested for their antimicrobial activity in collaboration with the Community for Open Antimicrobial Drug Discovery (CO-ADD).[68] The initial screening against microorganisms from Gram-negative bacteria (G –ve), Gram-positive bacteria (G +ve) and fungi groups at single drug concentration of 32 μg mL^{-1} (see ESI,† section 6) showed that all complexes were active (percentage of microorganism growth inhibition >80%) against at least one of the groups of microorganisms at that concentration, with the exception of complex **3a**, that was completely inactive. The ligands **1a–b** alone did not show any activity when tested at the same concentration against the same microorganisms.

Dose–response studies (in the concentration range of 0.25–32 μg mL^{-1}) of these complexes (except **3a**) were performed with the same microorganism species to determine the minimum inhibitory concentration (MIC) value.[69] The activity of complexes **2** and **4** was confirmed against a selection of the microorganisms, a G +ve bacteria, *Staphylococcus aureus* (*S. aureus*), and two yeasts, *Candida albicans* (*C. albicans*) and *Cryptococcus neoformans* (*C. neoformans*) *var. grubii* (Table 4, entries 1, 2 and 4, 5), with complex **2b** being the most active one, showing remarkably low MIC values of ≤0.25 μg mL^{-1} (Table 4, entry 2).[70,71] Although complex **3b** exhibited activity in the single dose assay, it showed no induction of growth inhibition at lower concentrations with any of the microorganisms tested (Table 4, entry 3).

Additionally, cytotoxicity of the new complexes was probed against human embryonic kidney cells (HEK-293) as the standard model for human cells used by the CO-ADD (Table 4). Cytotoxicity was expressed as a CC_{50} value – concentration at 50% cytotoxicity. The compounds with CC_{50} equal or lower than the maximum tested concentration (32 μg mL^{-1}) were

Table 3 Catalytic activity of complexes **2–4** in the nucleophilic addition of benzyl alcohol to allene-phthalimide **10**



Entry	Cat.	$AgNTf_2$ (mol%)	t (h)	**10**	**11a**	**12a**	**11b/12b**	**13**	Conv. (%)
1	**3a**	10	26	—	1.0	0.4	0.08	—	>99[a]
2	**3b**	10	27	—	1.0	0.24	0.07	—	>99[b]
3	**4b**	5	27	1.0	—	—	—	0.2	17
4	**2a**	10	24	1.0	—	—	—	—	0

[a] Isolated yield 67%. [b] Isolated yield 61%.

Table 4 Minimum inhibitory concentration (MIC, μg mL^{-1}) on selected microorganisms, in a range of 0.25–32 μg mL^{-1} of the drug (complexes **2–4**) concentration; CC_{50} (cytotoxicity, μg mL^{-1}) with HEK-293 cells; HC_{10} (haemolytic activity, μg mL^{-1}) with RBC cells

		G +ve	Yeast		Human	
		MIC [μg mL^{-1}]	MIC [μg mL^{-1}]			
Entry	Complex	*S. aureus*	*C. albicans*	*C. neoformans var. grubii*	CC_{50} [μg mL^{-1}] Embryonic kidney cells HEK-293	HC_{10} [μg mL^{-1}] Red blood cells RBC
1	**2a**	2	16	1	0.9567	>32
2	**2b**	≤0.25	≤0.25	≤0.25	1.066	>32
3	**3b**	>32	32	>32	>32	>32
4	**4a**	16	4	16	>32	>32
5	**4b**	16	4	≤0.25	>32	≤0.25

considered toxic. This included the most promising active Pd(II)-complex **2b** and the analogous **2a** that also showed good activity in the dose–response assay.

Au(III)-complexes **4a** and **4b** were moderately active against the microorganisms. However, **4b** showed haemolytic activity against human red blood cells (RBC),[72] which prevented further *in vivo* testing (using moth larvae *Galleria mellonella*).

Although complex **4a** seemed to be the most promising lead for antibiotic activity and non-toxic to selected human cells, the activity was moderate compared with current organometallics leads[15,73–76] and it was not selected for *in vivo* studies. However, these preliminary results, open the door for further studies with analogues of Au(III)-complexes **4** as promising novel antimicrobial drugs.[77–79]

Anticancer studies

In parallel to the antimicrobial studies, we investigated the preliminary *in vitro* anticancer activity of the new library of compounds **2–4** against the human breast adenocarcinoma cell line MDA-MB-231 (see ESI,† section 7), a triple-negative breast cancer cell line difficult to treat with current therapies and causing poorer outcomes for patients in comparison with other subtypes.[80]

The initial experiments were carried out with a concentration range of 0.19–100 μM over a 24 h incubation period, using the CellTiter-Blue assay to measure cell viability. Interestingly, only symmetric complexes of Pd(II) **2a** and Au(III) **4a** showed effects on MDA-MB-231 cell viability in the low micromolar range, while **2b** and **4b** were significantly less active. In particular, **2b** showed no activity even at the highest concentration tested (100 μM). Pt(IV) complexes **3a–b** did not show a clear sigmoidal trend of diminished cell survival, which could be attributed to solubility issues in the aqueous media, difficulties in the uptake into the cells or thwarted reduction to more active Pt(II) species, as we experienced in the chemical and electrochemical reduction experiments described above.

The free ligands **1a–b** did not show any effect on cell survival in line with the results seen in the antimicrobial activity studies (*vide supra*).

We validated the results of the more promising complexes **2a** and **4a** using the MTT assay, which measures cell metabolic activity, in a similar concentration range with a 24 h incubation period. The IC_{50} values for complexes **2a** and **4a** were determined to be 3.32 ± 1.22 and 2.50 ± 1.63 μM, using the MTT assay. These values are within the range obtained for reported Pd(II) (IC_{50} *ca.* 0.5–22.9 μM)[81–85] and Au(III) (IC_{50} *ca.* 0.3–11.2 μM)[86,87] complexes, and lower than reported values for cisplatin (IC_{50} *ca.* 25–50 μM depending on the study) against the MDA-MB-231 cell line. These results highlight the potential of our 3D metal complexes as new scaffolds for anticancer drug discovery.

As mentioned before, preliminary cytotoxicity results on embryonic kidney cells and red blood human cells were carried out as part of the antimicrobial study by our collaborators at CO-ADD (*vide supra*, Table 4). Neither complex showed haemolytic against human red blood cells. Although complex **2a** showed cytotoxicity towards embryonic kidney cells, complex **4a** did not, which could support further testing of this complex and new analogues, extending the study to other cancer cell lines and biocompatibility studies towards healthy human breast tissue.

DNA studies

The mechanism of action of many metal-based drugs is often, at least partially, ascribed to their interactions with DNA as an intracellular target (*e.g.* cisplatin, oxaliplatin, carboplatin).[88–92] We wanted to probe the possible source of the activity of the allene-containing metal complexes in anticancer assays by studying their interactions with different DNA types, not only with the standard double stranded helical structure, but also with non-canonical DNA secondary structures – such as G-quadruplexes and i-motifs.[93] These sequences are widespread throughout the human genome and can be found in promoter regions of their genes (for example the gene encoding death associated protein, DAP)[94] and in the telomeres.[95] In particular, the telomeric region of DNA and associated proteins have been linked to important functions such as DNA replication and protection against DNA damage.[96] As such, the link between dysfunctional telomeric DNA and disease *e.g.* cancer, is of much interest.[97]

The screening of the new complexes **2–4** and the free ligands **1a–b** was performed with several types of DNA: double helical DNA (DS), the i-motif (hTeloC) and G-quadruplex (hTeloG) structures from the human telomere and i-motif forming sequences from the promoter regions of hif-1-α and

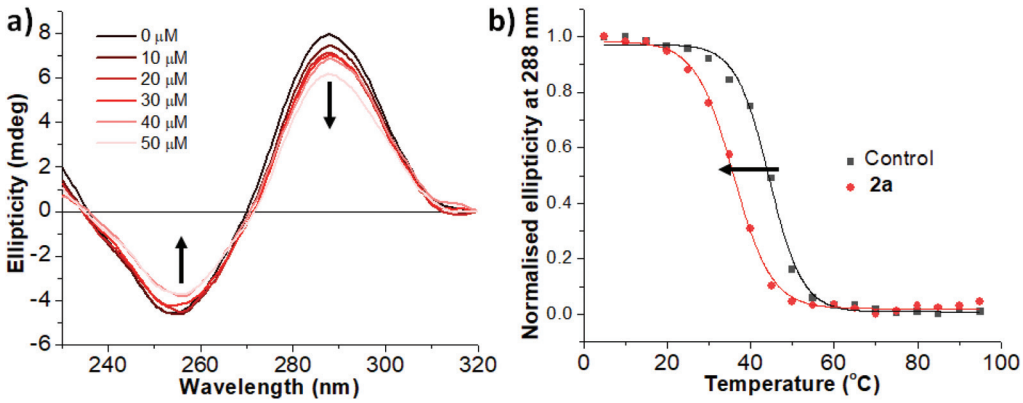


Fig. 4 Interaction of compound **2a** with i-motif DNA. (a) Circular dichroism spectroscopy of hTeloC (10 μM) in 10 mM sodium cacodylate buffer, 100 mM KCl, pH 5.5. 0–5 equiv. of compound **2a**. (b) CD melting experiment with hTeloC in the absence (black) and in the presence of 5 equiv. of compound **2a** (red).

DAP,[98] using fluorescence intercalator displacement (FID) assays with thiazole orange (TO) as the fluorescent probe (see ESI,† section 8.2).[99] The compounds were tested at a single concentration (2.5 μM, 5 equiv. in respect to DNA). In these assays, the ligands **1a–b** did not show displacement of the TO probe with any of the DNA structures, in line with the lack of activity observed in the antimicrobial and anticancer assays. This result provides further evidence that any observed bioactivity of the allene-derived metal complexes is probably not linked to the action of disassociated ligands if such de-complexation would take place under physiological conditions.

In contrast, Pd(II)-complexes **2a** and **2b** exhibited displacement of TO from hTeloG, showing partial affinity for the G-quadruplex sequence (35% and 48% for **2a** and **2b** respectively). These two Pd complexes were also the most active binders for all i-motif sequences, reaching displacement levels of 87% for **2b** with hTeloC.

Pt(IV) complexes **3a–b** were inactive on the level of the free allene ligands **1a–b**, and both Au(III)-complexes **4a–b** interacted with the DNA to a much lesser extent.[100] This would be in agreement with the mode of action of other Au(III) complexes, that are reported to be more prone to interact with protein targets.[101–106]

The more promising binders, Pd complexes **2a–b**, were further examined using circular dichroism (CD) spectroscopy (**2a**, Fig. 4).[107] By virtue of the significant number of studies of G-quadruplex binding compounds, there are a few examples of Pd complexes interacting with G-quadruplex DNA.[108–110] However, to the best of our knowledge, there are no studies that have looked at Pd complexes and their interaction with i-motif DNA.[111]

Pd-Complex **2a** was first titrated against hTeloC up to 5 : 1 ligand to DNA ratio (Fig. 4a).[112] The characteristic i-motif peaks were observed at around 255 nm (negative) and 288 nm (positive). The increasing excess of **2a** induced a visible hypochromic shift of both peaks, that is usually associated with an unfolding effect of the DNA secondary structure. CD-melting experiments were performed to quantify the effects on the stability of the DNA structures. A plot of normalised ellipticity *vs.* temperature for the peak at 288 nm allowed to find the values of the melting temperature for the hTeloC sequence with (red circle) and without (black square) the Pd complexes (Fig. 4b), and hence the change in melting temperature (ΔT_m) value of −8.1 ± 0.8 °C, which indicates destabilisation of this i-motif DNA structure in the presence of the complex. This is similar in potency to the destabilisation of i-motif caused by commonly-used G-quadruplex stabilising ligands.[113,114] Complex **2b** also showed a change in melting temperature (ΔT_m = −2.1 ± 0.1 °C) for hTeloC, albeit smaller than that for **2a**. Although there are many types of ligands capable of interacting with G-quadruplex DNA, compounds that interact with i-motif are very rare. Importantly these complexes did not significantly alter the stability of any other DNA structure examined, indicating some specificity for destabilising the i-motif form hTeloC. These unusual properties of our allene-containing Pd(II)-complexes and their different interaction with DNA structures, provide further indication of the promise of these new chemical scaffolds as leads for future medicinal applications.

Conclusions

In summary, we present here bis(pyridyl)allenes as robust ligands for novel Pd(II), Pt(IV) and Au(III) complexes with unique 3D architectures for application in two important areas, catalysis and medicinal chemistry. While Pd(II) and Pt(IV) complexes were easily obtained, the isolation of Au(III) complexes needed careful control of the reaction conditions to avoid cyclisation of the ligand framework to form Au(III) carbene derivatives. Crystal structure analysis of the Pd(II) complex **2b** showed an unusual *trans* geometry. The Pt(IV) complexes **3a–b** seemed to display octahedral geometry and *cis–trans* ligand isomerisation in solution. Electrochemical analysis showed quite negative Pt(IV)/Pt(II) reduction potentials, and cycloplatination occurred with **3b** when forcing chemical

reduction conditions were applied. The analogue Au(III) complexes **4a–b** containing intact allenes as ligands were isolated when the reaction of the ligands **1a–b** and Au(III) precursors was carried out at mild temperatures. Increasing temperature or longer reaction times gave full conversion to the Au(III) carbene complexes **5a–b** previously described by our group.

All the allene-based metal complexes were tested as catalysts in benchmark reactions for Pd (Mizoroki–Heck cross-coupling), Pt and Au (enyne cyclisation and nucleophilic addition to allenes). The symmetric Pd(II) complex **2a** performed well in the Mizoroki–Heck cross coupling of iodobenzene and styrene, giving good yields under standard reaction conditions without the need for additives. All the complexes performed well in the cycloisomerisation reaction of enynes. The Pd(II) and Pt(IV) complexes worked at higher temperatures in the presence of silver salts, while the Au(III) complexes worked at room temperatures. We observed a switch in selectivity towards different isomeric products depending on the metal complex used. The Au(III) complexes **4a–b** were the most active catalysts in alkoxycyclisation reaction of N-tethered enynes, while the Pt(IV) analogues **3a–b** performed well with all type of enynes and best in the nucleophilic addition to allenes to give double addition products. Further studies on the asymmetric version of the ligands and their application in asymmetric catalysis are undergoing in our laboratories.

Finally, studies of the new complexes as new antimicrobial and anticancer agents gave very promising preliminary results. Although Pt(IV) complexes did not show biological activity, possibly due to solubility issues or their difficulty of *in situ* reduction to potentially active Pt(II) derivatives, the Pd(II) and Au(III) analogues showed good activity in both antimicrobial and anticancer assays, as well as interaction with different DNA structures. In particular it is worth highlighting the Pd(II) complex **2a** that showed good activity in the antimicrobial and anticancer assays and an unusual destabilisation of the hTeloC i-motif DNA structure, which is notoriously hard to achieve, and could begin to explain its biological activity. This peculiar interaction with i-motif secondary structures makes this work highly novel and would open new avenues of research in the area of metallodrugs using Pd complexes with alternative biological mechanisms of action.

Author contributions

H. K. M. contributed to all the experimental work, writing of the original draft of the manuscript, later reviewing, editing and the whole ESI.† C. A. V. contributed to the MDA-MB-231 cell cultures, provided help with the anticancer assays, and reviewing of the manuscript. Y. X. carried out the FID and CD assays. M. J. M. provided training, supervision and analysis of the anticancer assays and contributed to writing, reviewing, and editing the manuscript. Z. A. E. W. provided guidance and analysis of the experiments with DNA, and contributed to writing, reviewing, and editing the manuscript. M. P. M. supervised the entire research with conceptualisation, analysis, and resources, contributed to the writing, and coordinated the reviewing and editing of the final manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The DNA and the anticancer experiments use synthetic DNA sequences and cultured human cell lines that do not have any associated ethic approvals. The antimicrobial work carried out by the CO-ADD was performed in strict accordance with The Australian Code for the Responsible Conduct of Research (2018). Human blood was sourced from the Australian Red Cross Blood Service with informed consent, and its use in haemolysis assays was approved by The University of Queensland Institutional Human Research Ethics Committee, Approval Number 2014000031.

Funding by the Faculty of Science at the University of East Anglia is gratefully acknowledged (H. K. M. and C. A. V.). C. A. V would like to thank Mr and Mrs Whittaker for her oncology fellowship. Authors would like to thank Dr John Fielden (UEA) for help with the electrochemistry, Dr Rianne Lord (UEA) for her help with the MTT assay, Dr Abdul-Sada (University of Sussex) for HRMS analysis and Stephen Boyer (London Metropolitan University) for elemental analysis. Antimicrobial screening was performed by CO-ADD (Community for Open Antimicrobial Drug Discovery), funded by the Welcome Trust (UK) and The University of Queensland (Australia).

Notes and references

1 F. Lovering, J. Bikker and C. Humblet, *J. Med. Chem.*, 2009, **52**, 6752–6756.
2 A. W. Hung, A. Ramek, Y. Wang, T. Kaya, J. A. Wilson, P. A. Clemons and D. W. Young, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 6799–6804.
3 F. Lovering, *MedChemComm*, 2013, **4**, 515–519.
4 C. H. Arrowsmith, J. E. Audia, C. Austin, J. Baell, J. Bennett, J. Blagg, C. Bountra, P. E. Brennan, P. J. Brown, M. E. Bunnage, C. Buser-Doepner, R. M. Campbell, A. J. Carter, P. Cohen, R. A. Copeland, B. Cravatt, J. L. Dahlin, D. Dhanak, A. M. Edwards, M. Frederiksen, S. V. Frye, N. Gray, C. E. Grimshaw, D. Hepworth, T. Howe, K. V. M. Huber, J. Jin, S. Knapp, J. D. Kotz, R. G. Kruger, D. Lowe, M. M. Mader, B. Marsden, A. Mueller-Fahrnow, S. Müller, R. C. O'Hagan, J. P. Overington, D. R. Owen, S. H. Rosenberg, R. Ross, B. Roth, M. Schapira, S. L. Schreiber, B. Shoichet, M. Sundström, G. Superti-Furga, J. Taunton, L. Toledo-Sherman, C. Walpole, M. A. Walters, T. M. Willson, P. Workman, R. N. Young and W. J. Zuercher, *Nat. Chem. Biol.*, 2015, **11**, 536–541.

5 R. H. Holm, P. Kennepohl and E. I. Solomon, *Chem. Rev.*, 1996, **96**, 2239–2314.
6 K. D. Mjos and C. Orvig, *Chem. Rev.*, 2014, **114**, 4540–4563.
7 C. N. Morrison, K. E. Prosser, R. W. Stokes, A. Cordes, N. Metzler-Nolte and S. M. Cohen, *Chem. Sci.*, 2020, **11**, 1216–1225.
8 J.-H. Xie and Q.-L. Zhou, *Acc. Chem. Res.*, 2008, **41**, 581–593.
9 W. Sommer and D. Weibel, *Chemfiles*, 2008, **8**, 1–92.
10 A. Pfaltz and W. J. Drury III, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 5723–5726.
11 W. W. Brandt, F. P. Dwyer and E. D. Gyarfas, *Chem. Rev.*, 1954, **54**, 959–1017.
12 N. Krause and A. S. K. Hashmi, in *Modern Allene Chemistry*, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2004, vol. 1, p. 54.
13 A. Hoffmann-Röder and N. Krause, *Angew. Chem., Int. Ed.*, 2004, **43**, 1196–1216.
14 M. A. Sierra, L. Casarrubios and M. C. de la Torre, *Chem. – Eur. J.*, 2019, **25**, 7232–7242.
15 A. Frei, J. Zuegg, A. G. Elliott, M. Baker, S. Braese, C. Brown, F. Chen, C. G. Dowson, G. Dujardin, N. Jung, A. P. King, A. M. Mansour, M. Massi, J. Moat, H. A. Mohamed, A. K. Renfrew, P. J. Rutledge, P. J. Sadler, M. H. Todd, C. E. Willans, J. J. Wilson, M. A. Cooper and M. A. T. Blaskovich, *Chem. Sci.*, 2020, **11**, 2627–2639.
16 K. Okamoto, Y. Kai, N. Yasuoka and N. Kasai, *J. Organomet. Chem.*, 1974, **65**, 427–441.
17 T. J. Brown, A. Sugie, M. G. Dickens and R. A. Widenhoefer, *Organometallics*, 2010, **29**, 4207–4209.
18 M. T. Quirós, M. P. Muñoz, J. Christensen and S. J. Coles, *Organometallics*, 2017, **36**, 318–330.
19 A. L. Colebatch, I. A. Cade, A. F. Hill and M. M. Bhadbhade, *Organometallics*, 2013, **32**, 4766–4774.
20 For other examples of allene containing phosphines as ligands for metal complexes see: E. V. Banide, J. P. Grealis, H. Müller-Bunz, Y. Ortin, M. Casey, C. Mendicute-Fierro, M. Cristina Lagunas and M. J. McGlinchey, *J. Organomet. Chem.*, 2008, **693**, 1759–1770 and ref. 21. No applications of these phosphine- and phosphine oxide-containing fluorenyl(allene) complexes have been reported.
21 S. Milosevic, E. V. Banide, H. Müller-Bunz, D. G. Gilheany and M. J. McGlinchey, *Organometallics*, 2011, **30**, 3804–3817.
22 S. Löhr, J. Averbeck, M. Schürmann and N. Krause, *Eur. J. Inorg. Chem.*, 2008, 552–556.
23 F. Cai, X. Pu, X. Qi, V. Lynch, A. Radha and J. M. Ready, *J. Am. Chem. Soc.*, 2011, **133**, 18066–18069.
24 A. Vanitcha, C. Damelincourt, G. Gontard, N. Vanthuyne, V. Mouriès-Mansuy and L. Fensterbank, *Chem. Commun.*, 2016, **52**, 6785–6788.
25 F. Cai, N. D. Thangada, E. Pan and J. M. Ready, *Organometallics*, 2013, **32**, 5619–5622.
26 S. Sentets, R. Serres, Y. Ortin, N. Lugan and G. Lavigne, *Organometallics*, 2008, **27**, 2078–2091.
27 A. Vanitcha, G. Gontard, N. Vanthuyne, E. Derat, V. Mouriès-Mansuy and L. Fensterbank, *Adv. Synth. Catal.*, 2015, **357**, 2213–2218.
28 H. K. Maliszewska, D. L. Hughes and M. P. Muñoz, *Dalton Trans.*, 2020, **49**, 4034–4038.
29 A. Fürstner and P. W. Davies, *Angew. Chem., Int. Ed.*, 2007, **46**, 3410–3449.
30 E. Jiménez-Núñez and A. M. Echavarren, *Chem. Rev.*, 2008, **108**, 3326–3350.
31 S. P. Nolan, *Acc. Chem. Res.*, 2011, **44**, 91–100.
32 C. Praveen, *Coord. Chem. Rev.*, 2019, **392**, 1–34.
33 B. W. Harper, A. M. Krause-Heuer, M. P. Grant, M. Manohar, K. B. Garbutcheon-Singh and J. R. Aldrich-Wright, *Chem. – Eur. J.*, 2010, **16**, 7064–7077.
34 B. Bertrand, M. R. M. Williams and M. Bochmann, *Chem. – Eur. J.*, 2018, **24**, 11840–11851.
35 X. Wang, X. Wang, S. Jin, N. Muhammad and Z. Guo, *Chem. Rev.*, 2019, **119**, 1138–1192.
36 A. R. Kapdi and I. J. S. Fairlamb, *Chem. Soc. Rev.*, 2014, **43**, 4751–4777.
37 N. C. Vieira, J. A. Pienkos, C. D. McMillen, A. R. Myers, A. P. Clay and P. S. Wagenknecht, *Dalton Trans.*, 2017, **46**, 15195–15199.
38 L. D. Jaques, S. L. McDarmont, M. M. Smart, C. D. McMillen, S. E. Neglia, J. P. Lee and J. A. Pienkos, *Inorg. Chem. Commun.*, 2020, **112**, 107722–107727.
39 T. Kawano, T. Shinomaru and I. Ueda, *Org. Lett.*, 2002, **4**, 2545–2547.
40 Interestingly the bis(phosphine)allene-Pd complex reported by Fensterbank (see ref. 27) shows a *cis* arrangement, a bent allene backbone (151.3°) and $\Delta\delta$ = 16.6 ppm of the central allene carbon in ^{13}C NMR.
41 R. D. Shannon, *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.*, 1976, **32**, 751–767.
42 R. Kuroda, S. Neidle, I. M. Ismail and P. J. Sadler, *Inorg. Chem.*, 1983, **22**, 3620–3624.
43 M. D. Hall and T. W. Hambley, *Coord. Chem. Rev.*, 2002, **232**, 49–67.
44 Y. Shi, S.-A. Liu, D. J. Kerwood, J. Goodisman and J. C. Dabrowiak, *J. Inorg. Biochem.*, 2012, **107**, 6–14.
45 D. Höfer, H. P. Varbanov, M. Hejl, M. A. Jakupec, A. Roller, M. Galanski and B. K. Keppler, *J. Inorg. Biochem.*, 2017, **174**, 119–129.
46 I. P. Beletskaya and A. V. Cheprakov, *Chem. Rev.*, 2000, **100**, 3009–3066.
47 A. B. Dounay and L. E. Overman, *Chem. Rev.*, 2003, **103**, 2945–2964.
48 C. C. C. Johansson Seechurn, M. O. Kitching, T. J. Colacot and V. Snieckus, *Angew. Chem., Int. Ed.*, 2012, **51**, 5062–5085.
49 C. Aubert, O. Buisine and M. Malacria, *Chem. Rev.*, 2002, **102**, 813–834.
50 L. Fensterbank and M. Malacria, *Acc. Chem. Res.*, 2014, **47**, 953–965.

51 M. T. Quirós and M. P. Muñoz, in *Au-Catalyzed Synthesis and Functionalization of Heterocycles*, Springer, Cham, 2016, vol. 46, pp. 117–174.
52 C. García-Morales and A. Echavarren, *Synlett*, 2018, 2225–2237.
53 D. Wang, R. Cai, S. Sharma, J. Jirak, S. K. Thummanapelli, N. G. Akhmedov, H. Zhang, X. Liu, J. L. Petersen and X. Shi, *J. Am. Chem. Soc.*, 2012, **134**, 9012–9019.
54 A. Franchino, M. Montesinos-Magraner and A. M. Echavarren, *Bull. Chem. Soc. Jpn.*, 2021, **94**, 1099–1117.
55 M. P. Muñoz, M. Méndez, C. Nevado, D. J. Cárdenas and A. M. Echavarren, *Synthesis*, 2003, 2898–2902.
56 L. Charruault, V. Michelet, R. Taras, S. Gladiali and J.-P. Genêt, *Chem. Commun.*, 2004, 850–851.
57 C. Nieto-Oberhuber, M. P. Muñoz, S. López, E. Jiménez-Núñez, C. Nevado, E. Herrero-Gómez, M. Raducan and A. M. Echavarren, *Chem. – Eur. J.*, 2006, **12**, 1677–1693.
58 N. Krause, Ö. Aksin-Artok, M. Asikainen, V. Breker, C. Deutsch, J. Erdsack, H.-T. Fan, B. Gockel, S. Minkler, M. Poonoth, Y. Sawama, Y. Sawama, T. Sun, F. Volz and C. Winter, *J. Organomet. Chem.*, 2012, **704**, 1–8.
59 M. P. Muñoz, *Chem. Soc. Rev.*, 2014, **43**, 3164–3183.
60 M. Paz Muñoz, M. C. de la Torre and M. A. Sierra, *Adv. Synth. Catal.*, 2010, **352**, 2189–2194.
61 M. P. Muñoz, M. C. de la Torre and M. A. Sierra, *Chem. – Eur. J.*, 2012, **18**, 4499–4504.
62 L. Cooper, J. M. Alonso, L. Eagling, H. Newson, S. Herath, C. Thomson, A. Lister, C. Howsham, B. Cox and M. P. Muñoz, *Chem. – Eur. J.*, 2018, **24**, 6105–6114.
63 M. T. Quirós, E. Gómez-Bengoa and M. P. Muñoz, *Pure Appl. Chem.*, 2020, **92**, 167–177.
64 Menggenbateer, M. Narsireddy, G. Ferrara, N. Nishina, T. Jin and Y. Yamamoto, *Tetrahedron Lett.*, 2010, **51**, 4627–4629.
65 M. P. Muñoz, *Org. Biomol. Chem.*, 2012, **10**, 3584–3594.
66 R. J. Harris, R. G. Carden, A. N. Duncan and R. A. Widenhoefer, *ACS Catal.*, 2018, **8**, 8941–8952.
67 There is one example of the di-Au(I)-mono phosphine complex reported by Fensterbank (see ref. 24) used as catalyst in the cyclisation of allenedienes.
68 M. A. T. Blaskovich, J. Zuegg, A. G. Elliott and M. A. Cooper, *ACS Infect. Dis.*, 2015, **1**, 285–287.
69 MIC is defined as the lowest concentration of the drug at which full inhibition of the bacteria and fungi growth has been detected (inhibition >80%) with the next highest concentration also exhibiting full growth inhibition. The compounds with MIC of 16 μg mL^{-1} or lower were deemed active.
70 The lowest MIC value against SA recently reported by CO-ADD was 0.5 μg mL^{-1} for an Ir complex. A Ru complex gave the lowest MIC value against Ca, 0.125 μg mL^{-1}. Few complexes were reported for activity against Cn, and a different Ir complex showed the lowest MIC value of 0.5 μg mL^{-1}. See ref. 15.
71 Previously reported antimicrobial activity of Pd complexes was lower (MIC > 100 μg mL^{-1} in G. Onar, C. Güsters, M. O. Karatas, S. Balcioglu, N. Akbay, N. Özdemir, B. Ates and B. Alici, *J. Organomet. Chem.*, 2019, **886**, 48–56; and MICs = 0.0197–5.0 mg mL^{-1} for bacteria in S. Y. Hussaini, R. A. Haque and M. R. Razali, *J. Organomet. Chem.*, 2019, **882**, 96–111).
72 The haemolytic activity against human red blood cells (RBC) was determined as a HC_{10} value – concentration at 10% haemolytic activity (Table 4). Values with the ">" sign indicate a sample with no haemolytic activity or samples with HC_{10} values above the maximum tested concentration.
73 M. Patra, G. Gasser and N. Metzler-Nolte, *Dalton Trans.*, 2012, **41**, 6350–6358.
74 B. Đ. Glišić and M. I. Djuran, *Dalton Trans.*, 2014, **43**, 5950–5969.
75 A. Frei, *Antibiotics*, 2020, **9**, 90–114.
76 B. Bertrand and A. Casini, *Dalton Trans.*, 2014, **43**, 4209–4219.
77 A. M. Mansour and O. R. Shehab, *Eur. J. Inorg. Chem.*, 2019, 2830–2838.
78 R. K. Mahato, A. K. Mahanty, S. Paul, V. Gopal, B. Perumalsamy, G. Balakrishnan, T. Ramasamy, D. Dharumadurai and B. Biswas, *J. Mol. Struct.*, 2021, **1223**, 129264–129274.
79 P. Chakraborty, D. Oosterhuis, R. Bonsignore, A. Casini, P. Olinga and D.-J. Scheffers, *ChemMedChem*, 2021, **16**, 3060–3070.
80 K. J. Chavez, S. V. Garimella and S. Lipkowitz, *Breast Dis.*, 2010, **32**, 35–48.
81 E. Ulukaya, F. Ari, K. Dimas, E. I. Ikitimur, E. Guney and V. T. Yilmaz, *Eur. J. Med. Chem.*, 2011, **46**, 4957–4963.
82 F. Ari, E. Ulukaya, M. Sarimahmut and V. T. Yilmaz, *Bioorg. Med. Chem.*, 2013, **21**, 3016–3021.
83 J. Haribabu, S. Srividya, D. Mahendiran, D. Gayathri, V. Venkatramu, N. Bhuvanesh and R. Karvembu, *Inorg. Chem.*, 2020, **59**, 17109–17122.
84 Z. Zhang, G. Du, G. Gong, Y. Sheng, X. Lu, W. Cai, F. Wang and G. Zhao, *J. Mol. Struct.*, 2021, **1232**, 130021–130027.
85 S. Kimani, S. Chakraborty, I. Irene, J. de la Mare, A. Edkins, A. du Toit, B. Loos, A. Blanckenberg, A. Van Niekerk, L. V. Costa-Lotufo, K. ArulJothi, S. Mapolie and S. Prince, *Biochem. Pharmacol.*, 2021, **190**, 114598–114615.
86 M. Williams, A. I. Green, J. Fernandez-Cestau, D. L. Hughes, M. A. O'Connell, M. Searcey, B. Bertrand and M. Bochmann, *Dalton Trans.*, 2017, **46**, 13397–13408.
87 S. Montanel-Pérez, R. Elizalde, A. Laguna, M. D. Villacampa and M. C. Gimeno, *Eur. J. Inorg. Chem.*, 2019, 4273–4281.
88 S. M. Cohen and S. J. Lippard, in *Progress in Nucleic Acid Research and Molecular Biology*, Academic Press, 2001, vol. 67, pp. 93–130.
89 C. K. Mirabelli, C.-M. Sung, J. P. Zimmerman, D. T. Hill, S. Mong and S. T. Crooke, *Biochem. Pharmacol.*, 1986, **35**, 1427–1433.

90 C. P. Saris, P. M. van de Vaart, R. C. Rietbroek and F. Bloramaert, *Carcinogenesis*, 1996, **17**, 2763–2769.
91 L. Messori, P. Orioli, C. Tempi and G. Marcon, *Biochem. Biophys. Res. Commun.*, 2001, **281**, 352–360.
92 S. G. Chaney, S. L. Campbell, B. Temple, E. Bassett, Y. Wu and M. Faldu, *J. Inorg. Biochem.*, 2004, **98**, 1551–1559.
93 S. Balasubramanian, L. H. Hurley and S. Neidle, *Nat. Rev. Drug Discovery*, 2011, **10**, 261–275.
94 E. P. Wright, J. L. Huppert and Z. A. E. Waller, *Nucleic Acids Res.*, 2017, **45**, 13095–13096.
95 A. T. Phan, *Nucleic Acids Res.*, 2002, **30**, 4618–4625.
96 W. Lu, Y. Zhang, D. Liu, Z. Songyang and M. Wan, *Exp. Cell Res.*, 2013, **319**, 133–141.
97 J. W. Shay and S. Bacchetti, *Eur. J. Cancer*, 1997, **33**, 787–791.
98 J. A. Brazier, A. Shah and G. D. Brown, *Chem. Commun.*, 2012, **48**, 10739–10741.
99 Q. Sheng, J. C. Neaverson, T. Mahmoud, C. E. M. Stevenson, S. E. Matthews and Z. A. E. Waller, *Org. Biomol. Chem.*, 2017, **15**, 5669–5673.
100 For an example of anticancer Pt(II) and Au(III) complexes with DNA interactions, see: J. J. Criado, J. L. Manzano and E. Rodríguez-Fernández, *J. Inorg. Chem.*, 2003, **96**, 311–320.
101 A. Casini, C. Hartinger, C. Gabbiani, E. Mini, P. J. Dyson, B. K. Keppler and L. Messori, *J. Inorg. Chem.*, 2008, **102**, 564–575.
102 A. Bindoli, M. P. Rigobello, G. Scutari, C. Gabbiani, A. Casini and L. Messori, *Coord. Chem. Rev.*, 2009, **253**, 1692–1707.
103 V. Milacic and Q. P. Dou, *Chem. Rev.*, 2009, **253**, 1649–1660.
104 E. Meggers, *Chem. Commun.*, 2009, 1001–1010.
105 A. Casini and L. Messori, *Curr. Top. Med. Chem.*, 2011, **11**, 2647–2660.
106 M. Dörr and E. Meggers, *Curr. Opin. Chem. Biol.*, 2014, **19**, 76–81.
107 M. Vorlíčková, I. Kejnovská, K. Bednářová, D. Renčiuk and J. Kypr, *Chirality*, 2012, **24**, 691–698.
108 E. Gao, M. Zhu, L. Liu, Y. Huang, L. Wang, C. Shi, W. Zhang and Y. Sun, *Inorg. Chem.*, 2010, **49**, 3261–3270.
109 M. A. Bork, C. G. Gianopoulos, H. Zhang, P. E. Fanwick, J. H. Choi and D. R. McMillin, *Biochemistry*, 2014, **53**, 714–724.
110 G. Onar, C. Gürses, M. O. Karataş, S. Balcıoğlu, N. Akbay, N. Özdemir, B. Ateş and B. Alıcı, *J. Organomet. Chem.*, 2019, **886**, 48–56.
111 A. Ali, M. Kamra, S. Roy, K. Muniyappa and S. Bhattacharya, *Bioconjugate Chem.*, 2017, **28**, 341–352.
112 We did not observe precipitation in these experiments. Partial precipitation causes scattering at 320 nm, which was not observed.
113 A. Pagano, N. Iaccarino, M. A. S. Abdelhamid, D. Brancaccio, E. U. Garzarella, A. Di Porzio, E. Novellino, Z. A. E. Waller, B. Pagano, J. Amato and A. Randazzo, *Front. Chem.*, 2018, **6**, 281–294.
114 M. A. S. Abdelhamid, A. J. Gates and Z. A. E. Waller, *Biochemistry*, 2019, **58**, 245–249.

Special Collection

Understanding the Potential *In Vitro* Modes of Action of Bis (β-diketonato) Oxovanadium(IV) Complexes

Baris Sergi[+],[b] Ipek Bulut[+],[b] Ying Xia,[c] Zoë A. E. Waller,[c, d] Yasemin Yildizhan,[e] Ceyda Acilan,*[f, g] and Rianne M. Lord*[a]

To understand the potential *in vitro* modes of action of bis(β-diketonato) oxovanadium(IV) complexes, nine compounds of varying functionality have been screened using a range of biological techniques. The antiproliferative activity against a range of cancerous and normal cell lines has been determined, and show these complexes are particularly sensitive against the lung carcinoma cell line, A549. Annexin V (apoptosis) and Caspase-3/7 assays were studied to confirm these complexes induce programmed cell death. While gel electrophoresis was used to determine DNA cleavage activity and production of reactive oxygen species (ROS), the Comet assay was used to determine induced genomic DNA damage. Additionally, Förster resonance energy transfer (FRET)-based DNA melting and fluorescent intercalation displacement assays have been used to determine the interaction of the complexes with double strand (DS) DNA and to establish preferential DNA base-pair binding (AT versus GC).

Introduction

Although the anticancer properties of inorganic compounds appeared in 1930,[1] the first indications for the antineoplastic effects of vanadium salts were not reported until 1965,[2] and more in depth studies were not conducted until 1980–1990s. Sabbioni et al. have since reported on the vanadium metabolism and the cytotoxicity of vanadium compounds.[3,4] In particular, the morphology transformations in mouse embryo after treatment with ammonium vanadate (V^{V}) or vanadyl sulfate (V^{IV}) salts has been studied, and it was noted that treatment with V^{V} alone, or in combination with diethylmaleate (DEM; a cellular glutathione (GSH)-depleting agent), reduces the complex from V^{V} to V^{IV} (measured by EPR spectroscopy). Although V^{IV}-compounds show lower toxicity in the same cell line,[4,5] it was later reported that vanadyl show no cytotoxic effects, whilst vanadate and pervanadate strongly inhibited cell development.[6]

Since the discovery of vanadium bromoperoxidase, a vanadium(V)-containing enzyme in marine algae,[7,8] the importance of vanadium in humans and diet has been an important research topic.[9] In the early 1980s, vanadium was reported to act as an insulin mimic and was able to normalize diabetic rats.[10,11] Since these reports, there has been a plethora of research on V-compounds for their insulin-like effects,[12–20] with groups studying their importance in different cell culture types.[21,22] Vanadium(V) was shown to stimulate glucose uptake and glucose oxidation,[23,24] and it was also postulated to inhibit tyrosine phosphate and bind to growth factors relating to cell proliferation.[25]

Over a decade later, bis(4,7-dimethyl-1,10-phenanthroline) sulfato oxovanadium(IV) (Metvan) was discovered, and to the best of our knowledge, it remains the most promising V^{IV} anticancer compound to be reported.[26] It exhibits nanomolar potency, and induces apoptosis in human leukemia cells, multiple myeloma cells and solid tumour cells (breast, glioblas-

[a] Dr. R. M. Lord
School of Chemistry
University of East Anglia
Norwich Research Park
NR4 7TJ Norwich (UK)
E-mail: r.lord@uea.ac.uk
[b] B. Sergi,[+] I. Bulut[+]
Gradute School of Health Sciences
Koç University
34450 Sariyer, Istanbul (Turkey)
[c] Y. Xia, Dr. Z. A. E. Waller
School of Pharmacy
University of East Anglia
Norwich Research Park
NR4 7TJ Norwich (UK)
[d] Dr. Z. A. E. Waller
School of Pharmacy
UCL
29–39 Brunswick Square, WC1E 6BT London (UK)
[e] Y. Yildizhan
Genetic Engineering and Biotechnology Institute
TUBITAK
41470 Kocaeli (Turkey)
[f] Dr. C. Acilan
School of Medicine
Koç University
34450 Sariyer, Istanbul (Turkey)
E-mail: cayhan@ku.edu.tr
[g] Dr. C. Acilan
Research Center for Translational Medicine
Koç University
34450 Sariyer, Istanbul (Turkey)
[+] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cmdc.202100152

Special Collection This article belongs to the joint Special Collection with the European Journal of Inorganic Chemistry, "Metals in Medicine".

toma, ovarian, prostate and testicular). It has been noted that the apoptosis is also associated with a loss of mitochondrial transmembrane and a depletion of GSH. Not only does this compound show high *in vitro* cytotoxicity, *in vivo* studies show no acute or subacute toxicity at 12.5–50 mg/kg.[26]

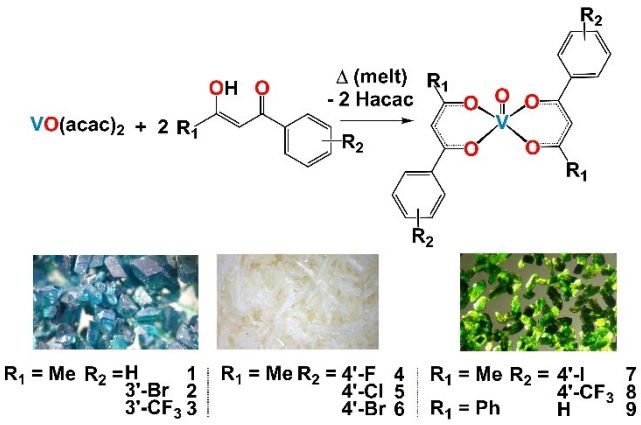
Since the discovery of Metvan, several research groups have studied V^{IV} complexes, showing they are better tolerated, more potent and have better cancer selectivity then clinical Pt(II) drugs.[27,28] In 2018, Ni et al. produced a small library of mixed ligand oxidovanadium(IV) complexes with polycarboxylates and *N*-heterocyclic ligands, with one compound showing high cytotoxicity, cell cycle S-, or G_2/M-phase arrest and cell death by apoptosis.[29] More recently, in 2020, Ribeiro et al. highlighted a range of VO(L) complexes (L = tridentate amino acid–pyridyl–phenol ligand), however, the complexes only exhibit moderate cytotoxicity values.[30] One V^{IV}-compound exhibited increased late apoptosis when compared to the analogous Cu^{II} complex, and it was able to nick and cleave plasmid DNA. Reports on non-oxidovanadium(IV) complexes have also emerged, with compounds exhibiting increased cytotoxicity, increased ROS formation and a decrease of the mitochondrial membrane potential.[31] Collectively, there are range of V^{IV}-compounds published, which all highlighted their importance in the treatment of different cancers. Importantly, the reports suggest both interactions with DNA and apoptosis by generation of ROS, are the major contributors to cytotoxicity and modes of action.

In 2019, we reported a series of bis(β-diketonato) oxovanadium(IV) complexes (**1**–**9**) through a facile and straightforward solid state synthesis, and show the complexes exhibit high cytotoxicity and selectivity towards cancer cells, with IC_{50} values up to 11.5× higher than cisplatin.[32] The complexes are also stable under physiological conditions and UV/Vis studies show interaction with BSA (bovine serum albumin). Since compound **9** was previously reported to have increased binding to both BSA, and inserted into the minor groove of the DNA duplex with a partial intercalation,[33] we have been interested in understanding the potential modes of action bis(β-diketonato) oxovanadium(IV) complexes. Herein, we report studies on the compound's cytotoxicity, interactions with DNA (plasmid, genomic and AT- and GC-rich sequences), production of reactive oxygen species (ROS) and modes of cell death by apoptosis and Caspase-3/7.

Results and Discussion

Synthesis and Characterization of bis(β-diketonato) oxovanadium(IV) complexes

We have recently highlighted the rapid (< 5 mins) dry-melt synthesis and anti-cancer screening of nine bis(β-diketonato) oxovanadium(IV) complexes (**1**–**9**, Scheme 1).[32] Alongside other researchers, we have reported that these vanadyl(IV) complexes have the potential to form vanadium(V) in solution. In order to address the compounds stability in DMSO, **1**–**9** and $VO(acac)_2$ have been studied by cyclic voltammetry (CV) in dry DMSO under a flow of argon. All compounds (~10 mM) display *quasi*-reversible behavior, where the forward and reverse electron transfer reactions occur with different rates. Notably, no CV spectra were obtained for the free ligands. In the presence of all vanadium compounds, there is a one-electron process for the β-diketonate ligands (two ligands), which is represented by the oxidation Epa_1 and reduction Epc_1. Followed by a weaker *quasi* reversible one electron process for the vanadium $V^{IV} \rightarrow V^{V}$, with oxidation Epa_2 and reduction Epc_2 (Table S1). Previous literature has highlighted the irreversible oxidation of $VO(acac)_2$ and VO(acac)L (L = aminophosphinic acid) when using 0.1 M NaCl (glass electrode),[34] whilst other reports have noted VO(acac)(abp) (abp = 2-acetylpyridine-benzoylhydrazone) has a strong abp ligand oxidation at 0.8 V but no observable oxidation for the vanadium (0.1 M PTBA/CH_2Cl_2).[35] However, the *quasi*-reversible nature of our vanadyl complexes is evident when the complexes are scanned at faster scan rates (100–800 mV/s, Figures S1–S13), where the rates of reaction are not sufficiently well match to maintain Nernstian equilibrium and ΔE_p increases (e.g. compound **5** in Figure 1A).



Scheme 1. The synthesis of bis(β-diketonato)vanadyl(IV) compounds **1**–**9**, with an image of the crystals obtained before and after the dry-melt method.[30]

When the CV measurements were conducted in other dry solvents, either dry dimethylformamide (DMF; compound **5** in Figure 1B) or dry acetonitrile (MeCN, $VO(acac)_2$ and compound **3** in Figure S6), the reversibility of the vanadyl(IV) is compromised. These CV measurements show the revisbility is solvent dependent, and this could be due to the interactions of the solvent with the metal center. Although all spectra were conducted using analytically pure vanadium sample from a solvent-free dry-melt reaction (Scheme 1); we have previously shown these complexes often crystallize with DMSO solvent in the 6th coordinate position. Figure 2 shows the packing of the complexes, when the 6th coordinate position is occupied by DMSO (**6**) or vacant (**8**) and highlight the interactions between two adjacent molecules. These interactions may have the ability to change the polarity of the V=O bond and alter the redox reversibility.

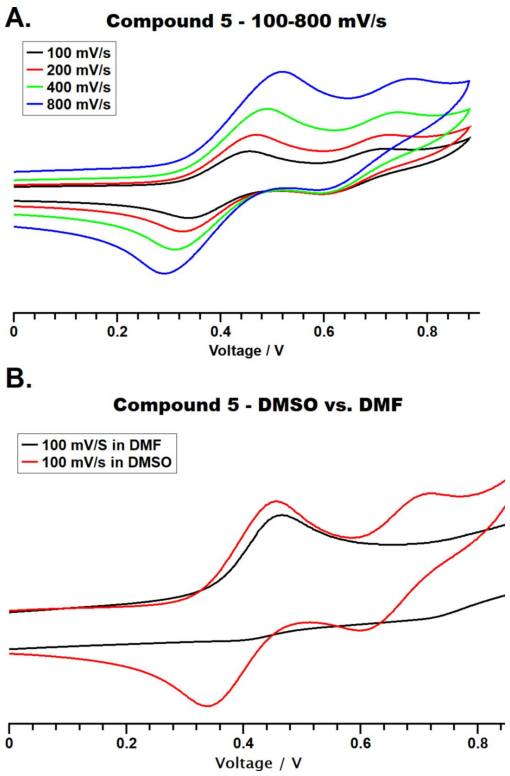


Figure 1. Cyclic voltammograms of compound **5** in A. dry DMSO/0.1 M NBu_4PF_6; scan rate = 100-800 mV/s; B. dry DMSO (red) or dry DMF (black)/ 0.1 m NBu_4PF_6; scan rate = 100 mV/s. All potentials are reported are referenced against ferrocene (Fc/Fc^+ = 0.0 V).

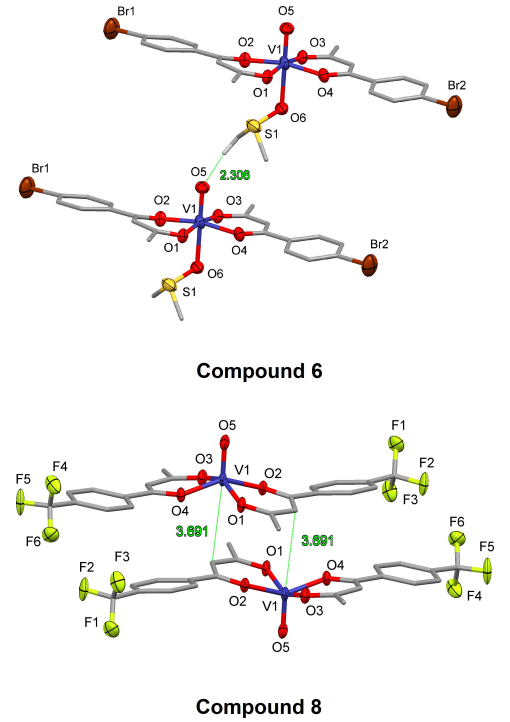


Figure 2. Packing diagram of compound **6** (DMSO-adduct) and compound **8** (adduct-free). Showing the interactions of the vanadyl with neighbouring molecules.[32]

Cytotoxicity Studies

The *in vitro* antiproliferation activity of complexes **1**–**9** was evaluated by the MTT assay at dosages between 0.1–100 μM, and repeated to confirm our previous cytotoxicity studies on these complexes.[32] Screening was conducted against of two human cancer cell lines, lung (A549) and pancreatic (MIA PaCa-2), and the normal retinal epithelium (RPE-1) cell line. The half maximal effective doses (IC_{50}) were calculated after cells were incubated with test complexes between 24 h and 96 h. In most cases, the complexes were either non-toxic or only moderately cytotoxic between 24–48 h and required longer incubation periods to become cytotoxic, therefore, only data for 72 h and 96 h incubation periods are presented in Table S2. A549 cells were selected for further cell-based studies due to their lower IC_{50} values and significant reduction in cell viability, whilst compounds **2**, **3**, **8** and **9** were selected for further *in vitro* analysis, due to their varied cytoxicity values across all cell lines and increased sensitivity towards A549 (Table S2).

Complex-DNA interaction by agarose gel mobility assay

The intercalating, nicking or cleaving capacity of DNA can be determined by gel electrophoresis using plasmid DNA. *In vitro* plasmid DNA assays were conducted by incubating 100 ng/mL plasmid DNA with varying concentrations (6.25–400 μM, lower concentrations are shown in Figure S16) of complexes **2**, **3**, **8** and **9**. The resulting fragments were separated by agarose gel electrophoresis and possible changes in the migration pattern of the DNA were assessed. In accordance with previous reports,[30] cisplatin treatment resulted in faster migration of the plasmid DNA as a result of its DNA-crosslinking activity, and no effect on DNA nicking. On the other hand, all the tested complexes significantly reduced the supercoiled form and exhibited nicking activity as visualized by the increase in closed circular band and decrease in the supercoiled form (Figure 3). These nicks did not appear to lead to double stranded DNA breaks (DSBs), since no linear band was detectable. As sodium azide (NaN_3) can be used to scavenger singlet oxygen ($^1O_2$), it can be used to determine if reactive oxygen species (ROS) were involved in these reactions. The addition of NaN_3 completely rescued and reversed all the effects after incubation with our complexes, even at the highest tested concentration, and this gives strong indication of the involvement of $^1O_2$ radicals (Figure 3, right lanes).

Complex-DNA Interactions

Förster resonance energy transfer (FRET) based DNA melting experiments were used to measure complex-induced stabilization or destabilization of DNA, by comparing the melting temperature of DNA (0.2 μM) in the absence and presence of our complexes. When the melting temperatures of the DNA (T_m) are higher in the presence of a compound, this indicates a stabilization of the DNA.[36,37] The ΔT_m of double strand (DS)

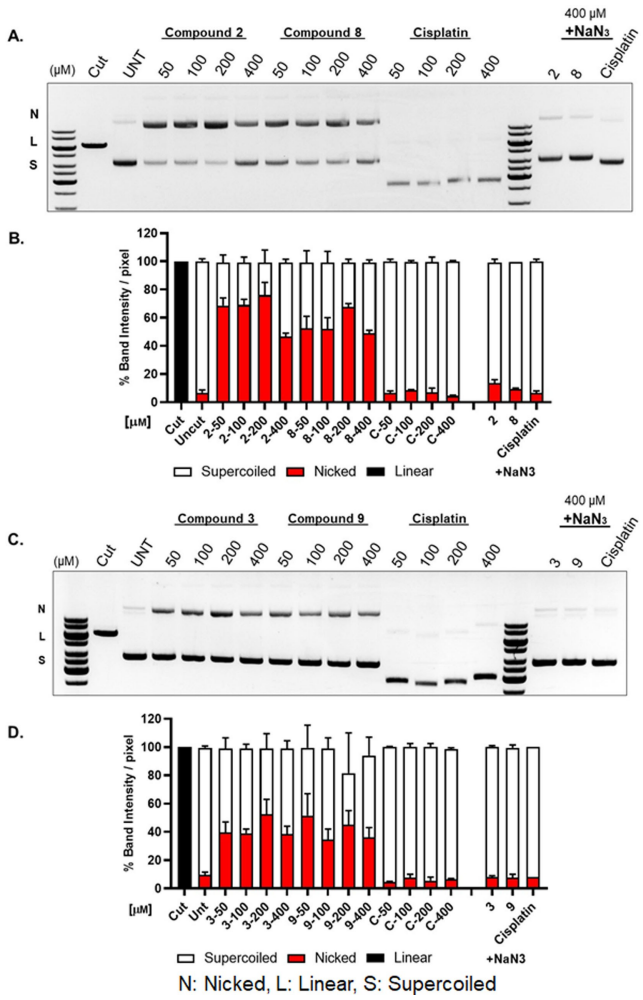


Figure 3. Gel electrophoresis *in vitro* analysis of compounds **2**, **3**, **8** and **9** and plasmid DNA. Plasmid DNA was incubated with different concentrations of compound (50 to 400 μM) for 24 h at room temperature. NaN_3 was used as a ROS scavenger. The intensity of the DNA bands in each lane was quantified and represented as mean ± SD (n = 2) for nicked circular, linear and super coiled. A and C: Representative gel images, B and D: Quantification of gels from A and C.

DNA was measured in the presence of 1 μM (5 eq.) of complexes **1**–**9** (Table 1 and Figures S17–S25). Under these conditions, there was no effect on the stability of DNA for any of the complexes examined. As we observe effects on DNA in our other experiments at a higher concentration of complex, we also performed the FRET melting at 25 μM (Table S3 and Figures S26–S30), consistent with our DNA fragmentation assay. Upon increasing the concentrations of complex **2**, **3**, **8** and **9** to 25 μM, there was no significant change in melting temperature of the DNA (Table 1). This indicates that the complexes do not stabilize or destabilize the DNA under these conditions. Given the results observed using gel electrophoresis we were interested in the relative binding affinity of the complexes to DS DNA. We performed a fluorescence intercalator displacement (FID) assay to determine the displacement of thiazole orange (TO) in the presence of the complexes.[38,39]

We performed the displacement assay in the presence of DS DNA, as well as sequences comprised of $(AT)_6$ and $(GC)_6$. This would reveal not only the affinity for DS DNA but assess whether there was any preference for AT- or GC-rich DNA regions. There was a range of relative binding affinities for the complexes (Table 1). Complexes **1** and **4** had very low relative binding, indicated by low displacement of TO from the DNA. The remainder of the complexes had moderate to high affinity for the DNA. Complex **9** showed the best displacement of DS DNA (27%) at 5 μM complex concentration (5 eq.). Additionally, complexes **5**, **7**, **8** and **9** showed a higher displacement for $(AT)_6$ compared to $(GC)_6$. Out of these complex **9** was found to have the strongest relative binding for $(AT)_6$ with a TO displacement of 47% at 1 μM of complex. However, complex **7** has the highest specificity for $(AT)_6$ over $(GC)_6$ ($(AT)_6$ = 32% vs. $(GC)_6$ = 16%). In line with our other experiments, additional studies were conducted with complexes **2**, **3**, **8** and **9** at the higher concentration of 25 μM (Figure 4), which showed further displacement of TO for complexes **2** and **9**. The relative binding of these complexes is consistent with the results observed in the DNA fragmentation assays. Taken together with the FRET melting data, we can conclude that these complexes bind but do not stabilize DS DNA, and this will contribute to the biological effects observed. Complexes **3**, **8** and **9** all show a preference for AT-rich regions of DNA, which may affect the regions of genomic DNA that they target.

Table 1. FRET melting ΔT_m with DS DNA (0.2 μM) and compounds **1**–**9** (1 μM or 25 μM). FID % TO displacement of 5 μM or 25 μM complex with 1 μM of a DNA buffer: 100 mM potassium chloride and 10 mM sodium cacodylate at pH 7.4. Errors represent the standard deviation from triplicate repeats.

	FRET Melting ΔT_m [°C]		TO Displacement D_{TO} [%]					
	DS		DS		AT6		GC6	
μM	1	25	5	25	1	25	1	25
1	0.0 ± 0.3	–	3 ± 2	–	7 ± 3	–	2 ± 1	–
2	0.0 ± 0.3	0.3 ± 0.2	9 ± 2	59 ± 3	26 ± 4	79 ± 9	20 ± 5	71 ± 4
3	−0.2 ± 0.7	0.7 ± 1.0	10 ± 2	15 ± 2	22 ± 4	34 ± 4	17 ± 4	20 ± 2
4	−0.8 ± 0.4	–	1 ± 2	–	8 ± 3	–	6 ± 1	–
5	−0.6 ± 0.2	–	5 ± 5	–	24 ± 2	–	15 ± 3	–
6	−0.3 ± 0.3	–	9 ± 9	–	12 ± 1	–	12 ± 1	–
7	−0.3 ± 0.3	–	13 ± 2	–	32 ± 1	–	16 ± 2	–
8	−0.4 ± 0.2	0.4 ± 0.3	16 ± 2	17 ± 2	30 ± 2	39 ± 4	21 ± 4	22 ± 3
9	−0.3 ± 0.3	0.0 ± 0.2	27 ± 4	36 ± 3	47 ± 3	66 ± 5	29 ± 2	42 ± 2

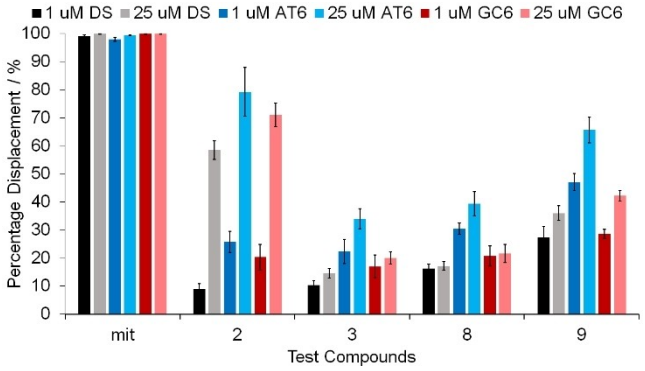


Figure 4. Thiazole orange (TO) displacements (%) for 1 μM (black = DS, dark blue = $(AT)_6$ and dark red = $(GC)_6$) and 25 μM (grey = DS, light blue = $(AT)_6$ and light red = $(GC)_6$) with test compounds: Mitoxantrone (mit, positive control) and compounds **2**, **3**, **8** and **9**. Error bars represent the standard deviation from triplicate repeats.

Reactive Oxygen Species (ROS)

As it is well documented that singlet $^1O_2$ species cause cell death, we hypothesized that the complexes may elevate intracellular ROS in cells, leading to loss of cell viability. To test that, cellular ROS levels were monitored using dihydroethidium, a well described reagent to detect oxidative stress in cell populations. Indeed, all complexes resulted in an increase in ROS, albeit with varying degrees (13–48% as opposed to 4% in untreated controls). In comparison, cisplatin induced ROS was on the lower end with a mild 19% increase (Figure 5A). The flow charts represent the distribution of two populations, M1 (blue) peaks indicate cell population without ROS (ROS−), while M2 (red) is cell population with ROS (ROS+). To further expand on the DNA damage induced in cells in response to our complexes, A549 cells (visualized by DAPI staining, Figure 5B, left panel) were stained for the presence of 8-oxo-Guanidine, which is the most common and abundant type of lesion exerted on the DNA in response to oxidation. As expected, oxidative DNA damage was elevated parallel to the increase in intracellular ROS (Figure 5B, middle panel). Interestingly, cells were also positive for γH2AX (Figure 5B, right panel), which is a widely used marker for DSBs, indicating that the complexes can trigger breaks in the cellular environment.

COMET assay

To further evaluate the genomic damage exerted by complexes **2**, **3**, **8** and **9**, a COMET assay was performed. A549 cells were treated for 48 h with 6.25, 12.5 and 25 μM of the test compounds, which lead to a dose-dependent increase in tail formation (Figure 5C). The damage was most pronounced at 25 μM which was comparable to the positive control (EMS at 12.5 μM, Figure 5D, top panel). Therefore, all compounds induced oxidative DNA stress, DSBs and genomic damage to

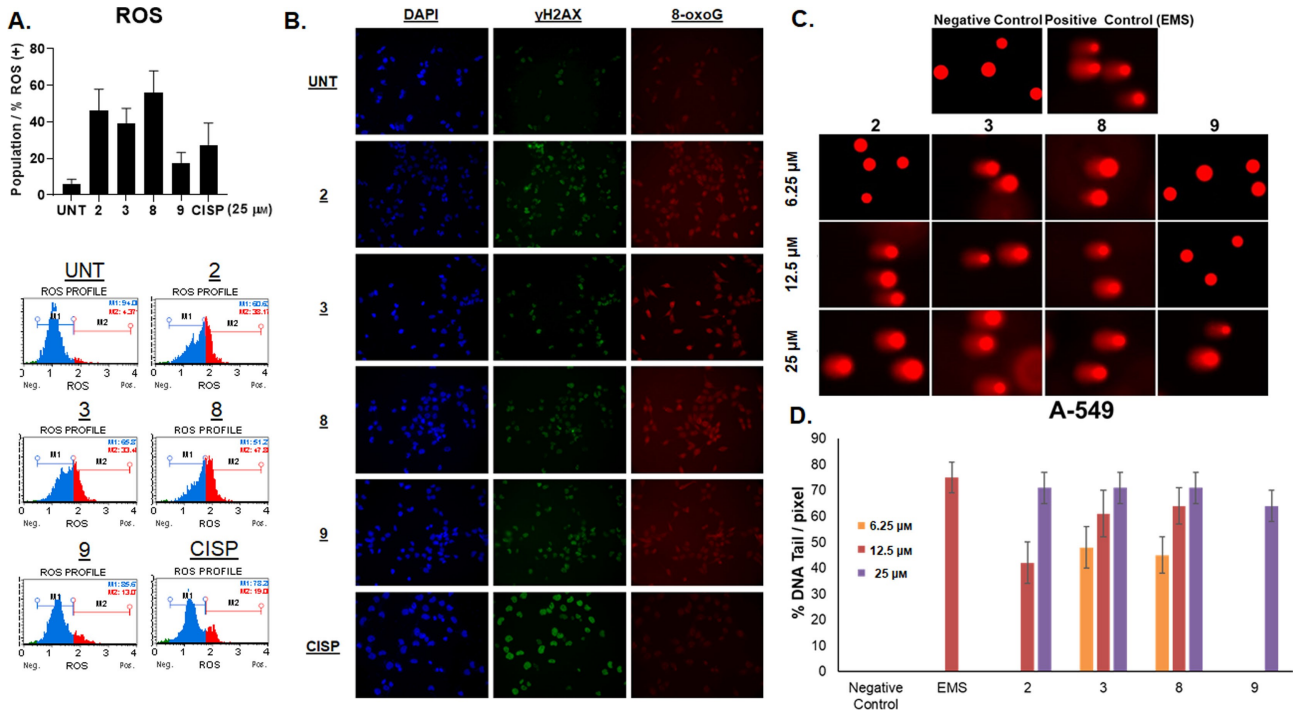


Figure 5. A. Top panel: Increase in the reactive oxygen species (ROS) upon treatment with **2**, **3**, **8**,**9** and cisplatin (25 μM, 48 h). Bottom panel: Representative flow profiles - M1 indicates cell population without ROS (ROS−) while M2 is cell population with ROS (ROS+). B. Microscopic visualization of DNA damage exerted by **2**, **3**, **8**, **9** and cisplatin. DNA was visualized through DAPI staining. 8-oxoG staining was used to determine oxidative DNA lesions, and γH2AX to determine DSBs. C. Images of nuclei from the COMET assay, showing quantitative genomic damage. Representative images of nuclei following treatment with **2**, **3**, **8** and **9** . D. Quantification of tail formation from the COMET assay.

considerable extends, possibly as a result of increased ROS induced by our compounds.

DNA fragmentation and condensation

In order to determine whether our complexes induce programmed cell death pathways, A549 cells were exposed to 12.5–50 μM of complexes **2**, **3**, **8** and **9** for 72 h and genomic DNA was collected. Indeed, each complex induced DNA ladder formation in a dose dependent manner, indicative of DNA fragmentation, which is a hallmark of apoptosis (Figure 6A). To further evaluate this result morphologically, cells were visualized under the microscope for presence of DNA shrinkage and fragmentation, and once again there were several examples of cells exhibiting these changes, supporting apoptosis as the mode of cell death (Figure 6B).

Apoptosis

In order to more quantitatively analyze apoptosis, an Annexin-V assay was used, which is a commonmethod to quantify the number of cells undergoing apoptosis, where flow cytometry is used to separate the cells depending on their uptake of certain dyes. It has previously been reported that when using an Annexin-V assay to assess apoptosis of V^{IV} complexes in A549 and HeLa cell lines, ~2% early stage apoptosis and ~18% late stage apoptosis was observed.[30] When compared to the control, all complexes cause an increase in both early and late apoptosis, however, complex **8** has the highest degree of late apoptosis, which was ~4× greater than cisplatin and 2× greater than recently published V^{IV} complexes (Figure 7A).[30] All complexes exhibit moderate early apoptosis, ranging from 23% (**9**) to 29% (**3**), which are comparable with cisplatin (30%), see Figure S31.

As the executioner Caspases (Caspase-3/7) play a role in the cleaving of proteins, which leads to apoptotic breakdown, the activity of Caspase-3/7 was determined after A549 cells were exposed to 25 μM of complexes **2**, **3**, **8** and **9**. Consistent with previous results,[30] Caspase-3/7 activity was drastically elevated in response to treatment with the complexes compared to the mock treated controls (5.8%) (Figure 7B). Gratifyingly, while cisplatin led to only a mild increase (~22%) in apoptosis, complexes **2**, **3** and **8** induced ~2–3× higher levels (68%, 58% and 68% respectively) and a slight increase by **9** (10%), possibly explaining the higher IC_{50} of these compounds (Figure S31). In summary, our cumulative evidence suggested apoptosis as the main form of cell death in response to complexes **2**, **3**, **8** and **9**, and their structure design warrants further investigation for the treatment of lung carcinomas.

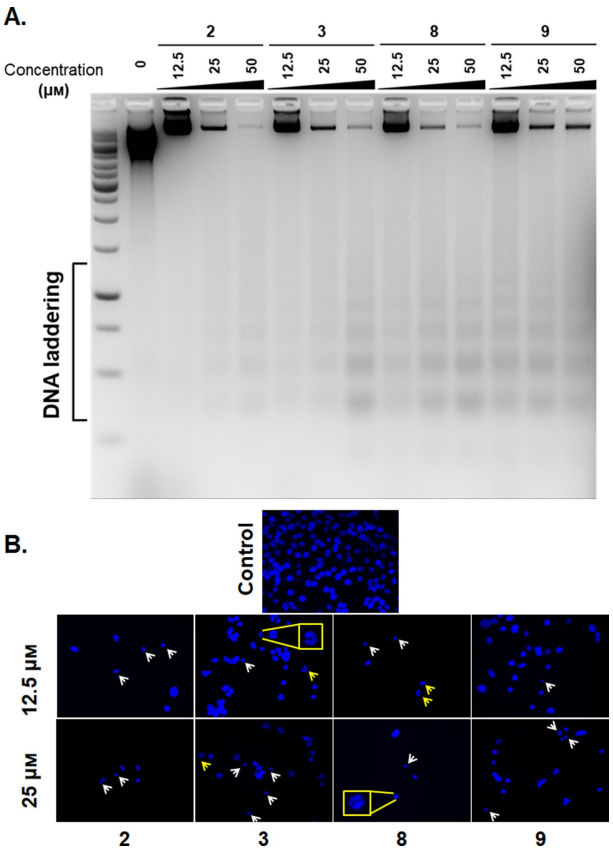


Figure 6. A. DNA fragmentation assay: A549 cells were exposed to indicated doses of **2**, **3**, **8** and **9**, and the genomic DNA was run on agarose gel. DNA laddering occurred in a dose dependent manner. B. Changes in DNA morphology: DNA shrinkage and fragmentation. Cells were exposed to 12.5 and 25 μM of **2**, **3**, **8** and **9** for 72 h and fixed. White arrows indicate DNA shrinkage, yellow arrows indicate DNA fragmentation. Insets show enlarged view of fragmented nuclei.

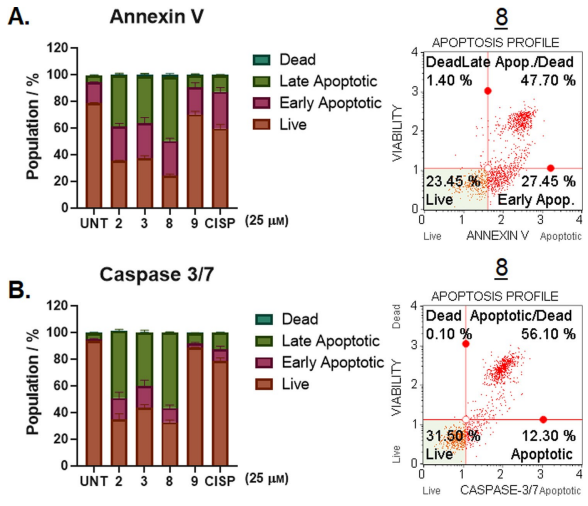


Figure 7. A. Flow cytometric analysis using Annexin-V staining with 7-AAD incorporation; B. Flow cytometric analysis of the activated Caspase-3/7 and simultaneous detection of dead cells by 7-AAD staining. In both cases, A549 cells were exposed to compounds **2**, **3**, **8** and **9** (25 μM) for 48 h. Percentages of cells in quadrant are shown in the right panel for compound **8** only, and given as non-apoptotic live (lower left), non-apoptotic dead (upper left), early apoptotic (lower right), and late-apoptotic (upper right).

Conclusion

We have reported herein, a study of the modes of action of bis (β-diketonato) oxovanadium(IV) complexes (**1**–**9**). Cytotoxicity studies were conducted against several cell lines, and highlight the human lung carcinoma cell line, A549 was sensitive to these compounds. Gel electrophoresis using plasmid DNA was conducted, and the tested complexes exhibit significant reductions in the supercoiled form and display DNA nicking activity. However, these nicks do not suggest double stranded DNA breaks (DSBs). The addition of NaN_3 (single oxygen scavenger), completely rescued and reversed all the effects after incubation with our compounds and gives a strong indication for the involvement of $^1O_2$ radicals.

Förster resonance energy transfer (FRET) based melting experiments show no significant changes in the DS DNA after incubation with 25 μM (125 eq.) of these bis(β-diketonato) oxovanadium(IV) complexes. Since the gel electrophoresis results highlighted DNA interactions, the potential binding affinities of the complexes to DS DNA were measured using a fluorescence intercalator displacement (FID) assay. At 5 μM complex concentrations (5 eq.), complex **9** shows the best displacement of DS DNA (27%), whilst an increase to 25 μM highlighted complex **2** with the best displacement (59%). Studies were conducted to assess the complexes preferentially binding to either AT- or GC-rich DNA, and complexes **3**, **8** and **9** all show a preference for AT-rich regions of DNA.

A549 cells were incubated with complexes **2**, **3**, **8** and **9**, and the presence of 8-oxo-Guanidine was measured,showing oxidative DNA damage was elevated parallel to the increase in intracellular ROS. Interestingly, cells were also positive for γH2AX, indicating that the complexes can trigger DNA breaks in the cellular environment. These compounds were also shown to exhibit dose-dependent increase in tail formation in genomic DNA when using the COMET assay, showing increases in DS DNA damage. This induction in oxidative DNA stress, DSBs and genomic damage, is possibly as a result of increased ROS induction. Finally, we have shown these compound have induced DNA ladder formation in a dose dependent manner and DNA shrinkage and fragmentation, which all support the Annexin-V and Caspase-3/7 results, and highlight increased early and late apoptosis and elevation of Caspase-3/7; which are all suggestive that apoptosis is the main form of cell death in A549 cells treated with these bis(β-diketonato) oxovanadium (IV) complexes.

Experimental Section

General: All ligands and complexes were synthesized under aerobic conditions. All chemicals were supplied by Sigma-Aldrich Chemical Co. Deuterated NMR solvents were purchased from Sigma-Aldrich Chemical Co. or Acros Organics. All ligands and complexes were prepared using a previously reported literature methods.[32]

Cyclic Voltammetry: Cyclic voltammetric measurements were carried out using Autolab PGStat 30 potentiostat/galvanostat. A single-compartment or a conventional three-electrode cell was used with a silver/silver chloride reference electrode (3 M NaCl, saturated AgCl), glassy carbon working electrode and Pt wire auxiliary electrode. Dimethylsulfoxide, DMSO or dimethylformamide, DMF was stored and used over molecular sieves, whilst acetonitrile (MeCN) was freshly distilled from CaH_2. Tetrabutylammonium hexafluorophosphate $[N(C_4H_9\text{-}n)_4][PF_6]$ was used as the supporting electrolyte. Solutions containing ca. 10 mM analyte (0.1 M electrolyte) were degassed by purging with argon, and spectra were collected with a constant flow of argon. All spectra were referenced to ferrocene/ferrocenium.

Cell Culture: A549 (ATCC, CCL-185), MIA PaCa-2 (ATCC, CRL-1420), and RPE-1 cells (ATCC, CRL-4000)) were all grown in Dulbecco's Modified Eagle Medium F-12 (Gibco #11320033) containing 10% FBS (Gibco, 10500064) and Penicillin-Streptomycin (10,000 U/mL) (Gibco, 15140122) in a 37 °C, 5% CO_2 incubator.

Cell viability: Stock solutions of all complexes were prepared in DMSO (10 mM) and aliquots were stored at −20 °C. 8×10^3 (MIA PaCa-2), 4×10^3 (A549) and 2×10^3 (RPE-1) cells were seeded on 96-well plates and incubated with serial dilutions of the complexes (freshly prepared, 0–100 μM) for 72 and 96 h (or 24 and 48 h, see *Supplementary Information*). Cell viability was measured via MTT assay. Briefly, 20 μL of MTT reagent (Sigma Aldrich, M-5655) was added on the cells (5 μg/mL, at 37 °C). After 4 h, formazan crystals were solubilized in 10% SDS, 0.01 M HCl (50 μL, overnight at 37 °C), and color formation was detected with Bio-Tek H1 Synergy microplate reader.

Agarose gel electrophoresis for complex/DNA interactions: The complex-DNA interactions were evaluated *in vitro* as described previously with minor changes.[40] Briefly, plasmid DNA (200 ng, pBOS-H2BGFP, BD Biosciences) was incubated with **2**, **3**, **8**, **9** or cisplatin for 24 h (50–400 μM, in a total reaction volume of 20 μL, (RT). NaN_3 (37.5 mM in ddH_2O, final concentration) was used for ROS scavenging activity. Plasmid DNA was linearized via digestion with BamHI (Thermo Scientific, ER0051), and samples were run on 1% agarose gel (100 V, 60 min). The experiment was performed with two biological repeats.

γH2AX and 8-oxoG staining: A549 cells were treated with 25 μM of **2**, **3**, **8**, **9** and cisplatin for 24 h and fixed in freshly prepared 4% paraformaldehyde (15 min at room temperature (RT)), permeabilized in 0.1% Triton X-100/PBS (1 h, RT), blocked in 0.2% gelatin (1 h, RT) and stained with γH2AX (Cell Signaling, #9718S, 1:400) or 8-oxo-Guanine (EMD Millipore, MAB3560, 1:100) antibodies overnight at 4 °C. Next day, cells were stained with Goat Anti-Rabbit IgG H&L (Alexa Fluor® 488) (ab150077) and Goat Anti-Mouse IgG H&L (Alexa Fluor® 594) secondary antibodies for 1 h, at RT. The cover slips were mounted on slides with VECTASHIELD Antifade Mounting Medium with DAPI (Vector Laboratories, H-1200) and imaged with Zeiss Axio Imaging M1 fluorescence microscopy under 40x magnification.

Detection of Reactive Oxygen Species: 2×10^5 A549 cells were seeded onto 6-well plates and treated with 25 μM of **2**, **3**, **8** and **9** for 48 h. Cells were trypsinized and centrifuged (300 g, 5 min). The pellet was resuspended with 1×Assay Buffer (MCH100111-2). Oxidative Stress Reagent (4700-1665) was diluted with 1×Assay Buffer in a 1:100 ratio to prepare intermediate solution. This solution was further diluted with 1×Assay Buffer in 1:80 ratio to prepare a working solution. 50 μL cell suspension was mixed with 150 μL working solution and incubated at 37 °C for 30 min. ROS activated cells were counted with Muse Cell Analyzer (Merck Millipore).

COMET assay: Genotoxicity of the complexes was evaluated using previously published protocols[41] with slight modifications. Briefly, 4×10^4 A549 cells were seeded on 24-well plates and incubated for 24 h (37 °C, 5% CO_2). Serial dilutions of **2**, **3**, **8**, or **9** (freshly prepared

at 6.25, 12.5, and 25 μM concentrations) were added on the cells and incubated for 48 h. DMSO (Sigma, D2650) was used as negative control, while 'Ethyl methanesulfonate' (EMS – Merck-Millipore #8.20774) served as positive control (40 mM, 1 h) under the same experimental conditions. At the end of incubation, cells were counted using a haemocytometer, and the resuspended in PBS to a final concentration of 1.6×10^4 cells/mL. Cells were then mixed with low melting agarose, spread over the slides, and subjected to lysis buffer (at dark, +4 °C, 24 h). The slides were washed with an alkaline carrier buffer (pH ≥ 13, 20 min), electrophoresed (13 V, 0.03 mA, 25 min), and stained with propidium iodide (10 μg/mL, #P4864 Sigma-Aldrich, 20 min). They were visualized by a fluorescence microscope (535 nm/617 nm wavelength, 40X magnification) and quantified using ImageJ software. All experiments were done in two biological repeats, and 50 individual cells/experiment were quantified.

Annexin-V staining and determination of Caspase-3/7 activity: 2×10^5 A549 cells were seeded onto 6-well plates and treated with 25 μM of **2**, **3**, **8** and **9** for 48 h. Cells were washed in PBS, trypsinized and analyzed using Muse Cell analyzer (Millipore). Annexin V/Dead Cell (Luminex, MCH100105) and Caspase-3/7 (Luminex, MCH100108) staining were performed using following manufacturer's instructions with the changes as previous reported.[29]

DNA Fragmentation assay: A modified protocol[40] was used for DNA fragmentation assay. Briefly, 2.5×10^5 cells were seeded onto 25 cm^3 flasks and treated with **2**, **3**, **8**, **9** at indicated doses for 72 h. Both floating and attached cells were collected and combined. The cells were washed with PBS and centrifuged (1000 g, 5 min, RT). The pellet was dissolved in 120 μL of lysis buffer [10 mmol/L Tris (pH 7.4), 100 nmol/L NaCl, 25 mmol/L EDTA, 1% N-lauryl sarcosine, and proteinase K (final concentration: 0.35 mg/mL)] by gently vortexing, and was incubated at 45 °C for 2 h. The lysates were further incubated for 1 h at RT following addition of 2 μL of RNAse A (10 mg/mL). The samples were resolved on 2% agarose gels (stained with Ethidium Bromide, 60 V, 5 h) and analyzed using a Bio-Rad Gel Imaging System.

Oligonucleotides: Oligonucleotides were purchased from Eurogentec and purified using reverse phase HPLC. The dry DNA was dissolved in ultrapure (100 μM for labelled oligos, 1 mM for unlabelled oligos) final concentrations were determined using a NanoDrop; further dilutions were carried out in 10 mM sodium cacodylate supplemented with 100 mM potassium chloride, pH 7.4. Samples were thermally annealed in a heat block at 95 °C for 5 min and cooled slowly to RT overnight.

FRET Based DNA Melting Experiments: Assessment of the ligand-induced change in melting temperature was performed using a fluorescence resonance energy transfer (FRET) DNA melting based assay.[34] The labelled oligonucleotide with a donor fluorophore FAM (6-carboxyfluorescein) and acceptor fluorophore TAMRA (6-carboxytetramethyl-rhodamine). DS_{FRET} FAM-d(TAT-AGC-TAT-A-HEG (18)-TAT-AGC-TAT-A)-TAMRA-3′). Strip-tubes (QIAgen) were prepared by combining 20 μL of 0.2 μM DNA with the respective complex. Control samples for each run were prepared with the same quantity of solvent with the DNA in buffer. Fluorescence melting curves were determined in a QIAgen Rotor-Gene Q-series PCR machine, using a total volume of 20 μL. Samples were held at 25 °C for 5 min then ramped to 95 °C, at increments of 1 °C, holding the temperature at each step for 1 min. Measurements were made with excitation at 470 nm and detection at 510 nm. Final analysis of the data was carried out using QIAgen Rotor-Gene Q-series software and Origin or Excel. T_m values were determined using the first derivative of the melting curves.

Fluorescence Intercalator Displacement (FID): The FID assay was performed on a BMG CLARIOstar plate reader using an excitation of 430 nm and emission was measured from 450 to 650 nm with the emission at 450 nm being normalised to 0%. 96-well plates (Corning 96 well solid black flat bottom plates) were used for this assay. 90 μL of thiazole orange (TO) at a concentration of 2 μL in 10 mM sodium cacodylate and 100 mM potassium chloride that was pH corrected to pH 7.4) was added to each well. The fluorescence was then measured at 450 nm with an excitation of 430 nm and normalised to 0%. 1 μM of DNA was added, shaken at 700 rpm in the plate reader for 30 s and left to equilibrate for 20 min. After equilibration the fluorescence was measured again and normalised to 100%. After that, additions in to each well (in triplicate) the complexes were added at the stated concentrations. The fluorescence was measured after each addition and normalised between the 0 and 100% levels previously determined. The percentage displacement of TO value (D_{TO}) was calculated from the displacement of TO after the addition of complex. DS = 5′-d[GGC-ATA-GTG-CGT-GGG-CGT-TAG-C]-3′ + complementary sequence. $(AT)_6$ = 5′-d[ATATATATATAT]-3′. $(GC)_6$ = 5′-d[GCGCGCGCGCGC]-3′ The sequences $(AT)_6$ and $(GC)_6$ are self-complementary so were annealed at 2 μM to give 1 μM of duplex overall.

Acknowledgements

This work was supported by the UKRI Future Leaders Fellowship which was awarded to RML [MR/T041315/1]. The authors would like to thank Dr. John Fielden and Dr. Kevin Vincent (UEA) for use of the CV equipment and discussion of the results. Biological studies were supported by Koç University School of Medicine (KUSOM) and the authors gratefully acknowledge use of the services and facilities of the Koç University Research Center for Translational Medicine (KUTTAM), funded by the Presidency of Turkey, Presidency of Strategy and Budget. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Presidency of Strategy and Budget.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: Apoptosis · β-diketonate complexes · Bioinorganic Chemistry · DNA interactions · Vanadium complexes

[1] O. H. Warburg, F. Dickens, *The Metabolism of Tumours*, Constable, London, **1930**.
[2] J. Kieler, A. Gromek, N. I. Nissen, *Acta Chir. Scand. Suppl.* **1965**, *343*, 154–164.
[3] E. Sabbioni, J. Rade, F. Bertolero, *J. Inorg. Biochem.* **1980**, *12*, 307–315.
[4] E. Sabbioni, G. Pozzi, A. Pintar, L. Casella, S. Garattini, *Carcinogenesis* **1991**, *12*, 47–52.
[5] E. Sabbioni, G. Pozzi, S. Devos, A. Pintar, L. Casella, M. Fischbach, *Carcinogenesis* **1993**, *14*, 2565–2568.
[6] A. M. Cortizo, V. C. Sálice, C. M. Vescina, S. B. Etcheverry, *BioMetals* **1997**, *10*, 127–133.
[7] H. Vilter, *Phytochemistry* **1984**, *23*, 1387–1390.
[8] Alison. Butler, J. V. Walker, *Chem. Rev.* **1993**, *93*, 1937–1944.
[9] A. Butler, M. J. Clague, G. E. Meister, *Chem. Rev.* **1994**, *94*, 625–638.
[10] Y. Shechter, S. J. D. Karlish, *Nature* **1980**, *284*, 556–558.
[11] C. E. Heyliger, A. G. Tahiliani, J. H. McNeill, *Science* **1985**, *227*, 1474–1477.

[12] D. C. Crans, M. Mahroof-Tahir, A. D. Keramidas, *Mol. Cell. Biochem.* **1995**, *153*, 17–24.
[13] D. C. Crans, J. J. Smee, E. Gaidamauskas, L. Yang, *Chem. Rev.* **2004**, *104*, 849–902.
[14] S. S. Amin, K. Cryer, B. Zhang, S. K. Dutta, S. S. Eaton, O. P. Anderson, S. M. Miller, B. A. Reul, S. M. Brichard, D. C. Crans, *Inorg. Chem.* **2000**, *39*, 406–416.
[15] D. C. Crans, A. M. Trujillo, P. S. Pharazyn, M. D. Cohen, *Coord. Chem. Rev.* **2011**, *255*, 2178–2192.
[16] D. C. Crans, J. T. Koehn, S. M. Petry, C. M. Glover, A. Wijetunga, R. Kaur, A. Levina, P. A. Lay, *Dalton Trans.* **2019**, *48*, 6383–6395.
[17] M. Melchior, S. J. Rettig, B. D. Liboiron, K. H. Thompson, V. G. Yuen, J. H. McNeill, C. Orvig, *Inorg. Chem.* **2001**, *40*, 4686–4690.
[18] K. H. Thompson, J. H. McNeill, C. Orvig, *Chem. Rev.* **1999**, *99*, 2561–2572.
[19] K. H. Thompson, J. Lichter, C. LeBel, M. C. Scaife, J. H. McNeill, C. Orvig, *J. Inorg. Biochem.* **2009**, *103*, 554–558.
[20] K. H. Thompson, C. Orvig, *Coord. Chem. Rev.* **2001**, *219–221*, 1033–1053.
[21] D. C. Crans, *Comments Inorg. Chem.* **1994**, *16*, 35–76.
[22] S. B. Etcheverry, A. M. Cortizo, in *Vanadium Compounds*, American Chemical Society, **1998**, pp. 270–276.
[23] J. B. Smith, *PNAS* **1983**, *80*, 6162–6166.
[24] K. G. Mountjoy, J. S. Flier, *Endocrinology* **1990**, *127*, 2025–204.
[25] K.-H. W. Lau, H. Tanimoto, D. J. Baylink, *Endocrinology* **1988**, *123*, 2858–2867.
[26] O. J. D'Cruz, F. M. Uckun, *Expert Opin Investig. Drugs* **2002**, *11*, 1829–1836.
[27] A. M. Evangelou, *Crit. Rev. Oncol. Hematol.* **2002**, *42*, 249–265.
[28] E. Kioseoglou, S. Petanidis, C. Gabriel, A. Salifoglou, *Coord. Chem. Rev.* **2015**, *301–302*, 87–105.
[29] L. Ni, H. Zhao, L. Tao, X. Li, Z. Zhou, Y. Sun, C. Chen, D. Wei, Y. Liu, G. Diao, *Dalton Trans.* **2018**, *47*, 10035–10045.
[30] N. Ribeiro, I. Bulut, B. Cevatemre, C. Teixeira, Y. Yildizhan, V. André, P. Adão, J. Costa Pessoa, C. Acilan, I. Correia, *Dalton Trans.* **2021**, *50*, 157–169.
[31] K. Dankhoff, A. Ahmad, B. Weber, B. Biersack, R. Schobert, *J. Inorg. Biochem.* **2019**, *194*, 1–6.
[32] M. Zegke, H. L. M. Spencer, R. Lord, *Chem. Eur. J.* **2019**, *25*, 12275–12280.
[33] M. Mohamadi, S. Yousef Ebrahimipour, M. Torkzadeh-Mahani, S. Foro, A. Akbari, *RSC Adv.* **2015**, *5*, 101063–101075.
[34] P. I. da S. Maia, V. M. Deflon, E. J. de. Souza, E. Garcia, G. F. de. Sousa, A. A. Batista, A. T. de. Figueiredo, E. Niquet, *Transition Met. Chem.* **2005**, *30*, 404–410.
[35] B. Kaboudin, K. Moradi, E. Safaei, H. Dehghan, P. Salehi, *Phosphorus, Sulfur, and Silicon and the Related Elements* **2012**, *187*, 1521–1527.
[36] E. Wright, H. Day, A. Ibrahim, J. Kumar, L. J. E. Boswell, C. Huguin, C. E. M. Stevenson, K. Pors, Z. A. E. Waller, *Sci. Rep.* **2016**, *6*, Article number: 39456.
[37] M. R. M. Williams, B. Bertrand, D. L. Hughes, Z. A. E. Waller, C. Schmidt, I. Ott, M. O'Connell, M. Searcey, M. Bochmann, *Metallomics* **2018**, *10*, 1655–1666.
[38] Q. Sheng, J. C. Neaverson, T. Mahmoud, C. E. M. Stevenson, S. E. Matthews, Z. A. E. Waller, *Org. Biomol. Chem.* **2017**, *15*, 5669–5673.
[39] P. Spence, J. Fielden, Zoë. A. E. Waller, *J. Am. Chem. Soc.* **2020**, *142*, 13856–13866.
[40] C. P. Matos, Y. Yildizhan, Z. Adiguzel, F. R. Pavan, D. L. Campos, J. C. Pessoa, L. P. Ferreira, A. I. Tomaz, I. Correia, C. Acilan, *Dalton Trans.* **2019**, *48*, 8702–8716.
[41] C. P. Matos, Z. Adiguzel, Y. Yildizhan, B. Cevatemre, T. B. Onder, O. Cevik, P. Nunes, L. P. Ferreira, M. D. Carvalho, D. L. Campos, F. R. Pavan, J. C. Pessoa, M. H. Garcia, A. I. Tomaz, I. Correia, C. Acilan, *Eur. J. Med. Chem.* **2019**, *176*, 492–512.
[42] O. Kacar, Z. Adiguzel, V. T. Yilmaz, Y. Cetin, B. Cevatemre, N. Arda, A. T. Baykal, E. Ulukaya, C. Acilan, *Anti-Cancer Drugs* **2014**, *25*, 17–29.

Manuscript received: March 2, 2021
Revised manuscript received: April 13, 2021
Accepted manuscript online: April 15, 2021
Version of record online: May 26, 2021